اندازه گیری، سنجش و ارزشیابی آموزشی

ویرایش ششم

دکتر علی‌اکبر سیف



۳۷۱/۲۶۴
الف ۹ س
۱۵

سرشناسه : سیف، علی‌اکبر، ۱۳۲۰ -
عنوان و نام پدیدآور : اندازه‌گیری، سنجش، و ارزشیابی آموزشی / تألیف علی‌اکبر سیف.
وضعیت ویراست : [ویراست ۶].
مشخصات نشر : تهران: نشر دوران، ۱۳۹۰.
مشخصات ظاهری : ۶۹۴ ص.: مصور، جدول، نمودار.
شابک : 978-600-6208-11-4
یادداشت : کتاب حاضر در ویراست قبلی با عنوان «روشهای اندازه‌گیری و ارزشیابی آموزشی» نیز منتشر شده است.
یادداشت : واژه‌نامه.
یادداشت : کتابنامه: ص. ۶۴۵-۶۴۷.
یادداشت : نمایه.
عنوان دیگر : روشهای اندازه‌گیری و ارزشیابی آموزشی.
موضوع : موفقیت تحصیلی ــ آزمون‌ها
موضوع : شاگردان ــ ارزشیابی
موضوع : آموزش و پرورش بر مبنای توانش ــ آزمون‌ها
رده‌بندی کنگره : ۱۳۹۰ ۸الف۹س/LC۱۰۳۴
رده‌بندی دیویی : ۳۷۱/۲۶۴
شماره کتابشناسی ملی: ۲۵۸۲۸۹۸

۱۲۲۳۹۷
۹۱/۷/۲

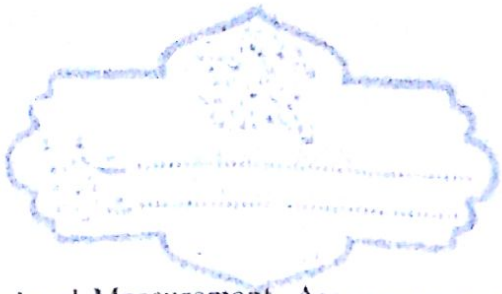
اندازه گیری، سنجش، و ارزشیابی آموزشی

ویرایش ششم

تألیف
دکتر علی‌اکبر سیف
استاد روانشناسی تربیتی
دانشگاه علامهٔ طباطبائی

ویژه کتاب - کتابخانه مرکزی شهرداری اصفهان

نشر دوران



Educational Measurement, Assessment, and Evaluation
Sixth Edition
Ali-Akbar Saif, Ph.D.
Professor of Educational Psychology
Alameh Tabatabaie University

Editions: 1996, 1998, 2003, 2006, 2008
Sixth Edition, First Printing, Winter, 2012



نشر دیدآوران

www.didavarbook.com
info@didavarbook.com

اندازه گیری، سنجش، و ارزشیابی آموزشی

دکتر علی‌اکبر سیف

ویرایش اول: پاییز ۱۳۷۵
ویرایش دوم: زمستان ۱۳۷۶
ویرایش سوم: پاییز ۱۳۸۲
ویرایش چهارم: تابستان ۱۳۸۵
ویرایش پنجم: پاییز ۱۳۸۷
چاپ اول ویرایش ششم، زمستان ۱۳۹۰
(چاپ مکرر سی و پنجم)
حروفنگاری: علم روز، لیتوگرافی: طیف‌نگار
چاپ منصور، صحافی ایران‌کتاب
تیراژ: ۵۱۵۰ نسخه

حق هرگونه چاپ و انتشار محفوظ است.

مرکز پخش:

کتاب دیدآور

۱) خیابان انقلاب، خیابان فخررازی، نرسیده به لبافی‌نژاد، کوچهٔ الوندی، شمارهٔ ۲، واحد ۴
تلفن ۶۶۹۶۱۶۴۱ - ۶۶۹۶۱۶۲۲ - تلفکس ۶۶۴۱۴۴۷۸

۲) خیابان انقلاب، خیابان فخررازی، خیابان روانمهر، پلاک ۸۴ ساختمان دیدآور
تلفن ۶۶۹۶۱۷۴۸

قیمت: ۱۶۵۰۰۰ ریال

پیشکش به پسرم:
امیرعزیز سیف

فهرست محتوایی

پیشگفتار ویرایش ششم ۲۱
پیشگفتار ویرایش اول ۲۳
راهنمای مطالعه و یادگیری مطالب کتاب ۲۵
هدفهای کلی کتاب ۲۷
فصل اول: تعریف اصطلاحات ۲۹
تعریف اندازه‌گیری ۳۰
تعریف آزمون ۳۱
تعریف آزمودن ۳۳
تعریف سنجش ۳۴
تعریف ارزشیابی ۳۵
مقایسهٔ اندازه‌گیری، آزمودن، سنجش، و ارزشیابی با یکدیگر ۳۷
ارزشیابی آموزشی و پژوهش آموزشی ۳۹
خلاصهٔ فصل ۴۰
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۴۱
فصل دوم: مقیاسهای اندازه‌گیری ۴۳
تعریف و انواع مقیاسهای اندازه‌گیری ۴۴
مقیاس اسمی ۴۴
مقیاس ترتیبی ۴۵
مقیاس فاصله‌ای ۴۷
مقایسهٔ مقیاس ترتیبی با مقیاس فاصله‌ای ۴۷
مقیاس نسبتی ۴۸
مقایسهٔ مقیاسهای اندازه‌گیری با یکدیگر ۴۹
مقیاس پیوسته و مقیاس گسسته ۵۱
مقیاس گسسته ۵۱

مقیاس پیوسته ۵۲
حدود واقعی اعداد ۵۳
گِرد کردن ۵۳
خلاصهٔ فصل ۵۴
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۵۵

فصل سوم: **دسته‌بندی آزمونهای مورد استفاده در روانشناسی و آموزش وپرورش** ۵۷
دسته‌بندی آزمونها با توجه به ویژگیهای مورد اندازه‌گیری ۵۸
آزمونهای سنجش توانایی ۵۸
پرسشنامه‌های سنجش عاطفی ۶۰
پرسشنامه‌های سنجش شخصیت ۶۱
ابزارهای سنجش علاقه و نگرش ۶۲
دسته‌بندی آزمونها با توجه به چگونگی تهیهٔ آنها ۶۳
ابزارهای اندازه‌گیری میزان‌شده ۶۳
آزمونهای معلم‌ساخته ۶۳
مقایسهٔ آزمونهای میزان شده با آزمونهای معلم‌ساخته ۶۳
استفاده از آزمونهای میزان شده و معلم ساخته در تصمیم‌گیریهای آموزشی ۶۴
دسته‌بندی روشهای سنجش با توجه به نوع سؤال یا ماده ۶۵
خلاصهٔ فصل ۶۸
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۶۸

فصل چهارم: **دسته‌بندی ارزشیابیهای آموزشی** ۷۱
دسته‌بندی ارزشیابیهای آموزشی در قالب رویکردهای مختلف ۷۲
دسته‌بندی ارزشیابیهای آموزشی در قالب الگوهای مختلف ۷۵
الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر هدف ۷۵
الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر مدیریت ۷۷
الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر مصرف‌کننده ۷۹
الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر نظر متخصصان ۸۱
الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر اختلاف نظرهای

متخصصان ۸۳
الگوهای ارزشیابی برخاسته از رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان ۸۳
دسته‌بندی ارزشیابیهای آموزشی با توجه به موضوع ارزشیابی ۸۹
ارزشیابی از دانشجویان و دانش‌آموزان ۸۹
ارزشیابی از برنامه‌های درسی و مواد آموزشی ۸۹
ارزشیابی از پروژه‌ها و برنامه‌های آموزشی ۹۰
ارزشیابی از آموزشگاه ۹۱
ارزشیابی از کارکنان آموزشگاه ۹۱
دسته‌بندی ارزشیابیهای آموزشی با توجه به ملاک مورد استفاده ۹۱
ارزشیابی وابسته به ملاک ۹۱
ارزشیابی وابسته به هنجار ۹۳
مقایسهٔ ارزشیابی وابسته به ملاک با ارزشیابی وابسته به هنجار ۹۴
دسته‌بندی ارزشیابیهای آموزشی با توجه به زمان و هدف استفاده از آنها ۹۶
ارزشیابی آغازین یا سنجش آغازین ۹۶
ارزشیابی تکوینی ۹۷
ارزشیابی تشخیصی ۹۹
ارزشیابی تراکمی ۱۰۱
دسته‌بندی ارزشیابیهای آموزشی با توجه به ارزشیابان ۱۰۳
خلاصهٔ فصل ۱۰۶
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۱۰۷

فصل پنجم: مراحل و فعالیتهای ارزشیابی آموزشی ۱۱۱
مراحل ارزشیابی آموزشی ۱۱۳
مرحلهٔ طراحی ۱۱۳
تحلیل موقعیت ۱۱۳
تعیین و توصیف هدفها ۱۱۴
توصیف پیش‌نیازها یا رفتارهای ورودی ۱۱۵
انتخاب و تولید وسایل اندازه‌گیری ۱۱۶
توصیف راهبردها یا استراتژیها ۱۱۶

انتخاب طرح پژوهشی ۱۱۷
تدارک یک برنامهٔ زمان‌بندی ۱۱۸
مرحلهٔ فرآیندی یا اجرایی ۱۱۹
مرحلهٔ فرآورده‌ای ۱۲۰
فعالیتهای ارزشیابی آموزشی ۱۲۲
خلاصهٔ فصل ۱۲۳
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۱۲۴

فصل ششم: تهیهٔ طرح ارزشیابی پیشرفت تحصیلی ۱۲۷
تعریف ارزشیابی پیشرفت تحصیلی ۱۲۸
هدفهای آموزشی یا هدفهای یادگیری ۱۲۸
استانداردهای تحصیلی ۱۲۹
مقایسهٔ هدف با استاندارد ۱۳۰
طبقه‌بندی هدفهای آموزشی ۱۳۰
حوزهٔ شناختی ۱۳۲
طبقه‌بندی تجدیدنظر شدهٔ حوزهٔ شناختی ۱۳۳
بُعد دانش ۱۳۴
بُعد فرایند شناختی ۱۳۵
حوزهٔ عاطفی ۱۳۹
حوزهٔ روانی ـ حرکتی ۱۴۰
جدول مشخصات ۱۴۱
نقش جدول مشخصات آزمون ۱۴۲
چگونگی توزیع سؤالها به بخشهای مختلف ۱۴۳
انواع جدولهای مشخصات ۱۴۵
انتخاب هدفهای مناسب برای گنجانیدن در آزمون ۱۴۹
خلاصهٔ فصل ۱۴۹
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۱۵۰

فصل هفتم: سنجش یادگیری به کمک آزمون صحیح ـ غلط ۱۵۳
تعریف و انواع آزمونهای صحیح ـ غلط ۱۵۴
تهیهٔ سؤالهای صحیح ـ غلط ۱۵۶
قواعد تهیهٔ سؤالهای صحیح ـ غلط ۱۵۹

	امتیازها و محدودیتهای آزمونهای صحیح ـ غلط	۱۶۴
	موارد استفادهٔ آزمونهای صحیح ـ غلط	۱۶۷
	خلاصهٔ فصل	۱۷۰
	سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل	۱۷۱
فصل هشتم:	**سنجش یادگیری به کمک آزمون جور کردنی**	۱۷۳
	تعریف و انواع آزمونهای جور کردنی	۱۷۴
	قواعد تهیهٔ سؤالهای جور کردنی	۱۷۶
	امتیازها و محدودیتهای آزمونهای جور کردنی	۱۸۲
	موارد استفادهٔ آزمونهای جور کردنی	۱۸۳
	خلاصهٔ فصل	۱۸۹
	سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل	۱۹۰
فصل نهم:	**سنجش یادگیری به کمک آزمون چند گزینه‌ای**	۱۹۱
	تعریف و انواع آزمونهای چند گزینه‌ای	۱۹۲
	قواعد تهیهٔ سؤالهای چند گزینه‌ای	۱۹۶
	آزمون چندگزینه‌ای رایانه‌ای	۲۱۳
	امتیازها و محدودیتهای آزمونهای چند گزینه‌ای	۲۱۳
	موارد استفادهٔ آزمونهای چند گزینه‌ای	۲۱۴
	خلاصهٔ فصل	۲۲۵
	سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل	۲۲۵
فصل دهم:	**سنجش یادگیری به کمک آزمون تشریحی**	۲۲۷
	تعریف و انواع آزمونهای تشریحی	۲۲۸
	قواعد تهیهٔ سؤالهای تشریحی	۲۳۰
	قواعد تصحیح پاسخهای سؤالهای تشریحی	۲۳۳
	امتیازها و محدودیتهای آزمونهای تشریحی	۲۳۷
	موارد استفادهٔ آزمونهای تشریحی	۲۳۹
	پرسش شفاهی	۲۴۲
	آزمون بیرون‌بردنی	۲۴۳
	خلاصهٔ فصل	۲۴۴
	سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل	۲۴۵
فصل یازدهم:	**سنجش یادگیری به کمک آزمون کوته پاسخ**	۲۴۷
	تعریف و انواع آزمونهای کوته پاسخ	۲۴۸
	آزمونهای کوته پاسخ تشکیل یافته از مسائل عددی ساده	۲۴۹

قواعد تهیهٔ سؤالهای کوته پاسخ ۲۵۰
قواعد تهیهٔ سؤالهای کوته پاسخ ویژهٔ مسائل عددی ۲۵۴
امتیازها و محدودیتهای آزمونهای کوته پاسخ ۲۵۵
موارد استفادهٔ آزمونهای کوته پاسخ ۲۵۶
خلاصهٔ فصل ۲۵۸
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۲۵۹

فصل دوازدهم: سنجش عملکردی ۲۶۲
تعریف و اهمیت سنجش عملکردی ۲۶۳
سنجش فرایند و فراوردهٔ یادگیری ۲۶۵
انواع سنجشهای عملکردی ۲۶۷
دسته‌بندی گرانلاند از آزمونهای عملکردی ۲۶۷
دسته‌بندی نیتکو از روشهای سنجش عملکردی ۲۷۳
قواعد تهیهٔ آزمونهای عملکردی ۲۷۵
امتیازها و محدودیتهای روشهای سنجش عملکردی ۲۸۳
موارد استفادهٔ روشهای سنجش عملکردی ۲۸۴
خلاصهٔ فصل ۲۸۸
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۲۸۹

فصل سیزدهم: سنجش با استفاده از کار پوشه ۲۹۱
تعریف کارپوشه ۲۹۲
هدفهای تهیهٔ کارپوشه ۲۹۳
کارپوشهٔ الکترونیکی ۳۰۱
محتوای کارپوشه ۳۰۴
قواعد تهیه و استفاده از کارپوشه ۳۰۶
امتیازها و محدودیتهای روش سنجش با استفاده از کارپوشه ۳۱۰
موارد استفادهٔ روش کارپوشه ۳۱۳
خلاصهٔ فصل ۳۱۷
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۳۱۹

فصل چهاردهم: سنجش مشاهده‌ای ۳۲۱
فهرست وارسی ۳۲۲
تعریف فهرست وارسی ۳۲۲
ویژگیهای فهرست وارسی ۳۲۳
قواعد تهیهٔ فهرست وارسی ۳۲۶

نمره‌گذاری فهرست وارسی ۳۲۷
مقیاس درجه‌بندی ۳۲۷
تعریف مقیاس درجه‌بندی ۳۲۷
ویژگیهای مقیاس درجه‌بندی ۳۲۸
انواع مقیاسهای درجه‌بندی ۳۲۹
قواعد تهیهٔ مقیاس درجه‌بندی ۳۳۴
نمره‌گذاری مقیاس درجه‌بندی ۳۳۶
واقعه‌نگاری ۳۳۶
تعریف واقعه‌نگاری ۳۳۶
ویژگیهای واقعه‌نگاری ۳۳۷
قواعد انجام واقعه‌نگاری و تهیهٔ گزارش آن ۳۳۹
امتیازها و محدودیتهای روشهای سنجش مشاهده‌ای ۳۴۱
موارد استفادهٔ روشهای سنجش مشاهده‌ای ۳۴۲
استفاده از روشهای مختلف سنجش مشاهده‌ای به صورت ترکیبی ۳۴۳
خلاصهٔ فصل ۳۴۴
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۳۴۵

فصل پانزدهم: **سنجش رفتاری** ۳۴۷
تعریف و ویژگیهای سنجش رفتاری ۳۴۷
انواع روشهای سنجش رفتاری ۳۵۲
سنجش محصولات ماندنی رفتار ۳۵۲
سنجش رویدادهای گذرای رفتار (ثبت مشاهده‌ای) ۳۵۳
سنجش قیاس‌پذیر ۳۶۳
سنجش موقعیتی ۳۶۴
نقش‌گزاری ۳۶۶
سنجش نهان‌یابی ۳۶۷
نظارت شخصی ۳۶۸
سنجش شناختی ـ رفتاری ۳۶۹
سنجش روانی ـ فیزیولوژیکی ۳۷۰
مصاحبهٔ رفتاری ۳۷۱
امتیازها و محدودیتهای روشهای سنجش رفتاری ۳۷۲
موارد استفادهٔ روشهای سنجش رفتاری ۳۷۴
خلاصهٔ فصل ۳۷۶
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۳۷۷

فصل شانزدهم: سنجش عاطفی ۳۷۹
تعریف ویژگیهای عاطفی ۳۸۰
اهمیت ویژگیهای عاطفی ۳۸۱
ویژگیهای روشهای سنجش عاطفی ۳۸۳
روشهای مختلف سنجش ویژگیهای عاطفی ۳۸۴
استفاده از پرسشنامه برای سنجش عواطف ۳۸۴
قواعد تهیه و اجرای پرسشنامه‌های سنجش عواطف ۳۸۹
روش تفکیک معنایی ۳۹۰
قواعد تهیهٔ پرسشنامه تفکیک معنایی ۳۹۱
مقیاس لیکرت ۳۹۲
نمره‌گذاری مقیاس لیکرت ۳۹۳
قواعد تهیهٔ پرسشنامه‌های عاطفی مبتنی بر مقیاس لیکرت ۳۹۸
استفاده از مصاحبه برای سنجش عواطف ۳۹۹
قواعد انجام روش مصاحبه ۴۰۰
گروه‌سنجی ۴۰۲
استفاده از یادداشتهای روزانه برای سنجش عواطف ۴۰۵
امتیازها و محدودیتهای روشهای سنجش عاطفی ۴۰۶
موارد استفادهٔ روشهای سنجش عاطفی ۴۰۶
خلاصهٔ فصل ۴۱۲
سؤال و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۴۱۳

فصل هفدهم: اجرا، نمره‌گذاری، و تحلیل آزمون ۴۱۵
شرایط اجرای آزمونهای کتبی ۴۱۶
ترتیب قراردادن سؤالهای آزمون به دنبال یکدیگر ۴۱۶
نوشتن دستورالعمل یا راهنمای آزمون ۴۱۸
اجرای آزمون ۴۱۹
نمره‌گذاری ۴۲۰
تحلیل سؤالهای آزمون ۴۲۱
مراحل تحلیل سؤال ۴۲۱
محاسبهٔ ضریب دشواری سؤال ۴۲۲
تفسیر ضریب دشواری ۴۲۴
محاسبهٔ ضریب تمیز سؤال ۴۲۶
تفسیر ضریب تمیز ۴۲۷
همبستگی سؤال با کل آزمون ۴۲۷

همبستگی میان سؤالهای آزمون ۴۲۸
رابطهٔ بین ضریب دشواری و ضریب تمیز ۴۲۹
تحلیل گزینه‌های انحرافی ۴۳۰
تحلیل سؤالهای آزمونهای تشریحی و آزمونهای عملکردی ۴۳۰
تجدید نظر و اصلاح آزمون ۴۳۳
تحلیل سؤالهای آزمونهای وابسته به ملاک ۴۳۳
تمیز بین پیش و پس از آموزش ۴۳۳
تمیز بین حد تسلط و غیر حد تسلط ۴۳۶
مثالی برای تحلیل سؤال یک آزمون وابسته به ملاک ۴۳۸
تحلیل آزمون به کمک نظریهٔ سؤال ـ پاسخ ۴۴۰
مقایسهٔ نظریهٔ کلاسیک با نظریهٔ سؤال ـ پاسخ آزمون‌سازی ۴۴۱
منحنی ویژگی سؤال و استفاده از آن برای تحلیل سؤال آزمون ۴۴۲
برازنده کردن منحنی ویژگی سؤال ۴۴۸
تحلیل گزینه‌های انحرافی به کمک منحنی ویژگی سؤال ۴۵۱
خلاصهٔ فصل ۴۵۲
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۴۵۲

فصل هجدهم: **استفاده از روشهای آماری برای تفسیر نمره‌های آزمون** ۴۵۷
تعریف متغیر، فراوانی، و توزیع فراوانی ۴۵۸
تهیهٔ توزیع فراوانی نمرات ۴۵۸
نمایش نمرات به صورت نمودار ۴۶۰
نمودار هیستوگرام ۴۶۱
نمودار چندضلعی ۴۶۲
نمودار تراکمی ۴۶۴
نمودار ستونی ۴۶۵
نمودار دایره‌ای ۴۶۷
نمودار تصویری ۴۶۸
اندازه‌های گرایش مرکزی ۴۶۸
نما ۴۶۹
میانگین ۴۶۹
ویژگیهای میانگین ۴۷۱
میانه ۴۷۲
اندازه‌های تغییرپذیری ۴۷۵
دامنهٔ تغییر ۴۷۶

واریانس و انحراف معیار ۴۷۶
محاسبهٔ انحراف معیار نمره‌های طبقه‌بندی شده ۴۷۸
همبستگی بین نمرات ۴۸۰
روش محاسبهٔ ضریب همبستگی گشتاوری پیرسون ۴۸۱
روش محاسبهٔ ضریب همبستگی رتبه‌ای اسپیرمن ۴۸۳
تفسیر ضریب همبستگی ۴۸۴
سایر همبستگی‌ها ۴۸۵
همبستگی دورشته‌ای نقطه‌ای ۴۸۶
همبستگی دورشته‌ای ۴۸۹
همبستگی فی ۴۹۱
همبستگی چهارخانه‌ای یا تتراکوریک ۴۹۲
خلاصهٔ فصل ۴۹۵
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۴۹۶

فصل نوزدهم: هنجارها و نیمرخها ۵۰۱

تعریف هنجار ۵۰۲
انواع هنجارها ۵۰۳
هنجار سنی ۵۰۳
هنجار کلاسی ۵۰۳
هنجار سنی نمایی ۵۰۴
هنجار سن ذهنی ۵۰۵
هوشبهر ۵۰۵
هنجار درصدی ۵۰۵
هنجارهای مربوط به نمرات معیار ۵۰۹
نمرهٔ Z ۵۰۹
نمرهٔ T ۵۱۰
نمرات SAT ۵۱۱
نمرات ۹ بخشی ۵۱۱
هوشبهر انحرافی ۵۱۲
نمرهٔ سنی معیار ۵۱۲
نمرهٔ معادل منحنی بهنجار ۵۱۳
مقایسهٔ نظامهای مختلف نمرات با منحنی بهنجار ۵۱۴
توزیع بهنجار و منحنی بهنجار ۵۱۵
نیمرخ ۵۱۶

خلاصهٔ فصل ۵۲۱
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۵۲۱

فصل بیستم: **روایی آزمون** ۵۲۵
تعریف و انواع روایی ۵۲۶
روایی محتوایی ۵۲۷
روش تعیین روایی محتوایی ۵۲۸
روایی صوری ۵۲۸
روش تعیین روایی صوری ۵۲۹
روایی ملاکی ۵۲۹
روایی پیش‌بینی ۵۳۰
روش تعیین روایی پیش‌بینی ۵۳۰
روایی همزمان ۵۳۲
روش تعیین روایی همزمان ۵۳۲
مقایسهٔ روایی پیش‌بینی با روایی همزمان ۵۳۲
ویژگیهای مطلوب ملاک ۵۳۳
روایی سازه ۵۳۴
روش تعیین روایی سازه ۵۳۵
معنای عشق: مثالی از تعیین روایی سازه ۵۳۷
مقایسهٔ ویژگیهای روواییهای مختلف با یکدیگر ۵۴۰
روایی آزمونهای وابسته به ملاک ۵۴۰
روایی محتوایی آزمونهای وابسته به ملاک ۵۴۱
روش تعیین روایی محتوایی آزمونهای وابسته به ملاک ۵۴۱
روش لاشه برای تعیین ضریب روایی محتوایی ۵۴۳
روایی آموزشی ۵۴۴
روایی پیش‌بینی آزمونهای وابسته به ملاک ۵۴۵
روش تعیین روایی پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی براساس هدف آموزشی دیگر ۵۴۵
روش تعیین روایی پیش‌بینی یادگیری در حد تسلط ۵۴۷
سایر استفاده‌ها از روایی پیش‌بینی آزمونهای وابسته به ملاک ۵۴۸
روایی سازهٔ آزمونهای وابسته به ملاک ۵۴۹
روش تعیین روایی سازهٔ آزمونهای وابسته به ملاک ۵۴۹
عوامل مؤثر بر روایی آزمون ۵۵۱
عوامل درونی مؤثر بر روایی آزمون ۵۵۱

عوامل بیرونی مؤثر بر روایی آزمون ۵۵۲
خلاصهٔ فصل ۵۵۳
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۵۵۴

فصل بیست و یکم: پایایی آزمون ۵۵۷
تعریف پایایی ۵۵۸
روشهای تعیین پایایی ۵۵۹
پایایی بین مصححان ۵۵۹
پایایی آزمون ۵۵۹
روش بازآزمایی ۵۶۰
روش فرمهای موازی ۵۶۰
روشهای همسانی درونی ۵۶۱
روش دو نیمه کردن آزمون ۵۶۱
روش کودر-ریچاردسون ۵۶۳
روش ضریب آلفا ۵۶۶
روش تعیین پایایی آزمونهای سرعت ۵۶۹
پایایی نمرات اختلاف ۵۷۱
عوامل مؤثر در افزایش پایایی آزمون (وابسته به هنجار) ۵۷۲
استفاده از ضریب پایایی آزمون برای برآورد نمرهٔ حقیقی ۵۷۲
خطای معیار اندازه‌گیری ۵۷۵
استفاده از خطای معیار اندازه‌گیری برای برآورد نمرهٔ حقیقی ۵۷۶
مفهوم پایایی در نظریهٔ سؤال ـ پاسخ ۵۷۷
رابطهٔ بین روایی و پایایی (فرمول اصلاح کاهش) ۵۷۸
پایایی روشهای درجه‌بندی یا رتبه‌بندی ۵۸۰
پایایی آزمونهای وابسته به ملاک ۵۸۲
ضریب پایایی آزمونهای وابسته به ملاک (روش درصد توافق) ۵۸۳
ضریب پایایی آزمونهای وابسته به ملاک (روش کاپا) ۵۸۴
ضریب پایایی آزمونهای وابسته به ملاک (روش لیوینگ استون) ۵۸۶
ضریب پایایی آزمونهای وابسته به ملاک (روش لیبرمن و مرندا) ۵۸۷
عوامل مؤثر در افزایش پایایی آزمون (وابسته به ملاک) ۵۸۸
پایایی روشهای سنجش مستقیم ۵۸۸
پایایی روش سنجش محصولات ماندنی رفتار یا سنجش فرآوردهٔ رفتار ۵۸۸
پایایی روش ثبت فراوانی رفتار ۵۸۹
پایایی روشهای سنجش طول مدت و دورهٔ نهفتگی رفتار ۵۸۹

پایایی روشهای ثبت فاصله‌ای و ثبت نمونه‌گیری زمان ۵۹۰
خلاصهٔ فصل ۵۹۱
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۵۹۱

فصل بیست و دوم: **ارزشیابی معلم** ۵۹۵
اهمیت ارزش معلم ۵۹۶
هدف ارزشیابی معلم ۵۹۶
تعریف ارزشیابی معلم ۵۹۷
روشهای ارزشیابی معلم ۵۹۷
مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلمان ۵۹۸
محاسن و معایب روش مصاحبه ۶۰۱
مشاهده از کلاس درس ۶۰۱
مشکل نمونه‌گیری از آموزش معلم به کمک مشاهده ۶۰۳
محاسن و معایب روش مشاهدهٔ کلاسی ۶۰۳
استفاده از نظر دانش‌آموزان و دانشجویان ۶۰۵
پرسشنامه‌های نظرخواهی از دانش‌آموزان ۶۰۶
پرسشنامه‌های نظرخواهی از دانشجویان ۶۰۸
انواع سؤالهای پرسشنامه‌های نظرخواهی ۶۱۵
شیوهٔ اجرا، نمره‌گذاری، و گزارش نتایج پرسشنامه‌های نظرخواهی ۶۱۵
محاسن و معایب روش استفاده از نظر دانش‌آموزان و دانشجویان ۶۱۶
استفاده از نظر والدین دانش‌آموزان و دانشجویان ۶۱۸
پرسشنامه‌های نظرخواهی از والدین ۶۱۹
اجرا و نمره‌گذاری پرسشنامه‌های نظرخواهی از والدین ۶۱۹
محاسن و معایب روش استفاده از نظر والدین ۶۲۱
استفاده از یادگیری دانش‌آموزان و دانشجویان ۶۲۲
استفاده از آزمونهای عملکرد آموزشی ۶۲۴
محاسن و معایب روش استفاده از یادگیری دانش‌آموزان و دانشجویان ۶۲۵
خودسنجی یا ارزشیابی از خود ۶۲۶
فرم درجه‌بندی شخصی ۶۲۷
گزارش شخصی ۶۲۷
مواد آموزش شخصی ۶۲۹
مشاهدهٔ آموزش همکاران ۶۳۰
بازخورد حاصل از ضبط صدا و ضبط تصویر ۶۳۰
خوداندیشی ۶۳۱

محاسن و معایب روش خودسنجی ۶۳۲
ارزشیابی به وسیلهٔ همکاران ۶۳۲
محاسن و معایب روش ارزشیابی به وسیلهٔ همکاران معلم ۶۳۴
ارزشیابی به کمک کارپوشه ۶۳۴
ویژگیهای کارپوشهٔ آموزشی ۶۳۴
هدفهای کارپوشهٔ آموزشی ۶۳۴
محتوای کارپوشهٔ آموزشی ۶۳۶
ارزشیابی از کارپوشهٔ آموزشی ۶۳۶
محاسن و معایب روش ارزشیابی به کمک کارپوشه ۶۳۷
بررسی آثار علمی معلم ۶۳۷
محاسن و معایب روش بررسی آثار علمی معلم ۶۳۹
خلاصهٔ فصل ۶۴۱
سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل ۶۴۱

مراجع فارسی ۶۴۵

مراجع انگلیسی ۶۴۷

فهرست موضوعی (نمایه) ۶۵۵

واژه‌نامه انگلیسی ـ فارسی ۶۶۹

واژه‌نامه فارسی ـ انگلیسی ۶۷۷

جواب سؤالهای عینی پایان فصلها ۶۸۵

پیشگفتار ویرایش ششم

به ویرایش ششم کتاب *اندازه‌گیری، سنجش، و ارزشیابی آموزشی* تغییراتی به شرح زیر اعمال شده است:

۱. تمایزگذاری بین آزمون و پرسشنامه در فصل ۱
۲. اصلاح مطالب فصل ۳ با توجه به تغییرات اعمال شده در فصل ۱
۳. جایگزینی شش فصل روشهای سنجش (فصل ۷ تا فصل ۱۲) با ده فصل تازه
۴. بازنویسی موضوع مربوط به هنجار و هنجاریابی
۵. افزودن یک روش تازهٔ تعیین روایی سازه به فصل روایی (فصل بیستم ویرایش جدید)
۶. افزودن یک روش تازهٔ تعیین پایایی به فصل پایایی (فصل بیست و یکم ویرایش جدید) و تکمیل فرمول محاسبهٔ ضریب آلفا در همین فصل

مهم‌ترین تفاوت ویرایش ششم کتاب با ویرایش قبلی جایگزینی ده فصل مربوط به روشهای سنجش یادگیری به جای شش فصل ویرایش قبلی است. این ده فصل عمدتاً از کتاب *سنجش فرآیند و فرآوردهٔ یادگیری: روشهای قدیم و جدید* به کتاب *اندازه‌گیری، سنجش، و ارزشیابی آموزشی* اضافه شده‌اند. فایدهٔ این کار بی‌نیاز کردن خوانندگان کتاب *اندازه‌گیری، سنجش، و ارزشیابی آموزشی* از مراجعه به کتاب *سنجش فرآیند و فرآوردهٔ یادگیری* است. از آنجا که فصلهای مربوط به روشهای سنجش در ویرایشهای قبلی خیلی کامل نبودند، استاد و دانشجو برای تکمیل آن فصلها نیاز داشتند به کتاب *سنجش فرآیند و فرآوردهٔ یادگیری* مراجعه کنند. اکنون با اضافه شدن فصلهای جدید به جای فصلهای قبلی، دیگر نیازی به مراجعه به کتاب *سنجش فرآیند و فرآوردهٔ یادگیری* نیست. به سخن دیگر، ویرایش ششم کتاب *اندازه‌گیری، سنجش، و ارزشیابی آموزشی*، هم از جهت روشهای سنجش و هم از لحاظ مباحث نظری و محاسبات روانسنجی یک کتاب کامل درسی است. در مقابل، برای کسانی که صرفاً به روشهای مختلف سنجش علاقه‌مندند و نیازی به مسائل و مفاهیم و محاسبات روانسنجی ندارند کتاب *سنجش فرآیند و فرآوردهٔ یادگیری: روشهای قدیم و جدید* پاسخگوی نیازشان خواهد بود.

خانم مهرناز اسحاق‌نیا در تصحیح فرمهای چاپی و تهیهٔ فهرستهای این ویرایش به من کمک کرده‌اند. از ایشان سپاسگزارم.

امیدوارم ویرایش تازهٔ کتاب مانند ویرایشهای قبلی مورد استفادهٔ خوانندگان قرار گیرد.

دکتر علی‌اکبر سیف

مهر ۱۳۹۰

پیشگفتار ویرایش اول

در سال ۱۳۶۳ کتاب نسبتاً مختصری به نام *اندازه‌گیری و ارزشیابی پیشرفت تحصیلی* از من به چاپ رسید. این کتاب مورد استفادهٔ استادان و دانشجویان درس سنجش و اندازه‌گیری در آموزش و پرورش واقع گردید و تاکنون ۱۰ بار تجدید چاپ شده است. از سه سال پیش کوشیدم تا این کتاب را مورد بازنگری قرار دهم و کم و کسرهای آن را برطرف نمایم و محتوای آن را با توجه به یافته‌های جدید علمی در زمینهٔ روانسنجی و اندازه‌گیری و ارزشیابی در روانشناسی و علوم تربیتی، اصطلاحاً، به روز در آورم. پس از جمع‌آوری مطالب و تألیف و تدوین آنها متوجه شدم که حجم و سازمان‌بندی مطالب به اندازه‌ای با کتاب قبلی متفاوت است و تغییرات ایجاد شده به حدی زیاد است که کتاب تازه را به هیچ وجه نمی‌توان صورت تجدید نظر شدهٔ کتاب قبلی به حساب آورد. لذا بر آن شدم تا کتاب قبلی را دست نخورده نگاه دارم برای مصارفی که تاکنون داشته است، یعنی درسنامهٔ درس مقدماتی سنجش و اندازه‌گیری در سطح کارشناسی (لیسانس) رشته‌های علوم تربیتی و دوره‌های تربیت معلم، و علاوه بر آن کتاب تازه‌ای فراهم آورم با مطالب و ساختاری متفاوت. کتاب حاضر محصول آن کوشش است. این کتاب به عنوان یک کتاب درسی (درسنامه) برای دروس اندازه‌گیری و سنجش، آزمونهای پیشرفت تحصیلی، و اندازه‌گیری و ارزشیابی آموزشی برای سالهای بالای دوره‌های کارشناسی و دوره‌های کارشناسی ارشد (فوق لیسانس) شاخه‌های مختلف علوم تربیتی و روانشناسی قابل استفاده است.

علاوه بر کتاب *اندازه‌گیری و ارزشیابی پیشرفت تحصیلی*، در سال ۱۳۷۱ کتاب دیگری از من با عنوان *اندازه‌گیری و سنجش در علوم تربیتی* توسط دانشگاه پیام نور انتشار یافت که آن نیز تاکنون چندین بار تجدید چاپ شده است. این کتاب و کتاب *اندازه‌گیری و ارزشیابی پیشرفت تحصیلی* تا اندازهٔ زیادی از لحاظ محتوا به هم شبیه‌اند و مخاطبان هردو دانشجویان سالهای پایین دوره‌های کارشناسی روانشناسی و علوم تربیتی و دوره‌های تربیت معلم هستند. چنانکه پیداست، کتاب حاضر از لحاظ حجم مطالب بسیار مفصل‌تر و از لحاظ موضوعهای مطرح شده بسیار متنوع‌تر از دو کتاب قبلی است. بنابراین، دسترسی به کتاب حاضر خواننده را از هر دو کتاب قبلی بی‌نیاز می‌کند.

در خلال مدتی که مشغول تدوین مطالب کتاب بودم، فصلهای مختلف آن را، به تدریج که

آماده می‌شدند، به دانشجویان کارشناسی ارشد رشته‌های روانشناسی تربیتی و سنجش و اندازه‌گیری آموزش دادم و به کمک بازخوردهای به دست آمده از آن کلاسها نواقص و ابهامهای موجود را برطرف نمودم. با این حال، چنانچه استادان دانشمند و دانشجویان ارجمند در جهت بهبود کم و کیف کتاب نظرها و پیشنهادهای اصلاحی خود را برای من ارسال بدارند، خواهم کوشید تا در ویرایشهای بعدی آنها را مورد استفاده قرار دهم. زحمات آنان را ارج خواهم نهاد و از پیش سپاسگزار محبتهایشان هستم.

دکتر علی‌اکبر سیف

فروردین ۱۳۷۵

راهنمای مطالعه و یادگیری مطالب کتاب

توجه به پیشنهادهای زیر شما را در یادگیری بهتر و بیشتر مطالب کتاب کمک خواهد کرد:

۱. از آنجا که در تنظیم مطالب کتاب یک روال منطقی رعایت شده، بهتر است فصلهای مختلف کتاب را به ترتیبی که در دنبال هم آمده‌اند مطالعه کنید. با این حال، چنانچه نیازی به خواندن همهٔ مطالب کتاب نداشته باشید، می‌توانید هر فصل یا هر بخش از هر فصل را که بخواهید به طور مستقل بخوانید و مطالب آن را مورد استفاده قرار دهید.

۲. پیش از شروع به مطالعهٔ هر فصل، هدفهای یادگیری آن فصل را که در آغاز فصل آمده‌اند به دقت مطالعه کنید. این هدفها از پیش به شما می‌گویند که پس از مطالعهٔ هر فصل چه اطلاعات و مهارتهایی را باید بیاموزید.

۳. پس از مطالعهٔ هر فصل، به فهرست هدفهای آن فصل بازگردید و ببینید که کدام‌یک از هدفها را به خوبی یاد نگرفته‌اید. برای هر یک از هدفهایی که در آن توفیق کامل به دست نیاورده‌اید مطالعهٔ بیشتری انجام دهید.

۴. در پایان هر فصل تعدادی سؤال و تمرین آمده است که جواب دادن به سؤالها و انجام تمرینها شما را در بهتر یادگرفتن مطالب کمک خواهد کرد. اگر پس از مطالعهٔ هر فصل نتوانستید به سؤال (سؤالهایی) جواب دهید یا تمرین (تمرینهایی) را حل کنید، به مطالب مربوط به آن در داخل فصل باز گردید و آنها را دوباره خوانی کنید.

۵. از آنجا که تسلط یابی بر فنون و روشهای اندازه‌گیری و ارزشیابی آموزشی نیاز به تمرین و فعالیت عملی دارد، بهتر است همراه با مطالعهٔ فصلهای مختلف کتاب روشهای معرفی شده را عملاً تمرین کنید. برای این منظور، همگام با یادگیری فصلهای مختلف کتاب، به تهیه و اجرای یک آزمون پیشرفت تحصیلی برای یکی از کتابهای درسی دورهٔ دبستان، راهنمایی، یا دبیرستان اقدام کنید، و نتایج حاصل را طبق اصول بحث شده در کتاب تحلیل و تفسیر نمایید.

۶. ضمن معرفی مطالب مختلف کتاب و بحث پیرامون آنها، به منابع و مآخذ مختلفی استناد شده است. فهرست و مشخصات کامل این منابع در پایان کتاب آمده است. مراجعه به این منابع و مطالعهٔ آنها در غنی‌تر ساختن دانش و اطلاعات شما در موضوعهای مختلف به شما کمک خواهد کرد. دانشجویانی که این کتاب را به صورت کتاب درسی مورد استفاده قرار می‌دهند

نباید یادگیری خود را به محتوای این کتاب محدود کنند. کتابها و مقالات تخصصی مختلفی که در حوزهٔ سنجش، اندازه‌گیری، و ارزشیابی آموزشی منتشر شده و می‌شوند منابع ارزشمندی هستند که باید مورد استفاده قرار گیرند. ما فهرست کتابها و مقاله‌های تخصصی مربوط به موضوعهای بحث شده در این کتاب را که در دسترس داشته و از آنها در تدوین مطالب کتاب حاضر سود برده‌ایم به همین منظور در پایان کتاب آورده‌ایم. شما نیز به هر تعداد از آنها که می‌توانید دسترسی پیدا کنید مراجعه نمایید و مطالبشان را به عنوان مکمل کتاب حاضر مورد استفاده قرار دهید.

موفق باشید

دکتر علی‌اکبر سیف

هدفهای کلی کتاب

در تدوین فصلهای مختلف این کتاب هدفهای کلی زیر مورد نظر بوده‌اند. امید است دانشجویان با خواندن مطالب کتاب و انجام فعالیتهای یادگیری به این هدفها نایل آیند.

۱. آشنا شدن با مفاهیم و اصطلاحات مهم حوزهٔ سنجش، اندازه‌گیری و ارزشیابی آموزشی
۲. آشنا شدن با انواع آزمونها و دیگر وسایل اندازه‌گیری در روانشناسی و آموزش و پرورش
۳. آشنا شدن با انواع ارزشیابیهای آموزشی
۴. کسب مهارت در طراحی، اجرا، و تفسیر نتایج ارزشیابیهای آموزشی، به ویژه ارزشیابی پیشرفت تحصیلی
۵. آشنا شدن با اصول و روشهای آماری مورد استفاده در تحلیل و تفسیر نتایج آزمونها
۶. آشنا شدن با ویژگیهای مطلوب آزمونها، به ویژه روایی و پایایی
۷. آشنا شدن با هنجارها و نیمرخها و طرز استفاده از آنها در تفسیر و گزارش نتایج آزمونهای مختلف
۸. آشنا شدن با ابزارها و وسایل مختلف سنجش هدفهای آموزشی حوزه‌های سه‌گانهٔ شناختی، عاطفی، و روانی ـ حرکتی، و کسب مهارت در تهیه، اجرا، تحلیل، و تفسیر نتایج وسایل سنجش هدفهای این حوزه‌ها
۹. آشنا شدن با منابع و مآخذ مختلف اندازه‌گیری و ارزشیابی آموزشی
۱۰. علاقه‌مند شدن به موضوعات و مباحث گوناگون اندازه‌گیری و ارزشیابی آموزشی
۱۱. کسب توانایی تهیه و اجرای آزمونهای پیشرفت تحصیلی و تحلیل و تفسیر نتایج آنها
۱۲. کسب توانایی بررسی انتقادی آزمونهای تهیه شده توسط دیگران و نشان دادن نقاط قوت و ضعف آنها

به منظور دستیابی به هدفهای کلی بالا، برای هر یک از فصلهای کتاب تعدادی هدف دقیق تهیه شده است، به این امید که دانشجویان، از راه یادگیری این هدفهای دقیق، در رسیدن به هدفهای کلی کتاب موفق بشوند.

۱

تعریف اصطلاحات

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. اندازه‌گیری را تعریف کند و برای اندازه‌گیری ویژگیهای فیزیکی و روانی مثالهایی ذکر نماید.

۲. آزمون را تعریف کند و رابطهٔ آن را با پرسشنامه توضیح دهد.

۳. تفاوت بین آزمون را به عنوان یک وسیلهٔ اندازه‌گیری ویژگیهای روانی و سایر وسایل اندازه‌گیری ویژگیهای فیزیکی توضیح دهد.

۴. آزمودن را تعریف کند و رابطهٔ آن را با اندازه‌گیری و آزمون توضیح دهد.

۵. سنجش را تعریف کند و با ذکر مثال رابطهٔ آن را با اندازه‌گیری و آزمودن توضیح دهد.

۶. ارزشیابی آموزشی را تعریف کند.

۷. مفهوم کیفیت را در ارزشیابی آموزشی تعریف کند.

۸. رابطهٔ میان اندازه‌گیری، آزمودن، سنجش، و ارزشیابی را توضیح دهد.

۹. پژوهش آموزشی و پژوهش ارزشیابی را تعریف کند.

۱۰. رابطهٔ بین پژوهش آموزشی و ارزشیابی آموزشی را توضیح دهد.

موضوع این کتاب *اندازه‌گیری، سنجش، و ارزشیابی آموزشی* است. بنابراین، لازم است ابتدا این اصطلاحات را روشن سازیم. به همین منظور، در فصل حاضر، نخست این اصطلاحات و مفاهیم وابسته به آنها را تعریف و تشریح می‌کنیم و بعد به مقایسه و توضیح روابط میان آنها می‌پردازیم.

تعریف اندازه‌گیری

در *اندازه‌گیری*[1]، *ویژگیها* یا *صفات* اشیاء و افراد تعیین می‌شوند و مقدار آن ویژگیها یا صفات به صورت عدد یا رقم گزارش می‌شود. بنابراین، می‌توان گفت: "اندازه‌گیری عبارت است از فرایندی که تعیین می‌کند یک شخص یا یک شیء چه مقدار از یک ویژگی برخوردار است" (گی[2]، ۱۹۹۱). نیتکو[3] (۲۰۰۱) اندازه‌گیری آموزشی را به نحو زیر تعریف کرده است: "شیوهٔ نسبت دادن یک عدد (که معمولاً به آن *نمره* می‌گویند) به یک صفت یا ویژگیِ معینِ یک شخص به گونه‌ای که آن عدد درجه‌ای از آن صفت یا ویژگی را که آن شخص از آن برخوردار است نشان دهد" (ص ۷).

چنان که از تعریفهای بالا پیداست، در اندازه‌گیری، ویژگی یا صفتی که اندازه‌گیری می‌شود به صورت عدد یا رقم در می‌آید. مثلاً، به جای اینکه بگوییم حسن نسبت به سن و قدش لاغر است، می‌گوییم حسن ۱۶ سال دارد، قدش ۱۷۰ سانتیمتر است، ولی تنها ۴۰ کیلوگرم وزن دارد. یا، به عوض اینکه بگوییم پروین از زهرا باهوش‌تر است، می‌گوییم هوشبهر پروین ۱۱۱ ولی هوشبهر زهرا ۱۰۲ است. در این مثالها، اعداد نسبت داده شده به افراد حاصلِ اندازه‌گیریهایی هستند که از ویژگیهای جسمی و روانی آنان به عمل آمده‌اند.

اندازه‌گیری فرایند یا شیوه‌ای دقیق و نظامدار است. نانالی[4] (۱۹۷۸) در تعریفی که از اندازه‌گیری به دست داده بر کاربرد *قاعده*[5] تأکید کرده است: "اندازه‌گیری دربرگیرندهٔ قواعدی است برای نسبت دادن اعداد به اشیاء [یا افراد] به گونه‌ای که صفتهایی از آنها را به صورت *کمیّت* نشان دهد" (ص ۳). منظور از کاربرد *قاعده* در اندازه‌گیری این است که روشهای نسبت دادن اعداد باید به‌طور روشن بیان شوند. مثلاً در نمره‌گذاری نتایج یک آزمون حساب، وقتی که *جوابهای درست دانش‌آموزان شمارش می‌شوند و به تعداد جوابهای درست به آنها نمره داده می‌شود*، شاهد نمونه‌ای از کاربرد قاعده در اندازه‌گیری هستیم. بنابراین، قواعد اندازه‌گیری "خط مشی‌هایی هستند برای نشان دادن مقدار (یا نوعی ویژگی دیگرِ) شیء مورد اندازه‌گیری" (کوهن[6] و سوردلیک[7]، ۲۰۰۲، ص ۷۳).

به مثال دیگری از کاربرد قاعده در اندازه‌گیری توجه کنید. فرض کنید می‌خواهید متغیر *سطح*

---

1. measurement
2. Gay
3. Nitko
4. Nunnally
5. rule
6. Cohen
7. Swerdlik

تحصیلات آموزشگاهی را اندازه‌گیری کنید. به این منظور، تصمیم می‌گیرید که تعداد سالهایی را که فرد در آموزشگاه با موفقیت تحصیل کرده است به عنوان شاخص تحصیلات آموزشگاهی مورد استفاده قرار دهید. بنابراین، کسانی که تحصیلات ابتدایی را به پایان رسانیده باشند نمرهٔ ۵ خواهند گرفت؛ و به همین منوال کسانی که تحصیلات دورهٔ راهنمایی را به پایان رسانیده باشند نمرهٔ ۸ (۸=۳+۵) و کسانی که تحصیلات دبیرستانی را به اتمام رسانیده باشند نمرهٔ ۱۲ دریافت خواهند کرد. باز هم به همین ترتیب، افرادی که مجموعاً ۱۰ سال تحصیل کرده باشند نمرهٔ ۱۰ و افرادی که، علاوه بر ۱۲ سال تحصیل پیش از دانشگاه، ۴ سال هم در دانشگاه تحصیل کرده باشند نمرهٔ ۱۶ (۱۶=۴+۱۲) خواهند گرفت.

در بعضی موارد، قاعدهٔ اندازه‌گیری تا حدی ساده و بدیهی است که توضیح مفصل آن ضرورت ندارد، مانند زمانی که از خط‌کش برای اندازه‌گیری طول یک میز استفاده می‌شود. اما قواعد مربوط به اندازه‌گیری ویژگیهای روانی و تربیتی افراد مانند هوش یا یادگیری دانش‌آموزان به بیان دقیق نیاز دارند. فایدهٔ بیان قواعد اندازه‌گیری این است که کمک می‌کند تا افراد مختلفی که می‌خواهند وسیلهٔ اندازه‌گیری را مورد استفاده قرار دهند آن را یکسان به کاربندند.

بیان قواعد اندازه‌گیری از جنبه‌های مهم **استاندارد کردن یا میزان کردن**[1] آزمونهای روانی و تربیتی است. بنا به گفتهٔ آیکن[2] (۱۹۹۱)، "هرگونه ابزار اندازه‌گیری استاندارد شده باید دارای راهنما یا دستور عمل روشنی برای اجرا و نمره‌گذاری باشد که به دقت به کار بسته شود و امکان تفسیر شخصی و سوگیری در نتایج وجود نداشته باشد" (ص ۸۷). بنابراین، از مشخصات مهم آزمونهای استاندارد شده یا میزان شده این است که افراد مختلفی که این وسائل اندازه‌گیری را به کار می‌بندند نتایج یکسانی به دست می‌آورند. به عنوان نمونه، در صورتی یک آزمون هوش به خوبی استاندارد شده است که افراد مختلفی که آن را با گروه معینی از دانش‌آموزان مورد استفاده قرار می‌دهند نمرات هوشی یکسانی به دست آورند.

تعریف آزمون

گفتیم که اندازه‌گیری یک فرایند است. این فرایند نیاز به وسیله‌ای دارد که به آن وسیلهٔ اندازه‌گیری می‌گویند. برای اندازه‌گیری ویژگیهای مختلف اشیاء و افراد از وسایل مختلفی استفاده می‌شود. مثلاً برای اندازه‌گیری ویژگیهای فیزیکی، مثل طول و وزن اشیاء یا قد و وزن افراد، از متر و ترازو استفاده می‌شود. متر و ترازو ویژگیهای فیزیکی را به طور مستقیم اندازه می‌گیرند، لذا به این دسته از وسایل اندازه‌گیری وسایل اندازه‌گیری مستقیم می‌گویند. علاوه بر ویژگیهای فیزیکی، رفتارهای آشکار یا محصولات رفتاری افراد را نیز می‌توان به‌طور مستقیم اندازه‌گیری کرد. مثلاً تعداد دفعاتی که

1. standardization 2. Aicken

دانش‌آموزی در یک جلسهٔ درس صندلی‌اش را ترک می‌کند (رفتار آشکار) یا تعداد مسائل ریاضی را که درست حل می‌کند (محصول رفتار) می‌توان به‌طور مستقیم اندازه گرفت. برخلاف ویژگیهای فیزیکی (جسمی) و رفتارهای آشکار که به طور مستقیم اندازه‌گیری می‌شوند، ویژگیها یا صفات روانی مانند هوش، خلاقیت، انگیزش، نگرش، یادگیری، و مانند اینها را باید به‌طور غیرمستقیم اندازه‌گیری کرد، زیرا هیچ کس به طور مستقیم به این ویژگیها دسترسی ندارد. معمول‌ترین وسیلهٔ اندازه‌گیری ویژگیها یا صفات روانی **آزمون**[1] است.

از آنجا که متخصص اندازه‌گیری به ویژگی روانی مورد نظرش دسترسی مستقیم ندارد تا مانند یک ویژگی جسمی یا یک رفتار آشکار آن را مستقیماً اندازه بگیرد، ناچار است به رفتارهایی از فرد که به گمان او مُعرف ویژگی روانی مورد نظر هستند مراجعه کند و به اندازه‌گیری آنها بپردازد. مثلاً رفتارهایی از افراد را که معرف هوشمندی یا کم‌هوشی هستند اندازه می‌گیرد و بر اساس آنها تعیین می‌کند که نمرهٔ هوش هر کسی چقدر است، یا با پرسیدن سؤالهایی از دانش‌آموزان کلاس تعیین می‌کند که هریک از آنها چه مقدار از دانش و مهارتهای درسی را که قرار بوده است یادبگیرد یادگرفته است.

رفتارهایِ مُعرفِ ویژگیهای روانی بسیار متنوع‌اند و بنابراین نمی‌توان همهٔ آنها را اندازه‌گیری کرد. مثلاً، اگر بخواهیم تمام رفتارهایی را که مُعرف هوشمندی یا یادگیری یک موضوع درسی‌اند فهرست کنیم در هیچ آزمونی نمی‌گنجند. بنابراین، متخصص اندازه‌گیری نمونه‌ای از رفتارهای مورد نظرش را انتخاب می‌کند و آزمون خود را بر پایهٔ آن نمونه بنا می‌نهد. پس، بنابه گفتهٔ گرانلاند[2] و لین[3] (۱۹۹۰)، "آزمون وسیله یا روشی نظام‌مند (یا نظام‌مند) برای اندازه‌گیری نمونه‌ای از رفتار است" (ص ۵). هر چه نمونهٔ رفتارهای انتخاب شده و گنجانیده شده در آزمون بیشتر معرف ویژگی روانی مورد نظر باشد، به همان میزان نتایج اندازه‌گیری دقیق‌تر و قابل اعتمادترند. بنابراین، در تعریف آزمون می‌توان گفت که *آزمون به مجموعه‌ای سؤال گفته می‌شود که غالباً به طور کتبی* (مداد و کاغذی)[4] *برای پاسخگویی در اختیار آزمون‌شوندگان قرار داده می‌شود.*

تعریف فوق از آزمون به اندازهٔ کافی گسترده هست که پرسشنامه[5] را هم شامل شود. از این‌رو، بعضی متخصصان روانسنجی بین آزمون و پرسشنامه تفاوت قائل نمی‌شوند. با این حال، این دو از جهاتی با هم فرق دارند، به گونه‌ای که می‌توان آنها را دو نوع ابزار اندازه‌گیری متفاوت تصور کرد. برای بحث مفصل دربارهٔ ویژگیهای متفاوت بین آزمون و پرسشنامه به اثر دیگر مؤلف (سیف، ۱۳۹۰) نگاه کنید. ما در اینجا تنها به تفاوت آنها از لحاظ موضوع مورد اندازه‌گیری اشاره می‌کنیم. تفاوت مهم بین آزمون و پرسشنامه در این است که به کمک آزمون تواناییها اندازه‌گیری می‌شوند، اما پرسشنامه ویژگیهای غیرتوانایی را می‌سنجد. به همین دلیل، به آزمون ابزار

1. test
2. Gronlund
3. Linn
4. paper-and-pencil
5. questionnaire

اندازه‌گیری **عملکرد بیشینه**[1] گفته می‌شود، در حالی‌که پرسشنامه ابزار اندازه‌گیری **عملکرد نوعی**[2] نام دارد. عملکرد بیشینه حداکثر توانایی فرد را در زمینهٔ مورد سنجش نشان می‌دهد؛ عملکرد نوعی وضع معمول فرد را معرفی می‌کند.

آزمونهایی که برای اندازه‌گیری عملکرد بیشینه مورد استفاده قرار می‌گیرند تعیین می‌کنند که افراد به چه خوبی از عهدهٔ انجام کارها برمی‌آیند، به چه خوبی مهارتها و موضوعهای مختلف را یاد گرفته‌اند، و شایستگی‌شان در زمینه‌های گوناگون چقدر است (کوالی[3]، ۲۰۱۰). نمونه آزمونهایی که عملکرد بیشینه را می‌سنجند عبارت‌اند از آزمونهای هوش و استعداد و پیشرفت تحصیلی. پرسشنامه‌های سنجش عملکرد نوعی برای سنجش توانایی‌ها به کار نمی‌روند، بلکه به کمک آنها وضع معمول و طبیعی افراد اندازه‌گیری می‌شود. نمونه ابزارهای سنجش عملکرد نوعی عبارت‌اند از پرسشنامه‌های شخصیت و نگرش و علاقه (کوالی، ۲۰۱۰).

به رغم وجود تفاوت در بین آزمون و پرسشنامه، آنها در تعریف مشترک‌اند. گفتیم که آزمون به نمونه‌ای از رفتارهای معرف یک ویژگی مانند هوش گفته می‌شود که معمولاً به صورت مجموعه‌ای سؤال کتبی (مداد و کاغذی) برای پاسخگویی در اختیار آزمون‌شوندگان قرار می‌گیرد. پرسشنامه نیز دربرگیرندهٔ نمونه رفتارهای معرف یک ویژگی مانند نگرش است که آن نیز غالباً (اما نه همیشه) به صورت کتبی برای پاسخگویی در اختیار پاسخ‌دهندگان قرار می‌گیرد. از لحاظ تعریف، می‌توان گفت که پرسشنامه "به مجموعه‌ای سؤال گفته می‌شود که هدف آن جمع‌آوری اطلاعات دربارهٔ ویژگیهایی از افراد است" (سیف ۱۳۹۰، ص ۲۷). وجه اشتراک دیگر آزمون و پرسشنامه این است که هر دو ابزار سنجش کمّی‌اند، یعنی نتیجهٔ اندازه‌گیری به وسیلهٔ آنها در قالب عدد و رقم (یعنی نمره) گزارش می‌شود.

تعریف آزمودن

بنابه تعریف، *وقتی که برای اندازه‌گیری یکی از ویژگیهای روانی یا تربیتی یک فرد یا گروهی از افراد از آزمون استفاده می‌شود به این فعالیت آزمودن*[4] *می‌گویند*. در تعریف آزمون، گفتیم که آزمون نمونه‌ای از رفتارهای موردنظر را اندازه می‌گیرد. از این رو، "اندازه‌گیری متغیرهای روانشناختی به وسیلهٔ تدابیر یا روشهایی که به منظور کسب نمونه‌هایی از رفتار طراحی شده‌اند آزمودن روانی یا روان‌آزمایی نام دارد" (کوهن و سوردلیک، ۲۰۰۲، ص ۴). هرچند که فعل آزمودن از اسم آزمون مشتق می‌شود، و گرچه ما آزمون را از پرسشنامه جدا فرض کردیم، اما آزمودن هم به کاربست آزمون و هم به استفاده از پرسشنامه برای سنجش ویژگیهای روانی گفته می‌شود. از این‌رو، ما هم به اندازه‌گیری یک ویژگی توانایی (مثلاً هوش) به کمک یک آزمون هوش و هم به اندازه‌گیری یک

1. maximum performance 2. typical performance 3. Coaly
4. testing

ویژگی غیرتوانایی (مثلاً شخصیت) به کمک یک پرسشنامهٔ شخصیت *آزمودن* می‌گوییم. از آنجاکه آزمون یک وسیلهٔ اندازه‌گیری است، هروقت آزمودن صورت می‌پذیرد یعنی آزمونی به‌کار می‌رود نوعی اندازه‌گیری انجام می‌شود، اما هر نوع اندازه‌گیری الزاماً آزمودن نیست. اندازه‌گیری از آزمودن مفهوم گسترده‌تری دارد، زیرا آزمودن نوع خاصی از اندازه‌گیری است. ممکن است اندازه‌گیری صورت پذیرد اما آزمونی در کار نباشد؛ یعنی اندازه‌گیری می‌تواند با استفاده از وسیلهٔ دیگری بجز آزمون انجام شود، مانند زمانی که یک معلم، از راه مشاهده و با شمردن تعداد دفعاتی که یک دانش‌آموز هم‌کلاسی کناردستی‌اش را آزار می‌دهد، رفتار پرخاشگرانهٔ او را اندازه می‌گیرد.

تعریف سنجش

اصطلاح دیگر نزدیک به *اندازه‌گیری* و *آزمودن* اصطلاح **سنجش**[1] است. سنجش از آزمودن و اندازه‌گیری مفهوم گسترده‌تری دارد. بنا به گفتهٔ نیتکو (۲۰۰۱)، "سنجش یک اصطلاح کلی است و به صورت فرایندی تعریف می‌شود که برای گردآوری اطلاعات موردنیاز تصمیم‌گیری دربارهٔ دانش‌آموزان، برنامه‌های درسی، و سیاستهای آموزشی مورد استفاده قرار می‌گیرد" (ص ۴). اِگِن[2] و کاوچاک[3] (۲۰۰۱) **سنجش کلاسی**[4] را به عنوان "تمامی فرایندهای موردنیاز برای تصمیم‌گیری دربارهٔ یادگیری دانش‌آموزان" (ص ۵۹۴) تعریف کرده‌اند.

چنان که دیدیم، وقتی که اندازه‌گیری صرفاً با استفاده از آزمون یا پرسشنامه انجام می‌شود *آزمودن* صورت می‌پذیرد، اما در *سنجش* وسایل و فنون مختلفی برای جمع‌آوری اطلاعات به کار می‌روند. از جملهٔ آنها می‌توان به آزمون، پرسشنامه، مقیاس درجه‌بندی، فهرست وارسی، کار آزمایشگاهی، پروژهٔ تحقیقی، امتحان شفاهی، تکلیف درسی، مصاحبه، و مشاهدهٔ عملکرد و رفتار دانش‌آموزان در موقعیتهای مختلف اشاره کرد. به عنوان نمونه، سانتروک[5] (۲۰۰۱) گفته است "مشاهدات غیررسمی می‌تواند دربارهٔ میزان انگیزش دانش‌آموزان نسبت به یادگیری موضوعهای درسی اطلاعات مفیدی به شما بدهد. همچنین اگر رابطهٔ خوبی با دانش‌آموزانتان داشته باشید پرسشهای مستقیم از آنان در یک گفتگوی خصوصی شما را از چگونگی نگرش آنان نسبت به درس و مدرسه آگاه خواهد ساخت" (ص ۴۹۸).

وولفلک[6] (۲۰۰۴) می‌گوید متخصصان ارزشیابی و اندازه‌گیری این روزها به طور فزاینده‌ای از اصطلاح سنجش برای توصیف فرایند جمع‌آوری اطلاعات دربارهٔ یادگیری دانش‌آموزان استفاده می‌کنند. او در این باره گفته است:

سنجش اصطلاح کلی‌تری از آزمودن و اندازه‌گیری است، زیرا سنجش دربرگیرندهٔ همهٔ راههای نمونه‌گیری و مشاهدهٔ مهارتها، دانش، و تواناییهای دانش‌آموزان است.... سنجش هم می‌تواند به

1. assessment
2. Eggen
3. Kauchak
4. classroom assessment
5. Santrock
6. Woolfolk

صورت رسمی انجام پذیرد، مانند اجرای آزمونها، و هم به صورت غیررسمی، مانند مشاهدهٔ اینکه چه کسی در کار گروهی به صورت رهبر گروه در خواهد آمد. سنجش هم می‌تواند به وسیلهٔ معلم کلاس طراحی و اجرا شود، و هم به وسیلهٔ مؤسسه‌های آموزشی ادارات محلی، ایالتی، و ملی. و امروزه سنجش از سؤالها و تمرینهای مداد و کاغذی بسیار فراتر رفته و مشاهدهٔ عملکرد، ایجاد کارپوشه، و کارهای دست‌ساز یادگیرندگان را نیز شامل می‌شود. (ص ۵۱۴)

اصطلاح سنجش بیشتر در حوزهٔ روانشناسی بالینی و آموزش و پرورش به کار می‌رود، زیرا در روانشناسی بالینی و آموزش و پرورش، علاوه بر آزمون و پرسشنامه، از وسایل و روشهای مختلف دیگری برای تعیین رفتارها و ویژگیهای روانی و تربیتی افراد استفاده می‌شود، مانند مصاحبه، ایفای نقش، استفاده از دستگاههای روانی ـ فیزیولوژیکی، و مانند اینها. آرندز[1] (۱۹۹۴) در توصیف سنجش در آموزش و پرورش مطالب زیر را بیان داشته است.

اصطلاح سنجش معمولاً به مجموعهٔ کاملی از اطلاعات که توسط معلمان در بارهٔ دانش‌آموزان و کلاسهایشان جمع‌آوری و دسته‌بندی می‌شوند اشاره می‌کند. اطلاعات مربوط به دانش‌آموزان را می‌توان به‌طور غیررسمی، مثلاً از راه مشاهده یا مبادلهٔ کلامی به دست آورد. همچنین این گونه اطلاعات را می‌توان به صورت رسمی، مثلاً از طریق دادن تکالیف، آزمونهای مختلف، و گزارشهای کتبی کسب کرد. افزون بر اینها، اطلاعات مربوط به کلاسهای درس و آموزش معلمان نیز می‌تواند بخشی از سنجش به حساب آید. (ص ۲۱۳)

تفاوت دیگر میان سنجش و اندازه‌گیری و آزمودن این است که هم اندازه‌گیری و هم آزمودن عموماً با کمیت سروکار دارند، اما سنجش الزاماً به کمیت منحصر نمی‌شود. به سخن دیگر، زمانی که اندازه‌گیری یا آزمودن صورت می‌پذیرد، نتیجه به صورت اعداد یا ارقام (کمیت) نشان داده می‌شود، اما نتیجهٔ سنجش می‌تواند به صورت غیر کمّی (توصیفی) گزارش شود. به عنوان مثال، وقتی که، به کمک یک آزمونِ ریاضی، یادگیری یک دانش‌آموز مشخص می‌شود، نتیجهٔ آن در قالب یک نمره که جنبهٔ کمّی دارد تعیین می‌گردد، اما می‌توان رفتار اجتماعی یک دانش‌آموز را در کلاس درس با عنوانهای پرخاشگر، سازگار، و مانند اینها توصیف کرد که جنبهٔ کیفی دارند.

تعریف ارزشیابی

اصطلاح **ارزشیابی**[2] (یا ارزیابی)، به طور ساده، به تعیین **ارزش**[3] برای هر چیزی یا **داوری ارزشی**[4] کردن گفته می‌شود. با این حال، تعریف جامع‌تری از ارزشیابی به شرح زیر می‌توان به دست داد: "ارزشیابی به یک فرایند نظامدار برای جمع‌آوری، تحلیل، و تفسیر اطلاعات گفته می‌شود به این

1. Arends 2. evaluation 3. value
4. value judgement

منظور که تعیین شود آیا هدفهای مورد نظر تحقق یافته‌اند یا در حال تحقق یافتن هستند و به چه میزانی" (گی، ۱۹۹۱، ص ۶).

یکی از ویژگیهای مهم ارزشیابی تعیین کیفیت[1] است. در فرایند ارزشیابی داوری ارزشی با توجه به کیفیت به عمل می‌آید. نیتکو (۲۰۰۱) کیفیت را در ارزشیابی پیشرفت تحصیلی به عنوان *دانش، مهارتها، و تواناییهایی که از دانش‌آموزان پس از آموزش انتظار می‌رود* تعریف کرده است. وی به عنوان مثالی از ارزشیابی پیشرفت تحصیلی گفته است فرض کنید شما دانش‌آموزی را در توانایی نوشتن، نسبت به پایهٔ تحصیلی‌اش، بسیار خوب ارزشیابی کرده‌اید. به منظور انجام این ارزشیابی شما توانایی نوشتن او را سنجش کرده‌اید. برای این کار، مجموعه نوشته‌های دانش‌آموز را مرور کرده‌اید، و نوشته‌های او را با نوشته‌های سایر دانش‌آموزان هم‌کلاسی‌اش و با معیارهای دیگری که برای نوشتن در سطح پایهٔ این دانش‌آموز می‌شناسید مقایسه کرده‌اید (تعیین کیفیت). سرانجام، داوری شما از توانایی نوشتن این دانش‌آموز به این نتیجه‌گیری انجامیده است که نوشته‌های او از کیفیت خوبی برخوردارند. پس توصیه می‌کنید که این دانش‌آموز می‌تواند در یک مسابقهٔ ملی مربوط به مقاله‌نویسی شرکت کند. بنابراین، ارزشیابی مبنای تصمیم‌گیری برای فعالیتهای عملی و اجرایی است.

مثال بالا در رابطه با ارزشیابی توانایی نوشتن نوعی ارزشیابی پیشرفت تحصیلی است که نمونه‌های دیگر آن می‌تواند ارزشیابی از پیشرفت دانش‌آموزان در درسهای ریاضی، علوم، تاریخ، و مانند اینها باشد. علاوه بر پیشرفت تحصیلی دانش‌آموزان، ارزشیابی آموزشی با موضوعهای دیگری نیز سروکار دارد. از جملهٔ آنها می‌توان عملکرد معلمان و مدیران، روشهای آموزشی، برنامه‌های درسی، دوره‌های آموزشی، مواد آموزشی، پروژه‌های آموزشی، گروه‌های آموزشی، و سازمانهای آموزشی را نام برد.

این‌گونه ارزشیابیهای آموزشی از ارزشیابی پیشرفت تحصیلی که در آن صرفاً موفقیتها و دستاوردهای یادگیری دانش‌آموزان و دانشجویان مورد نظر است معنی گسترده‌تری دارند. **ارزشیابی پیشرفت تحصیلی**[2] مفهومی بسیار نزدیک به مفهوم سنجش یادگیری است، اما **ارزشیابی آموزشی**[3] در مفهوم کلی آن معنایی بسیار وسیع دارد. به عنوان نمونه، مک بیت[4] و مک گلین[5] (۲۰۰۱) مفهوم ارزشیابی از آموزشگاه را که یک ارزشیابی آموزشی است به گونهٔ زیر توضیح داده‌اند:

> ارزشیابی از آموزشگاه باید نشان دهد که یادگیری چگونه در آن اتفاق می‌افتد و چگونه می‌تواند در این جهت بهبود یابد. این ارزشیابی باید مسائل مهمی چون مشورت و تصمیم‌گیری را که می‌تواند بر اخلاقیات کارکنان آموزشی تأثیر بگذارد درنظر بگیرد. این ارزشیابی باید امیدها و آرزوهای شاگردان و رضایت آنان از چگونگی برآورده شدن آن امیدها و آرزوها را نشان دهد. این ارزشیابی باید به

1. quality
2. achievement evaluation
3. educational evaluation
4. MacBeath
5. McGlynn

رضایت معلمان و سایر کارکنان آموزشگاه، کیفیت روابط میان اعضاء، و فرصتها و تساوی دسترسی به یکدیگر را شامل شود. این ارزشیابی باید ارتباط مدرسه با والدین، بار بودن به روی جامعه، انعطاف‌پذیری و پاسخگویی، و آمادگی برای تغییر را بسنجند. (ص ۸)

ورتن[1] و سندرز[2] (۱۹۸۷) دو تن از صاحب‌نظران در حوزهٔ ارزشیابی آموزشی گفته‌اند "در ارزشیابی به یک فعالیت رسمی گفته می‌شود که برای تعیین کیفیت، اثربخشی، یا ارزش یک برنامه، فراورده، پروژه، فرایند، هدف، یا برنامهٔ درسی به اجرا در می‌آید" (ص ۲۲). این متخصصان ارزشیابی آموزشی برای روشن‌ساختن تعریف خود مثال زیر را به دست داده‌اند: یک ارزشیاب برنامهٔ درسی[3] ابتدا هدفهای کلی یا غایتهای[4] برنامهٔ درسی را تعیین می‌کند و بعد، با استفاده از درون‌داد به دست آمده از گروههای مرجع مربوط، تعیین می‌کند که این هدفها برای دانش‌آموزان، پدران و مادران، و جامعه خوب هستند یا نه. پس از آن، ارزشیاب اطلاعات لازم را برای ارزشیابی از میزان تحقق یافتن آن هدفها و آثار جانبی ناشی از برنامهٔ درسی جمع‌آوری می‌کند. سپس به تحلیل و تفسیر اطلاعات به دست آمده می‌پردازد. و وقتی که این کار نیز انجام گرفت، آن گاه ارزشیاب در بارهٔ ارزش برنامهٔ درسی قضاوت یا داوری می‌کند و معمولاً به فرد یا افراد مسئول در تصمیم‌گیری‌های نهایی پیشنهادهایی ارائه می‌دهد.

گفتیم که ارزشیابی آموزشی در پی تعیین کیفیت است و در ارزشیابی پیشرفت تحصیلی کیفیت را به عنوان میزان دستیابی دانش‌آموزان و دانشجویان به دانش، مهارتها، و تواناییهایی که از آنها انتظار می‌رود تعریف کرده‌اند. در ارزشیابی آموزشی به طور کلی کیفیت ***به صورت تناسب داشتن موضوع مورد ارزشیابی با هدفهای موردنظر*** تعریف می‌شود. مثلاً وقتی که نتایج یک برنامهٔ درسی مورد ارزشیابی قرار می‌گیرد و معلوم می‌شود که اجزای آن برنامه هدفهای موردنظر را برآورده کرده است، می‌گوییم که آن برنامه از کیفیت مطلوبی برخوردار است.

مقایسهٔ اندازه‌گیری، آزمودن، سنجش، و ارزشیابی با یکدیگر

چنان‌که دیدیم، ارزشیابی شامل ***داوری ارزشی*** دربارهٔ کیفیتِ موضوع مورد ارزشیابی است که مبتنی بر تحلیل و تفسیر اطلاعات جمع‌آوری شده است. اطلاعات موردنیاز ارزشیابی آموزشی به کمک اندازه‌گیری (عموماً آزمودن) و یا سنجش گردآوری می‌شوند، و هرچه اطلاعات گردآوری شده جامع‌تر و کامل‌تر باشند ارزشیابی درست‌تری صورت می‌پذیرد.

به مثال زیر توجه کنید: فرض کنید معلمی می‌خواهد اثربخشی روش آموزشی خود و موفقیت دانش‌آموزانش را در درس علوم ارزشیابی کند. به این منظور، آزمونی برای درس علوم تهیه می‌کند و در اختیار دانش‌آموزان قرار می‌دهد تا به سؤالهای آن جواب دهند. بعد جوابهای

1. Worthen
2. Sanders
3. curriculum evaluator
4. goals

دانش‌آموزان را تصحیح می‌کند و با شمارش تعداد سؤالهایی که دانش‌آموزان درست جواب داده‌اند به آنان نمره می‌دهد. سپس بر پایهٔ نمرات به دست آمدهٔ دانش‌آموزان و انتظاراتی که از آنان می‌رفته است در بارهٔ مقدار یادگیری آنان و میزان توفیق خودش به داوری می‌پردازد.

در مثال بالا، مجموعه سؤالهایی که به دانش‌آموزان داده می‌شود آزمون به حساب می‌آید. اجرای آزمون و نمره گذاری آن آزمودن یا همان اندازه‌گیری به کمک آزمون نام دارد که در آن تبدیل نتایج آزمون به اعداد طبق یک قاعدهٔ مشخص (شمارش پاسخهای درست دانش‌آموزان) صورت پذیرفته است. و سرانجام داوری در بارهٔ نتایج آزمون یا نتایج اندازه‌گیری در ارتباط با موفقیت یادگیری دانش‌آموزان و توفیق آموزشی معلم ارزشیابی به شمار می‌آید. ناگفته پیداست که اگر نتایج آزمون مورد استفاده در مثال بالا مخدوش باشند، ارزشیابی انجام شده نیز قابل اعتماد نخواهد بود. بنابراین، ارزشیابیهای درست بر پایهٔ اطلاعات دقیق استوارند.

البته معلم می‌توانست، برای انجام ارزشیابی دقیق‌تری از میزان توفیق دانش‌آموزان در یادگیری و موفقیت خودش در آموزش، علاوه بر آزمون، از راههای دیگری نیز به جمع‌آوری اطلاعات بپردازد. مثلاً می‌توانست از دانش‌آموزان بخواهد تا، به صورت فردی یا در گروههای کوچک، یک پروژهٔ تحقیقی انجام دهند، در آزمایشگاه به آزمایشهای علمی بپردازند و نتایج کار خود را به صورت کتبی به معلم گزارش نمایند، یا هنگام اجرای آزمایش در آزمایشگاه عملکرد دانش‌آموزان را مشاهده کند و به کمک یک فهرست وارسی (چک لیست) یا مقیاس درجه‌بندی یادگیری آنان را تعیین کند. مجموعهٔ این اقدامات به اضافهٔ اجرای آزمون علوم سنجش یادگیری به حساب می‌آید که مبنای محکم‌تری را برای ارزشیابی فراهم می‌آورد. بنابراین، ارزشیابی آموزشی می‌تواند براساس شواهد و داده‌های گوناگون صورت پذیرد. بسته به شرایط و موقعیتهایی که با آن سروکار دارید، این شواهد و داده‌ها می‌توانند تنها کمّی باشند، کمّی و کیفی با هم باشند، یا حتی تنها کیفی باشند. نیتکو (۲۰۰۱) در این باره گفته است:

> شما می‌توانید دانش‌آموزان را براساس سنجش حاصل از مشاهدات نظام‌دار و توصیفهای کیفی ارزشیابی کنید ــ بدون آنکه به اندازه‌گیری اقدام نمایید. حتی زمانی که اطلاعات عینی [نتایج اندازه‌گیری] در دسترس هستند و مورد استفاده قرار می‌گیرند، متخصصان ارزشیابی باید آنها را با تجارب و دیدگاههای خود درهم آمیزند و آن گاه اتخاذ تصمیم کنند. بنابراین، در همهٔ ارزشیابیها، میزانی از ذهنیت، ناهمخوانی، و سوگیری وجود دارد. آزمودن و اندازه‌گیری، از آنجاکه از سایر روشهای سنجش عینی‌ترند، پاره‌ای از ناهمخوانیها و ذهنیتها را کاهش می‌دهند. اما گفتن اینکه استفاده از آزمودن و اندازه‌گیری (یا به طور کلی فنون کمّی اندازه‌گیری) ارزشیابی را خیلی زیاد بهبود می‌بخشد خود نوعی سوگیری نسبت به تکنولوژی است. (صص ۷ و ۸)

برای روشن‌تر ساختن تمایز بین سنجش و ارزشیابی، توضیحات آندرسون[1] و بورک[2] (۲۰۰۱) در رابطه با ویژگیهای عاطفی را می‌آوریم. در اینجا نیز سنجش به جمع‌آوری اطلاعات

1. Anderson 2. Burke

دربارهٔ ویژگیهای عاطفی گفته می‌شود؛ در مقابل، ارزشیابی به داوری یا قضاوت دربارهٔ ارزش آن ویژگیها یا داوری دربارهٔ مقدار آنها معنی می‌دهد.

بنابراین، در حالی که سنجش به ما می‌گوید شخص موردنظر از کدام ویژگیها برخوردار است (مثلاً، ارزشهای آموزشی این معلم کدام‌اند؟)، ارزشیابی قدر یا ارزش آن ویژگیها را مشخص می‌کند (مثلاً، آیا ارزشهای آموزشی این معلم ارزشهای درستی هستند؟). همچنین، در حالی که سنجش نشان می‌دهد که مقدار ویژگی عاطفی چقدر است (مثلاً، دانش‌آموزان به یک موضوع درسی خاص چه مقدار علاقه‌مندند؟)، ارزشیابی دربارهٔ مقدار آن ویژگی داوری می‌کند (مثلاً، آیا آن مقدار علاقه به موضوع درس موردنظر کافی است تا آن را در برنامهٔ درسی منظور کنیم؟). (ص ۷)

ارزشیابی آموزشی و پژوهش آموزشی

هم پژوهش آموزشی[1] و هم ارزشیابی آموزشی جزو فعالیتهای علمی به حساب می‌آیند، زیرا هر دو قویاً بر روشها و فنون بررسی تجربی[2] متکی‌اند. ورتن و سندرز (۱۹۸۷) در این باره گفته‌اند "اگر شخصی شاهد آزمونهایی که به دانش‌آموزان داده می‌شوند باشد یا ببیند که چگونه داده‌های مربوط به زمان صرف شده در تکلیف[3] در کلاس جمع‌آوری می‌شوند، به سختی خواهد توانست بگوید که فعالیت در حال انجام مربوط به یک کار پژوهشی است یا یک کار ارزشیابی" (ص ۲۵). با وجود این، بین پژوهش آموزشی و ارزشیابی آموزشی تفاوتهایی وجود دارند. عمده‌ترین آنها به قرار زیرند:

(۱) نخست اینکه، گرچه هم در ارزشیابی آموزشی و هم در پژوهش آموزشی اطلاعات جمع‌آوری می‌شوند و دانش کسب می‌گردد؛ با این حال، هدف ارزشیابی به‌طور عمده تصمیم‌گیری است ولی هدف پژوهش استنتاج است. به سخن دیگر، پژوهشگران عمدتاً به درک پدیده‌ها برای شناسایی بهتر آنها ولی متخصصان ارزشیابی یا ارزشیابان به فهم پدیده‌ها برای هدایت اعمال دیگران علاقه‌مندند.

(۲) دوم اینکه، در پژوهش آموزشی، مانند هر نوع پژوهش دیگر، تعمیم‌پذیری نتایج پژوهش از اهمیت خاصی برخوردار است. بنابراین، در فعالیت پژوهشی کوشش فراوان به عمل می‌آید تا یافته‌ها به موارد دیگر تعمیم یابند. در مقابل، ارزشیابی آموزشی بر یک پدیدهٔ آموزشی ویژه متمرکز است و تعمیم‌پذیری در آن نقشی ندارد. بنا به گفتهٔ پوفام[4] (۱۹۷۵) در ارزشیابی آموزشی "بر موقعیت موجود و تصمیماتی که دربارهٔ آن می‌توان گرفت تأکید می‌شود" (ص ۱۲).

(۳) سوم اینکه، ارزشیابی آموزشی با *ارزش* سروکار دارد، در حالی که پژوهش آموزشی مانند سایر پژوهشها، *خالی از ارزش* است. چنان‌که قبلاً گفته شد، هدف اصلی هرگونه فعالیت ارزشیابی، به ویژه ارزشیابی آموزشی، تعیین قدر و ارزش پدیدهٔ مورد ارزشیابی است تا اینکه به افراد

1. educational research 2. empirical inquiry 3. time-on-task
4. Popham

علاقه‌مند و مسئول کمک کند تا در بارهٔ آن پدیده تصمیمهای درستی اتخاذ نمایند. در مقابل، پژوهشگران به دنبال حقایق علمی هستند و هیچ گونه قدر و ارزش مثبت یا منفی برای حقایق به دست آوردهٔ خود قائل نمی‌شوند. به همین منظور، پس از آنکه پژوهشگر رابطهٔ بین دو یا چند پدیدهٔ مورد نظرش را کشف کرد، به هدفش دست یافته است و دیگر به دنبال ارزش‌گذاری آنها نخواهد بود؛ یعنی در مورد خوبی یا بدی یافته‌های خود به قضاوت و ارزش‌داوری نخواهد نشست.

چهارم اینکه، به اعتقاد بسیاری از صاحب‌نظران از جمله پین[1] (۱۹۹۴)، پژوهش وابسته به نظریه است اما ارزشیابی الزاماً مبنای نظری ندارد. به اعتقاد این صاحب‌نظران، هدف اصلی پژوهش تولید دانش تازه و افزودن آن به نظریه‌های موجود است. "بخش عمدهٔ مطالعات پژوهشی در علوم فیزیکی، زیست‌شناختی، و رفتاری با هدف گسترش یک نظریهٔ خاص انجام می‌گیرند، یا دست‌کم از نظریه استخراج می‌شوند.... [در حالی که] مطالعات ارزشیابی عموماً به منظور حل مشکلات عملی ویژه و کمک به تصمیم‌گیری صورت می‌پذیرند" (پین، ۱۹۹۴، ص ۱۱).

علاوه بر ارزشیابی آموزشی و پژوهش آموزشی اصطلاح **پژوهش ارزشیابی**[2] نیز در ادبیات ارزشیابی به چشم می‌خورد. روسی[3] (۱۹۸۲)، در رابطه با این مفهوم گفته است: "ارزشیابی‌ای که بر پایهٔ شواهد تجربی جمع‌آوری شده به راههایی که قابل تکرار باشند استوار است و در آن مسائل مربوط به روایی درونی، روایی بیرونی، و روایی سازه منظور می‌شود پژوهش ارزشیابی نام دارد" (ص ۶۱). به سخن دیگر، پژوهش ارزشیابی عبارت است از کاربرد روشهای پژوهش در مسائل ارزشیابی.

خلاصهٔ فصل

ما در این فصل اصطلاحات مهم حوزهٔ اندازه‌گیری و ارزشیابی آموزشی، یعنی *اندازه‌گیری*، *آزمون*، *پرسشنامه*، *آزمودن*، *سنجش*، *ارزشیابی*، *پژوهش*، و *پژوهش ارزشیابی* را تعریف کردیم. همچنین در این فصل رابطهٔ میان این اصطلاحات و شباهتها و تفاوتهای بین آنها را مورد بحث قرار دادیم. در ارتباط با تفاوت بین آزمون و پرسشنامه، گفتیم که آزمون ابزار اندازه‌گیری عملکرد بیشینه ولی پرسشنامه وسیلهٔ اندازه‌گیری عملکرد نوعی است. در مقایسهٔ میان آزمودن و سنجش و اندازه‌گیری، گفتیم که آزمودن نوعی اندازه‌گیری است که صرفاً به کمک آزمون یا پرسشنامه صورت می‌پذیرد و فقط جنبهٔ کمّی دارد، اما در سنجش هم از آزمون و سایر وسایل تبدیل اطلاعات به کمیت استفاده می‌شود و هم از روشهای کیفی.

در مقایسهٔ بین پژوهش آموزشی و ارزشیابی آموزشی، گفتیم که این دو فعالیت از این بابت که به‌طور نظامدار به جمع‌آوری داده‌ها و تفسیر آنها به منظور کسب دانش می‌کوشند هر دو جزو بررسیهای علمی به حساب می‌آیند، بنابراین می‌توانند ابزارها و روشها و فنون مشابهی را به کار بندند. اما پژوهش آموزشی و ارزشیابی آموزشی از جهت هدف و منظور با هم فرق دارند.

---

1. Payne
2. evaluation (evaluative) research
3. Rossi

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. اندازه‌گیری و آزمون دو اصطلاح مترادف‌اند.

الف ـ صحیح

ب ـ غلط

۲. کدام فعالیت زیر معرف سنجش مستقیم است؟

الف ـ تعیین تعداد رفتارهای پرخاشگرانهٔ یک کودک از راه مشاهده

ب ـ اندازه‌گیری هوشبهر یک دانش‌آموز با استفاده از یک آزمون هوشی

پ ـ سنجش میزان اضطراب یک فرد از طریق اجرای یک پرسشنامه

ت ـ برآورد توانایی یادگیرنده در حل مسائل ریاضی از طریق اجرای یک آزمون معلم ساخته

۳. کدام ویژگی آزمونهای استاندارد شده یا میزان شده تضمین می‌کند که اگر چندنفر آزمون واحدی را مورد استفاده قرار دهند نتایج یکسانی به دست خواهند آورد؟

الف ـ قواعد اندازه‌گیری

ب ـ هنجارهای آزمون

پ ـ نمرات استاندارد

ت ـ اجرای متحدالشکل

۴. کدام یک از بیانات زیر در بارهٔ اندازه‌گیری، آزمودن، و سنجش درست است؟

الف ـ اندازه‌گیری و آزمودن دو اصطلاح معادل هستند.

ب ـ اندازه‌گیری و آزمودن بخشی از سنجش یادگیری به حساب می‌آیند.

پ ـ سنجش نوعی اندازه‌گیری است که با وسایلی بجز آزمون انجام می‌شود.

ت ـ سنجش زیرمجموعهٔ اندازه‌گیری است.

۵. از اصطلاحات زیر کدام یک معنی محدودتری دارد؟

الف ـ اندازه‌گیری

ب ـ سنجش

پ ـ آزمودن

ت ـ ارزشیابی

۶. ارزشیابی آموزشی و پژوهش آموزشی در چه چیزی به هم شبیه‌اند؟

الف ـ تعمیم نتایج

ب ـ درک روابط بین امور

پ ـ ارزش داوری

ت ـ روشهای بررسی تجربی

۷. کدام مورد یا موردهای زیر ابزار یا ابزارهای کمّی اندازه‌گیری در روانشناسی و علوم تربیتی است؟

الف ـ آزمون

ب ـ پرسشنامه

پ ـ مصاحبه

ت ـ مشاهده

ث ـ الف و ب

ج ـ همهٔ موارد بالا

۸. تعریف نانالی از اندازه‌گیری را بنویسید و در بارهٔ کاربرد قواعد اندازه‌گیری توضیح دهید و برای آن مثالی ذکر نمایید.

۹. آزمون را تعریف کنید و رابطهٔ آن را با اندازه‌گیری توضیح دهید.

۱۰. روشهای اندازه‌گیری مستقیم و غیرمستقیم را توضیح دهید و برای هر یک مثالی ذکر کنید.

۱۱. سنجش را تعریف کنید و رابطهٔ آن را با اندازه‌گیری توضیح دهید.

۱۲. آزمودن را تعریف کنید و رابطهٔ آن را با اندازه‌گیری و سنجش توضیح دهید.

۱۳. ارزشیابی را تعریف کنید و رابطهٔ آن را با اندازه‌گیری، آزمودن، و سنجش توضیح دهید.

۱۴. پژوهش آموزشی و ارزشیابی آموزشی را با هم مقایسه کنید و شباهتها و تفاوتهای بین آنها را توضیح دهید.

۱۵. پژوهش ارزشیابی را تعریف کنید و رابطهٔ آن را با پژوهش آموزشی و ارزشیابی آموزشی توضیح دهید.

۲

مقیاسهای اندازه‌گیری

هدفهای یادگیری

از خواننده، انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. اصطلاح مقیاسهای اندازه‌گیری را تعریف کند.

۲. مقیاسهای اسمی، ترتیبی، فاصله‌ای، و نسبتی را با ذکر مثالی برای هریک از آنها توضیح دهد.

۳. عملیات مجاز ریاضی و آماری هریک از مقیاسهای اندازه‌گیری را توضیح دهد.

۴. ویژگیهای مقیاسهای چهارگانهٔ اندازه‌گیری را از لحاظ نحوهٔ کاربرد در موقعیتهای مختلف و با متغیرهای مختلف با هم مقایسه کند.

۵. مقیاس پیوسته و مقیاس گسسته را با ذکر مثالی برای هریک از آنها توضیح دهد.

۶. حدود واقعی اعداد را تعریف کند.

۷. گِرد کردن را توضیح دهد.

۸. اعداد اعشاری مختلفی را که به او داده می‌شود یا در محاسبات خود به دست می‌آورد طبق قاعده‌ای که در این فصل داده شده گِرد نماید.

در فصل پیش گفتیم که در اندازه‌گیریِ ویژگیها یا صفات اشخاص یا اشیاء اعداد یا نمراتی به آنها نسبت داده می‌شوند. مثلاً وقتی که یک پرسشنامهٔ نگرش‌سنج را به کار می‌بریم، برای تعیین نمرهٔ یک شخص در این پرسشنامه عددی برای او در نظر می‌گیریم. بسته به اینکه چه چیزی را اندازه می‌گیریم و چگونه آن را اندازه می‌گیریم، اعداد به دست آمده ویژگیهای متفاوتی خواهند داشت. این ویژگیهای متفاوت اعداد را مقیاسهای اندازه‌گیری می‌نامند که موضوع بحث این فصل است.

تعریف و انواع مقیاسهای اندازه‌گیری

بنا به تعریف، «**مقیاس**[1] مجموعه‌ای عدد (یا نمادهایی دیگر) است که ویژگیهای آنها بر ویژگیهای تجربی اشیایی که اعداد به آنها نسبت داده می‌شوند منطبق است» (کوهن و سوردلیک، ۲۰۰۲، ص ۷۳). **مقیاسهای اندازه‌گیری**[2] را به چهار نوع یا در چهار سطح دسته‌بندی کرده‌اند (استیونس[3]، ۱۹۵۱): **اسمی**[4]، **ترتیبی**[5]، **فاصله‌ای**[6]، **نسبتی**[7]. این طبقه‌بندی از ساده به پیچیده و به صورت سلسله مراتبی تنظیم شده است؛ یعنی اینکه، هر طبقه دارای تمامی ویژگیهای طبقات قبل به اضافهٔ ویژگیهای مخصوص به آن طبقه است. ویژگیهای مقیاسهای چهارگانهٔ اندازه‌گیری مربوط هستند به فرضهایی که در بارهٔ این مقیاسها قائل می‌شوند و نیز به اعمال ریاضی و آماری که می‌توان در بارهٔ آنها انجام داد. در زیر ویژگیهای این مقیاسها مورد بحث قرار می‌گیرند.

مقیاس اسمی

مقیاس اسمی پایین‌ترین سطح یا ساده‌ترین نوع اندازه‌گیری را معرفی می‌کند. این مقیاس به دو صورت مورد استفاده قرار می‌گیرد، یکی برای اسم‌گذاری و دیگری برای طبقه‌بندی.

۱. اسم‌گذاری. در این مقیاس از ارقام و اعداد برای اسم‌گذاری یا نام‌گذاری و تشخیص اشیاء و افراد استفاده می‌شود، بدون اینکه این اعداد هیچ گونه مفهوم و معنای ریاضی داشته باشند. برای مثال، شماره‌هایی که بر پشت پیراهن بازیکنان فوتبال نوشته می‌شوند و شمارهٔ خیابانهای شهرها مواردی از کاربرد اعداد برای اسم‌گذاری هستند.

۲. طبقه‌بندی. در مقیاس اسمی، علاوه بر نام‌گذاری، از اعداد و ارقام برای طبقه‌بندی اشیاء و افراد نیز استفاده می‌شود. باز هم در اینجا اعداد هیچ گونه مفهوم و معنای ریاضی ندارند و تنها برای مشخص کردن طبقه‌های مختلف طبقه‌بندی به کار می‌روند. فرض کنید یک زمین‌شناس انواع سنگها را در کیسه‌های مختلف با شماره‌های مختلفی دسته‌بندی می‌کند. در این دسته‌بندی،

---

1. scale
2. scales of measurement
3. Stevens
4. nominal
5. ordinal
6. interval
7. ratio

یک عدد معرف کیسهٔ حاوی سنگهای آهکی، عدد دیگر نشان‌دهندهٔ کیسهٔ محتوی سنگهای آتش‌زنه، و غیره خواهد بود. در اینجا، بجز اینکه اعداد مُعرف انواع مختلف سنگها هستند، هیچ معنی دیگری ندارند. مثال دیگری برای این مقیاس طبقه‌بندی افراد به هنجار و نابهنجار است که در آن مثلاً عدد ۱ فرد بهنجار و عدد ۲ فرد نابهنجار را معرفی می‌کند.

از آنجا که در تعریف اندازه‌گیری استفاده از قواعد ضروری فرض شده است، پاره‌ای از دانشمندان به استفاده از مقیاس اسمی به عنوان یک مقیاس اندازه‌گیری ایراد گرفته و گفته‌اند که این مقیاس فاقد قواعد اندازه‌گیری است. استیونس (۱۹۵۱) به این ایراد پاسخ داده و چنین اظهار داشته است:

> نام آن را هرچه می‌خواهید بگذارید، استفاده از اعداد برای نام‌گذاری طبقات مثالی از "نسبت دادن اعداد به اشیا و افراد بر طبق قاعده است." قاعده این است: یک عدد واحد را به طبقات مختلف نسبت ندهید و برای یک طبقهٔ واحد عددهای مختلف در نظر نگیرید. (ص ۲۶)

علاوه بر موردها و مثالهایی که در بالا از مقیاس اسمی ذکر کردیم، در علوم اجتماعی و رفتاری متغیرهای زیادی وجود دارند که به کمک مقیاس اسمی اندازه‌گیری می‌شوند، مانند جنس، نژاد، ملیت، مذهب، و مانند اینها. مثلاً جنسیت را به صورت زن = ۱؛ مرد = ۲؛ نژاد را به صورت سیاه = ۱، سفید = ۲، زرد = ۳؛ و غیره نشان می‌دهند.

عملیات مجاز آماری و ریاضی در بارهٔ مقیاس اسمی: عملیات مجاز دربارهٔ این مقیاس بسیار کم و به قرار زیرند:

عملیات مجاز آماری: شمارش فراوانی، یعنی تعیین تعداد موردهای درون هر طبقه و تعیین نما[1]، یعنی طبقه‌ای که بیشتر از دیگر طبقات عضو دارد.

عملیات مجاز ریاضی: انجام هیچ یک از چهار عمل اصلی جمع، تفریق، ضرب، و تقسیم در این مقیاس میسر نیست.

مقیاس ترتیبی

مقیاس ترتیبی یا رتبه‌ای مجموعه‌ای از **رتبه‌ها**[2] است. در این مقیاس، "(۱) دسته‌ای از افراد یا اشیاء، با توجه به یک صفت، از بزرگ به کوچک [یا بالعکس] مرتب می‌شوند، و (۲) معلوم نیست که هریک از این افراد یا اشیا، به مفهوم مطلق، چقدر از آن صفت را دارد، و نیز (۳) مشخص نیست که این اشیاء [یا افراد]، از لحاظ صفت مورد اندازه‌گیری، چه مقدار از یکدیگر

---

۱. مفهوم نما و سایر مفاهیم آماری که در این فصل از آنها نام برده می‌شود در فصل ۱۸ کتاب بحث شده‌اند.

2. ranks

فاصله دارند" (نانالی، ۱۹۷۸، ص ۱۴).

مثالی برای این مقیاس مرتب کردن دانش‌آموزان یک کلاس به ترتیب قد و شماره گذاری آنها از کوتاه‌ترین به بلندترین یا از بلندترین به کوتاه‌ترین فرد است. اگر هدف از اندازه‌گیری انتخاب بهترین فرد یا مناسب‌ترین شیء و از این قبیل باشد، این مقیاس مقیاسی مناسب است. در غیر این صورت، چون اکثر محاسبات ریاضی و آماری (مانند میانگین و انحراف معیار) در بارهٔ این مقیاس میسر نیست و بدون اینها (مخصوصاً میانگین و انحراف معیار) اندازه‌گیری چندان جنبهٔ علمی ندارد، برای سایر امور این مقیاس مفید نیست.

مثال دیگری برای مقیاس ترتیبی، پرسشنامهٔ استرس زندگی هولمز[1] و راهه[2] (۱۹۶۷) است. در این پرسشنامه تعداد تغییرات زندگی فرد در شش ماه گذشته با هم جمع می‌شوند (مثلاً ازدواج، تغییر محل سکونت، تغییر شغل، و غیره). چنین فرض می‌شود کسی که در این پرسشنامه نمرهٔ ۲۰ گرفته از کسی که نمره‌اش ۱۵ است استرس یا فشار روانی بیشتری را تجربه کرده است، که او هم از کسی که نمرهٔ ۱۰ دریافت کرده استرس بیشتری داشته است. بنابراین، با این پرسشنامه، افراد بر حسب میزان استرسی که در شش ماه گذشته تجربه کرده‌اند، که آن هم به توسط تغییرات حاصل در زندگی‌شان معلوم می‌گردد، به دنبال هم مرتب می‌شوند.

دقت کنید که دو مثال بالا، از لحاظ کاربرد اعداد یا نمرات، با هم فرق دارند. در مثال اول، رتبه‌های ۱، ۲، ۳، و ...، به کار می‌روند، در حالی که در مثال دوم اعداد مُعرف تعداد تغییرات در زندگی فرد هستند. با این حال، هر دو مُعرّفِ مقیاس ترتیبی هستند، زیرا در بارهٔ تفاوت بین نقاط مقیاس هیچ گونه اطلاعی موجود نیست؛ و این یک ویژگی مهم مقیاس ترتیبی است. برای نمونه، از روی نتایج به دست آمده از پرسشنامهٔ استرس‌سنج هولمز و راهه، ما فرض نمی‌کنیم که بین ۱۰ و ۱۵ تغییر در زندگی شخص همان مقدار استرس را تجربه می‌کند که بین ۱۵ و ۲۰ تغییر. این گونه تمایز تنها در مقیاس فاصله‌ای و نسبی یافت می‌شود.

عملیات مجاز آماری و ریاضی در بارهٔ مقیاس ترتیبی: در این مقیاس عملیات بیشتری از مقیاس اسمی مجازند. با وجود این، عملیات مورد نیاز یک مقیاس خوب اندازه‌گیری در این مقیاس نیز مجاز نیستند.

عملیات مجاز آماری: شمارش فراوانی، تعیین نما، محاسبهٔ میانه، محاسبهٔ درصدها، و محاسبهٔ ضریب همبستگی رتبه‌ای اسپیرمن[3].

عملیات مجاز ریاضی: هیچ یک از چهار عمل اصلی جمع، تفریق، ضرب، و تقسیم در این مقیاس میسر نیست.

1. Holmes
2. Rahe
3. Spearman

مقیاس فاصله‌ای

مقیاس فاصله‌ای مقیاسی است که در آن "(۱) رتبهٔ اشیاء [یا افراد]، با توجه به یک صفت مشخص است، (۲) معلوم است که اشیاء [یا افراد]، از لحاظ صفت مورد اندازه‌گیری چه مقدار از یکدیگر فاصله دارند، اما (۳) هیچ نوع اطلاعی راجع به مقدار مطلق صفت مورد نظر برای هیچ یک از اشیاء [یا افراد] در دست نیست" (نانالی، ۱۹۷۸، ص ۱۶). (چنان که از این تعریف معلوم است، وجه تمایز این مقیاس از مقیاس ترتیبی این است که در این مقیاس فاصله‌های بین واحدها معلوم و برابر هستند، در صورتی که در مقیاس ترتیبی فاصلهٔ بین واحدها معلوم نیست.)

اگر اندازهٔ قد دانش‌آموزان یک کلاس را در رابطه با قد کوچک‌ترین دانش‌آموز کلاس، نه با توجه به اندازهٔ واقعی قد آنها، اندازه‌گیری کنیم، یک مقیاس فاصله‌ای به دست خواهد آمد. در این مقیاس، کوتاه‌ترین دانش‌آموز نمرهٔ صفر خواهد گرفت، دانش‌آموزی که ۵ سانتی‌متر از دیگری بلندتر است نمرهٔ ۵، آنکه قدش از کوتاه‌ترین دانش‌آموز ۱۲ سانتی‌متر بلندتر است نمرهٔ ۱۲ می‌گیرد، و الی آخر. همچنین می‌توان فاصلهٔ قد این دانش‌آموزان را از قد متوسط دانش‌آموزان کلاس اندازه گرفت، بدین طریق که قد متوسط را صفر انتخاب کرد و کسانی که قد بلندتر از متوسط دارند را با اعداد مثبت و دانش‌آموزانی را که قد کوتاه‌تر از متوسط دارند با اعداد منفی مشخص کرد.

مثال دیگری برای مقیاس فاصله‌ای مقیاس به کار گرفته شده در دماسنجهای سانتی‌گراد و فارنهایت است. در هر دوی این دماسنجها صفر انتخابی یک صفر قراردادی است، نه یک صفر واقعی، اما فاصلهٔ ۵، ۱۰، ۱۵، و غیره در تمام طول مقیاس همواره یکسان است، یعنی مثلاً تفاوت دما بین ۱۰ و ۲۰ درجه برابر است با تفاوت دمای بین ۴۰ و ۵۰ درجه.

عملیات مجاز آماری و ریاضی دربارهٔ مقیاس فاصله‌ای: از آنجاکه دراین مقیاس فاصله‌های بین واحدها برابر هستند و در نتیجه این مقیاس از یکی از ویژگیهای مهم یک مقیاس خوب اندازه‌گیری برخوردار است، لذا همهٔ محاسبات آماری و غالب محاسبات ریاضی را در بارهٔ آن می‌توان انجام داد.

عملیات مجاز آماری: محاسبهٔ نما، میانه، میانگین، انحراف معیار، ضریب همبستگی رتبه‌ای اسپیرمن، و ضریب همبستگی گشتاوری پیرسون[1].

عملیات مجاز ریاضی: جمع و تفریق مجاز است؛ ضرب و تقسیم مجاز نیست.

مقایسهٔ مقیاس ترتیبی با مقیاس فاصله‌ای

چنان‌که دیدیم، بین مقیاس ترتیبی و مقیاس فاصله‌ای شباهتهای زیادی وجود دارند. فرض کنید در یک پرسشنامهٔ رضایت شغلی سؤالی نظیر این سؤال مطرح می‌شود: "شغل من رضایت‌بخش

1. Pearson

است"، که پاسخ‌دهنده برای آن اعداد از ۱ تا ۵ را که معرف کاملاً مخالفم (۱) مخالفم (۲)، نظری ندارم (۳)، موافقم (۴)، و کاملاً موافقم (۵) هستند بر می‌گزیند. این طبقات پاسخ، ترتیب توافق را نشان می‌دهد، اما در بارهٔ فاصلهٔ بین طبقات که لازمهٔ مقیاس فاصله‌ای است هیچ اطلاعی در اختیار ما نمی‌گذارد.

ابزارهای اندازه‌گیری نظیر پرسشنامه‌های نگرش‌سنج که در آن از سؤالهای نوع بالا استفاده می‌شود مقیاسهای ترتیبی هستند، اما معمولاً آنها را مانند مقیاسهای فاصله‌ای مورد استفاده قرار می‌دهند. از جمله اینکه میانگین نمرات همهٔ پاسخ‌دهندگان به یک سؤال را محاسبه و مثلاً رقم ۲/۷ را گزارش می‌کنند. یا اینکه مجموعهٔ سؤالها را جمع می‌کنند و میانگین هر یک از پاسخ‌دهندگان و نیز اندازه‌های تغییرپذیری کلی را محاسبه می‌نمایند. یعنی عملاً با مقیاس ترتیبی مانند مقیاس فاصله‌ای رفتار می‌شود. این عمل موجب بروز اختلاف نظرهایی در میان متخصصان روانسنجی شده است.

بروگاتا[1] و بوهرنستد[2] (۱۹۸۰) استدلال کرده‌اند که مقیاس ترتیبی شکل ضعیفی از مقیاس فاصله‌ای است که در آن اطلاعات مربوط به فاصلهٔ بین نقاط مقیاس "گم شده است." از آنجا که متغیرهای ترتیبی عموماً با آن دسته از شاخصهای آماری توصیف می‌شوند که فرض می‌کنند مقیاس مورد نظر پیوسته و عددی است، بروگاتا و بوهرنستد معتقدند که بهتر است متغیرهای ترتیبی را به صورت مقیاس فاصله‌ای ناکامل تصور کنیم تا اینکه آنها را سطح مستقلی از اندازه‌گیری بدانیم.

مقیاس نسبتی

مقیاس نسبتی یا نسبی مقیاسی است که در آن "(۱) رتبهٔ اشخاص، با توجه به یک صفت معلوم است، (۲) فاصلهٔ بین اشخاص معلوم است، (۳) علاوه بر اینها، فاصلهٔ حداقل یکی از اشخاص از یک صفر مطلق نیز معلوم است" (نانالی، ۱۹۷۸، ص ۱۷). مثلاً اگر قد دانش‌آموزان یک کلاس را با توجه به اندازهٔ واقعی قد آنها، نه در رابطه با قد کوتاهترین فرد یا فرد متوسط، تعیین کنیم مقیاس حاصل یک مقیاس نسبی است که دارای هر سه ویژگی بالاست. مهم‌ترین مقیاس نسبتی همان مقیاس اعداد از صفر تا بینهایت است.

اندازه‌گیری طول، وزن، حجم، و سایر صفات فیزیکی اشیاء با مقیاس نسبتی عملی است، اما اندازه‌گیری متغیرهای روانی و پرورشی با این مقیاس به ندرت میسر است. نانالی (۱۹۷۸) در این باره چنین اظهارنظر کرده است:

موارد بسیار کمی وجود دارند که در آنها تصور وجود مقیاسهای نسبی در اندازه‌گیریهای روانی معقول باشد. در مواردی که نمرات بر حسب خواص سادهٔ فیزیکی، مانند زمان و طول بیان می‌شوند،

1. Brogatta
2. Bohrnstedt

عملیات مجاز آماری و ریاضی دربارهٔ مقیاس نسبی: در این مقیاس همهٔ عملیات آماری و ریاضی مجاز است. در جدول ۱-۲ ویژگیهای مهم مقیاسهای اندازه‌گیری فوق داده شده‌اند.

مقایسهٔ مقیاسهای اندازه‌گیری با یکدیگر

برای روشن شدن تمایز بین مقیاسهای مختلف اندازه‌گیری، فرض کنید می‌خواهیم نظر بینندگان تلویزیون را در بارهٔ سریالها بپرسیم. نوع سؤالی که طرح می‌کنیم مقیاس معینی را مشخص می‌کند، مانند نمونه‌های زیر:

اسمی

کدام یک از سریالهای تلویزیونی را تماشا می‌کنید؟

سریال الف	سریال ب	سریال پ
۱	۲	۳

ترتیبی

سریال تلویزیونی که تماشا می‌کنید تا چه حد منعکس کنندهٔ واقعیتهای زندگی است؟

خیلی زیاد	زیاد	متوسط	کم	خیلی کم
۵	۴	۳	۲	۱

فاصله‌ای

هوشبهر (IQ) قهرمان سریال مورد علاقهٔ خود را حدس بزنید.

۹۲ ۹۴ ۹۶ ۹۸ ۱۰۰ ۱۰۲ ۱۰۴ ۱۰۶ ۱۰۸ ۱۱۰

نسبتی

هفته‌ای چند ساعت سریال تلویزیونی تماشا می‌کنید؟

۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹ ۱۰

جدول ۱-۲ خلاصهٔ ویژگیهای مهم مقیاسهای چهارگانهٔ اندازه‌گیری

مقیاس	ویژگیها				
	وجود ترتیب در طبقات	فواصل مساوی طبقات	وجود صفر مطلق	عملیات مجاز آماری	عملیات مجاز ریاضی
اسمی	–	–	–	نما (تعیین طبقه‌ای که بیشتر از بقیه عضو دارد)، فراوانی (تعیین تعداد اعضای هر طبقه)	هیچ یک از چهار عمل اصلی
ترتیبی	+	–	–	نما، فراوانی، میانه، ضریب همبستگی اسپیرمن	هیچ یک از چهارعمل اصلی
فاصله‌ای	+	+	–	همهٔ عملیات آماری	جمع، تفریق
نسبتی	+	+	+	همهٔ عملیات آماری	همهٔ عملیات ریاضی

هر چند که چهار مقیاس اندازه‌گیری پیش گفته از لحاظ تعریف و ویژگیها از هم قابل تمایزند، با این حال در کار بست به مسائل انسانی غالباً به این خوبی تمایز پذیر نیستند و لذا در استفاده از آنها باید بسیار دقیق بود. در صفحات پیش مشکل تمایز پذیری بین مقیاس ترتیبی و مقیاس فاصله‌ای را عنوان کردیم. در اینجا به مشکلات دیگری در بارهٔ این مقیاسها اشاره می‌کنیم. موقعیتی را در نظر بگیرید که در آن از گروهی دانش‌آموز آزمونی به عمل آورده‌ایم. نتایج این آزمون به هیچ وجه معرف مقیاس نسبتی نیست، زیرا در صورت پذیرفتن چنین فرضی باید بپذیریم که دانش‌آموزی که در این آزمون نمرهٔ صفر گرفته در بارهٔ مطلبی که آزمون از آن تهیه شده است هیچ اطلاعی ندارد و دانش‌آموزی که نمرهٔ ۸۰ گرفته است دو برابر دانش‌آموزی که نمرهٔ ۴۰ گرفته است اطلاع دارد، و الی آخر.

ممکن است محتاطانه‌تر عمل کنیم و بگوییم که مقیاس حاصل یک مقیاس فاصله‌ای است که در آن نمرهٔ صفر یک رقم قراردادی است (دانش‌آموزانی که نمرهٔ صفر گرفته‌اند چیزهایی در بارهٔ موضوع درس می‌دانند اما آزمون ما نتوانسته است آن اطلاعات را بسنجد)، ولی تفاوتهای مساوی نمرات معرف تفاوتهای مساوی در دانش افراد است. حتی می‌توان عاقلانه‌تر استدلال کرد و گفت که مقیاس ما یک مقیاس ترتیبی است که در آن نمرهٔ ۹۰ از ۸۵ بهتر است و ۸۵ نیز از ۷۵ بهتر است، اما تفاوتهای مساوی نمرات معرف تفاوتهای مساوی در اطلاعات دانش‌آموزان نیستند. به سخن دیگر، تفاوت بین دانش‌آموزانی که نمرات ۸۰ و ۷۰ گرفته‌اند از لحاظ مقدار اطلاعاتی که در بارهٔ موضوع مورد اندازه‌گیری دارند برابر تفاوت بین اطلاعات دانش‌آموزانی که نمرات ۴۰ و ۳۰ گرفته‌اند نیست.

بنا به توضیحات فوق، تصمیم‌گیری در بارهٔ اینکه چه نوع مقیاسی مورد استفادهٔ ما قرار دارد وابسته است به نحوهٔ استفادهٔ ما از آن مقیاس. به عنوان مثالی روشن‌تر از این مطلب، دمای خانه‌تان را در نظر بگیرید. ما قبلاً گفتیم که دماسنج نوعی وسیلهٔ اندازه‌گیری است که اندازه‌های حاصل از آن معرف مقیاس فاصله‌ای هستند. یعنی صفر آن قراردادی است، اما فاصله‌های مساوی بین درجات مختلف در سراسر مقیاس معرف دماهای مساوی هستند. البته این در صورتی درست است که ما دَما را به عنوان شاخصی از فعالیت ملکولی (ویژگی فیزیکی دَما) بسنجیم. اما اگر دَما را به عنوان شاخص آسایش (تأثیر آن بر اشخاص) در نظر بگیریم آن وقت مقیاس حاصل دیگر فاصله‌ای نیست بلکه یک مقیاس ترتیبی خواهد بود. برای شخصی که در اتاقی با دَمای ۱۷ درجه نشسته است، تغییر دَما از ۱۷ به ۱۸ درجه قابل تشخیص و مطلوب است، اما برای شخصی که در یک اتاق با دَمای ۲۳ درجه نشسته است، تغییر دما از ۲۳ به ۲۴ درجه قابل تشخیص نخواهد بود. کوتاه سخن اینکه، متغیر زیربنایی (مثلاً ویژگی فیزیکی یا آسایش) تعیین می‌کند که چه مقیاسی مورد استفاده قرار دارد، نه اعداد یا ارقام.

با توجه به توضیحات بالا، از آنجاکه در بارهٔ نوع مقیاس مورد استفاده در میان صاحب‌نظران توافق قطعی وجود ندارد، به شخص استفاده کننده از وسیلهٔ اندازه‌گیری مربوط است که تعیین کند چه مقیاسی را مورد استفاده قرار دهد. بنابراین، لازم است متخصص اندازه‌گیری پیش از تصمیم‌گیری در بارهٔ نوع مقیاس مورد استفاده مسئلهٔ مورد اندازه‌گیری و متغیرهای زیربنایی را به دقت بررسی نماید.

مقیاس پیوسته و مقیاس گسسته

علاوه بر مقیاسهای چهارگانهٔ اندازه‌گیری که توضیح داده شدند، دو نوع دیگر مقیاس اندازه‌گیری وجود دارند که در زیر آنها را معرفی می‌کنیم: این مقیاسها با عنوانهای **پیوسته**[1] و **گسسته**[2] شناخته شده‌اند.

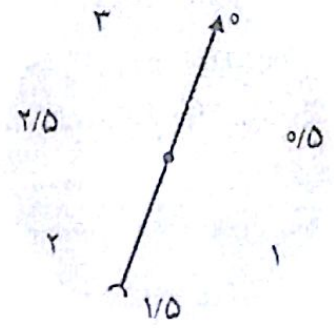
مقیاس گسسته

برای معرفی مقیاس گسسته مثالی می‌آوریم. فرض کنید از ما خواسته شده است تا تعداد کودکان هر خانوادهٔ ایرانی را تعیین کنیم. مقیاس اندازه‌گیری ما از صفر (بدون کودک) شروع خواهد شد و یک به یک تا مثلاً پانزده یا بیست پیش خواهد رفت. دقت کنید که در این مثال ارزش مقیاس از صفر تا بیست ما هر بار با یک عدد درست افزایش می‌یابد، یعنی ۰، ۱، ۲، ۳، ...، ۲۰. هر چند که در معرفی کُلی ممکن است بگوییم که متوسط خانوادهٔ ایرانی مثلاً ۳/۴ کودک دارد، اما مسلماً هیچ خانواده‌ای یافت نمی‌شود که این مقدار کودک داشته باشد، زیرا تعداد کودکان هر خانواده همیشه یک عدد درست است نه یک عدد اعشاری.

این گونه مقیاسها که در آنها متغیر مورد نظر تنها می‌تواند یکی از ارزشهای معین را به خود اختصاص دهد مقیاسهای گسسته یا مجزا[3] نام دارند. ویژگی مهم این نوع مقیاسها ***تساوی واحدهای شمارش*** آنها در طول مقیاس است.

مثلاً اگر ما تعداد کودکان یک خانواده را مطالعه می‌کنیم، صرف‌نظر از سن و جنس و سایر ویژگیها، هریک از کودکان در فراهم آوردن یک واحد شمارش با کودکان دیگر همتراز است. بنابراین، انجام تمامی عملیات اصلی ریاضی، یعنی جمع، تفریق، ضرب، و تقسیم، با این مقیاسها مجاز هستند. در نتیجه می‌توان گفت خانواده‌ای که چهار فرزند دارد دو برابر خانوادهٔ دو فرزندی فرزند دارد. متغیرهای دیگری که مُعرف مقیاس گسسته هستند عبارت‌اند از جنس (مذکر یا مؤنث)، وضعیت تأهل (مجرد یا متأهل)، سطح تحصیلات (۱ تا ۲۰ سال)، یا تعداد سؤالهایی که هریک از آزمون‌شوندگان در یک آزمون درست جواب می‌دهند.

1. continuous 2. discontinuous 3. discrete



شکل ۱-۲ نمونه‌ای از مقیاس گسسته که در آن متغیرها با نصف واحد تغییر می‌کنند (رونیون[1] و هابر[2]، ۱۹۷۶، ص ۲۴)

نکتهٔ دیگری را که باید در بارهٔ مقیاس گسسته تذکر بدهیم این است که این نوع مقیاس الزاماً با اعداد درست سروکار ندارد. صفحهٔ مدرج شکل ۱-۲ معرف مقیاس گسسته است، با این حال هریک از هفت رقم روی این صفحه از رقم قبلی و بعدی ۰/۵ رقم فاصله دارد. نکتهٔ اساسی در مثال صفحهٔ مدرج شکل ۱-۲ این است که متغیر مورد نظر نمی‌تواند بین ارقام ۰ و ۰/۵، ۰/۵ و ۱، ۱ و ۱/۵، و غیره دارای ارزشی باشد، درست همان‌طور که در مثال بالا هیچ خانواده‌ای نمی‌تواند بین اعداد اصلی ۱، ۲، ۳، ...، و ۲۰ ارزش دیگری که معرف تعداد فرزندان آن خانواده باشد به خود اختصاص دهد.

مقیاس پیوسته

برخلاف مقیاس گسسته یا مجزا، در مقیاس پیوسته متغیر مورد نظر می‌تواند در فاصلهٔ بین نقاط مقیاس هر تعدادی ارزش را به خود اختصاص دهد. باز هم به ارزشهای روی صفحهٔ مدرج شکل ۱-۲ (۰ تا ۳) نگاه کنید. اگر امکان داشت که این متغیر ارزشهای مختلف ۱/۷۵، ۲/۰۴، ۲/۳۰۴، و غیره را به خود اختصاص دهد در آن صورت با یک مقیاس پیوستهٔ اندازه‌گیری سروکار داشتیم. متغیرهایی که مُعرف مقیاس پیوسته هستند عبارت‌اند از سرعت، زمان، طول، وزن، هوش، پیشرفت تحصیلی، استعداد ریاضی، و نگرش.

لازم به گفتن است که گرچه اندازه‌گیری از متغیرهای گسسته یا مجزا همواره دقیق و قطعی است، اندازه‌های به دست آمده از متغیرهای پیوسته همیشه تقریبی هستند. مثلاً اگر بخواهیم قد دانش‌آموزان یک دبیرستان را اندازه بگیریم، هرگونه وسیلهٔ اندازه‌گیری که انتخاب کنیم تقریبی خواهد بود، زیرا همیشه تصور وسیلهٔ دقیق‌تری امکان‌پذیر است. یعنی اگر ما اندازهٔ قد دانش‌آموزان را تا دهم اعشاری اندازه‌گیری کرده باشیم، می‌توان مقیاس دیگری در نظر گرفت که قد این افراد را تا صدم اعشاری اندازه بگیرد.

ویژگی مهم مقیاسهای پیوسته *تساوی واحدهای اندازه‌گیری* است. بنابراین، اگر ما طول یک میز را به سانتیمتر اندازه می‌گیریم، در سراسر طول مقیاس یک سانتیمتر همواره یک چیز معنی می‌دهد.

1. Runyon 2. Haber

حدود واقعی اعداد

در توضیح مقیاس پیوسته، گفتیم که متغیرهای پیوسته می‌توانند تعداد بیشماری ارزش بین عددی به خود اختصاص دهند. بنابراین، ما هرگز نمی‌توانیم ارزش دقیق یک اندازهٔ بخصوص را مشخص کنیم، زیرا همیشه امکان دارد که یک وسیلهٔ اندازه‌گیریِ حساس‌تر قدری بر دقت اندازه‌گیری ما بیفزاید. به همین دلیل، گفتیم که ارزشهای عددی متغیرهای پیوسته همواره تقریبی هستند. با وجود این، می‌توان حدودی را تعیین کرد که در آن ارزش واقعی قرار دارد. بنا به تعریف، "**حدود واقعی**[1] ارزش عددی یک متغیر پیوسته برابر است با آن عدد به اضافه و منهای نصف واحد اندازه‌گیری" (رونیون و هابر، ۱۹۷۶، ص ۲۴).

باز هم برای توضیح دادن مطلب به چند مثال توجه کنید. فرض کنید با یک وزنهٔ خانگی وزن خودتان را تعیین می‌کنید. وقتی که روی این وسیلهٔ اندازه‌گیری قرار می‌گیرید عقربهٔ آن بین دو رقم قرار می‌گیرد. با این حال شما وزن خودتان را بر اساس نزدیک‌ترین رقم تعیین می‌کنید. مثلاً اگر عقربهٔ وسیلهٔ اندازه‌گیری وزن شما به طور تقریب در ربع سوم فاصلهٔ بین ۶۰ و ۶۱ کیلوگرم قرار بگیرد، شما وزن خودتان را ۶۱ کیلوگرم گزارش می‌کنید. در اینجا حدود واقعی وزن شما بین ۶۰/۵۰ و ۶۱/۵۰ قرار دارد. به قول رونیون و هابر (۱۹۷۶) اگر بخواهید وزن یک نهنگ را اندازه بگیرید، در آن صورت به جای یک کیلوگرم یک واحد بزرگتر اندازه‌گیری مثلاً ۱۰۰ کیلوگرم را انتخاب خواهید کرد. در نتیجه اگر وزن یک نهنگ را ۳۲۰۰۰ کیلوگرم گزارش کنید، منظورتان این است که نهنگ بین ۳۱۹۵۰ و ۳۲۰۵۰ کیلوگرم وزن دارد.

گرد کردن

یکی از مشکلاتی که متخصصان اندازه‌گیری در کار با اعداد غالباً با آن روبه‌رو می‌شوند تعداد زیاد ارقام اعشاری است. مثلاً اگر بخواهیم کسر $\frac{1}{3}$ را به صورت اعشاری در آوریم، عدد حاصل رقم ۰/۳۳۳۳۳+ با بینهایت ۳ بعد از علامت اعشاری خواهد بود. در چنین حالتی ما باید تصمیم بگیریم که اعشار را در جایی خاتمه دهیم. برای این منظور لازم است دو سؤال زیر را جواب دهیم:

۱. تا چند رقم اعشاری باید پیش برویم؟
۲. چگونه تصمیم بگیریم که آخرین رقم اعشاری چه رقمی باشد؟

برای سؤالهای بالا، رونیون و هابر (۱۹۷۶) قاعدهٔ زیر را پیشنهاد داده‌اند: "پس از هر عمل ریاضی، تا سه رقم اعشاری بیشتر از آنچه در اصل وجود داشته است پیش بروید و تا دو رقم اعشاری گرد کنید" (ص ۲۵). بنابراین قاعده، اگر عدد اصلی بدون اعشاری باشد، نتیجهٔ عملمان را

1. true limits

باید تا سه رقم اعشاری پیش ببریم و در رقم دوم گِرد کنیم. اگر عدد اصلی دارای یک رقم اعشاری باشد، تا چهار رقم اعشاری پیش می‌رویم و در رقم سوم گرد می‌کنیم، و الی آخر. برای گِرد کردن[1] اگر رقم بعداز رقم مورد نظر بیشتر از ۵ باشد به رقم مورد نظر یک رقم می‌افزاییم. اگر رقم بعد از رقم مورد نظر کمتر از ۵ باشد، رقم مورد نظر بدون تغییر باقی می‌ماند. به مثالهای زیر توجه کنید:

۶/۵۴۶ ⟶ ۶/۵۵
۶/۵۴۳ ⟶ ۶/۵۴
۱/۹۶۷ ⟶ ۱/۹۷
۱/۵۳۴ ⟶ ۱/۵۳

حال باید پرسید اگر رقمی که قرار است گِرد شود ۵ باشد چه اتفاقی می‌افتد؟ اگر بعد از ۵ ارقام اعشاری دیگری وجود داشته باشند، در آن صورت رقم مورد نظر در واقع بیشتر از ۵ است و طبق قانون فوق باید به رقم اعشاری قبلی یک رقم اضافه کنیم. اما اگر دقیقاً ۵ داشته باشیم، یعنی بعد از آن رقم دیگری وجود نداشته باشد یا ارقام بعدی همه صفر باشند دو حالت پیش می‌آید: (۱) رقم قبل از پنج یک رقم زوج است، در آن صورت چیزی به آن اضافه نمی‌کنیم، (۲) رقم قبل از ۵ فرد است، در این حالت ۱ رقم به آن می‌افزاییم.

۶/۵۴۵۰۰۱ ⟶ ۶/۵۵
۱۶/۸۹۵۰۱ ⟶ ۱۶/۹۰
۱/۹۶۵۲ ⟶ ۱/۹۷
۶/۵۴۵۰۰۰ ⟶ ۶/۵۴
۰/۰۰۵۰۰ ⟶ ۰/۰۰
۰/۰۱۵۰۰ ⟶ ۰/۰۲

خلاصهٔ فصل

مقیاسهای مهم اندازه‌گیری مورد بحث در این فصل عبارت‌اند از *مقیاس اسمی*، *مقیاس ترتیبی*، *مقیاس فاصله‌ای*، و *مقیاس نسبتی*. مقیاس اسمی صرفاً برای نام‌گذاری و طبقه‌بندی به کار می‌رود و در آن هیچ نوع عملیات ریاضی و آماری بجز تعیین نما و فراوانی مجاز نیست. مقیاس ترتیبی نشان‌دهندهٔ رتبه‌های افراد یک گروه است و در آن عملیات آماری شمارش فراوانی، تعیین نما، محاسبهٔ میانه، درصدها، و ضریب همبستگی رتبه‌ای اسپیرمن مجاز است، اما هیچ‌یک از اعمال ریاضی مجاز نیست. در مقیاس فاصله‌ای، علاوه بر رتبهٔ افراد، فاصلهٔ بین نمرات افراد نیز مشخص است. در این

1. rounding

مقیاس کلیهٔ عملیات آماری مجاز است، اما از عملیات ریاضی فقط جمع و تفریق مجاز است. مقیاس نسبتی کامل‌ترین مقیاس اندازه‌گیری است که معرف اعداد طبیعی است. در این مقیاس تمامی عملیات آماری و ریاضی مجاز است. حدود واقعی ارزش عددی یک متغیر پیوسته برابر است با آن عدد به اضافه و منهای نصف واحد اندازه‌گیری. ما در این فصل، به دنبال معرفی مقیاسهای چهارگانهٔ فوق، دو مقیاس معروف دیگر یعنی *مقیاس گسسته* و *مقیاس پیوسته* را نیز معرفی کردیم. در مقیاس گسسته طبقات مقیاس از هم مجزا هستند و معمولاً واحد به واحد جلو می‌روند و در آن اعشاری وجود ندارد، اما در مقیاس پیوسته متغیر مورد اندازه‌گیری می‌تواند در فاصلهٔ بین نقاط مقیاس هر تعدادی ارزش را به خود بگیرد. بنابراین در این مقیاس اعداد اعشاری وجود دارند. در آخرین قسمت فصل، *حدود واقعی اعداد* و *گِردکردن* را توضیح دادیم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. چه عاملی تعیین می‌کند که مقیاس اندازه‌گیری ما از کدام نوع باشد؟

الف ـ اعداد و ارقام به دست آمده

ب ـ چگونگی اندازه‌گیری

پ ـ رابطه اعداد و ارقام با یکدیگر

ت ـ متغیر مورد سنجش

ث ـ ب و ت

۲. در کدام یک از مقیاسهای اندازه‌گیری زیر همهٔ اعمال آماری مجاز است؟

الف ـ اسمی

ب ـ ترتیبی

پ ـ فاصله‌ای

ت ـ نسبتی

ث ـ فاصله‌ای و نسبتی

۳. اگر برای دسته‌بندی دانش‌آموزان یک ناحیه پسران را با ۱ و دختران را با ۰ نشان دهیم، به این سطح از اندازه‌گیری چه می‌گویند؟

الف ـ اسمی

ب ـ ترتیبی

پ ـ فاصله‌ای

ت ـ نسبتی

۴. رقم ۰/۰۲۵۰۰ پس از گرد کردن چند می‌شود؟

الف ـ ۰/۰۲

ب ـ ۰/۰۳

پ ـ ۰/۰۲۵۰۰

ت ـ ۰/۰۲۵

۵. اصطلاح مقیاس اندازه‌گیری را تعریف کنید.

۶. در مثالهای زیر ابتدا مقیاس اندازه‌گیری را تعیین کنید و بعد مشخص نمایید که متغیر گسسته است یا پیوسته.

الف ـ مسافتی که ۴۰ دانشجوی یک کلاس از منزل خود تا دانشکده طی می‌کنند.

ب ـ تعداد کودکانی که در ساعتهای مختلف روز متولد می‌شوند.

پ ـ تعداد رأی‌هایی که سه نمایندهٔ رئیس‌جمهوری دریافت کرده‌اند.

ت ـ تعداد دانشجویان سالهای مختلف یک دانشکده.

۷. مقیاسهای چهارگانهٔ اندازه‌گیری را با هم مقایسه کنید و شباهتها و تفاوتهای میان آنها را توضیح دهید.

۸. مقیاس پیوسته و گسسته را با ذکر مثالی برای هریک از آن دو با هم مقایسه کنید.

۹. حدود واقعی اعداد را تعریف کنید.

۱۰. اعداد زیر را تا دو رقم اعشاری گرد کنید.

الف ـ ۹۹/۹۹۵۰۰

ب ـ ۴۶/۴۰۵۰۱

پ ـ ۲/۹۶۵۰۰

ت ـ ۰/۰۰۵۰۱

ث ـ ۱۶/۴۶۵۰۰

ج ـ ۱/۰۵۴۹۹

چ ـ ۸۶/۲۱۳۹

ح ـ ۱۰/۰۰۵۰

۳

دسته‌بندی روشهای اندازه‌گیری و سنجش مورد استفاده در روانشناسی و آموزش و پرورش

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. انواع روشهای اندازه‌گیری و سنجش را با توجه به ویژگیهای مورد سنجش دسته‌بندی کند.
۲. تفاوتهای عمدهٔ بین سنجش توانایی و سنجش عاطفی را توضیح دهد.
۳. انواع آزمونهای توانایی را توضیح دهد و کاربردهای هریک از آنها را بیان کند.
۴. ویژگیهای مشترک و متفاوت آزمونهای استعداد و آزمونهای پیشرفت تحصیلی را توضیح دهد.
۵. انواع آزمونهای پیشرفت تحصیلی را توضیح دهد.
۶. انواع روشهای سنجش عاطفی را با ذکر مثال توضیح دهد.
۷. تفاوت بین روشهای سنجش علاقه و نگرش را توضیح دهد.
۸. دسته‌بندی روشهای سنجش را با توجه به چگونگی تهیهٔ آنها توضیح دهد.
۹. تفاوتها و شباهتهای بین آزمونهای میزان شده و معلم‌ساخته را توضیح دهد.
۱۰. موارد استفادهٔ آزمونهای میزان‌شده و معلم ساخته را در تصمیم‌گیریهای آموزشی توضیح دهد.

۳

دسته‌بندی روشهای اندازه‌گیری و سنجش مورد استفاده در روانشناسی و آموزش و پرورش

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. انواع روشهای اندازه‌گیری و سنجش را با توجه به ویژگیهای مورد سنجش دسته‌بندی کند.

۲. تفاوتهای عمدهٔ بین سنجش توانایی و سنجش عاطفی را توضیح دهد.

۳. انواع آزمونهای توانایی را توضیح دهد و کاربردهای هریک از آنها را بیان کند.

۴. ویژگیهای مشترک و متفاوت آزمونهای استعداد و آزمونهای پیشرفت تحصیلی را توضیح دهد.

۵. انواع آزمونهای پیشرفت تحصیلی را توضیح دهد.

۶. انواع روشهای سنجش عاطفی را با ذکر مثال توضیح دهد.

۷. تفاوت بین روشهای سنجش علاقه و نگرش را توضیح دهد.

۸. دسته‌بندی روشهای سنجش را با توجه به چگونگی تهیهٔ آنها توضیح دهد.

۹. تفاوتها و شباهتهای بین آزمونهای میزان شده و معلم‌ساخته را توضیح دهد.

۱۰. موارد استفادهٔ آزمونهای میزان‌شده و معلم ساخته را در تصمیم‌گیریهای آموزشی توضیح دهد.

ما در فصل نخست این کتاب گفتیم که متداول‌ترین روشهای اندازه‌گیری ویژگیها یا صفات روانی افراد *آزمون* و *پرسشنامه*اند. هرچند که در آموزش و پرورش و روانشناسی، علاوه بر آزمون و پرسشنامه، روشهای دیگر اندازه‌گیری و سنجش مورد استفاده قرار دارند، با این حال، آزمون و پرسشنامه از ابزارهای بسیار متداول‌اند. از تعیین میزان آموخته‌های دانش‌آموزان و دانشجویان گرفته تا برآورد ویژگیهای شخصیتی آنان، و تعیین نگرشها و علاقهٔ ایشان نسبت به امور مختلف همواره از ابزارهایی که عموماً نام آزمون و پرسشنامه بر آنها نهاده شده است استفاده می‌شود. بنابراین، آزمونها و پرسشنامه‌های مورداستفاده در روانشناسی و آموزش و پرورش بسیار متنوع‌اند و به همین منظور ما فصل حاضر را به دسته‌بندی انواع آنها اختصاص داده‌ایم.

دسته‌بندی روشهای اندازه‌گیری با توجه به ویژگیهای مورد اندازه‌گیری

روشهای اندازه‌گیری را می‌توان با توجه به ویژگیهای گوناگونی که اندازه می‌گیرند ابتدا به دو دستهٔ کلی **آزمونهای سنجش توانایی**[1] و **پرسشنامه‌های سنجش عاطفی**[2] تقسیم کرد و بعد هر یک از این دو دسته ابزار اندازه‌گیری را نیز به اجزای کوچکتری به شرحی که در زیر خواهد آمد دسته‌بندی نمود

آزمونهای اندازه‌گیری توانایی

آزمونهای توانایی، آموخته‌ها، مهارتها، و استعدادهای افراد را می‌سنجند. دسته‌ای از این آزمونها، تواناییهای **شناختی**[3] افراد را می‌سنجند و تعدادی دیگر تواناییهای **روانی ـ حرکتی**[4] را اندازه می‌گیرند. هریک از این دو دسته آزمون کاربرد خاص خود را دارد. آزمونهای تواناییهای شناختی بیشتر در زمینه‌های تحصیلی و فعالیتهای فکری و ذهنی مورد استفاده قرار می‌گیرند، اما آزمونهای روانی ـ حرکتی غالباً در موقعیتهای صنعتی و نظامی و تربیت‌بدنی به کار می‌روند. صرف‌نظر از کاربردهای اختصاصی، این دو دسته آزمون، به طور عمده، برای تعیین جنبه‌های مختلف تواناییهای افراد تهیه و اجرا می‌شوند. ویژگی مهم آزمونهای توانایی (شناختی و روانی ـ حرکتی) این است که هنگام اجرای آنها از آزمون‌شونده خواسته می‌شود تا حداکثر سعی خود را به کاربندد تا بهترین نمرهٔ ممکن را به دست آورد.

آزمونهای توانایی به دو دستهٔ آزمونهای **استعداد**[5] و آزمونهای **پیشرفت تحصیلی**[6] تقسیم می‌شوند. آزمونهای استعداد میزان توانایی یا آمادگی فرد را برای انجام دادن کارهایی که در پیش

1. ability
2. affective
3. cognitive
4. psychomotor
5. aptitude
6. academic (scholastic) achievement

ما در فصل نخست این کتاب گفتیم که متداول‌ترین روشهای اندازه‌گیری ویژگیها یا صفات روانی افراد **آزمون** و **پرسشنامه**اند. هرچند که در آموزش و پرورش و روانشناسی، علاوه بر آزمون و پرسشنامه، روشهای دیگر اندازه‌گیری و سنجش مورد استفاده قرار دارند، با این حال، آزمون و پرسشنامه از ابزارهای بسیار متداول‌اند. از تعیین میزان آموخته‌های دانش‌آموزان و دانشجویان گرفته تا برآورد ویژگیهای شخصیتی آنان، و تعیین نگرشها و علاقهٔ ایشان نسبت به امور مختلف همواره از ابزارهایی که عموماً نام آزمون و پرسشنامه بر آنها نهاده شده است استفاده می‌شود. بنابراین، آزمونها و پرسشنامه‌های مورداستفاده در روانشناسی و آموزش و پرورش بسیار متنوع‌اند و به همین منظور ما فصل حاضر را به دسته‌بندی انواع آنها اختصاص داده‌ایم.

دسته‌بندی روشهای اندازه‌گیری با توجه به ویژگیهای مورد اندازه‌گیری

روشهای اندازه‌گیری را می‌توان با توجه به ویژگیهای گوناگونی که اندازه می‌گیرند ابتدا به دو دستهٔ کلی **آزمونهای سنجش توانایی**[1] و **پرسشنامه‌های سنجش عاطفی**[2] تقسیم کرد و بعد هر یک از این دو دسته ابزار اندازه‌گیری را نیز به اجزای کوچکتری به شرحی که در زیر خواهد آمد دسته‌بندی نمود.

آزمونهای اندازه‌گیری توانایی

آزمونهای توانایی، آموخته‌ها، مهارتها، و استعدادهای افراد را می‌سنجند. دسته‌ای از این آزمونها، تواناییهای **شناختی**[3] افراد را می‌سنجند و تعدادی دیگر تواناییهای **روانی - حرکتی**[4] را اندازه می‌گیرند. هریک از این دو دسته آزمون کاربرد خاص خود را دارد. آزمونهای تواناییهای شناختی بیشتر در زمینه‌های تحصیلی و فعالیتهای فکری و ذهنی مورد استفاده قرار می‌گیرند، اما آزمونهای روانی - حرکتی غالباً در موقعیتهای صنعتی و نظامی و تربیت‌بدنی به کار می‌روند. صرف‌نظر از کاربردهای اختصاصی، این دو دسته آزمون، به طور عمده، برای تعیین جنبه‌های مختلف تواناییهای افراد تهیه و اجرا می‌شوند. ویژگی مهم آزمونهای توانایی (شناختی و روانی - حرکتی) این است که هنگام اجرای آنها از آزمون‌شونده خواسته می‌شود تا حداکثر سعی خود را به کاربندد تا بهترین نمرهٔ ممکن را به دست آورد.

آزمونهای توانایی به دو دستهٔ آزمونهای **استعداد**[5] و آزمونهای **پیشرفت تحصیلی**[6] تقسیم می‌شوند. آزمونهای استعداد میزان توانایی یا آمادگی فرد را برای انجام دادن کارهایی که در پیش

1. ability
2. affective
3. cognitive
4. psychomotor
5. aptitude
6. academic (scholastic) achievement

دارد و همچنین ظرفیت او را برای یادگیریهای مختلف می‌سنجند؛ اما آزمونهای پیشرفت تحصیلی دانش و مهارتهایی را که فرد تا لحظهٔ اجرای آزمون کسب کرده است اندازه می‌گیرند. به سخن دیگر، آزمونهای استعداد ناظر به آینده‌اند و برای پیش‌بینی موفقیتهای آتی فرد به کار می‌روند، در حالی که آزمونهای پیشرفت تحصیلی ناظر به گذشته‌اند و برای تعیین آموخته‌های قبلی فرد مورد استفاده قرار می‌گیرند. آزمونهای هوش و استعدادهای مختلف مُعرّف آزمونهای استعداد هستند و آزمونها و امتحانهایی که برای موضوعهای گوناگون درسی تهیه می‌شوند آزمونهای پیشرفت تحصیلی نام دارند.

آزمونهای استعداد و پیشرفت تحصیلی را به روشنی نمی‌توان از یکدیگر متمایز کرد، زیرا آنچه را که فرد در گذشته آموخته است غالباً می‌توان برای پیش‌بینی موفقیت یادگیری او در آینده مورد استفاده قرار داد. در واقع یکی از عوامل مهم پیش‌بینی یادگیری فرد در آینده، یادگیریهای قبلی او هستند. ثرندایک[1] و هیگن[2] (۱۹۷۹) در کتاب معروف خود با عنوان *اندازه‌گیری و ارزشیابی در روانشناسی و آموزش و پرورش*، در رابطه با آزمونهای استعداد و پیشرفت تحصیلی، به گونهٔ زیر نظر داده‌اند:

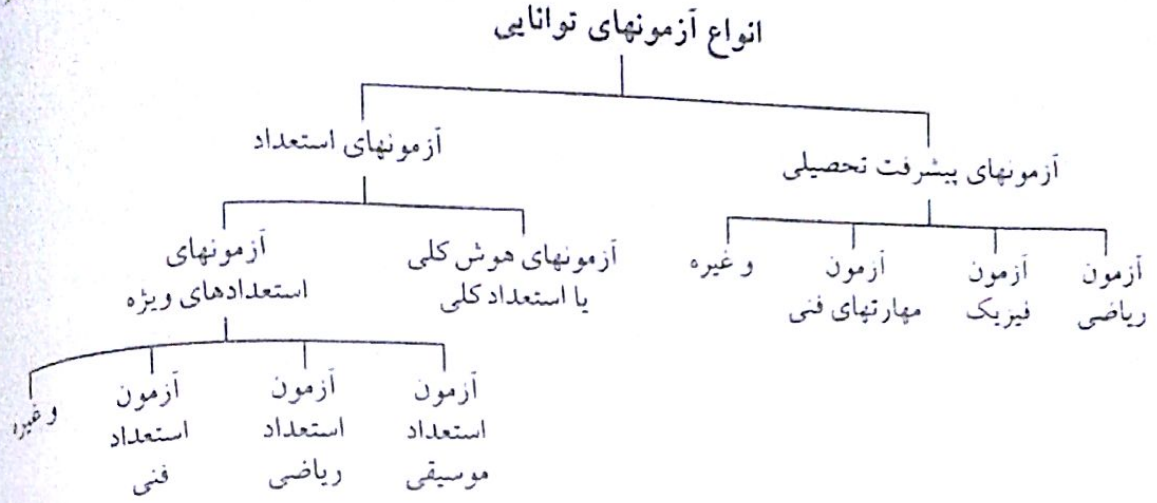
> باید به خاطر داشت که همهٔ آزمونهای توانایی، صرف‌نظر از نامی که دارند، تواناییهای موجود فرد را اندازه می‌گیرند. ما نمی‌توانیم مستقیماً توانایی "ذاتی" یا "ارثی" کسی را اندازه‌گیری کنیم. آنچه را که ما می‌توانیم در فرد اندازه بگیریم توانایی او برای انجام دادن اعمال در زمان حال است.... بنابراین، تمایز بین آزمونهای استعداد و پیشرفت تحصیلی به قصد و منظوری مربوط می‌شود که نتایج آزمون برای آن قصد و منظور مورد استفاده قرار می‌گیرند، نه به ماهیت یا محتوای آن آزمون. (ص ۲۴)

هرچند که این دو نوع آزمون از لحاظ ساخت و محتوا شباهتهای زیادی با یکدیگر دارند، اما از آنجا که آنها غالباً برای مقاصد مختلف به کار می‌روند، دسته‌بندی آزمونهای توانایی به استعداد و پیشرفت تحصیلی یک دسته‌بندی مفید است. چنان که قبلاً گفته شد، آزمونهای استعداد را معمولاً برای پیش‌بینی موفقیتهای آیندهٔ افراد به کار می‌بندند، در حالی که آزمونهای پیشرفت تحصیلی را برای تعیین مقدار آموخته‌ها یا تواناییهای موجود آزمون‌شوندگان مورد استفاده قرار می‌دهند.

آزمونهای پیشرفت تحصیلی با موضوع یا محتوایی که اندازه می‌گیرند مشخص می‌گردند و در نتیجه با عنوان آن موضوعها نام‌گذاری می‌شوند، مانند *آزمون پیشرفت ریاضی*، *آزمون پیشرفت نیزیک*، *آزمون پیشرفت مهارتهای فنی*، و جز اینها. اصطلاح آزمونهای استعداد نیز به تعدادی آزمون با مقاصد مختلف اطلاق می‌شود. معروف‌ترین آزمونهای استعداد آزمونهای هوش[3] یا استعداد کلی هستند. هدف اکثر آزمونهای هوش کلی اندازه‌گیری آن دسته از تواناییهای ذهنی است

---

1. Thorndike 2. Hagen 3. intelligence



شکل ۱-۳ دسته‌بندی انواع آزمونهای توانایی

که تقریباً در همهٔ انواع فعالیتهای فکری و شناختی ضروری‌اند.

علاوه بر آزمونهای هوش یا استعداد کلی که تواناییهای عمومی فرد را اندازه می‌گیرند، تعداد دیگری آزمون با نام آزمونهای استعدادهای ویژه وجود دارند که استعدادهای اختصاصی افراد را در زمینه‌های مختلف اندازه‌گیری می‌کنند. به تعبیر دیگر، هدف از کاربرد آزمونهای هوش کلی پیش‌بینی موفقیتهای همه‌جانبهٔ افراد است، در حالی که آزمونهای استعدادهای ویژه برای پیش‌بینی موفقیت یا شایستگی افراد در زمینه‌های فکری و عملی مشخص به کار می‌روند. آزمونهای استعدادهای ویژه با توجه به زمینه‌ای که استعداد افراد در آن سنجیده می‌شوند نام‌گذاری می‌شوند، مانند *آزمون استعداد ریاضی*، *آزمون استعداد فنی*، *آزمون استعداد موسیقی*، و غیره. شکل ۱-۳ انواع آزمونهای توانایی را نشان می‌دهد.

پرسشنامه‌های اندازه‌گیری عاطفی

آزمونهای شناختی و روانی - حرکتی تواناییهای فکری و عملی افراد را اندازه می‌گیرند، اما پرسشنامه‌های عاطفی ویژگیهای عاطفی و شخصیتی افراد را می‌سنجند. یکی از تفاوتهای عمدهٔ بین این دو دسته ابزار اندازه‌گیری این است که آزمونهای توانایی حداکثر عملکرد آزمون شونده یا همان عملکرد بیشینه را اندازه می‌گیرند، بدین معنی که نشان می‌دهند آزمون‌شونده در بهترین شرایط قادر به انجام چه کاری است، در حالی که پرسشنامه‌های عاطفی وضع موجود یا عملکرد نوعی آزمون‌شونده را در موقعیتهای طبیعی زندگی می‌سنجند. یکی دیگر از وجوه اختلاف این دو دسته ابزار اندازه‌گیری آن است که آزمونهای توانایی به‌طور عمده برای مقاصد پیش‌بینی موفقیتهای تحصیلی و شغلی مورد استفاده قرار می‌گیرند، اما پرسشنامه‌های عاطفی و شخصیتی بیشتر به امور مشاوره در حل مسائل روانی، تربیتی، و حرفه‌ای کمک می‌کنند.

پرسشنامه‌های سنجش عاطفی دارای انواع زیادی هستند. ما در اینجا تنها به مهم‌ترین آنها یعنی پرسشنامه‌های شخصیت، نگرش، و علاقه اشاره می‌کنیم.

پرسشنامه‌های اندازه‌گیری شخصیت

از شخصیت[1] تعریفهای زیادی به عمل آمده است. در واقع می‌توان شخصیت را به صورت مجموعهٔ رفتارهای فرد در موقعیتهای اجتماعی تعریف کرد (ایبل[2]، ۱۹۷۹). بعضی روانشناسان، علاوه بر رفتارهای آشکار، ویژگیهای غیررفتاری از جمله هوش، پیشرفت، قد و قامت، وضع سلامتی، کیفیت صدا، و وضع ظاهر را به این تعریف شخصیت اضافه می‌کنند. از این توصیف کلی شخصیت چنین برمی‌آید که روشهای سنجش شخصیت باید شامل متغیرهای مختلف شناختی، روانی ـ حرکتی، و عاطفی باشند. از جملهٔ این متغیرها می‌توان موفقیت، هوش، استعدادهای ویژه، علاقه، نگرش، ارزشها و غیره را نام برد. اما از آغاز نهضت روانسنجی تاکنون رسم بر این بوده است که کار اندازه‌گیری متغیرهایی چون هوش، استعداد، پیشرفت تحصیلی و بعضی ویژگیهای دیگر به عهدهٔ آزمونهای ویژهٔ آنها گذاشته شود و پرسشنامه‌های شخصیت به طور عمده به اندازه‌گیری متغیرهای انگیزش[3]، مزاج[4]، سازگاری[5]، منش[6]، و از این قبیل اختصاص یابد.

پرسشنامه‌ها و سایر ابزارهای اندازه‌گیری شخصیت را می‌توان به سه دستهٔ عمدهٔ زیر تقسیم کرد:

۱. **پرسشنامه‌های گزارش شخصی.** این ابزارها وسایلی هستند که در آنها فرد نظر شخصی‌اش را نسبت به خود بیان می‌کند.

۲. **پرسشنامه‌های جامعه‌سنجی.** در این وسایل اندازه‌گیری، افراد دیگر نظرشان را نسبت به فرد مورد سنجش ابراز می‌دارند.

۳. **فنون مشاهده‌ای.** در این نوع اندازه‌گیری رفتار فرد در یک موقعیت بخصوص مورد مشاهده قرار می‌گیرد.

راه دیگر دسته‌بندی ابزارهای سنجش شخصیت توجه به روش مطالعهٔ رفتار فرد یا نحوهٔ ارائهٔ محرکهاست. ابزارهای سنجش شخصیت را از این بابت می‌توان به دو دسته تقسیم کرد. در یکی از این دو دسته محرکها به صورت سازمان یافته ارائه می‌شوند؛ یعنی در آنها سؤالها طوری تهیه می‌شوند که همهٔ پاسخ‌دهندگان آنها را یکسان تعبیر می‌کنند (مثلاً "آیا همیشه در خواب و خیال به سر می‌برید؟"). در نوع دیگر ابزارهای سنجش[7]، محرکها به صورت سازمان نیافته ارائه می‌شوند، یعنی سؤالها به وسیلهٔ پاسخ‌دهندگان مختلف به گونه‌های متفاوت تعبیر و تفسیر می‌شوند. این نوع ابزارهای سنجش که غالباً به ابزارهای فرافکن معروف‌اند شامل تصاویر مبهم یا لکه‌های جوهر هستند که آزمون‌شونده باید آنها را تعبیر و تفسیر کند.

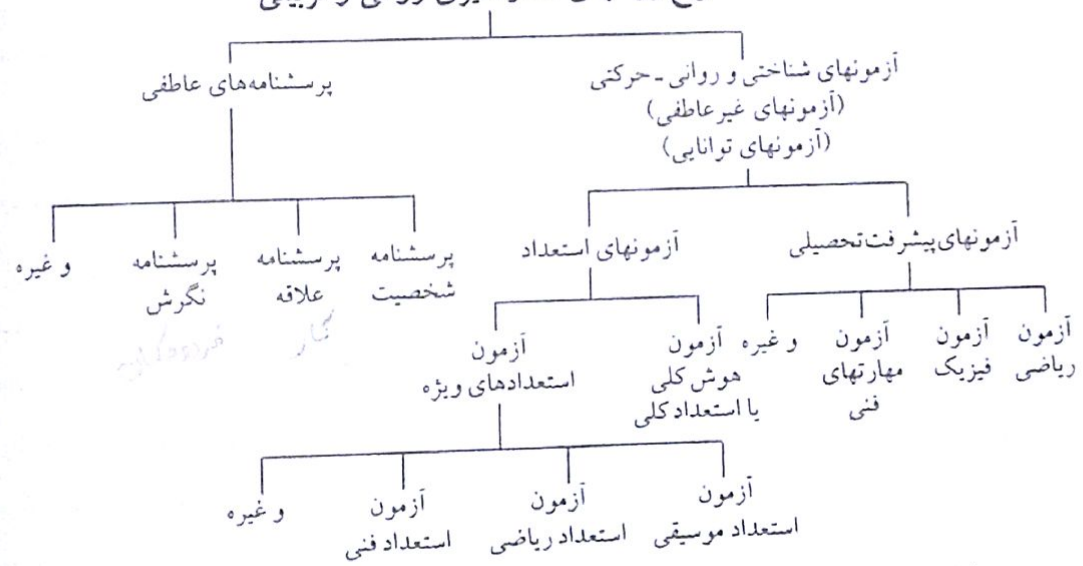
---

1. personality
2. Ebel
3. motivation
4. temperament
5. adjustment
6. character
7. projective

ابزارهای اندازه‌گیری علاقه و نگرش

دو بخش دیگر شخصیت علاقه[1] و نگرش[2] اند که به خاطر اهمیت زیادی که در امور مشاورهٔ شغلی و تربیتی دارند معمولاً به طور جداگانه سنجش می‌شوند. هم علاقه و هم نگرش به دوست داشتن‌ها و دوست نداشتن‌های افراد اشاره می‌کنند و هر دو به رجحانهای فرد در بارهٔ فعالیتها، نهادهای اجتماعی، یا گروهها ارتباط دارند. هر دو اینها شامل احساسهای شخصی نسبت به امور هستند. اما آنچه علاقه و نگرش را از هم مجزا می‌کند موضوع این علاقه‌ها یا بی‌علاقه‌گی‌هاست. علاقه به احساس فرد نسبت به یک *فعالیت* اشاره می‌کند، در حالی که نگرش حاکی از احساس فرد نسبت به یک *شیء*، یک *نهاد اجتماعی*، یا یک *گروه* است (مهرنز[3] و لهمان[4]، ۱۹۸۴، ص ۴۶۵). ساکس[5] (۱۹۹۷) در مقایسهٔ علاقه و نگرش با یکدیگر گفته است علاقه به *رجحان* یک *فعالیت* بر *سایر فعالیتها* اشاره می‌کند؛ در مقابل نگرش به معنی *ترجیح دادن گروهها*، *نهادها*، یا *اشیاء* است.

برای سنجش علاقه‌ها و نگرشهای افراد روشهای مختلفی وجود دارند. معروف‌ترین وسیلهٔ اندازه‌گیری در این دو زمینه پرسشنامه‌های علاقه و نگرش‌اند. پرسشنامه‌های علاقه جمله‌هایی در بارهٔ مشاغل و فعالیتهای گوناگون را شامل می‌شوند. پاسخ‌دهنده به هر یک از این سؤالها بر حسب علاقه‌ای که به آن حرفه یا فعالیت دارد پاسخ می‌دهد. پرسشنامه‌های نگرش‌سنج نیز حاوی تعدادی پرسش در بارهٔ نهادهای اجتماعی، گروههای نژادی و قومی، یا مفاهیم مختلف هستند که در اختیار آزمون‌شونده گذاشته می‌شوند و او باید نظر خود را نسبت به آنها مشخص کند. شکل ۲-۳ تقسیم‌بندی انواع ابزارهای سنجش توانایی و عاطفی را نشان می‌دهد.



شکل ۲-۳ دسته‌بندی انواع آزمونهای مورد استفاده در آموزش و پرورش و روانشناسی

1. interest
2. attitude
3. Mehrens
4. Lehman
5. Sax

دسته‌بندی ابزارهای اندازه‌گیری با توجه به چگونگی تهیهٔ آنها

علاوه بر ویژگیهای مختلف مورد اندازه‌گیری، ابزارهای اندازه‌گیری مورد استفاده در روانشناسی و آموزش و پرورش را با توجه به چگونگی تهیه و فرایند به وجود آمدن آنها نیز می‌توان دسته‌بندی کرد. از این لحاظ، آنها را به دو دستهٔ عمدهٔ ابزارهای اندازه‌گیری میزان‌شده یا استاندارد شده و ابزارهای اندازه‌گیری معلم ساخته تقسیم می‌کنند. در زیر آنها را توضیح می‌دهیم.

ابزارهای اندازه‌گیری میزان‌شده

ابزارهای اندازه‌گیری **میزان‌شده (معیار شده)**[1] ابزارهایی هستند که با استفاده از نمونه‌های بزرگ افراد و غالباً به وسیلهٔ مؤسسه‌ها یا بنگاههای آزمون‌سازی دولتی یا خصوصی و یا مراکز آموزشی و علمی و برای اندازه‌گیری پیشرفت تحصیلی، هوش، استعداد یا شخصیت گروههای بزرگ افراد تهیه می‌شوند. در آزمونهای میزان شده، ملاک نسبی[2] اندازه‌گیری مورد استفاده قرار می‌گیرد. هوگان[3] (۲۰۰۷) آزمونهای استاندارد شده یا میزان شده را به نحو زیر توضیح داده است:

> نخست، اصطلاح **استاندارد شده**، به وجود **روشهای هم‌شکل** اجرا و نمره‌گذاری اشاره می‌کند.... یعنی اینکه آزمون در هر شرایطی باید طبق روشهای مشخص شده اجرا شود و نتایج آن نیز طبق ضوابط ازپیش تعیین شده تفسیر گردند. دوم، **استاندارد شده** به این معنی است که آزمون دارای هنجار[4] است، مانند هنجارهای ملی که بر مبنای هزاران مورد محاسبه می‌شوند. در واقع، به فرایند جمع‌آوری این داده‌های هنجاری اغلب برنامهٔ استاندارد کردن آزمون می‌گویند. (صص، ۲۰-۱۹)

آزمونهای معلم‌ساخته

در مقابل آزمونهای استاندارد شده، آزمونهای معلم‌ساخته آزمونهایی هستند که معلمان برای سنجش پیشرفت تحصیلی دانش‌آموزان خود تهیه می‌کنند و در طول دورهٔ آموزش و یا در پایان دوره به اجرا در می‌آورند. این نوع آزمونها غالباً برای تعیین میزان توفیق یادگیرندگان در رسیدن به هدفهای دقیق آموزشی طرح‌ریزی می‌شوند و از این جهت معمولاً در آنها ملاک مطلق اندازه‌گیری به کار گرفته می‌شود. هرچند که ظاهر آزمونهای معلم‌ساخته و میزان‌شده خیلی به هم شبیه است، بین آنها تفاوتهایی اساسی وجود دارد که در زیر مختصراً به آنها اشاره می‌شود.

مقایسهٔ آزمونهای میزان‌شده با آزمونهای معلم‌ساخته

ثرندایک و هیگن (۱۹۷۹) بین آزمونهای میزان‌شده و معلم‌ساخته سه نوع تفاوت قائل شده‌اند. نخست اینکه تهیهٔ آزمونهای میزان‌شده، در مقایسه با آزمونهای معلم‌ساخته، به زمان و تخصص

---

1. standardized

۲. انواع ملاکهای اندازه‌گیری در فصل سوم توضیح داده شده‌اند.

3. Hogan

۴. norm (انواع هنجارها در فصل ۱۹ توضیح داده شده‌اند.)

بیشتری نیاز دارد) از آنجاکه این آزمونها برای بیش از یک بار استفاده تهیه می‌شوند مؤسسات آزمون‌سازی، برای تهیهٔ آنها، زمان و هزینهٔ زیادی به کار می‌گیرند. در تهیهٔ این آزمونها، برنامه‌های درسی مورد بررسی دقیق واقع می‌شوند، محتوای درسها و مهارتهای مورد اندازه‌گیری نمونه‌گیری می‌شوند، سؤالها به وسیلهٔ متخصصان نوشته و ویرایش می‌شوند، صورت مقدماتی آزمون به‌طور آزمایشی اجرا می‌شود، سؤالها تجزیه و تحلیل می‌شوند و مورد تجدید نظر قرار می‌گیرند، و پس از هنجاریابی[1] صورت نهایی آزمون تهیه و جهت استفادهٔ گسترده عرضه می‌شود. اما، برای تهیهٔ آزمونهای معلم‌ساخته، نیازی به گذر از این مراحل مفصل نیست، زیرا این آزمونها برای تعیین مقدار یادگیری دانش‌آموزان و دانشجویان درسهای بخصوص تهیه و اجرا می‌شوند، و پس از هر بار اجرا به سهولت قابل تجدید نظر و اصلاح هستند.

۲) دوم اینکه، از آنجاکه آزمونهای میزان‌شده برای استفاده در مناطق مختلف و مقایسهٔ دانش‌آموزان مدارس مناطق گوناگون با یکدیگر تهیه می‌شوند، هدفهای مختلف آموزشی مناطق مختلف را باید در این آزمونها منظور کرد. بنابراین، این نوع آزمونها ویژگیهای اختصاصی کلاسها و مدارس بخصوص را نادیده می‌گیرند و بر وجوه مشترک محتوای درسی مناطق مختلف تکیه می‌کنند. اما آزمونهای معلم‌ساخته با توجه به محتوا و هدفهای ویژهٔ درسهای بخصوص تهیه می‌شوند، لذا چنین آزمونهایی برای مقایسهٔ میزان پیشرفت دانش‌آموزان مناطق مختلف خیلی مفید نیستند.

۳) سوم اینکه اصطلاح "آزمون میزان‌شده" به وجود داده‌های هنجاری[2] اشاره می‌کند، و ناشران این نوع آزمونها، برای معنی‌دار کردن نمرات آزمون خود، هنجارهای لازم را تدارک می‌بینند. اما آزمونهای معلم‌ساخته نیازی به هنجار و هنجاریابی ندارند. مقایسهٔ بین گروههای مختلف دانش‌آموزان که به وسیلهٔ آزمونهای میزان‌شده سنجش می‌شوند وجود هنجارها را در این نوع آزمونها ضروری می‌سازد، در حالی که در آزمونهای معلم‌ساخته الزاماً مقایسه‌ای بین گروههای دانش‌آموزان صورت نمی‌گیرد، و هنجاریابی نیز ضرورت پیدا نمی‌کند. بنابراین، هدف اصلی آزمونهای میزان‌شده مقایسه است ــ مقایسهٔ یک فرد یا یک گروه با فرد یا گروه دیگر، یا مقایسهٔ جنبه‌های مختلف عملکرد یک فرد یا یک گروه با یکدیگر. بنا به گفتهٔ باتلر[3] و مکمان[4] (۲۰۰۶)، "این آزمونهای گسترده برای جمع‌آوری اطلاعات دربارهٔ یادگیری دانش‌آموزان مورد استفاده قرار می‌گیرند و در سراسر کلاسها یکسان اجرا می‌شوند تا امکان مقایسهٔ اطلاعات به‌دست آمده برای گروههای مختلف امکان‌پذیر باشد" (ص ۷).

استفاده از آزمونهای میزان‌شده و معلم‌ساخته در تصمیم‌گیریهای آموزشی

آزمونهای میزان‌شده و معلم‌ساختهٔ پیشرفت تحصیلی را می‌توان برای تصمیم‌گیریهای زیر مورد استفاده قرار داد:

---

۱. بحث هنجار و هنجاریابی در فصل ۱۹ آمده است.

2. normative data
3. Butler
4. McMunn

۱. تصمیمات آموزشی مربوط به کم و کیف جریان آموزش را باید بر مبنای آزمونهای معلم‌ساخته گرفت نه آزمونهای میزان‌شده، زیرا هدف عمدهٔ آزمونهای پیشرفت تحصیلی معلم‌ساخته بهبود جریان آموزش و بالا بردن سطح یادگیری دانش‌آموزان و دانشجویان است.

۲. تصمیمات مربوط به ارتقای دانش‌آموزان و دانشجویان را نیز باید بر مبنای آزمونهای معلم‌ساخته که دربرگیرندهٔ موضوعات مشخص درسی و هدفهای دقیق آموزشی است گرفت. آزمونهای میزان‌شده برای این منظور مفید نیستند.[1]

۳. تصمیمات تشخیصی مربوط به مشکلات یادگیرندگان و تصمیمات مربوط به رفع نواقص یادگیری و آموزش را می‌توان بر مبنای هر دو نوع آزمون معلم‌ساخته و میزان‌شده اتخاذ کرد.

۴. تصمیمات مربوط به جایابی یا پایه‌گزینی افراد به اندازه‌گیری وسیع‌تری نیازمندند و بنابراین بهتر است در این مورد از آزمونهای میزان‌شده استفاده کرد.

۵. تصمیمات مربوط به راهنمایی و مشاوره دانش‌آموزان معمولاً به مقایسه‌های هنجاری نیازمندند و بنابراین بهتر است برای این منظور از آزمونهای میزان‌شده استفاده کرد.

۶. از آنجا که تصمیمات مربوط به انتخاب بر مبنای مقایسهٔ افراد و گروهها گرفته می‌شوند، بنابراین در این مورد نیز بهتر است از آزمونهای میزان‌شده استفاده شود.

۷. تصمیمات مربوط به برنامه‌ریزی درسی نیز که به انتخاب یک برنامه از میان برنامه‌های مختلف منجر می‌شود به مقایسهٔ وسیع نیازمند است که آن هم از عهدهٔ آزمونهای میزان‌شده برمی‌آید. در این مورد می‌توان از آزمونهای محلی معلم‌ساخته نیز کمک گرفت.

۸. تصمیمات مربوط به ارزشیابی از میزان توفیق مدارس مختلف نیز مستلزم کسب نظری جامع و مقایسه‌ای در این باره است که یک آموزشگاه، در مقایسه با آموزشگاه دیگر، تا چه حد موفقیت کسب کرده است. برای این منظور نیز آزمونهای میزان‌شده مفیدتر هستند.

از آنجا که کتاب حاضر در درجهٔ اول برای استفادهٔ معلمان نوشته شده است، فنون مورد بحث در این کتاب به طرز تهیهٔ آزمونهای پیشرفت تحصیلی معلم‌ساخته اختصاص دارند. با وجود این، اصول و فنون مورد بحث در کتاب حاضر، کم و بیش، همان اصول مورد استفاده در تهیهٔ آزمونهای پیشرفت تحصیلی میزان‌شده هستند.

دسته‌بندی روشهای اندازه‌گیری و سنجش با توجه به نوع سؤال یا ماده

روشهای مختلف اندازه‌گیری و سنجش از جمله آزمونها را می‌توان با توجه به نوع **سؤال**[1] یا **ماده**[2] دسته‌بندی کرد. هرچند که در اشاره به روشهای مختلف اندازه‌گیری و سنجش اصطلاح *سؤال* متداول است، اما *ماده* معنای گسترده‌تری دارد. در اندازه‌گیری و سنجش، هم به سؤالِ آزمون و هم

1. question 2. item

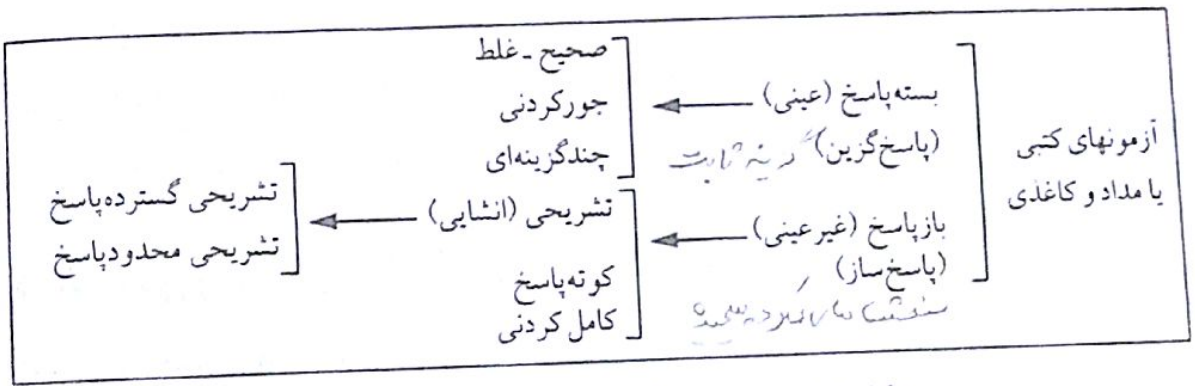
به هر تکلیف یا تمرینی که به آزمون شونده داده می‌شود **ماده** می‌گویند. ما در دنبالهٔ مطالب این کتاب از اصطلاحهای **ماده** و **سؤال** هر دو استفاده خواهیم کرد، اما منظورمان غالباً همان ماده است.

به طور کلی، وسایل مختلف اندازه‌گیری و سنجش یادگیری را با توجه به نوع سؤال یا ماده به **بسته‌پاسخ** و **بازپاسخ** تقسیم می‌کنند. در نوع بسته‌پاسخ معلم یا سنجشگر هم سؤال و هم پاسخ سؤال را تهیه می‌کند و آنها را در اختیار پاسخ‌دهنده می‌گذارد و از او می‌خواهد تا دربارهٔ آنها تصمیم بگیرد؛ مثلاً صحیح یا غلط بودن سؤال را تعیین کند (نوع صحیح ـ غلط)، از میان پاسخهای پیشنهادی آن را که درست است برگزیند (نوع چندگزینه‌ای)، یا تعدادی پرسش را با تعدادی پاسخ جور کند (نوع جورکردنی). در نوع بازپاسخ، دانش‌آموز خود پاسخ سؤال را تولید می‌کند، مانند سؤالهای تشریحی، کوته‌پاسخ، و کامل‌کردنی.

این تقسیم‌بندی ویژهٔ آزمونهای کتبی یا به اصطلاح **مداد و کاغذی** است که از آنها غالباً برای سنجش هدفهای شناختی استفاده می‌شود. شکل ۳ـ۳ این تقسیم‌بندی را نشان می‌دهد.

آزمونهای بسته پاسخ و بازپاسخ را به ترتیب **عینی**[1] و **غیرعینی**[2] نیز نامیده‌اند. تصحیح جواب سؤالهای آزمونهای عینی به طور کاملاً دقیق انجام می‌پذیرد و در آن نظر شخصی مصحح هیچ‌گونه دخالتی ندارد. به همین سبب، به این آزمونها صفت عینی داده شده است. در مقابل، تصحیح جواب سؤالهای آزمونهای غیرعینی معمولاً تحت تأثیر نظر شخصی مصحح قرار می‌گیرد، و چنان‌که در فصلهای بعدی این کتاب خواهیم دید عوامل مختلفی در چگونگی نمره‌گذاری معلم یا مصحح تأثیر می‌گذارند. به همین علت، آنها صفت غیرعینی (یا ذهنی) گرفته‌اند.

افزون بر اصطلاحات متداول بالا، برای دسته‌بندی و سهولت مطالعهٔ روشهای مختلف سنجش یادگیری، صاحب‌نظران عنوانهای دیگری را نیز مورد استفاده قرار داده‌اند. از جمله لین و گرانلاند (۲۰۰۰)، در دسته‌بندی روشهای مختلف سنجش، به جای اصطلاحات **بسته‌پاسخ** و **بازپاسخ**، از **آزمونهای گزینه ثابت**[3] و **سنجشهای عملکرد پیچیده**[4] استفاده کرده‌اند. منظور آنها از



شکل ۳-۳ دسته‌بندی آزمونهای کتبی یا مداد و کاغذی

1. objective
2. subjective
3. fixed-choice tests
4. complex-performance assessments

آزمونهای گزینه ثابت همان آزمونهای بسته پاسخ است که معروف‌ترین‌شان آزمونهای صحیح - غلط، چندگزینه‌ای، و جورکردنی است. لین و گرانلاند به عنوان نمونه برای سنجشهای عملکرد پیچیده موارد زیر را ذکر کرده‌اند: انشاءها یا مقاله‌های کتبی، مسائل ریاضی دارای راه‌حلهای چندمرحله‌ای، آزمایشهای آزمایشگاهی علوم، خلق آثار هنری، گزارشهای شفاهی، پروژه‌های مختلف، و نمایش کارهای انجام شده توسط یادگیرنده.

گونهٔ دیگر دسته‌بندی وسایل اندازه‌گیری و سنجش تقسیم آنها به دو گروه روشهای **انتخاب پاسخ (پاسخ‌گزین)**[1] و روشهای **ساختن پاسخ (پاسخ‌ساز)**[2] است. منظور از روشهای انتخاب پاسخ همان آزمونهای عینی یا بسته‌پاسخ هستند. از آنجا که در این آزمونها آزمون‌شونده پاسخ مورد نظر را برمی‌گزیند به آنها انتخاب پاسخ یا پاسخ‌گزین می‌گویند. در روشهای سنجش ساختنِ پاسخ، آزمون‌شونده خودْ پاسخ را می‌سازد یا تولید می‌کند. آزمونهای تشریحی نمونه‌ای از این روشها هستند.

اصطلاحات دیگری که اخیراً متداول شده و این روزها در همهٔ متون سنجش و اندازه‌گیری به کار می‌روند عبارت‌اند از: **سنجش جایگزین**[3]، **سنجش واقعی (سنجش اصیل)**[4]. منظور از سنجش جایگزین یا سنجش دیگر اشاره به روشهایی است که به جای روشهای قبلاً متداولِ عینی یا پاسخ‌گزین مورد استفاده قرار می‌گیرند (باتلر و مکمان، ۲۰۰۶). منظور از سنجش واقعی سنجش یا سنجش اصیل فعالیتها و تکلیفهای معنی‌داری است که در ارزشیابی از پیشرفت تحصیلی دانش‌آموزان به کار می‌رود و هدف آن نشان دادن توانایی یادگیرندگان در انتقال یادگیریها به فعالیتهای دنیای واقعی است (کوهن و سوردلیک، ۲۰۰۵). باتلر و مکمان (۲۰۰۶) در توضیح اصطلاح یادگیری اصیل گفته‌اند:

> وقتی که دانش‌آموزان یا دانشجویان در بحثهای سیاسی شرکت می‌کنند، برای روزنامهٔ مدرسه مطلب می‌نویسند، در جلسات گروهی مربوط به موضوعهای مختلف شرکت می‌جویند، یا پژوهشهای علمی انجام می‌دهند در تکالیف واقعی یا اصیل درگیر می‌شوند. دانش‌آموزان و دانشجویان زمانی بهتر یاد می‌گیرند که برای یادگیری یک دلیل شخصی داشته باشند و محیط یادگیری برایشان آشنا باشد. سنجش اصیل هم دلیل یادگیری و هم محیط یادگیری مناسب را فراهم می‌آورد. (ص ۶)

باز هم اصطلاح دیگری که در سالهای اخیر به وسیلهٔ صاحب‌نظران (از جمله هینتزه[5]، استونر[6]، و بول[7]، ۲۰۰۰)، به‌کار رفته، اصطلاح **سنجش قیاس‌پذیر**[8] است. "سنجش قیاس‌پذیر رفتار آزمون شونده را در موقعیتهای شبیه‌سازی شده با موقعیتهای زندگی واقعی که در آن رفتار مورد سنجش اتفاق می‌افتد اندازه می‌گیرد" (هینتزه، استونر، و بول، ۲۰۰۰، ص ۱۰۴). هدف از مشاهده و سنجش رفتار با روش قیاس‌پذیر پیش‌بینی وقوع رفتار در محیط زندگی واقعی است[9].

---

1. selective response
2. constructing response
3. alternative assessment
4. authentic assessment
5. Hintze
6. Stonner
7. Bull
8. analogue assessment

۹. توضیح بیشتر این روش و سایر روشهای مورد اشاره در این فصل را در فصلهای بعدی کتاب خواهید دید.

خلاصهٔ فصل

انواع آزمونها و دیگر روشهای سنجش یادگیری را به راههای مختلفی می‌توان دسته‌بندی کرد. یکی از این دسته‌بندیها با توجه به ویژگیهای مورد اندازه‌گیری صورت می‌گیرد. ما در این دسته‌بندی روشهای سنجش را ابتدا به دو دستهٔ روشهای سنجش توانایی و روشهای سنجش عاطفی تقسیم کردیم. روشهای سنجش توانایی نیز از دو دستهٔ روشهای سنجش پیشرفت تحصیلی و روشهای سنجش استعداد تشکیل می‌شوند. آزمونهای پیشرفت تحصیلی را به دو دستهٔ آزمونهای استاندارد شده یا میزان شده و آزمونهای معلم‌ساخته تقسیم می‌کنند.

ما همچنین آزمونها و روشهای مختلف سنجش یادگیری را با توجه به نوع سؤال یا ماده به گونه‌های مختلف زیر تقسیم کردیم: بسته پاسخ، بازپاسخ؛ عینی، غیرعینی (ذهنی)؛ انتخاب پاسخ، تولید پاسخ؛ و پاسخ‌گزین، پاسخ‌ساز. اصطلاحهای بسته‌پاسخ، عینی، انتخاب پاسخ، و پاسخ‌گزین معادل‌اند و عموماً آزمونهای صحیح ـ غلط، جورکردنی، و چندگزینه‌ای را شامل می‌شوند. اصطلاحهای ساختن پاسخ، بازپاسخ، پاسخ‌ساز، و غیرعینی نیز هم‌معنا هستند و عموماً سنجشهای تشریحی، عملکردی، و سنجش فعالیتهای فکری پیچیده را دربر می‌گیرند.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. "همهٔ آزمونهای توانایی، صرف‌نظر از نامی که دارند، توانایی فرد را برای انجام دادن عملی در زمان حال می‌سنجند."

الف ـ صحیح

ب ـ غلط

۲. همهٔ ابزارهای اندازه‌گیری زیر نمونه‌هایی از آزمونهای توانایی هستند بجز ابزار اندازه‌گیری:

الف ـ هوش

ب ـ شخصیت

پ ـ استعداد موسیقی

ت ـ ورودی دانشگاهها

۳. کدام یک از بیانات زیر در رابطه با علاقه و نگرش درست نیست؟

الف ـ علاقه به یک فعالیت ولی نگرش به یک نهاد مربوط می‌شود.

ب ـ موضوع علاقه و نگرش احساسهای فرد هستند.

پ ـ علاقه و نگرش دو نام متفاوت برای موضوع واحدی هستند.

ت ـ علاقه و نگرش دو جنبه از شخصیت هستند.

۴. انواع آزمونها را با توجه به ویژگیهای مورد اندازه‌گیری دسته‌بندی کنید.

۵. تفاوتهای عمدهٔ بین روشهای اندازه‌گیری توانایی و روشهای اندازه‌گیری عاطفی را توضیح دهید.
۶. انواع آزمونهای توانایی را توضیح دهید و کاربردهای هریک از آنها را بیان کنید.
۷. ویژگیهای مشترک و متفاوت آزمونهای استعداد و آزمونهای پیشرفت تحصیلی را توضیح دهید.
۸. انواع آزمونهای پیشرفت تحصیلی را توضیح دهید.
۹. انواع روشهای سنجش عاطفی را با ذکر مثال توضیح دهید.
۱۰. تفاوت بین روشهای سنجش علاقه و نگرش را توضیح دهید.
۱۱. آزمونهای میزان‌شده و معلم‌ساخته را با هم مقایسه کنید و شباهتها و تفاوتهای بین آنها را بیان کنید.
۱۲. آزمونهای میزان‌شده و معلم‌ساخته در کدام تصمیم‌گیریهای آموزشی به کار می‌روند؟ توضیح دهید.
۱۳. تفاوت بین آزمون و پرسشنامه را توضیح دهید و موارد استفادهٔ آنها را بیان کنید.
۱۴. ابزارهای اندازه‌گیری شخصیت را از لحاظ نوع وسیلهٔ اندازه‌گیری و نحوهٔ ارائهٔ محرکها دسته‌بندی کنید و هر دسته را به اختصار توضیح دهید.
۱۵. دسته‌بندی روشهای سنجش را با توجه به نوع سؤال یا ماده توضیح دهید.

۳

دسته‌بندی ارزشیابیهای آموزشی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. دسته‌بندی انواع ارزشیابیهای آموزشی را در قالب رویکردهای مختلف توضیح دهد.

۲. دسته‌بندی انواع ارزشیابیهای آموزشی را در قالب الگوهای مختلف توضیح دهد.

۳. الگوهای ارزشیابی زیر را توضیح دهد:

الگوی ارزشیابی تایلر

الگوی ارزشیابی سیپ

الگوی ارزشیابی یو سی ال ای یا الگوی ارزشیابی سی اس ای

الگوی ارزشیابی اسکریون

الگوی ارزشیابی هدف آزاد

الگوی ارزشیابی مبتنی بر نظر متخصصان

الگوی ارزشیابی مبتنی بر اختلاف‌نظر متخصصان

الگوی ارزشیابی سیمایی

۴. دسته‌بندی انواع ارزشیابیهای آموزشی را با توجه به موضوع ارزشیابی توضیح دهد.

۵. انواع ارزشیابیهای زیر را توضیح دهد:

ارزشیابی از دانشجویان و دانش‌آموزان

ارزشیابی از برنامه‌های درسی

ارزشیابی از آموزشگاه

ارزشیابی از کارکنان آموزشگاه

۶. دسته‌بندی انواع ارزشیابیهای آموزشی را با توجه به ملاک مورد استفاده در ارزشیابی توضیح دهد.

۷. ارزشیابی وابسته به ملاک و ارزشیابی وابسته به هنجار را توضیح دهد و کاربردهای آنها را بگوید.

۸. دسته‌بندی انواع ارزشیابیهای آموزشی را با توجه به زمان و هدف استفاده از آنها توضیح دهد.

۹. ارزشیابی آغازین، ارزشیابی تکوینی، ارزشیابی تشخیصی، و ارزشیابی تراکمی را توضیح دهد و ویژگیها و موارد استفادهٔ آنها را بگوید.

۱۰. ارزشیابی درونی و ارزشیابی بیرونی را توضیح دهد و رابطهٔ آنها را با ارزشیابی تکوینی و ارزشیابی تراکمی توضیح دهد.

در فصل قبل دسته‌بندی انواع روشهای اندازه‌گیری و سنجش یعنی آزمونها و پرسشنامه‌های مورد استفاده در آموزش و پرورش و روانشناسی را از نظر گذراندیم. در این فصل انواع ارزشیابیهای آموزشی را دسته‌بندی می‌کنیم. چنان‌که در فصل قبل عمل شد، در این فصل نیز ارزشیابیهای مختلف آموزشی از جهات گوناگون دسته‌بندی می‌شوند.

دسته‌بندی ارزشیابیهای آموزشی در قالب رویکردهای مختلف

صاحب‌نظران ارزشیابی آموزشی در دیدگاههایشان نسبت به اینکه ارزشیابی چیست و چگونه باید آن را انجام داد اختلاف نظر دارند. این اختلاف نظرها به ایجاد رویکردهای ارزشیابی آموزشی[1] مختلفی انجامیده است که در اینجا به اختصار آنها را توضیح می‌دهیم. ما برای معرفی رویکردها و الگوهای ارزشیابی از روش ورتن و سندرز (۱۹۸۷) استفاده خواهیم کرد. این صاحب‌نظران ارزشیابی تمامی رویکردهای ارزشیابی آموزشی را به شش دسته به شرح زیر تقسیم کرده‌اند:

رویکرد مبتنی بر هدف[2]. در رویکرد مبتنی بر هدف، هدفهای کلی و هدفهای دقیق[3] مورد تأکید قرار می‌گیرند و ارزشیابی آموزشی به دنبال این است که تعیین کند این هدفها به چه میزانی تحقق یافته‌اند.

رویکرد مبتنی بر مدیریت[4]. در رویکرد مبتنی بر مدیریت، هدفْ تشخیص و تحقق بخشیدن به نیازهای اطلاعاتی مدیرانِ تصمیم‌گیرنده است.

رویکرد مبتنی بر مصرف‌کننده[5]. در رویکرد مبتنی بر مصرف‌کننده، تأکید عمده بر کسب اطلاعات مربوط به فرآورده‌های کلی آموزشی است که توسط مصرف‌کنندگان آموزشی در انتخاب از میان برنامه‌های درسی مختلف، فرآورده‌های آموزشی، و نظایر اینها مورد استفاده قرار می‌گیرد.

رویکرد مبتنی بر نظر متخصصان[6]. در رویکرد مبتنی بر نظر متخصصان، تأکید اصلی بر کاربرد مستقیمِ نظر متخصصان در قضاوت در بارهٔ کیفیت فعالیتهای آموزش و پرورش است.

رویکرد مبتنی بر اختلاف نظر متخصصان. در رویکرد مبتنی بر اختلاف نظر متخصصان، اختلاف‌نظرهای ارزشیابان مختلف (موافق و مخالف) مورد تأکید قرار می‌گیرد.

---

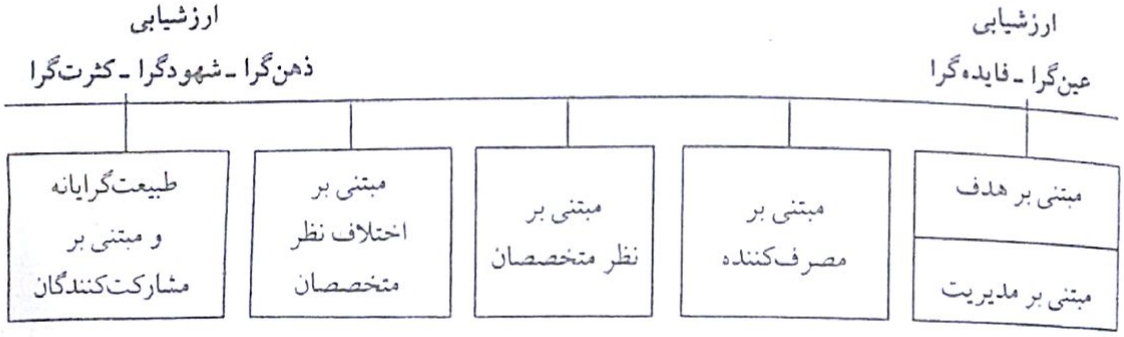
1. approachs to educational evaluation
2. objective-oriented approach
3. objectives
4. management-oriented approach
5. consumer-oriented approach
6. expertise-oriented approach

رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان[1]. در رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان، بررسی طبیعت‌گرایانه و کوشش مشارکت‌کنندگان در تعیین ارزشها، ملاکها، نیازها، و داده‌های ارزشیابی مورد تأکید قرار می‌گیرند.

* * * * *

دسته‌بندی رویکردهای شش‌گانهٔ ورتن و سندرز (۱۹۸۷) که در بالا معرفی شد، بر تقسیم‌بندی دوگانهٔ هاوس[2] (۱۹۸۳) از ارزشیابیهای آموزشی استوار است. این تقسیم‌بندی بیشتر مبنای فلسفی و ایدئولوژیکی دارد. هاوس انواع ارزشیابیهای آموزشی را در دو دستهٔ **عین‌گرا**[3] در مقابل **ذهن‌گرا**[4] یا **فایده‌گرا**[5] در مقابل **شهودگرا ـ کثرت‌گرا**[6] قرار داده است. در واقع این تقسیم‌بندی دوگانه نوعی پیوستار می‌سازد که در یک کرانهٔ آن عین‌گرایی یا فایده‌گرایی و در کرانهٔ دیگر آن ذهن‌گرایی یا شهودگرایی ـ کثرت‌گرایی قرار می‌گیرد. اگر دسته‌بندی شش‌گانهٔ رویکردهای ارزشیابی ورتن و سندرز را با دسته‌بندی دوگانهٔ هاوس مقایسه کنیم خواهیم دید که رویکردهای مبتنی بر هدف و مدیریت با کرانهٔ عین‌گرایی یا فایده‌گرایی مطابق‌اند و رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان بر کرانهٔ ذهن‌گرایی یا شهودگرایی ـ کثرت‌گرایی انطباق دارد، و بقیهٔ طبقات در حد فاصل بین این دو طبقه واقع می‌شوند (نگاه کنید به شکل ۱ـ۴).

در بارهٔ تقسیم‌بندی ارزشیابیهای آموزشی توسط هاوس (۱۹۸۳) به دو بُعد فایده‌گرایی و کثرت‌گرایی ـ شهودگرایی، باید گفت که رویکرد ارزشیابی فایده‌گرا ارزش برنامه‌های آموزشی را در تأثیرات کلی آنها می‌داند. همچنین این رویکرد ارزشیابی پیرو فلسفهٔ عینیت‌گرایی است. به سخن دیگر، در ارزشیابی فایده‌گرا، " بیشترین خوبی آن است که به بیشترین افراد فایده برساند" (ورتن و سندرز، ۱۹۸۷، ص ۴۸). یا به گفتهٔ هاوس (۱۹۸۳)، " فایده‌گرایی به این اندیشه که



شکل ۱-۴ توزیع شش رویکرد ارزشیابی ورتن و سندرز با توجه به دو بعد ارزشیابی فایده‌گرا در مقابل شهودگرا ـ کثرت‌گرای هاوس (ورتن و سندرز، ۱۹۸۷، ص ۶۱)

1. naturalistic and participant-oriented approach
2. House
3. objective
4. subjective
5. utilitarian
6. intuitionist-pluralist

شادمانی را در جامعه باید به حداکثر رسانید اشاره می‌کند" (ص ۴۹). بنابر این، در ارزشیابیهای وابسته به رویکرد فایده‌گرایی یا عینیت‌گرایی، متخصص ارزشیابی، با استفاده از میانگین نمرات آزمون یا شاخصهای دیگرِ "خوبی"، بر دستآوردهای کُلی گروه تأکید می‌ورزد تا "حداکثر خوبی را برای حداکثر افراد" تعیین نماید. در نتیجه بهترین برنامه‌های آموزشی از نظر این رویکردِ ارزشیابی آنهایی هستند که با توجه به ملاک یا ملاکهای انتخابی بیشترین دستاورد را فراهم می‌آورند.

در نقطهٔ مقابلِ رویکرد فایده‌گرایی یا عینیت‌گرایی، رویکردِ شهودگرایی ـ کثرت‌گرایی یا ذهن‌گرایی قرار دارد. رویکرد شهودگرایی ـ کثرت‌گرایی بر این اندیشه استوار است که ارزش به تأثیر برنامه بر فرد فرد افراد وابسته است نه بر اکثریت آنان. بنابراین رویکرد، ارزش حاکی از این است که بزرگترین خوبی مستلزم توجه به بهره‌مند شدن هریک از افراد از برنامهٔ مورد ارزشیابی است. به سخن دیگر، در رویکردهای کثرت‌گرایانهٔ ارزشیابی آموزشی، "احساسها[1] و دریافتهای[2] فردیِ ملاک خوبی و درستی به حساب می‌آیند" (هاوس، ۱۹۸۳، ص ۵۰). به همین سبب از لحاظ فلسفی به این رویکردها نام *ذهن‌گرایی* داده‌اند. بنابر آنچه گفته شد، در این رویکردِ ارزشیابی تأکید بر توزیع دستاوردها در میان همهٔ افراد و خرده‌گروههاست (مانند گروههای قومی)، لذا هیچ‌گونه شاخص عمومی برای "خوبی" وجود ندارد، بلکه مجموعه یا تعداد کثیری از ملاکها و داوریها به کار می‌روند و از این‌رو به این رویکرد ارزشیابی عنوان *کثرت‌گرایی* داده شده است. در ضمن، در این رویکرد ارزشیابی متخصص ارزشیابی به دنبال شاخصهای تمایل مرکزی نظیر میانگین و میانه نیست بلکه ارزشها و نیازهای مختلف را جمع‌آوری می‌کند. داده‌های این متخصصان ارزشیابی می‌توانند نمرات آزمونها باشند، اما در این رویکرد ارزشیابی "ارزشیابان اغلب داده‌های حاصل از مصاحبه‌های شخصی و نظرات مشارکت‌کنندگان در برنامه را ترجیح می‌دهند" (ورتن و سندرز، ۱۹۸۷، ص ۴۸).

متعادل ساختن و وزن دادن به داوریها و ملاکهای گوناگون در این رویکرد ارزشیابی عمدتاً جنبهٔ شهودی دارد، از این‌رو به آن عنوان *شهودگرایانه* داده‌اند. بنابراین رویکرد ارزشیابی، ارزش یا شایستگی یک برنامهٔ آموزشی عمدتاً وابسته به ارزشها و دیدگاههای همهٔ کسانی است که در بارهٔ برنامه قضاوت می‌کنند، و هر فرد یا گروه یک داور بر حق به حساب می‌آید. "هر شخصی برای خودش بهترین قاضی رویدادهاست" (هاوس، ۱۹۸۳، ص ۵۶). از لحاظ اجرایی، در ارزشیابیهای شهودگرا ـ کثرت‌گرا از تمامی افراد و گروههایی که تحت تأثیر برنامهٔ آموزشی قرار دارند به عنوان داوران استفاده می‌شود.

1. feelings
2. apprehensions

دسته‌بندی ارزشیابیهای آموزشی در قالب الگوهای مختلف

حال که با رویکردهای مختلف ارزشیابی آموزشی آشنا شدیم، آماده‌ایم تا **مدلها یا الگوهای ارزشیابی**[1] را مورد بحث قرار دهیم. هریک از رویکردهای ارزشیابی به یک یا چند الگوی ارزشیابی انجامیده است. ما در اینجا الگوهای مختلف را با توجه به رویکردهای شش‌گانهٔ ورتن و سندرز (۱۹۸۷) معرفی می‌کنیم. اما پیش از معرفی الگوهای ارزشیابی آموزشی، باید یادآوری کنیم که تعداد آنها بسیار زیاد است و در این کتاب امکان معرفی همهٔ آنها نیست. بنا به گفتهٔ ورتن و سندرز، در طول دو سه دههٔ اخیر بیشتر از ۵۰ الگوی مختلف ارزشیابی به وجود آمده‌اند. بنابراین، ما سعی خواهیم کرد تا معروف‌ترین آنها را معرفی نماییم. با این حال، چنانکه در فصل بعدی این کتاب خواهیم دید، انجام ارزشیابی آموزشی به نحو مطلوب هیچ نیازی به الگوگرایی ندارد، و نه تنها یادگیری ۵۰ الگوی ارزشیابی به هیچ وجه لازم نیست، بلکه چنانچه متخصص ارزشیابی مراحل اصلی ارزشیابی آموزشی را که در فصل بعد معرفی خواهیم کرد به درستی درک نماید هیچ نیازی ندارد که کار خود را در چهارچوب یک الگوی ارزشیابی خاص محدود سازد.

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر هدف

چنانکه پیشتر گفتیم، در رویکرد ارزشیابی مبتنی بر هدف ابتدا مقاصد یک فعالیت آموزشی تعیین می‌گردند، و بعد از آن در ارزشیابی کوشش به عمل می‌آید تا نشان داده شود که آن مقاصد تا چه اندازه تحقق یافته‌اند. در آموزش و پرورش فعالیت ارزشیابی می‌تواند از یک روز آموزش کلاسی تا تمامی نظام آموزشگاهی را شامل شود. نتایج به‌دست آمده از ارزشیابی مبتنی بر هدف را می‌توان برای اصلاح و تجدید نظر در مقاصد یا هدفهای فعالیت آموزشی، خود فعالیت آموزشی، یا روشها و تدابیری که برای سنجش تحقق مقاصد یا هدفها به کار رفته‌اند مورد استفاده قرار داد.

الگوی ارزشیابی تایلر. نخستین الگوی ارزشیابی آموزشی که تاریخ پیدایش آن به دههٔ ۱۹۳۰ میلادی بازمی‌گردد از آن رالف دبلیو. تایلر[2] است. به همین دلیل به آن **الگو یا رویکرد ارزشیابی تایلر یا نایلری**[3] می‌گویند.

مفهوم تایلر از ارزشیابی آموزشی این است که تعیین کند هدفهای آموزشیِ برنامهٔ آموزشگاه یا برنامهٔ درسی تا چه میزانی تحقق یافته‌اند. به همین منظور، تایلر در الگوی ارزشیابی خود مراحل زیر را پیشنهاد داده است:

۱. تعیین هدفهای کلی (غایتها) و هدفهای دقیق

۲. طبقه‌بندی غایتها و هدفها

---

1. models of evaluation 2. Ralph W. Tyler 3. Tylerian Evaluation Approach

۳. بیان هدفها به صورت رفتاری، یعنی قابل اندازه‌گیری
۴. یافتن موقعیتهایی که در آنها بتوان دست‌یابی به هدفها را نشان داد
۵. تولید یا تهیهٔ روشها و فنون اندازه‌گیری
۶. گردآوری داده‌های مربوط به عملکرد یادگیرندگان
۷. مقایسهٔ داده‌های مربوط به عملکرد با هدفهای رفتاری

در نتیجهٔ ارزشیابی، چنان‌که معلوم گشت که بین عملکرد و هدفها اختلاف وجود دارد، باید نواقص برطرف گردند، و بار دیگر ارزشیابی انجام گیرد. بنا به گفتهٔ ورتن و سندرز (۱۹۸۷)، "الگوی ارزشیابی تایلر، از لحاظ منطق درست، از نظر علمی قابل پذیرش، و توسط متخصصان ارزشیابی به سهولت قابل استفاده است، و لذا از زمان پیدایش تاکنون تأثیر عظیمی بر نظریه‌پردازان ارزشیابی داشته است" (ص ۶۳).

پس از تایلر کسان زیادی در راستای رویکرد ارزشیابی مبتنی بر هدف به نظریه‌پردازی و الگوسازی در ارزشیابی آموزشی پرداختند که از جملهٔ آنان می‌توان به **پارادایم ارزشیابی متفسل و مایکل**[1] (۱۹۶۷)، **رویکرد ارزشیابی هاموند**[2] (۱۹۷۳)، و **الگوی ارزشیابی اختلاف پرووس**[3] (۱۹۷۳) اشاره کرد.

الگوی ارزشیابی آموزشی تایلر حوزهٔ برنامه‌ریزی درسی را قویاً تحت تأثیر قرار داده است. تَبا[4] (۱۹۶۲)، یکی از همکاران نزدیک تایلر، برای برنامه‌ریزی درسی یک الگوی مبتنی بر هدف شامل مراحل زیر پیشنهاد داده است:

۱. تعیین نیازها
۲. بیان هدفها
۳. انتخاب محتوا
۴. سازمان دادن به محتوا
۵. انتخاب تجارب یادگیری
۶. سازمان دادن به تجارب یادگیری
۷. تعیین "چه چیزی" و "چگونگی" ارزشیابی

رابرت اف. میگر[5] در کتاب *تدارک هدفهای آموزشی* [۱۹۶۲، ترجمهٔ زنگنه و شمشیری (۱۳۵۵) و ۱۹۸۴] تکنولوژی ارزشیابی مبتنی بر هدف را ترویج نمود و روشهای دقیق بیان هدفهای

1. Metfessel and Michael's Evaluation Paradigm
2. Hammond's Evaluation Approach
3. Provus's Discrepancy Evaluation Model
4. Taba
5. Robert F. Mager

رفتاری را آموزش داد. تأکید بیش از اندازهٔ میگر بر بیان هدفهای دقیق رفتاری در دههٔ ۱۹۶۰ به مشاجرات و مباحثات تندی منجر شد که هنوز هم ادامه دارند.

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر مدیریت

آن دسته از الگوهای ارزشیابی که از رویکرد مبتنی بر مدیریت برخاسته‌اند به قصد کمک به مدیران و تصمیم‌گیرندگان به وجود آمده‌اند. پدیدآورندگان این الگوهای ارزشیابی آموزشی بر یک رویکرد سیستمی (نظامدار)[1] تأکید دارند که در آن در بارهٔ درون دادها، فرایندها، و برون‌دادها تصمیم‌گیری می‌شود. "مخاطب الگوهای ارزشیابی حاصل از رویکرد مبتنی بر مدیریت همواره تصمیم‌گیرندگان آموزش و پرورش هستند و جهت فعالیت ارزشیابی را خواسته‌های تصمیم‌گیرندگان، نیازهای اطلاعاتی، و ملاکهای اثربخشی تعیین می‌کنند" (ورتن و سندرز، ۱۹۸۷، ص ۷۷).

الگوی ارزشیابی سیپ[2]. پدیدآورندگان این الگوی ارزشیابی دانیل ال. استافل‌بیم[3] و همکاران او (۱۹۷۱) هستند. عنوان سیپ (CIPP) از حروف اول کلمات بافت یا موقعیت (Context)، درون‌داد (Input)، فرآیند (Process)، و فرآورده (Product) به دست آمده است، زیرا این الگوی ارزشیابی دارای چهارچوبی است که مدیران و تصمیم‌گیرندگان را در چهار نوع تصمیم‌گیری زیر کمک می‌کند:

۱. **ارزشیابی از بافت**[4]. از نظر استافل‌بیم و همکارانش (۱۹۷۱)، این مهمترین نوع ارزشیابی است. هدف این ارزشیابی فراهم آوردن یک زمینهٔ منطقی برای تعیین هدفهای آموزشی است. این مرحله از ارزشیابی همچنین شامل کوششهای تحلیل‌گرانه برای تعیین عناصر مربوط در محیط آموزشی و نیز کوشش در جهت شناسایی مشکلات، نیازها، و فرصتهای موجود در یک بافت یا موقعیت آموزشی است. به‌طور خلاصه، این مرحله از ارزشیابی به تصمیم‌گیری در زمینهٔ طراحی کمک می‌کند. روشهای ارزشیابی از بافت عمدتاً توصیفی و مقایسه‌ای هستند.

۲. **ارزشیابی از درون‌داد**[5]. در این مرحله از ارزشیابی اطلاعات موردنیاز در بارهٔ چگونگی استفاده از منابع به منظور دستیابی به هدفهای برنامه جمع‌آوری می‌شوند. در ضمن ارزشیابی از درون‌داد، ماهیت تواناییهای نظام آموزشی و استراتژیهای بالقوه برای رسیدن به هدفهایی که در نتیجهٔ ارزشیابی از بافت مشخص گردیدند تعیین می‌شوند. همچنین در این مرحله از ارزشیابی به تصمیم‌گیرندگان کمک می‌شود تا روشهایی را انتخاب و طراحی کنند که برای دستیابی به هدفها ضروری تشخیص داده شده‌اند.

---

1. system approach
2. The CIPP Evaluation Model
3. Daniel L. Stufflebeam
4. context evaluation
5. input evaluation

۳. **ارزشیابی از فرایند.** در این ارزشیابی کوشش می‌شود تا پاسخ پرسشهایی نظیر اینها تعیین گردند: آیا برنامه به خوبی درحال پیاده شدن است؟ چه موانعی بر سر راه موفقیت آن قرار دارند؟ چه تغییراتی ضروری هستند؟ پاسخ این پرسشها به کنترل و هدایت شیوه‌های اجرایی کمک می‌کند.

۴. **ارزشیابی از فرآورده.** در این مرحله معلوم می‌شود که نتایج به دست آمده چه هستند. سپس نتایج حاصل با هدفهای برنامه مقایسه می‌گردند و رابطهٔ بین انتظارات و نتایج واقعی مشخص می‌شوند. پس از بررسی کامل نتایج، اطلاعات لازم در اختیار تصمیم‌گیرندگان گذاشته می‌شود تا در صورت لزوم در بارهٔ ادامهٔ برنامه، متوقف ساختن آن، یا تجدید نظر در آن اتخاذ تصمیم کنند.

الگوی ارزشیابی یو سی ال ای[1]. این الگوی ارزشیابی از آن جهت با عنوان یو سی اِل اِی شهرت یافته که در دانشگاه کالیفرنیا واقع در لوس آنجلس پدید آمده است و حروف L, C, U، و A حروف اول کلمات تشکیل دهندهٔ نام این دانشگاه هستند. نام دیگر آن **الگوی سی اِس ای**[2] است که باز هم S، C، و E حروف اول **مرکز مطالعهٔ ارزشیابی**[3] **دانشگاه کالیفرنیا در لوس آنجلس** است. ماروین سی. الکین[4] (۱۹۶۹) هم مدیر مرکز مطالعهٔ ارزشیابی یو سی ال ای و هم واضع الگوی ارزشیابی یو سی ال ای است.

در الگوی ارزشیابی یو سی اِل اِی فرایند ارزشیابی آموزشی چنین تعریف شده است: "ارزشیابی عبارت است از فرایند تعیین انواع تصمیماتی که باید اتخاذ شوند، و نیز انتخاب، گردآوری، و تحلیل اطلاعات موردنیاز برای اخذ این تصمیمات؛ و گزارش کردن این اطلاعات به تصمیم‌گیرندگان" (پوفام، ۱۹۷۵، ص ۳۷). چنانکه از این تعریف برمی‌آید و چنانکه الکین (۱۹۶۹) خود گفته است مدل سی اِس ای بسیار شبیه به مدل سیپ است، با این تفاوت که آنچه را استافل بیم *ارزشیابی فرایند* نامیده الکین آن را تغییر داده است. الگوی سی اس ای یا یو سی ال ای از پنج نوع ارزشیابی زیر تشکیل یافته است:

۱. **سنجش نظام**[5]. این مرحله از ارزشیابی اطلاعات لازم را در بارهٔ حالت یا وضعیت نظام آموزشی گردآوری می‌کند (بسیار شبیه به مرحلهٔ ارزشیابی بافت در الگوی سیپ است).

۲. **طراحی برنامه**[6]. در این مرحله از ارزشیابی برنامه‌های ویژه‌ای برگزیده می‌شوند که تصور می‌شود در برآوردن نیازهای آموزشی مفیدند.

۳. **اجرای برنامه**[7]. در این مرحله اطلاعات ضروری در بارهٔ اینکه برنامهٔ انتخابی به گونه‌ای که در نظر بوده و به گروهی که می‌بایست داده شود معرفی گردیده است یا نه جمع‌آوری می‌گردد.

---

1. The UCLA Evaluation Model
2. The CSE Model
3. Center for the Study of Evaluation
4. Marvin C. Alkin
5. systems assessment
6. program planning
7. program implementation

۴. **بهسازی برنامه**[1]. منظور از این مرحله از ارزشیابی جمع‌آوری اطلاعات است در بارهٔ چگونگی کارکرد برنامه، چگونگی تحقق هدفهای واسطه‌ای یا موقتی، و اینکه بازده‌های پیش‌بینی نشده‌ای ظاهر شده‌اند یانه. (این مرحله شبیه به ارزشیابی فرآیندی الگوی سیپ است.)

۵. **تصدیق برنامه**[2]. در این مرحله اطلاعات مورد نیاز برای قضاوت دربارهٔ ارزش برنامه و امکان استفاده از آن در موقعیتهای دیگر فراهم می‌آیند.

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر مصرف‌کننده

ورتن و سندرز (۱۹۸۷) گفته‌اند مؤسسات یا افراد مستقلی که قبول مسئولیت کرده‌اند تا در بارهٔ فرآورده‌های آموزشی اطلاعات جمع‌آوری نمایند یا دیگران را در این کار یاری دهند رویکرد ارزشیابی مبتنی بر مصرف‌کننده را ترویج نموده‌اند. فرآورده‌های آموزشی عبارت‌اند از بسته‌های برنامهٔ درسی[3]، کارگاههای[4] آموزشی، رسانه‌های آموزشی[5]، فرصتهای آموزشی ضمن خدمت[6]، ارزشیابی از کارکنان[7]، تکنولوژی جدید[8]، نرم‌افزار و تجهیزات[9]، مواد آموزشی، و حتی خدماتی که به مدارس داده می‌شود. "فروش فرآورده‌های آموزشی در کشور آمریکا سالانه به ۵۰۰ میلیون دلار بالغ می‌شود" (ورتن و سندرز، ۱۹۸۷، ص ۸۷). بنابراین، رویکرد ارزشیابی مبتنی بر مصرف‌کننده والگوهای ارزشیابی برخاسته از آن عمدتاً برای کمک به مصرف‌کنندگان در انتخاب بهترین محصولات موجود در بازار فراهم آمده‌اند. به‌طور خلاصه، می‌توان گفت که در رویکرد مبتنی بر مصرف‌کننده ارزشیابی برابر است با تعیین میزان مطلوب بودن یک فرآوردهٔ آموزشی.

الگوی ارزشیابی اسکریون مایکل اسکریون[10] معروف‌ترین شخصیتی است که در زمینهٔ ارزشیابی آموزشی به قصد کمک به مصرف‌کنندگان کار کرده است. یکی از اندیشه‌های مهم اسکریون (۱۹۶۷) ایجاد تمایز بین **ارزشیابی تکوینی**[11] و **ارزشیابی تراکمی**[12] است. ما این دو نوع ارزشیابی را در دنبالهٔ مطالب این فصل به تفصیل توضیح خواهیم داد. در اینجا تنها به این نکته اشاره می‌کنیم که ارزشیابی تکوینی در حین اجرای برنامه و زمانی که هنوز امکان اصلاح مشکلات و رفع نواقص موجود است به کار می‌رود، یعنی زمانی که برنامه در حال شکل‌گیری یا تکوین است؛ اما ارزشیابی تراکمی در پایان کار و صرفاً برای اینکه معلوم گردد برنامهٔ آموزشی تا چه اندازه با هدفهای مورد نظر مطابقت دارد و نیازهای مصرف‌کنندگان را برآورده می‌سازد به اجرا در می‌آید. الگوی ارزشیابی اسکریون که برخاسته از رویکرد مبتنی بر مصرف کننده است عمدتاً یک ارزشیابی تراکمی است.

---

1. program improvement
2. program certification
3. curriculum packages
4. workshops
5. instructional media
6. in-service training opportunities
7. staff evaluation
8. new technology
9. software and equipment
10. Michael Scriven
11. formative evaluation
12. summative evaluation

ملاکهایی که اسکریون برای ارزشیابی از فرآورده‌ها به کار می‌گیرد به قرار زیرند:

۱. شواهد حاکی از تحقق یافتن هدفهای مهم آموزشی
۲. شواهد حاکی از تحقق یافتن هدفهای مهم غیرآموزشی (مانند هدفهای اجتماعی)
۳. نتایج حاصل از پی‌گیری[1]
۴. آثار جانبی پیش‌بینی نشده، مانند تأثیر بر معلم، همکاران معلم، سایر دانش‌آموزان، مدیران، والدین، مدرسه، مالیات‌دهنده، و سایر آثار جانبی مثبت یا منفی
۵. دامنهٔ استفاده (یعنی برای چه کسانی مفید است)
۶. ملاحظات اخلاقی (استفادهٔ نابجا از تنبیه یا محتوای بحث‌انگیز)
۷. هزینه‌ها

علاوه بر ملاکهای بالا، اسکریون (۱۹۷۴) یک فهرست وارسی[2] ارزشیابی از فرآورده‌های آموزشی به شرح زیر تهیه دیده است:

۱. **نیاز:** تعیین اینکه فرآورده به سلامت و بقای سیستم کمک می‌کند، یعنی شواهد مؤید نیاز.
۲. **بازار:** تعیین وجود یک طرح برای به مصرف رسانیدن فرآورده، یعنی بازار بالقوه.
۳. **عملکرد در اجراهای آزمایشی:** شواهد مربوط به عملکرد فرآورده یا برنامه در شرایط معمولی.
۴. **عملکرد ـ مصرف‌کنندهٔ حقیقی:** مناسب بودن فرآورده برای کسانی که تهیه شده است و احتمالاً به وسیلهٔ آنها مصرف خواهد شد (یعنی مصرف‌کنندگان حقیقی)، مانند معلمان، مدیران، و دانش‌آموزان.
۵. **عملکرد ـ مقایسه:** عملکرد فرآورده در مقایسه با فرآورده‌های رقیب آن.
۶. **عملکرد ـ در درازمدت:** اطلاعات مربوط به عملکرد در مدتی فراتر از زمان اجرای آزمایشی فرآورده.
۷. **عملکرد ـ آثار جانبی:** نتایجی بجز نتایج مربوط به هدفهای اولیه؛ یعنی نتایجی که توسط ارزشیابی هدف آزاد به دست آمده‌اند.
۸. **عملکرد ـ فرایند:** شواهد حاکی از اینکه فرآیندهای آموزشی با آنچه در فرآورده پیشنهاد شده است مطابقت دارند.
۹. **عملکرد ـ علّیت:** شواهد حاکی از اینکه نتایج به دست آمده در نتیجهٔ فرآورده یا برنامه به دست آمده‌اند.
۱۰. **عملکرد ـ معنی‌داری آماری:** شواهد آماری مربوط به اثربخشی فرآورده.

1. follow-up
2. checklist

۱۱. **عملکرد ـ معنی‌داری آموزشی:** اهمیت یا معنی‌داری تربیتی یا آموزشی از طریق داوران مستقل و متخصص، و قضاوتهای حاصل از تحلیل سؤالات و نمرات خام آزمونها، و نیز آثار جانبی، آثار درازمدت، و دستاوردهای تطبیقی، و استفادهٔ درست تربیتی حاصل می‌شود ـ به سخن دیگر از طریق موارد شمارهٔ ۳ تا شمارهٔ ۱۰ بالا.

۱۲. **هزینه‌ها و تأثیر هزینه‌ها:** برآورد هزینه‌های کلی برنامه و مقایسهٔ آن با هزینهٔ برنامه‌های رقیب.

۱۳. **حمایتهای دیگر:** ادامهٔ نظارت بر روی فرآورده یا برنامه و به روز درآوردن آن.

الگوی ارزشیابی هدف آزاد. یکی از پیشنهادهای اسکریون (۱۹۶۷، ۱۹۷۸) الگوی **ارزشیابی هدف آزاد**[1] در مقابل ارزشیابی مبتنی بر هدف است که قبلاً بحث کردیم. بنابر آنچه در ارزشیابی مبتنی بر هدف گفتیم، متخصص ارزشیابی هم مترصد کیفیت هدفهای برنامهٔ آموزشی و هم علاقه‌مند به این است که بداند تا چه میزانی آن هدفها تحقق یافته‌اند. در الگوی ارزشیابی هدف آزاد پیشنهادی اسکریون، علاوه بر اینها، بر بازده‌های پیش‌بینی نشدهٔ برنامهٔ آموزشی نیز تأکید می‌شود. به سخن دیگر، ارزشیابی هدف آزاد تمامی نتایج یک نوآوری آموزشی را، بدون هیچ‌گونه جهت‌گیری قبلی یا پیشداوری قبلی، سنجش و ارزیابی می‌کند.

اسکریون در دفاع از ارزشیابی هدف آزاد خود می‌گوید هدفهای آموزشی را نباید چشم بسته پذیرفت، بلکه باید آنها را نیز مانند هر چیز دیگری مورد ارزشیابی قرار داد. وی همچنین می‌گوید هدفهای آموزشی صرفاً الفاظ هستند و به ندرت مقاصد اصلی یک برنامه یا تغییرات مورد نظر را نشان می‌دهند. برای نمونه، تأسیس یک مرکز آموزش حرفه‌ای جدید به ایجاد مشاغل تازه خواهد انجامید که یک نتیجهٔ بسیار مطلوب است، اما چنین چیزی هرگز هدفِ از پیش تعیین شدهٔ مرکز آموزشی به حساب نمی‌آید. به باور اسکریون، نقش عمدهٔ ارزشیابی هدف آزاد کاستن از سوگیری و افزایش عینیت است. با این حال، "اسکریون ارزشیابی هدف آزاد را به عنوان جانشینی برای ارزشیابی مبتنی بر هدف پیشنهاد نداده است، بلکه آن را مکمل ارزشیابی مبتنی بر هدف می‌داند" (پوفام، ۱۹۷۵، ص ۲۸).

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر نظر متخصصان

در الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر نظر متخصصان آموزشی، اثربخشی مؤسسات، برنامه‌ها، فرآورده‌ها، و فعالیتهای آموزشی توسط متخصصان و صاحبان فن صورت می‌پذیرد. "برای نمونه، ارزش یک برنامهٔ درسی به وسیلهٔ متخصصان برنامه‌ریزی یا متخصصان موضوع درسی که برنامه را در عمل می‌بینند و محتوا و نظریهٔ یادگیری زیربنای آن را می‌سنجند ارزشیابی می‌شود" (ورتن و سندرز، ۱۹۸۷، ص ۹۸).

1. goal-free evaluation

ورتن و سندرز (۱۹۸۷) الگوهای ارزشیابی برخاسته از نظر متخصصان را در چهار طبقهٔ زیر دسته‌بندی کرده‌اند:

۱. نظامهای بررسی تخصصی رسمی
۲. نظامهای بررسی تخصصی غیررسمی
۳. بررسی توسط جلسات میزگرد ویژه
۴. بررسی توسط افراد خبره

این چهار نوع الگوی ارزشیابی مبتنی بر نظر متخصصان با توجه به ۵ سؤال زیر در جدول ۱ـ۴ ارائه شده‌اند.

۱. آیا یک ساختار برای انجام بررسی وجود دارد؟
۲. آیا معیارهای منتشرشده‌ای برای بررسی وجود دارند؟
۳. آیا بررسیها در فواصل زمانی از پیش تعیین‌شده‌ای انجام می‌شوند؟
۴. آیا بررسی توسط متخصصان مختلف انجام می‌گیرد؟
۵. آیا نتایج بررسی بر موقعیت و پایگاه آنچه یا آن کس که مورد ارزشیابی قرار می‌گیرد تأثیر خواهد داشت؟

موقعیتهایی که در آنها ارزشیابیهای مبتنی بر نظر متخصصان به کار می‌روند شامل موارد زیرند: امتحانات شفاهی دورهٔ دکتری، میزگردهایی که برای بررسی طرحهای پیشنهادی تحقیق تشکیل می‌شوند، بررسی پرونده‌های استخدامی و انجام مصاحبه‌های استخدامی، بررسی پرونده‌های

جدول ۱ـ۴ بعضی جنبه‌های چهار نوع الگوی ارزشیابی مبتنی بر نظر متخصصان (ورتن و سندرز، ۱۹۸۷، ص ۹۹)

نوع ارزشیابی مبتنی بر نظر متخصصان	ساختار موجود	معیار منتشرشده	برنامه از پیش‌تعیین‌شده	عقاید متخصصان مختلف	موقعیت متأثر از نتایج
نظامهای بررسی تخصصی رسمی	بله	بله	بله	بله	معمولاً
نظامهای بررسی تخصصی غیررسمی	بله	به‌ندرت	بعضی وقتها	بله	معمولاً
بررسی توسط جلسات میزگرد	نه	نه	نه	بله	بعضی وقتها
بررسی توسط افراد خبره	نه	نه	نه	نه	بعضی وقتها

علمی استادان دانشگاه برای صدور احکام قطعی استخدامی و ارتقاء، و بررسی مقالات رسیده برای چاپ در مجله‌های علمی توسط هیأت دبیران مجله.

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر اختلاف نظرهای متخصصان

الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر اختلاف نظرهای متخصصان بر این فرض استوار است که تمام متخصصان ارزشیابی خواسته و ناخواسته نظرها و تعصبات خود را در جریان ارزشیابی دخالت می‌دهند؛ لذا برای اینکه نتایج درست‌تری از ارزشیابی به دست آوریم باید از نظرهای موافق و مخالف چند تن از متخصصان و صاحب‌نظران استفاده کنیم. ورتن و سندرز (۱۹۸۷) موضوع را این‌گونه توضیح داده‌اند: "در حالی که غالب رویکردهای ارزشیابی می‌کوشند ناسوگیریها را کاهش دهند، رویکرد *مبتنی بر اختلاف نظرهای متخصصان* امیدوار است که با دخالت دادن نظرهای مثبت و منفی در ارزشیابی سوگیریها را متعادل سازد و از این طریق عدالت را برقرار نماید" (ص ۱۱۴).

بنابر توضیحات بالا، می‌توان گفت که ارزشیابی مبتنی بر اختلاف نظرهای متخصصان مجموعه‌ای از فعالیتهای مختلف ارزشیابی است که ذاتاً با هم تفاوت دارند. در این رویکرد یک ارزشیاب یا گروهی از ارزشیابان به عنوان مدافعان برنامه عمل می‌کنند و مثبت‌ترین جنبه‌های برنامه یا موضوع مورد ارزشیابی را معرفی می نمایند و از آن دفاع می‌کنند، در حالی که ارزشیاب دیگر یا گروهی دیگر از ارزشیابان نقش مخالف را بازی می‌کنند و نقاط ضعف برنامه را بزرگ جلوه می‌دهند. فرض بر این است که از ترکیب نظرهای موافق و مخالف نتیجهٔ درست‌تر و عادلانه‌تری در بارهٔ موضوع ارزشیابی به دست می‌آید.

کسان زیادی بنیان‌گذاران و پیروان این رویکرد ارزشیابی را تشکیل می‌دهند که به دلیل کمبود جا در این کتاب از ذکر نام و روش آنان خودداری می‌کنیم. خوانندگان می‌توانند برای این منظور به کتاب *ارزشیابی آموزشی: رویکردهای مختلف و رهنمودهای عملی*، اثر ورتن و سندرز (۱۹۸۷) مراجعه نمایند.

الگوهای ارزشیابی برخاسته از رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان

پدیدآورندگان و پیروان رویکرد ارزشیابی طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان چنین ادعا می‌کنند که غالب رویکردهای ارزشیابی موجود، به دلیل تأکید زیاد بر هدف‌گرایی و عینیت‌گرایی، روشهایی ماشینی و غیرانسانی را برگزیده‌اند. در ضمن، این متخصصان ارزشیابی می‌گویند روشهای متداول و سنتی ارزشیابی در مقیاسهای بسیار وسیعی انجام می‌شوند و لذا ارزشیابان شخصاً حتی یک‌بار در کلاسهای مورد ارزشیابی یا کلاسهایی که در فعالیت ارزشیابی شرکت دارند

پا نمی‌گذارند. در مقابل، رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان بر تجارب دست اول در موقعیتها و فعالیتهای تربیتی و آموزشی تأکید دارد. این رویکرد نسبتاً جدید در دهه‌های ۱۹۷۰ و ۱۹۸۰ به وجود آمد و پاگرفت.

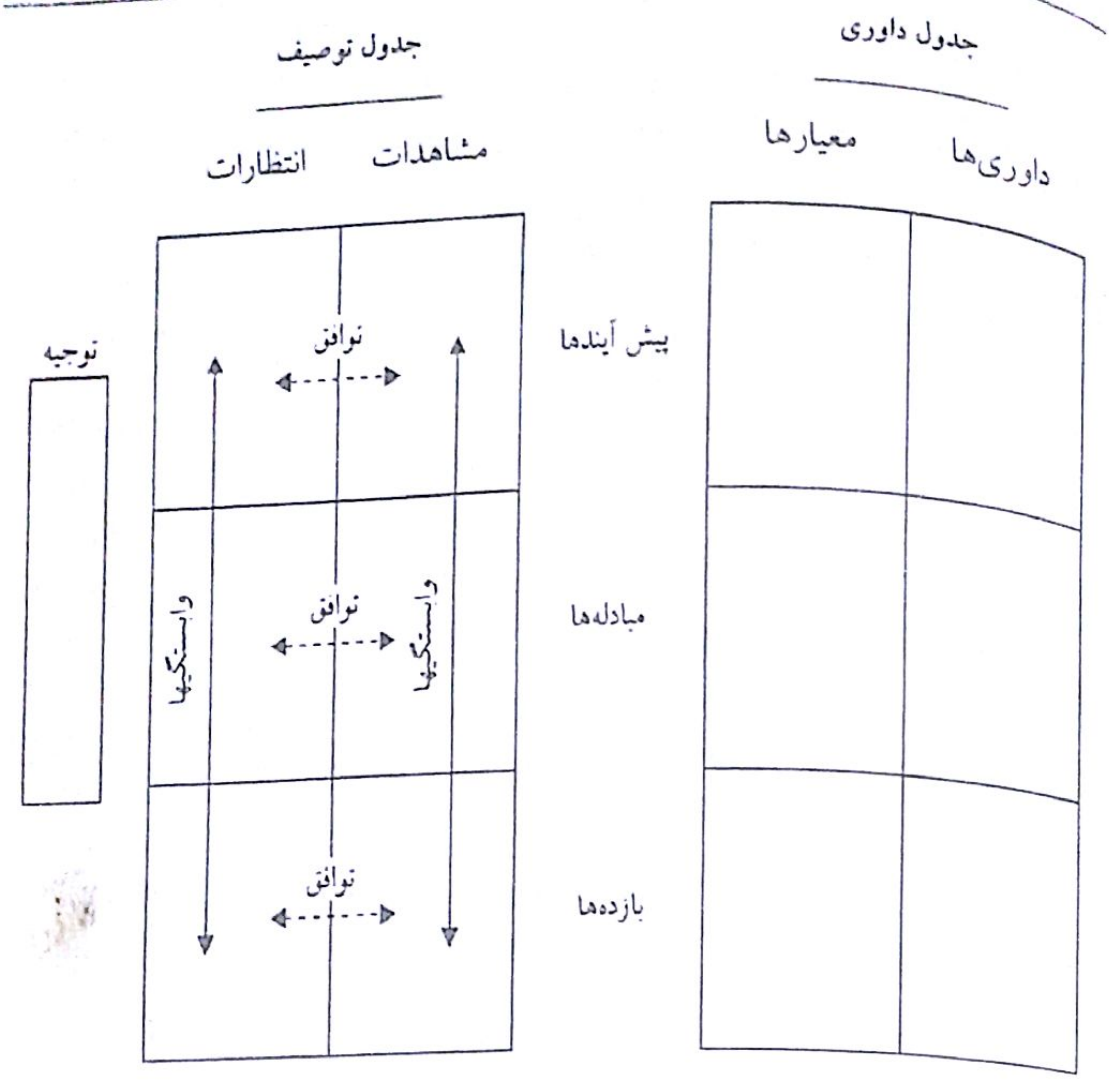
غالب کسانی که به ایجاد و استفاده از این رویکرد ارزشیابی خدمت کرده‌اند روشهای پژوهش و بررسی مبتنی بر مشاهدات طبیعی را به روشهای دقیق و کنترل‌شدهٔ آزمایشی ترجیح می‌دهند و از این رو به رویکرد آنها نام **طبیعت‌گرایانه** داده شده است. از سوی دیگر، چنانکه پیشتر گفته شد، این متخصصان بر دخالت کسانی که در موضوع مورد ارزشیابی شرکت دارند نیز تأکید می‌ورزند و از این جهت به رویکرد آنان نام رویکرد **مبتنی بر مشارکت‌کنندگان** داده‌اند.

الگوی ارزشیابی سیمایی[1]. یکی از معروف‌ترین الگوهای ارزشیابی برخاسته از رویکرد طبیعت‌گرایانه و مبتنی برمشارکت‌کنندگان رویکرد رابرت ای. استیک[2] (۱۹۶۷) به نام *الگوی سیمایی* است. در این الگوی ارزشیابی بر دو عمل اساسی ارزشیابی یعنی *توصیف*[3] و *داوری*[4] تأکید می‌شود و اینها دو سیما یا دو روی ارزشیابی به حساب می‌آیند، و بدین لحاظ به الگوی ارزشیابی استیک نام **سیمایی** داده‌اند. بنابراین، دو فعالیت عمدهٔ هر گونه مطالعهٔ ارزشیابی رسمی توصیف کامل و داوری در بارهٔ موضوع مورد ارزشیابی است. این دو عمل توصیف و داوری در سه مرحله از یک برنامهٔ آموزشی، یعنی *پیش‌آیندها*[5]، *مبادله‌ها*[6]، و *بازده‌ها یا پی‌آیندها*[7] انجام می‌گیرد. منظور استیک از پیش‌آیندها شرایط موجود پیش از آموزش است که می‌توانند به پی‌آیندها یا بازده‌ها بینجامند؛ منظور او از مبادله "توالی درگیریها" یا فرایند آموزش است؛ و بالاخره بازده‌ها یا پی‌آیندها برای استیک آثار برنامهٔ آموزشی به حساب می‌آیند. استیک برای کمک به ارزشیاب در سازمان دادن به داده‌های جمع‌آوری‌شده و تفسیر آنها طرح شکل ۲ـ۴ را ارائه داده است.

با توجه به طرح شکل ۲ـ۴، ارزشیاب آموزشی (۱) زمینه، دلایل توجیهی، و توصیف مربوط به دلایل ضروری بودن برنامه (از جمله نیاز به آن را) فراهم می‌آورد؛ (۲) فهرست پیش‌آیندهای مورد نظر (درون‌دادها، منابع، شرایط موجود)، مبادلات (فعالیتها، فرایندها)، و بازده‌ها را تهیه می‌کند؛ (۳) پیش‌آیندها، مبادلات، و بازده‌های مشاهده‌شده (از جمله مشاهدات مربوط به جنبه‌های پیش‌بینی‌نشده) را ثبت می‌کند؛ (۴) استانداردها (ملاکها، انتظارات، عملکرد برنامه‌های قابل مقایسه) را برای قضاوت در بارهٔ پیش‌آیندها، مبادله‌ها، و بازده‌ها به طور صریح بیان می‌کند؛ و (۵) داوریها یا قضاوتهای انجام‌گرفته در بارهٔ شرایط پیش‌آیند، مبادله‌ها، و بازده‌ها را ثبت می‌نماید.

---

1. Countenance Model
2. Robert E. Stake
3. description
4. judgement
5. antecedents
6. transactions
7. outcomes



شکل ۲ـ۴ طرح استیک از بیانات و داده‌های موردنیاز ارزشیابان آموزشی (به نقلِ ورتن و سندرز، ۱۹۸۷، ص ۱۳۱)

ارزشیاب با نگاه کردن به توافق بین انتظارات و مشاهدات و با نگاه کردن به وابستگی بازده‌ها به مبادله‌ها و پیش‌آیندها و نیز به وابستگی مبادله‌ها به پیش‌آیندها اطلاعات موجود در جدول ماتریس (شکل ۲ ـ ۴) را توصیف می‌کند.

برای روشن شدن ابعاد مختلف الگوی ارزشیابی استیک به مثال زیر از کتاب ***ارزشیابی آموزشی***، اثر پوفام (۱۹۷۵) توجه کنید. این یک مثال فرضی در بارهٔ یک معلم علوم اجتماعی در سطح دبیرستان است.

توجیه

معلم تصمیم گرفته است بر تمایز بین کارکردهای حکومت محلی، ایالتی، و مرکزی تأکید کند زیرا به اعتقاد او این تمایزها به دانش‌آموزان کمک می‌کنند تا مسئولیت شهروندی خود را بهتر درک کنند.

انتظارات

پیش‌آیندها: معلم می‌داند که فصل ۹ کتاب برای سه‌شنبه تعیین شده است.
مبادله‌ها: او تصمیم گرفته است که در این روز هم یک سخنرانی انجام دهد و هم بحث گروهی داشته‌باشد.
بازده‌ها یا نتایج: او پیش‌بینی می‌کند که دانش‌آموزان بتوانند در امتحانی که در روز چهارشنبه برگزار خواهد شد موفق بشوند.

مشاهدات

پیش‌آیندها: معلم مشاهده کرد که در روز سه‌شنبه شش دانش‌آموز غایب بودند.
مبادله‌ها: سخنرانی او آن قدر به طول انجامید که وقت زیادی برای بحث گروهی باقی نماند.
بازده‌ها: در امتحان روز چهارشنبه بیشتر از نصف دانش‌آموزان به یک سؤال مهم جواب غلط دادند.

معیارها (استانداردها)

پیش‌آیندها: معلم پیش‌بینی می‌کند که در روزی مانند سه‌شنبه که امتحانی برگزار نمی‌شود چند نفری غایب باشند.
مبادله‌ها: او معتقد است که سخنرانی‌هایش برای ۸۰ درصد دانش‌آموزان کلاس قابل فهم است.
بازده‌ها: معلمان همکار او در درس علوم اجتماعی معتقدند که سه‌چهارم دانش‌آموزان آنان در امتحاناتی نظیر امتحانات معلم مورد نظر موفق می‌شوند.

داوریها

پیش‌آیندها: معلم حالا که مطالب فصل ۹ را ارزیابی می‌کند آن را قدری گیج‌کننده می‌یابد.
مبادله‌ها: چند تن از دانش‌آموزان ضعیف به او گفتند که سخنرانی‌اش دارای ابهاماتی بوده است.
بازده‌ها: کمک‌معلمی که مأمور تصحیح برگه‌های امتحانی شاگردان بود گفت که تعداد زیادی از آنها در امتحان ضعیف عمل کردند.

علاوه بر الگوی ارزشیابی استیک، چند الگوی معروف دیگر وابسته به رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کننده وجود دارند. از جملهٔ آنها می‌توان به **ارزشیابی روشنگر**[1]، **ارزشیابی دموکراتیک**[2]، **ارزشیابی پاسخگو**[3]، و **ارزشیابی طبیعت‌گرایانه**[4] اشاره کرد. باز هم خوانندگان را برای کسب اطلاعات لازم در بارهٔ این الگوها به کتاب ***ارزشیابی آموزشی: رویکردهای مختلف و رهنمودهای عملی***، اثر ورتن و سندرز (۱۹۸۷) ارجاع می‌دهیم؛ ما در تهیهٔ مطالب این قسمت فصل حاضر از همین کتاب بسیار سود برده‌ایم. ویژگیهای مهم رویکردهای مختلف ارزشیابی آموزشی که تاکنون معرفی کردیم در جدول ۲ـ۴ با هم مقایسه شده‌اند.

---

1. illuminative evaluation
2. democratic evaluation
3. responsive evaluation
4. naturalistic evaluation

جدول ۲ـ۲ مقایسهٔ ویژگیهای مهم رویکردهای ارزشیابی آموزشی مختلف (اقتباس از اثر ورتن و سندرز، ۱۹۸۷، ص ۱۵۲)

۱. نام رویکرد	مبتنی بر هدف	مبتنی بر مدیریت	مبتنی بر مصرف‌کننده	مبتنی بر نظر متخصصان	مبتنی بر اختلاف‌نظر متخصصان	طبیعت‌گرایانه و مبتنی بر مشارکت کنندگان
۲. برخی از طرفداران:	تایلر، پسروورس، متفسل، مایکل، هاموند، یوفام، تاپا، بلوم، تالماج	استافل‌بیم، الکین، پرووس	اسکریون، کوموسکی	آیسنر، گروههای تأیید صلاحیت	وولف، اُونس، لویسن، کوریسکی	اسنیک، پاتون، گوبا، لینکولن، ریپی، مک‌دونالد، پارلت، هامیلتون
۳. هدف ارزشیابی:	تعیین میزان تحقق هدفها	رداری اطلاعات مفید برای مقاصد تصمیم‌گیری	تدارک اطلاعات مفید دربارهٔ فرآورده‌های آموزشی برای کمک به تصمیم‌گیری برای انتخاب و خرید این فرآورده‌ها	تدارک داوریهای حرفه‌ای دربارهٔ کیفیت موضوع مورد ارزشیابی	تدارک نظرهای موافق و مخالف و تعیین نقاط قوت و ضعف برنامه	فهمیدن و نشان دادن پیچیدگیهای برنامهٔ آموزشی و پاسخگویی به نیاز دست‌اندرکاران برنامه
۴. ویژگیهای شاخص:	تعیین هدفهای قابل اندازه‌گیری، استفاده از ابزارهای دقیق برای گردآوری اطلاعات، جستجو برای یافتن اختلافهای بین هدفها و عملکرد	کمک به تصمیم‌گیریهای منطقی، اجرای ارزشیابی در کلیهٔ مراحل تولید برنامه	استفاده از فهرست وارسی ملاکی برای تحلیل فرآورده‌ها، آزمودن فرآورده‌ها، آگاه ساختن مصرف کنندگان	داوری کردن بر اساس دانش و تجربهٔ فردی، استفاده از اتفاق نظر متخصصان	استفاده از اختلاف نظر متخصصان، تصمیم‌گیری براساس استدلالهای مخالف ابراز شده در جریان داوری	انعکاس دادن واقعیتهای چندگانه، استفاده از استدلال استقرایی، اکتشاف، و تجارب دست اول
۵. موارد استفاده:	تولید برنامهٔ درسی، نظارت بر یادگیری دانش‌آموزان، سنجش نیازها	تولید برنامهٔ آموزشی، نظامهای مدیریت آموزشی، طراحی برنامه، مسئولیت‌پذیری	گزارش به مصرف‌کننده، تولید فرآورده، انتخاب فرآورده برای توزیع	خودآموزی، صدور گواهینامه، امتحان کردن به وسیلهٔ کمیته، نقد	آزمودن برنامه‌ها یا مسائل بحث‌انگیز، جلسات سیاست‌گذاری	آزمودن و بررسی نوآوریها یا تغییراتی که در بارهٔ آنها اطلاعات کافی موجود نیست

دنبالهٔ جدول ۲ـ۴

۶. ملاکهای داوری برای ارزشیابی روش ارزشیابی:	اندازه‌پذیری هدفها؛ روایی و پایایی وسایل اندازه‌گیری	قابلیت استفاده، سهولت، مناسبت، و دقت فنی	فارغ بودن از سوگیری، دقت فنی، شواهد مربوط به اثربخش بودن و مورد نیاز بودن	استفاده از معیارهای شناخته‌شده، صلاحیت متخصصان یا خبرگان	تعادل، انصاف، فرصت وارسی مجدد	قابلیت پذیرش یا اعتبار، مناسب بودن، قابل تأیید بودن
۷. فواید:	سهولت استفاده، سادگی، تأکید بر بازده‌ها، پذیرش سطح بالا، اجبار کردن به داشتن هدف	جامعیت، حساسیت نسبت به نیاز اطلاعاتی کسانی که در موقعیت مدیریتی هستند، رویکرد سیستمی به ارزشیابی، استفاده از ارزشیابی در سراسر مراحل تولید برنامه، استفاده از منابع وسیع اطلاعاتی	تأکید بر اطلاعات مورد نیاز مصرف‌کننده، تأثیرگذاری بر تولیدکنندگان فرآورده‌ها، توجه به اثربخشی هزینه و قابلیت استفاده، موجود بودن فهرستهای وارسی	جامعیت، کارآمدی، سهولت استفاده، زمان‌بندی مناسب، تأکید بر قضاوت انسانی	جامعیت، وارسی دقیق ادعاها، معطوف بودن هدف به راه حل، روش ساختن جنبه‌های مختلف مسئله، استفاده از اطلاعات به دست‌آمده از منابع مختلف	تأکید بر توصیف و داوری، استفاده از استدلال استقرایی، استفاده از اطلاعات به دست‌آمده از منابع مختلف
۸. محدودیتها:	بیش از حد ساده گرفتن ارزشیابی و آموزش، تأکید صرف بر بازده آموزشی، کاهش‌گرایی، تأکید بیش از حد بر اجرای آزمون بر روی شاگردان	تأکید بر الگوی کارآمدی سازمانی و تولید، فرض نظم‌پذیری و پیش‌بینی‌پذیری در تصمیم‌گیری، پرهزینه بودن، تأکید بیش از حد بر نظر مدیران	هزینهٔ زیاد، امکان واپس زدن خلاقیت و نوآوری، عدم قابلیت وارسی مجدد	تکرارپذیری، حساس بودن نسبت به تعصبات شخصی، کمی شواهد و مدارک مؤید نتایج، وابستگی به صلاحیت متخصصان	خطاپذیر بودن داوران، هزینه و وقت زیاد، محدود بودن به اطلاعاتی که در دسترس‌اند	غیررهنمودی بودن، احتمال اینکه تحت تأثیر پدیده‌های غیرمعمول قرار بگیرد، احتمال هزینهٔ زیاد، فرضیه‌ساز بودن

دسته‌بندی ارزشیابیهای آموزشی با توجه به موضوع ارزشیابی

قبلاً در تعریف ارزشیابی گفتیم که ارزشیابی به این منظور به کار می‌رود که تعیین شود آیا هدفهای مورد نظر تحقق یافته‌اند یا نه. بنابراین، یکی از راههای تقسیم‌بندی انواع ارزشیابیها با توجه به هدف یا موضوع ارزشیابی است. از این لحاظ ارزشیابیهای مختلف آموزشی را می‌توان در گروههای مختلف زیر قرار داد.

ارزشیابی از دانشجویان و دانش‌آموزان

منظور از **ارزشیابی از دانشجویان و دانش‌آموزان**[1] عمدتاً ارزشیابی از **پیشرفت تحصیلی**[2] یا ارزشیابی از میزان **یادگیری**[3] آنان است. با این حال، در یک برنامهٔ کامل ارزشیابی دانشجویی یا دانش‌آموزی متغیرهای دیگری چون استعداد، هوش، شخصیت، نگرش، علاقه، و جز اینها نیز در نظر گرفته می‌شوند. برای این منظور، داده‌های زیادی جمع‌آوری می‌گردند و وسایل مختلف اندازه‌گیری مورد استفاده قرار می‌گیرند. از جمله وسایل اندازه‌گیری مورد استفاده می‌توان موارد زیر را نام برد: آزمونهای معلم‌ساخته و میزان‌شده، پروژه‌های کار عملی، گزارشهای شفاهی، و مشاهدات رسمی و غیررسمی.

استفاده از نتایج ارزشیابی دانش‌آموزی و دانشجویی، به ویژه نتایج ارزشیابی پیشرفت تحصیلی، نه‌تنها برای قضاوت در بارهٔ عملکرد یادگیرندگان، بلکه برای داوری در بارهٔ اثربخشی کیفیت آموزش و روشها و مواد آموزشی نیز مفید هستند. بنابراین، معلمان باید از این گونه ارزشیابی حداکثر استفاده را بنمایند و برای تجدید نظر در روش کار خود از نتایج آن سود ببرند.

عمده‌ترین ملاک قضاوت در ارزشیابی آموزشی هدفهای آموزشی از پیش‌تعیین‌شده هستند. معلمان، در آغاز هر درس خود، هدفهایی را در اختیار یادگیرندگان قرار می‌دهند و از آنان می‌خواهند تا در کوششهای یادگیری خود به آنها برسند. بنابراین، ارزشیابی پیشرفت تحصیلی که مهمترین نوع ارزشیابی آموزشی است تعیین می‌نماید که دانش‌آموزان یا دانشجویان تا چه میزانی به هدفهای آموزشی از پیش‌تعیین‌شده رسیده‌اند.

ارزشیابی از برنامه‌های درسی و مواد آموزشی

ارزشیابی آموزشی، گاهی به منظور تعیین اثربخشی **برنامهٔ درسی**[4] یا **مواد آموزشی**[5] انجام می‌شود و شامل ارزشیابی از عواملی چون روشهای آموزشی، کتابهای درسی، مواد دیداری ـ شنیداری، و تدارکات فیزیکی و سازمانی است (گی، ۱۹۹۱). "ارزشیابی از برنامهٔ درسی ممکن است شامل

---

1. student evaluation
2. academic achievement
3. learning
4. curriculum
5. instructional materials

ارزشیابی از یک برنامهٔ کامل و یا ارزشیابی از جنبهٔ کوچکی از یک برنامهٔ کامل، مانند استفاده از فیلمهای آموزشی، باشد" (ص ۱۶). نمونه‌هایی از عناصر برنامه‌های درسی که می‌توان از آنها ارزشیابی به عمل آورد عبارت‌اند از آموزش برنامه‌ای، آموزش انفرادی، آموزش به کمک کامپیوتر، و مانند اینها.

در ارزشیابی از برنامه‌های درسی و مواد آموزشی نیز عمده‌ترین ملاک پیشرفت تحصیلی یا میزان یادگیری دانش‌آموزان و دانشجویان است. اما علاوه بر این از ملاکهای دیگری چون انگیزش معلمان و یادگیرندگان نیز می‌توان سود جست. مثلاً می‌توان تعیین کرد که نظر معلمان نسبت به برنامه چگونه است، یا اینکه یادگیرندگان چگونه به آن واکنش می‌کنند. آیا ممکن است یادگیرندگان بیشتر یادبگیرند اما کمتر به درس و مدرسه علاقه‌مند شوند؟

ملاک دیگر ارزشیابی از برنامهٔ درسی و مواد آموزشی هزینهٔ آنهاست. هر برنامهٔ درسی که قرار است در یک مقیاس گسترده مورد استفاده قرار گیرد باید از لحاظ کارآمدی با هزینه‌ای که صرف آن می‌شود تناسب داشته باشد. با یکسان بودن سایر شرایط، یک برنامهٔ درسی جدید در صورتی از لحاظ هزینه کارآمد به حساب می‌آید که در یکی از جهات زیر قابل دفاع باشد:

- از لحاظ هزینه با سایر برنامه‌ها شبیه باشد، اما به پیشرفت بیشتری بینجامد.
- از لحاظ هزینه کمتر از سایر برنامه‌ها خرج بردارد، اما به پیشرفت بیشتر یا مشابهی بینجامد.
- اگر از برنامه‌های دیگر پرهزینه‌تر است، در عوض به پیشرفت بیشتری بینجامد.

ارزشیابی از پروژه‌ها و برنامه‌های آموزشی

علاوه بر برنامه‌های درسی و مواد آموزشی، برنامه‌ها[1] و پروژه‌های[2] آموزشی ویژه نیز مورد ارزشیابی واقع می‌شوند. منظور از پروژه‌ها و برنامه‌های آموزشی ویژه کوششهایی هستند موقعیتی که به عنوان بخشی از برنامه‌های معمول آموزشگاهها به حساب نمی‌آیند (گی، ۱۹۹۱). "آنها نوعاً ماهیت نوآورانه دارند و طول مدت حضورشان به موفقیتشان وابسته است" (ص ۲۲).

مفاهیم پروژه، برنامه، و برنامهٔ درسی از هم متمایزند. پین (۱۹۹۴، ص ۶) آنها را به گونهٔ زیر از هم تمیز داده است. پروژه یک کوشش مجزا و احتمالاً یکباره است که با استفاده از یک نوآوری می‌کوشد تا تغییری به وجود آورد. "این نوآوری می‌تواند به یک روش آموزش مهارتهای آزمایشگاهی... یا یک رویکرد بهبود نگرش دانش‌آموزان نسبت به یادگیری در یک روستا مربوط باشد" (ص ۶). بنابراین، در ارزشیابی از یک پروژهٔ آموزشی، ارزشیاب باید به دنبال پاسخ این پرسش باشد که آیا پروژهٔ موردنظر از جهاتی اثربخش بوده است یا نه. اگر جواب مثبت باشد آنگاه پروژه به صورت بخشی از برنامه در خواهد آمد. برنامه از پروژه گسترده‌تر و پیچیده‌تر است. "یک پروژهٔ خاص ممکن است یکی از چند نوآوری‌هایی باشد که به منظور کمک به رفع مشکل

1. programs
2. projects

دانش‌آموزان ابتدایی تدارک دیده شده است، در حالی که برنامه یک رویکرد چندوجهی برای حل مشکلات کلی آموزش و پرورش است" (پین، ۱۹۹۴، ص ۶). کوتاه سخن اینکه، به گفتهٔ پین، اصطلاح برنامه یک مشکل یا مسئله را پاسخ می‌دهد اما اصطلاح **پروژه** به یک منظور خاص مربوط می‌شود. برنامهٔ درسی از پروژه و برنامه جامع‌تر است. پین (۱۹۹۴) برنامهٔ درسی را به عنوان مجموعه‌ای از برنامه‌ها تعریف کرده است و مواردی چون مطالعات اجتماعی، خواندن، علوم، و ریاضیات را به عنوان برنامهٔ درسی ذکر کرده و گفته است که برنامهٔ درسی دارای دامنه و توالی است.

ارزشیابی از آموزشگاه

منظور از **ارزشیابی از آموزشگاه**[1] این است که تعیین شود هدفهای پیش‌بینی‌شده برای آموزشگاه به چه میزانی تحقق یافته‌اند و نقاط قوت و ضعف کدام‌اند. ارزشیابی از آموزشگاه دربرگیرندهٔ ارزشیابی از کل برنامهٔ آموزشگاه و مستلزم گردآوری اطلاعات لازم در بارهٔ همهٔ جنبه‌های آموزشگاه است.

یکی از جنبه‌های مهم ارزشیابی از آموزشگاه برنامهٔ آزمون آموزشگاه است. هرچه این برنامه جامع‌تر و کامل‌تر باشد اطلاعات به دست آمده نیز به همان نسبت قابل اعتمادتر و با ارزش‌تر خواهند بود.

مانند سایر ارزشیابیهای آموزشی، در ارزشیابی از آموزشگاه نیز پیشرفت تحصیلی دانش‌آموزان و دانشجویان مهم‌ترین بخش برنامهٔ آزمون و منبع کسب اطلاعات است، زیرا هدف عمدهٔ هر آموزشگاهی افزایش پیشرفت تحصیلی دانش‌آموزان یا دانشجویان آن است. اما، علاوه بر پیشرفت تحصیلی، نگرش یادگیرندگان نیز باید در نظر گرفته شود. همچنین اطلاعات به دست آمده از تغییرات حاصل در استعداد، شخصیت، نگرش، و علاقهٔ یادگیرندگان نیز نکته‌ای قابل توجه است. برای مثال، **مفهوم‌خود (خودپنداره)**[2] دانش‌آموزان و دانشجویان یک متغیر مهم آموزشی است و پژوهشهای زیادی نشان داده‌اند که این متغیر با پیشرفت تحصیلی رابطهٔ نزدیکی دارد و هرچه میزان پیشرفت بالاتر می‌رود سطح مفهوم‌خود دانش‌آموزان نیز به همان نسبت افزایش می‌یابد. از سوی دیگر معلوم گشته است که دانش‌آموزان ضعیف و دارای سطح پایینِ موفقیت از مفهوم‌خود ضعیف‌تری برخوردارند (بلوم[3]، ۱۹۸۲، ترجمهٔ سیف، ۱۳۷۴).

همچنین علاوه بر آزمونهای متداول پیشرفت تحصیلی و متغیرهای شخصیتی، کاربرد پرسشنامه‌های مختلف، مصاحبه، و مشاهده که از طریق آنها اطلاعات لازم از همهٔ افراد و کارکنان آموزشگاه به دست می‌آیند ضروری هستند. به‌ویژه از مدیران، معلمان، و مشاوران اطلاعات ضروری را باید جمع‌آوری کرد.

---

1. school evaluation 2. self-concept 3. Bloom

ارزشیابی از کارکنان آموزشگاه

مقصود از **ارزشیابی از کارکنان آموزشگاه**[1] ارزشیابی از همهٔ کسانی است که به طور مستقیم یا غیر مستقیم مسئول بازده‌های آموزشی مؤسسهٔ آموزشی هستند. هرچند که این ارزشیابی بیشتر به منظور تعیین میزان اثربخشی یا کارآمدی هیأت آموزشی یا معلمان به کار می‌رود و لذا مهم‌ترین نوع ارزشیابی از کارکنان آموزشگاه **ارزشیابی معلم**[2] است، اما سایر افرادی که در بالابردن سطح یادگیری و سایر جنبه‌های موفقیت تحصیلی مؤثرند، مانند مدیران و مشاوران، نیز مشمول این نوع ارزشیابی هستند.

در سالهای اخیر بر ارزشیابی از اثربخشی معلمان سطوح مختلف تحصیلی تأکید زیادی شده است. روشهای ارزشیابی معلم متنوع‌اند و غالباً موارد زیر را شامل می‌شوند:

- مصاحبه با معلم
- مشاهده از کلاس درس معلم
- نظر دانش‌آموزان و دانشجویان
- نظر والدین دانش‌آموزان و دانشجویان
- میزان یادگیری یا پیشرفت دانش‌آموزان و دانشجویان
- خودسنجی یا ارزشیابی از خود
- نظر همکاران معلم
- کارپوشه
- آثار علمی معلم

شاید پیچیده‌ترین نوع ارزشیابی، همین ارزشیابی از کارکنان، به‌ویژه ارزشیابی از کار معلمان است. علت پیچیدگی این روش ارزشیابی کم‌اعتباری و بی‌دقتی وسایل و روشهای اندازه‌گیری و سنجش مورد استفاده در این نوع ارزشیابی است، زیرا هیچ یک از منابع اطلاعاتی پیش‌گفته اطلاعات دقیق و بی‌غرضانه‌ای را که برای یک ارزشیابی سالم ضروری هستند به دست نمی‌دهد. بنابراین، متخصصان ارزشیابی معلم پیشنهاد می‌کنند برای اینکه ارزشیابی سالمتری از معلمان به عمل آید، باید از همهٔ منابع اطلاعاتی موجود داده‌های لازم را به دست آورد و قضاوت نهایی را براساس ترکیبی از آنها انجام داد.

دسته‌بندی ارزشیابیهای آموزشی با توجه به ملاک مورد استفاده

ارزشیابیهای آموزشی را، علاوه بر موضوع ارزشیابی، از لحاظ ملاک مورد استفاده نیز می‌توان تقسیم‌بندی کرد. معروف‌ترین تقسیم‌بندی برای ارزشیابیهای آموزشی تقسیم‌بندی آنها به ارزشیابی وابسته به ملاک و ارزشیابی وابسته به هنجار است.

ارزشیابی وابسته به ملاک

در ارزشیابی وابسته به ملاک یا **ارزشیابی ملاکی**[3]، معیار یا ملاک[4] ارزشیابی از پیش تعیین می‌شود و

1. school personnel evaluation
2. teacher evaluation
3. criterion-referenced evaluation
4. criterion

عملکرد یادگیرنده با توجه به آن ملاک مورد قضاوت قرار می‌گیرد. در این نوع ارزشیابی که بیشتر ویژهٔ پیشرفت تحصیلی دانش‌آموزان و دانشجویان است، نوعی ملاک مطلق که همان هدفهای آموزشی از پیش تعیین شده هستند مورد استفاده قرار می‌گیرد. در این نوع ارزشیابی، از آزمون شوندگان انتظار می‌رود که به هدفهای آموزشی معینی دست یابند یا به سطح بخصوصی از عملکرد برسند.

در استفاده از ملاک مطلق در ارزشیابی پیشرفت تحصیلی، میزان یادگیری یا پیشرفت یادگیرنده با توجه به آنچه آموزش داده شده است ارزشیابی می‌شود. یعنی کسب موفقیت در آزمونهایی که مبتنی بر ملاک مطلق هستند مستلزم یادگیری هدفهای آموزشی در سطحی است که معلم تعیین می‌کند. این نوع ارزشیابی تعیین می‌کند که آیا دانش آموزان یا دانشجویان مطالب و هدفهایی راکه قرار بوده است بیاموزند به میزان دلخواه آموخته‌اند یا نه. گلیزر[1] (۱۹۶۳) ویژگیهای آزمونهای وابسته به ملاک را به نحو زیر توصیف کرده است:

> مفهوم زیربنای اندازه‌گیریهای پیشرفت تحصیلی را این اندیشه می‌سازد که دانش‌اندوزی یک پیوستار تشکیل می‌دهد که در یک سوی آن هیچ‌گونه فراگیری وجود ندارد و در سوی دیگر آن فراگیری کامل قرار دارد. میزان پیشرفت یک فرد در جایی از این پیوستار قرار دارد که به وسیلهٔ رفتارهایی که او ضمن آزمون نشان می‌دهد مشخص می‌گردد. با استفاده از آزمونهای وابسته به ملاک پیشرفت تحصیلی، معلوم می‌شود که پیشرفت یادگیرنده تا چه اندازه به عملکرد مطلوب در نقطهٔ معینی از این پیوستار شبیه است. ملاکی که عملکرد دانش‌آموز، وقتی که به این طریق آزمون می‌شود، با آن مقایسه می‌گردد رفتاری است که هر نقطه را در طول این پیوستار مشخص می‌کند. (ص ۵۱۹)

برای روشن‌شدن مفهوم ارزشیابی وابسته به ملاک به مثالی توجه کنید. فرض کنید ملاک موفقیت در یک دورهٔ آموزش ماشین‌نویسی به نحو زیر بیان شده باشد:

بعد از ده هفته شرکت در یک دورهٔ آموزش ماشین‌نویسی، یادگیرنده باید بتواند در هر دقیقه ۶۰ کلمه را با حداکثر ۵ اشتباه از روی یک متن دست‌نویس ماشین کند.

در ارزشیابی عملکرد یادگیری در ارتباط با هدف بالا، چنانچه یادگیرنده توانسته باشد مطالبی راکه در اختیارش گذاشته می‌شود با حداکثر ۵ اشتباه در هر ۶۰ کلمه ماشین کند، گفته می‌شود که او به هدف آموزشی برنامه دست یافته و میزان پیشرفت او موفقیت‌آمیز بوده است.

ارزشیابی وابسته به هنجار

منظور از **ارزشیابی وابسته به هنجار** یا **ارزشیابی هنجاری**[2] آن نوع ارزشیابی است که به جای یک ملاک مطلق از پیش تعیین شده بر نوعی ملاک نسبی وابسته است. در این نوع ارزشیابی، اگر

1. Glazer
2. norm-referenced evaluation

بخواهیم پیشرفت تحصیلی گروهی دانش‌آموز را ارزشیابی کنیم، عملکرد دانش‌آموزان را با یکدیگر مقایسه می‌کنیم، نه با یک ملاک مشخص از پیش تعیین شده.

علت نام‌گذاری این نوع ارزشیابی به ارزشیابی هنجاری این است که در آن ملاک یا معیار ارزشیابی نوعی **هنجار**[1] گروهی است. بنابراین، در این نوع ارزشیابی می‌توان تعیین کرد که پیشرفت یک دانش‌آموز نسبت به دانش‌آموزان دیگر چه وضعی دارد، اما نمی‌توان مشخص کرد که پیشرفت یک دانش‌آموز نسبت به هدفهای آموزشی یا آنچه آموزش داده شده چگونه است. برای مثال، نمرهٔ ۱۶ یک دانش‌آموز اگر از ۹۰٪ نمرات سایر دانش‌آموزان بالاتر باشد نمرهٔ خوبی در ارزشیابی می‌شود، اما اگر همین نمره در حد ۱۰٪ آزمون شوندگان باشد نمرهٔ ضعیفی به حساب می‌آید. در این نوع ارزشیابی پیشرفت دانش‌آموزان برحسب رتبهٔ درصدی یا نوعی هنجار سنی یا کلاسی و در مقایسه با **گروه هنجار**[2] تعیین می‌شود. (انواع هنجارها و چگونگی محاسبهٔ آنها را در فصل پانزدهم بحث خواهیم کرد.)

مقایسهٔ ارزشیابی وابسته به ملاک با ارزشیابی وابسته به هنجار

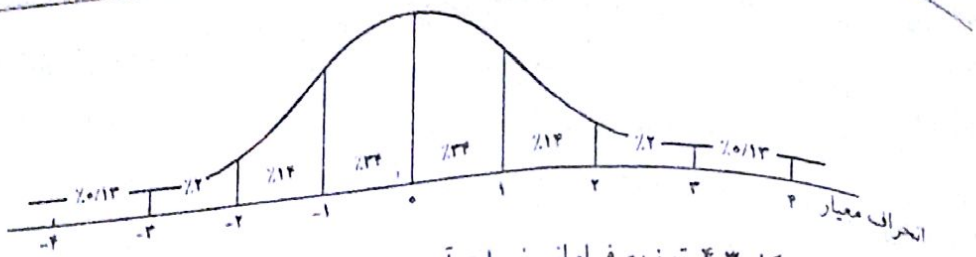
چنان که گفته شد، در ارزشیابی پیشرفت تحصیلی هنجاری هدف عمده این است که تعیین شود یک دانش‌آموز یا دانشجو، در مقایسه با دانش‌آموزان یا دانشجویان دیگر چه وضعی دارد؛ اما در ارزشیابی پیشرفت تحصیلی ملاکی منظور این است که مشخص شود یک دانش‌آموز یا دانشجو چه مقدار از آنچه راکه قرار بوده است بیاموزد آموخته است یا تا چه اندازه به انجام‌دادن وظیفه‌ای که به او محول شده تواناست.

ارزشیابی هنجاری و آزمونهای مبتنی بر ملاک نسبی در سنجش هوش و استعداد و نیز در سنجش شخصیت و سایر جنبه‌های عاطفی به‌کار می‌روند. در مقابل، ارزشیابی ملاکی و آزمونهای مبتنی بر ملاک مطلق به طور عمده در سنجش پیشرفت تحصیلی که به منظور تعیین میزان موفقیت یادگیرندگان در رسیدن به هدفهای آموزشی اجرا می‌شوند مورد استفاده قرار می‌گیرند.

به عنوان نمونه‌ای از ارزشیابی هنجاری می‌توان آزمونهای ورودی ویژهٔ داوطلبان ورود به آموزش عالی (کنکور سراسری) را موردی از ارزشیابی هنجاری نام برد. در این آزمونها، عملکرد هر داوطلب با توجه به عملکرد سایر داوطلبان مورد قضاوت قرار می‌گیرد و رتبه‌ای که هر داوطلب در این آزمونها کسب می‌کند تعیین‌کنندهٔ میزان موفقیت اوست. به همین سبب است که، بی‌توجه به سطح نمرات داوطلبان کنکور سراسری، تنها گنجایش کلاسهای دانشگاهی عامل مهم در انتخاب تعداد قبول‌شدگان افراد در کنکور است، و صرف‌نظر از سطح نمرات داوطلبان این تعداد تغییر نمی‌کند.

در مقابل، امتحانات نهایی مقاطع تحصیلی مختلف دبستان، راهنمایی، و دبیرستان بر نوعی ملاک مطلق که محتوای درسی و هدفهای آموزشی هستند متکی‌اند، و قبولی یا عدم قبولی

1. norm
2. normative group

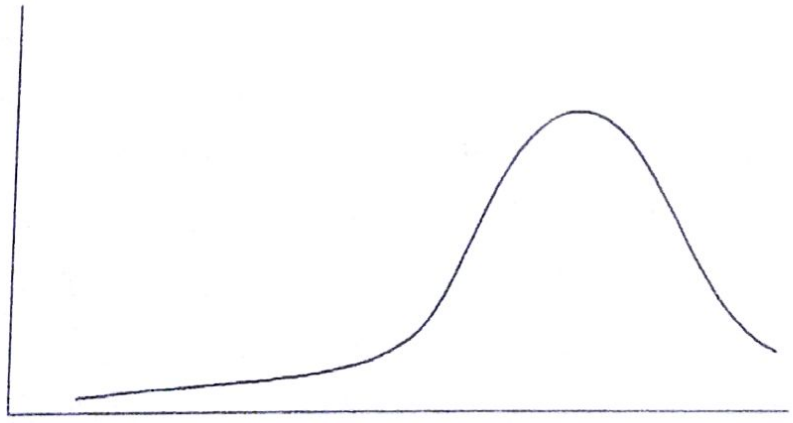


شکل ۳-۴ توزیع فراوانی نمرات آزمونهای هنجاری (توزیع بهنجار)

شرکت‌کنندگان در این امتحانات، صرفاً با توجه به این ملاک نسبتاً مطلق و مستقل تعیین می‌شود. به همین دلیل است که در این امتحانات، برای تعداد قبول‌شدگان درصد مشخصی از پیش تعیین نمی‌شود، بلکه حتی اگر همهٔ شرکت‌کنندگان به حد ملاک تعیین شده برسند قبول اعلام خواهند شد. در نتیجه می‌توان انتظار داشت که در یک زمان تعداد قبول‌شدگان بسیار زیاد و در زمان دیگر این تعداد کمتر از حد انتظار باشد.

یکی دیگر از تفاوتهای مهم میان آزمونهای هنجاری و آزمونهای ملاکی در توزیع فراوانی نمرات حاصل از اجرای آنها یافت می‌شود. نمرات آزمونهای هنجاری معمولاً به صورت توزیع طبیعی پراکنده می‌شوند، در این توزیع اکثر نمرات حول و حوش میانگین قرار دارند و تعداد کمی از آنها بسیار بالاتر از میانگین و به همان نسبت هم تعداد کمی از نمرات بسیار پایین‌تر از میانگین قرار دارند (شکل ۳-۴).

برخلاف نمرات آزمونهای هنجاری، توزیع فراوانی نمرات آزمونهای ملاکی یک منحنی بهنجار نمی‌سازد. بلکه این نمرات دارای توزیعی است که مُعرف منحنیهای دارای کجی منفی است (شکل ۴-۴). این نوع توزیعها نشان‌دهندهٔ بالابودن سطح همهٔ نمرات دانش‌آموزان است که نشان می‌دهد اکثریت بزرگی از آنان به اکثریت هدفهای آموزشی مورد نظر (ملاک مطلق) رسیده‌اند.



شکل ۴-۴ توزیع فراوانی نمرات آزمونهای ملاکی (توزیع کجی منفی)

دسته‌بندی ارزشیابیهای آموزشی با توجه به زمان و هدف استفاده از آنها

گونهٔ دیگر دسته‌بندی ارزشیابیها یا سنجشهای آموزشی با توجه به زمان و هدف استفاده از آنهاست. از این لحاظ، ارزشیابیهای آموزشی و آزمونهای مورد استفادهٔ آنها را به چهار دستهٔ ارزشیابی آغازین، ارزشیابی تکوینی، ارزشیابی تشخیصی، و ارزشیابی تراکمی تقسیم می‌کنند. در زیر هر یک از اینها را توضیح می‌دهیم.

ارزشیابی آغازین یا سنجش آغازین

نخستین ارزشیابی معلم که پیش از انجام فعالیتهای آموزشی او به اجرا درمی‌آید **ارزشیابی آغازین یا سنجش آغازین**[1] نامیده می‌شود. این نوع ارزشیابی به دو منظور یعنی برای پاسخ دادن به دو پرسش زیر مورد استفاده قرار می‌گیرد:

۱. آیا یادگیرندگان بر دانشها و مهارتهای پیش‌نیاز درسِ تازه از قبل تسلط یافته‌اند؟
۲. یادگیرندگان چه مقدار از هدفها و محتوای درس تازه را قبلاً یاد گرفته‌اند؟

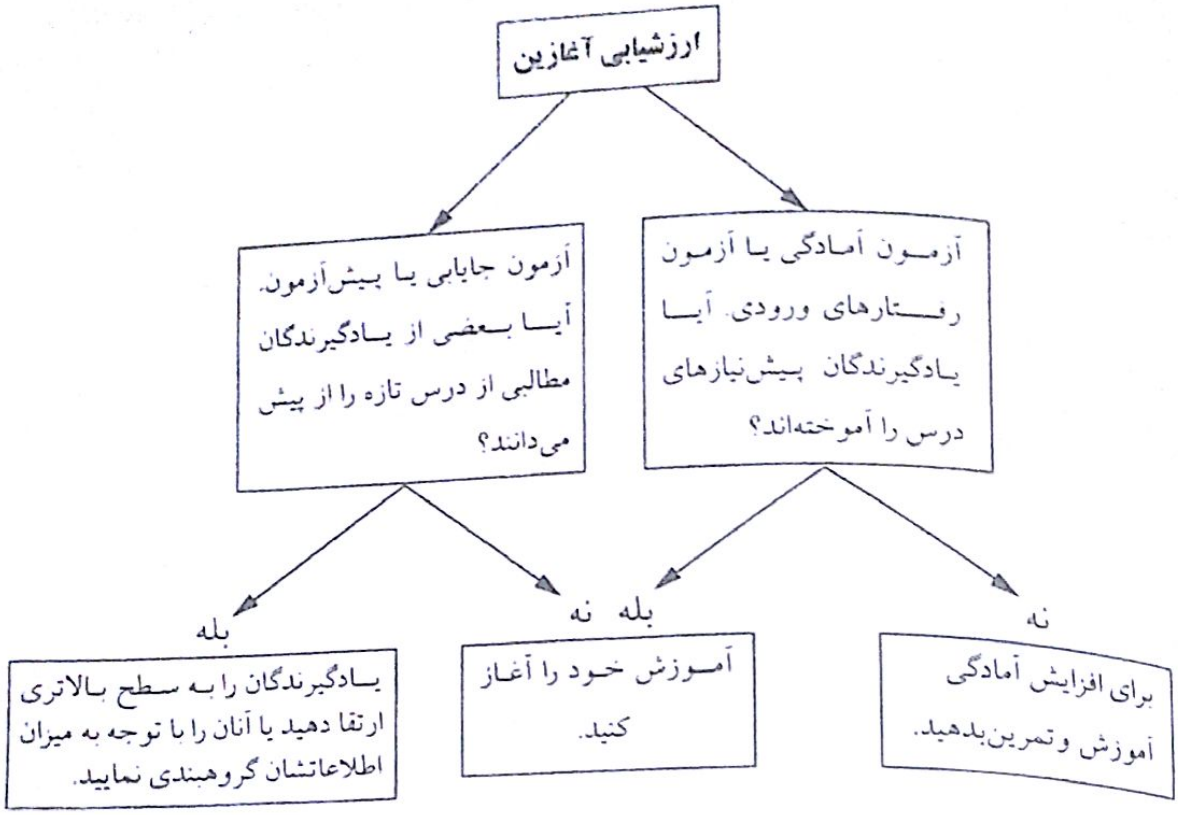
در مورد سؤال ۱، سنجش آغازین به منظور اندازه‌گیری **رفتارهای ورودی**[2] یا مهارتها و دانشهای پیش‌نیاز یادگیری درس تازه به کار می‌رود. از طریق سنجش رفتارهای ورودی معلوم می‌شود که یادگیرندگان برای یادگیری هدفها و محتوای درس تازه قبلاً آمادگیهای لازم را کسب کرده‌اند یا نه. "برای مثال، یک آزمون مربوط به مهارتهای محاسباتی را در آغاز درس جبر، یا یک آزمون دستور زبان انگلیسی را [برای یادگیرندگان انگلیسی‌زبان] در آغاز درس زبان آلمانی می‌توان مورد استفاده قرار داد" (گرانلاند، ۱۹۸۸، ص ۲).

در مورد سؤال ۲، به جای آزمون رفتارهای ورودی، از **آزمون جایابی** یا **آزمون پایه‌گزینی**[3] استفاده می‌شود. آزمون جایابی یا پایه‌گزینی که به آن **پیش‌آزمون**[4] نیز گفته می‌شود عمدتاً همان آزمون نهایی شامل هدفهای یادگیری یا هدفهای آموزشی درس یا ترجیحاً فرم موازی آن است. هدف از اجرای این آزمون تعیین میزان اطلاعات یادگیرندگان از مطالبی است که قرار است به آنان آموزش داده شوند.

متخصصان آموزشی پیشنهاد داده‌اند که بهتر است معلم در ارزشیابی آغازین خود ترکیبی از دو آزمون بالا یعنی آزمون رفتارهای ورودی و آزمون نهایی (که به آن پیش‌آزمون نیز گفته می‌شود) را با هم به اجرا درآورد. اگر نتایج آزمون رفتارهای ورودی نشان دادند که همه یا بعضی از یادگیرندگان برای یادگیری درس تازه آمادگی کافی ندارند، لازم است معلم پیش از آغاز درس تازه به جبران نواقص یادگیریهای قبلی یادگیرندگان یا رفع کم‌وکسر پیش‌نیازهای آنان اقدام نماید (سیف، ۱۳۶۸). اگر نتایج پیش‌آزمون نشان دادند که همه یا بعضی از یادگیرندگان مقداری از مطالب

---

1. preassessment
2. entry behaviors
3. placement test
4. pretest



شکل ۵-۴ نقش ارزشیابی آغازین در تصمیمات آموزشی

درس تازه را از قبل آموخته‌اند، آنگاه ممکن است لازم باشد که معلم در طرح درس خود تغییراتی بدهد، بعضی قسمتهای درس را برای بعضی از دانش‌آموزان حذف کند، و برای بعضی دانش‌آموزان دیگر سطح پیشرفته‌تری از آموزش را به اجرا درآورد. شکل ۵ـ۴ چگونگی استفاده از ارزشیابی آغازین را خلاصه‌وار نشان می‌دهد.

هرچند که ارزشیابی آغازین بسیار مفید است، اما همیشه ضروری نیست. معمولاً معلمی که چند واحد درسی متوالی را به گروه معینی از دانش‌آموزان آموزش می‌دهد، در آغاز هر واحد درسی تازه از یادگیریهای قبلی دانش‌آموزان در واحدهای قبلی به اندازهٔ کافی اطلاع دارد که نیازی به اجرای ارزشیابی آغازین نداشته باشد. اما اجرای ارزشیابی آغازین زمانی ضرورت حتمی دارد که معلم از دانشها و مهارتهای قبلی دانش‌آموزان خود اطلاع کافی نداشته باشد، یعنی موقعی که می‌خواهد به گروه جدیدی از دانش‌آموزان درس تازه‌ای را بدهد. در چنین موقعیتی ارزشیابی آغازین به معلم کمک می‌نماید تا کم‌وکسرهای یادگیریهای پیشین دانش‌آموزان خود را تشخیص داده آنها را برطرف کند و در قراردادن هریک از دانش‌آموزان در بهترین موقعیت یادگیری تصمیمات درست اتخاذ نماید.

ارزشیابی تکوینی

آنچه عمدتاً به منظور کمک به اصلاح موضوع مورد ارزشیابی، یعنی برنامه یا روش آموزشی،

مورد استفاده قرار می‌گیرد **ارزشیابی تکوینی**[1] نام دارد. پوفام (۱۹۷۵، ص ۱۴) مثالهای زیر را از کاربرد ارزشیابی تکوینی ذکر کرده است:

۱. صورتهای اولیهٔ یک مجموعه کتابچهٔ خودآموز

۲. یک برنامهٔ آموزش "باز" تازه تدوین شده که تدوین‌کنندگان آن می‌کوشند تا عناصر مؤثر آن را مشخص کنند

هدف از اجرای ارزشیابی تکوینی در مورد این‌گونه برنامه‌ها یا مواد آموزشی آگاه‌ساختن تولیدکنندگان برنامه از نواقص برنامهٔ خود و کمک به اصلاح آنهاست. هدف از کاربرد ارزشیابی تکوینی در رابطه با پیشرفت تحصیلی دانش‌آموزان و دانشجویان آگاهی یافتن از میزان و نحوهٔ یادگیری آنان برای تعیین نقاط قوت و ضعف یادگیری و نیز تشخیص مشکلات روش آموزشی معلم در رابطه با هدفهای آموزشی است. این ارزشیابی در طول دورهٔ آموزشی، یعنی زمانی که فعالیت آموزشی معلم هنوز در جریان است و یادگیری دانش‌آموزان در حال تکوین یا شکل‌گیری است انجام می‌شود. به همین دلیل به آن نام ارزشیابی تکوینی داده‌اند. بنابراین، معلم با استفاده از ارزشیابی تکوینی می‌تواند در زمانی که هنوز یادگیری در جریان است و امکان رفع مشکلات یادگیری دانش‌آموزان و برطرف کردن نواقص روش آموزشی خود او میسر است به انجام‌دادن چنین کاری اقدام کند.

در ارزشیابی تکوینی، درپایان هر واحد درسی، یک آزمون دقیق و مختصر که حاوی هدفهای آموزشی آن واحد است اجرا می‌شود و براساس نتایج حاصل معلوم می‌گردد که یادگیرندگان کدام یک از هدفهای آموزشی را یادگرفته‌اند و در یادگیری کدام یک از هدفهای آموزشی ناموفق مانده‌اند، تا معلم پیش از پرداختن به واحد درسی بعدی به رفع نواقص یادگیری دانش‌آموزان در واحد فعلی بپردازد. ضمناً معلم با مراجعه به نتایج این آزمونها از مشکلات روش آموزشی خود نیز مطلع می‌شود و پیش از آغاز آموزش واحد بعدی به رفع آن مشکلات اقدام می‌کند.

برای بهره‌گیری کامل از نتایج ارزشیابی تکوینی، بیان هدفهای آموزشی به صورت دقیق و رفتاری ضروری است. برای اینکه معلم بتواند در جریان یادگیری توفیق دانش‌آموزان در رسیدن به هدفهای مشخص واحدهای متوالی درس را ارزشیابی کند، باید آنها را به‌طور دقیق بیان نماید تا در موقع برخورد دانش‌آموزان با مشکل، به‌سادگی آن مشکلها را تشخیص دهد و به رفع آنها بکوشد.

بلوم و همکاران او در کتاب *راهنمای عملی برای ارزشیابی تکوینی و ارزشیابی تراکمی از آموخته‌های دانش‌آموزان* (۱۹۷۱) در رابطه با استفاده از ارزشیابی تکوینی در کلاسهای درسی معمولی مراحل زیر را پیشنهاد کرده‌اند:

۱. موضوع یا محتوای درس را به واحدهای کوچک درسی که آموزش هریک از آنها به بیشتر از یکی دو هفته وقت نیاز نداشته باشد تقسیم کنید.

---

1. formative evaluation

۲. بازده‌های یادگیری هریک از واحدهای درسی را برحسب انواع هدفهای آموزشی (دانستن، فهمیدن، کاربستن...) تعریف کنید.
۳. به تهیه و اجرای آزمونهای تکوینی مختصر در پایان هر واحد درسی اقدام کنید، و تعیین نمایید که دانش‌آموزان در هر واحد درسی به چه حدی از تسلط رسیده‌اند.
۴. نتایج آزمونهای تکوینی را تحلیل کنید تا مشکلات یادگیری هریک از دانش‌آموزان را تشخیص دهید.
۵. برای رفع مشکلات یادگیری دانش‌آموزان، آنان را به خواندن مطالب مشخص که در آن با مشکل مواجه شده‌اند و همچنین به استفاده از امکانات مختلف (کمک معلمان، آموزش برنامه‌ای، وسایل کمک‌آموزشی، و جز اینها) هدایت کنید.
۶. از نتایج ارزشیابی تکوینی برای بهبود شرایط و روشهای آموزشی استفاده کنید.

در ارزشیابی تکوینی، آزمونهایی که مورد استفاده قرار می‌گیرند مختصرند و به یک موضوع معین و یک یا چند هدف دقیق آموزشی منحصر می‌شوند. برای مثال، معلمی که در یکی از کلاسهای دبستان موضوع جمع اعداد دو رقمی را درس می‌دهد و هدف آموزشی خود را این‌گونه بیان کرده است: ***دانش آموزان باید بتوانند هر دو عدد دو رقمی راکه به آنان داده می‌شود درست با هم جمع کنند***، آزمون تکوینی او شامل تعدادی اعداد دورقمی خواهد بود که به دانش‌آموزان می‌دهد تا با هم جمع کنند، مانند اعداد زیر:

۵۰	۱۷	۹۰	۱۵	۶۱
+۲۰	+۱۴	+۱۱	+۱۷	+۱۲

از آنجاکه نتایج ارزشیابی تکوینی به‌طور عمده برای وارسی پیشرفت یادگیرندگان و تعیین نواقص روش آموزشی معلم به‌کار می‌روند، بهتر است از آنها به منظور نمره‌گذاری استفاده نشود. با این‌حال، در شرایط بخصوص می‌توان از نمره‌های این آزمونها در ارزشیابی نهایی از کار دانش‌آموزان و اثربخشی طرح آموزشی معلم استفاده کرد (نگاه کنید به سیف، ۱۳۶۸).

لازم به یادآوری است که آزمونهای مورد استفاده در ارزشیابی تکوینی آزمونهای ملاکی یا وابسته به ملاک مطلق هستند، زیرا این آزمونها برای اندازه‌گیری هدفهای دقیق آموزشی از پیش تعیین شده‌که همان هدفهای واحد درسی یا واحد آموزشی هستند تدوین می‌شوند. در ارزشیابی تکوینی هدفْ تعیین میزان توفیق یادگیرنده در رسیدن به تک‌تک هدفهای واحد درسی است و هیچ‌گونه مقایسه‌ای میان افراد مختلف مورد نظر نیست.

ارزشیابی تشخیصی

نوع دیگر ارزشیابی که آن نیز در جریان آموزش انجام می‌گیرد **ارزشیابی تشخیصی**[1] نام دارد. علت

1. diagnostic evaluation

این نام‌گذاری آن است که این ارزشیابی با هدف تشخیص مشکلات یادگیری دانش‌آموزان به کار می‌رود.

ارزشیابی تشخیصی زمانی مورد استفاده قرار می‌گیرد که معلم با مشکلات مبرم و مکرری در مورد یک یا چند دانش‌آموز روبه‌رو می‌شود که با روشهای اصلاحی معمول ارزشیابی تکوینی قابل رفع شدن نیستند (به عنوان نمونه، اگر دانش‌آموزی در درس خواندن، ریاضیات، یا درسی دیگر مرتباً شکست می‌خورد و معلم از طریق تغییر روش آموزشی خود یا تدابیر جبرانی دیگر از عهدهٔ حل مشکل برنمی‌آید، آن گاه اجرای یک آزمون تشخیصی و یافتن ریشه‌های مشکل ضروری است. لین و گرانلاند (۲۰۰۰) در مقایسهٔ ارزشیابی تشخیصی با ارزشیابی تکوینی گفته‌اند "اگر بخواهیم از قیاس با پزشکی استفاده کنیم، سنجش تکوینی [ارزشیابی تکوینی] برای مشکلات سادهٔ یادگیری جنبهٔ کمکهای اولیه را دارد ولی سنجش تشخیصی به دنبال کشف علتهای زیربنایی مشکلاتی است که به کمکهای اولیه پاسخ نمی‌دهند" (ص ۴۱).

نیتکو (۲۰۰۱) نیز در رابطه با ارزشیابی تشخیصی گفته است بعضی وقتها آموزشهایی که معلم به یک دانش‌آموز دارای مشکل یادگیری می‌دهد برای او اثربخش نیست، زیرا این دانش‌آموز به کمکهای ویژه نیاز دارد. "سنجشی که برخی از اطلاعات موردنیاز این‌گونه تصمیم‌گیری را فراهم می‌آورد سنجش تشخیصی نام دارد" (ص ۹). از جمله اطلاعاتی که ارزشیابی یا سنجش تشخیصی به دست می‌دهد جواب به این سؤال است که "معلم چه نوع فعالیت یادگیری را باید برای این‌گونه دانش‌آموزان برگزیند که مناسب وضع آنان باشد و دانش‌آموزان را کمک کند تا به هدفهای یادگیری موردنظر برسند؟" (ص ۹).

تفاوت آزمونهای تشخیصی با آزمونهای پیشرفت تحصیلی در آن است که "آزمونهای پیشرفت تحصیلی [مانند آزمونهای معمولی مورد استفادهٔ معلمان] پیشرفت یا توفیق دانش‌آموزان را در درسهای مختلف می‌سنجند، اما آزمونهای تشخیصی توصیف کاملی از نقاط قوت و ضعف یادگیرندگان در زمینه‌های مهارتی خاص به دست می‌دهند" (اِگن و کاوچاک، ۲۰۰۱، ص ۶۴۳). ارزشیابی تشخیصی، علاوه بر تشخیص دادن مشکلات یادگیری دانش‌آموزان، باید بتواند روشهای مناسب رفع مشکلات را نیز به معلمان نشان دهد. گفتن اینکه یک دانش‌آموز بخصوص در آزمون درس ریاضی ضعیف است مشکل‌گشا نیست. معلم درس ریاضی این دانش‌آموز به اطلاعات بیشتری نیاز دارد. دانش‌آموز در چه نوع سؤالهایی با اشکال مواجه شده است؟ آیا این سؤالها یک الگوی خاص را نشان می‌دهند؟ آیا این دانش‌آموز بیشتر در سؤالهای مربوط به جمع مشکل دارد یا تفریق؟ آیا در عمل انتقال با مشکل مواجه می‌شود یا گرفتاری او در جای دیگری است؟ بنا به گفتهٔ لین و گرانلاند (۲۰۰۰)، "هدف سنجش تشخیصی تعیین علتهای مشکلات مکرر یادگیری و تدوین نقشه برای اعمال اصلاحی است" (ص ۴۱). مهرنز و لهمان (۱۹۸۴) در این باره توضیحات زیر را داده‌اند:

یک آزمون تشخیصی خوب نه تنها معلم را از ضعف دانش‌آموز آگاه می‌سازد، بلکه زمینه‌های ضعف او را هم نشان می‌دهد ــ مانند ناتوانی او در درک واژگان یا جمع یا عمل انتقال. با این حال، نتایج این نوع ارزشیابی روابط علّی را نشان نخواهد داد. ممکن است معلم متوجه شود که مشکلات چه هستند اما نخواهد فهمید که علت آنها چیست؟ مثلاً فرض کنید معلوم شده که یک دانش‌آموز بخصوص در جبر ضعیف است. این ضعف ممکن است ناشی از ناتوانی ذهنی او، یا ضعف خواندن، یا ضعف بینایی، یا مشکلات روانی ــ حرکتی، یا مسائل عاطفی، یا سایر عوامل باشد. معلم برای رسیدن به یک راه حل منطقی باید این عوامل را در نظر بگیرد. در غیر این‌صورت، ممکن است مشکل به ظاهر و به فوریت حل شود، اما اگر ریشه‌های گرفتاری برطرف نشوند احتمالاً در سایر زمینه‌های درسی رخ خواهد نمود. (مهرنز و لِهمان، ۱۹۸۴، ص ۴۱۱)

ابزارهای مورد استفادهٔ سنجش تشخیصی معمولاً آزمونهای تشخیصی وابسته به هنجار ویژه هستند که برای همین منظور تهیه شده‌اند. علاوه براینها، از آزمونهای وابسته به ملاک و فنون مشاهده نیز استفاده می‌شود. چنان که قبلاً توضیح دادیم، در آزمونهای وابسته به هنجار عملکرد دانش‌آموز با عملکرد گروه هنجار مقایسه می‌شود، اما در آزمونهای وابسته به ملاک عملکرد دانش‌آموز با یک معیار مطلق مثلاً هدفهای یادگیری مقایسه می‌شود. به نظر سیفرت[1] (۱۹۹۱)، برای تشخیص مشکلات یادگیری دانش‌آموزان، آزمونهای تشخیصی وابسته به ملاک کارسازتر از آزمونهای تشخیصی وابسته به هنجار هستند، زیرا نتایج آزمونهای تشخیصی "باید نشان دهند که دانش‌آموزان چه می‌دانند [یا چه نمی‌دانند] نه اینکه در مقایسه با دانش‌آموزان دیگر چه وضعی دارند" (ص ۵۱۹).

آزمونهای تشخیصی استاندارد شده عموماً برای سطح دبستان تهیه شده‌اند و بیشتر موضوعهای ریاضیات و خواندن را شامل می‌شوند. این آزمونها معمولاً به صورت انفرادی اجرا می‌شوند و در قیاس با آزمونهای پیشرفت تحصیلی سؤالهای بیشتری را در بر می‌گیرند. برای نمونه یک آزمون تشخیصی درسِ خواندن شامل موارد زیر است: "تشخیص حروف، تحلیل کلمات، واژگان دیداری، واژگان در متن، و خواندن و فهمیدن" (اگن و کاوچاک، ۲۰۰۱، ص ۶۴۴).

ارزشیابی تراکمی

در **ارزشیابی تراکمی**[2]، تمامی آموخته‌های دانش‌آموزان در طول یک دورهٔ آموزشی تعیین می‌شوند و هدف آن نمره‌دادن به دانش‌آموزان و قضاوت در بارهٔ اثربخشی کار معلم و برنامهٔ درسی یا مقایسهٔ برنامه‌های مختلف درسی با یکدیگر است. از آنجا به این نوع ارزشیابی ***تراکمی*** یا مجموعی می‌گویند که به وسیلهٔ آن می‌توان یادگیریهای متراکم یا مجموع یادگیریهای دانش‌آموزان را در طول یک دورهٔ آموزشی اندازه‌گیری کرد. این نوع ارزشیابی معمولاً در پایان دورهٔ آموزشی به عمل می‌آید و به همین سبب به آن **ارزشیابی پایانی** نیز می‌گویند. با این حال، می‌توان از آن به دفعات

1. Seifert
2. summative evaluation

بیشتر نیز در طول دورهٔ آموزشی استفاده کرد. آزمونهایی که در ارزشیابی تراکمی مورد استفاده قرار می‌گیرند بسیار جامع و مفصل هستند. این آزمونها نمونهٔ کاملی از محتوای درس و هدفهای آموزشی را در بر می‌گیرند. لین و گرانلاند (۲۰۰۰) در رابطه با روشها و فنون مورد استفاده در ارزشیابی تراکمی گفته‌اند که "هدفهای آموزشی تعیین‌کنندهٔ آنها هستند، اما آنها معمولاً آزمونهای پیشرفت تحصیلی معلم ساخته، مشاهده، سنجش عملکردهای مختلف یادگیرندگان (مانند کار آزمایشگاهی و گزارش شفاهی)، و سنجش محصولات یادگیری (نوشته‌ها، رسم و نقاشی، گزارش تحقیقات) را شامل می‌شوند" (ص ۴۲).

هدف ارزشیابی تراکمی هم نمره‌دادن و اتخاذ تصمیمهای مختلف (ارتقای دانش‌آموزان به کلاسهای بالاتر، دادن امتیازهای تحصیلی، اعزام به دوره‌های عالی‌تر تحصیلی، اعطای بورس تحصیلی، و غیره) و هم تعیین اثربخشی طرح آموزشی و هدفهای معلم و میزان توفیق او در پیاده‌کردن مراحل مختلف طرح آموزشی است. بنابراین، آزمونهای مورد استفاده در ارزشیابی تراکمی هم آزمونهای هنجاری یا وابسته به ملاک نسبی هستند، و هم آزمونهای ملاکی یا وابسته به ملاک مطلق.

در رابطه با ارزشیابی از برنامه‌ها یا مواد آموزشی نیز باید گفت که ارزشیابی تراکمی در پایان برنامه، یعنی زمانی که به طور کامل اجرا شده و به نتیجه رسیده است، انجام می‌گیرد و هدف آن فراهم آوردن اطلاعات مبتنی بر داوریهای ارزشیابانه در بارهٔ برنامه برای مصرف‌کنندگان بالقوهٔ آن برنامه است. برای مثال، پس از آنکه یک برنامهٔ درسی به طور کامل آماده شد، با استفاده از یک نمونهٔ ملی از دانش‌آموزان، معلمان، و مدارس در سطحی که برنامهٔ درسی برای آن تدارک دیده شده است یک ارزشیابی تراکمی انجام می‌شود و نتایج آن در اختیار کسانی که آن برنامه را مورد استفاده قرار خواهند داد گذاشته می‌شود.

اگر ارزشیابیهای تراکمی و تکوینی را با هم مقایسه کنیم، خواهیم دید که مخاطبان و موارد استفادهٔ این دو نوع ارزشیابی با هم فرق دارند. "در ارزشیابی تکوینی، مخاطبان یا استفاده‌کنندگان از نتایج ارزشیابی طراحان و تولیدکنندگان برنامه هستند، در حالی که مصرف‌کنندگان نتایج ارزشیابی تراکمی کسانی هستند که برنامه را مورد استفاده قرار خواهند داد یا فعلاً از آن استفاده می‌کنند" (پوفام، ۱۹۷۵، ص ۱۶). علاوه بر این، ارزشیابی تکوینی به تصمیمات مربوط به توسعه و تولید برنامه کمک می‌کند، اما ارزشیابی تراکمی تصمیمات مربوط به ادامه، قطع، یا گسترش برنامه را یاری می‌دهد. در جدول ۳ـ۴ ارزشیابیهای پیش از آموزش، ضمن آموزش، و پس از آموزش را از جهات مختلف با هم مقایسه کرده‌ایم.

جدول ۴ـ۳ ویژگیهای مختلف ارزشیابیهای پیش از آموزش، ضمن آموزش، و پس از آموزش (اقتباس از آثار گرانلاند و لین، ۱۹۹۰، ص ۱۱۳ و لین و گرانلاند، ۲۰۰۰، ص ۱۴۲)

نوع ارزشیابی	ارزشیابی پیش از آموزش		ارزشیابی ضمن آموزش	ارزشیابی پس از آموزش	
نام ارزشیابی	آمادگی	جایابی	تکوینی	تشخیصی	تراکمی یا مجموعی
هدف ارزشیابی	سنجش رفتارها و مهارتهای پیش‌نیاز و جبران نواقص یادگیریهای پیشین	سنجش میزان آگاهی دانش‌آموزان از هدفها و محتوای درس و گماردن دانش‌آموزان به گروههای آموزشی مختلف	فراهم آوردن بازخورد برای معلم و تشخیص و رفع مشکلات احتمالی یادگیری دانش‌آموزان و نواقص آموزش معلم	تشخیص مشکلات ویژه و مکرر دانش‌آموزان که با اقدامات متداول معلمان قابل رفع نیستند و کمک به رفع آنها	نمره‌گذاری، صدور گواهینامهٔ تحصیلی، و داوری دربارهٔ صلاحیت حرفه‌ای معلمان
زمان ارزشیابی	آغاز ترم تحصیلی یا واحد آموزشی	آغاز ترم تحصیلی یا واحد آموزشی	به طور متناوب ضمن آموزش	در زمانهای لازم ضمن آموزش	پایان ترم تحصیلی یا واحد درسی
وسیلهٔ اندازه‌گیری	آزمون درحد تسلط وابسته به ملاک	آزمون درحد تسلط وابسته به ملاک و آزمون وابسته به هنجار	آزمون درحد تسلط وابسته به ملاک	آزمون در حد تسلط وابسته به ملاک و آزمون تشخیصی میزان شدهٔ وابسته به هنجار، و مشاهده	آزمون در حد تسلط وابسته به ملاک و آزمون وابسته به هنجار
ماهیت نمونه‌گیری	نمونهٔ محدودی از مهارتهای پیش‌نیاز	نمونهٔ بزرگی از همهٔ هدفهای درسی	نمونهٔ محدودی از تکالیف یادگیری	نمونهٔ محدودی از غلطهای ویژه	نمونهٔ بزرگی از همهٔ هدفهای درس
تأکید ارزشیابی	دانشها و مهارتهای پیش‌نیاز	هدفها و محتوای درس	بخش معینی از درس	غلطها و مشکلات ویژهٔ دانش‌آموزان	هدفها و محتوای درس
سطح دشواری سؤالها	نسبتاً ساده	ساده و دشوار	متناسب با مطالب آموزش داده شده	نسبتاً ساده	ساده و دشوار

دسته‌بندی ارزشیابیهای آموزشی با توجه به ارزشیابان

دو نوع ارزشیابی آموزشی دیگر که با ارزشیابی تکوینی و ارزشیابی تراکمی رابطهٔ نزدیکی دارند **ارزشیابی درونی**[1] و **ارزشیابی بیرونی**[2] هستند. منظور از درونی این است که ارزشیابی توسط کارکنان و دست‌اندرکاران تهیه و اجرای برنامه یعنی افراد درون برنامه انجام می‌شود، و منظور از بیرونی این است که ارزشیابی به وسیلهٔ اشخاصی که خارج از برنامه هستند انجام می‌شود. به طور کلی، ارزشیابی درونی بیشتر مناسب ارزشیابی تکوینی است و ارزشیابی بیرونی غالباً برای ارزشیابی تراکمی مفید است.

ارزشیابی درونی یا همان ارزشیابی تکوینی داوری دربارهٔ کیفیت یا ارزشِ مواد آموزشی، روشهای آموزشی، برنامه‌های درسی، یا برنامه‌های آموزشی در ضمن طراحی و اجرای آنهاست.

1. internal evaluation 2. external evaluation

هدف از این ارزشیابی کمک به شکل‌گیری، تجدیدنظر، اصلاح، و بهبود محصول آموزشی است. اصطلاح ارزشیابی درونی به جای ارزشیابی تکوینی غالباً در رابطه با مواد و برنامه‌های آموزشی به کار می‌رود، اما در ارتباط با فعالیتهای آموزشیِ معلمان عموماً از اصطلاح ارزشیابی تکوینی استفاده می‌شود. بنا به گفتهٔ نیتکو (۲۰۰۱)، هنگامی که معلمان از تجارب قبلی خود استفاده می‌کنند و در درسها یا مواد آموزشی تجدیدنظر به عمل می‌آورند به ارزشیابی تکوینی می‌پردازند. بعضی وقتها ما دربارهٔ انجام *ارزشیابی تکوینی از دانش‌آموزان* گفتگو می‌کنیم. معنی این سخن آن است که ما دربارهٔ کیفیت پیشرفت دانش‌آموزان از یک هدف یادگیری، زمانی که فرایند یادگیری هنوز در جریان است، داوری می‌کنیم. این‌گونه داوری به ما کمک می‌کند تا گامهای بعدی دانش‌آموزان را در یادگیری هدایت نماییم (ص ۹). بنابراین، ارزشیابی تکوینی یا ارزشیابی درونی عبارت است از "آن نوع ارزشیابی که هنگام جریان داشتن یک فرایند خلاق طرح و اجرا می‌شود و هدفش افزایش سطح رشد و پیشرفت عملکرد دانش‌آموزان یا برنامه‌های آموزشی است" (گلیکسون[1]، ۲۰۰۳، ص ۲۲۸).

برخلاف ارزشیابی درونی که در آن شخصِ طراح یا تولیدکنندهٔ برنامه به ارزشیابی از برنامهٔ خود می‌پردازد، ارزشیاب بیرونی معمولاً کسی است که از بیرون برای این منظور گمارده می‌شود و به برنامهٔ مورد ارزشیابی وابسته نیست. علاوه بر این، ارزشیابی درونی بیشتر برای این منظور به کار می‌رود که نشان دهد یادگیرندگان به هدفهای برنامه رسیده‌اند یا نه، در حالی که هدف از اجرای ارزشیابی بیرونی این است که معلوم گردد دانش‌آموزانی که با روش خاصی آموزش دیده‌اند نسبت به دانش‌آموزانی که برنامهٔ دیگری را گذرانده‌اند چه وضعی دارند (گی، ۱۹۹۱).

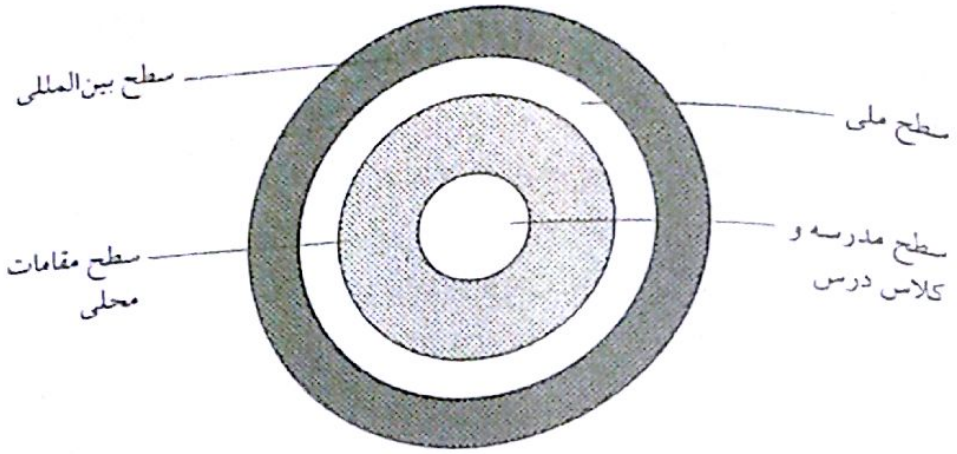
ارزشیابی بیرونی معادل ارزشیابی تراکمی به معنی وسیع آن است. این ارزشیابی داوری دربارهٔ کیفیت یا ارزشِ مواد آموزشی، روشهای آموزشی، برنامه‌های درسی، برنامه‌های آموزشی، یا نظامهای آموزشی است. ارزشیابی تراکمی یا ارزشیابی مجموعی نقاط قوت و ضعف را نشان می‌دهد و میزان توفیق برنامه، روش، یا نظام را در رسیدن به هدفهایش مشخص می‌کند. این ارزشیابی به مصرف کنندهٔ یک محصول آموزشی می‌گوید که آن محصول آموزشی در چه شرایطی و تا چه میزانی مفید و اثربخش است. چنان که قبلاً گفتیم، ارزشیابی تراکمی به وسیلهٔ معلمان و برای تعیین میزان توفیق دانش‌آموزان و دانشجویان در رسیدن به هدفهای یادگیری نیز انجام می‌گیرد. نیتکو (۲۰۰۱) در این باره می‌گوید،

> بعضی وقتها ما از **ارزشیابی تراکمی از دانش‌آموزان** گفتگو می‌کنیم. منظور ما از این سخن داوری دربارهٔ کیفیت یا ارزش پیشرفت دانش‌آموز بعد از پایان یافتن یا تکمیل فرایند آموزشی است. منظور کردن نمرات حرفی (الف، ب، و غیره) در گزارش پیشرفت دانش‌آموزان مثالی از ارزشیابی تراکمی است. (ص ۹)

مکبیت و مکگلین (۲۰۰۲) در ارتباط با ارزشیابی از آموزشگاه گفته‌اند، "ما اصطلاح 'ارزشیابی درونی' را به منظور وارسی و بازبینی جنبه‌های مختلف عملکرد آموزشگاه به وسیلهٔ

---

1. Gullickson



شکل ۶-۴ چهار سطح ارزشیابی (مک‌بیت و مک‌گلین، ۲۰۰۲، ص ۱۷)

صاحبان اصلی آن ــ یعنی کارکنان، دانش‌آموزان، و پدران و مادران ــ به کار می‌بریم" (ص ۱۴). آنان ارزشیابی درونی را با **ارزشیابی از خود (خودارزشیابی)**[1] معادل به کار می‌برند، ولی در رابطه با ارزشیابی بیرونی می‌گویند: "ارزشیابی بیرونی یعنی بازبینی و وارسی فعالیتهای آموزشگاه و گزارش نتایج آن به وسیلهٔ کسانی که اعضای سازمان آموزشگاه به حساب نمی‌آیند" (ص ۱۵).

مک بیت و مک گلین (۲۰۰۲) هر دو نوع ارزشیابی درونی و بیرونی را ضروری می‌دانند و ارتباط بین آنها را به صورت شکل ۶-۴ نشان داده‌اند. دقت کنید که در این شکل آموزشگاه در مرکز قرار داده شده، برای اینکه تأکید شود که خودِ مدرسه نیروی محرک اصلی برای ارزشیابی است. بنابراین، مدرسه باید به پرسشهای زیر بپردازد: "ما تا چه اندازه خوب هستیم؟" و "از کجا می‌دانیم؟"

خطاب این پرسشها به کسان زیادی است، اما مخاطب اصلی معلمان هستند. زیرا معلمان عامل اصلی‌اند. پس از آنکه معلمان عملکرد خود را مورد ارزشیابی قرار دادند، آن‌گاه دیگران نیز می‌توانند اظهارنظر کنند ــ منظور از دیگران همکاران، مدیران، و دانش‌آموزان است.

به همین ترتیب، مدیران به ارزشیابی از خودشان می‌پردازند و دیگران نیز آنان را مورد ارزشیابی قرار می‌دهند ــ منظور از دیگران هم مدیران بالاتر، بازرسان، و همکاران است، و هم معلمان، دانش‌آموزان، و پدران و مادران. همهٔ این جنبه‌های متفاوت در درون مفهوم "ارزشیابی از خود" قرار می‌گیرند.

لایهٔ دوم شکل ۶-۴ ارزشیابی بیرونی را شامل می‌شود. به همین منظور مقامات محلی می‌توانند مدارس خود را مورد ارزشیابی قرار دهند. این مقامات در ضمن ارزشیابی از مدرسه‌های زیرنظر خود و پاسخ دادن به این پرسش که "مدرسه‌های ما چقدر خوب‌اند؟" باید به این سؤال هم جواب دهند که "ما به عنوان مقامات محلی چقدر خوب هستیم؟"

---

1. self-evaluation

لایهٔ سوم شکل ۶-۴ ارزشیابی بیرونی در سطح سوم را نشان می‌دهد. در این نوع ارزشیابی سؤال زیر مورد نظر است "نظام آموزشی ما تا چه اندازه به حفظ و ارتقای کیفیت آموزش و پرورش مملکت کمک می‌کند؟" بنا به گفتهٔ مک بیت و مک گلین (۲۰۰۲)، "برای پاسخ دادن به این پرسش، وزارت آموزش و پرورش می‌تواند کمیسیونی از افراد مستقل تشکیل دهد تا این ارزشیابی را انجام دهند" (ص ۱۸).

سرانجام اینکه لایهٔ آخری نشان‌دهندهٔ ارزشیابی در سطح بین‌المللی است. سؤالی که در اینجا مطرح است این است که "در مقایسه با سایر کشورها وضع ما چگونه است؟" از جمله منابع و امکانات مورد استفاده در این نوع ارزشیابی می‌توان به سازمانهای جهانی مانند یونسکو (بخش آموزشی، علمی، و فرهنگی سازمان ملل متحد) تیمس (سومین مطالعهٔ ریاضیات و علوم بین‌المللی) و مانند اینها اشاره کرد. "به طور خلاصه، ارزشیابی از خود [ارزشیابی درونی] برای معلمان، مقامات محلی، و وزارتخانه‌های آموزش و پرورش است و بهترین نظامها عناصر ارزشیابی درونی و ارزشیابی بیرونی را با هم به کار می‌گیرند" (مک بیت و مک گلین، ۲۰۰۲، ص ۱۸).

هر یک از دو نوع ارزشیابی درونی و بیرونی نسبت به دیگری دارای محاسن و معایبی است. مثلاً ارزشیاب درونی خیلی بیشتر از هر ارزشیاب بیرونی در بارهٔ برنامه اطلاع دارد، اما ممکن است به سبب نزدیکی زیاد با برنامه نتواند کاملاً بی‌طرف بماند. بنابراین، اصولاً ارزشیابان بیرونی از ارزشیابان درونی عینی‌تر و بی‌طرفانه‌تر عمل می‌کنند، اما برای آنان مشکل است که به اندازهٔ ارزشیابان درونی از جریان برنامه مطلع باشند.

خلاصهٔ فصل

ما در این فصل انواع ارزشیابیهای آموزشی را از جهات مختلف و به صورتهای زیر دسته‌بندی کردیم:

۱. در قالب رویکردهای مختلف:
   الف ـ رویکرد مبتنی بر هدف
   ب ـ رویکرد مبتنی بر مدیریت
   پ ـ رویکرد مبتنی بر مصرف‌کننده
   ت ـ رویکرد مبتنی بر نظر متخصصان
   ث ـ رویکرد مبتنی بر نظرهای مختلف متخصصان
   ج ـ رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان

۲. در قالب الگوهای مختلف:
   الف ـ الگوی ارزشیابی تایلر
   ب ـ الگوی ارزشیابی سیپ

ب ـ الگوی ارزشیابی پو سی پل اِنْ یا سی اس ای
ت ـ الگوی ارزشیابی اسکریون
ث ـ الگوی ارزشیابی هدف آزاد
ج ـ الگوی ارزشیابی مبتنی بر نظر متخصصان
چ ـ الگوی ارزشیابی مبتنی بر اختلاف‌نظرهای متخصصان
ح ـ الگوی ارزشیابی سیمایی

۳. با توجه به موضوع ارزشیابی:
الف ـ ارزشیابی از دانش‌آموزان و دانشجویان
ب ـ ارزشیابی از برنامهٔ درسی
پ ـ ارزشیابی از آموزشگاه
ت ـ ارزشیابی از کارکنان آموزشگاه

۴. با توجه به ملاک مورد استفاده:
الف ـ ارزشیابی وابسته به ملاک
ب ـ ارزشیابی وابسته به هنجار

۵. با توجه به زمان و هدف:
الف ـ ارزشیابی آغازین
ب ـ ارزشیابی تکوینی
پ ـ ارزشیابی تشخیصی
د ـ ارزشیابی تراکمی

۶. با توجه به ارزشیابان:
الف ـ ارزشیابی درونی
ب ـ ارزشیابی بیرونی

در این فصل هریک از دسته‌بندیهای فوق از ارزشیابیهای آموزشی را تعریف کردیم و با ذکر مثال رابطهٔ بین اعضای هر دسته را با یکدیگر توضیح دادیم. در ضمن، چگونگی استفاده از این ارزشیابیها و موقعیتی را که می‌توان از آنها استفاده کرد نیز مورد بحث قرار دادیم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. نمره‌گذاری با توجه به منحنی طبیعی نوعی تفسیر نمرات آزمون با توجه به ملاک مطلق است.
الف ـ صحیح
ب ـ غلط

۲. اکثر آزمونهای استاندارد شده یا میزان شده ملاک نسبی را به کار می‌گیرند.

الف ـ صحیح

ب ـ غلط

۳. در الگوی ارزشیابی هدف آزاد بازده‌های پیش‌بینی‌نشدهٔ برنامهٔ آموزشی مورد داوری قرار می‌گیرند.

الف ـ صحیح

ب ـ غلط

۴. کدام‌یک از موارد زیر آزمون ملاکی است؟

الف ـ کنکور سراسری دانشگاههای ایران

ب ـ امتحان نهایی کلاس آخر دبستان

پ ـ پرسشنامهٔ نگرش‌سنج

ت ـ آزمونهای استعداد فنی

۵. در کدام‌یک از ارزشیابیهای زیر می‌توان از ملاک نسبی استفاده کرد؟

الف ـ تراکمی

ب ـ تشخیصی

پ ـ تکوینی

ت ـ جایابی

ث ـ الف و ب و ت

۶. ارزشیابی از خود یا خودارزشیابی معادل کدام اصطلاح زیر است؟

الف ـ ارزشیابی درونی

ب ـ ارزشیابی ملاکی

پ ـ ارزشیابی هنجاری

ت ـ ارزشیابی بیرونی

ث ـ ارزشیابی از فرایند

۷. نام دیگر رویکرد طبیعت‌گرایانهٔ ارزشیابی آموزشی چیست؟

الف ـ رویکرد مبتنی بر مصرف‌کننده

ب ـ رویکرد مبتنی بر مشارکت‌کنندگان

پ ـ رویکرد مبتنی بر مدیریت

ت ـ رویکرد مبتنی بر هدف

ث ـ رویکرد مبتنی بر نظر متخصصان

۸. در صورتی یک رویکرد ارزشیابی را با توجه به تأثیر آن بر اکثریت افراد مورد قضاوت قرار می‌دهند که آن رویکرد مبتنی بر ـــــــ باشد.

الف ـ مدیریت
ب ـ مشارکت‌کنندگان
پ ـ مصرف‌کننده
ت ـ نظر متخصصان
ث ـ هدف

۹. همهٔ موارد زیر را می‌توان جزو هدفهای ارزشیابی تراکمی دانست بجز:
الف ـ تشخیص مشکلات یادگیری دانش‌آموزان و کمک به رفع آنها
ب ـ تعیین مقدار یادگیری دانش‌آموزان در طول دورهٔ آموزشی
پ ـ قضاوت دربارهٔ اثربخشی کار معلمان و برنامه‌های درسی
ت ـ مقایسهٔ توفیق برنامه‌های مختلف آموزشی با یکدیگر

۱۰. به طور خلاصه ارزشیابیهای زیر را تعریف کنید و شباهتها و تفاوتهای بین آنها را توضیح دهید:
الف ـ ارزشیابی از دانشجویان و دانش‌آموزان
ب ـ ارزشیابی از برنامه‌های درسی و مواد آموزشی
پ ـ ارزشیابی از آموزشگاه
ت ـ ارزشیابی از کارکنان آموزشگاه

۱۱. الگوهای مختلف ارزشیابی را که در این فصل خواندید به اختصار توضیح دهید و شباهتها و تفاوتهای بین آنها را مشخص نمایید.

۱۲. انواع ارزشیابیهای مورد استفاده در پیش از آموزش، ضمن آموزش، و پس از آموزش را از جهات مختلف با هم مقایسه کنید.

۱۳. ارزشیابی درونی و ارزشیابی بیرونی را توضیح دهید و رابطهٔ آنها را با ارزشیابی تکوینی و ارزشیابی تراکمی بیان کنید.

۱۴. دسته‌بندی ارزشیابیهای آموزشی را در قالب رویکردهای مختلف توضیح دهید.

۱۵. شش رویکرد ارزشیابی ورتن و سندرز را در رابطه با دو بُعد ارزشیابی هاوس توضیح دهید.

۱۶. الگوهای ارزشیابی زیر را توضیح دهید:
الف ـ الگوی ارزشیابی تایلر
ب ـ الگوی ارزشیابی سیپ
پ ـ الگوی ارزشیابی یو سی ال ای
ت ـ الگوی ارزشیابی اسکریون
ث ـ الگوی ارزشیابی هدف آزاد
ج ـ الگوی ارزشیابی سیمایی

۵

مراحل و فعالیتهای ارزشیابی آموزشی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. مراحل مختلف ارزشیابی آموزشی را به‌ترتیب نام ببرد.
۲. مرحلهٔ طرح‌ریزی را تعریف کند و اقدامات مربوط به این مرحله را توضیح دهد.
۳. منظور از تحلیل موقعیت را توضیح دهد.
۴. غایت و هدف را تعریف کند و با ذکر مثال تفاوت بین آنها را توضیح دهد.
۵. هدفهای فرآیندی و فرآورده‌ای را تعریف کند و با ذکر مثال تفاوت بین آنها را توضیح دهد.
۶. توضیح دهد که در کدام ارزشیابیها بیشتر از کدام نوع هدفهای فرآیندی یا فرآورده‌ای استفاده می‌شود.
۷. رفتار ورودی را تعریف کند و رابطهٔ آن را با هدف برنامه توضیح دهد.
۸. نحوهٔ انتخاب یا تهیهٔ وسایل اندازه‌گیری را توضیح دهد.
۹. استراتژیها را توضیح دهد و برای هریک از آنها مثالی ذکر نماید.
۱۰. طرح پژوهشی را توضیح دهد و رابطهٔ پژوهش آزمایشی با ارزشیابی آموزشی را شرح دهد.
۱۱. منظور از تدارک یک برنامهٔ زمانی را در ارزشیابی توضیح دهد.
۱۲. مرحلهٔ فرآیندی یا اجرایی را تعریف کند و اقدامات مربوط به این مرحله را توضیح دهد.
۱۳. مرحلهٔ فرآورده‌ای یا تولیدی را تعریف کند و اقدامات مربوط به این مرحله را توضیح دهد.
۱۴. برای ارزشیابی پیشرفت تحصیلی یکی از سطوح آموزشی یک برنامهٔ کامل ارزشیابی تدارک ببیند و در آن کلیهٔ پیشنهادهای این فصل را منظور نماید.

ما در فصل پیش انواع ارزشیابیهای آموزشی را معرفی کردیم. در این فصل مراحل و فعالیتهای ارزشیابی آموزشی را مورد بحث قرار خواهیم داد. اما پیش از آن لازم است این نکته را مورد تأکید قرار دهیم که صرف‌نظر از اینکه چه چیزی مورد ارزشیابی قرار می‌گیرد و کدام الگوی ارزشیابی به کار بسته می‌شود فرایند ارزشیابی همواره یکسان است. زیرا در همهٔ انواع ارزشیابیهای آموزشی فعالیتهای زیربنای کار ارزشیابی بی‌تغییر باقی می‌مانند. به قول گی (۱۹۹۱)،

> گرچه ما از انواع مختلف ارزشیابیها بحث می‌کنیم، این بدان معنا نیست که فرایندهای مختلف ارزشیابی وجود دارند. صرف‌نظر از اینکه چه چیزی مورد ارزشیابی قرار می‌گیرد، فرایند اساسی ارزشیابی یکی است. در انواع مختلف ارزشیابیها چیزی که تغییر می‌کند موضوع مورد ارزشیابی، نحوهٔ کاربست فرایند ارزشیابی، و نوع تصمیماتی که گرفته می‌شوند هستند. بسته به اینکه چه چیزی مورد ارزشیابی قرار بگیرد، اطلاعات مختلفی جمع‌آوری می‌شوند، ملاکهای مختلفی به کار می‌روند، و تصمیمات مختلفی اتخاذ می‌گردند. اما فرایند اساسی ارزشیابی همیشه یکسان است و مفاهیم و اصول کلی اندازه‌گیری و ارزشیابی یکسانی به کار می‌روند. (ص ۱۴)

بنا به اظهارات گی (۱۹۹۱)، گفتگو در بارهٔ ارزشیابیهای مختلف، به اعتبار موضوعهای مختلف ارزشیابی یا الگوهای مختلف ارزشیابی، کاری نارواست.

> بسیاری کسان بر این باورند که مثلاً ارزشیابی از برنامه یا پروژهٔ آموزشی و ارزشیابی از آموزش (تدریس) فرایندهای کاملاً متفاوتی را شامل می‌شوند. زیادی تعداد و "تنوع" الگوهای ارزشیابی این فرض را تقویت کرده است. [برخلاف این فرض] همهٔ نظامها و الگوهای ارزشیابی شامل اجزای اساسی یکسانی هستند؛ یعنی تعیین و تشخیص غایتها و هدفها؛ انتخاب و یا تولید ابزارهای اندازه‌گیری؛ توصیف و اجرای راهبردها برای کسب اطلاعات؛ و تحلیل و تفسیر نتایج. (ص X)

در نتیجه، چه ما به ارزشیابی از یادگیری دانش‌آموزان و دانشجویان اقدام کنیم، چه ارزشیابی از یک دورهٔ آموزشی، چه ارزشیابی از یک پروژهٔ علمی و نظایر اینها، اصول و مراحل اساسی کار هیچ تغییری نمی‌کنند. لذا، چنانکه در فصل پیش اشاره کردیم، نباید به این فکر کنیم که کدام الگوی ارزشیابی را در کجا به کار بندیم، زیرا در آن صورت به تعداد چیزهای مورد ارزشیابی الگوی ارزشیابی خواهیم داشت. در واقع الگوگرایی در حوزهٔ ارزشیابی آموزشی به درجه‌ای از وخامت رسیده که گپهارت[1] در سال ۱۹۷۶ (به نقل از گی، ۱۹۹۱) در این باره گفته است، "ما به‌نقطه‌ای از بیهودگی رسیده‌ایم‌که دریک گفتگوی تازه ۲۷ الگوی 'مختلف' ارزشیابی نام برده شدند" (ص ۱۴). و حتی از این هم بدتر، ورتن و سندرز در سال ۱۹۸۷ گفته‌اند که "در طول دو دههٔ گذشته متجاوز از ۵۰ الگوی مختلف ارزشیابی ایجاد و توزیع شده‌اند" (ص ۴۳).

با توجه به توضیحات بالا، ما در این کتاب، به پیروی از گی (۱۹۹۱) و صرف نظر از الگوهای مختلف ارزشیابی و موضوعهای مختلف مورد ارزشیابی، فرایند ارزشیابی آموزشی را با توجه به

1. Gephart

مراحل اصلی آن توضیح می‌دهیم، و بر این باوریم که مراحل معرفی‌شده در اینجا برای ارزشیابی از تمامی موضوعهای آموزش و پرورش قابل کاربرد هستند.

مراحل ارزشیابی آموزشی

گفتیم که ارزشیابی آموزشی به تصمیم‌گیری برای منظورهای مختلف کمک می‌کند. این تصمیم‌گیریها هم در رابطه با هدفها هستند، هم راهبردها، هم روشها و ابزارهای اندازه‌گیری، و هم موارد دیگری نظیر اینها. گی (۱۹۹۱) این گونه تصمیم‌گیریهای مختلف را در سه دستهٔ زیر قرار داده است:

۱. مرحلهٔ طراحی
۲. مرحلهٔ فرآیندی
۳. مرحلهٔ فرآورده‌ای

مرحلهٔ طراحی با سؤالهایی از این نوع سر و کار دارد: "چه کاری انجام خواهیم داد؟" (زمان آینده). مرحلهٔ فرآیندی با این گونه سؤال سر و کار دارد: "چگونه آن را انجام می‌دهیم؟" (زمان حال). و مرحلهٔ فرآورده‌ای با این نوع سؤال سر و کار دارد: "چگونه آن را انجام دادیم؟" (زمان گذشته). در دنبالهٔ این فصل این مراحل را بیشتر توضیح می‌دهیم.

مرحلهٔ طراحی

چنان‌که گفتیم مرحلهٔ طراحی[1] شامل تصمیمهایی است در بارهٔ فعالیتهایی که قرار است انجام گیرند. به طور ویژه، در این مرحله اقدامات زیر صورت می‌پذیرند: تحلیل موقعیت، تعیین و توصیف هدفها، توصیف پیش‌نیازها، انتخاب یا تولید وسایل و ابزارهای مورد نیاز، توصیف استراتژیها یا راهبردها، انتخاب طرح یا طرحهای پژوهشی (در صورت نیاز)، و تدارک یک برنامهٔ زمان‌بندی. در زیر هر یک از این اقدامات را توضیح می‌دهیم.

تحلیل موقعیت

در نخستین گام مرحلهٔ طرح‌ریزی ارزشیابی آموزشی، موقعیت موجود را تحلیل می‌کنیم. از جمله کارهایی که برای این منظور باید انجام دهیم جمع‌آوری سوابق اطلاعاتی و برآورد و تعیین موانع موجود است. مثلاً اگر قرار است ارزشیابی از یک برنامهٔ آموزشی به عمل آید، مرور و بازنگری پژوهشهای مربوط به آن ضروری به نظر می‌رسد. از جمله فواید این گونه بازنگریِ سوابق پژوهشی این است که نشان خواهد داد کدام رویکردهای ارزشیابی مفید و کدام غیرمفید بوده‌اند. همچنین بازنگری پژوهشهای وابسته می‌تواند متخصص ارزشیابی را در شناسایی ابزارهای مورد

1. planning phase

نیاز جمع‌آوری و تحلیل داده‌های ضروری یاری دهد.

در گام نخست مرحلهٔ طرح‌ریزی، چنان که گفته شد، تشخیص موانع و محدودیتها و یک برآورد واقع‌بینانه از امکانات برای تعیین هدفهای ارزشیابی ضروری است. برای مثال، متخصص ارزشیابی در صورتی می‌تواند استفاده از ضبط ویدئویی را در کارگاه آموزشی در طرح ارزشیابی خود منظور نماید که مطمئن باشد فراهم آوردن چنین دستگاهی در حوزهٔ امکانات طرح او قرار دارد.

در ارزشیابی پیشرفت تحصیلی، به منظور تحلیل موقعیت موجود، معلم باید از آموخته‌های قبلی دانش‌آموزان یا دانشجویان خود اطلاعات کافی کسب نماید. همچنین او باید امکانات موجود را به دقت برآورد کند تا اینکه ابزارها و وسایل مناسب و در دسترس را به بهترین صورت مورد استفاده قرار دهد.

تعیین و توصیف هدفها

هدفهای آموزشی معمولاً در دو سطح بیان می‌شوند: (۱) کلی، (۲) جزئی و دقیق. هدفهای کلی را معمولاً **غایت**[1] و هدفهای جزئی را **هدف**[2] می‌نامند. غایتها یا هدفهای کلی، به سبب کلی بودن، مبهم‌اند و به سادگی قابل اندازه‌گیری نیستند. در مقابل، هدفهای جزئی دقیق، قابل تبدیل به کمیت، و اندازه‌پذیرند. غالباً ضرورت دارد که غایتها تحلیل شوند و به صورت هدفهای دقیق درآیند.

برای روشن شدن تفاوت بین غایت و هدف به نمونه‌های زیر توجه کنید:

غایت:

ارتقاء سطح معلومات ریاضی دانش‌آموزان

هدفهای دقیق:

دانش‌آموز بتواند:

۱. اعداد دو رقمی را با هم جمع کند.

۲. اعداد دو رقمی را از هم کم کند.

۳. اعداد دو رقمی را در هم ضرب کند.

۴. اعداد دو رقمی را بر هم تقسیم کند.

هدفها نیز به دو دستهٔ کلی (۱) **هدفهای فرایندی** و (۲) **هدفهای فراورده‌ای** یا **تولیدی** تقسیم می‌شوند. هدفهای فرایندی به کوششها و اقداماتی اشاره می‌کنند که در ضمن اجرای برنامه انجام خواهند گرفت. در مقابل، هدفهای فراورده‌ای به بازده‌ها و نتایج کوششها و اقدامات اشاره می‌کنند.

در ارزشیابی از برنامه‌ها و پروژه‌های آموزشی هر دو نوع هدف آموزشی بالا به کار می‌روند، اما در ارزشیابی از پیشرفت تحصیلی یادگیرندگان غالباً هدفهای فراورده‌ای مورد استفاده قرار می‌گیرند. چنین فرض می‌شود که فرایندهای معینی به فراورده‌های خاصی می‌انجامند. بنابراین،

1. goal
2. objective

در ضمن ارزشیابی از یک برنامه یا پروژهٔ آموزشی هم باید فراورده‌های حاصل را ارزشیابی کرد و هم فرایندهای مورد استفاده را.

چنان که گفتیم، در ارزشیابی از پیشرفت تحصیلی یادگیرندگان غالباً هدفهای فراورده‌ای مورد استفاده قرار می‌گیرند. با این حال، در اینجا نیز می‌توان، علاوه بر هدفهای فراورده‌ای، از هدفهای فرایندی نیز سود برد.

هدفهای فراورده‌ای عمدتاً بازده‌های مورد انتظار را وصف می‌کنند و معمولاً به صورت تغییرات مطلوبی که در رفتار یادگیرندگان (دانش‌آموزان و دانشجویان) رخ خواهند کرد بیان می‌شوند. در مقابل، هدفهای فرایندی کمتر با رفتار یادگیرنده و بیشتر با راهبردها و فعالیتهایی که برای ایجاد هدفهای فراورده‌ای تدارک دیده می‌شوند و به اجرا درمی‌آیند سر و کار دارند. بنابراین، چنین فرض می‌شود که تحقق یافتن هدفهای فرایندی یعنی تحقق یافتن هدفهای فراورده‌ای.

گی (۱۹۹۱) مثالهای زیر را برای دو نوع هدف بالا ذکر کرده است. یک هدف فراورده‌ای ممکن است چنین بیان شود: "حضور یادگیرندگان در مدرسه به میزان ۲۰ درصد افزایش خواهد یافت." یک هدف فرایندی در رابطه با هدف فراورده‌ای بالا می‌تواند چنین بیان شود: "معلم در طول سال تحصیلی دست‌کم دو بار به خانهٔ هر یک از یادگیرندگان سرکشی خواهد کرد." در بیان این هدف چنین فرض می‌شود که تماس نزدیک معلم با خانه و خانوادهٔ یادگیرنده به حضور بیشتر یادگیرنده در مدرسه خواهد انجامید.

بنابر آنچه گفته شد، هدفها به فعالیتهای بعدی ارزشیابی جهت می‌دهند و نهایتاً در ارزشیابی میزان تحقق این هدفها سنجش می‌شود. پس، چون هدفها اساس فعالیتهای بعدی ارزشیابی را تشکیل می‌دهند، باید خود آنها بر حسب ربط داشتن، اندازه‌پذیر بودن، مهم بودن، و سایر ویژگیهای مطلوب هدفهای آموزشی ارزشیابی شوند.

توصیف پیش‌نیازها یا رفتارهای ورودی

تحقق هدفهای آموزشی به وجود **رفتارهای ورودی**[1] وابسته است، زیرا رفتار ورودی به آمادگی فرد برای یادگیری هدفهای آموزشی اشاره می‌کند. رفتارهای ورودی همان **پیش‌نیازها**[2] هستند که یادگیرنده پیش از شروع به یادگیری هدفهای تازه باید آنها را کسب کرده باشد. آموزش و ارزشیابی درست مستلزم توصیف و اندازه‌گیری این پیش‌نیازهاست.

برای تهیه و توصیف رفتارهای ورودی، معلم باید ابتدا به تهیهٔ فهرست هدفهای آموزشی اقدام کند و بعد معلوم نماید که یادگیری هر یک از آن هدفها چه پیش‌نیازهایی لازم دارد. معمولاً انجام چنین کاری به روش **تحلیل تکلیف**[3] نیاز دارد (نگاه کنید به سیف، ۱۳۸۰، فصل ششم).

به عنوان نمونه‌ای ساده از تحلیل تکلیف، اگر معلمی بخواهد تقسیم را به دانش‌آموزان خود

1. entry behaviors 2. prerequisites 3. task analysis

آموزش دهد، باید تحلیل کند که عمل تقسیم بر چه اعمالی استوار است. از آنجا که عمل تقسیم به اعمال ضرب، جمع، و تفریق وابسته است، توانایی انجام این اعمال به صورت پیش‌نیاز برای رفتارهای ورودی برای هدفِ *انجام دادن* تقسیم در خواهد آمد.

بنا بر آنچه گفته شد، آموزش و ارزشیابی نظامدار مستلزم تهیه، تدوین، و اندازه‌گیری پیش‌نیازها یا رفتارهای ورودی هستند. معمولاً سنجش رفتارهای ورودی در آغاز سال تحصیلی یا در ابتدای آموزش یک درس تازه انجام می‌گیرد.

انتخاب و تولید وسایل اندازه‌گیری

غالباً برای جمع‌آوری اطلاعات ضروری در ارزشیابی آموزشی وسایل یا ابزارهایی مورد نیازند. از جمله می‌توان آزمونهای پیشرفت تحصیلی، آزمونهای نگرش‌سنج، یا فرمهای ثبت مشاهده‌ای را نام برد. در صورت موجود بودن وسایل، باید به بررسی و انتخاب مناسب‌ترین آنها اقدام کرد. اگر وسایل مورد نیاز موجود نباشند، آنگاه باید آنها را تهیه کرد.

تهیهٔ وسایل اندازه‌گیری و سنجش، نظیر آزمونها و پرسشنامه‌ها، کاری است که به تخصص، دقت، و صلاحیت نیاز دارد. بنابراین، این مسئولیت را باید به متخصصان آن واگذار کرد. هر چند غالب معلمان، چه آنانی که آموزش اصول تهیه و استفاده از آزمونهای مختلف را دیده‌اند و چه آنانی که از چنین فرصتی بی‌بهره بوده‌اند، به سهولتْ اقدام به تهیهٔ آزمونهای پیشرفت تحصیلی می‌کنند، چنین کاری اگر صرفاً برای استفادهٔ کلاسی خود معلمان باشد چندان مشکل‌آفرین نیست، اما برای استفاده در برنامه‌های جدی ارزشیابی که بر اساس یافته‌های آنها تصمیمات مهم و سرنوشت‌سازی گرفته خواهند شد، تهیهٔ وسایل اندازه‌گیری ویژهٔ سنجش فرآیندها و فرآورده‌های برنامهٔ مورد ارزشیابی باید به عهدهٔ متخصصان گذاشته شود.

توصیف راهبردها یا استراتژیها

منظور از **استراتژی** یا **راهبرد**[1] فعالیت یا تدبیری است که به منظور کمک به تحقق‌یافتن هدفهای برنامه به‌کار می‌رود؛ هم می‌توان از استراتژیهای آموزشی نام برد و هم از استراتژیهای برنامهٔ درسی و غیره.

گی (۱۹۹۱) برای توصیف مفهوم استراتژی مثال آموزشی زیر را ذکر کرده است. فرض کنید می‌خواهیم موضوع چندجمله‌ای را در ریاضی به دانش‌آموزان آموزش بدهیم. هدفهای آموزشی ما عبارت‌اند از: (۱) تشخیص مفهوم یک‌جمله‌ای (مثلاً $2x^2$، y^3، یا $7z$)، دوجمله‌ای (مثلاً $4x + 6y^2$)، و چندجمله‌ای (بیشتر از دوجمله‌ای)، (۲) تعیین درجهٔ آنها، و (۳) جمع کردن آنها.

نخستین تصمیم استراتژیکی این است که تعیین کنیم توالی ارائهٔ این موضوعها چگونه باشد.

---

1. strategy

این تصمیم، سوای آنچه قرار است آموزش داده شود، همواره نخستین تصمیمی است که باید اتخاذ گردد. مسلماً ترتیب ارائهٔ موضوعها نباید به صورت تصادفی باشد. در مثال حاضر، یک استراتژی معقول این است که ابتدا اطلاعات ضروری را در بارهٔ یک‌جمله‌ایها آموزش بدهیم؛ یعنی اول مفهوم یک‌جمله‌ای را آموزش بدهیم، بعد درجهٔ یک‌جمله‌ای، و بالاخره جمع یک‌جمله‌ایها را. پس از آن، می‌توانیم همهٔ این اطلاعات را در بارهٔ دوجمله‌ایها آموزش بدهیم. و آخرسر می‌توانیم این اطلاعات را در بارهٔ چندجمله‌ایها نیز آموزش بدهیم.

استراتژی معقول دوم این است که ابتدا مفاهیم را آموزش بدهیم (یعنی یک‌جمله‌ای چیست، دوجمله‌ای چیست، و سه‌جمله‌ای چیست). بعد نحوهٔ تعیین درجهٔ هر یک از آنها را آموزش بدهیم. و سرانجام اینکه چگونگی جمع هر یک از آنها را آموزش بدهیم.

برای اینکه تعیین کنیم هدفهای آموزشی یک درس به چه ترتیبی آموزش داده شوند، لازم است روش تحلیل تکلیف را به کار بریم. در صفحات قبلی توضیح دادیم که روش تحلیل تکلیف برای تهیهٔ رفتارهای ورودی یا پیش‌نیاز ضروری است. این روش همچنین برای استخراج هدفهای دقیق از غایتهای کلی مفید است و علاوه بر آن سلسله مراتب هدفهای آموزشی یک درس را نشان می‌دهد و به معلم کمک می‌کند تا تشخیص دهد ترتیب آموزش هدفهای مختلف چگونه باشد.

نمونهٔ دیگری از استراتژی آموزشی **مرور یا بازنگری**[1] مطالب است. بسیاری از پژوهشهای انجام‌شده در این باره نشان داده‌اند که مرور مطالب درسی موجب نگهداری بیشتر آنها در حافظه می‌شوند (برای مثال نگاه کنید به آزوبل[2] و یوسف[3]، ۱۹۶۵). همچنین تمرین و بازخورد استراتژیهای مفید دیگری هستند. پاره‌ای از استراتژیهای آموزشی، نظیر بازخورد دادن، مرور کردن، و تمرین کردن برای موضوعهای درسی و هدفهای آموزشی مختلف قابل استفاده هستند، اما برخی دیگر از استراتژیها مانند دسته‌بندی یادگیرندگان و ایفای نقش تنها برای هدفهای خاصی مفیدند.

انتخاب طرح پژوهشی

چنان‌که در فصل نخست این کتاب گفتیم، ارزشیابی آموزشی از لحاظ شیوهٔ کار و استفاده از ابزارها و عینیت می‌تواند خیلی زیاد شبیه به پژوهش آموزشی باشد، اما ارزشیابی آموزشی الزاماً پژوهش آموزشی نیست. با این حال، ممکن است متخصص ارزشیابی بخواهد یا نیاز داشته باشد همزمان با انجام یک پروژهٔ ارزشیابی به پژوهش نیز اقدام کند. مثلاً در ارزشیابی از یک برنامهٔ درسی جدید ممکن است ارزشیاب بخواهد جواب سؤالهای زیر را پیدا کند:

۱. آیا دانش‌آموزان به هدفهای تعیین‌شده رسیده‌اند؟
۲. آیا دانش‌آموزانی که از برنامهٔ فعلی استفاده کرده‌اند از دانش‌آموزانی که برنامهٔ دیگری را مورد استفاده قرار دادند بیشتر آموخته‌اند؟

1. review 2. Ausubel 3. Youssef

پاسخ دادن به سؤال اول یک فعالیت ارزشیابی است، اما پاسخ دادن به سؤال دوم یک فعالیت پژوهشی است و نیاز به کاربرد یک طرح پژوهشی دارد، زیرا پیدا کردن پاسخ این سؤال مستلزم استفاده از گروههای مختلف دانش‌آموزان و مقایسهٔ بین آنهاست که این یک کار پژوهشی است. به طور کلی،

اساس آزمایشگری[1] این است که اگر شما با دو گروه یا بیشتر که در متغیرهای مورد نظر (مانند توانایی) معادل هستند شروع کنید و برخورد یکسانی با آنها داشته باشید و برای تمامی عواملی که ممکن است بر یکی از دو گروه تأثیر بگذارد کنترل اعمال نمایید ــ بجز اینکه هر یک از گروهها روش یا برنامهٔ متفاوتی را دریافت کند ــ و آن‌گاه اگر پس از گذشت زمانی معقول گروهها عملکردهای متفاوت داشته باشند، تفاوت حاصل احتمالاً ناشی از اثربخشی متفاوت روشهای اِعمال‌شده خواهد بود. (گی، ۱۹۹۱، ص ۳۴)

طرحهای پژوهشی بسیار متفاوت‌اند و غالباً به دو گونهٔ **طرحهای آزمایشی**[2] و **طرحهای شبه‌آزمایشی**[3] تقسیم می‌شوند. برای کسب اطلاعات لازم در این باره به کتابهای روش پژوهش، از جمله به کتاب *طرحهای پژوهش آزمایشی و شبه‌آزمایشی*، اثر کمپبل[4] و استنلی[5] (۱۹۶۳)، مراجعه کنید.

تدارک یک برنامهٔ زمان‌بندی

تهیه و تدارک یک برنامهٔ زمان‌بندی که برآورد واقع‌بینانه‌ای از جریان پیشرفت کار ارائه دهد برای هر فعالیت جدی آدمی، از جمله ارزشیابی آموزشی، امری واجب است. اساساً یک برنامهٔ زمان‌بندی دربرگیرندهٔ فهرستی از فعالیتهای اصلی مربوط به اقدامات پیشنهادی و زمان شروع و خاتمهٔ هر یک از آن فعالیتهاست.

به‌هنگام تدارک برنامهٔ زمان‌بندی، زمان‌پیش‌بینی‌شده برای هر مرحله یا هر اقدام مربوط به ارزشیابی باید به‌قدر کافی باشد، به گونه‌ای که بدون نیاز به عجلهٔ غیرضروری و در فرصت کافی هر کاری به موقع و با حوصله تکمیل گردد. دست‌اندرکاران ارزشیابی آموزشی، بر اساس تجارب شخصی خود، یافته‌اند که ملحوظ کردن مقداری زمان اضافی در خلال برنامهٔ زمان‌بندی گاه بسیار مفید است، زیرا خیلی از اوقات رویدادهای غیرمنتظره‌ای رخ می‌نمایند که نیاز به رسیدگی و لذا صرف وقت دارند. به همین منظور، بهتر است در تهیهٔ برنامهٔ زمان‌بندی تاریخ پایان گرفتن آخرین اقدام یا مرحلهٔ ارزشیابی را قدری جلوتر از موعدی که رسماً تعیین شده، یعنی مثلاً در زمانی که طرح ارزشیابی پیش‌بینی شده است، در نظر گرفت.

گی (۱۹۹۱) برای تدارک جدول زمانی **روش چارتی گانت**[6] را پیشنهاد داده است. در این روش فعالیتهایی که قرار است انجام گیرند در سمت راست کاغذ فهرست‌وار به دنبال یکدیگر نوشته می‌شوند و زمان انجام تمامی پروژه در بالای صفحه قرار داده می‌شود. در این روش یک نمودار ستونی نیز مورد استفاده قرار می‌گیرد و در آن تاریخ آغاز و انجام هر فعالیت معین نشان داده

1. experimentation
2. experimental designs
3. quasi – experimental designs
4. Campbel
5. Stanley
6. Gantt chart method

می‌شود. چنین چارتی به متخصص یا گروه متخصصان ارزشیابی یا ارزشیابان اجاره می‌دهد تا به سادگی هم تصویر کلی کاری را که در پیش رو دارند ببینند و هم فعالیتهای مختلفی را که قرار است انجام دهند به طور انفرادی ملاحظه کنند.

مرحلهٔ فرآیندی یا اجرایی

دومین مرحلهٔ ارزشیابی آموزشی **مرحلهٔ فرآیندی**[1] است. در این مرحله تصمیمهای مربوط به اقدامات عملی یا اجرایی گرفته می‌شوند. از جملهٔ نخستین اقدامات در این مرحله عبارت‌اند از اجرای پیش آزمون و در صورتی که ارزشیابی پیشرفت تحصیلی در کار باشد اجرای آزمون رفتار ورودی. بر اساس نتایج حاصل از این آزمونها معلوم می‌شود هدفهایی که در مرحلهٔ قبل تعیین شده‌اند مناسب بوده‌اند یا نیاز به تجدید نظر دارند.

پس از اجرای پیش آزمون و یا آزمون رفتار ورودی، استراتژیها و فعالیتهای طراحی‌شده به سبکی که از قبل تعیین شده است به اجرا درمی‌آیند. اطلاعات جمع‌آوری‌شده در ضمن این مرحله بازخورد لازم را در این باره به دست می‌دهند که مراحل اجرایی به نحوی که از پیش تعیین شده به اجرا درمی‌آیند یا در آن اشکالاتی موجود است.

هدفهای اساسی این مرحله از ارزشیابی این هستند که معلوم شود آیا کوششها آن گونه که منظور بوده است به کار گرفته شده‌اند تا میزان توفیق هدفهای فرایندی را تعیین کند و نیز راههایی را که در آن بهبود امکان‌پذیر است تعیین نماید. اگر همزمان چند استراتژی یا راهبرد با هم به کار می‌روند، باید تعیین شود که کدام مؤثرتر است و کدام اثربخشی لازم را ندارد. برای مثال، "معلم ممکن است نتیجه گیری کند که وقت کمتری باید صرف مثالهایی که برای دانش‌آموزان بر روی تابلو حل می‌کند نماید و در عوض به حل مسئله توسط گروههای کوچک دانش‌آموزان زمان بیشتری اختصاص دهد، زیرا این فعالیتی است که دانش‌آموزان بیشتر دوست دارند و از آن بهرهٔ بیشتری می‌برند" (گی، ۱۹۹۱، ص ۳۵).

هر چند که باید کوشش شود تا تمام اقدامات طبق برنامه‌های پیش‌بینی‌شده به اجرا درآیند، ایجاد تغییرات جزئی درحین اجرای برنامهٔ ارزشیابی ممکن‌است لازم آید. هدف اصلی ارزشیابی در مرحلهٔ فرایندی هدایت فعالیتهای آتی برای بهبود فعالیتهای آموزش و یادگیری است.

در این مرحله تأکید بر پیشرفت فردی دانش‌آموزان یا پیشرفت گروهی آنان به نوع ارزشیابی وابسته است. در ارزشیابی از دانش‌آموزان و دانشجویان لازم است برای هر یک از یادگیرندگان بازخورد و تقویت تدارک دیده شود، در حالی که در ارزشیابی از برنامهٔ درسی و مواد آموزشی و پروژه‌ها تأکید باید بر بهبود استراتژیهایی باشد که به کل گروه دانش‌آموزان یا دانشجویان مربوط می‌شود.

گی (۱۹۹۱) گفته است که مرحلهٔ فرایندی ارزشیابی همان چیزی است که ***ارزشیابی تکوینی***

1. process phase

نامیده می‌شود و ما در فصل پیش آن را توضیح دادیم. پوفام (۱۹۷۵) هدف اصلی ارزشیابی تکوینی را کمک به اصلاح برنامه‌های آموزشی می‌داند و برای آن مثالهای زیر را به دست می‌دهد: (۱) صورت اولیهٔ مجموعه‌ای از دفترچه‌های خودآموز، یا (۲) یک برنامهٔ تازه تولیدشدهٔ آموزش از راه دور که هنوز در حال آزمایش و اصلاح است. "متخصصان ارزشیابی می‌کوشند که این گونه برنامه‌ها را ارزشیابی نمایند تا تولیدکنندگان برنامه را در اصلاح نواقص آموزش خود یاری دهند" (ص ۱۴).

۲) مرحلهٔ فرآورده‌ای

در مرحلهٔ فرآورده‌ای[1] یا تولیدی، برخلاف مرحلهٔ فرایندی، تصمیمات در پایان برنامه یا پروژه یا در پایان دوره‌ای یا مقطعی از آن، یعنی زمانی که فعالیتها کامل می‌شوند، اتخاذ می‌گردند. نام دیگر این مرحلهٔ ارزشیابی، *ارزشیابی تراکمی* است که در فصل پیش آن را توضیح دادیم. در این مرحله از ارزشیابی نتایج پس‌آزمونهای[2] پیشرفت تحصیلی، نگرش، و نظایر آنها مورد استفاده و تحلیل قرار می‌گیرند. هدف اصلی این مرحله جمع‌آوری داده‌هایی است که متخصص ارزشیابی را در تصمیم‌گیریهایش در بارهٔ اثربخشی کلی روش آموزشی، برنامهٔ درسی، پروژهٔ آموزشی، و غیره کمک نماید. در این مرحله تعیین می‌شود که هدفهای از پیش تعیین‌شده تا چه میزانی تحقق یافته‌اند. علاوه بر هدفهای از پیش تعیین‌شده، نتایج غیرقابل پیش‌بینی نیز تحلیل می‌شوند. برای مثال، "ممکن است معلوم شود برنامه‌ای که برای بهبود مفهوم خود[3] اجرا شده همچنین بر افزایش میزان حضور شرکت‌کنندگان تأثیر مثبت داشته است. این بازده، هر چند از پیش منظور نشده، شایستهٔ گزارش کردن و بررسی آینده است" (گی، ۱۹۹۱، ص ۳۶). در این مرحله از ارزشیابی آموزشی، علاوه بر تحلیل و تفسیر داده‌ها، تهیهٔ گزارش نهایی که در آن هدفها، روشها، و بازده‌های فعالیتهای ارزشیابی منظور شده باشد نیز لازم است. البته در ارزشیابی از پیشرفت تحصیلی دانش‌آموزان و دانشجویان تعیین نمرات و گزارش آن به خود یادگیرندگان و اولیا معمولاً کفایت می‌کند.

گی (۱۹۹۱) سه راه عمدهٔ استفاده از نتایج مرحلهٔ فراورده‌ای را به نحو زیر توضیح داده است: نخست اینکه، این نتایج برای همهٔ کسانی که ذینفع هستند بازخورد و رهنمون فراهم می‌کند؛ از جمله دانش‌آموزان و دانشجویان، معلمان، و مجریان برنامه. دوم اینکه، این نتایج برای تصمیم‌گیرندگان خارجی مانند والدین، مدیران، منابع مالی، و نظایر آنها بازخورد فراهم می‌آورند. سوم اینکه، بسته به نوع ارزشیابی، گروههای مختلفی می‌توانند از نتایج ارزشیابی استفاده کنند. برای مثال، اطلاع از عملکرد دانش‌آموزان برای راهنمایی و مشاوره بسیار مفید است؛ نتایج

---

1. product phase
2. posttests
3. self - concept

ارزشیابی از برنامهٔ درسی مورد استفادهٔ مدیران مدارس واقع می‌شود؛ و پژوهشگران آموزشی و سایر متخصصان ارزشیابی نیز می‌توانند از نتایج ارزشیابی استفاده کنند.

به طور خلاصه، در مرحلهٔ فراورده‌ای یا تولیدی ارزشیابی تصمیمات مربوط به اثربخشی موضوع مورد ارزشیابی اتخاذ می‌گردند. نتایج این مرحله از ارزشیابی برای اشخاص داخل و خارج سیستم بازخورد لازم را فراهم می‌آورند و مبنای تصمیمات مربوط به فعالیتها را در آینده تشکیل می‌دهند. پوفام (۱۹۷۵) می‌گوید "نتایج این مرحله از ارزشیابی به دیگران کمک می‌کند تا تصمیم بگیرند که آیا برنامهٔ آموزشی مورد ارزشیابی را ادامه دهند، خاتمه دهند، تغییر دهند، یا مورد تأکید قرار دهند" (ص ۳۶).

جدول ۱ـ۵ تصمیمها و فعالیتهای سه مرحلهٔ اصلی ارزشیابی را به طور خلاصه ارائه می‌دهد. برای کسب اطلاعات بیشتر دربارهٔ مراحل ارزشیابی آموزشی به کتاب ***اندازه‌گیری و ارزشیابی آموزشی***، اثر گی (۱۹۹۱) مراجعه کنید.

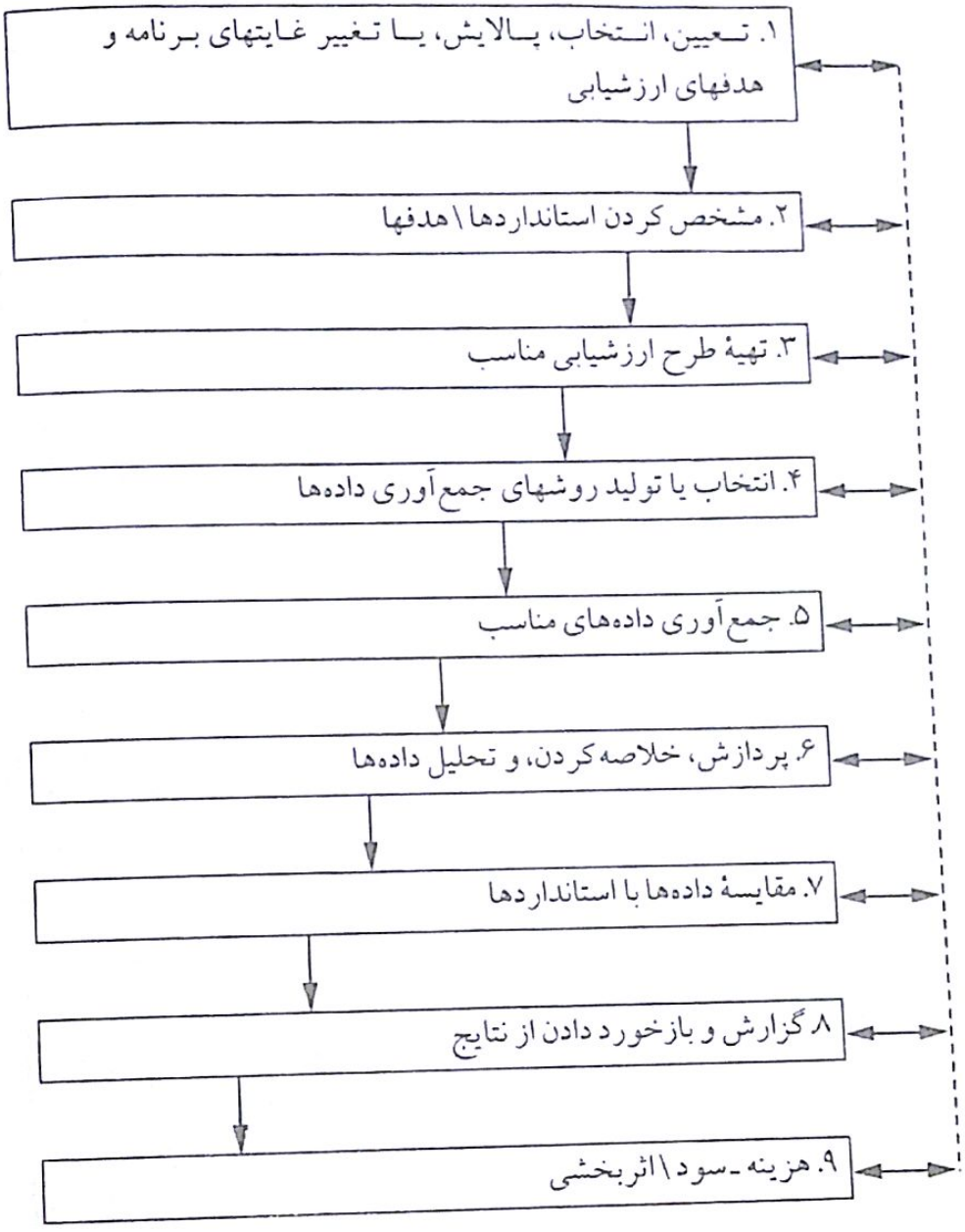
جدول ۱ـ۵ خلاصهٔ تصمیمها و فعالیتهای سه مرحلهٔ اصلی ارزشیابی (اقتباس از اثر گی، ۱۹۹۱، ص ۳۷)

مرحلهٔ ارزشیابی	زمان اجرا	نوع تصمیمها	فعالیتهای ویژه
طراحی	پیش از اجرای آموزش، برنامه، پروژه، و غیره	۱. نتایجی که قرار است به دست آیند ۲. اعمالی که باید انجام شوند	۱. تحلیل موقعیت ۲. تعیین و توصیف هدفها ۳. توصیف پیش‌نیازها ۴. انتخاب یا تهیهٔ وسایل اندازه گیری ۵. توصیف استراتژیها یا راهبردها ۶. انتخاب طرح پژوهشی ۷. تدارک برنامهٔ زمان‌بندی
فرایندی یا اجرایی	ضمن اجرای آموزش، برنامه، پروژه، و غیره	۱. به چه میزانی مراحل طرح‌ریزی شده به اجرا درمی‌آیند؟ ۲. چه تغییراتی باید داده شوند تا پیشرفت لازم به دست آید؟	۱. اجرای پیش‌آزمون ۲. اجرای آزمون رفتار ورودی ۳. سنجش مناسب بودن هدفها ۴. جمع‌آوری داده‌ها به طور دوره‌ای، مثل اجرای آزمون ۵. تحلیل اثربخشی استراتژیها
فراورده‌ای یا تولیدی	پس از اجرای آموزش، برنامه، پروژه، و غیره	۱. اثربخشی کلی آموزش، پروژه، برنامه، و غیره ۲. اقدامات آتی	۱. جمع‌آوری داده‌های مربوط به هدفها، مثلاً اجرای پس‌آزمون ۲. جمع‌آوری داده‌های مربوط به بازده‌های پیش‌بینی‌نشده ۳. تحلیل و تفسیر داده‌ها ۴. تهیهٔ گزارش

فعالیتهای ارزشیابی آموزشی

مراحل ارزشیابی آموزشی که در قسمت قبلی این فصل توضیح داده شدند برای همهٔ انواع ارزشیابیهای آموزشی پیشنهاد شده‌اند. پین (۱۹۹۴) برای ارزشیابی از پروژه‌ها و برنامه‌های آموزشی **فعالیتهای ارزشیابی آموزشی** را توضیح داده است. ما در اینجا این فعالیتها را به طور خلاصه معرفی می‌کنیم.

شکل ۱-۵ فعالیتهای ارزشیابی از برنامه و پروژهٔ آموزشی را خلاصه‌وار نشان می‌دهد. خطوط نقطه‌چین حاکی از این است که اطلاعات در بین فعالیتها می‌تواند مشترک باشد و تصمیمها مرتباً در حال تجدیدنظر و تغییرند. در ارزشیابی تراکمی همهٔ فعالیتهای شکل ۱-۵ به طور



شکل ۱-۵ خلاصهٔ فعالیتهای معمول در فرایند ارزشیابی (پین، ۱۹۹۴، ص ۱۳)

تامل و یکباره انجام می‌شوند، اما در ارزشیابی تکوینی این فعالیتها به طور دوره‌ای و تکراری صورت می‌پذیرند. بسته به نیاز ارزشیابی از برنامه یا پروژه و نیز بسته به رویکرد روش‌شناختی مورد استفادهٔ ارزشیاب، توالی فعالیتهای ۹ گانه می‌تواند تغییر کند. "ارزشیاب متکی به رویکرد سنتی مبتنی بر هدف تمام فعالیتهای شکل ۱-۵ را کمابیش به همان صورت انجام خواهد داد، در حالی که ارزشیاب هدف آزاد ممکن است فعالیتهای شمارهٔ ۱، ۲، و (یا) ۳ را نادیده بگیرد و صرفاً از راه مشاهده، پرسشنامه، مصاحبه، و غیره شروع به گردآوری داده‌ها کند" (پین، ۱۹۹۴، ص ۱۲).

چنان‌که در شکل ۱-۵ دیده می‌شود، نخستین فعالیت در ارزشیابی از برنامه یا پروژهٔ آموزشی تعیین، انتخاب، پالایش، یا تغییر غایتها یا هدفهای کلی برنامه و هدفهای دقیق ارزشیابی است. فعالیت بعدی مشخص کردن استانداردها یا معیارهای ارزشیابی است. سؤالی که ارزشیاب باید در این باره به آن جواب دهد این است که "بر چه اساسی باید به داوری ارزشی بپردازم؟" پین (۱۹۹۴) می‌گوید معیار می‌تواند به یک فرد مربوط باشد (مثلاً، آیا فلان دانش‌آموز ۷۵٪ واژگان درس امروز مربوط به گیاهان را یادگرفته است؟) یا یک گروه یا یک نهاد (آیا ۸۰٪ دانش‌آموزان کلاس پنجم روستایی ۸۰٪ پایتختهای ۵۰ کشور مهم را یاد گرفته‌اند؟).

پس از تعیین استاندارد یا معیار، فعالیتهای بعدی برای جمع‌آوری داده‌های لازم صورت می‌گیرد تا با توجه به معیار تعیین شده هدفها مورد ارزشیابی قرار گیرند. استانداردها می‌توانند پیش یا پس از جمع‌آوری داده‌ها مشخص گردند، اما تعیین و تشخیص آنها حتماً بر تصمیم‌گیری مقدم است.

پس از جمع‌آوری، تحلیل، و تفسیر داده‌ها و مقایسهٔ داده‌ها با معیارها و انجام داوری ارزشی و تهیهٔ گزارش نهایی نوبت به تحلیل هزینه‌ها می‌رسد. پین (۱۹۹۴) می‌گوید تحلیل هزینه یکی از فعالیتهای بسیار مهم ارزشیابی است که غالباً مورد غفلت قرار می‌گیرد. در تحلیل هزینه‌های مربوط به برنامه‌ها و پروژه‌های آموزشیِ اثربخش هم هزینه‌های مالی باید منظور شوند و هم هزینه‌ها برحسب منابع انسانی به حساب آیند. سؤالهای مربوط به هزینه و سود[1] (مانند اینکه آیا برنامه به فرد یا جامعه سودی رسانیده است؟) و سؤالهای مربوط به هزینه و اثربخشی[2] (مانند اینکه آیا سرمایه‌گذاری ارزش پولی را که برای آن هزینه شده است دارد؟) باید پس از انجام ارزشیابی کلی پاسخ داده شوند.

خلاصهٔ فصل

ما در این فصل مراحل مختلف ارزشیابی آموزشی را از آغاز تا به انجام به شرح زیر توضیح دادیم:

۱. مرحلهٔ طرح‌ریزی شامل اقدامات زیر:

الف ـ تحلیل موقعیت

1. cost-benefit
2. cost-effectiveness

ب ـ تعیین و توصیف هدفها
پ ـ توصیف رفتارهای ورودی
ت ـ انتخاب و تولید وسایل اندازه‌گیری
ث ـ توصیف استراتژیها یا راهبردها
ج ـ انتخاب طرح پژوهشی
چ ـ تدارک یک برنامهٔ زمان‌بندی

۲. مرحلهٔ فرآیندی یا اجرایی

۳. مرحلهٔ فرآورده‌ای یا تولیدی

مراحل سه‌گانهٔ بالا، مراحل اصلی تصمیم‌گیری در فرایند ارزشیابی هستند. مرحلهٔ طرح‌ریزی با این نوع سؤال سر و کار دارد: "چه کاری انجام *خواهیم داد*؟" مرحلهٔ فرآیندی به جواب دادن به سؤالهایی از این نوع می‌پردازد: "چگونه آن کار را انجام *می‌دهیم*؟" و مرحلهٔ فرآورده‌ای به پرسشهایی از این قبیل پاسخ می‌دهد: "چگونه آن کار را انجام *دادیم*؟"

علاوه بر مراحل ارزشیابی آموزشی، در این فصل برای ارزشیابی برنامه‌ها و پروژه‌های آموزشی ۹ مرحله به شرح زیر توضیح داده‌ایم:

۱. تعیین، انتخاب، پالایش، یا تغییر غایتهای برنامه و هدفهای ارزشیابی
۲. مشخص کردن استانداردها \ هدفها
۳. تهیهٔ طرح ارزشیابی مناسب
۴. انتخاب یا تولید روشهای جمع‌آوری داده‌ها
۵. جمع‌آوری داده‌های مناسب
۶. پردازش، خلاصه کردن، و تحلیل داده‌ها
۷. مقایسهٔ داده‌ها با استانداردها
۸. گزارش و بازخورد دادن از نتایج
۹. هزینه و سود \ اثربخشی

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. هرچند که الگوهای ارزشیابی آموزشی متنوع‌اند، اما فعالیتهای اصلی کار ارزشیابی کم و بیش یکسان‌اند.

الف ـ صحیح
ب ـ غلط

۲. متخصص ارزشیابی زمانی به انتخاب یک طرح پژوهشی اقدام می‌کند که بخواهد:
الف- برای تصمیم‌گیری مدیران اطلاعاتی را دربارهٔ میزان موفقیت کارکنان فراهم آورد.
ب- رابطهٔ بین دو یا چند متغیر مورد مطالعه را پیدا کند.
پ- تعداد دانش آموزانی را که به حد تسلط رسیده‌اند معلوم نماید.
ت- تعیین کند که آیا برنامهٔ مورد ارزشیابی موفق بوده است یا نه.

۳. مراحل مختلف ارزشیابی آموزشی را به ترتیب نام ببرید.

۴. اقدامات مهم مربوط به مرحلهٔ طراحی را توضیح دهید.

۵. غایت و هدف را تعریف کنید، برای هر یک مثالی بزنید، و تفاوتهای بین آنها را توضیح دهید.

۶. با چه روشی می‌توان غایتها را به هدفهای دقیق تبدیل کرد؟ آن روش را با ذکر مثالی توضیح دهید.

۷. هدفهای فرآیندی و فرآورده‌ای را توضیح دهید، برای هر یک مثالی بزنید و رابطهٔ آنها را با ارزشیابی آموزشی توضیح دهید.

۸. استراتژی را توضیح دهید و برای آن مثالی ذکر کنید.

۹. مرحلهٔ فرایندی یا اجرایی را تعریف کنید و اقدامات مربوط به این مرحله را توضیح دهید.

۱۰. فعالیتهای نُه گانهٔ پیشنهادی پین برای ارزشیابی از پروژه‌ها و برنامه‌های آموزشی را به اختصار توضیح دهید.

۱۱. برای ارزشیابی پیشرفت تحصیلی یکی از سطوح آموزشی یک برنامهٔ کامل ارزشیابی بنویسید و در آن کلیهٔ پیشنهادهای این فصل را منظور نمایید.

۹

تهیهٔ طرح ارزشیابی پیشرفت تحصیلی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. ارزشیابی پیشرفت تحصیلی را تعریف کند و آن را با تعریف کلی ارزشیابی آموزشی که در فصل اول آمده است مقایسه نماید.

۲. توضیح دهد که برای تهیهٔ طرح ارزشیابی پیشرفت تحصیلی چه اقداماتی ضروری هستند.

۳. هدف آموزشی را تعریف کند و فواید استفاده از آن را توضیح دهد.

۴. طبقه‌بندی سه‌گانهٔ هدفهای آموزشی را به اختصار توضیح دهد.

۵. طبقه‌بندی تجدیدنظر شدهٔ حوزهٔ شناختی را به اختصار توضیح دهید.

۶. جدول مشخصات را تعریف کند و اجزای آن را توضیح دهد.

۷. عواملی را که به هنگام توزیع سؤالها در میان بخشها و هدفهای مختلف باید در نظر گرفت توضیح دهد.

۸. انواع جدولهای مشخصات را با هم مقایسه کند.

۹. برای یکی از کتابهای یکی از دوره‌های تحصیلی یک جدول مشخصات کامل تهیه کند، به نحوی که در آن اجزای محتوایی، هدفهای آموزشی، درصد سؤالها، و تعداد واقعی سؤالها برای یک آزمون ۶۰ سؤالی را وارد نماید.

ما در فصل چهارم این کتاب در معرفی انواع ارزشیابیهای آموزشی گفتیم که یکی از ارزشیابیهای آموزشی مهم ارزشیابی از پیشرفت تحصیلی دانش‌آموزان و دانشجویان در درسهای گوناگون و در دوره‌های مختلف تحصیلی است. موضوع اصلی کتاب حاضر نیز این نوع ارزشیابی است. به همین منظور، ما از این پس در این کتاب به طور عمده چگونگی تهیه و مراحل اجرا و استفاده از ارزشیابی پیشرفت تحصیلی را بحث و بررسی خواهیم کرد.

تعریف ارزشیابی پیشرفت تحصیلی

در فصل اول، اصطلاح ارزشیابی آموزشی را به این صورت تعریف کردیم: *ارزشیابی به یک فرایند نظام‌دار برای جمع‌آوری، تحلیل، و تفسیر اطلاعات گفته می‌شود به این منظور که تعیین شود آیا هدفهای مورد نظر تحقق یافته‌اند یا در حال تحقق یافتن هستند و به چه میزانی.* ارزشیابی پیشرفت تحصیلی را می‌توان به صورت زیر تعریف کرد: *سنجش عملکرد یادگیرندگان و مقایسهٔ نتایج حاصل با هدفهای آموزشی از پیش تعیین‌شده، به منظور تصمیم‌گیری در این باره که آیا فعالیتهای آموزشی معلم و کوششهای یادگیری دانش‌آموزان یا دانشجویان به نتایج مطلوب انجامیده‌اند و به چه میزانی.* بنابراین، در ارزشیابی پیشرفت تحصیلی دو اقدام اساسی ضروری‌اند: (۱) تعیین هدفهای آموزشی، (۲) سنجش عملکرد یادگیرندگان. ما نیز در دنبالهٔ مطالب این کتاب ابتدا به بحث از هدفهای آموزشی و به‌دنبال آن به‌معرفی وسایل اندازه‌گیری پیشرفت تحصیلی اقدام خواهیم کرد.

هدفهای آموزشی یا هدفهای یادگیری

در فصل پیش گفتیم که در آموزش و پرورش دو نوع هدف وجود دارند: کلی و دقیق. هدفهای کلی غایت نام دارند. غایتها بیاناتی هستند کلی و آرمانی که قصد و منظور از ایجاد و ادارهٔ دوره‌های آموزشی یک جامعه (مانند دبستان، دبیرستان، دانشگاه، و غیره) را نشان می‌دهند. این هدفهای کلی معمولاً به وسیلهٔ برنامه‌ریزان در سطح وزارت تهیه و تدوین می‌شوند و به صورت مقاصد آرمانی دوره‌های آموزشی در اختیار معلمان و سایر مربیان قرار می‌گیرند. در مقابل غایت یا هدف کلی، نوع دیگری هدف وجود دارد که دقیق است و آن را غالباً معلمان و متخصصان آموزشی با توجه به غایتهای کلی و با استفاده از محتوای درسهای مختلف تهیه می‌کنند. همین هدفهای دقیق آموزشیِ برخاسته از غایتهای کلی هستند که هدفهای درسهای مختلف معلمان را تشکیل می‌دهند. بنابراین، در ارزشیابی آموزشی نیز عملکرد یادگیرندگان (دانش‌آموزان و دانشجویان) با توجه به این گونه هدفهای دقیق آموزشی سنجش می‌شود، و همین هدفها هستند که ملاک قضاوت توفیق معلمان و یادگیرندگان به حساب می‌آیند. به طور ساده، هدفهای آموزشی یا هدفهای یادگیری یک درس یا یک بحث *منظور و مقصودی است که معلم از آموزش خود و یادگیرنده از یادگیری خود دارند.*

هم برای آموزش و هم برای ارزشیابی، لازم است هدفهای کلی به هدفهای دقیق تجزیه شوند، به گونه‌ای که اطمینان حاصل گردد که یادگیرنده از راه یادگیری هدفهای دقیق به هدف کلی

خواهد رسید. ما در فصل پیش یک نمونه هدف کلی و هدفهای دقیق مشتق شده از آن را ارائه دادیم. در زیر نیز به نمونهٔ دیگری توجه کنید.

هدف کلی

کسب مهارتهای مورد نیاز استفاده از وسائل معمولی برای اندازه‌گیری طول، حجم، و وزن.

هدفهای دقیق

دانش‌آموز بتواند:

۱. با استفاده از خط‌کش، طول اشیاء را برحسب سانتیمتر اندازه بگیرد.
۲. به کمک ترازو، وزن اشیاء را برحسب گرم اندازه بگیرد.
۳. با استفاده از یک استوانهٔ مدرج، حجم مایعات را برحسب سانتیمتر مکعب اندازه بگیرد.

در تجزیهٔ هدفهای کلی به هدفهای جزئی، نباید زیاده‌روی کرد، زیرا هدفهای خیلی جزئی مفید نیستند. صاحب‌نظران سنجش و ارزشیابی، در رابطه با میزان کلی و جزئی بودن هدفها، گفته‌اند هدفهای آموزشی را نه خیلی کلی و نه خیلی جزیی بلکه تا آن حد کلی بنویسید که آموزش‌پذیر و سنجش‌پذیر باشند. "به طور آشکار، درجهٔ *اندازه‌پذیری* کلید شناسایی هدفهای آموزشی خوب است. هدفی که اندازه‌پذیر است هرچه گسترده‌تر باشد بهتر است" (پوفام، ۲۰۰۲، ص ۹۸).

استانداردهای تحصیلی

اصطلاح دیگری که در آموزش و پرورش کاربردی شبیه به هدف دارد **استاندارد**[1] است. استاینمنتز[2] (۱۹۸۸) می‌گوید "منظور ما از استاندارد یک فهرست، توصیف، یا بازنمایی از کیفیتها یا ویژگیهایی است که یک شیء باید داشته باشد. به سخن دیگر، توصیفی از اینکه چیزی چگونه باید باشد استاندارد نامیده می‌شود" (ص ۸۰). بنا به تعریف لین و گرانلاند (۲۰۰۰)، **استانداردهای تحصیلی**[3] "بیاناتی هستند که مشخص می‌کنند چه چیزی باید آموزش داده شود و دانش‌آموزان چه چیزی را باید یاد بگیرند" (ص ۶).

استانداردها معمولاً به شکلی کلی بیان می‌شوند؛ از این‌رو به غایتهای پرورشی یا هدفهای کلی آموزشی شبیه‌اند. به همین سبب، برای هر استاندارد یک یا چند **نشانگر**[4] یا **سنجه**[5] می‌نویسند تا آموزش‌پذیر و سنجش‌پذیر بشود. به نمونهٔ زیر توجه کنید:

استاندارد: دانش‌آموزان فرایند خواندن را به طور مؤثر مورد استفاده قرار می‌دهند.

سنجه: دانش‌آموزان، براساس عنوان و شکلهای یک متن، پیش‌بینی می‌کنند که آن متن دربارهٔ

---

1. standard
2. Steinmentz
3. academic standards
4. indicator
5. bench mark

چه چیزی صحبت می‌کند.

(منبع: برنامهٔ درسی مربوط به مشاغل بهداشت، مدارس کارولینای شمالی، پایه‌های ۱۲-۹، به نقل باتلر و مکمان، ۲۰۰۶، ص ۲۴)

استانداردهای تحصیلی به دو گونه‌اند: (۱) **استانداردهای محتوایی**[1] و (۲) **استانداردهای عملکردی**[2]. استانداردهای محتوایی بیاناتی هستند دربارهٔ واقعیتها، مفاهیم، اصول، روشها، تعمیمها، و نظریه‌های مربوط به موضوع درسی. استانداردهای عملکردی بیاناتی هستند دربارهٔ کارهایی که دانش‌آموزان پس از یادگیری استانداردهای محتوایی می‌توانند انجام دهند (لین و گرانلاند، ۲۰۰۰).

سنجه‌ها، در عمل، همان استانداردهای عملکردی هستند، زیرا با استفاده از آنها می‌توان نشان داد که یادگیرندگان به استانداردهای محتوایی رسیده‌اند یا نه. از این‌رو، در بسیاری از متون (از جمله اثر باتلر و مکمان، ۲۰۰۶)، اصطلاحات سنجه و استاندارد عملکردی هم معنا به کار رفته‌اند.

مقایسهٔ هدف با استاندارد

استاندارد و هدف بسیار به هم شبیه‌اند. به ویژه اینکه، استانداردها به هدفهای کلی و سنجه‌ها یا نشانگرها به هدفهای دقیق آموزشی می‌مانند. باتلر و مکمان (۲۰۰۶) گفته‌اند بیانات مربوط به استانداردها از لحاظ دامنه گسترده‌اند، اما سنجه‌ها بر رفتارهای مشخص دانش‌آموزان تأکید می‌ورزند. در بعضی مدارس آمریکا از استانداردها با اصطلاح *هدف کلی* یاد می‌شود و سنجه‌ها را *هدف آموزشی* می‌نامند. "سنجه‌ها توضیح می‌دهند که دانش‌آموزان چه کارهایی باید انجام دهند تا معلوم گردد که به استانداردها رسیده‌اند" (ص ۲۳).

هدف و استاندارد را از نگاه دیگری نیز می‌توان با هم مقایسه کرد. آن هم این است که استانداردها را عموماً کسانی در مراکز تصمیم‌گیری آموزش و پروش و در خارج از کلاس درس تعیین می‌کنند، اما هدفها را هنوز هم عمدتاً معلمان درسهای مختلف و برای استفاده در کلاسهای درس خودشان تهیه و تدوین می‌نمایند[3].

طبقه‌بندی هدفهای آموزشی

هدفهای آموزشی بسیار متنوع‌اند و، برای سهولت مطالعه، آنها را به گونه‌های مختلف طبقه‌بندی کرده‌اند. معروف‌ترین طبقه‌بندی از هدفهای آموزشی به وسیلهٔ گروهی از متخصصان آموزش و

---

1. content standards
2. performance standards

۳. برای اطلاعات بیشتر دربارهٔ استانداردهای تحصیلی به اثر دیگر مؤلف *سنجش، فرآیند و فرآوردهٔ یادگیری: روشهای قدیم و جدید* (۱۳۸۷) نگاه کنید.

ارزشیابی پیشرفت تحصیلی تهیه شده و به نام بنجامین اس. بلوم که سرپرستی این گروه را به عهده داشته معروف گشته است (بلوم، انگلهارت[1]، فرست[2]، هیل[3]، و کراتول[4]، ۱۹۵۶، ترجمهٔ سیف و علی آبادی، ۱۳۶۸). در این طبقه‌بندی، هدفهای آموزشی ابتدا به سه طبقهٔ کلی با نامهای حوزهٔ شناختی[5]، حوزهٔ عاطفی[6]، و حوزهٔ روانی-حرکتی[7] تقسیم شده‌اند و هر حوزه نیز دارای طبقه‌بندیهای مخصوص به خود است.

هدفهای حوزهٔ شناختی به جریانهایی که با فعالیتهای ذهنی و فکری آدمی سر و کار دارند مربوط می‌شوند. از این لحاظ، این حوزه مهم‌ترین حوزهٔ یادگیری به شمار می‌آید، زیرا اکثریت فعالیتهای تحصیلی آموزشگاهها و غالب موضوعهای درسی و هدفهای آموزشی به این حوزه مربوط می‌شوند. آزمونهای مورد نیاز برای سنجش هدفهای حوزهٔ شناختی آزمونهای توانایی‌های شناختی هستند که غالباً به صورت کتبی تهیه و اجرا می‌شوند.

حوزهٔ عاطفی شامل هدفهایی است که تغییرات حاصل در علاقه‌ها، نگرشها و ارزشها، و نیز رشد ارج‌شناسی و سازگاری را نشان می‌دهد. هر چند که همهٔ معلمان در ضمن آموزش هدفهای حوزهٔ شناختی، چه مستقیم و چه غیرمستقیم، در آموزش هدفهای حوزهٔ عاطفی نیز می‌کوشند، با این حال، چون محتوای برنامه‌های درسی را بیشتر هدفهای حوزهٔ شناختی تشکیل می‌دهند، کاربرد حوزهٔ عاطفی در سطح آموزشگاهها به اندازهٔ کاربرد حوزهٔ شناختی شهرت نیافته است.

حوزهٔ روانی-حرکتی به زمینهٔ مهارتهای حرکتی یا حرکات بدنی مربوط است. به سخن دیگر، هر فعالیتی که علاوه بر جنبهٔ روانی دارای جنبهٔ جسمانی نیز باشد در این حوزه جای دارد. تفاوت عمدهٔ هدفهای این حوزه با هدفهای حوزهٔ شناختی در این است که هدفهای حوزهٔ شناختی صرفاً به فعالیتهای ذهنی ختم می‌شوند، اما هدفهای حوزهٔ روانی-حرکتی هم به فعالیتهای ذهنی و عاطفی و هم به فعالیتهای جسمی نیاز دارند. کاربرد هدفهای حوزهٔ روانی-حرکتی نیز، مانند حوزهٔ عاطفی، به اندازهٔ کاربرد حوزهٔ شناختی معروفیت نیافته است.

برخلاف حوزهٔ شناختی، هدفهای حوزهٔ عاطفی و حوزهٔ روانی-حرکتی با آزمونهای کتبی متداول به خوبی قابل سنجش نیستند. چنان‌که در فصل یازدهم خواهیم دید، سنجش هدفهای حوزهٔ عاطفی و حوزهٔ روانی-حرکتی با وسایلی به غیر از آزمونهای مورد استفاده در حوزهٔ شناختی امکان‌پذیر است.

حال که با ویژگیهای کلی حوزه‌های سه‌گانهٔ یادگیری طبقه‌بندی بلوم آشنا شدیم، آماده‌ایم تا این حوزه‌ها را با جزئیات بیشتری مرور کنیم. ما در اینجا طبقات اصلی این حوزه‌های یادگیری را معرفی می‌کنیم. خوانندگان می‌توانند برای کسب اطلاعات بیشتر در بارهٔ این طبقه‌بندیها و خرده طبقات آنها به منابع و مراجعی که مورد اشارهٔ ما قرار خواهند گرفت، به‌ویژه به کتاب *طبقه‌بندی*

---

1. Engelhart
2. Furst
3. Hill
4. Krathwohl
5. cognitive domain
6. affective domain
7. psychomotor domain

هدفهای پرورشی، تألیف بلوم و همکاران (۱۹۵۶)، ترجمهٔ سیف و علی‌آبادی (۱۳۶۸)، و کتاب روانشناسی پرورشی نوین، تألیف سیف (۱۳۸۷) مراجعه نمایند.

حوزهٔ شناختی

بلوم و همکاران (۱۹۵۶، ۱۳۶۸) هدفهای حوزهٔ شناختی را در شش طبقهٔ کلی طبقه‌بندی کرده‌اند. هر یک از این طبقه‌ها یک نوع فرایند شناختی را توصیف می‌کند. بنا به آنچه که در زیر خواهد آمد، این شش طبقه به طور سلسله مراتبی از عینی و ساده به انتزاعی و پیچیده مرتب شده‌اند.

دانش[1]. دانش شامل یادآوری[2] (بازخوانی[3] و بازشناسی[4]) امور است. در واقع دانش عبارت است از تکرار پاسخهایی که قبلاً در موقعیت یادگیری تمرین و آموخته شده‌اند.

هدفهای نمونه:

یادگیرنده بتواند متغیر مستقل و متغیر وابسته را تعریف کند.
یادگیرنده بتواند سه ویژگی مهم هدفهای رفتاری کامل را ذکر کند.
یادگیرنده بتواند انواع آزمونهای عینی را نام ببرد.

فهمیدن[5]. توانایی درک منظور یا مقصود یک مطلب. فهمیدن به پاسخهایی بیشتر از آنچه قبلاً تمرین و آموخته شده‌اند نیاز دارد. این پاسخهای اضافی عبارت‌اند از ترجمه یا برگردان، تفسیر و تعبیر، خلاصه کردن، و کشف شباهتها و تفاوتها.

هدفهای نمونه:

یادگیرنده بتواند دلایل جنگ جهانی دوم را توضیح دهد.
یادگیرنده بتواند یک صفحه متن انگلیسی را به فارسی ترجمه کند.
یادگیرنده بتواند مثالهایی برای غذاهای پرکالری و کم‌کالری ذکر کند.

کار بستن[6]. توانایی استفاده از امور انتزاعی، قواعد و قوانین، اصول، اندیشه‌ها، و روشها در موقعیتهای عینی و عملی.

هدفهای نمونه:

یادگیرنده بتواند یک رژیم مناسب غذایی برای خودش تعیین کند.
یادگیرنده بتواند مسائل ریاضی مربوط به درس خود را حل نماید.
یادگیرنده بتواند اصول تقویت را در آموزش به کاربندد.

---

1. knowledge
2. remember
3. recall
4. recognition
5. comprehension
6. application

تحلیل[1]. توانایی شکستن یک مطلب یا موضوع به اجزا یا عناصر تشکیل‌دهندهٔ آن.

هدفهای نمونه:
یادگیرنده بتواند اجزای مختلف یک گزارش تحقیقی را مشخص کند.
یادگیرنده بتواند در یک مقاله عقاید را از واقعیتها متمایز سازد.
یادگیرنده بتواند هدف اصلی یک مطلب داستانی را تشخیص داده آن را بیان نماید.

ترکیب[2]. توانایی پهلوی هم گذاشتن عناصر و اجزا برای ایجاد یک الگو یا ساختار نو. ترکیب نامی است که بلوم و همکاران او (۱۹۵۶، ۱۳۶۸) برای آفرینندگی یا خلاقیت[3] به کار برده‌اند.

هدفهای نمونه:
یادگیرنده بتواند خاطرات تعطیلات تابستانی خود را به صورت یک داستان بنویسد.
یادگیرنده بتواند برای تدریس یک موضوع یک طرح آموزشی ابتکاری ارائه دهد.
یادگیرنده بتواند با استفاده از یافته‌های تعدادی پژوهش در یکی از زمینه‌های یادگیری آموزشگاهی یک نظریهٔ جدید آموزشی تدوین کند.

ارزشیابی[4]. توانایی قضاوت یا داوری کمّی و کیفی در بارهٔ امور، با توجه به ملاکهای معین.

هدفهای نمونه:
یادگیرنده بتواند یک داستان یا نمایشنامه را نقد و بررسی کند.
یادگیرنده بتواند یک پایان‌نامهٔ دانشجویی را با توجه به اصولی که در انجام و تدوین این‌گونه فعالیتهای پژوهشی باید رعایت گردند مورد داوری قرار دهد.
یادگیرنده بتواند یک مقالهٔ انتقادی در بارهٔ استفاده از تنبیه بدنی در مدارس بنویسد.

طبقه‌بندی تجدیدنظرشدهٔ حوزهٔ شناختی

اخیراً چندنفر از صاحب‌نظران آموزشی (آندرسون[5]، کراتول، و همکاران، ۲۰۰۱) در طبقه‌بندی حوزهٔ شناختی معروف به طبقه‌بندی بلوم تجدیدنظر کرده و طبقه‌بندی تازه‌ای با نام ***یک طبقه‌بندی برای یادگیری، آموزش، و سنجش*** به وجود آورده‌اند. پین (۲۰۰۳) در رابطه با این طبقه‌بندی گفته است:

دلیل عمدهٔ ایجاد طبقه‌بندی تازه آسان ساختن ارتباط در میان معلمان، تولیدکنندگان برنامه‌های درسی، و متخصصان سنجش و ارزشیابی بوده است. طبقه‌بندی اصلی [طبقه‌بندی بلوم] در میان آزمون‌سازان

1. analysis
2. synthesis
3. creativity
4. evaluation
5. Anderson

و ارزشیابان آموزشی پذیرش عام یافته بوده، زیرا نیاز آنان را برای بیان هدفهای دقیق برآورده می‌ساخته است. شکل تازهٔ طبقه‌بندی نیز هم بر فرایند آموزش ـ یادگیری و هم بر سنجش نتایج تأکید می‌کند. (ص ۱۱۶)

در طبقه‌بندی تازهٔ حوزهٔ شناختی یک *بُعد دانش* و یک *بُعد فرایندشناختی* وجود دارد. *بُعد دانش* شامل دانش امور واقعی، دانش مفهومی، دانش روندی، و دانش فراشناختی است. *بُعد* فرایندشناختی دربرگیرندهٔ به یادآوردن، فهمیدن، به کاربستن، تحلیل کردن، ارزشیابی کردن، و آفریدن است. طبقه‌های هر دو بُعد این طبقه‌بندی به صورت سلسله مراتبی، یعنی از عینی به انتزاعی و از ساده به پیچیده، تنظیم یافته‌اند. در زیر به توضیح مختصر طبقه‌ها و خرده طبقه‌های این طبقه‌بندی همراه با نمونه‌هایی از آنها توجه کنید. (نمونه‌ها از اصل کتاب و کتاب *سنجش آموزش کاربردی* تألیف دیوید آلن پین، ۲۰۰۳ صص ۱۱۷-۱۱۶ اقتباس شده‌اند.)

بُعد دانش

الف ـ دانش امور واقعی[1]. دانش عناصر اساسی موردنیاز یادگیرنده برای آشنا شدن با یک رشتهٔ علمی یا حل مسائل مربوط به آن

الف الف دانش اصطلاحات[2]

نمونه:

دانش اصطلاحات علمی (مانند قسمتهای مختلف یک یاخته)

الف ب دانش اجزاء و عناصر خاص[3]

نمونه:

دانش نامها، مکانها، و رویدادهای مهم

ب ـ دانش مفهومی[4]. دانش مقوله‌ها، طبقه‌ها، و روابط میان آنها

ب الف دانش طبقه‌بندیها و طبقه‌ها[5]

نمونه:

دانش انواع گوناگون ادبیات

ب ب دانش اصلها و تعمیمها[6]

نمونه:

دانش قوانین اساسی فیزیک

ب پ دانش نظریه‌ها، الگوها، و ساختها[7]

---

1. factual knowledge
2. knowledge of terminology
3. knowledge of specific detailes and elements
4. conceptual knowledge
5. knowledge of classifications and categories
6. knowledge of principles and generalizations
7. knowledge of theories, models, and structures

نمونه: دانش ساختار مدیریتی حکومت محلی

ب ـ دانش روندی[1]. دانش چگونه انجام دادن کارها

ب الف دانش مهارتها و الگوریتمهای خاص یک رشته[2]

نمونه: دانش مهارتهای مورد استفاده در نقاشی با آبرنگ

ب ب دانش فنها و روشهای خاص یک رشته[3]

نمونه: دانش روشهای پژوهش متناسب با علوم اجتماعی

ب ت دانش معیارهای تعیین زمان استفاده از روشهای مناسب[4]

نمونه: دانش معیارهای تعیین اینکه برای داده‌های حاصل از یک آزمایش خاص کدام روش آماری را باید به کار بست.

ت ـ دانش فراشناختی[5]. دانش شناخت به طور کلی و دانش فرد دربارهٔ شناخت خودش

ت الف دانش راهبردی (استراتژیک)[6]

نمونه: این دانش که تکرار اطلاعات یکی از راههای حفظ اطلاعات است.

ت ب دانش مربوط به تکالیف شناختی، شامل دانش موقعیتی و دانش مشروط[7]

نمونه: این دانش که درک یک کتاب تخصصی علمی دشوارتر از درک یک کتاب عمومی است.

ت ث دانشِ خود[8]

نمونه: این دانش که شخص در بعضی زمینه‌ها مطلع و در بعضی دیگر بی‌اطلاع است.

بُعد فرایندشناختی

۱. **به یاد آوردن**[9]. بازیابی اطلاعات از حافظهٔ درازمدت

۱/۱ بازشناسی[10] (تشخیص)

---

1. procedural knowledge
2. knowledge of subject-specific skills and algorithms
3. knowledge of subject-specific techniques and methods
4. knowledge of criteria for determining when to use appropriate procedures
5. metacognitive knowledge
6. strategic knowledge
7. knowledge about cognitive tasks, including contexual and conditional knowledge
8. self knowledge
9. remember
10. recognizing

نمونه:
تشخیص نام نویسندگان مشهور
۱/۲ بازخوانی[1] (بازیابی)

نمونه:
یادآوری تاریخ جنگهای جهانی مهم

۲. فهمیدن[2]. ساختن معنی از یک ارتباط
۲/۱ تفسیر کردن[3] (تغییر بیان، معنی کردن، ترجمه کردن)

نمونه:
برگرداندن یک جمله از زبان انگلیسی به زبان فارسی
۲/۲ مثال آوردن[4] (توضیح دادن)

نمونه:
گفتن نمونه‌هایی از سبکهای مختلف نقاشی
۲/۳ طبقه‌بندی کردن[5]

نمونه:
طبقه‌بندی کردن انواع اختلالهای روانی براساس مشاهده یا توصیف
۲/۴ خلاصه کردن[6] (به صورت چکیده درآوردن)

نمونه:
خلاصه کردن وقایع تاریخی که به صورت تصویر نشان داده شده‌اند
۲/۵ استنباط کردن[7]

نمونه:
در یادگیری یک زبان خارجی، استنباط اصول دستوری از مثالها
۲/۶ مقایسه کردن[8]

نمونه:
درک رویدادهای تاریخی از راه مقایسهٔ آنها با موقعیتهای مشابه
۲/۷ تبیین کردن[9]

نمونه:
بیان علتهای پیدایش همزمان پروتستانیسم و سرمایه‌داری در قرن شانزدهم میلادی

۳. به کاربستن[10]. استفاده از یک روش در یک موقعیت خاص
۳/۱ اجرا کردن[11]

---

1. recalling
2. understand
3. interpreting
4. exemplifying
5. classifying
6. summurizing
7. inferring
8. comparing
9. explaining
10. apply
11. executing

نمونه:
تقسیم یک عدد صحیح چندرقمی بریک عدد صحیح چندرقمی دیگر
۳/۲ مورد استفاده قراردادن[1]

نمونه:
توانایی حل کردن مسائل مالی شخصی مختلف

۴. **تحلیل کردن**[2]. شکستن مواد یا مطالب به عناصر تشکیل‌دهنده و توصیف روابط میان آنها
۴/۱ متمایز کردن[3] یا ویژه‌سازی

نمونه:
تعیین نکات مهم و غیرمهم گزارشهای تحقیقی
۴/۲ سازمان دادن[4]

نمونه:
بیان یک استدلال حاوی مطالب موافق و مخالف دربارهٔ پرداخت دستمزد به دانشجویان دانشگاه
۴/۳ نسبت دادن[5]

نمونه:
تعیین ارزشهای مذهبی نویسندهٔ یک مقاله از محتوای مقالهٔ او دربارهٔ اخلاق

۵. **ارزشیابی کردن**[6]. داوری کردن با استفاده از معیار یا ملاک
۵/۱ وارسی کردن[7] (آزمون کردن)

نمونه:
تعیین اینکه نتیجه‌گیریها به طور منطقی حاصل داده‌های گزارش شده هستند یا نه
۵/۲ نقد کردن[8] (داوری کردن)

نمونه:
قضاوت کردن در این باره که کدام‌یک از دو روش آماری موجود برای تحلیل یک مجموعه داده مناسب‌تر است

۶. **آفریدن (خلق کردن)**[9]. قرار دادن عناصر یا اجزاء در یک الگو یا ساختار نو
۶/۱ تولید کردن[10] (فرضیه ساختن)

نمونه:
ساختن یک فرضیه برای توجیه داده‌های مشاهده شده
۶/۲ طرح‌ریزی کردن[11]

---

1. implementing
2. analyze
3. differentiating
4. organizing
5. attributing
6. evaluate
7. checking
8. critiquing
9. create
10. generating
11. planning

نمونه:
طراحی یک برنامهٔ ارزشیابی برای تعیین اثربخشی برنامه‌های آموزش هنر در یک مدرسه

۶/۳ پدید آوردن[1] (ساختن)

نمونه:
ساختن یک وسیلهٔ سنجش برای ارزشیابی از درس علوم اجتماعی پایهٔ سوم راهنمایی

طبقه‌بندی تجدیدنظر شدهٔ حوزهٔ شناختی را می‌توان برای مقاصد مختلفی مورد استفاده قرارداد. از جملهٔ آنها می‌توان موارد زیر را نام برد: "کمک به معلمان در انتخاب فعالیتهای آموزشی، تدوین برنامه‌های درسی، و تحلیل هدفهای یادگیری" (پین، ۲۰۰۳، ص ۱۱۷). در جدول ۲-۶ با استفاده از تعدادی هدف آموزشی و طبقه‌بندی آنها چگونگی استفاده از طبقه‌بندی شناختی تجدیدنظر شده را نشان داده‌ایم.

جدول ۲-۶ نمونه هدفهای یادگیری مربوط به طبقه‌های مختلف طبقه‌بندی دوبُعدی آندرسون و کراتول (اقتباس از آثار آندرسون، کراتول، و همکاران، ۲۰۰۱، پین، ۲۰۰۳)

بُعد دانش	بُعد فرایند شناختی					
	۱ به یاد آوردن	۲ فهمیدن	۳ به کار بستن	۴ تحلیل کردن	۵ ارزشیابی کردن	۶ آفریدن
الف - دانش امورواقعی	دانشجو بتواند تاریخ تأسیس اولین آزمایشگاه روانشناسی را از حفظ بگوید.					
ب - دانش مفهومی		دانشجو بتواند طبقه‌بندی تجدیدنظر شدهٔ حوزهٔ شناختی‌را به زبان خودش توضیح دهد.		دانش‌آموز بتواند اعداد گویا و گنگ را از هم تشخیص دهد.		
پ - دانش روندی			دانشجو بتواند ضریب همبستگی دو دسته نمره را حساب کند.			
ت - دانش فراشناختی		دانش‌آموز بتواند تشخیص دهد که در چه موضوعهایی توانایی یادگیری بیشتری دارد.			دانش‌آموز بتواند طرحی‌برای حل کردن مسائل مربوط به قانون اُهم انتخاب کند که با سطح فعلی درک او بیشترین تناسب را داشته باشد.	دانشجو بتواند طرح تازه‌ای برای سنجش یادگیری در یک درس متناسب‌با نیازهای معلم و دانش‌آموزان یک کلاس تهیه کند.

1. producing

حوزهٔ عاطفی

کراتول، بلوم، و ماسیا[1] (۱۹۶۴) هدفهای حوزهٔ عاطفی را طبقه‌بندی کرده‌اند. این طبقه‌بندی رامهایی را توصیف می‌کند که در آنها یادگیرندگان از ارزشها و نگرشهایی که رفتار آدمی را هدایت می‌نمایند آگاه می‌شوند و آنها را می‌پذیرند. مانند حوزهٔ شناختی، طبقه‌های حوزهٔ عاطفی نیز به صورت سلسله مراتبی، یعنی از ساده به پیچیده، مرتب شده‌اند. این طبقه‌بندی شامل پنج طبقهٔ اصلی به شرح زیر است:

دریافت کردن[2] **(توجه کردن**[3]**)**. حساسیت نسبت به وجود بعضی پدیده‌ها و محرکهای محیطی و میل به دریافت کردن آنها یا توجه کردن به آنها.

هدفهای نمونه:

یادگیرنده مایل است به اتفاقاتی که در کلاس درس می‌افتند توجه نماید.

یادگیرنده می‌پذیرد که دانش‌آموزان نژادهای مختلف وجود دارند.

پاسخ دادن[4]. توجه فعالانه به پدیده‌ها و انجام کارهایی در بارهٔ آنها، یا فعالیت حاکی از پذیرش یک عقیده یا خط‌مشی.

هدفهای نمونه:

یادگیرنده به طور داوطلبانه کتابها و مجله‌های خاص کودکان را مطالعه می‌کند.

یادگیرنده با دانش‌آموزان نژادهای مختلف به فعالیت گروهی می‌پردازد.

ارزش گذاردن[5]. احساس یا باور پایدار حاکی از ارزشمند بودن یک چیز، یک اندیشه، یک شخص، یا یک گروه، یا ابراز عقیده یا نگرش در بارهٔ ارزش یک چیز.

هدفهای نمونه:

یادگیرنده برای دفاع از یک موضوع اجتماعی که برای آن ارزش قائل است نظر خود را در بارهٔ آن به صورت نامه‌ای برای درج در مطبوعات ارسال می‌دارد.

یادگیرنده در بحث گروهی از عقاید یکی از دوستان خود که به آن باور دارد دفاع می‌کند.

سازمان دادن به ارزشها[6]. دسته‌بندی و نظامدار کردن ارزشها به گونه‌ای که بعضی از آنها مهمتر از بعضی دیگر جلوه کنند.

هدفهای نمونه:

یادگیرنده نیاز به یافتن نوعی تعادل بین رشد اقتصادی و حفظ منابع طبیعی را تشخیص می‌دهد.

---

1. Masia
2. receiving
3. attending
4. responding
5. valuing
6. organization of values

یادگیرنده باورهای خود را در هم می‌آمیزد و یک نظام ارزشی برای خود به وجود می‌آورد

شخصیت پذیرفتن[1]. تبدیل نظام ارزشی به صورت یک سبک زندگی یا یک دیدگاه فلسفی. این کار از درهم‌آمیزی صِرف ارزشها و تعیین روابط بین آنها فراتر می‌رود و نوعی فلسفهٔ پایدار زندگی را شامل می‌شود.

هدفهای نمونه:

یادگیرنده برای خود یک فلسفهٔ پایدار زندگی برمی‌گزیند.

یادگیرنده در موقعیتهای مختلف متناسب با نظام ارزشی خودش رفتار می‌کند.

حوزهٔ روانی ـ حرکتی

برخلاف طبقه‌بندی حوزهٔ شناختی که ویژهٔ کتاب بلوم و همکاران (۱۹۵۶، ۱۳۶۸) و طبقه‌بندی حوزهٔ عاطفی که مختص کتاب کراتول و همکاران (۱۹۶۴) است، برای حوزهٔ روانی ـ حرکتی تعدادی طبقه‌بندی پدید آمده‌اند. ما در اینجا خلاصهٔ طبقه‌بندی سیمپسون[2] (۱۹۷۲) را می‌آوریم.

طبقه‌بندی سیمپسون از حوزهٔ روانی ـ حرکتی شامل هفت طبقهٔ اصلی و تعدادی خرده طبقه است. ما باز هم در اینجا طبقات اصلی این طبقه‌بندی را معرفی می‌کنیم. مانند دو حوزهٔ دیگر، طبقه‌های حوزهٔ روانی ـ حرکتی نیز به صورت سلسله مراتبی، یعنی از ساده به پیچیده، مرتب شده‌اند.

ادراک (حسی)[3]. استفاده از حواس برای هدایت کنشهای حرکتی.

هدف نمونه:

یادگیرنده با چشیدن غذا (استفاده از حس چشایی) مزهٔ آن را تشخیص می‌دهد.

آمادگی یا آمایه[4]. آماده بودن برای انجام یک عمل یا یک رشته اعمال معین.

هدف نمونه:

یادگیرنده آماده است به توپی که به سوی او شوت (پرتاب) می‌شود ضربه بزند.

پاسخ هدایت شده[5]. عمل کردن به کمک هدایت یک الگو، یا اعمالی که در مراحل اولیهٔ یادگیری انجام می‌شوند.

هدف نمونه:

یادگیرنده یک راکت تنیس را مطابق با آنچه مربی انجام می‌دهد حرکت می‌دهد.

مکانیسم[6] **یا عادت**. انجام اعمالی که نسبتاً مشخص و عادتی و خودکار هستند.

---

1. characterization
2. Simpson
3. perception
4. set
5. guided response
6. mechanism

هدف نمونه:
یادگیرنده بتواند اولین سِروْ را در منطقه سِروْ در ۷۰ درصد از موارد بگیرد.

پاسخ پیچیدهٔ آشکار[1]. انجام رشته اعمالی که تا حدودی به صورت خودکار یا عادتی درآمده‌اند یا ترکیب تعدادی از عادتها.

هدف نمونه:
یادگیرنده بتواند طول یک استخر را شنا کند.
دقت کنید که شناکردن یک پاسخ پیچیده است، در حالی که هریک از حرکتهای فردی شنا یک عادت به حساب می‌آید.

انطباق[2]. استفاده از مهارتهای قبلاً آموخته‌شده برای انجام تکالیف تازه اما مربوط به تکالیف گذشته.

هدف نمونه:
یادگیرنده نوشته‌های دارای درجات دشواری مختلف را با سرعتهای متفاوت تایپ می‌کند.

ابتکار[3]. خلق الگوهای تازهٔ حرکتی برای حل مسائل غیرمعمول یا برخورد با موقعیتهای جدید.

هدف نمونه:
یادگیرنده بتواند یک قطعه آهنگ موسیقی را بسازد و آن را اجرا کند.

جدول مشخصات

در آغاز این فصل گفتیم که ارزشیابی پیشرفت تحصیلی مستلزم انجام دو فعالیت است: (۱) تعیین و توصیف هدفهای آموزشی یا هدفهای یادگیری، و (۲) سنجش یا اندازه‌گیری عملکرد یادگیرندگان. در صفحات قبل هدفهای آموزشی را معرفی کردیم؛ از این پس چگونگی اندازه‌گیری یا سنجش عملکرد یادگیرندگان را بحث می‌کنیم. از آنجاکه یکی از وسایل مهم اندازه‌گیری پیشرفت تحصیلی آزمونهای پیشرفت تحصیلی به حساب می‌آیند، لازم است روش تهیهٔ آزمونهای پیشرفت تحصیلی را توضیح دهیم. مهم‌ترین گام در تهیهٔ آزمونهای پیشرفت تحصیلی تهیهٔ **جدول مشخصات**[4] برای موضوعی است که قرار است آزمون آن تهیه بشود.

جدول مشخصات یک جدول دوبعدی است با تعدادی ردیف و ستون، متناسب با محتوا و هدفهای آموزشی درس. در ردیف بالای جدول که بُعد محتوا نام دارد اجزای مطالب درس نوشته می‌شوند و در ستون کنار آن که بُعد هدف نامیده می‌شود فهرست هدفهای آموزشی قرار می‌گیرند.

---

1. complex overt response
2. adaptation
3. origination
4. table of specifications

به بُعد محتوا بُعد تجلیات[1] نیز می‌گویند، زیرا این بُعد معرف نمودها یا تجلیات بعد محتوا است. اصطلاح تجلیات غالباً برای معرفی بُعد عمودی جدول مشخصات پرسشنامه‌های سنجش عملکرد نوعی به کار می‌رود، ولی برای معرفی این بُعدِ جدول مشخصات آزمونهای سنجش عملکرد بیشینه، از جمله آزمونهای پیشرفت تحصیلی، از اصطلاح بعد هدف استفاده می‌شود[2].

جدول ۳-۶ یک جدول مشخصات کلی را نشان می‌دهد که در آن تمامی محتوای یک درس و همهٔ هدفهای یادگیری سه حوزهٔ شناختی، عاطفی، و روانی حرکتی پیش‌بینی شده است.

چنان که در جدول مشخصات ۳-۶ می‌بینید، در بُعد محتوا همهٔ بخشها و فصلهای کتاب یا درسی که قرار است سنجش شود و در بُعد هدف تمامی طبقه‌های حوزه‌های سه گانهٔ یادگیری درنظر گرفته می‌شوند. افزون بر اینها، در درون خانه‌های جدول مشخصات نیز تعداد سؤالهایی که برای سنجش هدفها و محتوای هر بخش به کار خواهند رفت نوشته می‌شود.

جدول ۳-۶ الگوی کلی جدول مشخصات آزمونهای پیشرفت تحصیلی

بُعد هدف	بُعد محتوا	بخش ۱: فصل اول	بخش ۱: فصل دوم	بخش ۱: فصل سوم	بخش ۱: ... و غیره	بخش ۲: فصل اول	بخش ۲: فصل دوم	بخش ۲: فصل سوم	بخش ۲: ... و غیره	بخش ۳: فصل اول	بخش ۳: فصل دوم	بخش ۳: فصل سوم	بخش ۳: ... و غیره	... و غیره
هدفهای حوزهٔ شناختی: دانش، فهمیدن، کاربستن، ... و غیره		تعداد سؤالها				تعداد سؤالها				تعداد سؤالها				
هدفهای حوزهٔ عاطفی: دریافت کردن، پاسخ دادن، ارزش‌گذاردن، ... و غیره		تعداد سؤالها				تعداد سؤالها				تعداد سؤالها				
هدفهای حوزهٔ روانی-حرکتی: ادراک، آمادگی، پاسخ هدایت شده، ... و غیره		تعداد سؤالها				تعداد سؤالها				تعداد سؤالها				

1. manifestation

۲. برای توضیحات بیشتر نگاه کنید به اثر دیگر مؤلف: ساختن ابزارهای اندازه گیری متغیرهای پژوهشی (سیف، ۱۳۹۰).

نقش جدول مشخصات آزمون

یک آزمون خوب پیشرفت تحصیلی آزمونی است که به بهترین شکل منعکس‌کنندهٔ تمامی هدفهای آموزشی و همهٔ محتوای برنامهٔ درسی باشد. از لحاظ نظری، بهترین آزمون برای یک درس یا یک موضوع آزمونی است که تمام مطالب و کلیهٔ هدفهای آموزشی آن درس را شامل شود. اما اگر درس یا موضوع مفصل باشد تهیهٔ چنین آزمونی غیرعملی است. بنابراین، به جای گنجانیدن همهٔ سؤالهای ممکنِ تمامی محتوا و هدفها در آزمون، تهیهٔ کنندهٔ آزمون به انتخاب تعدادی سؤال نمونه که معرف هدفها و محتوای درس باشند اقدام می‌کند. در صورتی چنین آزمونی یک آزمون خوب یا به اصطلاح یک آزمون دارای **روایی**[1] خواهد بود که سؤالهای آن حاوی نمونهٔ درستی از هدفها و محتوای درس باشند. جدول مشخصات به معلم یا آزمون‌ساز کمک می‌کند تا از مجموع بزرگی از سؤالهای ممکن نمونه‌ای را که به‌خوبی مُعرف محتوا و هدفهای درس باشند انتخاب کند. بنا به آنچه در *استانداردهای آزمودن روانی و تربیتی* (انجمن پژوهش آموزشی آمریکا، انجمن روانشناسی آمریکا، و شورای ملی اندازه‌گیری در آموزش و پرورش، ۱۹۹۹) آمده است:

> وقتی که از یک آزمون به عنوان شاخص پیشرفت در یک حیطهٔ آموزشی استفاده می‌شود، لازم است شواهد مربوط به اینکه آن آزمون نمونه‌ای از مجموع دانشهای موردنظر فراهم می‌آورد و فرایندهای موجود در بُعد هدفها را فرا می‌خواند ارائه شود. چنین تحلیلی باید آن جنبه‌هایی را که آزمون معرف آنهاست و آنهایی را که نمی‌تواند معرف باشد آشکار سازد.... جدول مشخصات آزمون به ایجاد تعادل بین آموزش و سنجش کمک می‌کند. (به نقلِ پین، ۲۰۰۳، ص ۱۰۶)

چگونگی توزیع سؤالها به بخشهای مختلف

بعد از تکمیل بعد محتوا و بُعد هدف جدول مشخصات، آزمون‌ساز باید تعیین کند که سؤالهای آزمون خود را به چه صورتی در میان هدفها و عنوانهای درس توزیع نماید؛ یعنی تعداد سؤالهایی را که به هر یک از موضوعهای درسی و هدفهای آموزشی اختصاص خواهند داشت تعیین نماید و این تعداد را یا به صورت تعداد واقعی سؤالها یا به صورت درصد در خانه‌های جدول مشخصات وارد کند. برای تعیین درصد سؤالهایی که به بخشهای مختلف محتوا و هدفها تعلق می‌گیرند جواب سؤالهای زیر را باید تعیین کرد: (۱) کدام بخش از محتوا و کدام دسته از هدفها دارای حجم بیشتری است و وقت بیشتری صرف آموزش آنها شده است؟ (۲) کدام هدفها از لحاظ نگهداری در حافظه و کاربرد در آینده ارزش بیشتری دارند؟ (۳) کدام قسمت یا قسمتهای درس یا موضوع درسی از اهمیت بیشتری برخوردارند؟ نحوهٔ توزیع سؤالهای آزمون در میان بخشها و هدفهای مختلف باید متناسب با جواب این سؤالها باشد.

---

۱. validity (بحث مربوط به روایی آزمون در فصل ۲۰ کتاب آمده است.)

چنانچه بخشهای مختلف محتوا از اهمیت یکسانی برخوردار باشند و در توزیع سؤالها در میان بخشهای مختلف تنها بخواهیم مقدار ساعتهای تدریس را برای هر بخش منظور کنیم، در آن صورت می‌توانیم از فرمول زیر استفاده نماییم:

نسبت ساعتهای تدریس هر بخش = تعداد ساعتهایی که صرف تدریس بخش شده / تعداد کل ساعتهای تدریس

برای روشن شدن مطلب، از اطلاعات جدول ۴ـ۶ استفاده می‌کنیم. در این جدول، کتاب علوم پنجم دبستان به ۹ بخش تقسیم شده است، و همچنین تعداد ساعتهایی که به تدریس هر بخش اختصاص داده شده نیز آمده است. اگر بخواهیم یک آزمون ۶۴ سؤالی برای این کتاب تهیه کنیم و متناسب با ساعات تدریس سؤالها را در میان بخشهای مختلف تقسیم نماییم، به گونهٔ زیر عمل می‌کنیم:

نسبت ساعتهای تدریس بخش اول = $\frac{۱۳}{۶۹}$ = ۰/۱۹

از آنجا که در مجموع ۶۴ سؤال خواهیم داشت، لازم است ۰/۱۹ را در کل سؤالها یعنی ۶۴ ضرب کنیم تا تعداد سؤالهای بخش اول تعیین شود.

تعداد سؤالهای بخش اول = ۰/۱۹ × ۶۴ = ۱۲/۱۵ ~ ۱۲

برای تعیین نسبت ساعتهای تدریس و تعداد سؤالهای بخشهای دیگر کتاب نیز همین محاسبات را ادامه می‌دهیم. ارقام جدول ۴ـ۶ این‌گونه به دست آمده‌اند.

جدول ۴ـ۶ اجزای محتوایی کتاب علوم پنجم دبستان (اقتباس از مؤسسهٔ تحقیقات و برنامه‌ریزی علمی و آموزشی وزارت آموزش و پرورش، ۱۳۵۳)

عنوان بخش	ساعتهای تدریس بخش	نسبت ساعتهای تدریس بخش	تعداد سؤالهای بخش
۱. طبقه‌بندی موجودات	۱۳	۰/۱۹	۱۲
۲. مواد از چه تشکیل شده‌اند؟	۸	۰/۱۲	۸
۳. هوا در اطراف ما	۶	۰/۰۸	۵
۴. پیش‌بینی وضع هوا	۸	۰/۱۲	۸
۵. آتش چیست؟	۵	۰/۰۷	۵
۶. بدن شما چگونه کار می‌کند؟	۱۰	۰/۱۵	۹
۷. ستارگان	۷	۰/۱۰	۶
۸. پرواز در هوا	۷	۰/۱۰	۶
۹. اکتشافات فضایی	۵	۰/۰۷	۵
جمع	۶۹	۱	۶۴

انواع جدولهای مشخصات

هم در تهیهٔ آزمونهای وابسته به ملاک و هم در تهیهٔ آزمونهای وابسته به هنجار استفاده از جدول مشخصات ضروری است. با این حال، تفاوت عمدهٔ بین این دو نوع جدول مشخصات در آن است که در بُعد هدف جدول مشخصات آزمونهای وابسته به ملاک هدفهای دقیق آموزشی به طور کامل وارد می‌شوند (نگاه کنید به جدولهای ۵-۶ و ۶-۶)، اما در بُعد هدف جدول مشخصات آزمونهای وابسته به هنجار، معمولاً به جای هدفهای دقیق آموزشی، اصطلاحات و عبارتهای کلی‌تری به کار می‌روند که مُعرف طبقات هدفهای آموزشی یا انواع یادگیری هستند (مانند جدول ۷-۶).

جدول ۵-۶ جدول مشخصات آزمون تراکمی وابسته به ملاک درس هواشناسی مربوط به دورهٔ راهنمایی تحصیلی (اقتباس از اثر لین و گرانلاند، ۲۰۰۰، ص ۵۶۲)

بُعد محتوا / بُعد هدف	فصل ۱: فشار هوا			فصل ۲: باد			فصل ۳: دمای هوا			فصل ۴: رطوبت و بارندگی				فصل ۵: ابرها				تعداد و درصد سؤالها
	اندازه‌گیری و گزارش فشار هوا	عوامل مؤثر بر فشار هوا	فشار هوا و تغییرات هوا	اندازه‌گیری سرعت و جهت باد	عوامل مؤثر بر سرعت و جهت باد	ملاحظم مربوط به گزارش سرعت وجهت باد	اندازه‌گیری و گزارش دمای هوا	عوامل مؤثر بر دمای هوا	رابطهٔ دما و شکل‌گیری آب و هوا	اندازه‌گیری و گزارش رطوبت	عوامل مؤثر بر رطوبت	شکلهای مختلف بارندگی	اندازه‌گیری و گزارش بارندگی	انواع ابرها	ویژگی ابرها	عوامل مؤثر بر شکل‌گیری ابرها	ملاحظم مربوط به انواع ابرها	
دانش	۳		۳	۳		۳	۳		۳	۳			۳	۶			۶	۱۸ ٪۳۰
فهمیدن	۳			۶			۳			۶				۳			۳	۲۱ ٪۳۵
کاربستن	۳			۶			۳			۶				۳			۳	۲۱ ٪۳۵
تعداد سؤالها	۹			۱۵			۹			۱۵				۱۲				۶۰
درصد سؤالها	٪۱۵			٪۲۵			٪۱۵			٪۲۵				٪۲۰				٪۱۰۰

دانش
۱. تعریف هریک از اصطلاحات درس را بنویسد.
۲. اصطلاحات مربوط به هریک از عناصر هواشناسی را تشخیص دهد.
۳. هر علامتی را به عنصر هوایی آن ربط دهد.
۴. معنی هریک از علامتها را مشخص کند.
۵. عناصر مؤثر بر آب و هوا را نام ببرد.
۶. ابزارهای مورد استفاده برای اندازه‌گیری عناصر هوا را نام ببرد.
۷. نام هریک از انواع ابرها را با توصیف ویژگیهای آن ربط دهد.

فهمیدن
۸. چگونگی تشکیل ابرها را توضیح دهد.
۹. چگونگی تأثیر ابرها را بر آب و هوا و دما توضیح دهد.
۱۰. ویژگیهای یک شرایط آب و هوایی معین را توصیف کند.
۱۱. گزارش محتمل و گزارش غیرمحتمل هوا را از هم تشخیص دهد.
۱۲. براساس تغییر شرایط هوا را پیش‌بینی کند.
۱۳. هوای یک منطقهٔ معین را شرح دهد.

کاربستن
۱۴. با استفاده از یک توصیف کتبی از شرایط موجود، یک نقشهٔ هوایی بکشد.
۱۵. تغییرات فشار هوا را اندازه‌گیری و گزارش کند.
۱۶. رطوبت نسبی را اندازه‌گیری کند.

جدول ۶-۶ جدول مشخصات آزمون تراکمی وابسته به‌ملاک درس سلول (اقتباس از اثر نیتکو، ۲۰۰۱، ص ۴۷۳)

بُعد محتوا / بُعد هدف		درس ۱- قسمتهای مهم سلول		درس ۲- سلول حیوانی در مقایسه با سلول گیاهی		درس ۳- غشاء سلولی			درس ۴- تقسیم سلولی			تعداد سؤالها	درصد سؤالها
		هسته	سیتوپلاسم	غشاء سلولی	شباهتها	تفاوتها	کارهای حیاتی	انتشار	موادی که به وسیلهٔ سلول انتشار می‌یابد	مراحل تقسیم	کروموزوم و DNA	تقسیم سلولی حیوانی و گیاهی	
یادآوری	دانش امور واقعی: ۱. قسمتهای مختلف سلول را نام ببرد. ۲. موادی را که به وسیلهٔ غشاء سلول انتشار می‌یابند نام ببرد.		X					X	۳ سؤال		۲ سؤال		۷	٪۳۵
	دانش مفهومی: ۳. انتشار را تعریف کند. ۴. تقسیم سلولی، کروموزوم، و DNA را تعریف کند.							X		X				
نهیدن	دانش مفهومی: ۵. تفاوت بین انتشار و اکسیداسیون را توضیح دهد. ۶. از روی عکس واقعی سلول حیوانی یا گیاهی قسمتهای مختلف تقسیم سلولی را مشخص کند. ۷. تفاوتهای بین سلول حیوانی و سلول گیاهی را توضیح دهد. ۸. دیوارهٔ سلولی و غشاء سلولی را توصیف کند. ۹. تفاوت بین تقسیم سلولی را در گیاهان و حیوانها توضیح دهد. ۱۰. ازروی یک طرح خطی قسمتهای تقسیم سلولی را مشخص کند.		۳ سؤال	X X X X	۴ سؤال		X	۲ سؤال	X	X	۳ سؤال		۱۲	٪۶۰
کاربستن	دانش مفهومی: ۱۱. با دردست داشتن تعداد کروموزومهای یک سلول پیش از تقسیم، تعداد کروموزومهای هر سلول را پس از تقسیم تعیین کند.									X	۱ سؤال		۱	٪۵
تعداد سؤالها			۵		۴			۵			۶		۲۰	
درصد سؤالها			٪۲۵		٪۲۰			٪۲۵			٪۳۰			٪۱۰۰

۱. هدفهای این جدول مشخصات براساس طبقه‌بندی تازهٔ حوزهٔ شناختی طبقه‌بندی شده‌اند.

دقت کنید که در جدول ۷-۶، در بُعد هدف، به جای هدفهای آموزشی، طبقات هدفهای آموزشی با سطوح یادگیری آمده‌اند که دقیقاً با طبقه‌بندی انواع هدفهای آموزشی معروف به طبقه‌بندی بلوم یکسان نیستند، اما با آنها قابل انطباق‌اند. در این جدول، یادآوری اطلاعات واقعی مربوط به طبقهٔ دانش است. همچنین انجام دادن اعمال ریاضی و حل کردن مسائل در طبقهٔ کاربستن، نشان دادن درک اندیشه‌ها و مفاهیم در طبقهٔ فهمیدن، ابتکار و بینش در طبقهٔ ترکیب، و سرانجام فرایندهای عالی‌تر ذهنی در طبقه‌های ارزشیابی، ترکیب، و تحلیل قرار دارند. ویژگی دیگر این جدول مشخصات آن است که، به جای تعداد یا درصد سؤالها در درون خانه‌های جدول،

جدول ۷-۶ جدول مشخصات آزمون ریاضی وابسته به هنجار مربوط به پایه‌های هفتم تا نهم (از مجموعه آزمونهای مرکز خدمات آزمون آمریکا، به نقل گرانلاند، ۱۹۷۱)

محتوا / سطوح یادگیری	عدد و اعمال ریاضی	هندسه و اندازه گیری	رابطه‌ها، تابع‌ها، نمودارها	برهان، اثبات	احتمال و آمار	جملات ریاضی	مجموعه‌ها، سیستمهای ریاضی	موارد استفاده	تعداد کل سؤالها
۱. یادآوری اطلاعات واقعی	شمارهٔ سؤال ۳۸ ,۴۹						۳۱		۳
۲. انجام دادن اعمال ریاضی	۲ ,۲۴ ,۳۲ ,۳۶		۱۲			۸			۶
۳. حل کردن مسائل معمولی	۱ ,۲۵	۱۰ ,۳۰ ,۴۶						۴۷	۶
۴. درک اندیشه‌ها و مفاهیم ریاضی	۴ ,۷ ,۹ ,۱۱ ,۱۳ ,۲۷ ,۲۸	۱۷ ,۱۸ ,۲۱ ,۲۲ ,۳۴ ,۳۵ ,۳۹ ,۴۸ ,۵۰			۶ ,۱۹ ,۲۰	۱۵		۵ ,۱۴ ,۲۹	۲۳
۵. ابتکار و بینش	۲۳ ,۴۰ ,۴۵	۱۶ ,۲۶ ,۳۷ ,۴۴	۴۱						۸
۶. فرایندهای عالی‌تر ذهنی	۳	۴۳		۴۲			۳۳		۴
تعداد کل سؤالها	۱۹	۱۷	۲	۱	۳	۲	۲	۴	۵۰

شمارهٔ سؤالهای آزمون نوشته شده است. امتیاز این روش آن است که علاوه بر مشخص شدن تعداد سؤالها برای هریک از خانه‌های جدول، شمارهٔ سؤالها نیز مشخص می‌شوند.

نکته‌ای راکه در تهیهٔ جدول مشخصات آزمون باید درنظر گرفت این است که انواع هدفهای آموزشی معلم یا تمام بازده‌های یادگیری باید ملحوظ شوند. به سخن دیگر، اگر درس علاوه بر هدفهای شناختی، دارای هدفهای عاطفی و (یا) هدفهای روانی - حرکتی است، آنها نیز باید نوشته شوند. البته چنان که در فصلهای بعدی این کتاب خواهیم دید، برای سنجش هر دسته از هدفها روشهای خاصی وجود دارند که باید به تناسب نوع هدف از روشهای ویژهٔ آنها استفاده کرد.

علاوه بر دو نوع جدول مشخصاتی که توضیح دادیم، نوع دیگری جدول مشخصات وجود دارد که ویژهٔ ارزشیابی تکوینی است. ما ارزشیابی تکوینی را در فصل چهارم کتاب توضیح دادیم. در آنجا گفتیم که ابزار ارزشیابی تکوینی آزمونهای مختصری هستند که معلم در آغاز هر جلسهٔ درس و پیش از آموزش درس جدید از یادگیرندگان به عمل می‌آورد و هدف آن کسب اطمینان از تسلط یابی یادگیرندگان بر هدفهای درس گذشته است. بنابراین، هر معلمی، در طول یک ترم یا سال تحصیلی، به تعداد واحدها یا جلساتی که داشته است آزمون تکوینی تهیه و اجرا می‌کند که برای هریک از آنها نوشتن یک جدول مشخصات لازم است.

از آنجا که یک آزمون تکوینی دربرگیرندهٔ محتوا و هدفهای یک جلسهٔ درس است، جدول مشخصات آن به مفصلی جدول مشخصات آزمون تراکمی نیست. معمولاً در بعد محتوای جدول مشخصات آزمونهای تکوینی تنها یک موضوع واحد و در بعد هدف آنها فقط چند هدف نوشته می‌شوند. جدول ۸-۶ نمونه‌ای از این جدولهاست. چنان که می‌بینید، در بعد محتوای این جدول

جدول ۸ـ۶ جدول مشخصات یک آزمون تکوینی ۱۰ سؤالی، شامل یک موضوع درسی و ۵ هدف آموزشی

بُعد هدف / بعد محتوا	آزمون جورکردنی	
سطوح یادگیری	هدفهای یادگیری	تعداد سؤالها
	از دانشجو انتظار می‌رود پس از مطالعه ویادگیری مطالب این درس بتواند:	
دانش	۱. آزمون جورکردنی را تعریف کند.	۱
دانش	۲. انواع آزمونهای جورکردنی را توضیح دهد.	۲
دانش	۳. محاسن و معایب آزمونهای جورکردنی را شرح دهد.	۲
فهمیدن	۴. با مقایسهٔ آزمونهای جورکردنی و آزمونهای چندگزینه‌ای، تفاوتها و شباهتهای آنها را از لحاظ کاربردی توضیح دهد.	۲
کاربستن	۵. برای سنجش میزان آموخته‌های دانش‌آموزان یک درس یک آزمون جورکردنی تهیه کند.	۳
		۱۰

یک موضوع (آزمون جورکردنی) و در بعد هدف آن ۵ هدف نوشته شده است. در ضمن برای آزمونی که قرار است در این درس به کار رود ۱۰ سؤال پیش‌بینی شده.

زمانی که بخواهیم یک آزمون وابسته به ملاک بسازیم که توسط بیش از یک نفر مورد استفاده قرار بگیرد (از جمله آزمونهای مورد استفاده در سنجش یادگیری در حد تسلط) معمولاً از جدول مشخصات مفصل‌تری در مقایسه با آنچه برای تهیهٔ جدول مشخصات گفته شد استفاده می‌کنیم. این جدول مشخصات معمولاً توصیفی از سؤالهای آزمون همراه با نمونه سؤالاتی که ویژگیهای سؤالها را نشان می‌دهد تشکیل شده است.

انتخاب هدفهای مناسب برای گنجانیدن در آزمون

هدفهای آموزشی و بازده‌های یادگیری مربوط به موضوعهای مختلف و سطوح گوناگون آموزش بسیار متنوع‌اند، و بلوم و همکاران (۱۹۵۶) آنها را در طبقه‌بندی خود به سه حوزهٔ شناختی، عاطفی، و روانی ـ حرکتی طبقه‌بندی کرده‌اند. آزمونهای پیشرفت تحصیلی متداول که به آزمونهای کتبی یا آزمونهای مداد و کاغذی معروف‌اند به طور عمده با حوزهٔ شناختی که اجزای آن را در صفحات قبل معرفی کردیم سروکار دارند. هدفهای آموزشی و بازده‌های یادگیری متعلق به حوزه‌های غیرشناختی، از جمله نگرشها، علاقه‌ها، و ارج‌نهادن‌ها (هدفهای حوزهٔ عاطفی)، و مهارتهای مختلف، مثلاً انجام آزمایش در آزمایشگاه، کار در کارگاه، فعالیتهای ورزشی، و ایجاد ارتباط کلامی و غیرکلامی (هدفهای حوزهٔ روانی ـ حرکتی) به کمک وسایل دیگر سنجش از جمله پرسشنامه، مشاهده، مقیاس درجه‌بندی[1]، فهرست وارسی[2]، واقعه‌نگاری[3]، ثبت حضور و غیاب[4]، گزارشهای کار عملی، و مانند اینها سنجش می‌شوند. بنابراین، در درسهایی که هدفهای آموزشی و یادگیری متنوعی دارند باید همهٔ انواع هدفهای آموزشی را در نظر گرفت و آنها را در جدول مشخصات درس واردکرد و در صورت لزوم برای هر دسته از هدفهای حوزه‌های مختلف آزمون یا آزمونهای مجزایی مورد استفاده قرار داد. از آنجاکه متداول‌ترین آزمونهای مورد استفاده در آموزش و پرورش آزمونهای کتبی ویژهٔ حوزهٔ شناختی هستند، ما ابتدا این آزمونها را معرفی می‌کنیم و سپس به معرفی سایر انواع آزمونها می‌پردازیم.

خلاصهٔ فصل

ما در این فصل در جهت معرفی طرح ارزشیابی پیشرفت تحصیلی موضوعهای زیر را بحث کردیم:

۱. هدفهای آموزشی
۲. طبقه‌بندی سه‌گانهٔ حوزه‌های شناختی، عاطفی، و روانی ـ حرکتی هدفهای آموزشی

---

1. rating scale
2. checklist
3. anecdotal record
4. inventory

۳. جدول مشخصات و انواع آن
۴. دسته‌بندی انواع آزمونهای پیشرفت تحصیلی

جدول مشخصات که تهیهٔ آن مهم‌ترین مرحلهٔ تهیهٔ طرح ارزشیابی پیشرفت تحصیلی به حساب می‌آید یک جدول دوبعدی است که در یک بعد آن اجزای محتوای درس و در بعد دیگرش فهرست هدفهای درس قرار داده می‌شوند. در این جدول، علاوه بر اجزای محتوای درس و هدفهای آموزشی، تعداد یا درصد سؤالهایی که به هریک از هدفهای آموزشی و به هریک از بخشهای درس تعلق می‌گیرد نیز وارد می‌شود. در جدول مشخصات آزمونهای وابسته به هنجار، هدفهای آموزشی یا هدفهای یادگیری در قالب مفاهیم کلی وارد می‌شوند، در حالی که در جدول مشخصات آزمونهای وابسته به ملاک هدفهای دقیق آموزشی وارد می‌شوند.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. هدف و استاندارد در اصل بیان‌کنندهٔ یک مفهوم‌اند.

الف ـ صحیح

ب ـ غلط

۲. " دانش‌آموز بتواند کتابهای کتابخانهٔ شخصی خود را براساس نام نویسندهٔ آن طبقه‌بندی کند."

هدف فوق مُعرف کدام سطح یادگیری است؟

الف ـ به یاد آوردن دانش مفهومی

ب ـ به کاربستن دانش مفهومی

پ ـ به کار بستن دانش روندی

ت ـ تحلیل دانش مفهومی

ث ـ آفریدن دانش روندی

۳. اولین گام در تهیهٔ طرح آزمونهای پیشرفت تحصیلی چیست؟

الف ـ تهیهٔ جدول مشخصات آزمون

ب ـ تعیین نوع سؤالات آزمون

پ ـ تعیین نوع مقیاس اندازه‌گیری

ت ـ توزیع سؤالات در میان بخشهای مختلف درس

۴. " دانش‌آموز با استفاده از فرمولهای ریاضی بتواند مساحت باغچهٔ خانهٔ خود را که به شکل ذوزنقه است درست حساب کند".

هدف آموزشی بالا معرف کدام‌یک از طبقات زیر است؟

الف ـ دانش

ب ـ فهمیدن
پ ـ کاربستن
ت ـ تحلیل
ث ـ ترکیب
ج ـ ارزشیابی

۵. طبقه‌بندیهای جدید و قدیم حوزهٔ شناختی را به اختصار توضیح دهید.

۶. طبقه‌بندی حوزهٔ عاطفی را به اختصار توضیح دهد.

۷. برای هریک از طبقه‌های حوزه‌های شناختی و حوزهٔ عاطفی یک هدف نمونه بنویسید.

۸. فواید رسم جدول مشخصات در تهیهٔ طرح درس و در تهیهٔ آزمون را توضیح دهید.

۹. برای یکی از کتابهای درسی دورهٔ ابتدایی، راهنمایی، یا دبیرستان یک جدول مشخصات رسم کنید. در این جدول هدفهای آموزشی خود را با توجه به طبقه‌های شش‌گانهٔ طبقه‌بندی بلوم دسته‌بندی کنید. فرض کنید می‌خواهید یک آزمون ۲۰ سؤالی برای این فصل بنویسید. تعداد سؤالهایی را که به هر هدف آموزشی و به هر موضوع درسی اختصاص می‌دهید در خانه‌های داخل جدول بنویسید.

هدفهای زیر نمونه‌هایی از هدفهای آموزشی حوزه‌های سه گانهٔ یادگیری هستند. این هدفها را به دقت مطالعه کنید و تعیین نمایید که هرکدام از آنها به کدام یک از حوزه‌های یادگیری تعلق دارد.

هدف آموزشی	نام حوزهٔ مربوط به هدف را در اینجا بنویسید
۱۰. از دانشجوی درس روانشناسی تربیتی انتظار می‌رود در پایان دورهٔ آموزشی خود بتواند روشها و فنون مختلف آموزشی را در ارتباط با آموزش هدفهای مختلف مورد نقد و بررسی قرار دهد.	
۱۱. از دانش‌آموز درس حرفه و فن انتظار می‌رود بتواند با استفاده از وسایل مورد نیاز یک میز مطابق با الگویی‌که به او داده می‌شود بسازد.	

نام حوزهٔ مربوط به هدف را در اینجا بنویسید

هدف آموزشی

۱۲. از دانشجو انتظار می‌رود در بحثهای گروهی برای عقاید دیگران که با نظر او مخالف‌اند احترام قائل شود.

۱۳. از دانش‌آموز انتظار می‌رود تمام مسئله‌های ریاضی را که به او داده می‌شود درست حل کند.

۱۴. از دانش‌آموز انتظار می‌رود شعری را که در درس فارسی او آمده است به زبان ساده خلاصه کند.

۱۵. از دانش‌آموز انتظار می‌رود فاصلهٔ ۱۰۰۰ متر را، بدون افتادن از روی زین، دوچرخه‌سواری کند.



سنجش یادگیری به کمک آزمون صحیح ـ غلط

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمون صحیح ـ غلط را تعریف کند.

۲. ویژگیهای سؤالهای صحیح ـ غلط را توضیح دهد.

۳. انواع سؤالهای صحیح ـ غلط را توضیح دهد و برای هر یک مثالی بزند.

۴. چگونگی تهیهٔ سؤالهای صحیح ـ غلط را توضیح دهد.

۵. قواعد تهیهٔ سؤالهای صحیح ـ غلط را شرح دهد.

۶. امتیازها و محدودیتهای آزمونهای صحیح ـ غلط را توضیح دهد.

۷. موارد استفادهٔ آزمونهای صحیح ـ غلط را توضیح دهد.

۸. برای هدفهای یادگیری سطح پایین یک درس یک آزمون صحیح ـ غلط با رعایت کلیهٔ پیشنهادهای داده شده در این فصل تهیه نماید.

در فصل ۳ توضیح دادیم که یک دسته از روشهای شناخته شدهٔ سنجش روشهای عینی یا گزینش پاسخ نام دارند. روشهای عینی سنجش یادگیری از سه نوع صحیح ـ غلط، جورکردنی، و چندگزینه‌ای تشکیل می‌شوند. ما در این فصل نوع صحیح ـ غلط را توضیح می‌دهیم و دو فصل بعدی را به ترتیب به آزمونهای جورکردنی و چندگزینه‌ای اختصاص خواهیم داد.

تعریف و انواع آزمونهای صحیح ـ غلط

در آزمونهایی که صحیح ـ غلط[1] نام دارند هر سؤال از یک جمله تشکیل می‌شود که آزمون‌شونده (دانش‌آموز یا دانشجو) آن را به یکی از صورتهای صحیح ـ غلط، درست ـ نادرست، بله ـ نه، موافق ـ مخالف، واقعیت ـ عقیده، و مانند اینها پاسخ می‌دهد. اما از آنجا که این نوع سؤالها غالباً به صورتی نوشته می‌شوند که آزمون‌شونده یا پاسخ‌دهنده به صورت صحیح یا غلط به آنها جواب می‌دهد، به سؤالهای صحیح ـ غلط شهرت یافته‌اند. به نمونه‌های زیر توجه کنید:

نوع صحیح ـ غلط

مادهٔ سبز موجود در برگ گیاهان کلروفیل نام دارد.

صحیح (غلط)

نوع صحیح ـ غلط با اصلاح

راهنمایی: اگر گفتهٔ زیر درست است پاسخ "صحیح" را انتخاب کنید (دور آن خط بکشید)؛ اگر غلط است "غلط" را انتخاب کنید و برای کلمه‌ای که زیر آن خط کشیده شده کلمهٔ دیگری برگزینید که جمله را درست کند.

"استفاده از نیروی بخار موجب انقلابی در حمل و نقل در قرن هفدهم شد."

صحیح (غلط)

اصلاح: نوزدهم

(ایبل، ۱۹۷۲، ص ۱۰۳)

در واقع، این نوع سؤال ترکیبی از سؤال صحیح ـ غلط و سؤال کوته‌پاسخ است. قسمت اول که تصمیم‌گیری دربارهٔ درستی یا نادرستی سؤال است بخش صحیح ـ غلط و قسمت دوم که یافتن کلمهٔ درست برای کلمهٔ نادرست و نوشتن آن در جواب سؤال است بخش کوته‌پاسخ آن است (به فصل ۱۱ نگاه کنید).

نوع بله ـ نه

امکان دارد بدون یادگیری پیش نیازهای اساسی یک درس مطالب آن درس را یاد گرفت؟

بله (نه)

نوع بله ـ نه با توضیح

یک نظرسنجی از ۵۰۰ نفر عضو حزب دموکرات و ۵۰۰ نفر عضو حزب جمهوری‌خواه به عمل آمده و در آن

1. true-false

سؤال زیر پرسیده شده است: "حکومتی بهترین است که کمترین حکومت را می‌کند". نتیجهٔ این نظرخواهی به شرح زیر اعلام شده است:

	درصد موافق	درصد مخالف
مردان عضو حزب دموکرات	٪۱۲	٪۳۵
زنان حزب دموکرات	٪۳	٪۱۴
مردان حزب جمهوری خواه	٪۳۸	٪۱۲
زنان حزب جمهوری خواه	٪۲۸	٪۷

من معتقدم که این نظرسنجی ثابت می‌کند که اکثریت مردم طالب این هستند که حکومت‌ها کمتر حکومت کنند.

نظر من درست است؟ (بله) نه

اگر جواب شما "نه" است، توضیح دهید چرا من اشتباه می‌کنم.

(نیتکو، ۲۰۰۱، ص ۱۳۶)

نوع موافق ـ مخالف

راهنمایی: با کشیدن دایره‌ای به دور "موافق" یا "مخالف" نظر خود را نسبت به گفتهٔ زیر مشخص کنید.

"هر کسی آن درود عاقبت کار که کشت". (موافق) مخالف

نوع واقعیت ـ عقیده

بیان زیر یک حقیقت علمی است یا یک عقیدهٔ شخصی؟

"در بعضی سیارگانی که به دور ستارگان دوردست می‌چرخند نوعی زندگی هوشمندانه یافت می‌شود."

واقعیت علمی (عقیدهٔ شخصی)

(لین و گرانلاند، ۲۰۰۰، ص ۱۷۹)

نوع چند صحیح ـ غلط (نوع خوشه‌ای)

کشور ایران هم مرز است با کشور:

	صحیح	غلط
۱ ـ عراق	(صحیح)	غلط
۲ ـ قزاقستان	صحیح	(غلط)
۳ ـ افغانستان	(صحیح)	غلط
۴ ـ سوریه	صحیح	(غلط)

سؤالهای چند صحیح ـ غلط شبیه به سؤالهای چند گزینه‌ای هستند، با این تفاوت که در سؤالهای چند گزینه‌ای آزمون شونده از میان گزینه‌های پیشنهادی تنها یکی را برمی‌گزیند، اما در سؤالهای چند صحیح ـ غلط لازم است به همهٔ گزینه‌ها جواب بدهد. بنابراین، به هر یک

سؤال چندگزینه‌ای، صرف نظر از تعداد گزینه‌های آن، تنها ۱ نمره تعلق می‌گیرد، اما برای هر سؤال چند صحیح ـ غلط، به تعداد انتخابهای آن، نمره منظور می‌شود. به همین دلیل، در تنظیم سؤالهای یک آزمون به تنه یا صورت سؤالهای چند صحیح ـ غلط شماره داده نمی‌شود، اما همه گزینه‌ها شماره گذاری می‌شوند.

نکتهٔ دیگری که در رابطه با سؤالهای چند صحیح ـ غلط باید به یاد بسپارید این است که در هر مجموعه همهٔ انتخابهای صحیح یا غلط باید به موضوع واحدی مربوط باشند. همچنین امکان دارد که همهٔ انتخابهای یک مجموعه سؤال چند صحیح ـ غلط صحیح باشند یا غلط. به نمونهٔ دیگری از این نوع سؤالها توجه کنید.

مدافعان برده‌داری در کشور آمریکا پیش از جنگ داخلی با توسل به گفته‌های زیر از عقیدهٔ خود دفاع می‌کردند. هر یک از گفته‌ها را بررسی کنید و ببینید که با موازین دموکراسی همخوانی دارد یا نه؟

۱ ـ برده‌داری کار درستی است زیرا در تمام طول تاریخ وجود داشته است. صحیح (غلط)

۲ ـ برده‌داری کار درستی است زیرا مورد قبول کسانی بوده است که قانون اساسی آمریکا را نوشته‌اند. صحیح (غلط)

۳ ـ برده‌داری کار درستی است زیرا مورد قبول فیلسوف سرشناس یونان باستان، ارسطو، بوده است. صحیح (غلط)

(ایبل، ۱۹۷۲، ص ۱۷۴)

تهیهٔ سؤالهای صحیح ـ غلط

صاحب‌نظران سنجش در رابطه با تهیهٔ سؤالهای صحیح ـ غلط قواعد مختلفی را به دست داده‌اند. ما در اینجا قواعدی را که اگر رعایت شوند بر کیفیت سؤالهای صحیح ـ غلط می‌افزایند مورد بحث قرار می‌دهیم. اما پیش از معرفی این قواعد به پیشنهادهای رابرت ایبل (۱۹۷۲) یکی از طرفداران آزمونهای صحیح ـ غلط در این‌باره توجه کنید.

ایبل در رابطه با تهیهٔ سؤالهای صحیح ـ غلط گفته است ابتدا قسمتی از موضوع درسی را که قابل طرح در یک پاراگراف باشد انتخاب کنید. "این قسمت از موضوع یا این جزء دانش باید به سادگی قابل توصیف در یک پاراگراف واحد مانند آنچه در کتابهای درسیِ خوب یافت می‌شود باشد" (ص ۱۷۲). سپس دربارهٔ آن موضوع جمله یا گزاره‌ای با ویژگیهای زیر بنویسید.

۱. حاوی اندیشهٔ مهمی باشد.

۲. به وسیلهٔ اشخاص صاحب‌نظر قابل دفاع باشد.

پاراگراف
گرفت یا گرفتگی زمانی رخ می‌دهد که یک جرم نجومی مانع تابش بخشی از یا تمامی نور جرم نجومی دیگری می‌شود. در خورشید گرفتگی، ماه در فاصلهٔ بین خورشید و زمین قرار می‌گیرد و مانع رسیدن نور خورشید به زمین می‌شود. در ماه گرفتگی، سایهٔ زمین بر روی ماه می‌افتد و مانع بازتابش نور خورشید بر سطح کرهٔ ماه می‌شود.

گزاره
گرفتگی (مثلاً خورشید گرفتگی یا ماه گرفتگی) در منظومهٔ شمسی بر اثر افتادن سایهٔ یک جرم نجومی بر جرم نجومی دیگر حاصل می‌شود.

شکل ۱ـ۷ یک پاراگراف مطلب و گزارهٔ استخراج شده از آن (ایبل، ۱۹۷۲، ص ۱۷۳)

۳. توسط کسانی که از موضوع مطرح شده اطلاع کافی ندارند قابل جواب دادن نباشد؛ یعنی براساس اطلاعات عمومی نتوان به آن پاسخ درست داد.

در شکل ۱ـ۷ به یک نمونه پاراگراف و جمله یا گزارهٔ حاصل از آن توجه کنید.

ایبل (۱۹۷۲) می‌گوید سؤال‌نویسان باتجربه به ندرت گزاره‌ای را که سؤال از آن استخراج می‌شود عملاً می‌نویسند، بلکه به هنگام نوشتن سؤال آن را در ذهن خود دارند. با این حال، وی توصیه می‌کند هنگامی که در نوشتن یک سؤال صحیح ـ غلط به مشکل برمی‌خورید از خود بپرسید که "این سؤال قرار است چه چیزی را بسنجد؟" یا "گزاره‌ای که براساس آن من می‌خواهم سؤالم را بنویسم چیست؟"

گام بعدی تبدیل گزارهٔ حاصل از پاراگراف به سؤال صحیح ـ غلط است. برای این منظور، ۵ پیشنهاد ارائه شده است. این پیشنهادها همراه با نمونه سؤالهای حاصل از پاراگراف و گزارهٔ شکل ۱ـ۷ در زیر توضیح داده می‌شوند.

پیشنهاد ۱ ـ اندیشهٔ اصلی را در قالب کلمات دیگری تغییر دهید و صورت صحیح یا غلط سؤال را بنویسید.

صورت صحیح
وقتی که یک جرم نجومی مانند ماه مانع تابش قسمتی از نور ستاره‌ای مانند خورشید بر سطح سیاره‌ای مانند زمین می‌شود، می‌گویند گرفتی رخ داده است.

صورت غلط
وقتی که نور ستاره‌ای مانند خورشید از سطح سیاره‌ای مانند زمین به سطح جرم دیگری مانند ماه منعکس می‌شود، می‌گویند گرفتی رخ داده است.

پیشنهاد ۲ ـ تنها بخشی از اندیشهٔ اصلی را به صورت صحیح یا غلط بنویسید نه تمام آن را.

صورت صحیح	صورت غلط
همهٔ گرفتها از سایه تشکیل می‌شوند.	اگر پرتوهای نور قابل فهم شدن نباشند، گرفتها رخ نمی‌دهند.

پیشنهاد ۳ ـ اندیشهٔ اصلی را به اندیشهٔ دیگری تغییر دهید.

صورت صحیح	صورت غلط
ماه گرفتگی تنها زمانی رخ می‌دهد که ماه قرص کامل باشد.	ماه گرفتگی تنها زمانی رخ می‌دهد که ماه نو باشد.

پیشنهاد ۴ ـ پیامدهای اندیشهٔ اصلی را مشخص کنید.

صورت صحیح	صورت غلط
پیش‌بینی گرفتها مستلزم داشتن اطلاع دربارهٔ حرکات مداری اجرام موردنظر است.	پیش‌بینی گرفتها مستلزم داشتن اطلاع دربارهٔ شیب محور زمین است.

پیشنهاد ۵ ـ اثر موقعیتهای متفاوت (حتی غیرممکن) را استنتاج کنید.

صورت صحیح	صورت غلط
اگر جرمهای نجومیِ درخشان مانند ماه وجود نداشتند گرفتها قابل مشاهده نبودند.	در حضور ستارگان درخشان گرفتها رخ نمی‌دهند.

نمونه سؤالهای بالا نشان می‌دهند که از یک اندیشهٔ واحد می‌توان سؤالهای صحیح ـ غلط متنوعی درست کرد. البته سؤالهای بالا تنها سؤالهایی نیستند که از پاراگراف فوق دربارهٔ خسوف یا کسوف می‌توان نوشت. مسلماً سؤالهای بیشتری نیز می‌توان دربارهٔ آن طرح کرد. چیزی که نباید فراموش شود این است که "ماده‌های صحیح ـ غلط صرفاً به آزمودن یادآوری مطالب خلاصه نمی‌شوند، بلکه می‌توانند مسائل تازه را در اختیار آزمون شونده قرار دهند و او را وادار به تفکر و حل مسئله نمایند" (ایبل، ۱۹۷۲، ص ۱۷۵).

دربارهٔ کیفیت صورت درست و صورت غلط سؤالهای صحیح ـ غلط، ایبل (۱۹۷۲) اظهار می‌دارد که آزمون ساز باید بتواند دربارهٔ صورت صحیح سؤال بگوید: "درستی این بیان بر همه کس آشکار نیست، اما من می‌توانم شواهدی دال بر درستی آن به دست دهم." همچنین دربارهٔ صورت غلط سؤال باید بتواند بگوید: "این بیان یک غلط آشکار به نظر می‌رسد، اما من می‌توانم شواهدی دال بر اینکه آن واقعاً غلط است به دست دهم. اگر سؤالهای شما این آزمون را با موفقیت بگذرانند احتمالاً سؤالهای صحیح ـ غلط خوبی هستند" (ص ۱۷۵).

قواعد تهیهٔ سؤالهای صحیح ـ غلط

الزون بر راهنماییهای بالا، برای تهیهٔ سؤالهای صحیح ـ غلط، از سوی متخصصان اندازه‌گیری و سنجش پیشرفت تحصیلی، قواعدی پیشنهاد شده است. ما در اینجا به توضیح آنها می‌پردازیم.

۱. موضوعهای مهم و در ارتباط با هدفهای آموزشی را در سؤالهای خود بیاورید. سعی کنید هر یک از سؤالهای شما یک موضوع مهم مربوط به یک هدف آموزشی را بسنجد. هرچند که این پیشنهاد مخصوص سؤالهای صحیح ـ غلط نیست و به همهٔ انواع سؤالهای پیشرفت تحصیلی مربوط می‌شود، با این حال هنگام نوشتن سؤالهای صحیح ـ غلط بیشتر به آنها توجه کنید، زیرا غالباً در این گونه سؤالهاست که نکات جزئی و کم‌اهمیت امکان طرح شدن می‌یابند. بهترین وسیلهٔ رعایت این پیشنهاد استفاده از جدول مشخصات است که ما آن را در فصل پیش (فصل ۶) توضیح دادیم.

نمونهٔ ضعیف:

بعضی چیزها در بعضی چیزهای دیگر حل می‌شوند. (صحیح) غلط

نمونهٔ بهتر:

در یک لیتر آب گرم نمک بیشتری حل می‌شود تا در یک لیتر آب سرد. (صحیح) غلط

(ایبل، ۱۹۷۲، ص ۱۷۶)

سؤال ضعیف بالا یک بیان کلی است و مُعرف هیچ نکتهٔ مهمی نیست. در عوض، نمونهٔ سؤالِ بهتر فهم یک رابطهٔ مهم را می‌سنجد.

۲. سؤالهایی طرح کنید که درک و فهم را بسنجند نه یادآوری کلمه به کلمهٔ مطالب را. تکرار مطالب کتاب در سؤالهای امتحانی تنها محفوظات دانش‌آموزان را می‌سنجد. برای سنجش درک و فهم باید مطالب را بازسازی کنید و آنها را در قالبهای دیگری بجز آنچه در کتاب آمده است بنویسید.

نمونهٔ ضعیف:

در یک مثلث قائم‌الزاویه مربع وتر برابر است با مجموع مربعهای دو ضلع دیگر. (صحیح) غلط

نمونهٔ بهتر:

اگر وتر یک مثلث قائم‌الزاویهٔ متساوی‌الساقین ۷ سانتیمتر است، هر یک از دو ضلع دیگر آن باید از ۵ سانتیمتر بزرگ‌تر باشد. صحیح (غلط)

(ایبل، ۱۹۷۲، ص ۱۷۷)

سؤال ضعیف بالا در برگیرندهٔ یک جملهٔ آشنای کتابهای ریاضی است و لذا صرفاً اطلاعات حفظی دانش‌آموزان را می‌سنجد. در مقابل، نمونهٔ سؤال بهتر که از راه بازنویسی همان موضوع ولی با بیان دیگری درست شده می‌تواند درک و فهم دانش‌آموز را از موضوع مورد نظر بسنجد. در زیر به نمونه‌های دیگری از همین نوع سؤالها توجه کنید:

نمونهٔ ضعیف:

برای هر عملی یک عکس‌العمل برابر ولی مخالف با آن وجود دارد.

صحیح غلط

نمونهٔ بهتر:

وقتی که دستی دری را با نیروی معینی می‌فشارد، در با همان نیرو به دست فشار وارد می‌آورد.

صحیح غلط

(ایبل، ۱۹۷۲، ص ۱۷۶)

۳. **سؤالهایی طرح کنید که درست یا غلط بودن آنها قابل دفاع باشد.** هر سؤال باید به گونه‌ای نوشته شود که کسانی که از موضوع آن اطلاع کافی دارند به طور آشکار آن را درست یا غلط بدانند. به سخن دیگر، سؤال باید برای پاسخ دهندهٔ آگاه از موضوع سؤال کاملاً درست یا کاملاً غلط باشد.

نمونهٔ ضعیف:

ستارگان از خود نورهایی تولید می‌کنند که سوسو می‌زند.

صحیح غلط

نمونهٔ بهتر:

سوسو زدن نور ستارگان به علت حرکت جو زمین است.

صحیح غلط

(ایبل، ۱۹۷۲، ص ۱۷۷)

یک شخص مطلع از موضوع، می‌تواند سؤال ضعیف بالا را با توجه به دلایل زیر مورد تردید قرار دهد: نور منتشره از ستاره سوسو نمی‌زند. نور ستاره نسبتاً ثابت است، اما به علت تغییرات موجود در جو زمین نوری که از ستارگان به چشمان ما می‌رسد سوسو می‌زند. "سؤال ضعیف و غیرقابل دفاع بالا یا ناشی از دانش اندک نویسندهٔ سؤال یا به سبب بی‌دقتی او در نوشتن سؤال است" (ایبل، ۱۹۷۲، ص ۱۷۸).

۴. **از جمله‌های پیچیده و چند قسمتی استفاده نکنید.** جمله‌های پیچیده و طولانی حواس آزمون شوندگان را پرت می‌کنند و باعث می‌شوند که توجه او به جای نکتهٔ اصلی سؤال به مسائل فرعی جلب شود.

نمونهٔ ضعیف:

جین آستین، داستان‌سرای آمریکایی، متولد ۱۷۹۰، یک نویسندهٔ پرکار بود و شهرتش به خاطر داستان معروف او غرور و تعصب است که در سال ۱۸۲۰ انتشار یافت.

صحیح (غلط)

نمونهٔ بهتر:

شهرت جین آستین به خاطر داستان غرور و تعصب اوست.

صحیح (غلط)

(پین، ۲۰۰۳، ص ۱۹۴)

۵. **در هر سؤال بیشتر از یک موضوع قرار ندهید.** هر سؤال باید تنها یک موضوع مهم را شامل شود. قرار دادن بیش از یک موضوع در یک سؤال، به ویژه اگر یکی درست و دیگری غلط باشد، وسیلهٔ مناسبی برای سنجش یادگیری نیست. پوفام (۲۰۰۲) در این باره گفته است "حضور دو مفهوم در یک سؤال، حتی اگر هر دو صحیح یا هر دو غلط باشند، باعث سردرگمی دانش‌آموز می‌شود، و نتایجی به دست می‌دهد که منجر به استنباطهای غلطی دربارهٔ دانش‌آموزان خواهد شد" (ص ۱۳۱). نول[1] و اسکانل[2] (۱۹۷۲، ص ۲۱۴)، برای این مورد مثالهای زیر را آورده‌اند:

آب و هوای انگلستان، به سبب نفوذ گلف استریم اقیانوس اطلس، از بیشتر کشورهای هم‌عرض جغرافیایی آن سردتر است.

صحیح غلط

بیان فوق مبهم است، زیرا دو موضوع را دربر می‌گیرد که یکی از آنها درست و دیگری غلط است. بهتر است آن را به دو سؤال مستقل مانند نمونه‌های زیر تبدیل کنیم:

۱ـ آب و هوای انگلستان از بیشتر کشورهای هم‌عرض جغرافیایی آن سردتر است. صحیح (غلط)

۲ـ آب و هوای انگلستان تحت تاثیر گلف‌استریم اقیانوس اطلس قرار دارد. (صحیح) غلط

نیتکو (۲۰۰۱، ص ۱۴۱) نیز در ارتباط با این سفارش که در سؤالهای صحیح ـ غلط بیشتر از یک موضوع قرار ندهید مثالهای زیر را ذکر کرده است.

نمونهٔ ضعیف:

رود مونونگاهلا به سوی شمال جریان می‌یابد و در کولومبوس به رود الگهنی می‌پیوندد و در آنجا با همدیگر رود اوهایو را به وجود می‌آورند.

صحیح (غلط)

نمونهٔ بهتر:

رود مونونگاهلا و رود الگهنی به هم می‌پیوندند و رود اوهایو را به وجود می‌آورند. صحیح (غلط)

1. Noll 2. Scannell

نیتکو (۲۰۰۱) در رابطه با سؤال ضعیف بالا می‌گوید دانش‌آموز ممکن است به یکی از دلایل موجه زیر برای این سؤال جواب غلط را انتخاب کند: (۱) ممکن است (اشتباهاً) فکر کند که رود مونونگاهلا به سوی شمال جریان ندارد. (۲) ممکن است نداند که به هم پیوستن رودخانه‌ها در پیتسبورگ اتفاق می‌افتد. (۳) ممکن است چیزی دربارهٔ این سه رود نداند. بنابراین، دانش‌آموز به سؤال درست جواب می‌دهد بدون اینکه اطلاعات لازم برای درست جواب دادن را داشته باشد.

پس، بهتر است از گنجانیدن بیشتر از یک موضوع در هر سؤال پرهیز شود. حتی اگر دو موضوع هر دو صحیح یا هر دو غلط باشند نیز دشواری به وجود می‌آورد، زیرا اگر یک آزمون شونده از هیچ یک از دو موضوع آگاه نباشد و آزمون شوندهٔ دیگری از یکی از آندو مطلع باشد، نتیجهٔ سؤال برای هر دو نفر یکی خواهد بود. تنها استثناء برای این قاعده زمانی است که بخواهید رابطهٔ علت و معلولی بین دو موضوع را بسنجید.

۶. **در صورت امکان از زبان کمّی و دقیق استفاده کنید نه از زبان کیفی و غیردقیق.** کلماتی مانند اندکی، بسیاری، کوچک، بزرگ، و مهم مبهم و قابل تفسیرند مگر اینکه با یک معیار مقایسه به کار بروند.

نمونهٔ ضعیف:

در انتخابات ریاست جمهوری سال ۱۹۹۲، بسیاری از مردم به بیل‌کلینتون رای دادند.

نمونهٔ بهتر:

در انتخابات ریاست جمهوری سال ۱۹۹۲، بیشتر از ۶۰ درصد رای‌دهندگان به بیل‌کلینتون رای دادند.

(پین، ۲۰۰۳، ص ۱۹۲)

۷. **تا آنجا که ممکن است از طرح سؤالهای منفی پرهیز کنید.** جواب دادن به سؤالهای منفی معمولاً دشوار است. پوفام (۲۰۰۲) می‌گوید فرض کنید در یک سؤال صحیح ـ غلط از شما خواسته شده که دربارهٔ درست یا غلط بودن بیان زیر تصمیم بگیرید:

جامعهٔ ملل بلافاصله پس از پایان یافتن جنگ جهانی دوم تشکیل نشد. (صحیح) غلط

جواب درست این سؤال صحیح است، زیرا جامعهٔ ملل پیش از جنگ جهانی دوم تشکیل شده بود. اما وجود نشد در سؤال بعضی دانش‌آموزان را دچار سردرگمی می‌کند. حتی اگر بدانند که جامعهٔ ملل پیش از شروع جنگ جهانی دوم وجود داشته، احتمالاً به سؤال جواب غلط خواهند داد. بهتر است سؤال را به یکی از صورتهای زیر درآورید:

۱ ـ *جامعهٔ جهانی پیش از شروع جنگ جهانی دوم تشکیل شده بود.* (صحیح) غلط

۲ ـ جامعهٔ جهانی بلافاصله پس از پایان جنگ جهانی دوم تشکیل شد. صحیح (غلط)

با توجه به توضیحات بالا، تا آنجا که ممکن است در نوشتن سؤالهای صحیح ـ غلط از کاربرد عبارتهای منفی پرهیز کنید. استفاده از منفی مضاعف به هیچ وجه جایز نیست.

۸. از کلمات راهنما یا اشاره کننده به جواب درست استفاده نکنید. بعضی وقتها کاربرد کلمه‌هایی مانند ممکن است، همیشه، همه، هرگز و هیچ صحیح یا غلط بودن یک سؤال را کاملاً آشکار می‌سازد، به گونه‌ای که پاسخ‌دهندهٔ ناآگاه از موضوع مورد آزمون به سؤال جواب درست می‌دهد. به نمونه‌های زیر نگاه کنید:

۱ ـ همهٔ شهرهای بزرگ با قطار به هم مرتبط‌اند. صحیح (غلط)

۲ ـ هیچ نظام آموزشگاهی با کمکهای مالی حمایت نمی‌شود. صحیح (غلط)

۳ ـ امکان دارد که هر زاویه‌ای را دو نیم کرد. (صحیح) غلط

۴ ـ غیرممکن است که یک کیلومتر را در ۲ دقیقه دوید. صحیح (غلط)

۵ ـ انقلابها همیشه به هدفهای مطلوب اجتماعی انجامیده‌اند. صحیح (غلط)

۶ ـ در یک دموکراسی جنگ هرگز توجیه نمی‌شود. صحیح (غلط)

(ساکس، ۱۹۹۷، ص ۸۷)

استفاده از کلمه‌های همه، هیچ، غیرممکن، همیشه، و هرگز، در سؤالهای ۱، ۲، ۴، ۵، و ۶ بالا غلط بودن آنها را آشکار می‌سازد. در مقابل، آزمون شونده سؤال ۳ را نیز به راحتی صحیح تشخیص می‌دهد، زیرا کمتر فعالیتی غیرممکن است. بنا به گفتهٔ ساکس (۱۹۹۷)، "در حالی که استفاده از کلمه‌های راهنما ممکن است دانش‌آموزان را تشویق کند تا به صورت آزمون دهندگان مجرب درآیند، اما کمک زیادی به سنجش هدفهای درس نمی‌کند" (ص ۸۷). در زیر به نمونه‌های دیگری از این سؤالها توجه کنید:

نمونهٔ ضعیف:

یک "برد مادر[1]" معیوب ممکن است باعث خاموش شدن کامپیوتر شخصی بشود. (صحیح) غلط

نمونهٔ بهتر:

یک "برد مادر" معیوب باعث خاموش شدن کامپیوتر شخصی می‌شود. (صحیح) غلط

(پین، ۲۰۰۳، ص ۱۹۱)

پین (۲۰۰۳) می‌گوید آزمون‌شوندگان مجرب با توجه به کلمه‌های "ممکن است" به راحتی

1. mother board

پاسخ سؤال ضعیف بالا را پیدا خواهند کرد. البته اگر در طول آزمون به تساوی از این گونه کلمات در سؤالهای صحیح و غلط استفاده شود احتمال حدس زدن جواب درست سؤال با توجه به آنها کاهش خواهد یافت.

۹. **سعی کنید تعداد سؤالهای غلط بیشتر از تعداد سؤالهای درست باشد.** ایبل (۱۹۷۲) در توجیه این پیشنهاد می‌گوید دانش‌آموزان وقتی که در انتخاب بین جواب صحیح و جواب غلط تردید می‌کنند ترجیح می‌دهند جواب صحیح را انتخاب کنند. "چندین پژوهشگر دریافته‌اند که بیانات غلط تا حدی بهتر از بیانات صحیح بین دانش‌آموزان با پیشرفت بالا و دانش‌آموزان با پیشرفت پایین تمیز قائل می‌شوند" (ص ۱۸۰). ایبل برای سؤالهای غلط رقم ۶۷ درصد و برای سؤالهای درست رقم ۳۳ درصد را پیشنهاد داده است.

آمایهٔ پاسخ[1] نام دارم بنابه تعریف، آمایهٔ پاسخ به گرایش آزمون شونده در دادن پاسخ به سؤالها با توجه به شکل سؤال نه محتوای آن گفته می‌شود. منطق حکم می‌کند که اگر تعداد سؤالهای صحیح و تعداد سؤالهای غلط در یک آزمون صحیح ـ غلط مساوی باشند از مشکل آمایهٔ پاسخ جلوگیری شود، "اما شواهدی موجودند که نشان می‌دهند سؤالهای غلط قدرت تمیز بیشتری دارند و شاید نسبت ۶۰ به ۴۰ به نفع سؤالهای غلط بهترین پیشنهاد در این‌باره باشد" (پین، ۲۰۰۳، ص ۱۹۱).

۱۰. **طول سؤالهای صحیح و سؤالهای غلط را تقریباً هم‌اندازه انتخاب کنید.** کسانی که در نوشتن سؤالهای صحیح ـ غلط مهارت کافی ندارند، غالباً سؤالهای صحیح را از سؤالهای غلط طولانی‌تر می‌نویسند. این کار باعث می‌شود که دانش‌آموزان متوجه مطلب بشوند و بدون اینکه از موضوع طرح شده در سؤال اطلاع داشته باشند به آن درست پاسخ دهند. بنابراین، سعی کنید سؤالهای صحیح و سؤالهای غلط را هم‌اندازه بنویسید یا اگر به ناچار طول یکی از دیگری بیشتر می‌شود طول آنها را به تناوب و به یک نسبت کم و زیاد کنید.

امتیازها و محدودیتهای آزمونهای صحیح ـ غلط

بر سر استفاده از آزمونهای صحیح ـ غلط برای سنجش میزان یادگیری دانش‌آموزان و دانشجویان در میان متخصصان سنجش اختلاف‌نظر وجود دارد. برخی استفاده از آن را مفید می‌دانند. از جملهٔ کسانی که این آزمونها را مفید تشخیص داده‌اند، ایبل (۱۹۷۲) است. او، در دفاع از این آزمونها، گفته است "دلیل عمدهٔ استفاده از سؤالهای صحیح ـ غلط این است که آنها وسیلهٔ ساده و مستقیمی برای اندازه‌گیری بازده‌های اساسی آموزش رسمی هستند" (ص ۱۶۴). پین (۲۰۰۳) نیز در این باره می‌گوید:

1. response set

سؤال صحیح ـ غلط یک شیوهٔ بالقوه ارزشمند برای جمع‌آوریِ اطلاعات است.... استفاده از سؤالهای صحیح ـ غلط دارای امتیازهایی است که مهم‌ترین آنها کارآمدی است. معلم می‌تواند تعداد زیادی از این سؤالها را در یک جلسهٔ امتحانی مورد استفاده قرار دهد. این کار امکان نمونه‌گیری از دامنهٔ بزرگی از مطالب درسی را به منظور سنجش دانش یادگیرنده فراهم می‌آورد. (ص ۱۹۰)

به طور کلی، می‌توان امتیازها یا محاسن سؤالهای صحیح ـ غلط را به صورت زیر خلاصه کرد:

۱. تهیهٔ این نوع سؤالها کار نسبتاً آسانی است.
۲. تصحیح جواب آزمون‌شوندگان به این نوع سؤالها به سادگی و سرعت و با عینیت انجام‌پذیر است.
۳. به این سبب که می‌توان تعداد زیادی سؤال صحیح ـ غلط را در یک جلسهٔ امتحان مورد استفاده قرار داد، بخش گسترده‌ای از محتوای درس را می‌توان با این سؤالها سنجش کرد.

مخالفان آزمونهای صحیح ـ غلط نیز ایرادهایی به آن وارد آورده‌اند. از جمله هارپر و هارپر[1] (۱۹۹۰) و لین و گرانلاند (۲۰۰۰) گفته‌اند، به خلاف تصور مدافعان آزمونهای صحیح ـ غلط، تهیهٔ این آزمونها کار آسانی نیست. "ساختن سؤالهای صحیح ـ غلط که بازده‌های مهم یادگیری را بسنجند نیازمند مهارت فراوان است" (لین و گرانلاند، ۲۰۰۰، ص ۱۸۱). هارپر و هارپر (۱۹۹۰) نیز گفته‌اند، "تجارب دردناک نشان داده‌اند که سؤالهای صحیح ـ غلط به آن سادگی که به نظر می‌رسد نیستند و تهیهٔ آنها غالباً به مشکل می‌انجامد" (ص ۲۱۳).

در ردّ امتیاز دیگرِ عنوان شده برای سؤالهای صحیح ـ غلط، یعنی اینکه چون در یک جلسهٔ امتحانی تعداد زیادی از آنها را می‌توان مورد استفاده قرار داد پس حجم بزرگی از محتوای درس را به وسیلهٔ این سؤالها می‌توان در یک جلسهٔ امتحانی سنجش کرد، نیز نظرهایی داده شده است. از جمله گفته شده که همهٔ موضوعهای درسی را نمی‌توان به صورت عبارتهای صحیح یا غلط درآورد. بنا به گفتهٔ لین و گرانلاند (۲۰۰۰)، "در همهٔ موضوعهای درسی، از جمله علوم‌اجتماعی، زمینه‌هایی وجود دارند که دربارهٔ آنها نوشتن بیانات مطلقاً درست یا مطلقاً غلط که قابل تبدیل به سؤالهای صحیح ـ غلط باشند امکان‌پذیر نیست. تنها موضوعهای کم اهمیت را می‌توان به این گونه بیانات تبدیل کرد" (ص ۱۲۸).

ایراد دیگری که به سؤالهای صحیح ـ غلط گرفته‌اند میزان بالای حدس‌پذیری آنهاست. به طور متوسط، نصف همهٔ آزمون‌شوندگان می‌توانند تنها با حدس زدن به یک سؤال صحیح ـ غلط درست جواب دهند. یا یک آزمون‌شونده صرفاً با حدس زدن می‌تواند ۵۰ درصد سؤالهای صحیح ـ غلط یک آزمون را درست جواب دهد. حتی آزمون شوندگان بی‌اطلاعِ خوش اقبال ممکن است بیشتر از نصف سؤالها را درست جواب دهند.

هارپر و هارپر (۱۹۹۰) در ارتباط با امکان حدس‌پذیری بالای آزمونهای صحیح ـ غلط

1. Harper

گفته‌اند آمار سادهٔ نظریهٔ احتمال نشان می‌دهد که از هر ۱۰۰ آزمون‌شونده یک نفر خوش اقبال می‌تواند به گونهٔ زیر کسب نمره نماید:

از هر ۱۰ سؤال به ۸ سؤال درست جواب دهد. (٪۸۰)
از هر ۵۰ سؤال به ۳۳ سؤال درست جواب دهد. (٪۶۶)
از هر ۱۰۰ سؤال به ۶۲ سؤال درست جواب دهد. (٪۶۳)
از هر ۲۰۰ سؤال به ۱۱۶ سؤال درست جواب دهد. (٪۵۸)

افزون بر ایرادهای بالا، عیب بزرگ آزمونهای صحیح ـ غلط این است که برای سنجش هدفهای آموزشی و بازده‌های مهم یادگیری محدودیت دارند. لین و گرانلاند (۲۰۰۰) در این‌باره گفته‌اند:

سؤالهای صحیح ـ غلط برای سنجش بازده‌های یادگیری فراتر از دانش چندان مفید نیستند. تنها استثناهای این قاعده تمیز دادن بین واقعیت از عقیده و تشخیص رابطهٔ علت و معلولی است. این دو بازده یادگیری احتمالاً مهم‌ترین بازده‌هایی هستند که با این نوع سؤالها سنجش‌پذیرند. بسیاری از بازده‌های دیگر یادگیری را می‌توان با گونه‌های دیگر آزمونهای انتخاب پاسخ، به ویژه نوع چندگزینه‌ای، به نحو مؤثرتری اندازه‌گیری کرد. (ص ۱۸۲)

پین (۲۰۰۳) نیز دربارهٔ استفاده از سؤالهای صحیح ـ غلط گفته است استفاده از این سؤالها در موقعیتهایی مناسب است که در آنها تنها دو پاسخ وجود دارند. وی به عنوان نمونه سؤال زیر را مثال زده است.

درهای خروجی اضطراری مدرسه باید به سمت درون باز شوند. صحیح (غلط)

(پین، ۲۰۰۳، ص ۱۹۰)

لین و گرانلاند (۲۰۰۰) استفاده از سؤالهای صحیح ـ غلط را به دو موقعیت محدود کرده‌اند. یکی موقعیتی که در آن تنها دو شِق چپ ـ راست، بیشتر ـ کمتر، بالا ـ پایین، و مانند اینها وجود دارند. موقعیت دیگر زمانی است که آزمون‌شونده باید بین عقیده از واقعیت، علت از معلول، خرافه از علم، مربوط از نامربوط، معتبر از غیرمعتبر، و مانند اینها تمیز قائل شود.

☆ ☆ ☆ ☆ ☆

با توجه به توضیحات بالا، لازم است کاربرد سؤالهای صحیح ـ غلط به حداقل کاهش یابد، و عمدتاً برای سنجش هدفهای یادگیری سطوح پایین از آنها استفاده شود. علاوه بر آزمونهای صحیح ـ غلط، آزمونها و ابزارهای متنوع دیگری وجود دارند که به نحو شایسته‌تری هدفهای سطح بالای یادگیری را می‌سنجند. ما در فصلهای بعدی این کتاب به معرفی آنها خواهیم پرداخت.

موارد استفادهٔ آزمونهای صحیح ـ غلط

برچند که ما کمترین استفاده را از آزمونهای صحیح ـ غلط به معلمان سفارش دادیم؛ با این حال، دقت در تهیهٔ این آزمونها می‌تواند دامنهٔ استفاده از آنها را گسترش دهد. ما در این قسمت نمونه‌هایی از کاربرد آزمونهای صحیح ـ غلط را که از سوی متخصصان برای سنجش هدفهای گوناگون آموزشی تهیه شده‌اند معرفی می‌کنیم.

یادآوری دانش امور واقعی

۱ ـ میانه بهترین شاخص گرایش مرکزی است.

صحیح (غلط)

جنگ جهانی اول:

۲ ـ به تأسیس جامعهٔ ملل انجامید. (صحیح) غلط

۳ ـ منجر به تقسیم آلمان به دو قسمت شد. صحیح (غلط)

۴ ـ به تأسیس دولت مستقل اسرائیل انجامید. صحیح (غلط)

۵ ـ تنها جنگی است که در آن از گازهای سمی به مقیاس وسیعی استفاده شد. (صحیح) غلط

یادآوری دانش مفهومی

۶ ـ می‌توان به کمک تقویت منفی رفتار مطلوب یک دانش‌آموز را افزایش داد. (صحیح) غلط

یادآوری دانش روندی

۷ ـ برای محاسبهٔ ضریب همبستگی یک سؤال با کل آزمون از روش همبستگی دونقطه‌ای رشته‌ای استفاده می‌کنیم. (صحیح) غلط

یادآوری دانش فراشناختی

۸ ـ یکی از روشهای موفق یادگیری کتابهای درسی علمی این است که ابتدا با یک مرور مقدماتی با موضوعهای مختلف یک آشنایی کلی کسب کنیم و بعد برای یادگیری کامل آنها یک برنامهٔ زمان‌بندی تهیه نماییم. (صحیح) غلط

فهمیدن دانش مفهومی

۹ ـ با افزودن تعدادی سؤال مشابه سؤالهای یک آزمون به آن آزمون هم روایی و هم پایایی آزمون بالا می‌رود.

صحیح (غلط)

۱۰ - تفاوت بین نمره‌های خام منطبق با صدکهای چهل و پنجم و پنجاه و پنجم از تفاوت بین نمره‌های خام منطبق با صدکهای پنجم و پانزدهم کمتر است. صحیح (غلط)

فهمیدن دانش روندی

۱۱ - یکی از راههای خوب استخراج فرضیه‌های علمی مراجعه به نظریه‌های موجود در زمینهٔ مورد پژوهش است.

(صحیح) غلط

۱۲ - در استفاده از ضریب همبستگی در تحلیل داده‌های پژوهشی، هم می‌توان رابطهٔ زمانی بین متغیرها و هم رابطهٔ علت و معلولی را تعیین کرد.

صحیح (غلط)

کاربستن دانش امور واقعی

۱۳ - اگر در یک اتاق دربسته در یک یخچال روشن را باز بگذاریم درجهٔ حرارت اتاق کاهش خواهد یافت.

صحیح (غلط)

کاربستن دانش روندی

برای نمره‌های ۸، ۹، ۱۰، ۱۱، و ۱۲ ———

۱۴ - میانگین برابر است با ۱۰ (صحیح) غلط

۱۵ - میانه برابر است با ۱۰ (صحیح) غلط

۱۶ - نما برابر است با ۱۰ صحیح (غلط)

۱۷ - دامنه برابر است با ۸ صحیح (غلط)

۱۸ - با در دست داشتن فرمول $K=\frac{PV}{T}$، اگر $K=2$، $P=5$، و $T=50\ C$ باشد، آن‌گاه $V=20$ خواهد شد.

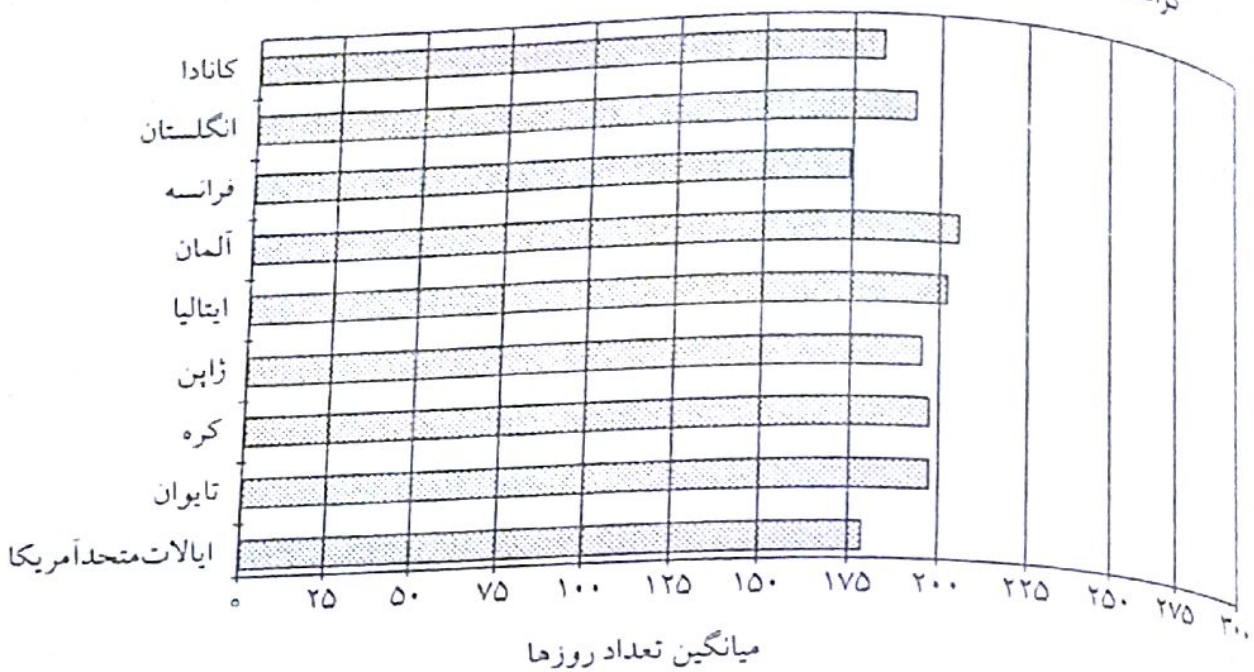
(صحیح) غلط

۱۹ - یک آزمون صحیح ـ غلط ۳۰ سؤالی در دست داریم که ضریب پایایی آن ۰/۶۰ است. برای افزودن بر پایایی این آزمون می‌خواهیم ۱۰ سؤال مشابه سؤالهای موجود به آن اضافه کنیم. ضریب پایایی آزمون جدید ۴۰ سؤالی ما ۰/۶۷ خواهد بود.

(صحیح) غلط

تحلیل دانش امور واقعی

میانگین تعداد روزهایی که دانش‌آموزان ۱۳ سالهٔ کشورهای مختلف در طول یک سال تحصیلی به مدرسه می‌روند. (اطلاعات مربوط به سالهای ۹۱ - ۱۹۹۰، نقل شده در لین و گرانلاند، ۲۰۰۰، ص ۲۲۴)



راهنمایی: سؤالهایی که در دنبال می‌آیند به اطلاعات موجود در شکل بالا اشاره می‌کنند. هر یک از آنها را بخوانید و پاسخ خود را مطابق با دستورالعمل زیر علامت بزنید:

صحیح: اگر براساس اطلاعات موجود در شکل سؤال صحیح است.
غلط: اگر براساس اطلاعات موجود در شکل سؤال غلط است.
نامشخص: اگر اطلاعات موجود در شکل برای تشخیص دادن صحیح و غلط بودن سؤال کافی نیست.

۲۰ - تعداد روزهای رفتن دانش‌آموزان به مدرسه در ایالات متحد کمتر از کشورهای دیگر است.

صحیح (غلط) نامشخص

۲۱ - تعداد روزهای رفتن دانش‌آموزان به مدرسه در سه کشور آسیایی بیشتر از بقیه کشورهاست.

(صحیح) غلط نامشخص

۲۲ - دانش‌آموزان آمریکایی از دانش‌آموزان ژاپنی ساعتهای کمتری را در مدرسه سپری می‌کنند.

صحیح غلط (نامشخص)

حوزهٔ عاطفی

دریافت کردن (علاقه)

۲۳ - دوست دارید در یک کلوپ ورزشی عضو شوید؟ بله نه

۲۴ - اگر یکی از اعضای خانوادهٔ من به موسیقی رادیو گوش دهد من ترجیح می‌دهم اتاق را ترک کنم. موافق مخالف

پاسخ دادن

۲۵ - من تکالیف درسی‌ام را با میل و رغبت انجام می‌دهم. موافق مخالف

۲۶ - برای شما غیرممکن است که یک ساعت مطالعه کنید و حوصله‌تان سرنرود؟ بله نه

ارزش‌گذاری (نگرش)

۲۷ - سازمان ملل یک نهاد بی‌خاصیت است. موافق مخالف

ارزش‌گذاری (باور)

۲۸ - حکومتها باید مواظب مردم باشند. موافق مخالف

سازمان دادن

۲۹ - خواندن کتاب برای من به صورت یک سرگرمی تازه درآمده است. موافق مخالف

خلاصهٔ فصل

سؤال صحیح ـ غلط از یک جمله تشکیل می‌شود که آزمون‌شونده (دانش‌آموز یا دانشجو) باید آن را به صورت صحیح یا غلط جواب دهد. معروف‌ترین شکل این آزمونها همان نوع صحیح ـ غلط است. با این حال، ما در این فصل سؤالهای صحیح ـ غلط را به صورتهای دیگر زیر نیز معرفی کرده‌ایم:

- نوع صحیح ـ غلط با اصلاح
- نوع چند صحیح ـ غلط یا نوع خوشه‌ای
- نوع بله ـ نه
- نوع بله ـ نه با توضیح
- نوع موافق ـ مخالف
- نوع واقعیت ـ عقیده

در سؤالهای صحیح ـ غلط با اصلاح، آزمون‌شونده، علاوه بر صحیح یا غلط تشخیص دادن

سؤال، باید برای سؤال غلط جواب درست تهیه کند. در نوع بله ـ نه، سؤال به صورتی طرح می‌شود که با جواب بله یا نه بهتر از صحیح یا غلط می‌توان به آن پاسخ داد. در نوع بله ـ نه با توضیح، علاوه بر انتخاب پاسخ بله یا نه، باید برای آن توضیح فراهم شود. در نوع موافق ـ مخالف، آزمون‌شونده نظر خودش را به صورت موافق یا مخالف سؤال اعلان می‌کند. و بالاخره در نوع واقعیت ـ عقیده، آزمون‌شونده باید تعیین کند که آنچه در سؤال آمده یک واقعیت است یا عقیدهٔ شخصی.

برای تهیهٔ سؤالهای صحیح ـ غلط صاحب‌نظران سنجش قواعد و پیشنهادهای مختلفی ارائه داده‌اند. ما در این فصل آنها را همراه با مثالهای مختلف توضیح داده‌ایم.

در رابطه با امتیازها و محدودیتهای آزمونهای صحیح ـ غلط، نظرهای مثبت و منفی زیادی اظهار شده است. از جمله نظرهای مثبت به قرار زیرند:

۱. به سادگی قابل تهیه‌اند.
۲. با عینیت و سهولت می‌توان پاسخها را تصحیح کرد.
۳. نمونهٔ بزرگی از محتوا و هدفهای آموزشی به دست می‌دهند.
۴. در یک جلسهٔ امتحان تعداد زیادی از آنها را می‌توان مورد استفاده قرار داد.

در مقابل، سؤالهای صحیح ـ غلط از سوی منتقدان با ایرادهایی روبه‌رو شده‌اند. مهم‌ترین ایرادها به قرار زیرند:

۱. سؤالهای صحیح ـ غلط، برخلاف انتظار، به سادگی قابل تهیه نیستند.
۲. برای همهٔ موضوعهای درسی مناسب نیستند.
۳. عمدتاً هدفهای سطح پایین آموزشی را می‌سنجند.
۴. میزان حدس‌پذیری آنها بالاست.

با وجود ایرادهای فوق، اگر در تهیهٔ سؤالهای صحیح ـ غلط دقت کافی به کار بسته شود، دامنهٔ استفادهٔ آنها را می‌توان از سطح دانش و یادآوری فراتر برد. ما در پایان این فصل نمونه‌هایی از سؤالهای صحیح ـ غلط برای هدفهای متنوع آموزشی مثال آورده‌ایم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. دست کم می‌توان ۷ نوع سؤال صحیح ـ غلط را نام برد.

الف ـ صحیح

ب ـ غلط

۲. براساس آنچه در این فصل خواندید، می‌توان از سؤالهای صحیح ـ غلط برای سنجش توانایی

حل مسئله در ریاضیات استفاده کرد؟

الف ـ بله

ب ـ نه

۳. در آزمونهای صحیح ـ غلط، برای اینکه سؤالها قدرت تمیز بیشتری داشته باشند، بهتر است نسبت سؤالهای صحیح به غلط چند باشد؟

الف ـ ۵۰ به ۵۰

ب ـ ۲۵ به ۷۵

پ ـ ۳۳ به ۶۷

ت ـ ۲۰ به ۸۰

۴. آزمون صحیح ـ غلط را تعریف کنید و با ذکر مثال انواع آن را شرح دهید.

۵. چگونگی تهیهٔ سؤالهای صحیح ـ غلط را توضیح دهید.

۶. قواعد تهیهٔ سؤالهای صحیح ـ غلط را با ذکر مثال توضیح دهید.

۷. کاربردهای سؤالهای صحیح ـ غلط را توضیح دهید.

۸ امتیازها و محدودیتهای سؤالهای صحیح ـ غلط را توضیح دهید.

۹. در رابطه با جدول مشخصات یا طرح سنجش خود که در تمرین مربوط به فصل قبل تهیه کردید، برای هدفهایی که با سؤالهای صحیح ـ غلط قابل سنجش‌اند، یک آزمون صحیح ـ غلط بسازید. سعی کنید آزمون شما تا حد امکان از سطح دانش یا یادآوری محض فراتر برود.

۸

سنجش یادگیری به کمک آزمون جورکردنی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمون جورکردنی را تعریف کند.
۲. ویژگیهای سؤالهای جورکردنی را توضیح دهد.
۳. انواع سؤالهای جورکردنی را شرح دهید و برای هر یک مثالی بزنند.
۴. قواعد تهیهٔ سؤالهای جورکردنی را توضیح دهد.
۵. امتیازها و محدودیتهای سؤالهای جورکردنی را توضیح دهد.
۶. موارد استفادهٔ آزمونهای جورکردنی را توضیح دهد.
۷. برای هدفهای یادگیری یک درس که با آزمونهای جورکردنی قابل سنجش‌اند یک آزمون جورکردنی تهیه نماید.

ما در فصل قبل یکی از روشهای عینی (یا گزینش پاسخ) سنجش یادگیری به نام آزمون صحیح ـ غلط را توضیح دادیم. در این فصل نوع دیگر این روشها یعنی آزمون جورکردنی را معرفی می‌کنیم. آگاهی یافتن معلم از ویژگیها و چگونگی تهیهٔ سؤالهای جورکردنی او را در استفادهٔ بهینه از این سؤالها به عنوان مکمل سؤالهای صحیح ـ غلط یاری خواهد داد.

تعریف و انواع آزمونهای جورکردنی

هر آزمون جورکردنی[1] از تعدادی تمرین[2] (سؤال) تشکیل می‌شود و هر تمرین نیز شامل سه بخش زیر است: (۱) راهنمای انجام دادن تمرین؛ (۲) فهرست محرکها یا پرسشها؛ (۳) فهرست پاسخها. آزمون‌شونده، با توجه به توضیحات راهنما، فهرست پاسخها را با فهرست پرسشها جور می‌کند. نگاه کنید به نمونهٔ زیر:

راهنمایی: در جاهای خالی مقابل اکتشافات و اختراعات ستون سمت راست زیر، حروف اول مربوط به کاشفان و مخترعان ستون سمت چپ را بنویسید. هر پاسخ را بیشتر از یک بار به کار نبرید.

(پ)	۱ ـ پنی‌سیلین	الف ـ ماری کوری
(ث)	۲ ـ پرتو	ب ـ دی باکی
(الف)	۳ ـ رادیوم	پ ـ فلمینگ
(ج)	۴ ـ واکسن فلج	ت ـ هاروی
(ب)	۵ ـ قلب مصنوعی	ث ـ رونتگن
(ت)	۶ ـ گردش خون	ج ـ سالک
		چ. پاستور

نمونهٔ بالا متداول‌ترین الگوی سؤالهای جورکردنی است، اما سؤالهای جورکردنی به این الگو منحصر نمی‌شوند و الگوهای دیگری را نیز می‌توان مورد استفاده قرار داد. برای مثال، می‌توان نقشه‌ای را که در آن نقاط مختلفی با شماره یا حرف مشخص شده‌اند در اختیار دانش‌آموزان گذاشت و از آنان خواست تا شماره‌ها یا حرفها را با نام شهرها، دریاچه‌ها، رودخانه‌ها، یا مواردی از این قبیل جور کنند. یا می‌توان شکل یک دستگاه برقی را که در آن قسمتهای مختلف دستگاه شماره‌گذاری شده‌اند به دانش‌آموزان داد و از آنان خواست تا نام قسمتهای مختلف دستگاه را به شماره‌های داخل شکل ارتباط دهند.

در تدوین سؤالهای جورکردنی، رابطهٔ پرسشها و پاسخها را می‌توان به گونه‌های دیگری تعیین کرد. از جمله اینکه می‌توان از آزمون‌شوندگان خواست تا از هر پاسخ خطی به پرسش

1. matching test 2. exercise

مربوط به آن بکشند. هر روشی را که مفید تشخیص می‌دهید به کار بندید، اما همواره سعی کنید که آن را در راهنمای تمرین یا سؤال به روشنی توضیح دهید.

نوع معروفی از سؤالها یا تمرینهای جورکردنی، نوع رده‌بندی[1] یا طبقه‌بندی است که به نوع مادهٔ فهرست اصلی[2] نیز شهرت دارد (نیتکو، ۲۰۰۱). در این نوع سؤالها، یک طرح رده‌بندی (مثلاً قسمتهای جمله، انواع جانوران، دوره‌های تاریخی، انواع گیاهان، انواع واکنشهای شیمیایی، و غیره) ارائه می‌شود و از دانش‌آموزان خواسته می‌شود تا هر سؤال را طبق این طرح رده‌بندی کنند. در آزمونهای نوع جورکردنی رده‌بندی یا مادهٔ فهرست اصلی هر یک ماده (سؤال) یا تمرین دارای سه قسمت زیر است: (۱) راهنمایی برای پاسخ دادن به سؤال، (۲) فهرست اصلی گزینه‌ها (که همان فهرست یا ستون پرسشهاست)، و (۳) فهرست پاسخها. چنان‌که گفتیم، دانش‌آموز، برای پاسخ دادن به این‌گونه سؤالها یا تمرینها، باید هر یک از پرسشها را که با شماره مشخص شده است بخواند، و بعد برای یافتن جواب آن در میان فهرست پاسخهایی که با حروف نشان داده شده‌اند پاسخ مناسب را بیابد.

نول و اسکانل (۱۹۷۲) و مهرنز و لهمان (۱۹۸۴) می‌گویند این نوع سؤالها بیشتر برای سنجش توانایی انتقال یادگیری، توضیح، و درک مطالب مناسب‌اند. به مثال زیر توجه کنید:

راهنمایی: در ستون اول نام دستوری قسمتهای مختلف یک جمله داده شده. در جای خالی مقابل هر جملهٔ ستون دوم حرف مربوط به نام دستوری کلمه‌ای را که با حروف سیاه نوشته شده است بنویسید.

الف ـ صفت

ب ـ قید

پ ـ حرف اضافه

ت ـ حرف ربط

(پ) ۱ ـ کسی در مدرسه نبود **جز** ناظم.

(ت) ۲ ـ **تا** درس نخوانی چیزی نمی‌آموزی.

(ت) ۳ ـ دیروز به کتابفروشی رفتم **و** دو جلد کتاب خریدم.

(الف) ۴ ـ همهٔ درهای **بسته** باز خواهند شد.

(الف) ۵ ـ پرویز از احمد دانش‌آموز **کوشاتری** است.

.

.

.

(ب) ۷ ـ پروین برای جبران سهل‌انگاری گذشته‌اش **سخت** می‌کوشد.

---

1. classification 2. master list item

سؤالهای جورکردنی رده‌بندی از نوع بالا با سؤالهای جورکردنی معمولی از لحاظ تعداد مادّه‌های فهرست پرسش و پاسخ فرق دارند. از آنجا که در سؤالهای رده‌بندی از یک پاسخ بیشتر از یک بار می‌توان استفاده کرد، در هر تمرین تعداد پرسشها می‌تواند بیشتر از تعداد پاسخها باشد، در صورتی که در سؤالهای جورکردنی معمولی تعداد پاسخها بیشتر از تعداد پرسشهاست.

یکی از امتیازهای سؤالهای جورکردنی از نوع رده‌بندی یا مادّهٔ فهرست اصلی این است که در تصحیح جوابها می‌توان از پاسخنامه‌های جداگانه شبیه آنچه برای آزمونهای چندگزینه‌ای به کار می‌رود استفاده کرد. به نمونهٔ زیر توجه کنید:

راهنمایی: در پرسشهای زیر شما باید تاثیر ورزش را بر فرایندها و مواد مختلف بدن تعیین کنید. فرض کنید که در ارگانیسم هیچ تغییری بجز آنچه در نتیجهٔ ورزش رخ می‌دهد اتفاق نمی‌افتد. برای هر پرسش مطابق با توضیحات زیر در پاسخنامه علامت بگذارید.

الف ـ اگر نتیجهٔ ورزش قطعاً **افزایش** چیزی است که در پرسش آمده.

ب ـ اگر نتیجهٔ ورزش قطعاً **کاهش** چیزی است که در پرسش آمده.

پ ـ اگر ورزش هیچ‌گونه تاثیر چشم‌گیری بر چیزی که در پرسش آمده ندارد یا تاثیر آن غیرقابل **پیش‌بینی** است.

	الف	ب	پ
۱ ـ ضربان قلب	■	□	□
۲ ـ فشار خون	■	□	□
۳ ـ مقدار قند موجود در خون	□	■	□
۴ ـ مقدار هوای اضافی در ششها	□	■	□
۵ ـ و غیره			

(ایبل، ۱۹۷۲، ص ۱۰۵)

☆ ☆ ☆ ☆ ☆

نکته‌ای را که لازم است در رابطه با سؤالهای جورکردنی یادآور شویم این است که فهرست پرسشها و فهرست پاسخها را هم می‌توان در دو ستون موازی در کنار هم قرار داد و هم می‌توان آنها را پشت سر هم نوشت (مانند مثال بالا). آنجا که طول عبارتهای فهرست پرسش و فهرست پاسخ کوتاه است و دو ستون در کنار هم جا می‌گیرند بهتر است آنها را به طور موازی پهلوی هم قرار دهید، اما در جایی که مطالب فهرست پرسش یا فهرست پاسخ مفصل است و در کنار هم جا نمی‌گیرند آنها را به دنبال یکدیگر بیاورید.

قواعد تهیهٔ سؤالهای جورکردنی

صاحب‌نظران سنجشِ یادگیری در رابطه با چگونگی تهیهٔ سؤالهای جورکردنی قاعده‌هایی را پیشنهاد داده‌اند که ما در این قسمت به توضیح آنها می‌پردازیم. در بحث از سؤالهای جورکردنی

ما اصطلاح تکلیف را مورد استفاده قرار داده‌ایم. منظور ما از تکلیف مجموعه‌ای از پرسشها و پاسخهای همگون است که آزمون شونده باید آنها را با هم جور کند. این مجموعه بر روی هم بیشتر از یک سؤال محسوب می‌شود. در واقع تعداد سؤالهای هر تکلیف برابر تعداد پرسشهای آن است.

۱. **اطلاعات کامل دربارهٔ چگونگی جور کردن پرسشها و پاسخها را در راهنمای سؤال بنویسید.** هرچند که در آزمونهای جورکردنی نحوهٔ عمل جورکردن پرسشها با پاسخها نسبتاً آشکار است، اما باز هم لازم است در راهنمای سؤالها چگونگی این کار را توضیح دهید. این توضیحات از سردرگمی دانش‌آموزان جلوگیری می‌کند. با این حال، از نوشتن توضیحات خیلی مفصل در راهنمای سؤالها که موجب به هدر دادن مقداری از وقت آزمون شوندگان می‌شود پرهیز کنید.

به طور کلی، در راهنمای سؤالها سه دسته اطلاع زیر را قرار دهید: (۱) اساس رابطهٔ بین پرسشها و پاسخها یا مبنای جور کردن آنها؛ (۲) روش مشخص کردن پاسخها یا جور کردن پرسشها و پاسخها؛ و (۳) آیا از یک پاسخ تنها یک بار استفاده بشود یا بیشتر از یک بار. در راهنمای سؤالهایی که تاکنون در این فصل مثال زدیم موارد سه گانهٔ بالا دیده می‌شوند. در زیر به نمونهٔ دیگری از راهنمای سؤال که هر سه مورد بالا را شامل است توجه کنید.

راهنمایی: در زیر دو ستون یا دو فهرست مطلب وجود دارند. ستون اول در برگیرندهٔ عناوین داستانهای کوتاه و ستون دوم شامل نامهای نویسندگان است. برای هر یک از داستانهای ستون سمت راست نام یکی از نویسندگان ستون سمت چپ را برگزینید. حرف مقابل نام نویسنده را در جای خالی مقابل عنوان داستان بنویسید. تعداد نامهای نویسندگان بیشتر از تعداد عنوانهای داستانهاست. بنابراین، تعدادی نام نویسنده اضافی خواهید آورد. همچنین ممکن است نام یک نویسنده برای بیشتر از یک داستان انتخاب شود.

(نول و اسکانِل، ۱۹۷۲، ص ۸)

هنگام اجرای آزمونهای جورکردنی، علاوه بر توضیحات کتبی، اگر لازم بود، توضیحات شفاهی هم بدهید. این عمل به ویژه در رابطه با دانش‌آموزان سنهای پایین‌تر مفید است. برای این منظور، می‌توانید مثالی بر روی تابلو بنویسید. همچنین می‌توانید از دانش‌آموزان بخواهید تا، به جای نوشتن حرفهای مربوط به پاسخها در مقابل پرسشها، از هر پاسخ خطی به پرسش مربوط به آن وصل کنند.

۲. **موضوعهای مهم و در ارتباط با هدفهای آموزشی را در سؤالها قرار دهید.** ما در ارتباط با تهیهٔ سؤالهای صحیح ـ غلط بر این نکته تأکید کردیم. در اینجا نیز آن را یادآور می‌شویم. هر تمرین جورکردنی باید یک موضوع مهم مربوط به یک هدف آموزشی را بسنجد. از قراردادن مطالب جزئی، بی‌اهمیت، و نامربوط به محتوای درس و هدفهای یادگیری بپرهیزید. در تدارک همهٔ روشهای سنجش این مطلب را به یاد داشته باشید.

می‌دهید، پرسشهای تمرین جورکردنی را از ۱۶ تا ۲۰ شماره گذاری کنید. در این حالت ارزش یک تمرین جورکردنی ۵ پرسشی شما ۵ نمره خواهد بود. (ما در سؤالهای نمونه‌ای که در این فصل آورده‌ایم همواره این نکته را در نظر گرفته‌ایم.)

۹. **فهرست پرسشها و فهرست پاسخها را به طور منطقی مرتب کنید.** اگر فهرست پرسشها و پاسخهای یک تمرین جورکردنی به گونه‌ای منطقی مرتب شده باشند، دانش‌آموزان مطلع از موضوع بدون صرف وقت پاسخ درست هر پرسش را پیدا خواهند کرد. در غیر این صورت، ممکن است وقت زیادی صرف یافتن آن بکنند. علاوه بر این، مرتب کردن فهرست پرسشها و فهرست پاسخها موجب می‌شود که دانش‌آموزان نتوانند اشاره‌های احتمالی را که منجر به یافتن پاسخهای درست می‌شود کشف کنند.

برای مرتب کردن مطالب، نامها را به ترتیب حروف الفبا، تاریخها را به ترتیب زمان، و اعداد را از بزرگ به کوچک یا به عکس مرتب کنید. به نمونهٔ زیر توجه نمایید:

راهنمایی: در جاهای خالی مقابل آثار ستون سمت راست، حروف مربوط به نویسندگان آن آثار را که در ستون سمت چپ نوشته شده‌اند قرار دهید.

(ج)	۳ ـ شاهنامه	الف ـ	بوعلی‌سینا
(پ)	۱ ـ بوستان	ب ـ	حافظ
(ث)	۵ ـ کیمیای سعادت	پ ـ	سعدی
(الف)	۴ ـ شفا	ت ـ	عروضی
(ت)	۲ ـ چهارمقاله	ث ـ	غزالی
		ج ـ	فردوسی

۱۰. **در برگه‌های امتحانی، همهٔ پرسشها و پاسخهای یک سؤال را در یک صفحه قرار دهید.** رعایت این سفارش اجرای آزمونهای جورکردنی را آسان می‌کند. اگر نصف مطالب یک تمرین در یک صفحه و نصف دیگر آن در صفحه‌ای دیگر باشد، مراجعهٔ مکرر از این صفحه به آن صفحه موجب حواس‌پرتی و صرف وقت زیاد از دانش‌آموزان می‌شود. در ضمن، ورق زدن برگه‌های امتحانی توسط دانش‌آموزان مختلف باعث تولید سروصدای زیاد و ایجاد مزاحمت می‌شود. بنابراین، سؤالها را طوری تنظیم کنید که تمام مطالب دو ستون پرسش و پاسخ یک تمرین در یک صفحه قرار بگیرند تا آزمون‌شوندگان مجبور نباشند برای یافتن پاسخهای مورد نظر مرتباً برگه‌ها را ورق بزنند. همچنین قرار گرفتن بخشی از پاسخها در صفحهٔ دیگر ممکن است موجب شود که دانش‌آموزان آنها را نادیده بگیرند.

۱۱. **در نوشتن صورت سؤالها ابتکاری عمل کنید.** لازم نیست در نوشتن سؤالهای

۳. پرسشها و پاسخهای هر تمرین را متجانس انتخاب کنید. منظور از تجانس این است که مجموعهٔ پرسشها و پاسخهای یک تمرین به یک مفهوم، طبقه، یا زمینهٔ واحد مربوط باشند. در زیر به نمونه‌ای از سؤالهای نامتجانس توجه کنید.

راهنمایی: در جای خالی مقابل هر شماره در ستون سمت راست، حرف مربوط به مطلب ستون سمت چپ را که با آن جور درمی‌آید بنویسید. هر پاسخ را بیشتر از یک بار به کار نبرید.

(پ)	۱ ـ پنی‌سیلین را کشف کرد.	الف ـ آرمسترانگ
(ب)	۲ ـ آمریکا را کشف کرد.	ب ـ کریستف کلمب
(الف)	۳ ـ اولین فضانوردی که بر ماه قدم نهاد.	پ ـ فلمینگ
(ت)	۴ ـ اولین کاشف که دماغهٔ امیدنیک را دور زد.	ت ـ گلن
		ث ـ مارلان
		ج ـ نیوتون

تجانس یا همگونی یک امر نسبی است و به همین دلیل طراحان آزمون ممکن است در نوشتن سؤالهای آزمونهای جورکردنی با مشکل مواجه شوند، زیرا آنچه برای یک گروه از آزمون‌شوندگان متجانس است ممکن است برای گروهی دیگر نامتجانس جلوه کند. لین و گرانلاند (۲۰۰۰، ص ۱۸۸) در این‌باره سؤال زیر را ذکر کرده‌اند.

راهنمایی: در جای خالی مقابل هر شماره در ستون سمت راست حرف مربوط به مطلب ستون سمت چپ را که با آن جور درمی‌آید بنویسید. هر یک از نامهای ستون سمت چپ را می‌توانید یک بار یا بیشتر از یک بار به کار برید یا اصلاً آن را مورد استفاده قرار ندهید.

(الف)	۱ ـ مخترع تلفن	الف ـ الکساندر گراهام‌بل
(ب)	۲ ـ کاشف امریکا	ب ـ کریستوفر کولومبوس
(پ)	۳ ـ اولین فضانورد امریکایی که زمین را دور زد	پ ـ جان گلن
(ت)	۴ ـ اولین رئیس‌جمهور امریکا	ت ـ ابراهام لینکن
		ث ـ فردیناند مگلان
		ج ـ جورج واشنگتن
		چ ـ الی ویتنی

تمرین بالا از آنجا که تنها انسانها را شامل می‌شود متجانس است. این تمرین ممکن است برای دانش‌آموزان ابتدایی متجانس به نظر برسد، اما برای دانش‌آموزان سالهای بالاتر که قدرت تشخیص بیشتری دارند خیلی نامتجانس به نظر خواهد رسید زیرا شامل مکتشفان، مخترعان، و رؤسای جمهور است. بنابراین، برای استفاده در کلاسهای سطح بالاتر باید، به جای این تمرین، تمرین دیگری تهیه کرد که تنها شامل مخترعان، مکتشفان، یا رؤسای جمهور باشد. یعنی تمرین را می‌توان

به سه تمرین متفاوت تبدیل کرد. بنابراین، به همان نسبتی که بخواهیم قدرت تمیز یا تشخیص سؤالهای جورکردنی را بالا ببریم، باید بر تجانس یا همگونی عناصر پرسش و پاسخ آنها بیفزاییم.

۴. **طول فهرست پرسشها و پاسخها را کوتاه بگیرید.** فهرست کوتاه مطالب هم برای معلم بهتر است هم برای دانش‌آموزان. معلم در تهیهٔ مطالب کمتر بیشتر می‌تواند تجانس آنها را رعایت کند. از لحاظ مفید بودن برای دانش‌آموزان، خواندن فهرستهای کوتاه‌تر مطالب دقت و تمرکز حواس را بیشتر امکان‌پذیر می‌سازد.

متخصصان سنجش و ارزشیابی پیشرفت تحصیلی رقمهای ۴ تا ۱۰ ماده را برای فهرست پرسشها و پاسخها پیشنهاد کرده‌اند. پوفام (۲۰۰۲) و نول و اسکانل (۱۹۷۲) بالاترین رقم را برای تعداد پرسشها و پاسخها ۱۰ تا ۱۲ ذکر کرده‌اند. پوفام در رابطه با طولانی بودن فهرست پرسشها و فهرست پاسخها می‌گوید "مشکل فهرستهای طولانی این است که دانش‌آموزان آن قدر وقت صرف جداکردن پاسخ مناسب برای یک پرسش می‌کنند که ممکن است یادشان برود که به دنبال چه چیزی می‌گردند" (ص ۱۴۲).

۵. **توضیحات مفصل را در فهرست پرسشها قرار دهید و در فهرست پاسخها مطالب کمتری بنویسید.** قراردادن مطالب مختصر در ستون پاسخها به کارآمدی سؤال کمک می‌کند. این امر باعث می‌شود که دانش‌آموزان ابتدا مطالب طولانی‌تر پرسشها را به دقت مطالعه کنند، و آن‌گاه به سرعت پاسخها را بخوانند و تصمیم خود را در رابطه با انتخاب آنها بگیرند.

دانش‌آموز، هنگام جواب دادن به یک تمرین جورکردنی، ابتدا یک پرسش را می‌خواند، بعد از میان فهرست پاسخها به دنبال پاسخ مناسب برای آن پرسش می‌گردد. برای هر پرسش این کار را تکرار می‌کند. یعنی به تعداد پرسشها تمامی فهرست پاسخها را می‌خواند. بنابراین، بهتر است مطالب مفصل در فهرست پرسشها ـ که دانش‌آموز هر یک از آنها را تنها یک بار می‌خواند ـ بیاید، و مطالب کوتاه (کلمه‌ها، نمادها، و غیره) در فهرست پاسخها ـ که بارها باید خوانده شود ـ قرار گیرد. به نمونهٔ زیر توجه کنید:

راهنمایی: در جای خالی مقابل هر پرسش در ستون سمت راست حرف مربوط به ضریب پایایی یکی از موارد ستون سمت چپ را بنویسید. هر یک از حرفهای مربوط به ستون سمت چپ را می‌توانید یک بار یا چند بار به کار برید، یا هرگز از آن استفاده نکنید.

(ث) ۱ ـ نمرات استانفورد ـ بینه با نمرات آزمون وکسلر — الف ـ ثبات

همبستگی دارند.

(ت) ۲ ـ روشی که غالباً برای تخمین پایایی آزمونهای — ب ـ تعادل

معلم ساخته به کار می‌رود.

(ت) ۳ ـ معمولاً کمترین برآورد پایایی را به‌دست می‌دهد. — پ ـ ثبات و تعادل

(ت) ۴ ـ نتایج تحت تاثیر آزمونهای سرعت قرار دارند.

(الف) ۵ ـ آزمون استعداد تفضیلی یک بار به دانش‌آموزان ت ـ دو نیمه کردن
سال اول دبیرستان و بار دوم به همان دانش‌آموزان
در سال آخر دبیرستان داده می‌شود. ث ـ هیچ یک از موارد بالا

(مهرنز و لهمان، ۱۹۷۳، ص ۲۶۰)

۶. از نوشتن تمرینها یا سؤالهایی که پرسشها و پاسخهای آنها کاملاً جور می‌شوند پرهیز کنید. منظور از این پیشنهاد آن است که تمرینهایی ننویسید که در آنها هر پاسخ تنها یک بار با یک پرسش جور می‌شود و تعداد پاسخها و پرسشهای آن برابر است. در این گونه تمرینها، اگر دانش‌آموز پاسخ همهٔ پرسشها بجز یکی از آنها را بداند، آخری را نیز خودبه‌خود جواب خواهد داد. یا اگر در یکی از پاسخها اشتباهی صورت گیرد در آخرین پاسخ هم همین اشتباه تکرار خواهد شد. اگر تعداد پاسخها از تعداد پرسشها بیشتر باشد این مشکل برطرف خواهد شد. همچنین سعی کنید پاسخهایی بنویسید که بیش از یک بار مورد استفاده قرار گیرند.

بنابراین، در نوشتن سؤالهای جورکردنی، لازم است تعداد پاسخها از تعداد پرسشها بیشتر باشد، یا اینکه بتوان هر پاسخ را برای بیشتر از یک پرسش مورد استفاده قرار داد (مانند سؤالهای رده‌بندی که در اوایل این فصل معرفی شدند). همچنین می‌توان سؤال را به گونه‌ای طراحی کرد که در آن بعضی پاسخها یک بار و بعضی دیگر بیشتر از یک بار مورد استفاده قرار گیرند، و بعضی از آنها اصلاً قابل استفاده نباشند. این مطلب باید در راهنمای سؤال توضیح داده شود.

۷. هر یک از پاسخها را به گونه‌ای انتخاب کنید که برای تمام پرسشها درست جلوه کند. سؤال را به گونه‌ای طرح نکنید که پاسخهای غلط کاملاً مشخص باشند، طوری که دانش‌آموز به راحتی بتواند آنها را تشخیص داده کنار بگذارد. چنان که قبلاً توضیح دادیم، هر پاسخی چه درست و چه غلط باید برای دست‌کم یک پرسش درست جلوه کند. اگر قاعده‌ای را که دربارهٔ متجانس بودن پرسشها و پاسخها قبلاً توضیح دادیم رعایت کنید این منظور برآورده می‌شود، زیرا معمولاً پاسخهای متجانس برای فهرستی از پرسشهای متجانس درست به نظر می‌رسند.

۸. پرسشها را با شماره و پاسخها را با حرف مشخص کنید. در تمرینهای مربوط به آزمونهای جورکردنی هر پرسش یک نمرهٔ جداگانه دریافت می‌کند. بنابراین، هر پرسش باید دارای یک شماره باشد و به عنوان یک سؤال مستقل منظور شود. اگر در یک آزمون از انواع مختلف سؤالها استفاده می‌کنید، همان طور که برای هر یک از آنها یک شماره اختصاص می‌دهید، برای هر یک از پرسشهای سؤالهای جورکردنی نیز یک شماره اختصاص دهید. مثلاً، اگر ۱۵ سؤال اول آزمون شما صحیح ـ غلط است و به دنبال آنها یک تمرین جورکردنی دارای ۵ پرسش قرار

می‌دهید، پرسشهای تمرین جورکردنی را از ۱۶ تا ۲۰ شماره گذاری کنید. در این حالت ارزش یک تمرین جورکردنی ۵ پرسشی شما ۵ نمره خواهد بود. (ما در سؤالهای نمونه‌ای که در این فصل آورده‌ایم همواره این نکته را در نظر گرفته‌ایم.)

۹. **فهرست پرسشها و فهرست پاسخها را به طور منطقی مرتب کنید.** اگر فهرست پرسشها و پاسخهای یک تمرین جورکردنی به گونه‌ای منطقی مرتب شده باشند، دانش‌آموزان مطلع از موضوع بدون صرف وقت پاسخ درست هر پرسش را پیدا خواهند کرد. در غیر این صورت، ممکن است وقت زیادی صرف یافتن آن بکنند. علاوه بر این، مرتب کردن فهرست پرسشها و فهرست پاسخها موجب می‌شود که دانش‌آموزان نتوانند اشاره‌های احتمالی را که منجر به یافتن پاسخهای درست می‌شود کشف کنند.

برای مرتب کردن مطالب، نامها را به ترتیب حروف الفبا، تاریخها را به ترتیب زمان، و اعداد را از بزرگ به کوچک یا به عکس مرتب کنید. به نمونهٔ زیر توجه نمایید:

راهنمایی: در جاهای خالی مقابل آثار ستون سمت راست، حروف مربوط به نویسندگان آن آثار را که در ستون سمت چپ نوشته شده‌اند قرار دهید.

(ج)	۳ ـ شاهنامه	الف ـ	بوعلی‌سینا
(پ)	۱ ـ بوستان	ب ـ	حافظ
(ث)	۵ ـ کیمیای سعادت	پ ـ	سعدی
(الف)	۴ ـ شفا	ت ـ	عروضی
(ت)	۲ ـ چهارمقاله	ث ـ	غزالی
		ج ـ	فردوسی

۱۰. **در برگه‌های امتحانی، همهٔ پرسشها و پاسخهای یک سؤال را در یک صفحه قرار دهید.** رعایت این سفارش اجرای آزمونهای جورکردنی را آسان می‌کند. اگر نصف مطالب یک تمرین در یک صفحه و نصف دیگر آن در صفحه‌ای دیگر باشد، مراجعهٔ مکرر از این صفحه به آن صفحه موجب حواس‌پرتی و صرف وقت زیاد از دانش‌آموزان می‌شود. در ضمن، ورق زدن برگه‌های امتحانی توسط دانش‌آموزان مختلف باعث تولید سروصدای زیاد و ایجاد مزاحمت می‌شود. بنابراین، سؤالها را طوری تنظیم کنید که تمام مطالب دو ستون پرسش و پاسخ یک تمرین در یک صفحه قرار بگیرند تا آزمون‌شوندگان مجبور نباشند برای یافتن پاسخهای مورد نظر مرتباً برگه‌ها را ورق بزنند. همچنین قرار گرفتن بخشی از پاسخها در صفحهٔ دیگر ممکن است موجب شود که دانش‌آموزان آنها را نادیده بگیرند.

۱۱. **در نوشتن صورت سؤالها ابتکاری عمل کنید.** لازم نیست در نوشتن سؤالهای

جورکردنی (یا هر نوع سؤال دیگری) صرفاً از قالبهای متداول پیروی نمایید. با حفظ اصل اساسی حاکم بر نوع سؤال (در آزمونهای جورکردنی، عمدتاً سنجش رابطه) ابتکار به خرج دهید و سؤالهای تازه و متنوع بسازید. در زیر به نمونه‌ای از سؤالهای جورکردنی ابتکاری توجه کنید.

راهنمایی: شما دربارهٔ زمین و حرکات آن، آن‌گونه که واقعاً وجود دارد، اطلاعاتی کسب کرده‌اید. در این سؤال باید آثار شرایطی کاملاً فرضی را شناسایی کنید. در مقابل شمارهٔ هر سؤال در پاسخنامه، برحسب موارد زیر علامت بزنید.

الف ــ در صورتی که زمین نسبت به مدار خود مایل نبود سؤال درست بود.

ب ــ در صورتی که مدار به جای بیضی دایره بود سؤال درست بود.

پ ــ در صورتی که زمین به جای شرق به طرف غرب می‌چرخید سؤال درست بود.

ت ــ در صورتی که زمین با همین جرم نصف قطر فعلی‌اش را داشت سؤال درست بود.

ث ــ در صورتی که زمین ماه نداشت سؤال درست بود.

۱ ــ طول همهٔ روزهای شمسی مساوی می‌بودند.

۲ ــ اشیاء روی زمین چهار برابر وزن فعلی خود را داشتند.

۳ ــ خط استوای کرهٔ آسمانی و مدار خورشید یکی بود.

۴ ــ خورشید از شرق غروب می‌کرد.

۵ ــ ستارهٔ شمالی دیگری باید انتخاب می‌شد.

۶ ــ نیروی جاذبه چهار برابر بیشتر از حالا می‌بود.

۷ ــ سرعت حرکت زمین بر روی مدارش در ضمن سال تغییر نمی‌کرد.

۸ ــ ما خیلی کمتر از حالا دربارهٔ ماهیت خورشید اطلاع می‌داشتیم.

۹ ــ طول روز و شب در تمام طول سال در همهٔ عرضهای جغرافیایی برابر بود.

(از بلوم و همکاران، ۱۹۵۶، ترجمهٔ سیف و علی‌آبادی، ۱۳۶۸، ص ۱۴۸)

امتیازها و محدودیتهای آزمونهای جورکردنی

صاحب‌نظران سنجش و اندازه‌گیری، در رابطه با آزمونهای جورکردنی، نظرهای موافق و مخالف ابراز داشته‌اند. به طور کلی، امتیازهای استفاده از سؤالهای جورکردنی را می‌توان به صورت زیر خلاصه کرد:

۱. از آنجاکه پاسخ دادن به این سؤالها زمان زیادی لازم ندارد، تعداد زیادی از آنها را می‌توان در یک زمان معین مورد استفاده قرار داد. بنابراین، استفاده از این نوع سؤالها، مانند سؤالهای عینی دیگر، نمونهٔ نسبتاً بزرگی از محتوای درس و هدفهای آموزشی فراهم می‌آورد.

۲. تصحیح سؤالهای جورکردنی به سرعت و سهولت امکان‌پذیر است و نیازی به مصحح

متخصص ندارد. علاوه بر این، تصحیح این سؤالها به طور کاملاً عینی امکان‌پذیر است و در آن امکان دخالت نظر شخصی مصحح وجود ندارد.

۳. امکان حدس زدن جواب درست آزمونهای جورکردنی کمتر از سایر آزمونهای گزینش پاسخ، به ویژه آزمونهای صحیح ـ غلط، است.

۴. از سؤالهای جورکردنی برای سنجش هدفهای یادگیری سطح بالاتری از سؤالهای صحیح ـ غلط می‌توان استفاده کرد.

در مقابل فایده‌های بالا، برای آزمونهای جورکردنی محدودیتهایی ذکر کرده‌اند. مهم‌ترین آنها به قرار زیرند:

۱. از آنجا که در یادگیری تداعی بین امور حفظ طوطی‌وار دخالت زیادی دارد و معلمان غالباً در تهیهٔ این نوع سؤالها بر روابط حفظی بین امور تأکید می‌کنند، این آزمونها اگر به دقت تهیه نشوند ممکن است دانش‌آموزان را به یادگیری مطالب سطحی تشویق کنند.

۲. بعضی وقتها پیداکردن مجموعه‌ای پرسش که به اندازهٔ کافی شبیه به هم باشند تا بتوان برای آنها مجموعه پاسخهای همگونی پیداکرد کار دشواری است.

۳. علاوه بر معایب بالا، بعضی از سؤالهای جورکردنی معمولی که به وسیلهٔ معلمان تهیه می‌شوند دارای عیبهای زیر نیز هستند:

الف ـ مبهم بودن راهنمای سؤال

ب ـ طولانی بودن فهرست پرسشها و پاسخها

پ ـ مبهم بودن بیان پرسشها

موارد استفادهٔ آزمونهای جورکردنی

آزمونهای جورکردنی، در قیاس با آزمونهای صحیح ـ غلط، برای سنجش دامنهٔ وسیع‌تری از هدفهای آموزشی قابل استفاده‌اند. عمده‌ترین مورد استفادهٔ آزمونهای جورکردنی سنجش هدفهایی است که روابط بین امور را شامل می‌شوند ـ به ویژه سنجش توانایی آزمون‌شوندگان در تشخیص روابط میان موضوعهای نسبتاً مشابه یا مربوط به هم. در این سؤالها، می‌توان از دانش‌آموز خواست تا تاریخها را به رویدادها، مؤلفان را به کتابها، ابزارها را به موارد استفادهٔ آنها، و مانند اینها ربط داد. لین و گرانلاند (۲۰۰۰) نمونهٔ روابط مهم از لحاظ آموزش معلمان را در فهرست زیر آورده‌اند:

اشخاص ..... دستاوردها

تاریخها ..... رویدادهای تاریخی

اصطلاحات ..... تعاریف

قواعد ..... مثالها
نمادها ..... مفاهیم
نویسندگان ..... عنوانهای کتابها
کلمات خارجی ..... معادلهای زبان مادری
ماشینها ..... موارد استفاده
گیاهان یا جانوران ..... طبقه‌بندیها
اصلها ..... نمونه‌ها
اشیاء ..... نامهای اشیاء
اجزاء ..... کارکردها

علاوه بر موارد بالا، از آزمونهای جورکردنی می‌توان برای سنجش توانایی تشخیص روابط در شکلها و نمودارها نیز استفاده کرد. برای نمونه می‌توان از دانش‌آموزان خواست تا شکلها را با کلمات جور کنند یا موقعیتهای مختلف را در نقشه‌ها، نمودارها، طرحها، و مانند اینها تشخیص دهند. "صرف‌نظر از صورت ظاهر ارائهٔ سؤال، وظیفهٔ دانش‌آموز این است که دو چیز را که بر اساس نوعی منطق با هم جور درمی‌آیند به هم ربط دهد" (لین و گرانلاند، ۲۰۰۰، ص ۱۸۷).

علاوه بر سنجش توانایی تشخیص روابط میان امور، از آزمونهای جورکردنی نوع رده‌بندی یا مادهٔ فهرست اصلی می‌توان برای سنجش درک مفاهیم نیز استفاده کرد. نیتکو (۲۰۰۱) در این‌باره گفته است "برای اینکه از این نوع سؤالها (جورکردنی نوع فهرست اصلی) در سنجش درک مفهوم استفاده نمایید، مثالهایی را که به دانش‌آموزان می‌دهید تا دسته بندی کنند نه باید مثالهایی که در کلاس مورد استفاده قرار داده‌اید باشند نه مثالهایی که در کتاب درسی آمده‌اند" (ص ۱۷۳). ثرندایک، کانینگهام[1]، ثرندایک، و هیگن (۱۹۹۱)، دربارهٔ سؤالهای جورکردنی نوع مادهٔ فهرست اصلی، گفته‌اند "این‌گونه سؤالها را می‌توان با اطمینان کامل برای سنجش کاربستن، فهمیدن، و تفسیر کردن مورد استفاده قرار داد. این سؤالها راه مناسبی برای کشف و تعیین دامنه‌ای از یادگیری مفاهیم مختلف فراهم می‌آورند" (ص ۲۳۸).

همچنین از آزمونهای جورکردنی (باز هم از نوع ماده یا سؤال فهرست اصلی و نوع رده‌بندی) می‌توان برای سنجش توانایی تحلیل استفاده کرد. "سؤال نوع فهرست اصلی راه مناسبی برای سنجش تواناییهای زیر است: (۱) تحلیل متنهای مختصر، جدولها، یا نمودارها، و (۲) تشخیص تفسیر یا استنتاج مناسب از مطالب تفسیری" (نیتکو، ۲۰۰۱، ص ۱۷۴). در زیر به نمونه‌هایی از سؤالهای جورکردنی برای سنجش سطوح مختلف توجه کنید:

1. Cunningham

یادآوری دانش امور واقعی

راهنمایی: در جای خالی مقابل نام هر یک از نویسندگان ستون سمت راست، حرف مربوط به عنوان یکی از کتابهای ستون سمت چپ را قرار دهید. هر یک از حروف ستون سمت چپ را می‌توانید یک بار یا بیش از یک بار مورد استفاده قرار دهید یا اصلاً از آن استفاده نکنید. مورد اول به عنوان نمونه جواب داده شده است.

ستون نویسندگان		ستون عنوان کتابها
(الف)	× - آرتور هیلی	الف - چرخها
(ت)	۱ - اریک سیگال	ب - خیابان اصلی
(ث)	۲ - آیرونیگ والاس	پ - بادهای جنگ
(ح)	۳ - جوزف وامباگ	ت - داستان عاشقانه
(پ)	۴ - هرمان ووک	ث - هفت دقیقه
(ج)	۵ - لئون اُرس	ج - کیوبی هفت
(خ)	۶ - پرل باک	چ - پله‌های چرخان
(ج)	۷ - الکس هیلی	ح - فرماندهان جدید
		خ - خاک خوب
		د - ریشه‌ها

(مهرنز و لهمان، ۱۹۸۴، ص ۱۴۰)

یادآوری دانش مفهومی

راهنمایی: ستون سمت چپ دربرگیرندهٔ اصطلاحات مختلف روانشناسی یادگیری و ستون سمت راست شامل موقعیتهای مختلف مُعرف آن اصطلاحات است. با گذاشتن حروف اول مقابل مطالب ستون سمت چپ در جاهای خالی مقابل شماره‌های مربوط به مطالب ستون سمت راست، آنها را با هم جور کنید. از هر یک از موارد الف و ب و پ می‌توانید یک بار یا بیشتر از یک بار استفاده کنید یا اصلاً از آن استفاده نکنید.

(الف)	۸ - ارائهٔ تقویت کنندهٔ مثبت پس از رفتار	الف - تقویت مثبت
(ب)	۹ - حذف تقویت کنندهٔ منفی پس از رفتار	ب - تقویت منفی
(پ)	۱۰ - ارائهٔ تقویت‌کنندهٔ منفی پس از رفتار	پ - تنبیه
(ث)	۱۱ - حذف تقویت‌کنندهٔ مثبت پس از رفتار	ت - محروم کردن
		ث - جریمه کردن

فهمیدن دانش مفهومی

راهنمایی: مطالب موجود در ستون سمت راست معرف انواع روایی‌های داده شده در ستون سمت چپ است. در جای خالی مقابل هر یک از پرسشهای ستون سمت راست حرف مربوط

به یکی از پاسخهای ستون سمت چپ را بنویسید. از هر یک از پاسخها می‌توانید یک یا چند بار استفاده کنید یا اصلاً آن را مورد استفاده قرار ندهید.

(الف) ۱۲ - برای ساختن یک آزمون دیکته، معلم از کلماتی که محتوایی که دانش‌آموزان در نوشتن انشاء غالباً آنها را غلط می‌نویسند استفاده می‌کند.

(ب) ۱۳ - معلمی می‌خواهد مشخص کند که اشکالات لغتی دانش‌آموزان در دیکته همان اشکالات لغتی آنان در نوشتن انشاء است.

(پ) ۱۴ - در دفترچهٔ راهنمای یک آزمون هوش بین نمرات حاصل از آن آزمون و آزمون هوش استانفورد ـ بینه ضریب همبستگی ۰/۷۶+ گزارش شده است.

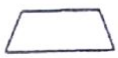
(پ) ۱۵ - میزان تولید کارگران با درجه‌بندی سرپرست کارگاه از موفقیت کارگران مقایسه شده است.

(الف) ۱۶ - معلم به تهیهٔ جدول مشخصات اقدام کرده است.

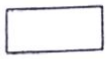
الف - روایی
ب - روایی سازه
پ - روایی ملاکی
ت - هیچ یک از موارد بالا

(مهرنز و لهمان، ۱۹۷۳، ص ۲۶۳)

راهنمایی: درستون سمت راست پنج شکل هندسی و در ستون سمت چپ نامهای شکلهای هندسی مختلف داده شده‌اند. با قرار دادن حرف اول نام شکلهای فهرست سمت چپ در مقابل شکلهای فهرست سمت راست آنها را با هم جور کنید. از هر پاسخ تنها یک بار می‌توانید استفاده کنید یا هرگز استفاده نکنید.

(پ) ۱۷ - 

(الف) ۱۸ - 

(ب) ۱۹ - 

(ج) ۲۰ - 

(ت) ۲۱ - 

الف - مربع
ب - مستطیل
پ - ذوزنقه
ت - لوزی
ث - استوانه
ج - متوازی‌الاضلاع
چ - مثلث
ح - هرم

کاربستن دانش مفهومی

راهنمایی: در مقابل شمارهٔ هر جفت بیان زیر یکی از حرفهای الف، ب، یا پ را مطابق با توضیحات داده شده قرار دهید:

الف - اگر افزایش در مورد اول همراه با افزایش در مورد دوم است.
ب - اگر افزایش در مورد اول همراه با کاهش در مورد دوم است.
پ - اگر افزایش در مورد اول اثر چندانی بر مورد دوم ندارد.

(پ) ۲۲ ــ تعداد موشهای صحرایی در یک منطقه از قطب شمال
تعداد گوزنها در همان منطقه

(الف) ۲۳ ــ تعداد گل‌سنگها در یک منطقه از قطب شمال
تعداد گوزنها در همان منطقه

(الف) ۲۴ ــ درجهٔ حرارت محیط زندگی یک جانور
درجهٔ حرارت بدن همان جانور

(ب) ۲۵ ــ مقدار کربناتهای محلول در آب یک رودخانه
تعداد صدفها در همان رودخانه

(ب) ۲۶ ــ فشردگی خاک در یک منطقه
مقدار جذب آب به وسیلهٔ خاک پس از یک بار باران سنگین

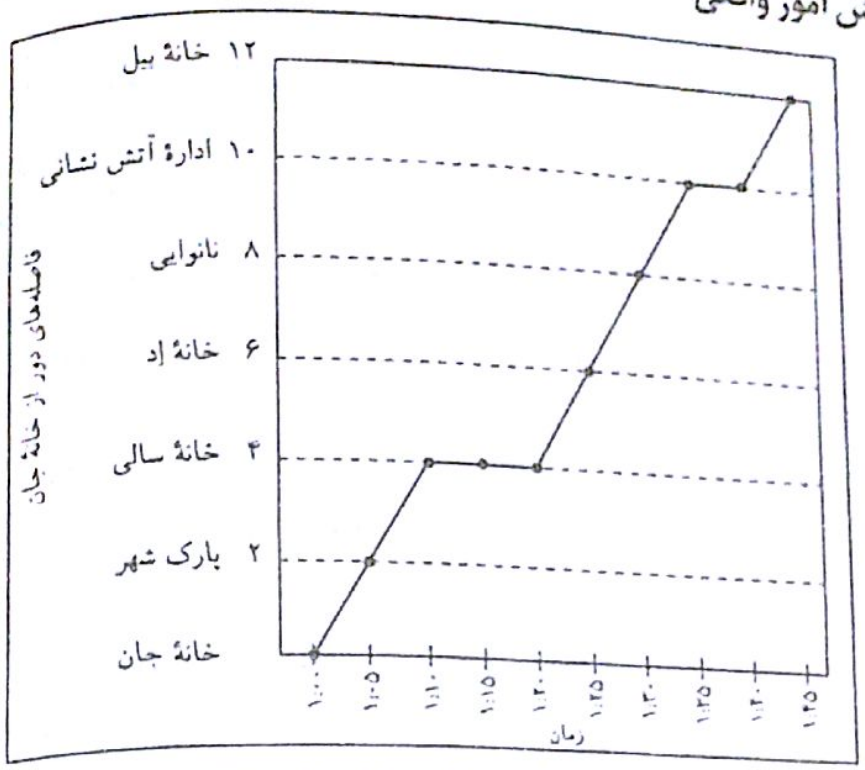
(الف) ۲۷ ــ فراوانی آتش‌سوزی در یک جنگل کاج
تعداد درختهای سپیدار همان جنگل

(الف) ۲۸ ــ ارتفاع محیط زندگی یک جانور
میزان تقسیم گلبولهای سرخ خون آن جانور

(بلوم و همکاران، ۱۹۵۶، ترجمهٔ سیف و علی‌آبادی، ۱۳۶۸، ص ۱۴۴)

بلوم و همکاران (۱۳۶۸) در معرفی سؤال بالا گفته‌اند "دانش‌آموز باید اصول بوم‌شناسی را به یاد آورد و اصول مناسب را در موقعیتهایی که در سؤال آمده به کار بندد.... موقعیتها واقعی هستند، اما احتمالاً دانش‌آموز با آنها آشنایی ندارد" (ص ۱۴۴).

تحلیل دانش امور واقعی



این شکل نشان می‌دهد که جان در ساعت ۱ خانه‌اش را ترک کرده و در ساعت ۱:۴۵ به خانهٔ بیل رسیده است. شکل نشان می‌دهد که جان در زمانهای مختلف در کجاها بوده است.

راهنمایی: بیانات شماره گذاری شدهٔ زیر نظرهای دانش‌آموزان مختلف را دربارهٔ نمودار بالا نشان می‌دهد. هر یک از آنها را بخوانید و تعیین کنید که اطلاعات موجود در نمودار با آن بیانات همخوانی دارند یا نه. با توجه به توضیحات زیر پاسخ دهید:

الف – اگر اطلاعات موجود در نمودار با بیان مورد نظر همخوانی دارد.
ب – اگر اطلاعات موجود در نمودار با بیان مورد نظر مغایر است.
پ – اگر اطلاعات موجود درنمودار نه با بیان موردنظر همخوانی دارد و نه با آن مغایر است.

(الف) ۲۹ – جان فاصلهٔ بین پارک شهر و خانه‌اش را دویده یا سریع طی کرده.
(الف) ۳۰ – جان در سر راه به خانهٔ بیل در خانهٔ سالی توقف کرده.
(ب) ۳۱ – جان در سر راه به خانهٔ بیل در خانهٔ اِد توقف کرده.
(ب) ۳۲ – جان پیش از رسیدن به خانهٔ بیل در نانوایی برای خرید توقف کرده.
(پ) ۳۳ – جان پس از گذشتن از خانهٔ سالی سریع‌تر از زمانی که به خانهٔ خود رسید راه رفت.

(نیتکو، ۲۰۰۱، ص ۱۷۵)

خلاصهٔ فصل

آزمون جورکردنی از تعدادی تمرین تشکیل می‌شود و هر تمرین شامل سه قسمت زیر است: (۱) راهنمای انجام دادن تمرین یا جواب دادن به پرسشها، (۲) فهرست پرسشها یا محرکها، و (۳) فهرست پاسخها. سؤالهای جورکردنی دارای انواع مختلفی هستند که ما در این فصل انواع مهم آنها یعنی نوع معمولی، نوع رده‌بندی یا نوع فهرست اصلی را توضیح دادیم.

در نمونه‌های معمولی سؤالهای جورکردنی، فهرستی از پرسشها و فهرستی از پاسخها در یک تمرین آورده می‌شود و معمولاً تعداد پاسخها از تعداد پرسشها بیشتر است. در سؤالهای جورکردنی نوع رده‌بندی یا نوع فهرست اصلی تعداد پرسشها می‌تواند از تعداد پاسخها بیشتر باشد، زیرا در این سؤالها از یک پاسخ می‌توان بیشتر از یک بار استفاده کرد.

برای تهیهٔ سؤالهای جورکردنی، صاحب نظران قواعدی را پیشنهاد داده‌اند. ما در این فصل مهم‌ترین آنها را توضیح داده‌ایم. در ضمن توضیح این قواعد، مثالهای مختلفی از انواع آزمونها نیز آورده‌ایم.

در رابطه با امتیازها و محدودیتهای آزمونهای جورکردنی، نظرهای مثبت و منفی چندی اظهار شده است. از جمله نظرهای مثبت در رابطه با سؤالهای جورکردنی، گفته شده است که:

۱. نمونهٔ بزرگی از محتوا و هدفهای درسی را در یک آزمون جورکردنی می‌توان قرار داد.
۲. تصحیح پاسخها به سرعت و سهولت امکان‌پذیر است.
۳. امکان حدس زدن پاسخ درست زیاد نیست.
۴. در سنجش هدفها با سؤالهای جورکردنی تا سطح بالاتری از آنچه سؤالهای صحیح ـ غلط می‌سنجند می‌توان رفت.

برای آزمونهای جورکردنی محدودیتهایی نیز ذکر شده است. محدودیتهایی که برای این آزمونها عنوان شده به قرار زیرند:

۱. سؤالهای جورکردنی ممکن است حفظ طوطی‌وار و یادگیری سطحی را تشویق کنند.
۲. بعضی وقتها پیدا کردن مجموعه‌ای پرسش متجانس که بشود برای آن سؤالهای چندگزینه‌ای نوشت دشوار است.
۳. همچنین مبهم بودن راهنمای سؤال، طولانی بودن فهرست پرسشها و پاسخها، و مبهم بودن بیان پرسشها می‌تواند از محدودیتهای این سؤالها باشد.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. با افزودن بر طول فهرست پرسشها و پاسخهای یک تمرین جورکردنی می‌توان تشخیص پرسشها و پاسخها را افزایش داد.

الف ـ صحیح

ب ـ غلط

۲. در تهیهٔ سؤالهای جورکردنی کدام قاعدهٔ زیر ذکر نشده است؟

الف ـ هم مطالب پرسشها و هم مطالب پاسخها متجانس باشند.

ب ـ هم طول فهرست پرسشها و هم طول فهرست پاسخها کوتاه باشند.

پ ـ مطالب بیشتری در فهرست پاسخها نوشته شوند.

ت ـ علاوه بر توضیحات کتبی، در صورت لزوم توضیحات شفاهی نیز داده شود.

ث ـ هر یک از پاسخها باید برای تمام پرسشها درست جلوه کند.

۳. در یک سؤال جورکردنی ۸ پرسش آمده است که باید با ۹ پاسخ دربارهٔ تاریخها، شهرها و کشورها جور شوند. اِشکال این سؤال در چیست؟

الف ـ تعداد زیاد پرسشها

ب ـ تعداد زیاد پاسخها

پ ـ تجانس زیاد پاسخها

ت ـ تجانس کم پاسخها

۴. ویژگیهای سؤالهای جورکردنی را توضیح دهید.

۵. با ذکر مثال، انواع سؤالهای جورکردنی را توضیح دهید.

۶. امتیازها و محدودیتهای سؤالهای جورکردنی را توضیح دهید.

۷. سؤالهای جورکردنی و صحیح ـ غلط را از لحاظ موارد کاربرد با هم مقایسه کنید.

۸. در رابطه با طرح سنجش خود که در تمرین مربوط به فصل ۶ تهیه کردید، برای هدفهایی که با سؤالهای جورکردنی قابل سنجش‌اند یک آزمون جورکردنی بسازید.

۹

سنجش یادگیری به کمک آزمون چندگزینه‌ای

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمون چندگزینه‌ای را تعریف کند.
۲. ویژگیهای سؤالهای چندگزینه‌ای را توضیح دهد.
۳. انواع سؤالهای چندگزینه‌ای را شرح دهید و برای هریک مثالی بزنند.
۴. قواعد تهیهٔ سؤالهای چندگزینه‌ای را توضیح دهد.
۵. منابعی را که برای نوشتن گزینه‌های انحرافی می‌توان از آنها کمک گرفت نام ببرد.
۶. امتیازها و محدودیتهای سؤالهای چندگزینه‌ای را توضیح دهد.
۷. موارد استفادهٔ آزمونهای چندگزینه‌ای را توضیح دهد.
۸. برای هدفهای آموزشی مختلف سؤالهای چندگزینه‌ای بنویسد.
۹. با توجه به ویژگیهای سؤالهای چندگزینه‌ای، سؤالهای معیوبی را که در اختیار شما گذاشته می‌شوند اصلاح کند.

موضوع فصل حاضر آزمون چندگزینه‌ای است. آزمون چندگزینه‌ای نوع سومی از روشهای سنجش گزینش پاسخ یا بسته پاسخ است. ما دو نوع دیگر آن یعنی آزمون صحیح ـ غلط و آزمون جورکردنی را در دو فصل قبلی این کتاب توضیح دادیم. آزمونهای چندگزینه‌ای از آزمونهای صحیح ـ غلط و جورکردنی معروف‌تر و پراستفاده‌ترند. در دنبالهٔ مطالب به بحث مفصل این آزمونها توجه کنید.

تعریف و انواع آزمونهای چندگزینه‌ای

هر آزمون چندگزینه‌ای[1] از تعدادی سؤال درست می‌شود و هر سؤال چندگزینه‌ای از سه قسمت اصلی زیر تشکیل می‌یابد: (۱) **تنهٔ سؤال**[2]، (۲) **گزینهٔ کلید**[3]، و (۳) **گزینه‌های انحرافی**[4]. تنهٔ سؤال متن یا صورت سؤال را تشکیل می‌دهد و دربرگیرندهٔ مسئله یا موضوعی است که سؤال باید آن را بسنجد. گزینهٔ کلید همان پاسخ سؤال است که آزمون‌شوندهٔ آگاه از موضوع سؤال آن را برمی‌گزیند. گزینه‌های انحرافی پیشنهادهای غلطی هستند که نقش آنها منحرف کردن آزمون‌شوندگان ناآگاه از موضوع سؤال است. شکل ۱ـ۹ قسمتهای اصلی یک سؤال چندگزینه‌ای را نشان می‌دهد.

اصطلاحات بالا متداول‌ترین اصطلاحات مورد استفاده در سؤالهای چندگزینه‌ای هستند. با این حال، در متون مختلف از اصطلاحات دیگری نیز استفاده می‌شود که آشنایی با آنها مفید به نظر می‌رسد. برای گزینهٔ کلید "پاسخ کلید"، "پاسخ درست" یا صرفاً "کلید"؛ برای پاسخهای انحرافی "پاسخهای پیشنهادی"، "انتخابها"؛ و برای گزینه‌های انحرافی "گزینه‌های ردگم‌کنی" استفاده می‌شوند.

تعداد گزینه‌های سؤالهای چندگزینه‌ای (گزینهٔ درست و گزینه‌های انحرافی) از ۲ تا ۵ تغییر می‌کند. از لحاظ نظری، هرچه تعداد گزینه‌ها بیشتر باشد امکان حدس زدن کمتر است. اما به سبب اینکه پیدا کردن بیش از سه گزینهٔ انحرافی کار دشواری است، آزمونهای چهارگزینه‌ای که در آنها امکان حدس زدن جواب درست یک در چهار است شهرت بیشتری کسب کرده‌اند.

تنهٔ سؤال	بیت زیر چه معنی می‌دهد؟ ز سُم ستوران در آن پهن دشت زمین شد شش و آسمان گشت هشت
گزینهٔ کلید	الف – بر اثر گرد و خاک حاصل از سم چهارپایان یک لایه از زمین به آسمان چسبید.
گزینه‌های انحرافی	ب – به دلیل آشفتگی ناشی از هرج و مرج میدان نبرد همه به اشتباه محاسبه گرفتار شدند. پ – نیروی خارج از تصوری که رزمندگان به نمایش گذاشتند حقارت زمین در برابر عظمت آسمانها را نمایان‌تر کرد. ت – برخورد سهمگین اسبان واسب‌سواران بایکدیگر چشمان بینندگان را تیره وتار کرد.

شکل ۱ـ۹ قسمتهای مختلف یک سؤال چندگزینه‌ای

1. multiple-choice 2. stem 3. keyed option 4. distractors

هر چند که همهٔ سؤالهای آزمونهای چندگزینه‌ای از سه قسمت اصلی نشان داده در شکل ۱ـ۹ درست می‌شوند، اما انواع مختلفی دارند و معلم علاقه‌مند به تهیهٔ سؤالهای ابتکاری می‌تواند از آنها بهره بگیرد. در زیر گونه‌های شناخته شدهٔ آزمونهای چندگزینه‌ای را توضیح می‌دهیم.

(۱) نوع پرسشی

در سؤالهای نوع پرسشی آزمونهای چندگزینه‌ای، تنهٔ سؤال را یک جملهٔ پرسشی تشکیل می‌دهد و گزینه‌ها (پیشنهادی و کلید) پاسخهای آن پرسش هستند.

مثال

ژان والژان اولین‌بار که محکوم به اعمال شاقه شد به چیزی دزدیده بود؟

الف ـ شمعدانهای اسقف

(×) ب ـ یک قرص نان

پ ـ چند قطعه چوب

ت ـ گاو یک پیرزن

ث ـ پوشش سکوی خطابهٔ کلیسا

(بلوم و همکاران، ۱۹۵۶، ترجمهٔ سیف و علی‌آبادی، ۱۳۶۸، ص ۸۸)

(۲) نوع ناتمام

در سؤالهای نوع ناتمام آزمونهای چندگزینه‌ای، تنهٔ سؤال یک جملهٔ ناتمام است و گزینه‌ها (پیشنهادی و کلید) تکمیل‌کنندهٔ آن هستند.

مثال

مهم‌ترین ویژگی هدفهای آموزشی رفتاری این است که ـــــ .

(×) الف ـ برحسب عملکرد قابل اندازه‌گیری نوشته می‌شوند

ب ـ مهارتهای سطح بالای یادگیری را شامل می‌شوند

پ ـ مطالب جزئی محتوای درس را دربر می‌گیرند

ت ـ دانش‌آموزان به راحتی آنها را می‌آموزند

(۳) نوع تنها گزینهٔ درست

در این نوع سؤالها، همهٔ گزینه‌ها بجز گزینهٔ کلید به طور قطع غلط هستند. بنابراین، از این سؤالها بیشتر در ریاضیات و علوم می‌توان استفاده کرد.

مثال

چه کسی چرخ خیاطی را اختراع کرد؟

الف ـ فولتون

(×) ب ـ هوو

پ ـ سینگر

ت ـ وایت

ث ـ ویتنی

(نیتکو، ۲۰۰۱، ص ۱۴۷)

نوع بهترین گزینه

در این نوع سؤالها، همهٔ گزینه‌ها به طریقی درست هستند، اما یکی از آنها از همه درست‌تر است. این نوع سؤالهای چندگزینه‌ای در زمینه‌های علوم‌انسانی و علوم اجتماعی که برای سؤالها معمولاً جواب قطعی یافت نمی‌شود مفیدند.

مثال

بنا به تعریف، می‌توان گفت که یادگیری عبارت است از ـــــ .

الف ـ کسب اطلاعات تازه در نتیجهٔ آموزش

(×) ب ـ ایجاد تغییرات نسبتاً پایدار در توان رفتاری یادگیرنده

پ ـ ایجاد تغییرات مطلوب و موردنظر در رفتار یادگیرنده

ت ـ توانایی به خاطر سپردن و یادآوری تجارب پیشین

نوع منفی

در این نوع سؤالها، تنهٔ سؤال به صورت منفی بیان می‌شود و همهٔ گزینه‌ها بجز یکی از آنها درست هستند.

مثال

کدام مورد زیر از قواعد تهیهٔ آزمونهای چندگزینه‌ای نیست؟

الف ـ مربوط بودن همهٔ گزینه‌ها به موضوعی واحد

ب ـ سنجش تنها یک هدف با هر سؤال

(×) پ ـ قرار دادن حداکثر دو مطلب در هر سؤال

ت ـ استفاده نکردن از گزینهٔ "همهٔ موارد بالا"

یکی از مشکلات تهیهٔ سؤالهای چندگزینه‌ای این است که بعضی وقتها تهیهٔ تعدادی گزینهٔ انحرافی غلط که به ظاهر درست جلوه کنند بسیار دشوار است. در این گونه موارد، چنانچه تهیهٔ گزینه‌های درست آسان‌تر باشد می‌توان از سؤالهای نوع منفی که در آنها همهٔ گزینه‌ها بجز یکی درست‌اند استفاده کرد، مانند نمونهٔ بالا.

نوع جایگزینی

در سؤالهای نوع جایگزینی، در متن یا تنهٔ سؤال غلطهایی گنجانیده می‌شود و آزمون شونده باید آنها را با گزینه‌های پیشنهادی جایگزین کند.

مثال

"به درستی که از نیروهای آموزش و پرورش باید به طور کامل استفاده شود تا جوانان از ماهیت واقعی خطراتی که دموکراسی را تهدید می‌کنند آگاه شوند **زیرا**(۱) *هیچ جای دیگری بجز آموزشگاه نمی‌تواند* **فرصتهای خوب**(۲) **یا بهتری** *در اختیار جوانان قرار دهد تا* **به طور واقعی**(۳) *به کُنه مسائل پی ببرند."*

مطلب بالا را بخوانید و با توجه به سؤالهایی که در زیر آمده برای قسمتهایی که با شماره مشخص شده‌اند جایگزین تعیین کنید.

۱ ـ
(×) الف ـ ، زیرا
ب ـ ؛ زیرا
پ ـ ـ زیرا
ت ـ هیچ علامتی قبل از زیرا لازم نیست.

۲ ـ
الف ـ فرصتهای خوب یا خوب‌تری
ب ـ فرصتهای بهتر یا بهتری
پ ـ فرصتهای بهتر یا بدتری
(×) ت ـ فرصتهای بهتری

۳ ـ
الف ـ به صورت منطقی
(×) ب ـ به طور منطقی
پ ـ به طور عقلانی
ت ـ به صورت علمی

(نیتکو، ۲۰۰۱، ص ۱۴۷)

۷ نوع ناتمام پیشنهادی

در این نوع سؤالها، آزمون شونده با توضیحاتی که در متن سؤال می‌آید راهنمایی می‌شود و او باید با توجه به آن راهنماییها جواب درست سؤال را پیدا کند.

مثال

به سیبی که مزه‌اش ترش و شیرین است چه نوع سیبی گفته می‌شود؟ (جواب یک کلمهٔ سه حرفی است). کدام گزینهٔ زیر حرف اول آن کلمه است؟

الف ـ ت

(×) ب ـ م

پ ـ ش

ت ـ ک

۸ نوع پاسخ ترکیبی

در این نوع سؤالها، جمله‌های یک پاراگراف به هم‌ریخته‌اند و از آزمون‌شونده خواسته می‌شود تا آنها را پشت سر هم مرتب کند.

مثال

جمله‌های زیر را به چه ترتیبی باید پشت سرهم قرار دهیم تا از مجموع آنها یک پاراگراف مطلب معنی‌دار به دست آید؟

جمله‌ها

۱....

۲....

۳....

۴....

۵....

گزینه‌ها:

الف ـ ۱، ۲، ۳، ۴، ۵

ب ـ ۵، ۲، ۳، ۱، ۴

پ ـ ۴، ۲، ۳، ۱، ۵

ت ـ ۵، ۴، ۳، ۲، ۱

قواعد تهیهٔ سؤالهای چندگزینه‌ای

قواعدی که در اینجا برای تهیهٔ سؤالهای چندگزینه‌ای پیشنهاد می‌شوند براساس این طرز فکر تدوین شده‌اند که یک سؤال چندگزینه‌ای باید آزمون شونده را با تکلیفی که هم مهم و هم کاملاً

برای او قابل فهم است مواجه سازد. هر سؤال باید طوری نوشته شود که تنها به وسیلهٔ کسانی که هدف مورد اندازه گیری (جواب سؤال) را می‌دانند به درستی پاسخ داده شود، و کسانی که تسلط کامل بر مطلب ندارند نتوانند به آن پاسخ درست بدهند.

۱. **هر سؤال را به گونه‌ای انتخاب کنید که یک موضوع مهم یا یک هدف آموزشی را اندازه بگیرد.** در طرح هر سؤال، نویسندهٔ سؤال باید از خود بپرسد که "آیا دانستن و ندانستن پاسخ سؤال در صلاحیت آزمون‌شونده، از لحاظ مطلب مورد آزمایش، تأثیر دارد یا نه؟" اگر جواب این سؤال مثبت بود آن‌گاه باید به گنجانیدن موضوع در سؤال اقدام کند. بنابراین، هر سؤال باید یک هدف آموزشی را بسنجد و شامل بخش مهمی از محتوای درس باشد.

نیتکو (۱۹۸۳) در این باره می‌گوید ابتدا یک هدف آموزشی یا یک مطلب مهم را انتخاب کنید و تصمیم بگیرید که برای آن چند سؤال تهیه خواهید کرد. بعد برای نوشتن هر سؤال اقدامات زیر را انجام دهید.

الف ـ با نوشتن یک جملهٔ سؤالی یا یک جملهٔ ناتمام تنهٔ سؤال را بنویسید.
ب ـ جواب سؤال یا بخش ناتمام جمله را به صورت گزینهٔ کلید بنویسید.
پ ـ گزینه‌های انحرافی را به صورتی بنویسید که نظر دانش‌آموزان نامطلع را به خود جلب کنند.

به مطلب و هدف آموزشی زیر و سؤال نمونه‌ای که برای آن آماده شده است توجه کنید.

مطلب درسی: اگر زمان آزمون محدود باشد، یک آزمون چندگزینه‌ای از یک آزمون تشریحی نمونهٔ بهتری از هدفها و محتوا را به دست می‌دهد.

هدف آموزشی: دانشجو بتواند با ذکر دلیل توضیح دهد که، برای استفاده در یک زمان محدود، آزمونهای چندگزینه‌ای از آزمونهای تشریحی مناسب‌ترند.

سؤال نمونه:

زمانی که وقت آزمون محدود و مقدار مطلبی که باید سنجش شود گسترده است، سؤالهای چندگزینه‌ای از سؤالهای تشریحی بهترند، زیرا سؤالهای چندگزینه‌ای ــــ .

الف ـ راحت‌تر تصحیح می‌شوند
ب ـ امکان کمی برای حدس زدن به آزمون‌شونده می‌دهند
(×) *پ ـ نمونهٔ بزرگ‌تری از مطالب را شامل می‌شوند*
ت ـ تواناییهای سطح بالای یادگیرندگان را می‌سنجند

۲. **بیشتر از یک هدف یا یک مطلب در هر سؤال قرار ندهید.** هر سؤال باید تنها به یک مطلب یا

یک هدف آموزشی مربوط باشد. اگر بیش از یک مطلب یا یک هدف در سؤال مطرح شود، نه تنها سؤال را پیچیده می‌کند بلکه اگر آزمون‌شونده در پاسخگویی به سؤال موفق نباشد معلوم نیست که کدام یک از مطالب را ندانسته و علت عدم موفقیت او چه بوده است.

۳. **سؤالها را کاملاً روشن و در حد درک آزمون‌شوندگان بنویسید.** استفاده از جمله‌ها و عبارتهای مبهم در سؤالهای چندگزینه‌ای ممکن است باعث شود که آزمون‌شوندگان با وجود اینکه جواب درست را می‌دانند به علت ابهام و دشواری متن سؤال نتوانند به آن جواب درست بدهند. به ویژه سؤالهایی که برای دانش‌آموزان سطحهای پایین‌تر تحصیلی نوشته می‌شوند باید با زبانی ساده و در خور درک آنان باشند. سؤالهایی که در آنها کلمه‌ها و عبارتهای دشوار و غیرقابل درک دانش‌آموزان به کار می‌روند نمی‌توانند به خوبی توانایی موردنظر را سنجش کنند، زیرا اگر دانش‌آموزان در جواب دادن به این نوع سؤالها موفق نباشد معلوم نمی‌شود که علت شکست آنان ندانستن جواب سؤال یا ناتوانی در درک سؤال است. میزان دشواری یک سؤال باید به محتوا و هدف آموزشی مورد سنجش مربوط باشد نه به کلمه‌ها و عبارتهایی که در آن به کار می‌روند. تنها در مواردی که سؤال برای اندازه‌گیری توانایی خواندن و فهمیدن یا توانایی کلامی آزمون‌شوندگان نوشته می‌شود استفاده از ساختمان پیچیدهٔ دستوری و لغات و عبارات دشوار مجاز است.

۴. **از تکرار مطالب در گزینه‌ها خودداری کنید.** تا آنجا که ممکن است پاسخها را مختصر و کوتاه بنویسید. ضمناً سعی کنید مطالب را یک بار آن هم در متن یا تنهٔ سؤال بیاورید و در صورت امکان از تکرار آن در پاسخها خودداری نمایید.

مثال

سؤال ضعیف:

در آزمونهای عینی، اصطلاح عینی ــــــ .

الف ـ به روش مشخص کردن نتایج یادگیری مربوط است

ب ـ به روش انتخاب محتوای آزمون مربوط است

(×) پ ـ به روش تصحیح جوابها مربوط است

ت ـ به روش ارائهٔ مطالب مربوط است

سؤال بهتر:

در آزمونهای عینی، اصطلاح عینی به روش ــــــ مربوط است.

الف ـ مشخص کردن نتایج یادگیری

ب ـ انتخاب محتوای آزمون

(×) پ ـ تصحیح جوابها

ت ـ ارائهٔ مطالب

سؤال باز هم بهتر:

در آزمونهای عینی، اصطلاح عینی به چه چیزی مربوط است؟

الف ـ مشخص کردن نتایج یادگیری

ب ـ انتخاب محتوای آزمون

(×) پ ـ تصحیح جوابها

ت ـ ارائهٔ مطالب

فایدهٔ رعایت قاعدهٔ بالا و جلوگیری از تکرار مطالب در گزینه‌ها صرفه‌جویی در وقت خواندن برای آزمون‌شوندگان و نیز صرفه‌جویی در وقت نوشتن و تایپ کردن برای طراح سؤال و تکثیرکنندهٔ آزمون است.

۵. **مطالب اصلی سؤال را به طور کامل در تنهٔ سؤال بنویسید.** تنهٔ سؤال باید شامل مطالب مهم سؤال باشد، به گونه‌ای که آزمون‌شونده، پس از خواندن آن و پیش از مراجعه به گزینه‌ها، اندیشهٔ اصلی مورد سؤال را درک کند و بداند که سؤال به چه مطلبی مربوط است. مشکل سؤالهایی که این نکته را در نظر نمی‌گیرند این است که آزمون‌شونده، پیش از خواندن گزینه‌ها، متوجه نخواهد شد که سؤال دربارهٔ چه مطلبی نوشته شده است. یک راه کسب اطمینان از این امر آن است که روی گزینه‌ها را بپوشانید و تنهٔ سؤال را به تنهایی بخوانید تا متوجه شوید که مسئله یا مطلب به طور کامل بیان شده است یا نه. اگر جواب شما منفی است در تنهٔ سؤال تجدید نظر کنید.

مثال

سؤال ضعیف:

جدول مشخصات ـــــ .

الف ـ نشان می‌دهد که چگونه می‌توان از آزمون برای بهبود امر آموزش و یادگیری استفاده کرد

(×) ب ـ نمونهٔ بهتری از محتوای درس و هدفهای آموزش به دست می‌دهد

پ ـ روش مورد استفاده در تصحیح پاسخهای آزمون را مشخص می‌کند

ت ـ عینیت آزمون را افزایش می‌دهد

سؤال بهتر:

مهم‌ترین فایدهٔ استفاده از جدول مشخصات در تهیهٔ آزمونهای پیشرفت تحصیلی آن است که ـــــ .

الف ـ آموزش و یادگیری را بهبود می‌بخشد

(×) ب ـ نمونهٔ بهتری از محتوای درس و هدفهای آموزش به دست می‌دهد

پ ـ روش مورد استفاده در تصحیح پاسخهای آزمون را مشخص می‌کند

ت ـ عینیت آزمون را افزایش می‌دهد

۶. بکوشید تا همهٔ گزینه‌های یک سؤال متجانس و به موضوع واحدی مربوط باشند. همهٔ گزینه‌های یک سؤال باید از لحاظ محتوا همگون یا متجانس باشند نه اینکه هر یک از آنها به مطلب جداگانه‌ای اشاره کند. در چنین حالتی سؤال از صورت چندگزینه‌ای خارج شده و هر یک از گزینه‌ها به صورت یک سؤال صحیح ـ غلط درمی‌آید که از دیگر گزینه‌ها مستقل است. نمونهٔ معمول این‌گونه سؤالها سؤالی است که تنهٔ آن این جمله را شامل می‌شود: "کدام یک از مطالب یا بیانات زیر درست است؟" و گزینه‌های آن هر یک به مطلب جداگانه‌ای اشاره می‌کند.

مثال

سؤال ضعیف:

کدام یک از موارد زیر درست است؟

(×) الف ـ سؤالهای انشایی بهترین نوع سؤال برای ارزشیابی هدفهای طبقهٔ آفریدن هستند.

ب ـ تهیهٔ جدول مشخصات برای آزمونهای انشایی مفید نیست.

پ ـ سؤالهای عینی را تنها در کلاسهای دانشگاه می‌توان به اجرا درآورد.

ت ـ هدفهای رفتاری از دیرباز مورد استفادهٔ معلمان قرار داشته‌اند.

سؤال بهتر:

در رابطه با استفاده از سؤالهای انشایی و ارزشیابی هدفهای مختلف آموزشی، کدام مورد زیر درست است؟

(×) الف ـ بهترین نوع سؤال برای ارزشیابی هدفهای طبقهٔ آفریدن هستند.

ب ـ برای ارزشیابی هدفهای هیچ یک از طبقات حوزهٔ شناختی، مفید نیستند.

پ ـ بین سؤالهای انشایی و عینی از لحاظ ارزشیابی هدفهای گوناگون چندان تفاوتی وجود ندارد.

ت ـ بهتر است از سؤالهای انشایی به ندرت استفاده شود.

تجانس گزینه‌ها یکی از راههای کنترل سطح دشواری سؤالهاست. قاعدهٔ کلی این است که هر چه گزینه‌ها متجانس‌تر باشند به همان نسبت سؤال دشوارتر است. البته چنان که در مورد سؤالهای جورکردنی نیز گفته شد تجانس یک امر نسبی است. گزینه‌هایی که برای یک گروه دانش‌آموز متجانس به نظر می‌رسند ممکن است برای گروه دیگر (معمولاً گروه سطح بالاتر) از تجانس کمتری برخوردار باشند.

۷. سؤال را طوری بنویسید که پاسخ درست تنها پاسخ درست یا قطعاً درست‌ترین پاسخ باشد. قبلاً گفتیم که گزینهٔ درست هر سؤال یا باید تنها گزینهٔ درست در میان سایر گزینه‌ها باشد یا از گزینه‌های انحرافی به طور قطع درست‌تر یا بهتر باشد.

مثال

سؤال ضعیف:

یک پسربچهٔ کوررنگ، کوررنگی خود را از چه کسی به ارث می‌برد؟

الف ـ پدر

ب ـ مادر

پ ـ پدرِ مادر

ت ـ پدرِ پدر

ث ـ نیاکان دور

سؤال بهتر:

یک پسربچهٔ کوررنگ به احتمال بسیار زیاد کوررنگی خود را از چه کسی به ارث می‌برد؟

الف ـ پدر

(×) ب ـ مادر

پ ـ پدرِ پدر

ت ـ مادرِ مادر

ث ـ نیاکان دور

(ثرندایک و هیگن، ۱۹۶۹)

ثرندایک و هیگن دربارهٔ سؤالهای بالا اظهار می‌دارند که گرچه در سؤال ضعیف گزینهٔ ب درست است، گزینه‌های پ و ت و ث هم کاملاً غلط نیستند. در سؤال تجدیدنظر شده (سؤال بهتر) اضافه شدن کلمهٔ احتمالاً گزینه ب را از دیگر گزینه‌ها درست‌تر می‌سازد.

۸. **برای اندازه‌گیری فرایندهای پیچیدهٔ ذهنی از موقعیتهای تازه استفاده کنید.** باید به خاطر داشت که مسائل و موقعیتهایی که در کلاس و کتاب برای آموزش فعالیتهای ذهنی چون کاربستن، ترکیب، و از این قبیل به کار رفته‌اند دیگر نمی‌توانند برای سنجش این تواناییها مورد استفاده قرار گیرند. اگر این مطالب در سؤالها تکرار شوند، تنها دانش را می‌سنجند. بنابراین، در طرح سؤالهایی که به منظور سنجش تواناییهای بالاتر از سطح دانش به کار می‌روند از موقعیتها و مثالهای تازه‌ای که با مثالهای کتاب و موقعیتهای به کار رفته در کلاس متفاوت‌اند استفاده کنید. ثرندایک و هیگن (۱۹۶۹) به عنوان نمونه‌ای برای روشن کردن این مورد، دو سؤال زیر را معرفی کرده‌اند. سؤال اول یک سؤال در سطح یادآوری است که محفوظات یادگیرنده را می‌سنجد، ولی سؤال دوم یک سؤال کاربردی است که در آن یک موقعیت تازه ارائه شده است.

سؤال سطح یادآوری:

کدام یک از موارد زیر وقتی که در بدن مصرف می‌شود بیشترین کالری را تولید می‌کند؟

(×) الف ـ یک گرم چربی

ب ـ یک گرم شکر

پ ـ یک گرم نشاسته

ت ـ یک گرم پروتئین

سؤال کاربردی:

کدام‌یک از موارد زیر اگر ازغذاهای روزانه حذف شود بیشترین مقدارکاهش کالری را به دنبال خواهد داشت؟

(×) الف ـ یک قاشق کره

ب ـ یک قاشق شکر

پ ـ یک قرص نان

ت ـ یک تکه‌مرغ آب‌پز

در زیر به سؤال دیگری که برای اندازه گیری توانایی کاربستن اصول نوشته شده و در آن یک موقعیت جدید مطرح گردیده است توجه کنید:

فرض کنید ۴ لیتر رنگ برای رنگ کردن مساحتی برابر ۴۰ مترمربع کافی باشد. چند لیتر رنگ برای رنگ کردن ۴ دیوار اتاقی با طول و عرض ۳ × ۱۰ متر لازم است؟

الف ـ ۶

ب ـ ۸

(×) پ ـ ۱۲

ت ـ ۱۸

۹. گزینه‌های انحرافی را طوری بنویسید که توجه آزمون‌شوندگان بی‌اطلاع از موضوع سؤال را به خود جلب کنند. هدف از گزینه‌های انحرافی در سؤالهای چندگزینه‌ای این است که توجه آزمون‌شوندگان بی‌اطلاع از موضوع مورد سؤال را به خود جلب کند. این گزینه‌ها باید ظاهری منطقی داشته باشند و در یک بررسی سطحی درست جلوه کنند. یعنی نباید آشکارا غلط باشند، آن‌گونه که هر کسی بتواند غلط بودن آنها را تشخیص دهد.

مثال

سؤال ضعیف:

امیرکبیر صدراعظم کدام یک از شاهان زیر بود؟

الف ـ شاه عباس

ب ـ سلطان محمود

(×) پ ـ ناصرالدین شاه

ت ـ نادرشاه

سؤال بهتر:

امیرکبیر صدراعظم کدام یک از شاهان زیر بود؟

الف ـ مظفرالدین شاه

ب ـ محمدعلیشاه

(×) پ ـ ناصرالدین شاه

ت ـ احمد شاه

برای نوشتن گزینه‌های انحرافی پیشنهادهای زیر را از لین و گرانلاند (۲۰۰۰) مورد توجه قرار دهید:

- از غلطهای معمول دانش‌آموزان استفاده کنید.
- از کلماتی که مهم به نظر می‌رسند (مانند با اهمیت، معنی‌دار، و دقیق) و به تنهٔ سؤال ربط پیدا می‌کنند استفاده کنید. اما در این کار افراط ننمایید.
- از کلماتی که با تنهٔ سؤال تداعی کلامی دارند استفاده کنید (مانند، سیاست‌مدار، سیاسی).
- از زبان کتاب درسی و عبارتهای دیگری که ظاهری درست‌نما دارند استفاده کنید.
- از پاسخهای غلطی که از کج فهمی‌ها و بی‌دقتی‌های دانش‌آموزان ناشی می‌شوند استفاده کنید (مثلاً اینکه فراموش می‌کنند که دسیمتر را به سانتیمتر تبدیل نمایند).
- از گزینه‌های انحرافی که از لحاظ محتوا متجانس و شبیه به گزینهٔ درست هستند استفاده کنید.
- از گزینه‌های انحرافی که در شکل ظاهر با تنهٔ سؤال موازی و از لحاظ دستوری با آن هماهنگی دارند استفاده کنید.
- گزینه‌های انحرافی را از لحاظ طول، واژگان، ساختار جمله، و پیچیدگی موضوع شبیه به گزینهٔ درست بنویسید.

احتیاط: گزینه‌های انحرافی باید پاسخ‌دهندگان بی‌اطلاع از موضوع سؤال را منحرف کنند، اما نباید برای پاسخ‌دهندگان مسلط بر موضوع سؤال گمراه کننده باشند.

۱۰ ـ گزینه‌های هر سؤال را طوری بنویسید که از نظر دستوری و جمله‌بندی به نحو درست مکمل متن سؤال باشند. یکی از معایب سؤالهای چندگزینه‌ای افراد تازه کار این است که بین تنهٔ سؤال و گزینه‌های پیشنهادی از لحاظ دستوری هماهنگی وجود ندارد، و این ناهماهنگی در رابطه با گزینه‌های غلط بیشتر به چشم می‌خورد. در نتیجه، این امر به صورت یک راهنما برای

آزمون‌شونده درمی‌آید و متوجه می‌شود که گزینه‌های غلط آنهایی هستند که از نظر دستوری تنهٔ سؤال را کامل نمی‌کنند. بنابراین، سعی کنید سؤالها را طوری بنویسید که همهٔ گزینه‌ها از لحاظ دستوری با تنهٔ سؤال هماهنگی داشته باشند.

این نکته به ویژه در رابطه با سؤالهایی که متن آنها نیمه‌تمام و پاسخها مکمل آن هستند باید دقت رعایت شود. همچنین در سؤالهایی که آخرین گزینه‌های آنها **همهٔ موارد بالا** یا **هیچ یک از موارد بالا** است غالباً این مشکل دیده می‌شود. به سؤال زیر که از این لحاظ اشکال دارد توجه کنید:

نزدیک به هفتاد درصد جمعیت مردم یک مملکت افرادی هستند که هوشبهر آنان ــــــ.

الف ـ بین ۱۰۰ و ۱۵۰ است

ب ـ بین ۵۰ و ۱۰۰ است

(×) پ ـ بین ۸۵ و ۱۱۵ است

ت ـ هیچ کدام درست نیست

دقت کنید که اگر عبارت "هیچ کدام درست نیست" را در دنبالهٔ تنهٔ سؤال قرار دهیم، یک جملهٔ درست و منطقی به دست نمی‌آید. بنابراین، یا کلاً باید سؤال بالا را از لحاظ شکل ظاهری تغییر دهیم یا گزینهٔ ت را به گونه‌ای تغییر دهیم که از لحاظ دستوری مکمل تنهٔ سؤال باشد. سؤال زیر نمونهٔ دیگری از این سؤالات ضعیف است.

چرا زندگی شهری و روستایی در ناحیهٔ کنار فزر به هم وابسته است؟

(×) الف ـ شهرها بازار فروش محصولات روستایی است

ب ـ وابستگی کارخانه‌ها به محصولات کشاورزی

پ ـ وجود بازارهای هفتگی

ت ـ همهٔ موارد بالا

پیداست که هیچ یک از گزینه‌های سؤال بالا، به ویژه گزینهٔ ت، از نظر دستوری و جمله‌سازی جواب درست و منطقی سؤال نیست. برای اصلاح این سؤال باید شکل کلی سؤال را تغییر داد. باز هم به مثالی دیگر توجه کنید.

سؤال ضعیف:

یک مبدّل برق (ترانسفورماتور) را می‌توان برای ــــــ به کار برد.

الف ـ ذخیره‌سازی برق

(×) ب ـ افزایش ولتاژ یک جریان برق متناوب

پ ـ انرژی برقی را به انرژی مکانیکی تبدیل می‌کند

ت ـ جریان متناوب را به جریان مستقیم تبدیل می‌کند

سؤال بهتر:

یک مبدل (ترانسفورماتور) برقی را می‌توان برای ـــــ به کار برد.

الف ـ ذخیره‌سازی برق

(×) ب ـ افزایش ولتاژ یک جریان متناوب

پ ـ تبدیل انرژی برقی به انرژی مکانیکی

ت ـ تغییر جریان متناوب به جریان مستقیم

(لین و گرانلاند، ۲۰۰۰)

۱۱. از به کار بردن اشاره‌های دستوری نامربوط خودداری کنید. اشاره‌هایی از قبیل زمان افعال، مفرد و جمع بودن آنها، و یا حروف تعریف a و an در زبان انگلیسی را در سؤالها آن گونه که آزمون شونده با توجه به این اشاره‌ها گزینهٔ درست را انتخاب کند به کار نبرید. در ضمن اگر تنهٔ سؤال و گزینهٔ صحیح از بعضی جهات مشابه باشند ممکن است آزمون شونده جواب درست را با توجه به تشابه ظاهری پیدا کند، بدون اینکه واقعاً پاسخ سؤال را بداند. به سؤال زیر که از این جهت دارای اشکال است توجه کنید:

اسناد مالکیت املاک مردم در کجا ثبت می‌شوند؟

(×) الف ـ ادارهٔ کل ثبت اسناد و املاک

ب ـ ادارهٔ کل ثبت احوال

پ ـ وزارت امور اقتصادی و دارایی

ت ـ وزارت علوم و آموزش عالی

به نمونهٔ دیگری از این گونه سؤالهای ضعیف توجه کنید.

پایتخت کشور مکزیک کجاست؟

الف ـ آکاپولکو

ب ـ مریاد

(×) پ ـ شهر مکزیک

ت ـ واشنگتن

۱۲. در سؤالات منفی کلمات منفی را برجسته جلوه دهید. برجسته جلوه دادن کلمات منفی به آزمون‌شونده کمک می‌کند تا متوجه منفی بودن سؤال باشد. اگر آزمون به صورت چاپی در اختیار آزمون‌شونده قرار می‌گیرد، می‌توان از طریق سیاه نوشتن کلمات منفی آنها را برجسته کرد، اما در سؤالهای آزمونهایی که به صورت دستنویس یا تایپی در اختیار آزمون‌شونده قرار می‌گیرند می‌توان زیر کلمات منفی خط کشید. به مثال زیر توجه کنید.

در طرح سؤالهای چندگزینه‌ای تمامی موارد زیر را باید رعایت کرد بجز:

(×) الف ـ کوتاه کردن متن اصلی سؤال یا طولانی کردن پاسخها

ب ـ نوشتن متن اصلی سؤال به صورت یک جملهٔ استفهامی

پ ـ خط کشیدن زیر عبارت منفی یا برجسته کردن آن به طریقی دیگر

ت ـ استفاده از موقعیتهای واقعی در سؤالهای کاربردی

۱۳. از نوشتن سؤالهایی که در آنها متن سؤال منفی و گزینه‌ها هم منفی هستند، یعنی منفی مضاعف، احتراز کنید. سؤالهایی که در آنها هم متن سؤال منفی است و هم گزینه‌ها با عبارات و کلمات منفی نوشته می‌شوند بسیار دشوار هستند، و خواندن آنها آزمون شونده را گیج و سردرگم می‌کند.

مثال

سؤال ضعیف:

در رابطه با استفاده از سؤالهای انشایی در ارزشیابی هدفهای آموزشی، نمی‌توان گفت که ــــــ.

(×) الف ـ سؤالهای انشایی برای ارزشیابی هدفهای طبقهٔ ترکیب بهترین نوع نیستند

ب ـ سؤالهای انشایی برای ارزشیابی همهٔ طبقات حوزهٔ شناختی مفید نیستند

پ ـ سؤالهای انشایی و عینی تفاوت زیادی با یکدیگر ندارند

ت ـ استفاده از سؤالهای عینی برای هدفهای دانش همیشه صلاح نیست

شکل بهتر این سؤال در صفحه ۲۰۰ آمده است.

۱۴. از کاربرد عبارتهایی مانند «همهٔ آنچه در بالا گفته شد»، «تمامی موارد بالا» و غیره خودداری کنید. سؤالهایی که در آنها این گونه عبارتها به کار می‌روند، به ویژه اگر به جای گزینهٔ درست سؤال به کار روند، بسیار آسان هستند زیرا اگر آزمون‌شونده بداند که حداقل دو گزینه در گزینه‌ها درست است، چه از درست و غلط بودن بقیهٔ گزینه‌ها اطلاع داشته باشد چه نداشته باشد، گزینهٔ صحیح را انتخاب خواهد کرد. به جای این‌گونه سؤالها می‌توانید از سؤالهای دیگری نظیر آنچه در زیر آمده است استفاده کنید.

مثال

در ترکیب هوا در اتاقی که هیچ‌گونه منفذی به بیرون ندارد و تنها تعدادی گیاه در آن قرار دارند چه تغییراتی رخ می‌دهد؟

الف ـ اکسیژن کاهش می‌یابد.

ب ـ دی‌اکسید کربن افزایش می‌یابد.

پ ـ اکسیژن افزایش می‌یابد.

ت ـ هم دی‌اکسیدکربن و هم اکسیژن افزایش می‌یابند.

(×) ث ـ اکسیژن کاهش می‌یابد ولی دی‌اکسیدکربن افزایش می‌یابد.

(مهرنز و لهمان، ۱۹۸۴)

۱۵. تا حد امکان از کاربرد گزینهٔ «هیچ یک از موارد بالا» پرهیز کنید. صاحب‌نظران سنجش و اندازه‌گیری استفاده از گزینهٔ "هیچ‌کدام" یا "هیچ یک از موارد بالا" را پیشنهاد نمی‌کنند. بهترین مورد استفادهٔ این گزینه در ارتباط با سؤالهای درسهای ریاضیات است؛ با این حال، بعضی از متخصصان استفاده از این گزینه را حتی با سؤالهای ریاضی هم مفید نمی‌دانند. از جمله پین (۲۰۰۳) در رابطه با نگرانی خود مورد زیر را ذکر کرده است:

سؤال ضعیف:

دو ضلع مجاور زاویهٔ قائمهٔ یک مثلث قائم‌الزاویه ۳ و ۴ سانتیمتر است. مساحت این مثلث چندسانتیمتر مربع است؟



الف ـ ۷

ب ـ ۱۲

پ ـ ۲۵

(×) ت ـ هیچ یک از موارد بالا

جواب سؤال بالا ۶ سانتیمتر است و دانش‌آموزی که جواب سؤال را می‌داند گزینهٔ ت را انتخاب خواهد کرد. اما دانش‌آموز دیگری که مسئله را اشتباه حل می‌کند (مثلاً به جای مساحت مثلث وتر آن را که ۵ سانتیمتر است محاسبه می‌نماید و جواب ت را انتخاب می‌کند) از یک راه غلط به یک جواب درست می‌رسد. بنابراین، گزینهٔ "هیچ یک از موارد بالا" فقط زمانی درست عمل می‌کند که از آن به عنوان گزینهٔ غلط استفاده شود و گزینهٔ درست در میان سایر گزینه‌ها باشد.

سؤال بهتر:

دو ضلع مجاور زاویهٔ قائمهٔ یک مثلث قائم‌الزاویه ۳ و ۴ سانتیمتر است. مساحت این مثلث چندسانتیمتر مربع است؟



(×) الف ـ ۶

ب ـ ۷

پ ـ ۱۲

ت ـ هیچ یک از موارد بالا

در غیر از مسائل و محاسبات ریاضی، استفاده از گزینهٔ "هیچ یک از موارد بالا" پیشنهاد نمی‌شود، زیرا در سؤالهایی که "هیچ یک از موارد بالا" گزینهٔ درست است و دانش‌آموز آن را انتخاب می‌کند و به ما نشان می‌دهد که هیچ‌کدام از گزینه‌های پیشنهادی درست نیست، باز هم نخواهیم فهمید که او جواب درست سؤال را می‌داند یا نه. به مثال زیر توجه کنید:

مرکز استان مازندران کدام‌یک از شهرهای زیر است؟

الف ـ چالوس
ب ـ آمل
پ ـ بابل
(×) ت ـ هیچ‌کدام

دانش‌آموزی که گزینهٔ ت (گزینهٔ درست) سؤال بالا را انتخاب می‌کند می‌داند که شهرهای چالوس، آمل، و بابل مرکز استان مازندران نیستند، اما بر ما معلوم نمی‌شود که او جواب درست سؤال را که شهر ساری است می‌داند یا نه. "نکتهٔ دیگری که موضوع را پیچیده‌تر می‌کند این است که دانش‌آموزان قوی بعضی وقتها در گزینهٔ صحیح اشکال می‌بینند و به همین دلیل گزینهٔ 'هیچ‌کدام' را انتخاب می‌کنند" (ریس[1]، براون[2]، و اسمیت[3]، ۲۰۰۵، ص ۴۷).

۱۶. سؤالها را مستقل از یکدیگر بنویسید. پاسخ دادن به یک سؤال نباید شرط لازم برای پاسخ دادن به سؤال دیگری باشد، و جواب یک سؤال نباید به پاسخ دادن به سؤال دیگری کمک بکند. فرض کنید در یک آزمون ریاضی دو سؤال پشت سر هم آمده باشند که در سؤال اول از آزمون‌شونده خواسته شده باشد تا مساحت یک شکل را تعیین کند و در سؤال دوم از او خواسته شده باشد که هزینهٔ نوعی کفپوش را برای مساحت تعیین شده به دست آورد. اگر جواب آزمون‌شونده به سؤال اول غلط باشد، جواب سؤال دوم نیز الزاماً غلط خواهد بود. اگر آزمون‌شونده نحوهٔ محاسبهٔ هزینهٔ کفپوش را بداند اما، به علت یک اشتباه کوچک در محاسبه، سؤال اول را غلط جواب دهد، برای هیچ یک از دو سؤال نمره‌ای نخواهد گرفت.

۱۷. از طرح سؤالهای گمراه‌کننده بپرهیزید. استفاده از سؤالهای گمراه کننده در آزمونهای پیشرفت تحصیلی به هیچ وجه جایز نیست. اغلب اتفاق می‌افتد آزمون‌شوندگان مطلع‌تر که اطلاعات وسیع‌تری دارند به وسیلهٔ این‌گونه سؤالها گمراه می‌شوند. این‌گونه سؤالها با هدف اصلی آزمونهای پیشرفت تحصیلی که اندازه‌گیری میزان فهم و دانش آزمون‌شوندگان است مغایرت دارد.

مثال

سؤال گمراه کننده:

با توجه به اینکه نیروی جاذبهٔ کرهٔ زمین شش برابر نیروی جاذبهٔ کرهٔ ماه است، اگر جرم جسمی در روی زمین ۶۰ کیلوگرم باشد، جرم این جسم در کرهٔ ماه چند کیلوگرم است؟

الف ـ ۰/۱۶
ب ـ ۶
پ ـ ۱۰
(×) ت ـ ۶۰

1. Race 2. Brown 3. Smith

توضیح مقدمهٔ تنهٔ سؤال بالا ممکن است برای تعدادی از آزمون‌شوندگان گمراه کننده باشد. برای اصلاح این سؤال باید آن توضیح را به کلی حذف کرد.

سؤال بهتر:

اگر جرم جسمی در روی زمین ۶۰ کیلوگرم باشد، جرم این جسم در کرهٔ ماه چند کیلوگرم است؟

الف ـ ۰/۶

ب ـ ۶

پ ـ ۱۰

(×) ت ـ ۶۰

۱۸. دو گزینهٔ متضاد را که یکی از آنها درست است به کار نبرید. هرگاه در میان گزینه‌ها تنها یک زوج گزینه متضاد به کار رود که یکی از آنها گزینهٔ درست سؤال باشد، احتمالاً آزمون‌شونده انتخاب خود را به همین دو گزینه محدود خواهد کرد، زیرا تصور می‌کند که ممکن نیست هر دو پاسخ غلط باشند. در این‌گونه موارد، نمی‌توان سؤال را چندگزینه‌ای دانست، زیرا آن سؤال مانند سؤالهای صحیح ـ غلط عمل می‌کند و احتمال حدس زدن جواب صحیح نیز افزایش می‌یابد. بنابراین، بهتر است یا به طور کلی از کاربرد متضادها صرف‌نظر شود یا دو زوج متضاد برای سؤالهای چهار گزینه‌ای مورد استفاده قرار گیرد.

مثال

سؤال ضعیف:

وقتی که شباهت بین محرک غیرشرطی و محرک شرطی افزایش می‌یابد،

(×) الف ـ تعمیم بیشتر اتفاق می‌افتد، اما تمیز دشوارتر می‌شود.

ب ـ تمیز بیشتر اتفاق می‌افتد، اما تعمیم دشوارتر می‌شود.

پ ـ تعمیم و تمیز کمتر اتفاق می‌افتند.

ت ـ اتفاق تازه‌ای رُخ نمی‌دهد.

سؤال بهتر:

وقتی که شباهت بین محرک غیرشرطی و محرک شرطی افزایش می‌یابد،

(×) الف ـ تعمیم بیشتر اتفاق می‌افتد، اما تمیز دشوارتر می‌شود.

ب ـ تمیز بیشتر اتفاق می‌افتد، اما تعمیم دشوارتر می‌شود.

پ ـ تعمیم و تمیز کمتر اتفاق می‌افتند.

ت ـ تعمیم و تمیز بیشتر اتفاق می‌افتند.

۱۹. طول گزینهٔ درست را در سؤالهای مختلف تغییر دهید. گزینهٔ درست در مقایسه با گزینه‌های انحرافی باید طوری نوشته شود که به هیچ‌وجه نمایان نباشد. گزینهٔ درست نباید در چندین

سؤال پشت سر هم از سایر گزینه‌ها بلندتر یا کوتاه‌تر باشد، زیرا ممکن است آزمون‌شونده را، صرف‌نظر از اینکه جواب درست سؤال را می‌داند یا نمی‌داند، در پیداکردن گزینهٔ درست کمک کند. بنابراین، سعی کنید گاهی گزینهٔ درست را بلندتر از سایر گزینه‌ها، گاه کوتاه‌تر از آنها، و اغلب هم‌طراز با آنها بنویسید. اما مواظب باشید که در چند سؤال پشت سر هم گزینهٔ درست کوتاه‌تر یا بلندتر از سایر گزینه‌ها نباشد.

۲۰. محل گزینهٔ درست را در میان گزینه‌های انحرافی به طور تصادفی انتخاب کنید. اگر در تعیین محل گزینه‌های درست یک آزمون نوعی نظم و ترتیب حاکم باشد، یا اگر گزینه‌های صحیح در جای معینی قرار داده شوند، ممکن است آزمون شونده این مطلب را کشف کند و بدون اینکه جواب سؤالی را بداند گزینهٔ درست را انتخاب کند. ساده‌ترین راه برای انتخاب محل گزینه‌های درست استفاده از شمارهٔ صفحات یک کتاب است. کتابی را به طور دلخواه باز کنید و به صفحه‌ای که شمارهٔ آن فرد است نگاه کنید. اگر صفحه‌ای را که باز کرده‌اید شماره‌اش ۱ است، محل پاسخ درست را **الف** انتخاب کنید و اگر شماره‌اش ۳ است، ب، و همچنین برای شماره‌های ۵، پ، ۷، ت، و ۹، ث را منظور کنید. این روش را برای همهٔ سؤالها به کار برید.

رقم پایانی شمارهٔ صفحهٔ فرد	محل گزینهٔ درست
۱	الف
۳	ب
۵	پ
۷	ت
۹	ث

روش دیگر قرار دادن گزینه‌ها در دنبال تنهٔ سؤال استفاده از نوعی نظم و ترتیب طبیعی است که این خود سبب می‌شود که گزینهٔ انحرافی در سؤالهای مختلف جای تصادفی داشته باشد. مثلاً می‌توان گزینه‌ها را به ترتیب حروف الفبا به دنبال هم قرار داد (اگر از یک کلمه تشکیل یافته‌اند)، یا به ترتیب بزرگی و کوچکی اعداد (اگر از اعداد تشکیل شده‌اند)، یا به ترتیب بلندی یا کوتاهی طول پاسخها. به مثال زیر توجه کنید:

مثال

مقدار X در معادلهٔ $X^2 + 1 = 5$ چقدر است؟

الف ـ ۱

(×) ب ـ ۲

پ ـ ۳

ت ـ ۴

۲۱. برای هر سؤال بین ۳ تا ۵ گزینه در نظر بگیرید. در سؤالهای چندگزینه‌ای کاربرد ۳ تا ۵ گزینه تفاوت زیادی ندارد. البته از لحاظ نظری هرچه تعداد گزینه‌ها بیشتر باشد احتمال حدس زدن پاسخ سؤال کمتر است. اما از آنجا که نوشتن تعداد زیادی گزینهٔ انحرافی دشوار است، گاهی اوقات تعداد گزینه‌های کمتر یعنی ۳ تا ۴ گزینه مناسب‌تر است. به ویژه خواندن تعداد کمتری گزینه وقت کمتری از آزمون‌شونده را می‌گیرد، و این امر سبب می‌شود که در یک جلسهٔ امتحان تعداد بیشتری سؤال با گزینه‌های کمتری مورد استفاده قرار گیرد.

مهرنز و لهمان (۱۹۸۴) پیشنهاد کرده‌اند که بهتر است برای اندازه‌گیری پیشرفت تحصیلی دانش‌آموزان سطوح پایین از سؤالهای دارای گزینه‌های کمتر استفاده شود. آنان براساس تجربه‌های خود نظر داده‌اند که دانش‌آموزان کلاس دوم از عهدهٔ پاسخ دادن به سؤالهای سه گزینه‌ای برمی‌آیند، دانش‌آموزان کلاس سوم و چهارم می‌توانند سؤالهای ۳ تا ۴ گزینه‌ای را جواب دهند، و سؤالهای ۴ تا ۵ گزینه‌ای بیشتر برای دانش‌آموزان کلاس ششم و بالاتر مناسب‌اند.

۲۲. در آزمونهای تکوینی از گزینهٔ «نمی‌دانم» استفاده کنید. هرچند که دربارهٔ کاربرد این گزینه میان صاحب‌نظران اختلاف عقیده وجود دارد، اما می‌توان از آن در سؤالهای آزمونهای تکوینی که برای شناسایی مشکلات یادگیری و آموزش به کار می‌روند استفاده کرد. چنان‌که قبلاً در این کتاب (فصل ۴) توضیح داده شد، آزمونهای تکوینی به عنوان کمکی برای آموزش و به منظور تشخیص مشکلات یادگیری دانش‌آموزان و آموزش معلمان مورد استفاده قرار می‌گیرند و هدف از آنها نمره دادن به دانش‌آموزان نیست. در چنین حالتی به نفع یادگیرنده است که اگر نمی‌تواند به سؤالی جواب دهد، به جای حدس زدن کورکورانه، صادقانه عمل کند و گزینهٔ "نمی‌دانم" را انتخاب نماید. زیرا اگر حدس بزند و حدس او درست باشد، بی آنکه جواب سؤال را بداند، در مرحلهٔ بالاتری از آموزش قرار داده خواهد شد که ممکن است برای او مفید نباشد. بنابراین، در آزمونهایی که به منظور تشخیص مشکلات آموزش و یادگیری به کار می‌روند، نه برای تعیین نمره، گنجانیدن گزینهٔ "نمی‌دانم" هم به نفع دانش‌آموز است و هم به نفع معلم. البته باید به دانش‌آموزان گفته شود که هرگاه جواب سؤالی را نمی‌دانند به جای اینکه حدس بزنند گزینهٔ "نمی‌دانم" را انتخاب کنند.

مثال:

در معادلهٔ $\frac{x}{۲} - \sqrt{۹} = ۱$ مقدار x چقدر است؟

الف ـ ۳

ب ـ ۴

(×) پ ـ ۸

ت ـ ۲

ث ـ نمی‌دانم

۲۳. در سؤالهایی که تنهٔ آنها جملهٔ ناتمام است، جای خالی را در قسمت آخر جمله قرار دهید. موضوع اصلی هر سؤال باید در تنهٔ سؤال باشد، به گونه‌ای که آزمون شونده بعد از مطالعهٔ آن و پیش از مراجعه به گزینه‌ها منظور اصلی سؤال را بفهمد. یکی از سفارشهای صاحب‌نظران در این باره قرار دادن جای خالی در انتهای تنهٔ سؤال است نه در اوایل یا اواسط آن. به نمونه‌های زیر توجه کنید.

مثال

سؤال ضعیف:

طبق نظر دیباکی ـــــ عامل اصلی بیماری قلبی است.

(×) الف ـ کلسترول

ب ـ ضربهٔ

پ ـ سیگار کشیدن

ت ـ استرس

سؤال بهتر:

طبق نظر دیباکی عامل اصلی بیماری قلبی ـــــ است.

(×) الف ـ کلسترول

ب ـ ضربهٔ

پ ـ سیگار کشیدن

ت ـ استرس

(مهرنز و لهمان، ۱۹۸۴)

چنان که دیده می‌شود، در سؤال ضعیف بالا جای خالی وسط سؤال مانع از ادامهٔ مطالعهٔ آزمون‌شونده می‌شود و موجب سردرگمی او می‌گردد. در سؤال بهتر جای خالی سؤال در آخر تنهٔ سؤال آمده است و آزمون‌شونده پس از خواندن سؤال و پیش از پرداختن به گزینه‌ها از هدف سؤال به خوبی آگاه می‌شود.

۲۴. تا آنجا که ممکن است در تنهٔ سؤال به جای جملهٔ ناتمام از جملهٔ پرسشی استفاده کنید. متخصصان سنجش و اندازه‌گیری پیشرفت تحصیلی پیشنهاد کرده‌اند که آزمون‌ساز (به ویژه آزمون‌ساز تازه‌کار) کوشش کند تا تنهٔ سؤالهای خود را در قالب جمله‌های سؤالی یا پرسشی طرح کند و از عبارتهای ناتمام پرهیز نماید. جملهٔ پرسشی هدف سؤال را به طور آشکارتری بیان می‌کند تا جملهٔ ناتمام. پین (۲۰۰۳) در این باره گفته است "یکی از مشکلات استفاده از جمله‌های ناتمام این است که جملهٔ ناتمام موجب می‌شود که آزمون‌شونده سؤالی را که در ذهن آزمون‌ساز بوده است بازسازی کند، که خودِ این ممکن است باعث به هدر رفتن وقت و نیروی آزمون شونده گردد" (۱۹۶).

به رغم نگرانیهای فوق، چنان‌که طرح سؤال در قالب جملهٔ پرسشی خیلی طولانی‌تر از طرح آن به صورت عبارت ناتمام باشد، بهتر است از همان شکل جملهٔ ناتمام استفاده کنید. اما سعی نمایید که تمام گزینه‌های پیشنهادی تکمیل کنندهٔ جملهٔ ناتمام باشند.

آزمون چندگزینه‌ای رایانه‌ای

یکی از کاربردهای ابتکاری سؤالهای چندگزینه‌ای استفاده از رایانه (کامپیوتر) است. برای این منظور، مجموعهٔ متنوعی از سؤالهای چندگزینه‌ای در یک زمینهٔ علمی شامل سؤالهای دارای درجه‌های دشواریِ از ساده تا دشوار و ضریبهای تمیز خوب (سؤالهایی که گروههای قوی و ضعیف آزمون شونده را از هم جدا می‌کند) انتخاب و به دستگاه داده می‌شود. در جریان اجرای آزمون، ابتدا آزمون‌شونده سؤالی از رایانه دریافت می‌کند. اگر توانست به سؤال پاسخ دهد، سؤال نسبتاً دشوارتری دریافت می‌کند، و همچنان این جریان ادامه می‌یابد. اگر آزمون شونده به سؤال اول جواب غلط بدهد، سؤال بعدی او یک سؤال ساده‌تر خواهد بود، و این جریان نیز تا دادن جواب درست ادامه می‌یابد. رایانه برنامه‌ریزی شده است که سطح توانایی آزمون‌شونده را پیدا کند و تا رسیدن به این سطح به سؤال پرسیدن ادامه می‌دهد. "این‌گونه آزمودن تاکنون در حوزهٔ آموزش پزشکی بیشترین مورد استفاده را داشته و معلوم گشته است که وقتی با آزمونهای سنتی به کار می‌رود روش بسیار مفید و معتبری است" (ریس، براون، و اسمیت، ۲۰۰۵، ص ۵۴).

امتیازها و محدودیتهای آزمونهای چندگزینه‌ای

آزمونهای چندگزینه‌ای تمام محاسن آزمونهای صحیح ـ غلط و جورکردنی را دارند. علاوه بر آن، بنا به گفتهٔ ایبل (۱۹۷۹)،

> این آزمونها قادر به اندازه‌گیری غالب بازده‌های یادگیری از دانش گرفته تا فهمیدن، قضاوت کردن، حل مسئله، ارائهٔ پیشنهادهای عملی، و پیش‌بینی امور هستند. تقریباً هرگونه درک و فهم یا توانایی را که بتوان با آزمونهای دیگر ـ کوته پاسخ، کامل کردنی، صحیح ـ غلط، جورکردنی، یا تشریحی ـ اندازه گرفت می‌توان با آزمونهای چندگزینه‌ای سنجش کرد. (ص ۱۳۵)

به طور کلی امتیازهای آزمونهای چندگزینه‌ای را می‌توان به شرح زیر بیان کرد:

۱. این آزمونها از سایر آزمونهای عینی انعطاف‌پذیرترند. آنها علاوه بر دانش، توانایی استدلال، قضاوت، و بسیاری بازده‌های مهم دیگر یادگیری را می‌سنجند.

۲. با استفاده از سؤالهای چندگزینه‌ای می‌توان در یک زمان محدود تعداد زیادی از هدفهای آموزشی و بخش مهمی از محتوای درس را اندازه‌گیری کرد.

۳. آزمونهای چندگزینه‌ای نسبت به آزمونهای صحیح ـ غلط کمتر امکان حدس زدن کورکورانه را به آزمون‌شونده می‌دهند.

۴. پاسخهای این سؤالها به سادگی و با عینیت کامل قابل تصحیح هستند.
۵. دانش‌آموزان و معلمان این نوع آزمونها را به سایر آزمونهای عینی ترجیح می‌دهند.
۶. اگر گزینه‌های انحرافی این آزمونها با توجه به اشتباهات و کج فهمی‌های متداول دانش‌آموزان تهیه شوند، منبع بسیار مناسبی برای تشخیص مشکلات یادگیری دانش‌آموزان خواهند بود.

آزمونهای چندگزینه‌ای از چند نظر مورد انتقاد قرار گرفته‌اند. ما در اینجا این انتقادها را به عنوان محدودیتهای آزمونهای چندگزینه‌ای برمی‌شماریم.

۱. ساختن این آزمونها بسیار دشوار است. از جمله اینکه، معلمان غالباً نمی‌توانند تعدادی گزینهٔ انحرافی خوب برای سؤالهای چندگزینه‌ای انتخاب کنند.
۲. معلمان اغلب سؤالهای چندگزینه‌ای را در حد سنجش اطلاعات جزئی و کم‌اهمیت می‌نویسند. (البته این ایرادی است که به همهٔ آزمونهای عینی از جمله به آزمونهای چندگزینه‌ای می‌گیرند.)
۳. در مقایسه با بعضی آزمونهای دیگر، خواندن این آزمونها و پیداکردن گزینهٔ درست مستلزم صرف وقت نسبتاً زیادتری است، به ویژه اگر گزینه‌ها خیلی شبیه به هم ساخته شوند.
۴. زمانی که به پاسخهای غلط نمرهٔ منفی داده می‌شود دانش‌آموزانی که خطر می‌کنند از سایر دانش‌آموزان نمره‌های بهتری می‌گیرند. همچنین دانش‌آموزانی که در پاسخ دادن به این گونه سؤالها خبره و باتجربه یا به اصطلاح آزمون آشنا[1] هستند از دانش‌آموزان تازه‌کار نمره‌های بهتری به دست می‌آورند.
۵. دانش‌آموزان قوی بیش از دانش‌آموزان معمولی قادر به پیداکردن اشکالات، پیچیدگی‌ها، و نظرگاههای واگرا (غیرمعمولی) در سؤالها هستند، اما از آنجاکه تنها یک پاسخ صحیح در هر سؤال وجود دارد، بابت این تیزبینی‌های خود نه تنها تشویق نمی‌شوند بلکه با دادن جوابی که مورد نظر طراح سؤال نیست ممکن است تنبیه هم بشوند.

موارد استفادهٔ آزمونهای چندگزینه‌ای

آزمونهای چندگزینه‌ای بهترین ابزارهای سنجش گزینش پاسخ یا بهترین نوع آزمونهای بسته پاسخ هستند. سؤالهای این آزمونها می‌توانند هدفهای آموزشی مختلفی را اندازه بگیرند و با محتوای غالب موضوعهای درسی سازگاری دارند (لین و گرانلاند، ۲۰۰۰). از آنجاکه سؤالهای چندگزینه‌ای بسیار انعطاف‌پذیرند و قابلیت کاربرد متنوعی دارند، تاکنون اکثریت آزمونهای استاندارد شده از آنها استفاده کرده‌اند. سؤالهای چندگزینه‌ای عموماً برای سنجش هدفهای یادآوری، فهمیدن، و کاربستن مفیدند. با این حال، اگر در تهیهٔ آنها دقت و ابتکار به خرج دهید

1. test-wise

می‌توانید یادگیریهای پیچیده‌تر را نیز بسنجید. ما در ادامهٔ این قسمت، با استفاده از منابع معتبر اندازه‌گیری و سنجش یادگیری، نمونه‌های مختلفی از سؤالهای چندگزینه‌ای را که برای سنجش هدفهای گوناگون تهیه شده‌اند معرفی می‌کنیم.

یادآوری دانش امور واقعی

۱ ـ قطبهای آهنربا معمولاً به این صورت نامگذاری می‌شوند: _____.

(x) الف ـ منفی و مثبت

ب ـ قرمز و آبی

پ ـ غربی و شرقی

ت ـ شمالی و جنوبی

ث ـ آند و کاتد

(بلوم و همکاران، ۱۹۵۶، ترجمهٔ سیف و علی‌آبادی، ۱۳۶۸، ص ۹۰)

۲ ـ بیت زیر از کدام منبع است؟

ز سُم ستوران در آن پهن‌دشت زمین شد شش و آسمان گشت هشت

(x) الف ـ شاهنامه فردوسی

ب ـ بوستان سعدی

پ ـ هفت‌پیکر نظامی

ت ـ مثنوی مولوی

یادآوری دانش مفهومی

۳ ـ در نظریهٔ اسکینر تقویت‌کننده چگونه تعریف می‌شود؟

الف ـ محرکی که مورد علاقهٔ ارگانیسم است و آن را بر محرکهای دیگر ترجیح می‌دهد.

(x) ب ـ رویدادی که وقتی بعد از رفتار می‌آید احتمال بروز مجدد آن را افزایش می‌دهد.

پ ـ هر پاسخی که در سلسله مراتب پاسخهای مختلف ارگانیسم در ردهٔ بالا قرار دارد.

ت ـ هر چیزی که وقتی به دنبال رفتار ارگانیسم می‌آید موجب کاهش سائق شود.

۴ ـ در حوزهٔ عاطفی نگرش به کدام طبقه مربوط است؟

الف ـ دریافت کردن

ب ـ پاسخ دادن

(x) پ ـ ارزش‌گذاری

ت ـ سازمان

ث ـ شخصیت پذیرفتن

یاداوری دانش روندی

۵ ـ مراحل زندگی مگس خانگی به ترتیب عبارت‌اند از:

الف ـ کرم، تخم، شفیره، مگس بالغ

ب ـ شفیره، کرم، تخم، مگس بالغ

پ ـ شفیره، تخم، کرم، مگس بالغ

ت ـ تخم، کرم، مگس بالغ، شفیره

(×)ث ـ تخم، کرم، شفیره، مگس بالغ

(بلوم و همکاران، ۱۹۶۸، ص ۹۳)

فهمیدن دانش امور واقعی

راهنمایی: به نمودار زیر دقت کنید و سؤالهای بعد آن را با توجه به اطلاعات موجود در نمودار جواب دهید.

۶ ـ "پول خرج شده برای رفاه و آموزش و پرورش در سال ۱۹۳۰ کمتر از پولی بوده که قبل از ۱۹۲۹ به همین منظور خرج شده است."

در رابطه با بیان فوق چگونه می‌توان داوری کرد؟

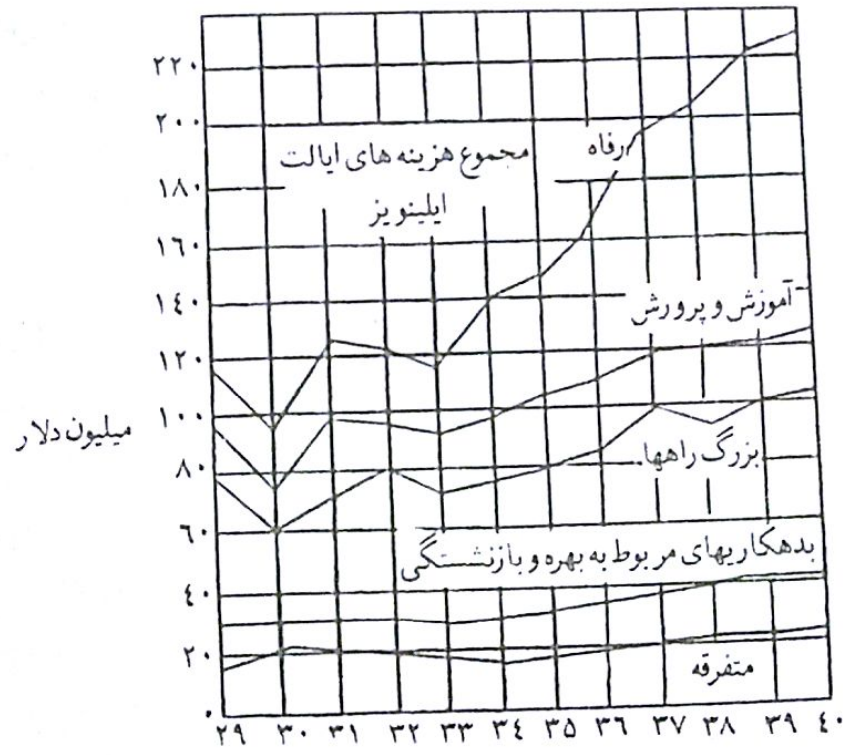
(×)الف ـ احتمالاً درست است.

ب ـ احتمالاً غلط است.

پ ـ قطعاً درست است.

ت ـ قطعاً غلط است.

ث ـ نمی‌توان نظر داد.



۷ـ بین سالهای ۱۹۳۷ و ۱۹۳۹ بیکاری در ایالت افزایش یافت.
الف ـ احتمالاً درست است.
ب ـ احتمالاً غلط است.
پ ـ قطعاً درست است.
ت ـ قطعاً غلط است.
(x)ث ـ نمی‌توان نظر داد.

(اقتباس براساس بلوم و همکاران، ۱۳۶۸، صص ۱۲۹ـ۱۲۸)

فهمیدن دانش مفهومی

۸ـ "قلب انسان در مرحلهٔ بسیار ابتدایی رشد دارای دو حفره است."
بیان فوق مؤید کدام بخش از نظریهٔ تکامل زیست‌شناسی است؟
الف ـ کالبدشناسی تطبیقی
ب ـ فیزیولوژی تطبیقی
(x)پ ـ رویان‌شناسی
ت ـ دیرین‌شناسی

(اقتباس براساس بلوم و همکاران، ۱۳۶۸، ص ۱۰۰)

راهنمایی: بیتهای زیر را بخوانید و سؤالهای بعد آن را جواب دهید.

۱	حسن تو دائم بدین قرار نماند	مست تو جاوید در خمار نماند
۲	ای گل خندان نو شکفته نگه دار	خاطر بلبل که نوبهار نماند
۳	حسن دلاویز پنجه‌ایست نگارین	تا به قیامت بر او نگار نماند
۴	عاقبت از ما غبار ماند زنهار	تا ز تو بر خاطری غبار نماند
۵	پار گذشت آنچه دیدی از غم و شادی	بگذرد امسال و همچو پار نماند
۶	هم بدهد دور روزگار مرادت	ور ندهد دور روزگار نماند
۷	سعدی شوریده بی‌قرار چرایی	در پی چیزی که برقرار نماند

۹ـ در سه سطر اول شعر بالا، سعدی ـــــــ.
الف ـ معشوق خود را از همهٔ زیبایان بهتر می‌داند
(x)ب ـ می‌گوید زیبایی زودگذر است و نباید مغرور آن شد
پ ـ حسن و زیبایی را پدیده‌ای جاودانی فرض کرده است
ت ـ بلبل را ستایشگر زیبایی گل بهاری می‌داند

۱۰ـ منظور شاعر در سطرهای ۴ و ۵ این است که ـــــــ.

(x)الف ـ چون همه چیز زودگذر است، بهتر آن است که کسی را از رفتار ما کلوزتی حاصل نشود

ب ـ هیچ سالی به سال دیگر شبیه نیست، پس بهتر است از هر موقعیت حداکثر استفاده را ببریم

پ ـ سال قبل آرزوهای ما برآورده نشدند، باید سعی کنیم که امسال به مراد خود برسیم

ت ـ ما همه عاقبت به غباری از خاک بدل می‌شویم، به جز زیبایی که جاودانه باقی می‌ماند

۱۱ ـ به شکل‌های پایین نگاه کنید. آنها دو بطری هم‌اندازه‌اند که از یک نوع مایع پر شده و درشان با بادکنک بسته شده است.

بطری الف را حرارت می‌دهیم و بادکنک در آن به حالتی که می‌بینید در می‌آید. در این صورت کدام‌یک از گفته‌های زیر در مورد آنها درست است؟

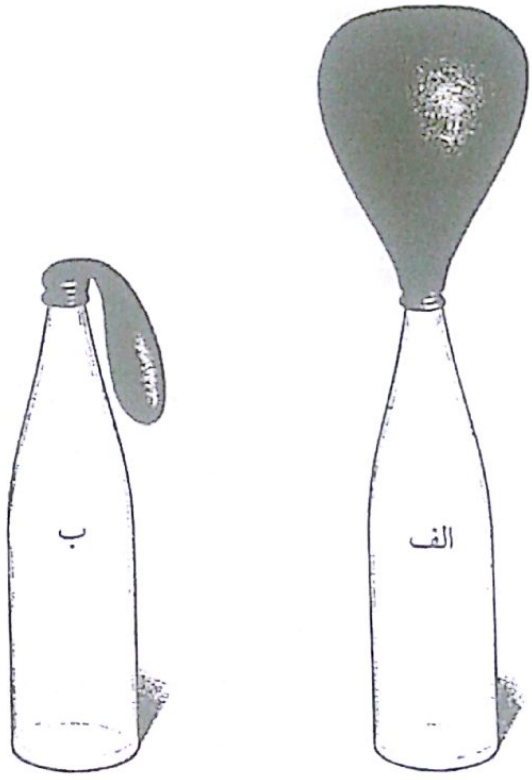
الف ـ چگالی (غلظت) هوای بطری الف از چگالی هوای بطری ب بیشتر است.

(x)ب ـ چگالی هوای بطری الف از چگالی هوای بطری ب کمتر است.

پ ـ چگالی هوای هر دو بطری مساوی است.

ت ـ برای مقایسهٔ چگالی دو بطری اطلاعات کافی در دست نیست.

(اگن و کاوچاک، ۲۰۰۱، ص ۶۰۴)



کاربستن دانش مفهومی

۱۲ ـ اگر حجم مقدار معینی از گاز ثابت نگه‌داشته شود فشار آن را چگونه می‌توان کاهش داد؟

(x) الف ـ کاهش دما

ب ـ افزایش دما

پ ـ افزایش چگالی

ت ـ کاهش چگالی

ث ـ دادن گرما

(بلوم و همکاران، ۱۳۶۸، ص ۱۵۰)

۱۳ ـ فرض کنید آسانسوری با شتاب گرانش "g" پایین می‌آید. اگر فردی که سوار آسانسور است سعی کند که یک توپ لاستیکی را به سمت بالا پرتاب نماید، حرکت توپ نسبت به آسانسور چه وضعی خواهد داشت؟

الف ـ در نقطه‌ای که فرد آن را آزاد می‌کند ثابت می‌ماند.

ب ـ به بالای آسانسور رفته، در آنجا ثابت می‌ماند.

پ ـ اصلاً بالا نمی‌رود، بلکه به کف آسانسور می‌افتد.

(x) ت ـ بالا می‌رود، به سقف می‌خورد و برمی‌گردد، و با سرعتی ثابت پایین می‌افتد.

ث ـ بالا می‌رود، به سقف می‌خورد و باز می‌گردد، و با سرعت فزاینده‌ای پایین می‌افتد.

(بلوم و همکاران، ۱۳۶۸، ص ۱۴۶)

کاربستن دانش روندی

۱۴ ـ x و y می‌توانند کاری را به کمک هم در ۱۵ روز انجام دهند. آنها ۶ روز با هم کار می‌کنند. بعد x از کار دست می‌کشد و y آن را ادامه می‌دهد، و کار در ۳۰ روز تمام می‌شود. y به تنهایی در چند روز می‌تواند کار را تمام کند؟

(x) الف ـ ۳۰

ب ـ ۴۰

پ ـ ۵۰

ت ـ ۶۰

ث ـ هیچکدام از موارد بالا

(بلوم و همکاران، ۱۳۶۸، ص ۱۴۳)

۱۵ ـ می‌خواهیم بدانیم که یک روش جدید آموزشی در قیاس با روشهای معمول معلمان به یادگیری دانش‌آموزان کمک می‌کند یا نه. برای این منظور بهتر است _____.

الف ـ به کمک یک پرسشنامه نظر معلمان را جویا شویم

ب ـ از طریق مصاحبه با پدران و مادران دانش‌آموزان عقیدهٔ آنان را بسنجیم

پ ـ از راه مشاهده چگونگی برخورد دانش‌آموزان با مواد درسی را تعیین کنیم
(×)ت ـ به کمک یک آزمایش، تأثیر روش جدید آموزش را با روشهای قدیم مقایسه کنیم

تحلیل دانش مفهومی

۱۶ ـ گالیله مسئلهٔ شتاب سرعت اجسام در حال سقوط را به وسیلهٔ غلتاندن گلوله‌ها بر روی سطوح بسیار صاف با زاویهٔ شیب فزاینده مورد بررسی قرار داد، زیرا او برای اندازه‌گیری فاصله‌های زمانی کوتاه هیچ ابزاری در اختیار نداشت. او از اطلاعات به دست آمده به استخراج نتایجی دربارهٔ سقوط آزاد اجسام پرداخت. کدام یک از موارد زیر یک فرض ضمنی مربوط به استخراج او را نشان می‌دهد؟

الف ـ مقاومت هوا در سقوط آزاد قابل چشم‌پوشی است.

ب ـ اجسام با شتاب ثابت سقوط می‌کنند.

(×)پ ـ شتاب مشاهده شده در سطوح شیب‌دار برابر با شتاب در سقوط آزاد است.

ت ـ سطوح مستوی بدون اصطکاک هستند.

ث ـ سطح قائم و سطحی که تقریباً قائم است بر روی گلولهٔ غلتان اثر تقریباً مساوی دارند.
(بلوم و همکاران، ۱۳۶۸، ص ۱۶۸)

راهنمایی: متن زیر را بخوانید و سه سؤال بعد آن را از روی متن جواب دهید.

"(۱) هاملت از روح پدر مقتول خود دستور می‌گیرد که از قاتل، کلادیوس، انتقام بکشد. (۲) او به فوریت قادر به انجام این کار نیست، زیرا دلایل کافی را که نشان دهد کلادیوس قاتل پدر اوست در دست ندارد. (۳) هاملت در جریان کشف دلیل، ناخواسته به پادشاه اجازه می‌دهد تا به شک او پی ببرد. (۴) در جریان ادامهٔ نمایش، هاملت قادر به انتقام گرفتن نیست، زیرا موقعیت مناسبی برای انجام این کار نمی‌یابد. (۵) در پایان نمایش، هاملت در یک مبارزهٔ تن به تن که به توسط کلادیوس تدارک دیده شده شرکت می‌کند و نتیجهٔ آن مرگ قهرمان و دشمن او و دیگر شخصیتهای نسبتاً مهم داستان است."

۱۷ ـ بحث و ارزشیابی از متن فوق بیشتر حول نکته‌ای که در جمله (جمله‌های) زیر آمده است دور می‌زند. آن جمله (جمله‌ها) کدام است؟

الف ـ ۱

ب ـ ۲ و ۳

پ ـ ۲ و ۴

ت ـ ۵

۱۸ ـ موقتاً فرض کنید که با مطلب فوق کاملاً موافقید. در بحث از بخشهای مختلف نمایش، کدام یک از موارد زیر را کمتر مورد توجه قرار می‌دهید؟

الف ـ صحبتهای هاملت با روح پدرش در پردهٔ اول

ب ـ گذشت زمان بین پرده‌های اول و دوم

پ ـ نمایش در درون نمایش

ت ـ عزیمت هاملت به انگلستان

ث ـ گذشت زمان کوتاه بین نمایش درون نمایش و عزیمت هاملت به انگلستان

۱۹ ـ کدام یک از بیانات زیر دربارهٔ هاملت با موضع‌گیری کل مطلب بالا همخوانی کمتری دارد؟

الف ـ هاملت مرد عمل است.

ب ـ هاملت طبیعتاً فردی متفکر است و عادت به رویارویی با مشکلات از طریق عمل مستقیم ندارد.

پ ـ هاملت یک روشنفکر است که فردی زیرک‌تر از او، اگرچه کم دانش‌تر، از لحاظ کارهای عملی بر او امتیاز دارد.

ت ـ هاملت به طور طبیعی فردی حساس و خوش‌خلق است، اما در طول مدتی که به وسیلهٔ چهار پرده از نمایش ارائه می‌شود حالتی پریشان و افسرده دارد ـ شرایطی که به علت مرگ پدر رخوت شدیدی را در او ایجاد کرده است.

(بلوم و همکاران، ۱۳۶۸، صص ۱۶۹ـ۱۶۸)

۲۰ ـ یکی از استادان درس آمار می‌گوید "در یک آزمون یک نمرهٔ Z برابر است با رتبهٔ درصدی حدود ۸۴". این استاد کدام یک از فرضهای زیر را مبنای گفتهٔ خود قرار داده است؟

الف ـ نمره‌های آزمون از ۰ تا ۱۰۰ را شامل می‌شوند.

ب ـ انحراف معیار نمره‌های آزمون ۱۳/۴ است.

(x) پ ـ توزیع نمره‌های آزمون توزیع طبیعی است.

ت ـ آزمون موردنظر دارای پایایی و روایی است.

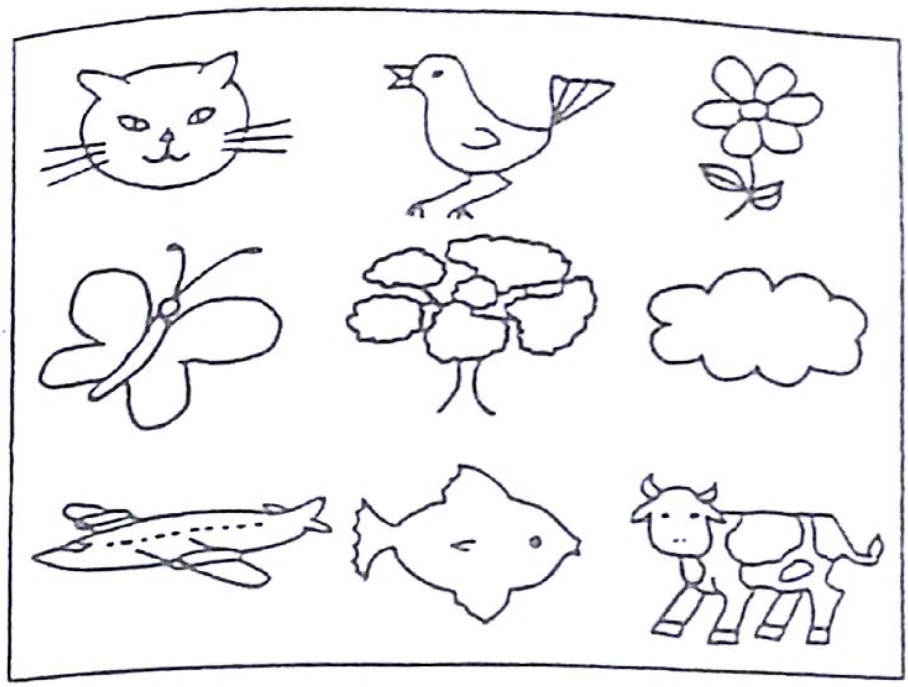
(وولفلک، ۲۰۰۴، ص ۵۵۰)

سؤالهای قابل استفاده با دانش‌آموزان پیش‌دبستانی و دبستانی

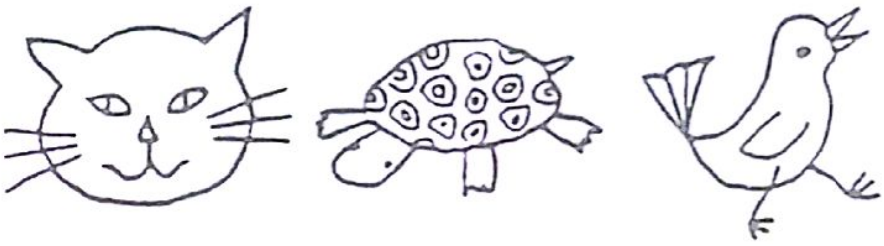
برای سهولت تصحیح پاسخهای آزمون‌شوندگان به سؤالهای گزینش پاسخ (به ویژه چندگزینه‌ای) از پاسخنامه‌های جداگانه که هم با ماشین و هم با کلید دستی قابل تصحیح‌اند استفاده می‌شود. اما در استفاده از این آزمونها با دانش‌آموزان خردسال، به ویژه پیش‌دبستانی و سالهای آغاز دبستان، این کار چندان عملی نیست. در این سالها بهترین راه ثبت پاسخها این است که از آزمون‌شوندگان خواسته شود تا بر روی پاسخ نوعی علامت (مثلاً +) قرار دهند. به

نمونه‌های زیر توجه کنید:

۲۰ ـ بر روی شکلهای چیزهای زنده یک علامت (+) بگذارید.
۲۱ ـ بر روی شکلهای حیوانها دو تا علامت (++) بگذارید.
۲۲ ـ بر روی شکل حیوانهایی که پستاندارانند سه تا علامت (+++) بگذارید.



۲۳ ـ کدام یک از شکلهای زیر یک پیشی را نشان می‌دهد؟ روی آن یک علامت (+) بگذارید.



۲۴ ـ یکی از شکلهای زیر نشان می‌دهد که وقتی چیزی را حرارت می‌دهیم و گرم می‌شود چه اتفاقی می‌افتد. روی آن یک علامت (+) بگذارید.



در استفاده از آزمونهای عینی (چندگزینه‌ای) با دانش‌آموزان سالهای سوم دبستان به بالا

می‌توان از پاسخنامه‌های سادهٔ معلم ساخته استفاده کرد. اما در استفاده از این سؤالهای چندگزینه‌ای با دانش‌آموزان خردسال سعی کنید از گزینه‌های کمتری (۳ گزینه) استفاده کنید. سؤالهای زیر از جمله سؤالهای مورد استفاده با دانش‌آموزان سالهای آخر دبستان است که در آنها تعداد گزینه‌ها اندک و استفاده از پاسخنامهٔ جداگانه امکان‌پذیر است.

۲۵ ـ وقتی که ساعتی بر روی زمین قرار بگیرد به گونه‌ای که ساعت ۱۲ شمال را نشان دهد، ساعت ۹ نشان‌دهندهٔ چیست؟

الف ـ جنوب

ب ـ شرق

(x) پ ـ غرب

۲۶ ـ اگر ۱۲ بر روی صفحهٔ ساعت شما شمال را نشان دهد، کدام یک از ساعتهای زیر جنوب شرقی را نشان می‌دهد؟

(x) الف ـ ۴:۳۰

ب ـ ۷:۳۰

پ ـ ۱۰:۳۰

۲۷ ـ در یک روز روشن به شخصی که در ساحل دریا ایستاده و به یک کشتی که در چندکیلومتری او در دریا قرار دارد نگاه می‌کند، چه چیزی را می‌بیند؟

الف ـ تمامی کشتی

(x) ب ـ تنها قسمت بالایی کشتی

پ ـ تنها قسمت پایینی کشتی

هنگام استفاده از آزمونهای چندگزینه‌ای با دانش‌آموزان دورهٔ راهنمایی به بالا، هم می‌توان از انواع پاسخنامه‌های دستی و پاسخنامه‌های قابل تصحیح با ماشین استفاده کرد و هم می‌توان تعداد گزینه‌ها را تا ۴ و ۵ گزینه افزایش داد. سؤالهای زیر نمونه‌هایی از سؤالهای مورد استفاده با دانش‌آموزان دورهٔ راهنمایی تحصیلی است.

راهنمایی: سؤالهای زیر را با توجه به شکل پایین پاسخ دهید. این شکل به لایه‌های زمین‌شناسی اشاره می‌کند. شکلهای مختلف زمین با اعداد رومی نشان داده شده‌اند.

۲۸ ـ فسیل با احتمال کمتری در کجا تشکیل می‌شود؟

(x) الف ـ III

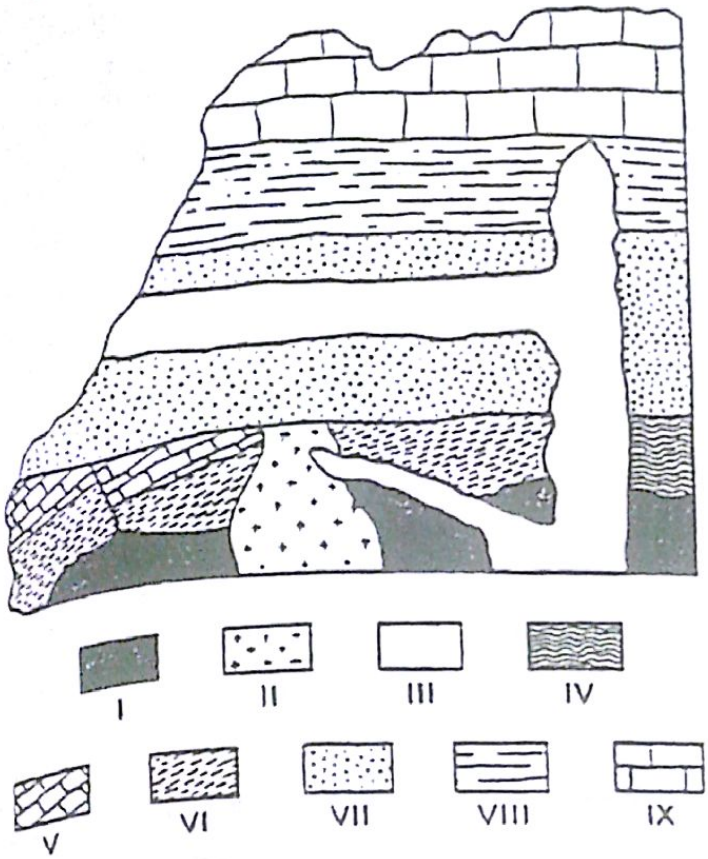
ب ـ V

پ ـ VII

ت ـ VIII

ث ـ IX

۲۹ ـ جوان‌ترین لایه زمین کدام است؟



الف ـ I

ب ـ II

پ ـ III

ت ـ IV

(x) ث ـ V

۳۰ ـ قدیمی‌ترین لایه زمین کدام است؟

(x) الف ـ I.

ب ـ II

پ ـ III

ت ـ IV

ث ـ V

حوزهٔ عاطفی (علاقه)

۳۱ ـ در جلسهٔ بعدی درس علوم اجتماعی دوست دارم ـــــــ.

الف ـ تنها کار کنم

ب ـ با یک یا دو نفر از دوستانم کار کنم

پ ـ با یک گروه کار کنم

۳۲ ـ برای درس علوم، زمان چهل دقیقه ـــــ .

الف ـ خیلی زیاد است

ب ـ کافی است

پ ـ کم است

۳۳ ـ من درس زبان را از درس ریاضیات ـــــــ دوست دارم.

الف ـ خیلی بیشتر

ب ـ کمی بیشتر

پ ـ خیلی کمتر

ت ـ کمی کمتر

خلاصهٔ فصل

آزمون چندگزینه‌ای معروف‌ترین و پرکاربردترین آزمون نوع گزینش پاسخ است. هر سؤال چندگزینه‌ای از سه قسمت زیر تشکیل یافته است: (۱) تنهٔ سؤال، (۲) گزینهٔ کلید، و (۳) گزینه‌های انحرافی. تنهٔ سؤال متن یا صورت سؤال را تشکیل می‌دهد. گزینهٔ کلید پاسخ درست سؤال است، و گزینه‌های انحرافی پاسخهای پیشنهادی غلط هستند که ظاهری درست نما دارند و آزمون‌شوندگان نامطلع از موضوع سؤال را به خود جلب می‌کنند.

سؤالهای چندگزینه‌ای انواع مختلفی دارند. ما در این فصل گونه‌های زیر را توضیح داده‌ایم: نوع پرسشی، نوع ناتمام، نوع تنها گزینهٔ درست، نوع بهترین گزینه، نوع منفی، نوع جایگزینی، و نوع پاسخ ترکیبی.

صاحب‌نظران سنجش و اندازه‌گیری پیشرفت تحصیلی قاعده‌های مختلفی را برای تهیهٔ سؤالهای چندگزینه‌ای پیشنهاد داده‌اند. ما در این فصل مهم‌ترین این پیشنهادها را با ذکر نمونه سؤالهایی در ۲۴ قاعده توضیح داده‌ایم.

دربارهٔ آزمونهای چندگزینه‌ای امتیازها و محدودیتهای مختلفی گفته شده است. مهم‌ترین امتیازها را می‌توان به صورت زیر بیان کرد: (۱) انعطاف‌پذیری، (۲) توانایی نمونه‌گیری بزرگ محتوا و هدفها، (۳) امکان حدس زدن کورکورانه کمتر از آزمونهای صحیح ـ غلط، (۴) قابل تصحیح بودن پاسخها با سهولت و عینیت، (۵) ترجیح داشتن بر سایر آزمونهای عینی از سوی معلمان و دانش‌آموزان، و (۶) فراهم آوردن امکانی برای تشخیص مشکلات یادگیری دانش‌آموزان.

در رابطه با محدودیتهای این آزمونها نیز مطالبی عنوان شده است. عمده‌ترین محدودیتها به شرح زیرند: (۱) دشواری ساخت، (۲) سنجش اطلاعات جزئی و کم‌اهمیت، (۳) نیاز به صرف وقت زیادتر از بعضی آزمونهای دیگر برای یافتن جواب سؤالها، (۴) نمرهٔ کمتر گرفتن دانش‌آموزان قوی‌تر در آزمونهایی که در آنها نمرهٔ منفی منظور می‌شود، و (۵) زیان دیدن دانش‌آموزان دارای تفکر واگرا.

سؤالهای چندگزینه‌ای موارد استفاده‌های فراوانی دارند و می‌توان آنها را برای سنجش تواناییهای از سطح یادآوری تا سطح تحلیل مورد استفاده قرار داد. در این فصل نمونه کاربردهای این سؤالها در سطوح مختلف یادگیری داده شده‌اند.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. بهتر است صورت سؤالهای چندگزینه‌ای را به صورت جملهٔ ناتمام بنویسیم تا جملهٔ پرسشی.

الف ـ صحیح

ب ـ غلط

۲. جنگ اول جهانی از ـــــ تا ـــــ به طول انجامید.

الف ـ ۱۹۱۰ تا ۱۹۱۲

(x) ب ـ ۱۹۱۴ تا ۱۹۱۸

پ ـ ۱۹۱۵ تا ۱۹۲۰

ت ـ ۱۹۲۵ تا ۱۹۲۸

ث ـ ۱۹۳۹ تا ۱۹۴۵

سؤال فوق چه نوع سؤالی است؟

الف ـ تنها گزینهٔ درست

ب ـ ناتمام

پ ـ جایگزینی

ت ـ الف و ب

ث ـ ب و پ

۳. نقش گزینهٔ انحرافی در سؤالهای چندگزینه‌ای چیست؟

الف ـ افزودن بر سطح دشواری سؤال

ب ـ گمراه کردن آزمون‌شوندگان بی‌اطلاع از موضوع سؤال

پ ـ گمراه کردن آزمون‌شوندگانی که دقت کافی ندارند

ت ـ افزودن بر قدرت تمیز سؤال

۴. کدام مورد زیر را در رابطه با سؤالهای چندگزینه‌ای اصلاً **نباید** به کار برد؟

الف ـ استفاده از گزینهٔ "همهٔ موارد بالا"

ب ـ استفاده از گزینهٔ "هیچ یک از موارد بالا"

پ ـ نوشتن همهٔ مطالب سؤال در تنهٔ سؤال

ت ـ اختصاص دادن هر سؤال فقط به یک هدف

۵. با ذکر یک مثال، ویژگیهای سؤالهای چندگزینه‌ای را توضیح دهید.

۶. قواعد تهیهٔ سؤالهای چندگزینه‌ای را با ذکر مثال شرح دهید.

۷. امتیازها و محدودیتهای سؤالهای چندگزینه‌ای را توضیح دهید.

۸ موارد کاربرد آزمونهای چندگزینه‌ای را توضیح دهید.

۹. برای طرح سنجش خود که به عنوان تمرین فصل ۶ تهیه کرده‌اید، یک آزمون چندگزینه‌ای متناسب با هدفهایی از آن درس که با این نوع آزمون بهتر سنجیده می‌شود بسازید.

۱۰

سنجش یادگیری به کمک آزمون تشریحی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمون تشریحی را تعریف کند.

۲. آزمونهای تشریحی محدود پاسخ و گسترده پاسخ را توضیح دهد.

۳. قواعد تهیهٔ آزمونهای تشریحی را با ذکر مثال برای هر قاعده توضیح دهد، و هنگام نوشتن آزمونهای خود این قواعد را رعایت کند.

۴. قواعد تصحیح پاسخهای سؤالهای آزمونهای تشریحی را توضیح دهد و هنگام تصحیح ورقه‌های امتحانی خود این قواعد را رعایت کند.

۵. امتیازها و محدودیتهای آزمونهای تشریحی را توضیح دهد و راههای پیشنهادی برای کاستن از محدودیتهای این نوع آزمونها را بیان کند.

۶. موارد استفادهٔ آزمونهای تشریحی را توضیح دهد.

۷. پرسش شفاهی را توضیح دهد و شباهتهای میان پرسشهای شفاهی و پرسشهای کتبی را بگوید.

۸. موارد استفادهٔ پرسشهای شفاهی را توضیح دهد.

۹. سؤالهای تشریحی مورد استفادهٔ معلمان را، با توجه به ویژگیهای مطلوب آزمونهای تشریحی که در این فصل می‌آموزد، مورد بررسی قرار دهید و نظر انتقادی خود را دربارهٔ آنها بنویسد.

۱۰. برای هدفهای سطح بالای آموزشی یک درس سؤالهای تشریحی بنویسد.

در سه فصل قبلی این کتاب روشهای سنجش مبتنی بر انتخاب پاسخ یا آزمونهای عینی را معرفی کردیم. چند فصل باقیمانده را به روشها و فنون سنجش مبتنی بر ساختن پاسخ اختصاص خواهیم داد. شناخته‌شده‌ترین روش سنجش مبتنی بر ساختن پاسخ سنجش تشریحی یا آزمون انشایی است که موضوع بحث فصل حاضر کتاب است.

تعریف و انواع آزمونهای تشریحی

(هرچند که آزمون تشریحی (انشایی)[1] نزد همهٔ ما به خوبی شناخته شده است، باز هم چند ویژگی مهم آن را یادآور می‌شویم: (۱) هیچ پاسخ واحدی برای سؤالهای این نوع سنجش نمی‌توان در نظر گرفت که صددرصد کامل و به طور قطع درست باشد، (۲) پاسخها از لحاظ درجهٔ کیفیت یا درستی باهم فرق دارند، و (۳) در این نوع سنجش (به ویژه آزمونهای تشریحی گسترده پاسخ) به آزمون‌شونده آزادی عمل زیادی داده می‌شود (استالناکر[2]، ۱۹۵۱، به نقلِ مهرنز و لهمان، ۱۹۷۳، ۲۰۶).

آزمونهای تشریحی را، با توجه به آزادی عمل آزمون‌شونده در پاسخ دادن به سؤالهای آزمون، به دو دسته تقسیم می‌کنند. ویژگیهای این دو نوع آزمون را در زیر توضیح می‌دهیم.

آزمون تشریحی گسترده پاسخ

در آزمونهای تشریحی گسترده پاسخ[3]، هیچ‌گونه محدودیتی برای آزمون‌شونده منظور نمی‌شود، بلکه او عملاً آزاد است تا هر طور که مایل باشد پاسخ خود را بپروراند و سازمان دهد. آزمون شونده در پاسخ دادن به سؤالهای این نوع آزمون، هم از لحاظ زمان پاسخدهی و هم از جهت مقدار پاسخ، آزادی کامل دارد. در زیر به نمونه‌هایی از سؤالهای گسترده پاسخ توجه کنید.

- *آنچه را که به نظر شما باید در یک برنامهٔ امتحانی یک آموزشگاه گنجانیده شود بنویسید. نوع آزمون مناسب برای این برنامه را با ذکر دلایل منطقی خود توضیح دهید.*

(مهرنز و لهمان، ۱۹۸۴، ص ۹۶)

- *شرطی شدن کلاسیک پاولفی و شرطی شدن کنشگر اسکینری را با هم مقایسه کنید. در توضیحات خود از کاربرد هریک از این دو نوع شرطی شدن در زمینه‌های آموزشی و بالینی مثالهایی بیاورید.*

(کوهن و سوردلیک، ۲۰۰۵، ص ۲۰۵)

- *تأثیر قانون مندل را بر تحول زیست‌شناسی به عنوان یک علم شرح دهید.*

(لین و گرانلاند، ۲۰۰۰، ص ۲۳۸)

---

1. essay test 2. Stalnaker 3. extended response

آزمون‌شونده در پاسخ دادن به این سؤالها آزاد است تا مطالبی را ارائه دهد که به نظر او منطقی و مفید هستند و مطالب خود را به گونه‌ای که می‌خواهد سازمان دهد. همچنین آزمون‌شونده باید اطلاعات و دانشهای خود را ارزیابی کند و اندیشه‌های خود را به نحوی منطقی، منسجم، و بدیع روی کاغذ آورد. بنابراین، سؤالهای آزمونهای تشریحی گسترده پاسخ برای سنجش هدفهای تحلیل و آفرینندگی (بالاترین طبقه‌های طبقه‌بندی هدفهای آموزشی، حوزهٔ شناختی) مناسب‌ترین سؤالها هستند.

آزمون تشریحی محدود پاسخ

در آزمونهای تشریحی **محدود پاسخ**[1]، آزمون شونده در دادن پاسخ به سؤالها آزادی کامل ندارد، بلکه صورت سؤال او را ملزم می‌سازد تا پاسخ خود را در چهارچوب شرایطی خاص محدود کند. همچنین این سؤالها برای پاسخ آزمون‌شونده، هم از لحاظ زمان پاسخدهی و هم از نظر مقدار پاسخ، محدودیتهایی قائل می‌شوند. به نمونه‌های زیر از این نوع سؤالها توجه کنید:

- *پاولف در آزمایشهای خود مشاهده کرد سگی که شرطی شده بود تا با شنیدن صدای زنگ ترشح بزاق کند بعضی وقتها از این کار سر باز می‌زد. علت این امر چیست؟ جواب خود را حداکثر در نصف صفحه بنویسید.*

(اقتباس براساس مِهرنز و لهمان، ۱۹۸۴، ص ۹۶)

- *توضیح دهید چرا فشارسنج (بارومتر) یکی از ابزارهای مفید برای پیش‌بینی هواست. جواب خود را در یک پاراگراف بنویسید.*

(لین و گرانلاند، ۲۰۰۰، ص ۲۳۷)

این سؤالها آزمون‌شونده را محدود به جواب دادن دربارهٔ یک موضوع خاص می‌کنند، و از این لحاظ به آنها محدود پاسخ می‌گویند. افزون بر این، پاسخ آزمون‌شونده به نصف صفحه و یک پاراگراف محدود شده است.

ما در آزمونهای محدود پاسخ، با محدود کردن پاسخ آزمون‌شوندگان به یک موضوع مشخص، تصحیح پاسخهای آنان را هماهنگ‌تر و آسان‌تر می‌کنیم. اما این محدودیت سبب می‌شود که آزمون‌شونده نتواند توانایی خود را در ترکیب کردن اندیشه‌ها و بیان آنها به صورتی منسجم و منطقی ابراز کند. بنابراین، سؤالهای آزمون محدود پاسخ برای اندازه‌گیری توانایی یادگیرندگان در سطوح فهمیدن، کاربستن، و تحلیل مناسب‌اند، ولی برای سنجش توانایی ترکیب (خلاقیت) و ارزشیابی چندان مفید نیستند.

1. restricted response

قواعد تهیهٔ سؤالهای تشریحی

امتیاز مهم آزمونهای تشریحی این است که آزمون‌شوندگان را وامی‌دارند تا اندیشه‌های خود را به طور منطقی، منسجم، و سازمان یافته بیان کنند و در این کار خلاقیت خود را نشان دهند. نویسندهٔ سؤالهای تشریحی باید همواره این نکتهٔ مهم را در نظر داشته باشد. علاوه بر این، سؤالهای تشریحی را باید به گونه‌ای نوشت که تا حد امکان از معایب آنها که بعداً توضیح داده خواهند شد کاسته شوند. قواعد زیر به منظور تحقق این منظورها پیشنهاد شده‌اند.

۱. **دقت کنید که سؤالها به طور مستقیم به هدفهای آموزشی مربوط شوند.** از آنجاکه در آزمونهای تشریحی تعداد سؤالها کمتر از آزمونهای عینی (مثلاً صحیح ـ غلط یا چندگزینه‌ای) است، دقت عمل در نوشتن این نوع سؤالها بیشتر ضروری است، زیرا بد بودن یک سؤال در یک آزمون تشریحی ۵ سؤالی خیلی بیشتر از بد بودن یک سؤال در یک آزمون عینی ۵۰ سؤالی به نتیجهٔ ارزشیابی لطمه می‌زند. برای افزودن بر دقت خود، سعی کنید هر سؤال را با توجه به یک هدف دقیق آموزشی بنویسید. بنابراین، پیش از هر اقدام دیگر، جدول مشخصات موضوعی را که می‌خواهید برای آن آزمون تهیه کنید آماده سازید و هدفهای آموزشی را در آن وارد کنید. سطوح یادگیری هدفها یعنی طبقه‌هایی را که این هدفها به آنها تعلق دارند نیز مشخص کنید.

۲. **استفاده از آزمونهای تشریحی را تنها به اندازه‌گیری هدفهایی محدود کنید که با سایر انواع آزمونها به خوبی قابل اندازه‌گیری نیستند.** از سؤالهای تشریحی برای اندازه‌گیری هدفهای دانش استفاده نکنید. این‌گونه سؤالها را به هدفهای طبقات بالای طبقه‌بندی هدفهای آموزشی اختصاص دهید. سؤالهای تشریحی گسترده پاسخ بهترین سؤالها برای طبقهٔ تحلیل، ترکیب (آفرینندگی) و ارزشیابی هستند، اما از سؤالهای محدود پاسخ می‌توان برای طبقات پایین‌تر مانند کاربستن و فهمیدن استفاده کرد.

۳. **صورت سؤالهای تشریحی را با عبارات و کلمات واضح و روشن بنویسید و از کلی‌گویی و ابهام در بیان بپرهیزید.** کلمات و عبارات روشن منظور سؤال را بهتر نشان می‌دهند. کلمات و عبارات "تعریف کنید"، "حل کنید"، "خلاصه کنید" و "طبقه‌بندی کنید" از کلمات "بحث کنید" و "بررسی کنید" روشن‌تر هستند. اگر لازم بود که از "بحث یا بررسی کنید" استفاده شود، باید به طور دقیق بنویسید که چه نکاتی باید مورد بحث و بررسی قرار گیرند. همچنین در کاربرد کلمات "توضیح دهید" و "مقایسه کنید" بهتر است منظور خود را به صراحت بیان نمایید، در غیر این صورت ممکن است آزمون‌شوندگان برداشتهای متفاوتی از این کلمات داشته باشند و به گونه‌های مختلف به سؤالها پاسخ دهند.

۴. **از کاربرد کلمات "چه کسی"، "چه وقت"، "کجا"، و جز اینها بپرهیزید.** این کلمات تنها

یادآوری معلومات را اندازه می‌گیرند. به جای این کلمات، بهتر است از کلمات "چرا"، "چگونه"، "به چه دلیل" و از این قبیل استفاده کنید. این کلمات در مقایسه با کلمات قبلی هدفهای سطح بالاتری را می‌سنجند.

۵. **تا حد امکان از سؤالهای تازه و موقعیتهای جدید استفاده کنید.** سؤالهایی که در کتاب یا کلاس توضیح داده شده‌اند صرفاً معلومات و حافظهٔ آزمون‌شونده را می‌سنجند، اما استفاده از موقعیتهای جدید در طرح سؤالها قدرت کاربرد و درک و فهم را اندازه می‌گیرند. در طرح سؤال برای سنجش کاربستن و خلاقیت، رعایت این نکته بسیار ضروری است.

۶. **سؤالهای مربوط به موضوعات و عقاید بحث‌انگیز را طوری طرح کنید که از آزمون‌شونده بخواهند تا شواهد لازم برای مستند کردن عقیدهٔ انتخابی را بیان کند، نه اینکه از او بخواهید تا صرفاً عقاید شخصی خودش را شرح دهد.** این نوع سؤالها باید به طور روشن از پاسخ دهنده بخواهند تا موضوع انتخابی خود را با دلایل علمی و منطقی مستند سازد. پاسخ دهنده باید بداند که پاسخ داده شده تنها به این خاطر که او نظر خاصی را می‌پسندد نمی‌تواند مورد پذیرش قرار گیرد.

۷. **به آزمون‌شوندگان حق انتخاب چند سؤال از میان تعدادی سؤال را ندهید.** از همهٔ آزمون‌شوندگان بخواهید تا به تمامی سؤالها پاسخ دهند. این امر به شما اجازه می‌دهد تا با کاربرد ملاکی یکنواخت به تصحیح اوراق بپردازید و پاسخها را بهتر با یکدیگر مقایسه کنید. اگر آزمون‌شوندگان به سؤالهای مختلف جواب دهند امکان مقایسهٔ آنها با یکدیگر به طور کامل میسر نخواهد بود.

از آنجا که یکی از معایب عمدهٔ سؤالهای تشریحی عدم نمونه گیری درست از آموخته‌های آزمون‌شوندگان است، و از آنجا که هنگام دادن حق انتخاب به آزمون‌شوندگان آنان سؤالهایی را جواب می‌دهند که آمادگی بیشتری برای آنها دارند، دادن حق انتخاب به آزمون‌شوندگان مشکل نمونه گیری را حادتر می‌کند.

مهرنز و لهمان (۱۹۸۴) مشکلات دادن حق انتخاب چند سؤال از میان تعدادی سؤال را در سه مورد زیر خلاصه کرده‌اند:

الف) تهیهٔ سؤالهایی که از نظر درجهٔ دشواری همسنگ باشند کار آسانی نیست.

ب) آزمون‌شوندگان توانایی انتخاب سؤالهایی را که جواب آنها را بهتر می‌دانند ندارند.

پ) از آنجا که آزمون‌شوندگان قوی‌تر بیشتر به سراغ سؤالهای دشوار می‌روند، ممکن است داشتن حق انتخاب در نهایت به ضرر این گونه افراد تمام شود.

با این حال، در مواردی که هدف سؤالها اندازه گیری میزان یادگیری مطالب نیست، بلکه

منظور از آزمون صرفاً اندازه‌گیری مهارت پاسخ‌دهندگان در نوشتن و پروراندن مطالب است، دادن حق انتخاب تعدادی سؤال از میان سؤالهای تشریحی ممکن است قابل دفاع باشد.

۸. **برای پاسخ دادن به سؤالها، زمان کافی در نظر بگیرید و زمان هر سؤال را نیز به طور جداگانه مشخص کنید.** وقتی را که بدین منظور تعیین می‌کنید باید هم شامل دادن جواب باشد و هم فرصتی برای فکر کردن آزمون‌شونده در نظر بگیرد. از آنجا که دادن وقت کافی ممکن است سبب شود که تعداد کمی سؤال در هر آزمون به کار رود، بهتر است برای اندازه‌گیری هدفهای سطح بالای آموزشی در هر دفعه امتحان تعداد سؤالهای کمتری به کار برید، و مشکل نمونه‌گیری را با انجام آزمونهای متعدد و مکرر برطرف کنید.

۹. **با نوشتن سؤالهایی که به جواب کوتاه نیاز دارند، تعداد سؤالها را افزایش دهید تا از مشکل ضعف نمونه‌گیری آزمونهای تشریحی کم کنید.** هرچه تعداد سؤالهای یک آزمون بیشتر باشد نمونهٔ بهتری از محتوا و هدفهای درس را شامل خواهد شد و نتیجهٔ بهتری به دست خواهد داد. در ضمن سؤالهای دارای پاسخ کوتاه روشن‌تر هستند و جواب دادن به آنها از سوی آزمون‌شوندگان مختلف یکنواخت‌تر انجام می‌گیرد.

۱۰. **عواملی را که در ارزشیابی آزمونهای تشریحی دخالت خواهید داد از پیش تعیین کنید و آنها را به اطلاع آزمون‌شوندگان برسانید.** ما این موضوع را در بحث پیرامون قواعد تصحیح جواب سؤالهای تشریحی نیز توضیح خواهیم داد. هدف از بیان آن در اینجا این است که آزمون‌شوندگان پیش از جواب دادن به یک سؤال بدانند که چه چیزی را باید در جوابهای خود منظور کنند که نظر طراح آزمون را تأمین کند. برای مثال، مشخص کنید که آیا عواملی چون املای درست کلمات، انشای درست مطالب، نکات دستوری، خوش خطی و روشنی بیان در ارزشیابی جواب سؤالات مؤثر خواهند بود یا نه. به یاد داشته باشد که، بجز در مواردی که توانایی نوشتن، املاء و انشاء را امتحان می‌کنید، از دخالت دادن این عوامل در تصحیح اوراق امتحانات تشریحی بپرهیزید.

۱۱. **از روش کارپوشه برای تکمیل آزمونهای تشریحی استفاده کنید.** یکی از اشکالات آزمونهای تشریحی نمونه‌گیری محدود آنها از محتوا و هدفهای درس است. برای غلبه بر این مشکل در زمینهٔ انشاء یا مقاله‌نویسی، از دانش‌آموزان بخواهید تا در طول یک مدت نسبتاً طولانی، مثلاً در طول یک ترم یا سال تحصیلی، تعدادی انشاء بنویسند. چند انشاء مختلف از یک انشاء واحد که در یک جلسهٔ امتحانی نوشته می شود بهتر می‌تواند توانایی دانش‌آموزان را بسنجد. شما می‌توانید این مجموعه انشاء را که در طول ترم تحصیلی جمع‌آوری شده‌اند ارزشیابی کنید. (روش سنجش با کارپوشه را به تفصیل در فصل ۱۳ توضیح خواهیم داد.)

قواعد تصحیح پاسخهای سؤالهای تشریحی

برخلاف آزمونهای عینی که در آنها اندازه گیری توانایی آزمون شوندگان مستقل از نظر شخصی تصحیح کنندگان ورقه‌ها صورت می‌گیرد، در آزمونهای تشریحی بخش اعظم اندازه گیری توانایی مورد سنجش به شخصی که پاسخهای سؤالها را می‌خواند و به آنها نمره می‌دهد وابسته است. بنابراین، دقت روش اندازه گیری با آزمونهای تشریحی به مقدار زیاد به نحوهٔ تصحیح برگه‌های امتحانی و دقت عمل مصححان مربوط است. دلیل عمدهٔ مخالفت برخی صاحب‌نظران با کاربرد این آزمونها نیز از بی‌ثباتی نمره گذاری آنها ناشی می‌شود. لذا، رعایت نکات زیر که به منظور بهبود کیفیت نمره گذاری این نوع آزمونها پیشنهاد شده از عوامل بسیار مؤثر است.

۱. **پاسخهای سؤالهای تشریحی را تنها براساس هدفی که در سؤال گنجانیده شده است تصحیح کنید.** از دخالت دادن نکات دستوری، املایی، انشایی، خوش‌نویسی، و جز اینها خودداری کنید. این نکات را تنها وقتی که امتحان مربوط به دستور زبان فارسی، املاء، و خوش‌نویسی است منظور کنید. زمانی که علاوه بر محتوا و مطالب درسی، اندازه گیری نکات دستوری فوق نیز مورد نظر است، برای هر مورد نمرهٔ جداگانه‌ای در نظر بگیرید. این مطلب باید هنگام امتحان به اطلاع آزمون‌شوندگان برسد. در هر صورت، معلم باید دقت کند که عوامل فوق ارزشیابی او را از محتوا تحت تأثیر قرار ندهند.

۲. **پاسخها را سؤال به سؤال تصحیح کنید نه برگه به برگه.** یعنی ابتدا پاسخ سؤال اول همهٔ برگه‌ها یا ورقه‌ها را تصحیح کنید، بعد پاسخهای سؤال دوم همهٔ برگه‌ها را، و همین طور تا آخر. این روش امکان مقایسهٔ پاسخهای آزمون‌شوندگان مختلف به یک سؤال را فراهم می‌آورد، کار نمره گذاری به آن سؤال را آسان می‌سازد، و امکان کاربرد یک ملاک واحد را به صورت هماهنگ برای کلیهٔ افراد میسر می‌کند.

۳. **هنگام تصحیح برگه‌های امتحانی از شناسایی نام صاحبان آنها خودداری کنید.** اگر معلم در هنگام تصحیح برگه‌های امتحانی صاحبان آنها را نشناسد از دخالت نظر شخصی در نمره‌دادن که یکی از معایب بزرگ آزمونهای تشریحی است کاسته خواهد شد. برای این منظور از سربرگهای ورقه‌ها که مشخصات پاسخ‌دهندگان بر روی آنها نوشته می‌شود و در وقت تصحیح از آنها جدا می‌گردد استفاده کنید. همچنین می‌توان از دانش‌آموزان خواست تا نام خود را در پشت ورقه‌ها بنویسند یا از نوعی رمز عددی استفاده کنند. حتی پیشنهاد می‌شود که پیش از شروع به تصحیح ورقه‌ها آنها را خوب به هم بزنید و مخلوط کنید، تا ترتیب تقدم و تأخر آنها به هم بخورد و بر شما معلوم نباشد که برگه‌های رو از آن دانش‌آموزان قوی است یا ضعیف.

۴. **در صورت امکان، از یکی دو نفر از همکارانتان بخواهید تا سؤالهایی را که شما تصحیح کرده‌اید تصحیح کنند.** این روش تنها راه بررسی عینیت و دقت ارزشیابی شما از پاسخهای دانش‌آموزان است. این کار باید بدون آگاهی طرفین از نمرات یکدیگر و بدون مشورت با هم صورت پذیرد. این روش نمره‌گذاری در شرایطی که قرار است نتایج ارزشیابی برای تصمیمهای مهم و غیرقابل بازگشت به کار رود، مانند انتخاب افراد برای ادامهٔ تحصیل یا برای دریافت سایر امتیازهای تحصیلی، بسیار بااهمیت است. در این مورد، میانگین نمرات مصححان مختلف از نمرات هر یک از آنان دقیق‌تر خواهد بود. ایبل (۱۹۷۹) پیشنهاد می‌کند که اگر تصحیح همهٔ برگه‌ها توسط سایر اشخاص ممکن نیست؛ دست کم نمونه‌ای از آنها را برای تصحیح به کس دیگری واگذار کنید.

۵. **تمام پاسخهای آزمون‌شوندگان به یک سؤال را در یک نشست و بدون وقفهٔ زمانی تصحیح کنید.** انسانها از روزی به روز دیگر و حتی از ساعتی به ساعتی دیگر تغییر می‌کنند. اگر پاسخهای تعدادی از آزمون‌شوندگان را امروز و تعدادی دیگر را روز بعد تصحیح کنید، ممکن است عوامل مختلف در نحوهٔ ارزشیابی شما تأثیر بگذارند. مشکلات خانوادگی، حالات جسمانی، و اتفاقات غیرمنتظره در این کار دخالت می‌کنند. البته این بدان معنی نیست که معلم نباید به منظور تنفس و رفع خستگی برای مدت کوتاهی از کار دست بکشد. منظور این است که در شرایط مختلف به تصحیح پاسخهای آزمون‌شوندگان به یک سؤال واحد نپردازد. وقتی که تصحیح پاسخ یک سؤال را آغاز می‌کنید تمام پاسخهای آزمون‌شوندگان به آن سؤال را تصحیح کنید و اگر خواستید می‌توانید بقیهٔ سؤالها را برای وقت دیگری بگذارید. به هر حال، اگر ناچار شدید تصحیح تعدادی از پاسخهای سؤالی را که مقداری از آنها را تصحیح کرده‌اید به وقت دیگری واگذارید، در آغاز شروع مجدد به ارزشیابی، تعدادی از برگه‌هایی را که قبلاً تصحیح کرده‌اید بازخوانی کنید تا با نحوهٔ نمره‌گذاری و شیوهٔ ارزشیابی قبلی خود آشنا شوید و برگه‌های باقیمانده را نیز به همان سبک ارزشیابی کنید.

۶. **به نمرات سؤالهایی که قبلاً تصحیح کرده‌اید نگاه نکنید.** نمرات سؤالهایی را که قبلاً تصحیح کرده‌اید در جایی قرار دهید که هنگام تصحیح سؤالهای بعدی دور از دید شما باشند. عدم دسترسی به نمرات سؤالهای قبلی موجب می‌شود که معلم هر سؤال تازه را به طور مستقل و بدون تأثیر گرفتن از نمرات سؤالهای قبلی تصحیح کند. یک راه انجام این کار آن است که نمرات را بر روی یک صفحهٔ مجزا بنویسید و پس از جواب دادن به تمام پاسخهای یک سؤال برگه‌های آزمون را به طور کامل به هم بزنید. اگر پاسخ سؤال در حد یک صفحه باشد، می‌توان هر سؤال را بر روی یک صفحه نوشت و از پاسخ دهندگان خواست که جواب هر سؤال را در همان صفحه بنویسند. این روش برای آزمونهای ورودی که در آنها طراحان سؤالها و مصححان متفاوت هستند بسیار مفید است.

۷. بر روی برگه‌های آزمون اشتباهات دانش‌آموزان را تصحیح کنید و اظهارنظرهای خود را بنویسید. اظهارنظرها و توضیحات کتبی معلمان بر روی برگه‌های امتحانی دانش‌آموزان منجر به بهبود عملکرد آنان می‌شود (پیج[1]، ۱۹۵۸؛ وولفلک، ۱۹۸۷). یکی از هدفهای آزمون کردن و ارزشیابی پیشرفت تحصیلی دانش‌آموزان فراهم آوردن اطلاعات یا بازخورد دربارهٔ پیشرفت آنان به منظور ایجاد انگیزش و آموزش غیرمستقیم است. این کار همچنین به معلم کمک می‌کند تا الگوی اشتباهات دانش‌آموزان را کشف کند و به نقائص روش آموزش خود و روش یادگیری دانش‌آموزان پی ببرد.

نکتهٔ دیگری که دربارهٔ دادن بازخورد از نتیجهٔ آزمون به دانش‌آموزان پیشنهاد شده، مرور کردن سؤالها و جوابها با فردفرد دانش‌آموزان است. چند دقیقه صحبت کردن با هر دانش‌آموز دربارهٔ چگونگی پاسخها و نقاط قوت و ضعف او بسیار مفید و آموزنده است. این صحبت شفاهی از هر نوع اظهارنظر کتبی مفیدتر است.

۸. با نوشتن یک نمونه یا الگو برای هر سؤال به عنوان کلید، از دخالت عوامل نامربوط جلوگیری کنید. در امتحانات تشریحی محدود پاسخ تهیهٔ یک کلید پیش‌ساخته عملی‌تر از انجام این کار در امتحانات گسترده‌پاسخ است. با این حال، سعی کنید برای سؤالهای گسترده‌پاسخ نیز یک الگوی پاسخ درست کنید که دست‌کم در آن نکات مهمی را که می‌خواهید به آنها امتیاز بدهید معلوم نمایید. این کلید در یکنواختی نمره‌گذاری کمک زیادی به مصحح می‌کند. در تهیهٔ پاسخهای نمونهٔ خود سعی کنید پاسخ هر سؤال را به چند قسمت تقسیم نمایید و برای هر قسمت امتیاز جداگانه‌ای منظور کنید و به پاسخهای دانش‌آموزان با توجه به این امتیازها نمره بدهید. دقت کنید که در تقسیم نمرهٔ کل هر پاسخ به اجزای آن کلیهٔ عوامل مهم را در نظر بگیرید. مثلاً برای سازمان دادن پاسخ، مثالهای داده شده، روابط بین اجزاء، دلایل ذکر شده، و از این قبیل امتیازهای لازم را منظور کنید.

۹. از راهنمای نمره‌گذاری استفاده کنید. برای نمره‌گذاری و تصحیح پاسخ سؤالهای تشریحی **راهنمای نمره‌گذاری**[2] بسیار مفید است. "راهنمای نمره‌گذاری به یک مجموعه قواعد منسجم گفته می‌شود که برای سنجش کیفیت پاسخ یا عملکرد دانش‌آموز به کار می‌رود. ما معمولاً برای سنجش یک هدف یادگیری یا بخشی از عملکرد از یک راهنما استفاده می‌کنیم، اما عملکردهای پیچیده از چند قسمت یا چند هدف تشکیل می‌شوند؛ لذا برای این منظور چند راهنمای نمره‌گذاری لازم‌اند" (نیتکو، ۲۰۰۱، صص ۲۴۳-۲۴۲). ما در زیر مهم‌ترین راهنماهای نمره‌گذاری را توضیح می‌دهیم.

1. Page 2. scoring rubric

روش تحلیلی

در روش تحلیلی[1] نمره‌گذاری که به آن روش امتیازبندی نیز گفته می‌شود پاسخ نمونهٔ خود را به اجزای کوچک‌تری تقسیم کنید و برای هر جزء مشخص نمره یا امتیاز جداگانه‌ای در نظر بگیرید. همچنین در این روش برای عواملی چون "قدرت بیان"، "ساختمان منطقی پاسخ" و "ذکر دلایل" امتیازهای جداگانه‌ای منظور کنید.

روش کلی

در روش کلی یا سراسری[2] نمره‌گذاری که به آن روش درجه‌بندی نیز گفته می‌شود، پاسخ نمونه به اجزاء و قسمتهای کوچک تقسیم نمی‌شود بلکه تنها به صورت یک معیار به کار می‌رود. در این روش، معلم تمامی پاسخ هر فرد به یک سؤال را می‌خواند و یک برداشت کلی از آن کسب می‌کند و بعد این برداشت کلی را به یک نمره تبدیل می‌نماید. در این روش به هیچ عامل واحدی امتیاز خاصی داده نمی‌شود، بلکه همهٔ عوامل مورد نظر قرار می‌گیرند و کل پاسخ یکباره مورد قضاوت واقع می‌شود.

در روش سراسری می‌توان شیوه‌های مختلفی را برای نمره‌گذاری پیشنهاد کرد: می‌توان یک نمرهٔ دوگانه مورد استفاده قرار داد (قبول ـ مردود)، یا مقیاس را تا ۵ واحد به شرح زیر افزایش داد.

خیلی خوب	۵
خوب	۴
متوسط	۳
ضعیف	۲
خیلی ضعیف	۱

ایبل (۱۹۷۹) دربارهٔ نمره‌گذاری با روش کلی گفته است که در مجموع این روش نمره‌گذاری آسان‌تر و سریع‌تر از روش تحلیلی است، اما دلایل توجیهی روشنی برای دادن نمرات به دست نمی‌دهد و برای دانش‌آموزان نیز تصویر روشنی از چگونگی کامل نبودن جوابشان فراهم نمی‌آورد. از این بابت روش تحلیلی نمره‌گذاری بهتر است.

گی (۱۹۹۱) گفته است در حالی که هر دو روش تحلیلی و کلی نمره‌گذاری را می‌توان هم با آزمونهای وابسته به ملاک به کار بست و هم با آزمونهای وابسته به هنجار، اما روش تحلیلی برای آزمونهای وابسته به ملاک مناسب‌تر است، زیرا نمرات بیشتر بر پاسخهای واقعی آزمون‌شوندگان مبتنی هستند، و روش کلی بیشتر مناسب آزمونهای وابسته به هنجار است. او

1. analytical 2. global

همچنین می‌گوید، زمانی که می‌خواهید روش نمره‌گذاری را با آزمونهای وابسته به هنجار به کار بندید، ابتدا نمونه‌ای از پاسخها را مطالعه کنید تا نسبت به کیفیت کل سؤالها یک برداشت حاصل نمایید، بعد هر پاسخ را به عنوان خیلی خوب، خوب، متوسط، ضعیف، یا خیلی ضعیف مورد داوری قرار دهید.

ویرسما[1] و جورس[2] (۱۹۹۰) بهترین شیوهٔ نمره‌گذاری را ترکیب دو روش فوق دانسته‌اند. این صاحب‌نظران گفته‌اند ابتدا همهٔ پاسخها را با روش کلی تصحیح کنید تا یک احساس کلی نسبت به کیفیت و گسترهٔ آنها کسب نمایید. با این کار می‌توانید پاسخها را در طبقات نسبتاً کلی دسته‌بندی کنید و سپس پاسخها را به طور تحلیلی تصحیح نمایید و این بار بر نکات یا اجزای دقیق‌تر تأکید کنید. به باور ویرسما و جورز، این شیوهٔ نمره‌گذاریِ ترکیبی سبب جلوگیری از معایب هر دو روش می‌شود و روش نمره‌گذاری معلم را قوت می‌بخشد.

روش ویژگیهای اصلی

روش دیگر نمره‌گذاری سؤالهای آزمونهای تشریحی روش **ویژگیهای اصلی**[3] نام دارد. در این روش، معلم یا مصحح ویژگیهای اصلی پاسخ دانش‌آموز یا دانشجو به هر سؤال را سنجش می‌کند و برای هر یک از آنها نمره‌های ۴ تا ۰ را که معرف عالی تا غیرقابل‌قبول‌اند منظور می‌نماید. ساکس (۱۹۹۷)، به عنوان نمونه، گفته است اگر در سؤال تشریحی معلم از دانش‌آموز خواسته است تا نامه‌ای دوستانه بنویسد، هنگام نمره‌گذاری، باید به دنبال ویژگیهای زیر باشد: بذله گویی، توصیف دقایق، و دوستانه بودن. مطالب مربوط به راهنمایی استفاده از یک برنامهٔ کامپیوتری را می‌توان براساس روشنی بیان، سازمان مطالب، و مفید بودن مثالها ارزشیابی کرد. همچنین یک نامهٔ متقاعد کننده یا مجاب‌گر باید بتواند خواننده را متقاعد کند یا نگرش او را تغییر دهد.

امتیازها و محدودیتهای آزمونهای تشریحی

صاحب‌نظران و متخصصانْ نظرهای موافق و مخالف زیادی نسبت به کاربرد آزمونهای تشریحی ابراز داشته‌اند. مدافعان این نوع آزمونها امتیازهای آنها را به شرح زیر برشمرده‌اند:

۱. تهیهٔ آزمونهای تشریحی از تهیهٔ آزمونهای عینی به ویژه آزمونهای چندگزینه‌ای، آسان‌تر است.
۲. این نوع آزمون بهترین وسیلهٔ موجود برای سنجش توانایی آزمون‌شونده در پروراندن جواب سؤالها و سازمان دادن مطالب و بیان آنهاست.
۳. این آزمونها توانایی تولید پاسخها را می‌سنجند نه توانایی انتخاب پاسخها را.
۴. آزمونهای تشریحی از آزمونهای عینی موقعیتهای واقعی‌تری را در اختیار آزمون‌شوندگان می‌گذارند.

---

1. Wiersma 2. Jurs 3. primary traits

۵. آزمونهای تشریحی بر روش مطالعهٔ یادگیرندگان تأثیر مثبت به جای می‌گذارند. اگر یادگیرندگان از پیش بدانند که با استفاده از آزمونهای تشریحی یادگیری آنان سنجش خواهد شد، با احتمال بیشتری در میان مطالب به جستجوی روابط خواهند پرداخت و اطلاعات را به راههای معنی‌دار سازمان خواهند داد (اِگن و کاوچاک، ۲۰۰۱، ص ۶۰۷). به عنوان نمونه، در پژوهشی که به منظور بررسی تأثیر روشهای متفاوت سنجش بر چگونگی مطالعه و یادگیری دانشجویان دانشگاه انجام گرفت (فتح‌آبادی، ۱۳۸۵)، معلوم شد که استفاده از امتحانات تشریحی دانشجویان را به سمت رویکرد عمقی مطالعه سوق می‌دهد.

۶. استفاده از آزمونهای تشریحی عادتهای مطالعهٔ بهتری را در دانش‌آموزان و دانشجویان تشویق می‌کند.

در مقابل امتیازهای بالا، برای آزمونهای تشریحی محدودیتهایی را نیز ذکر کرده‌اند. مهم‌ترین آنها به قرار زیرند:

۱. یک آزمون تشریحی نمونهٔ کوچکی از محتوای درس و هدفهای آموزشی را اندازه می‌گیرد.

۲. تصحیح پاسخهای تشریحی نمی‌تواند با دقت و عینیّت انجام گیرد.

۳. تصحیح پاسخهای تشریحی بسیار وقت‌گیر است.

معایب ذکر شدهٔ بالا به مقدار زیادی از کارآیی آزمونهای تشریحی می‌کاهند. مشکل نمونه‌گیری محدود برای این نوع آزمونها یک مشکل جدی است. از آنجا که در یک زمان محدود امتحانی نمی‌توان تعداد زیادی سؤال تشریحی به آزمون‌شوندگان داد، لذا همیشه تعداد سؤالهای یک آزمون تشریحی، به ویژه آزمونهای تشریحی گسترده پاسخ، بسیار محدود است. در نتیجه آزمونهای تشریحی معمولاً نمونهٔ ضعیفی از محتوا و هدفهای درس را اندازه می‌گیرند. بنابراین، لازم است به هنگام کاربرد این آزمونها، تا آنجا که ممکن است از سؤالهای محدود پاسخ استفاده کرد که تعداد بیشتری از آنها را در یک جلسهٔ امتحان می‌توان مورد استفاده قرار داد؛ و به جای یک نوبت چندین نوبت امتحان به عمل آورد، تا نمونهٔ بزرگ‌تری از محتوا و هدفهای آموزشی سنجش شوند.

مشکل دیگر آزمونهای تشریحی ذهنی بودن نمره‌گذاری مصححان و یکنواخت نبودن ارزشیابی آنان است. عوامل بسیار زیادی در این زمینه دخالت می‌کنند که باعث می‌شوند نمره‌گذاری مصححان مختلف به یک برگهٔ امتحانی یا حتی نمره‌گذاری یک مصحح به یک برگه در زمانهای متفاوت یکسان نباشد.[1] برای کاستن از این مشکل بهتر است هنگام نوشتن صورت سؤالهای این آزمونها دقت لازم را به کار بندید تا تمام آزمون‌شوندگان منظور سؤال را درک کنند و در نوشتن پاسخهای خود یکنواخت عمل نمایند. همچنین در تصحیح پاسخهای سؤالهای

---

۱. برای کسب اطلاعات جامع در این باره به کتاب روانشناسی ارزشیابی تحصیلی [تألیف نوازه (Noizet) و کاورنی (Caverni)، ۱۹۷۸، ترجمهٔ گنجی، ۱۳۶۴] مراجعه کنید.

این آزمونها لازم است از یک شیوهٔ نمره‌گذاری یکنواخت که از قبل تعیین می‌شود استفاده کنید.

سومین عیب آزمونهای تشریحی، یعنی وقت‌گیر بودن تصحیح پاسخ سؤالها، یک مشکل جدی دیگر است. برخلاف پاسخهای آزمونهای عینی که به وسیلهٔ ماشین یا افراد غیرمتخصص قابل تصحیح است، تصحیح برگه‌های امتحانی آزمونهای تشریحی نیاز به افراد متخصص دارد. بنابراین، زمانی که هدفهای آموزشی معلم با آزمونهای عینی قابل اندازه‌گیری است، بهتر است از این آزمونها استفاده کند و آزمونهای تشریحی را صرفاً به هدفهای سطح بالای یادگیری اختصاص دهد.

به رغم این مشکلات، آزمونهای تشریحی از ابزارهای مهم سنجش یادگیری دانش‌آموزان و دانشجویان هستند. چنان که در صفحات قبلی این کتاب توضیح دادیم، بسیاری از هدفهای آموزشی را نمی‌توان با سؤالهای گزینش پاسخ یا پاسخ‌گزین سنجش کرد و روشهای سنجش تولید پاسخ یا پاسخ‌ساز همواره مورد نیاز معلم‌اند.

تنها به کمک روشهای سنجش مبتنی بر تولید پاسخ که آزمونهای تشریحی از انواع خوب آنها هستند می‌توان توانایی سازمان‌دهی، پروراندن مطالب، خلاقیت، و قدرت بیان یادگیرندگان را سنجش کرد. حتی با هدفهایی که ممکن است با روشهای سنجش مبتنی بر گزینش پاسخ قابل اندازه‌گیری باشند، مانند یادآوری، فهمیدن، حل مسئله و تحلیل، بازهم این روشها صرفاً توانایی بازشناسی یادگیرندگان را می‌سنجند نه بازخوانی را. با توجه به اینکه بازخوانی جنبهٔ مهم یادگیری و یادآوری یادگیرندگان است، معلمان در فعالیتهای سنجش و اندازه‌گیری خود ناگزیر از کاربرد روشهای سنجش وابسته به تولید پاسخ از جمله روشهای سنجش تشریحی هستند.

موارد استفادهٔ آزمونهای تشریحی

چنان‌که گفتیم، از آزمونهای تشریحی می‌توان برای سنجش انواع مختلف فرایندها و فراورده‌های یادگیری دانش‌آموزان و دانشجویان سود برد. جدول ۱ـ۱۰ یادگیریهایی را که می‌توان با این آزمونها سنجش کرد نشان می‌دهد. این جدول الگوی کلی سؤالهای تشریحی را که برای مقاصد گوناگون سنجش قابل استفاده‌اند نشان می‌دهد. در زیر به چند نمونهٔ واقعی از سؤالهای تشریحی که برای سنجش هدفهای مختلف نوشته شده‌اند توجه کنید.

فهمیدن دانش امور واقعی

۱. توضیح دهید چرا اندکی پس از آنکه ظرفی را بر روی شمعی قرار می‌دهیم آن شمع خاموش می‌شود. (لین و گرانلاند، ۲۰۰۰، ص ۲۴۴)

۲. تأثیر نمره‌های خیلی بالا و خیلی پایین را بر میانه و میانگین با هم مقایسه کنید.

جدول ۱ـ۱۰ انواع یادگیریهایی که به وسیلهٔ آزمونهای تشریحی محدود پاسخ و گسترده پاسخ سنجش می‌شوند (اقتباس براساس لین و گرانلاند، ۲۰۰۰، ص ۲۴۰)

نوع سؤال تشریحی	یادگیریهایی که می‌توان سنجش کرد
سؤال تشریحی محدودپاسخ	توانایی:
	تبیین روابط علت و معلولی
	توصیف چگونگی کاربرد اصول
	ارائهٔ دلایل لازم
	تدوین فرضیه‌های قابل دفاع
	تدوین نتیجه گیریهای معتبر
	بیان فرضهای لازم
	توصیف محدودیت داده‌ها
	تبیین روشها و شیوه‌ها
	(و انجام کارهای مشابه براساس توانایی دانش‌آموز)
سؤال تشریحی گسترده‌پاسخ	توانایی:
	تولید، سازمان دادن، و بیان اندیشه‌ها
	ترکیب کردن یادگیریهای حاصل از زمینه‌های متفاوت
	آفریدن شکلهای نو یا اصیل (مانند طراحی یک آزمایش تازه)
	خلاصه کردن (مانند خلاصه کردن یک داستان)
	ساختن داستانهای ابتکاری
	تبیین مفاهیم یا اصول (مثلاً نوشتن یک مقالهٔ تشریحی)
	متقاعد کردن خواننده (مثلاً نوشتن یک مقالهٔ متقاعدکننده)
	و انجام کارهای مشابه براساس توانایی دانش‌آموز برای نوشتن یک مقاله به منظور هدفی خاص

فهمیدن دانش مفهومی

۳. دلایلی را که موجب شده بود تا آمریکاییان اولیه در دههٔ ۱۷۸۰ به غرب سفر کنند توضیح دهید.

۴. در حداکثر ۵ سطر امتیازها و محدودیتهای آزمونهای تشریحی را خلاصه کنید.

۵. معنی گفتهٔ زیر را توضیح دهید:

"اگر نمی‌توانید حرارت را تحمل کنید، از آشپزخانه خارج شوید."

(لین و گرانلاند، ۲۰۰۰، ص ۲۴۴)

کاربستن دانش روندی

۶. فرض کنید تولدتان را جشن گرفته‌اید و ۹ نفر از دوستان خود را دعوت کرده‌اید. دو پیتزای بزرگ سفارش داده‌اید اما هر یک از آنها به چهار قسمت تقسیم شده است. مشکل شما چیست؟ چه تدبیری می‌اندیشید تا هر نفر یک تکه پیتزا داشته باشد؟

(کوبیسزین[1] و بوریچ[2]، ۲۰۰۳، ص ۱۳۱)

۷. وزن شما در سطح کرهٔ ماه و در سطح خورشید بیشتر یا کمتر از وزن شما در سطح زمین خواهد بود؟ دلایل خود را بنویسید.

(پین، ۲۰۰۳، ص ۲۳۵)

۸. طرحی تهیه کنید که به وسیلهٔ آن بتوانید تعیین نمایید که آیا طرفداران دو حزب معروف سیاسی به طور مساوی در تمام شهر پراکنده‌اند یا طرفداران هر یک از این دو حزب در محلهٔ خاصی متمرکزند.

(اقتباس براساس نیتکو، ۲۰۰۱، ص ۱۸۷)

تحلیل دانش مفهومی

۹. مقالهٔ پیوست را بخوانید. بیاناتی را که نشان‌دهندهٔ عقیدهٔ نویسنده‌اند مشخص کنید. دلایل خود را ذکر نمایید.

(نیتکو، ۲۰۰۱، ص ۱۹۲)

۱۰. رنگهای آبی و خاکستری به آب و هوای سرد مربوط‌اند. چه رنگهای دیگری به آن ربط دارند؟ این رنگها بر شکلی که می‌خواهید بکشید چه تأثیری خواهند گذاشت؟

(کوبیسزین و بوریچ، ۲۰۰۳، ص ۱۳۱)

ارزشیابی دانش مفهومی

۱۱. در تعیین رفتار انسانها طبیعت بیشتر نقش دارد یا تربیت؟ نظر خود را با ذکر دلیل توضیح دهید.

(پین، ۲۰۰۳، ص ۲۳۵)

آفریدن دانش مفهومی

۱۲. نامه‌ای به مدیر مدرسه‌تان بنویسید و موافقت او را برای بردن دانش‌آموزان به یک گردش علمی جلب کنید.

(لین و گرانلاند، ۲۰۰۰، ص ۲۴۴)

---

1. Kubiszyn 2. Borich

۱۳. با مشاهدهٔ ابرهای موجود در آسمان دربارهٔ وضع طوفانی که در راه است چه پیش‌بینی‌ای می‌کنید؟ (چه فرضیه‌ای می‌سازید؟) چه چیزی در این پیش‌بینی به شما کمک می‌کند؟ (کوبیسزین و بوریچ، ۲۰۰۳، ص ۱۳۱)

۱۴. برای اجرا با بی‌سوادها، روانشناسان آزمونهایی ساخته‌اند که به صورت ایما و اشاره قابل استفاده‌اند، و در آنها تصاویر به جای نوشته به کار رفته است. شما می‌توانید برای اجرا با دانش‌آموزانی که قادر به خواندن نیستند روش دیگری ابداع کنید؟ روش خود را به طور کامل و همراه با مثال توضیح دهید. (مهرنز و لهمان، ۱۹۷۳، ص ۲۰۹)

پرسش شفاهی

یکی از تدابیر متداول و مؤثر معلمان در ارزشیابیهای غیررسمی از یادگیری دانش‌آموزان، به منظور دادن بازخورد به آنان، تشویق و ترغیب آنان در یادگیری، و رفع مشکلات یادگیری آنان استفاده از **پرسش شفاهی**[1] است. معمولاً پرسش شفاهی را نمی‌توان نوعی آزمون به حساب آورد که از نتایج آن بتوان برای نمره‌گذاری میزان آموخته‌های دانش‌آموزان استفاده کرد. اما از این نوع پرسش برای بهبود روشهای یادگیری دانش‌آموزان و شیوه‌های آموزشی معلم می‌توان به خوبی استفاده کرد، زیرا بازخوردهای فوری برای معلم و دانش‌آموز فراهم می‌آورد.

می‌توان پرسش شفاهی را نوعی سؤال تشریحی دانست که به صورت غیررسمی طرح و جواب آن هم به صورت شفاهی در کلاس ارائه می‌شود. البته امتحانات رسمی شفاهی را می‌توان به گونه‌ای ترتیب داد که از نتایج آنها برای نمره‌گذاری دانش‌آموزان استفاده شود. اما از آنجا که این نوع امتحانات دارای همان مشکلات امتحانات تشریحی هستند، و علاوه بر آن بسیار وقت‌گیر و غیردقیق‌تر از امتحانات تشریحی کتبی‌اند، در شرایط معمولی کاربرد آنها به عنوان وسیله‌ای برای ارزشیابی و نمره‌گذاری توصیه نمی‌شود. بهتر است پرسشهای شفاهی را در همان محدودهٔ کلاس درس و به عنوان وسیله‌ای برای وارسی نحوهٔ پیشرفت یادگیرندگان و تدبیری برای بازخورد دادن به آنان به کار برد.

مهرنز و لهمان (۱۹۸۴) کاربرد پرسشهای شفاهی را برای مقاصد تشخیصی پیشنهاد می‌کنند. از آنجا که در گفت‌وشنود رو در رو امکان هدایت گفتگو به مسیرهای دلبخواهی میسر است، معلم می‌تواند از این طریق دانش‌آموزان را به بیان مشکلات خود و یافتن اشکالات یادگیری هدایت نماید. "یک پرسشگر ماهر می‌تواند پاسخهایی را در دانش‌آموزان برانگیزاند که نشان‌دهندهٔ فرایندهای فکری او باشند" (ص ۱۲۱).

1. oral question

از امتیازهای دیگر پرسشهای شفاهی این است که دانش‌آموز می‌تواند از معلم بخواهد تا نکات مبهم سؤال را برای او کاملاً روشن کند. قبلاً گفتیم که یکی از اشکالات بعضی از آزمونهای تشریحی کتبی مبهم بودن متن سؤالهاست. بسیاری از اوقات دانش‌آموزان نه به خاطر بی‌اطلاعی بلکه به سبب ابهام در صورت سؤالها به آنها جواب غلط می‌دهند. در پرسشهای شفاهی نیز ابهام ممکن است وجود داشته باشد، اما چون پرسشگر و پاسخ‌دهنده رودرروی هم قرار دارند امکان برطرف شدن ابهام موجود است.

یکی دیگر از موارد استفادهٔ پرسش شفاهی یا امتحان شفاهی کاربرد آن با دانش‌آموزان معلول جسمی است. برخی از دانش‌آموزان به خاطر معلولیت دست یا چشم قادر به دادن جوابهای کتبی نیستند، لذا در چنین شرایطی بهترین روش آزمون شفاهی است.

استفاده از پرسشهای شفاهی در موقعیتهای خاص، مانند انجام مصاحبه با داوطلبان شغل معلمی، یا داوطلبان ورود به دوره‌های دکتری، یا علاقه‌مندان به استخدام در هیأت‌های علمی دانشگاهها، می‌تواند در شناخت توانایی افراد، به ویژه نگرشها و واکنشهای رفتاری آنان، کمک کند. به عنوان نمونه، در جلسهٔ پرسش و پاسخ شفاهی، اگر داوطلب در پاسخ دادن به سؤالها تردید کند، برای یافتن کلمه‌های مناسب مِن و مِن کند، یا علائم تنش از خود نشان دهد بر سنجش شایستگی یا اعتماد به نفس او تأثیر منفی به جای می‌گذارد. داوطلب شغل معلمی در مصاحبه و گفتگوی شفاهی با هیأت داوران باید بتواند هنگام صحبت کردن از دانش خود استفاده نماید، درست و معقول صحبت کند، و بر اعصاب خود مسلط باشد.

به طور کلی، می‌توان گفت امتحان شفاهی بخش مکمل امتحان کتبی است. زمانی که، علاوه بر دانش و مهارتهایی که آزمون‌شونده در برگهٔ کتبی امتحانی خود نشان می‌دهد، سنجش مهارتهای گفتاری و تواناییهای ایجاد واکنش مناسب در جمع و نیز کنترل هیجان ضروری است، می‌توان از پرسش شفاهی به عنوان یک وسیلهٔ مؤثر ارزشیابی سود جست.

آزمون بیرون بردنی

نوع دیگر آزمون تشریحی **آزمون بیرون بردنی**[1] نام دارد. در این امتحان یا آزمون، دانش‌آموز یا دانشجو در ساعت معینی (مثلاً ۸ صبح) سؤالها را دریافت می‌کند و با خود به کتابخانه، منزل، اتاق کامپیوتر، یا هرکجا که بخواهد می‌برد. طبعاً می‌تواند از هر منبعی کمک بگیرد و با هرکس که بخواهد مشورت کند. در ساعت معینی که از قبل اعلام شده، مثلاً ۴ بعدازظهر، جواب سؤالها را تحویل می‌دهد. ریس، براون، و اسمیت (۲۰۰۵) می‌گویند این نوع آزمونها از لحاظ روایی بسیار خوب هستند، زیرا بر آنچه در زندگی واقعی اتفاق می‌افتد منطبق‌اند. "البته مشکل اعتماد

1. take-away exam

کردن دربارهٔ آنها وجود دارد ــ اما این هم با آنچه در زندگی واقعی رخ می‌دهد منطبق است» (ص ۵۴). البته زمان امتحان بیرون بردنی را می‌توان به بیشتر از یک روز، مثلاً تعطیلات آخر هفته، گسترش داد.

خلاصهٔ فصل

آزمون تشریحی به آزمونی گفته می‌شود که در آن صورت سؤالهای آزمون از سوی معلم یا ممتحن ساخته و در اختیار آزمون‌شوندگان یا دانش‌آموزان گذاشته می‌شود و آزمون‌شوندگان جواب سؤالها را تهیه و در برگهٔ امتحانی می‌نویسند. آزمونهای تشریحی به دو دستهٔ معروف گسترده پاسخ و محدودپاسخ تقسیم می‌شوند. آزمون گسترده‌پاسخ شامل سؤالهایی است که چه از لحاظ نوع پاسخ و چه از لحاظ مقدار پاسخ عملاً هیچ‌گونه محدودیتی برای پاسخدهی آزمون‌شوندگان قائل نمی‌شود. آزمون محدودپاسخ آزمونی است که سؤالهای آن هم از نظر نوع پاسخ و هم از لحاظ مقدار پاسخ محدود می‌شود.

صاحب‌نظران سنجش و اندازه‌گیری دربارهٔ چگونگی تهیهٔ سؤالهای تشریحی و نحوهٔ تصحیح پاسخهای آنها قواعدی را پیشنهاد داده‌اند که ما در این فصل آنها را توضیح داده‌ایم. از نکات مهمی که در رابطه با این آزمونها گفته شده این است که استفاده از آنها را به سطوح بالای هدفهای آموزشی چون تحلیل، ارزشیابی، و آفریدن محدود کنید و آنها را برای سنجش معلومات مورد استفاده قرار ندهید.

برای آزمونهای تشریحی امتیازها و محدودیتهایی را مطرح کرده‌اند. از جمله امتیازها به قرار زیرند:

۱. تهیهٔ آزمونهای تشریحی از تهیهٔ آزمونهای عینی آسان‌تر است.
۲. این آزمونها بهترین وسیلهٔ سنجش توانایی پروراندن و سازمان دادن مطالب‌اند.
۳. توانایی تولید پاسخ را می‌سنجند نه گزینش پاسخ را.
۴. موقعیتهای واقعی‌تری را به آزمون شوندگان ارائه می‌دهند.
۵. بر روش یادگیری دانش‌آموزان و دانشجویان تأثیر مثبت به جای می‌گذارند.

در مقابل امتیازهای بالا، برای آزمونهای تشریحی محدودیتهایی را نیز ذکر کرده‌اند. خلاصهٔ آنها به قرار زیر است:

۱. یک آزمون تشریحی نمونهٔ کوچکی از محتوای درس و هدفهای آموزشی را اندازه می‌گیرد.
۲. تصحیح پاسخهای تشریحی با دقت کافی امکان‌پذیر نیست.
۳. تصحیح پاسخهای سؤالهای تشریحی بسیار وقت‌گیر است.

موارد استفادهٔ آزمونهای تشریحی فراوان‌اند و می توان از آنها برای سنجش هدفهای سطوح مختلف استفاده کرد. اما صاحب‌نظران استفاده از سؤالهای تشریحی گسترده‌پاسخ را برای تحلیل، ارزشیابی، و آفرینندگی، و استفاده از سؤالهای تشریحی محدودپاسخ را برای فهمیدن، کاربستن، و تحلیل سفارش داده‌اند.

آزمون شفاهی یا پرسش شفاهی نوعی آزمون تشریحی است که برای مقاصد تشخیصی مشکلات یادگیری دانش‌آموزان و دادن بازخورد به معلم و دانش‌آموزان مفید است. از این نوع آزمون در مصاحبه‌های رودررو، برای سنجش توانایی صحبت در حضور جمع و واکنشهای هیجانی و اجتماعی داوطلبان مشاغلی مانند معلمی یا داوطلبان ورود به دوره‌های دکتری و تخصصی، می‌توان استفاده کرد. البته در ارتباط با دانش‌آموزان و دانشجویان معلول جسمی، مانند نابینایان، می‌توان از این آزمونها برای نمره‌دادن نیز سود برد.

آزمون بیرون‌بردنی نیز نوع دیگری از آزمون تشریحی است که در آن آزمون‌شونده سؤالهای امتحانی را از معلم دریافت می‌کند و آن را به خارج از کلاس می‌برد و با استفاده از منابع موجود در کتابخانه یا هر منبع دیگری سرفرصت به آنها جواب می‌دهد. امتیاز این سؤالها بالا بودن سطح روایی آنهاست، اما مشکل عمدهٔ این روش اعتمادپذیری جوابهای داده شده به سؤالهاست.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. در تصحیح پاسخ سؤالهای تشریحی، بهتر است ابتدا جواب همهٔ سؤالهای برگهٔ اول را تصحیح کنیم، بعد سؤالهای برگهٔ دوم، و الی آخر.

الف ـ صحیح

ب ـ غلط

۲. سؤال زیر چه نوع هدف آموزشی را می‌سنجد؟

"در یک پاراگراف، به طور خلاصه توضیح دهید که چرا فشارسنج بهترین ابزار پیش‌بینی وضع هواست."

(گرانلاند، ۱۹۷۱)

الف ـ یادآوری

ب ـ فهم

پ ـ تحلیل

ت ـ ارزشیابی

ث ـ آفرینندگی

۳. سؤال بالا (شمارهٔ ۲) چه نوع سؤالی است؟

الف ـ تشریحی گسترده‌پاسخ

ب ـ تشریحی محدودپاسخ

۴. سؤال تشریحی زیر چه نوع هدف آموزشی را می‌سنجد؟

"به زمانی در زندگی خود فکر کنید که با مشکلی رو به رو بودید، چیزی بر سر راه شما قرار داشت و می‌بایست بر آن فائق می‌آمدید. دربارهٔ این مشکل داستانی بسازید و آن را برای کلاس بگویید."
(بلوم، ۱۹۵۹، ترجمهٔ سیف و علی‌آبادی، ۱۳۶۸)

الف ـ فهمیدن
ب ـ به کار بستن
پ ـ تحلیل کردن
ت ـ ارزشیابی کردن
ث ـ آفریدن

۵. سؤال بالا (شمارهٔ ۴) چه نوع سؤالی است؟

الف ـ تشریحی گسترده‌پاسخ
ب ـ تشریحی محدودپاسخ

۶. گفته شده است در آزمونهای تشریحی به آزمون‌شوندگان حق انتخاب تعدادی سؤال از میان مجموعه‌ای سؤال را ندهید. دلایل این کار را توضیح دهید، و شرایطی را که می‌توان این حق را به آزمون‌شوندگان داد مشخص کنید.

۷. روشهای مختلف نمره‌گذاری سؤالهای تشریحی را توضیح دهید و بگویید کدام روش در چه موقعیتی مناسب‌تر است؟

۸. قواعد تهیهٔ سؤالهای تشریحی را شرح دهید.

۹. قواعد تصحیح پاسخهای تشریحی را شرح دهید.

۱۰. سؤالهای تشریحی گسترده‌پاسخ و محدودپاسخ را از جنبه‌های مختلف با هم مقایسه کنید.

۱۱. امتیازها و محدودیتهای آزمونهای تشریحی را توضیح دهید.

۱۲. برای طرح سنجش خود که به عنوان تمرین فصل ۶ تهیه کرده‌اید یک آزمون تشریحی متناسب با هدفهایی از آن درس که با این نوع آزمونها بهتر سنجیده می‌شوند بسازید.

۱۳. آزمون شفاهی را توضیح دهید و موارد استفادهٔ آن را بنویسید.

۱۴. آزمون بیرون‌بردنی را تعریف کنید و خوبی و بدی آن را توضیح دهید.

۱۵. فرض کنید می‌خواهید یک مصاحبهٔ استخدامی برای داوطلبان معلمی برگزار کنید. برای سنجش نگرش معلمان نسبت به شغل معلمی چه سؤالهایی از داوطلبان خواهید پرسید؟ آنها را بنویسید.

۱۱

سنجش یادگیری به کمک آزمون کوته‌پاسخ

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمون کوته‌پاسخ را تعریف کند و تفاوت میان آزمون کوته‌پاسخ، آزمون انشایی، و آزمون عینی را توضیح دهد.

۲. انواع آزمونهای کوته‌پاسخ را با ذکر مثالی برای هر نوع آن توضیح دهد.

۳. امتیازها و محدودیتهای آزمونهای کوته‌پاسخ را توضیح دهد.

۴. قواعد تهیهٔ سؤالهای کوته‌پاسخ را با ذکر مثال برای هر قاعده توضیح دهد.

۵. با توجه به ویژگیهای مطلوب آزمونهای کوته‌پاسخ، سؤالهای کوته‌پاسخ معیوبی راکه به شما داده می‌شود اصلاح کند.

۶. سؤالهای عددی را تعریف کند، تفاوت آنها را با سؤالهای آزمونهای کوته‌پاسخ توضیح دهد، و قواعد تهیه و کاربرد آنها را با ذکر مثال شرح دهد.

۷. موارد استفادهٔ آزمونهای کوته‌پاسخ را توضیح دهد.

۸. برای هدفهای آموزشی قابل سنجش با سؤالهای کوته‌پاسخ سؤالهای امتحانی کوته‌پاسخ بنویسد.

حد وسط بین آزمون تشریحی و آزمون عینی آزمون کوته‌پاسخ است. شباهت این نوع آزمون با آزمون تشریحی آن است که در هر دو آنها صورت سؤال به وسیلهٔ طراح سؤال تهیه می‌شود و جواب آنها را آزمون‌شونده تهیه می‌کند، یعنی هر دو از نوع تولید پاسخ هستند. با وجود این، جواب سؤالهای آزمونهای کوته‌پاسخ آن قدر مختصر است که این آزمونها را به کلی از آزمونهای تشریحی مجزا می‌سازد و به آزمونهای عینی نزدیک می‌کند. افزون بر این، چون تصحیح پاسخ سؤالهای این آزمونها با دقّت و عینیت بیشتری انجام می‌پذیرد، گروهی از متخصصان اندازه‌گیری و ارزشیابی آنها را در ردیف آزمونهای عینی قرار داده‌اند.

تعریف و انواع آزمونهای کوته‌پاسخ

منظور از آزمون کوته‌پاسخ[1] مجموعه‌ای سؤال مختصر است که غالباً برای سنجش هدفهای آموزشی سطح پایین طرح می‌شوند. سؤالهای آزمون کوته‌پاسخ از آزمون‌شونده می‌خواهند تا کلمه، عبارت، جمله، عدد، یا علامتی را در پاسخ به یک سؤال یا تکمیل یک جمله بنویسند.

آزمونهای کوته‌پاسخ را به سه دستهٔ **پرسشی**، **کامل‌کردنی**، و **تشخیصی یا تداعی** تقسیم می‌کنند. به نمونه سؤالهایی برای هر یک از این سه دسته آزمون توجه کنید:

۱ **نوع پرسشی**

- *مرکز استان مازندران کجاست؟ (ساری)*
- *یک میلی‌متر چند میکرون است؟ (۱۰۰۰)*

۲ **نوع کامل‌کردنی**

- *بزرگترین دریاچهٔ دنیا دریاچهٔ ـــــ است. (خزر)*
- *(۶) ـــــ = ۲: (۸ + ۴)*

۳ **نوع تشخیصی یا تداعی**

- *بعد از نام هر استان، نام مرکز آن استان را بنویسید.*

خوزستان ـــــ (اهواز)

خراسان ـــــ (مشهد)

مرکزی ـــــ (اراک)

لرستان ـــــ (خرم‌آباد)

- *بعد از عناصر شیمیایی زیر، علامت شیمیایی آنها را بنویسید.*

1. short-answer test

باریوم ــ (Ba)

کلسیم ــ (Ca)

پتاسیم ــ (K)

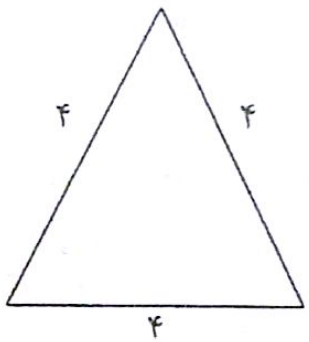
روی ــ (Zn)

چنان‌که می‌بینید، در آزمونهای کوته‌پاسخ نوع پرسشی، سؤال به صورت یک جملهٔ استفهامی طرح می‌شود و آزمون‌شونده پاسخ کوتاهی به آن می‌دهد. در آزمونهای کوته‌پاسخ نوع کامل‌کردنی، سؤال به صورت یک جملهٔ ناقص نوشته می‌شود که در آن یک جای خالی وجود دارد و آزمون‌شونده با پرکردن جای خالی به سؤال جواب می‌دهد. و در آزمونهای کوته‌پاسخ تشخیصی یا تداعی، سؤال به صورت یک موضوع یا مطلب ارائه می‌شود و آزمون‌شونده باید جواب آن را تداعی کند.

آزمونهای کوته‌پاسخ تشکیل یافته از مسائل عددی ساده

ایبل (۱۹۷۹) آزمونهایی را که از سؤالهای مربوط به محاسبات عددی تشکیل می‌شوند نوع ویژه‌ای از آزمونهای کوته‌پاسخ می‌داند و به آنها نام آزمونهای **مسائل عددی**[1] داده است. او می‌گوید "در حالی که مسائل عددی را می‌توان به صورت سؤالهای چندگزینه‌ای نوشت، اما اغلب به صورت سؤالهای کوته‌پاسخ نوشته می‌شوند" (ص ۱۷۱). این‌گونه سؤالها در ریاضیات، علوم، حسابداری، و زمینه‌های دیگری که با کاربرد اعداد سروکار دارند قابل استفاده‌اند. امتیاز مهم آزمونهای مسائل عددی این است که می‌توان آنها را در موقعیتهای تازه به کار بست و توانایی فهم و کاربستن آزمون‌شوندگان را مورد سنجش قرار داد. به نمونه‌هایی از این نوع سؤالها توجه کنید:

- در تساوی $۳x + ۶ = ۳۰$، مقدار x چقدر است؟
- یک کودک ۷ سال و ۷ ماهه، براساس آزمون هوش استنفورد ـ بینه دارای سن هوشی ۸ است. هوشبهر او را حساب کنید.
- در مثلث زیر اندازهٔ ضلعها معلوم است، اندازهٔ هر یک از زاویه‌های داخلی را برحسب درجه حساب کنید.



1. numerical problems

قواعد تهیهٔ سؤالهای کوته‌پاسخ

۱. هر سؤال را با توجه به یک موضوع مهم بنویسید. از قرار دادن موضوعها و مطالب بی‌اهمیت و خیلی جزئی در سؤالها بپرهیزید. ثرندایک، کانینگهام، ثرندایک، و هیگن (۱۹۹۱) دو سؤال زیر را به عنوان نمونه سؤالهایی که در آنها موضوعهای بی‌اهمیت مطرح شده مثال زده‌اند:

- در سال ۱۹۵۰، در کشور آمریکا چند مورد بیماری آبله مشاهده شده است؟
- در سال ۱۹۵۰، تنها یک نفر از ــــ نفر مردم آمریکا از دندانهای خود مواظبت کامل به عمل آورده‌اند. (ص ۲۵۴)

۲. صورت سؤال را کاملاً روشن بنویسید به گونه‌ای که به پاسخ مشخص و معینی نیاز داشته باشد. از مبهم و چندپهلو نوشتن صورت سؤالها بپرهیزید. ایبل (۱۹۷۹) می‌گوید یک مشکل معمولی سؤالهای کوته‌پاسخ این است که سؤالی که به نظر طراح آن پاسخ مشخصی را می‌طلبد از سوی آزمون‌شوندگان پاسخهای درست دیگری دریافت می‌کند. برای مثال، سؤال "ذغال سنگ چیست؟" که پاسخ آن از نظر طراح سؤال "نوعی سوخت" است، ممکن است از سوی آزمون‌شوندگانِ مختلف جوابهای زیر را دریافت کند که همه درست هستند:

یک مادهٔ قابل اشتعال

کربن ناخالص

بقایای سفت شدهٔ مواد گیاهی

برای رفع ابهام این سؤال و روشن بیان کردن موضوع مورد نظر و نیز برای جلوگیری از دشواریهای بعدی تصحیح پاسخها، سؤال را باید طوری نوشت که پاسخ واحد و مشخصی را از همهٔ آزمون‌شوندگان بخواهد. برای مثال به نمونه سؤالهای زیر توجه کنید.

- ذغال سنگ غالباً به چه منظوری به کار می‌رود؟
- ذغال سنگ از چه موادی درست شده است؟
- ذغال سنگ از چه عنصر شیمیایی درست می‌شود؟

به مثالهای دیگری در زیر توجه کنید:

سؤالهای ضعیف:

برگ تنباکو حاوی چه چیزی است؟ (نیکوتین)

برگ تنباکو حاوی (نیکوتین) است.

سؤالهای بهتر:

نام مادهٔ سمی برگ تنباکو چیست؟ (نیکوتین)

نام مادهٔ سمی برگ تنباکو (نیکوتین) است.

(ثرندایک و همکاران، ۱۹۹۱، ص ۲۵۵)

بنابراین، بکوشید تا صورت سؤال را کاملاً آشکار و خالی از ابهام بنویسید. پوفام (۲۰۰۲) توصیه می‌کند که همواره سعی کنید خودتان را، به طور ذهنی، به جای دانش‌آموزانتان قرار دهید و آن‌گاه ببینید که آنها سؤال شما را چگونه تفسیر خواهند کرد.

۳. **صورت سؤال را عیناً از روی مطالب کتاب ننویسید.** نقلِ مطالب کتاب موجب تشویق دانش‌آموزان به حفظ کردن مطالب و مانع درک و فهم آنان می‌شود. پیشنهاد نیتکو (۲۰۰۱) برای نوشتن صورت سؤالهای کوته‌پاسخ این است که "ابتدا به پاسخ موردنظرتان فکر کنید و بعد سؤالی طرح نمایید که آن پاسخ تنها پاسخ درست آن باشد" (ص ۱۳۴).

۴. **در سؤالهایی که پاسخهای آنها اعداد هستند، واحد مقیاس و میزان دقتی را که در محاسبات باید رعایت شوند مشخص کنید.** سؤالها باید واحدهای عددی و میزان دقت لازم را مشخص کنند. این کار تکلیف آزمون‌شوندگان را مشخص می‌کند و موجب می‌شود تا در وقت کسانی که در محاسبات خود بالاتر از دقت مورد انتظار معلم پیش می‌روند، صرفه‌جویی شود.

سؤال ضعیف:

مقدار عدد π چقدر است؟ (۳/۱، ۳/۱۴، ۳/۱۴۱)

سؤال بهتر:

مقدار عدد π تا سه رقم اعشار چقدر است؟ (۳/۱۴۱)

برای رعایت قاعدهٔ بالا، راهنماییهای لازم را فراهم آورید. به نمونهٔ زیر که نوعی راهنمایی برای یک آزمون است توجه کنید:

راهنمایی:

این آزمون شامل ۶۰ سؤال است که بعضی از آنها به انجام دادن محاسبات نیاز دارند. برای سؤالهایی که احتیاج به محاسبه دارند باید: (۱) جوابهای خود را تا دو رقم اعشاری نشان دهید، و (۲) آنها را به واحدهای مناسب چون پوند، گرم، یا ولت بنویسید.

(مهرنز و لهمان، ۱۹۸۴، ص ۱۳۴)

۵. **در سؤالهای کوته پاسخ کامل‌کردنی، تنها کلمه‌ها و عبارتهای مهم را حذف کنید.** چنان که دیدیم، سؤال کامل کردنی نوع خاصی از سؤال کوته‌پاسخ است. بنا به گفتهٔ کوهن و سوردلیک (۲۰۰۵)،

"نام دیگر سؤال کامل کردنی سؤال کوته‌پاسخ است" (ص ۲۰۴). در نوشتن این نوع سؤالها، از آزمون‌شوندگان بخواهید تا به جنبه‌های مهم آنچه خوانده‌اند پاسخ دهند نه به جنبه‌های بی‌اهمیت و خیلی جزئی. در سؤالهای کامل‌کردنی، جای خالی باید مطالب مهم را شامل شود، یعنی مطالبی که دانستن آنها از سوی آزمون‌شوندگان مهم است. جای خالی نباید به اجزای فرعی جمله مربوط باشد.

سؤال ضعیف:

کبد گلوکز اضافی را به صورت گلیکوژن در خود ـــــ می‌کند. (ذخیره)

سؤال بهتر:

در سوخت و ساز طبیعی بدن، گلوکز اضافی خون به صورت ـــــ ذخیره می‌شود. (گلیکوژن)

(ثرندایک و همکاران، ۱۹۹۱، ص ۲۵۶)

۶. **در سؤالهای کامل‌کردنی که برای ارزشیابی پیشرفت تحصیلی به کار می‌روند تعداد زیادی جای خالی منظور نکنید.** وقتی که در سؤالی بیش از یک جای خالی قرار می‌گیرد، سؤال مبهم و ناخوانا می‌شود. این‌گونه سؤالها ممکن است برای آزمونهای هوش و خلاقیت مناسب باشند، اما برای سنجش یادگیری کارساز نیستند. ثرندایک و هیگن (۱۹۷۹) می‌گویند "این نوع سؤالها ممکن است برای اندازه‌گیری آفرینندگی یا اصالت اندیشهٔ آزمون‌شوندگان در مورد کلمات مفید باشند" (ص ۲۶). در سؤالهای کامل‌کردنی گاه به عوض یک جای خالی می‌توان از دو جای خالی استفاده کرد، آن هم در شرایطی که صورت سؤال طوری طرح شود که هیچ‌گونه ابهامی در آن نباشد. به نمونه‌های زیر توجه کنید.

سؤال ضعیف:

ـــــ و ـــــ روشهای تصحیح آزمونهای تشریحی هستند. (تحلیلی) و (کلی)

سؤال بهتر:

دو روش معمول تصحیح آزمونهای تشریحی روش ـــــ و روش ـــــ هستند. (تحلیلی) و (کلی)

۷. **جای خالی سؤالهای کامل‌کردنی را تا آنجا که ممکن است در قسمت پایانی سؤال قرار دهید.** اگر جای خالی در قسمت پایانی سؤال باشد، آزمون‌شونده می‌تواند، پیش از رسیدن به جای خالی، از موضوع مورد سؤال اطلاع کامل حاصل کند. در حالی که اگر جای خالی در اوایل یا اواسط صورت سؤال قرار داشته باشد، تا زمانی که آزمون‌شونده تمام صورت سؤال را می‌خواند ممکن است مطلب مهم و اساسی سؤال را فراموش کند. این مورد به ویژه دربارهٔ مطالب پیچیده و نیز در رابطه با کودکان خردسال که دامنهٔ توجه و حافظهٔ اندکی دارند صادق است. به نمونه‌های زیر توجه کنید.

سؤال ضعیف:

در آزمایشهای شرطی کردن پاولفی ـــــ به پاسخی گفته می‌شود که حیوان از آن طریق به صدای زنگ واکنش می‌کند. (پاسخ شرطی)

سؤال بهتر:

در آزمایشهای شرطی کردن پاولفی واکنش حیوان به صدای زنگ را ـــــ می‌گویند. (پاسخ شرطی)

۸. **جای خالی سؤالهای کامل‌کردنی را همواره یک اندازه تعیین کنید.** اگر جای خالی هر سؤال کامل‌کردنی متناسب با طول پاسخ مورد انتظار باشد، دانش‌آموزان ممکن است به این مطلب پی ببرند و این امر موجب کمک به آنها در حدس زدن جواب درست باشد. بنابراین، بهتر است جای خالی در همهٔ سؤالها یک اندازه تعیین شود.

۹. **در سؤالهای کامل‌کردنی، از کاربرد اشاره‌های دستوری و موارد دیگری که جواب سؤال را مشخص می‌کنند خودداری کنید.** اشاره‌هایی از قبیل زمان افعال، مفرد و جمع بودن آنها، یا در زبان انگلیسی حرفهای a و an، و از این قبیل را به کار نبرید. ایبل (۱۹۷۹، ص ۱۶۸) سؤال زیر را به عنوان مثالی که در آن اشاره‌ای در صورت سؤال منجر به حدس زدن پاسخ سؤال می‌شود ذکر کرده است.

- *موتور کشتی بخار با نیروی ـــــ کار می‌کند. (بخار)*

۱۰. **تا حد امکان، به جای سؤالهای کامل‌کردنی، از سؤالهای پرسشی استفاده کنید.** به طور کلی، سؤالهای کوته‌پاسخ نوع پرسشی از سؤالهای نوع کامل‌کردنی بهترند، زیرا: (۱) تهیهٔ سؤالهای نوع پرسشی آسان‌تر است و این سؤالها دارای ابهام کم‌تری هستند، و (۲) پاسخ دادن به سؤالهای پرسشی برای دانش‌آموزان آسان‌تر است، زیرا آنان غالباً به پرسشهای معلمان پاسخ می‌دهند نه به جمله‌های ناتمام. به سخن دیگر، آشنایی دانش‌آموزان با جمله‌های پرسشی خیلی بیشتر از جمله‌های ناتمام است. به ویژه وقتی که تغییردادن جای خالی سؤالهای کامل کردنی از اوایل صورت سؤال به قسمت آخر آن غیرعملی است، بهتر است صورت سؤال را از نوع کامل‌کردنی به نوع پرسشی تغییر دهید. انجام دادن این کار به راحتی صورت می‌گیرد، زیرا سؤال کامل‌کردنی در واقع از نظر ماهیت همان سؤال پرسشیِ تغییر شکل یافته است. به نمونهٔ زیر توجه کنید.

سؤال کامل کردنی:

ـــــ پایتخت کشور انگلستان است. (لندن)

سؤال پرسشی:

نام پایتخت کشور انگلستان چیست؟ (لندن)

افزون بر آنچه گفته شد، معنی سؤالهایی که به صورت پرسشی تهیه می‌شوند روشن‌تر از زمانی است که به صورت کامل‌کردنی هستند. به سؤالهای زیر توجه کنید:

سؤال کامل کردنی:

نویسندهٔ داستان آلیس در سرزمین عجایب ـــ است. (لوئیس کارول)

سؤال پرسشی:

نام نویسندهٔ داستان آلیس در سرزمین عجایب چیست؟ (لوئیس کارول)

نیتکو (۲۰۰۱) می‌گوید به سؤال کامل‌کردنی بالا چندین جواب می‌توان داد، از جمله: "داستان‌سرا"، "ریاضی‌دان"، "انگلیسی" و مانند اینها، اما جواب سؤال پرسشی فوق تنها همان یک جواب داده شده است.

قواعد تهیهٔ سؤالهای کوته‌پاسخ ویژهٔ مسائل عددی

علاوه بر قواعد کلی تهیهٔ سؤالهای آزمونهای کوته‌پاسخ که قبلاً معرفی شدند، برای تهیهٔ سؤالهای مسائل عددی نیز قواعد زیر پیشنهاد شده‌اند (ایبل، ۱۹۷۹).

۱. **تا حد امکان از اعداد ساده استفاده کنید.** هدف سؤالهای مسائل عددی آزمون کردن فهم دانش‌آموز از جریان یا فرایند محاسبات است نه دقت در محاسبات عددی. اگر می‌خواهید دقت محاسبات دانش‌آموز را سنجش کنید سؤالهای مجزایی به آن اختصاص دهید و در این سؤالها اعداد پیچیده را آن طور که می‌خواهید مورد استفاده قرار دهید.

۲. **تا حد امکان مسائلی را طرح کنید که دارای جواب صحیح هستند.** این کار نیاز به محاسبات پیچیدهٔ اعداد اعشاری و عدم اطمینان آزمون‌شونده را از اینکه محاسبات خود را تا چند عدد اعشاری ادامه دهد از بین می‌برد.

۳. **میزان دقت مورد انتظار را تعیین کنید.** اگر دانش‌آموزان دربارهٔ آنچه از آنها خواسته شده است نامطمئن باشند و اگر در این‌باره حدس غلط بزنند، اندازه‌گیری شما از میزان توانایی آنان غیردقیق خواهد بود.

۴. **اگر لازم است آزمون‌شوندگان در پاسخ خود واحد اندازه‌گیری را نشان دهند، این مطلب را به اطلاع آنان برسانید.** اگر از ذکر این نکته خودداری کنید، ممکن است سبب شود که بعضی از آزمون‌شوندگان به آن توجه نکنند و بی‌جهت از نمرهٔ آنان کاسته شود.

۵. **در صورت امکان، مسائل پیچیده و چندمرحله‌ای را به تعدادی مسئلهٔ یک مرحله‌ای ساده تقسیم**

کنید. ایبل (۱۹۷۹، ص ۱۷۱) می‌گوید "این تصور اشتباهی است که هرچه یک مسئله پیچیده‌تر باشد توانایی آزمون‌شونده را بهتر می‌سنجد؛ معمولاً عکس این درست است." هر مسئلهٔ پیچیده‌ای شامل چند مرحله و تعدادی محاسبه است. هر یک از اینها را می‌توان به صورت یک مسئلهٔ مجزا درآورد. حل کردن کل مسئله به حل کردن این اجزاء وابسته است.

۶. **تا آنجا که امکان دارد مسائل عددی را به زبان ساده و به طور مختصر بیان کنید.** این کار مستلزم دادن اطلاعات کافی در صورت مسئله و توضیحات روشن و دقیق است. همچنین از توضیحات اضافی و تکرار کلمات و مطالب خودداری کنید.

امتیازها و محدودیتهای آزمونهای کوته‌پاسخ

یکی از امتیازهای آزمونهای کوته‌پاسخ سهولت تهیهٔ سؤالها و تصحیح پاسخهای آنهاست. مهرنز و لهمان (۱۹۸۴) می‌گویند زمانی که دانش تعاریف یا اصطلاحات فنی دانش‌آموزان را می‌سنجید بهتر است از آنان بخواهید تا جواب سؤال را خود تهیه کنند نه آنکه جوابها را از میان تعدادی جواب پیشنهاد شده انتخاب نمایند. آزمونهای کوته‌پاسخ از این امتیاز برخوردارند. علاوه بر این، آزمونهای کوته‌پاسخ، در مقایسه با آزمونهای عینی، نظیر آزمونهای صحیح ـ غلط یا چندگزینه‌ای، دارای مزایای دیگری به شرح زیرند:

۱. تهیه و اجرای آنها آسان است.
۲. تقلب را کاهش می‌دهند.
۳. اطلاعات تشخیصی بیشتری را در اختیار معلمان می‌گذارند.

امتیاز دیگر آزمونهای کوته‌پاسخ این است که در آنها حدس کورکورانه که از معایب عمدهٔ آزمونهای عینی است وجود ندارد. چنان که پیش‌تر گفتیم، دانش‌آموزی که به یک آزمون صحیح ـ غلط جواب می‌دهد پنجاه درصد شانس دارد که حدس بزند، و در آزمونهای چهارگزینه‌ای این احتمال ۲۵ درصد است، اما در آزمونهای کوته‌پاسخ هیچ‌گونه احتمال حدس زدن وجود ندارد.

در مقابل امتیازهای بالا، آزمونهای کوته‌پاسخ دارای مشکلاتی هستند که کاربرد آنها را محدود می‌کند. از آنجا که پاسخ سؤالهای این نوع آزمونها بسیار مختصرند، نمی‌توان از آنها برای سنجش هدفهای سطح بالای یادگیری استفاده کرد، و کاربرد آنها به اندازه‌گیری هدفهای سطح دانش و حداکثر فهمیدن و کاربستن (آن هم بیشتر در رابطه با مسائل عددی) محدود می‌شود. مشکل دیگر این است که استفادهٔ زیاد از این‌گونه آزمونها سبب تشویق یادگیرندگان به حفظ کردن اطلاعات جزئی و کم اهمیت درس خواهد شد، زیرا در این نوع آزمونها بیشتر این‌گونه اطلاعات سنجیده می‌شوند.

همچنین تصحیح و نمره‌گذاری پاسخهای این آزمونها به دقت و سرعت آزمونهای عینی

امکان‌پذیر نیست. آزمون‌شوندگان در پاسخ دادن به سؤالهای آزمونهای کوته پاسخ ممکن است یکنواخت عمل نکنند و در جواب سؤالها پاسخهای مختلفی بدهند. در نتیجه مصحح باید تصمیم بگیرد که جوابها درست‌اند، غلط‌اند، یا نیمه درست‌اند. این امر ممکن است به عدم دقت در تصحیح برگه‌های امتحانی و کاستن از میزان روایی و پایایی آزمون منجر شود. ساکس (۱۹۹۷) برای روشن کردن مطلب سؤال زیر را مثال زده است.

- نخستین رئیس‌جمهور ایالات متحد آمریکا چه کسی بود؟

ظاهراً معلم انتظار دارد که دانش‌آموز به سؤال بالا جواب جورج واشنگتن را بدهد. اما جوابهای دیگری مانند یک ژنرال که در ویرجینیا متولد شد جوابهای بالقوه درستی هستند. بنابراین، "ساختن سؤالهای کوته‌پاسخ یا کامل‌کردنی که تنها یک پاسخ درست و غیرمبهم داشته باشند دشوار است" (ساکس، ۱۹۹۷، ص ۱۱۸).

موارد استفادهٔ آزمونهای کوته‌پاسخ

آزمونهای کوته‌پاسخ برای اندازه‌گیری هدفهای شناختی سطوح پایین، به ویژه یادآوری دانش، مفیدند. مهرنز و لهمان (۱۹۷۳) بهترین مورد استفاده را برای آزمونهای کوته‌پاسخ سنجشِ دانشِ تعاریف و اصطلاحات فنی دانسته‌اند. بنا به گفتهٔ این متخصصان، "وقتی که ما علاقه‌مند به سنجش دانش تعاریف یا اصطلاحات فنی دانش‌آموزان هستیم، اگر از آنان بخواهیم تا جواب خود را تولید کنند بیشتر از کارمان مطمئن می‌شویم تا زمانی که از آنان می‌خواهیم تا جواب خود را انتخاب نمایند" (ص ۲۵۶).

البته در برخی مواقع از سؤالهای کوته‌پاسخ برای اندازه‌گیری هدفهای آموزشی سطح بالاتر نیز می‌توان استفاده کرد. گرانلاند و لین (۱۹۹۰) موارد زیر را ذکر کرده‌اند:

۱. توانایی تفسیر سادهٔ اطلاعات و کاربرد قواعد، مثلاً شمارش تعداد هجاهای یک کلمه؛ نشان دادن دانش ارزش مکانی اعداد؛ تشخیص دادن اندامهای یک موجود زنده از روی عکس؛ کاربرد تعریف یک مثلث متساوی‌الساقین.
۲. توانایی حل کردن مسائل عددی در ریاضیات و علوم.
۳. توانایی دستکاری نمادهای ریاضی و حل کردن معادله‌های ریاضی و شیمیایی.

ایبل (۱۹۷۹) معتقد است که سؤالهای آزمونهای کوته‌پاسخ به طور عمده با کلمات و اعداد سروکار دارند. این سؤالها نامهای اشخاص، مکانها، اشیاء، فرایندها، رنگها، و غیره را دربر می‌گیرند. سؤالهای کوته‌پاسخ همچنین اعداد مُعرف تاریخها، فاصله‌ها، هزینه‌ها، و جمعیتها را شامل می‌شوند. اگر آنها یک عبارت را شامل شوند، معمولاً عبارتی کوتاه است. "سؤالهایی که

پاسخهای طولانی‌تری را دربرمی‌گیرند، مانند 'سه دلیل برای... ذکر کنید' یا 'ویژگیهای .... را نام ببرید' بهتر است در طبقهٔ آزمونهای تشریحی محدود پاسخ طبقه‌بندی شوند" (ص ۱۶۶). بنابراین، از نظر ایبل همهٔ آزمونهای کوته‌پاسخ برای اندازه گیری اطلاعات واقعی به کار می‌روند.

با وجود این، اگر سؤالهای آزمونهای کوته‌پاسخ به گونه‌ای نوشته شوند که مسائل مختصر ریاضی و تبدیل و تفسیر مطالب را شامل شوند، می‌توان سطح آنها را از دانش به فهمیدن و کاربستن افزایش داد. برای مثال به سؤال کوته‌پاسخ زیر توجه کنید.

- *چه عددی است که دو برابر آن به اضافهٔ ۵ می‌شود ۳۹؟ (۱۷)*

برای حل کردن این مسئله، دانش‌آموز ابتدا باید عدد مورد نظر را مجهول فرض کند. بعد باید مسئله را از صورت کلامی به صورت معادله‌ای مانند $۲X + ۵ = ۳۹$ تبدیل نماید (تفسیر کردن). آنگاه با کاربرد فرمولی که برای حل کردن این‌گونه مسائل آموخته است به حل مسئله بپردازد (کاربستن).

ما در دنبالهٔ مطالب این فصل تعدادی سؤال کوته‌پاسخ واقعی را که برای سنجش هدفهای یادگیری مختلف تهیه شده‌اند می‌آوریم. خواندن دقیق آنها می‌تواند به شما در شناخت بهتر این‌گونه سؤالها کمک کند.

یادآوری دانش امور واقعی

۱. نمایندگان مجلس شورای اسلامی برای یک دورهٔ ـــــ انتخاب می‌شوند.

۲. مرکز استان همدان کجاست؟

یادآوری دانش مفهومی

۳. در طبقه‌بندی تجدیدنظرشدهٔ حوزهٔ شناختی بالاترین سطح دانش چه نام دارد؟

۴. در نظریهٔ ثرندایک، بنا به اصل گسترش اثر، تقویت یک پاسخ هم موجب نیرومندی آن پاسخ و هم نیرومندی پاسخهای مجاور آن می‌شود. در نظریهٔ شرطی‌سازی کلاسیک پاولفی این اصل چه نام دارد؟

یادآوری دانش روندی

۵. برای تعیین بار الکتریکی مثبت و منفی از چه وسیله‌ای استفاده می‌کنند؟

۶. وقتی که رتبه‌های دو درس دانش‌آموزان یک کلاس را در اختیار داریم، از چه روشی می‌توانیم همبستگی بین آنها را حساب کنیم.

فهمیدن دانش مفهومی

۷. اگر دمای گازی ثابت نگه داشته شود در حالی که فشار وارده بر آن افزایش می‌یابد، حجم آن چه تغییری می‌کند؟

۸. در عدد ۵۲۱ رقم ۵ چه ارزشی را نشان می‌دهد؟

کاربستن دانش روندی

۹. در معادلهٔ $X^2 + X - ۱ = ۲۰$، مقدار X چقدر است؟

۱۰. قیمت یک بسته پنیر یک کیلویی ۱۲۰۰ تومان است. قیمت بستهٔ ربع کیلویی همان پنیر ۴۰۰ تومان است. اگر به جای چهار بستهٔ ربع کیلویی یک بستهٔ یک کیلویی پنیر بخرید چند تومان به نفع شماست؟

خلاصهٔ فصل

آزمون کوته‌پاسخ در بین آزمونهای عینی و تشریحی قرار دارد. آزمون کوته‌پاسخ تعدادی سؤال مختصر را دربرمی‌گیرد که پاسخهای آنها به یک عدد، علامت، کلمه، عبارت، یا جمله محدود می‌شود.

آزمونهای کوته‌پاسخ را به سه دستهٔ پرسشی، کامل‌کردنی، و تشخیصی (تداعی) طبقه بندی می‌کنند. در آزمون کوته‌پاسخ نوع پرسشی سؤال به صورت یک جملهٔ پرسشی طرح می‌شود و آزمون‌شونده پاسخ کوتاهی به آن می‌دهد. در آزمون کوته‌پاسخ نوع کامل‌کردنی سؤال به صورت یک جملهٔ ناقص نوشته می‌شود که در آن یک جای خالی وجود دارد و آزمون‌شونده با پرکردن جای خالی به سؤال جواب می‌دهد. و در آزمون کوته‌پاسخ نوع تشخیصی یا تداعی صورت سؤال ارائه می‌شود و آزمون‌شونده باید جواب آن را تداعی کند. علاوه بر اینها، نوع دیگری از آزمونهای کوته‌پاسخ وجود دارد که به آزمونهای کوته‌پاسخ تشکیل یافته از مسائل عددی شهرت دارند. این‌گونه آزمون در ریاضیات، علوم، حسابداری، و زمینه‌های دیگری که با کاربرد اعداد سروکار دارند قابل استفاده است.

صاحب‌نظران سنجش و اندازه‌گیری برای تهیهٔ سؤالهای کوته‌پاسخ قواعدی را به دست داده‌اند. ما در این فصل این قاعده‌ها را با ذکر مثال توضیح داده‌ایم.

برای آزمونهای کوته‌پاسخ امتیازها و محدودیتهایی را قائل شده‌اند. مهم‌ترین امتیازهای این سؤالها به قرار زیرند:

۱. سهولت تهیه و اجرا
۲. عینی بودن تصحیح پاسخ سؤالها در مقایسه با آزمونهای تشریحی
۳. قابلیت استفاده با مسائل سادهٔ عددی

۴. کاهش دادن تقلب
۵. فراهم آوردن اطلاعات تشخیصی بیشتر از آزمونهای عینی

در مقابل امتیازهای بالا، برای آزمونهای محدودپاسخ محدودیتهایی را نیز ذکر کرده‌اند. در زیر به مهم‌ترین آنها توجه کنید:

۱. یادگیریهای سطح پایین را می‌سنجند
۲. استفادهٔ زیاد از آنها سبب تشویق دانش‌آموزان به یادگیری اطلاعات جزئی می‌شود
۳. در مقایسه با آزمونهای عینی، پاسخ سؤالها با سرعت کمتری قابل تصحیح است

در رابطه با موارد استفادهٔ آزمونهای عینی، باید گفت که این آزمونها عمدتاً برای سنجش هدفهای سطوح پایین یادگیری مفیدند. با این حال، می‌توان این گونه سؤالها را طوری تهیه کرد که هدفهای طبقه‌های فهمیدن و کاربستن را نیز سنجش کنند. به ویژه از آزمونهای تشکیل یافته از مسائل سادهٔ عددی به خوبی می‌توان در سطح کاربستن استفاده کرد.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. در سؤالهای آزمونهای کوته‌پاسخ کامل کردنی بهتر است جای خالی را در چه قسمتی از صورت سؤال قرار دهیم؟

الف ـ ابتدای سؤال
ب ـ وسط سؤال
پ ـ انتهای سؤال
ت ـ هر جا که ایجاب کند

۲. امتیاز مهم آزمونهای مسائل عددی این است که _____.

الف ـ می‌توانند هدفهای سطح کار بستن را بسنجند
ب ـ به سادگی قابل تهیه هستند
پ ـ از آزمونهای کوته‌پاسخ معمولی برای سنجش هدفهای دانش مفیدترند
ت ـ چون صرفاً با اعداد سروکار دارند هرگز دانش محض را نمی‌سنجند

سؤالهای کوته‌پاسخ زیر را به دقت بخوانید و آنها را برحسب طبقه‌بندی جدید حوزهٔ شناختی طبقه‌بندی کنید.

سؤال	نام طبقه
۳. فرمول نمک طعام را بنویسید.	.....
۴. قانون علمی را در یک سطر و به زبان ساده توضیح دهید.	.....
۵. جنگ جهانی اول از ـــــ تا ـــــ به طول انجامید.	.....
۶. در معادلهٔ زیر مقدار X را حساب کنید. $۴X - ۱۸ = -۲$	.....
۷. اگر درجهٔ حرارت یک گاز را هنگامی که فشار آن را افزایش می‌دهیم ثابت نگه داریم، حجم آن چه تغییری خواهد کرد؟	.....
۸. هدف غیررفتاری زیر را به یک هدف رفتاری تبدیل کنید: "دانش‌آموز معنی قانون شکست نور را بفهمد"	.....
۹. اگر بخواهید درخت بید باغچهٔ منزلتان را تکثیر کنید چه روشی را به کار می‌برید؟	.....
۱۰. مرکز استان فارس ـــــ است.	.....
۱۱. براساس روند افزایش جمعیت ایران در ۲۰ سال گذشته، جمعیت ایران در ۱۰ سال آینده چقدر خواهد بود؟	.....

۱۲. انواع سؤالهای کوته‌پاسخ را نام ببرید و برای هر یک مثالی ذکر کنید.

۱۳. امتیازها و محدودیتهای آزمونهای کوته‌پاسخ را توضیح دهید.

۱۴. برای سنجش هدفهای کاربستن از چه نوع سؤالهای کوته‌پاسخ می‌توان استفاده کرد؟ یک مثال بزنید.

۱۵. در این فصل یکی از توصیه‌هایی که به طراح سؤالهای کوته‌پاسخ شده این است که "تا حد امکان، به جای سؤالهای کامل‌کردنی، از سؤالهای پرسشی استفاده کنید". توضیح دهید که چرا این توصیه داده شده و فواید آن چیست؟

۱۶. سؤالهای کوته‌پاسخ و سؤالهای تشریحی محدودپاسخ را با هم مقایسه کنید.

۱۷. برای طرح سنجش خود که به عنوان تمرین فصل ۶ تهیه کرده‌اید، یک آزمون کوته‌پاسخ متناسب با هدفهایی از آن درس که با این نوع آزمونها بهتر سنجیده می‌شوند بسازید.

۱۲

سنجش عملکردی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. آزمونهای عملکردی و آزمونهای کتبی (مداد و کاغذی) را با هم مقایسه کند.
۲. سنجش واقعی یا اصیل را تعریف کند.
۳. سنجش عملکردی را تعریف کند و ویژگیهای آن را توضیح دهد.
۴. تقسیم‌بندی گرانلاند از انواع آزمونهای عملکردی را با ذکر مثال توضیح دهد.
۵. آزمونهای عملکردی مختلف مورد نظر گرانلاند را با هم مقایسه کند.
۶. منظور از سنجش قیاس‌پذیر را توضیح دهد.
۷. تقسیم‌بندی نیتکو از روشهای سنجش عملکردی را با ذکر مثال توضیح دهد.
۸. روش سنجش ساختارمند را با روش سنجش کتبی عملکردی مقایسه کند.
۹. روش سنجش در موقعیتهای طبیعی و روش سنجش نمونهٔ کار را با هم مقایسه کند.
۱۰. قواعد تهیهٔ آزمونهای عملکردی را توضیح دهد.
۱۲. نمونهٔ هدفهایی را که بهتر است با آزمونهای عملکردی سنجش شوند بنویسد.
۱۳. روشهای نمره گذاری آزمونهای عملکردی را توضیح دهد.
۱۴. موارد استفادهٔ سنجش عملکردی را توضیح دهد.
۱۵. برای هدفهای آموزشی قابل سنجش با روشهای سنجش عملکردی سؤالهای عملکردی بنویسد.

ما در چند فصل قبلی این کتاب آزمونهای کتبی یا به اصطلاح آزمونهای مداد و کاغذی را معرفی کردیم. این آزمونها عمدتاً برای اندازه‌گیری هدفهای آموزشی و بازده‌های یادگیری حوزهٔ شناختی مناسب‌اند. از آنجا که هدفهای حوزهٔ شناختی بخش مهمی از موضوعهای درسی همهٔ دوره‌های تحصیلی را شامل می‌شوند، آزمونهای کتبی شهرت زیادی کسب کرده‌اند. اما استفاده از این آزمونها برای اندازه‌گیری هدفهای سایر حوزه‌های یادگیری، یعنی حوزهٔ روانی ـ حرکتی و حوزهٔ عاطفی و حتی برای اندازه‌گیری تمامی هدفهای حوزهٔ شناختی، به تنهایی، کافی نیست، و بدین منظور لازم است از تدابیر دیگری نیز سود جست. گرانلاند و لین (۱۹۹۰) تعدادی یادگیری را که با آزمونهای کتبی یا مداد و کاغذی به خوبی قابل سنجش نیستند و برای سنجش آنها روشها و فنون دیگری به غیر از این آزمونها مورد نیازند در جدول ۱ـ۱۲ خلاصه کرده‌اند.

جدول ۱-۱۲ یادگیریهایی که به خوبی با آزمونهای متداول کتبی قابل سنجش نیستند (گرانلاند و لین، ۱۹۹۰، ص ۳۷۶)

انواع یادگیری	رفتارهای مُعرف
مهارتها	سخن گفتن، نوشتن، گوش دادن، خواندن شفاهی، انجام آزمایش در آزمایشگاه، رسم کردن، نواختن آلات موسیقی، ژیمناستیک، مهارتهای شغلی، مهارتهای مطالعه کردن، و مهارتهای اجتماعی
عادتهای کاری	توانایی در برنامه‌ریزی، استفادهٔ درست از زمان، استفاده از وسائل، استفاده از منابع، نشان دادن ویژگیهایی چون ابتکار، خلاقیت، پشتکار، و قابل اعتماد بودن
نگرشهای اجتماعی	علاقه‌مندی به رفاه دیگران، احترام گذاشتن به قانون، احترام گذاشتن به اموال دیگران، حساس بودن نسبت به مسائل اجتماعی، احترام گذاشتن به نهادهای اجتماعی، میل به کوشیدن برای پیشبرد اجتماعی
علاقه‌ها	اظهار علاقه نسبت به فعالیتهای آموزشی، مکانیکی، هنری، علمی، اجتماعی، تفریحی، شغلی
ارج‌گذاری یا قدردانی	ابراز رضایت و شادمانی نسبت به طبیعت، موسیقی، هنر، ادبیات، مهارتهای جسمانی، و خدمتهای برجستهٔ اجتماعی
سازگاری	رابطه با همسالان، واکنش به ستایش و سرزنش، واکنش به مرجع قدرت، ثبات هیجانی، سازگاری اجتماعی

چنان‌که دیده می‌شود، هیچ‌یک از آزمونهایی که قبلاً در این کتاب معرفی شدند به تنهایی نمی‌تواند این یادگیریها را به خوبی سنجش کند. بنابراین، لازم است روشها و فنون دیگری را معرفی کنیم که به کمک آنها بهتر بتوان فرایندها و فراورده‌های این یادگیری‌ها را سنجش کرد. ما در این فصل و چند فصل بعدی روشهایی را معرفی خواهیم کرد که به کمک آنها هم می‌توان یادگیریهای دانش‌آموزان و دانشجویان را در حوزه‌های روانی - حرکتی و عاطفی سنجش کرد و هم فرایندها و فراورده‌های فکری سطح بالای شناختی را که با آزمونهای سنتی مداد و کاغذی به تنهایی و به سهولت قابل اندازه گیری نیستند مورد ارزشیابی قرار داد. از آنجا که به کمک این روشهای سنجش غالباً می‌توان در موقعیتهای طبیعی و شرایط واقعی به سنجش یادگیری اقدام کرد به آنها نام **سنجش واقعی (اصیل)**[1] را داده‌اند. بنا به گفتهٔ پوفام (۲۰۰۲)، "بعضی وقتها ممکن است ببینید که صاحب‌نظران آموزش و سنجش برای توصیف سنجش عملکردی از اصطلاحهای دیگری استفاده کنند. مثلاً آنها ممکن است اصطلاح سنجش اصیل یا واقعی را به کار برند زیرا این سنجش با تکالیف زندگی واقعی و غیر آموزشگاهی بیشتر سروکار دارد" (ص ۱۷۵). سنجش واقعی را در فصل ۵ توضیح داده‌ایم.

تعریف و اهمیت سنجش عملکردی

یک دسته از روشهای سنجش واقعی یا اصیل **سنجش عملکردی**[2] نام دارد. در سنجش عملکردی یا **آزمونهای عملکردی**[3] فرایندها و فراورده‌های یادگیری دانش‌آموزان و دانشجویان به طور مستقیم سنجش می‌شوند. کوبیسزین و بوریچ (۲۰۰۳) در این باره گفته‌اند "معلمان می‌توانند از آزمونهای عملکردی برای سنجش یادگیریهای شناختی پیچیده، نگرشها، و مهارتهای اجتماعی در موضوعهای تحصیلی مختلف، مانند علوم، مطالعات اجتماعی، و ریاضیات استفاده کنند. برای این منظور، معلمان باید موقعیتهایی را تدارک ببینند که در آن رفتار یادگیرندگان را به طور مستقیم مشاهده و سنجش نمایند" (ص ۱۵۴).

در سالهای اخیر بر استفاده از سنجش عملکردی تأکید زیادی شده است. بنا به گفتهٔ پوفام (۲۰۰۲)، "در ضمن سالهای اول دههٔ ۱۹۹۰ میلادی تعداد قابل توجهی از سیاست‌گذاران آموزش و پرورش شیفتهٔ سنجش عملکردی شدند" (ص ۱۷۳). یکی از دلایل عمدهٔ تأکید و توجه به این روش سنجش پیشرفتهای نسبتاً تازه در روانشناسی شناختی و تأثیر چشم‌گیر آن بر جریانهای آموزش و پرورش بوده است. بنا به گفتهٔ ثرندایک، کانینگهام، ثرندایک، و هیگن (۱۹۹۱)، "علاقهٔ روزافزون به استفاده از اصول روانشناسی شناختی در کلاس درس نیاز به تشویق و پرورش فرایندهای عالی فکری و مهارتهای تفکر انتقادی را سبب شده است. باور عمومی این است که این فرایندها و مهارتهای عقلی را با فنون ارزشیابی عملکرد و محصول

1. authentic assessment 2. performance assessment 3. performance tests

عملکرد بهتر می‌توان سنجش کرد» (ص ۲۷۱). همچنین اسپرینتهال[1]، اسپرینتهال، و اُجا[2] (۱۹۹۴) بیان داشته‌اند که علاقه‌مندی تازهٔ متخصصان آموزشی و روانشناسان پرورشی به سنجش عملکرد یادگیرندگان امکان سنجش مهارتهای شناختی سطح بالا و تکالیف مهم مانند توانایی نوشتن، ایجاد ارتباطهای کلامی، و حل مسئله را فراهم آورده است. دمبو[3] (۱۹۹۴)، دربارهٔ اینکه چرا در سالهای اخیر آزمونهای عملکردی گسترش یافته‌اند و از شهرت آزمونهای عینی کاسته شده، دلایل زیر را ذکر کرده است:

۱. نظریه‌های شناختی یادگیری بر آموزش تأثیر گذاشته‌اند و لذا مهارتهای فکری پیچیده مورد تأکید قرار گرفته‌اند. در گذشته نظریه‌های رفتاری یادگیری افکار معلمان را نسبت به آموزش تحت تأثیر قرار می‌دادند، و از آنجا که در نظریه‌های رفتاری اعتقاد بر این است که یادگیری در مراحل کوچک انجام می‌شود، لذا نظام ارزشیابی گذشته بر آزمونهای عینی که دانشهای اساسی و خُرد را می‌سنجند مبتنی بود.

۲. نظریه‌های شناختی جدید همچنین بر جنبه‌های فکری و خود نظم‌دهی یادگیرنده تأکید می‌کنند. لذا امروزه بیشتر توجه‌ها معطوف به این است که چگونه یادگیرندگان دانش را تفسیر می‌کنند و به کار می‌گیرند تا مسائل پیچیده را حل کنند. این‌گونه مهارتها با آزمونهای عینی، مثلاً آزمونهای چندگزینه‌ای، قابل سنجش نیستند.

۳. پژوهشهای تازه نشان می‌دهند که یادگیری و انگیزش بر هم تأثیر زیادی دارند؛ یادگیرنده ممکن است دانش فراوانی داشته باشد، اما نخواهد از آن استفاده کند. بنابراین، پژوهشگران بر این باورند که در زمینهٔ نوشتن باید یادگیرندگان را وادار کرد تا نسبت به کار خود بیندیشند و کار خود را ارزشیابی نمایند (ارزشیابی شخصی). این کار به آنان کمک خواهد کرد تا معیارهای سطح بالایی برای خود برگزینند.

علت دیگر تأکید بر آزمونهای عملکردی یا سنجش واقعی در سالهای اخیر اعتراضهای شدیدی بوده است که نسبت به آزمونهای سنتی عینی (به ویژه آزمونهای چندگزینه‌ای) ابراز شده. بنا به گفتهٔ منتقدین، آزمونهای سنتی مهارتهایی را می‌سنجند که در دنیای واقعی مانند ندارند (وولفلک، ۲۰۰۴).

از دانش‌آموزان خواسته می‌شود تا سؤالهایی را جواب دهند یا مسائلی را حل کنند که هرگز دوباره با آنها روبه‌رو نخواهند شد. از آنان می‌خواهند تا این کارها را تنها انجام دهند ــ بدون اتکاء به هیچ نوع ابزار یا منبعی و در شرایط کاملاً محدود زمانی. زندگی واقعی اصلاً به اینها شبیه نیست. حل مسائل مهم به وقت کافی نیاز دارد و غالباً نیازمند استفاده از منابع، مشورت با دیگران، و ادغام مهارتهای اساسی با خلاقیت و تفکر سطح بالاست.

1. Sprinthall 2. Oja 3. Dembo

در پاسخ به این انتقادها، نهضت **سنجش واقعی** متولد شد. هدف این اقدام سنجش یادگیریهای واقعی، مهم، و پیچیده بوده است. به این رویکرد نامهای دیگر **سنجش مستقیم** یا **سنجش جایگزین** نیز داده شده است. این اصطلاحات نشان‌دهندهٔ شیوه‌هایی هستند که نوع دیگری به غیر از آزمونهای عینی چهارگزینه‌ای را دربر می‌گیرند و عملکرد دانش‌آموزان را به طور مستقیم و در تکالیف زندگی واقعی می‌سنجند. (ص ۵۳۸)

ساکس (۱۹۹۷) در رابطه با این آزمونها می‌گوید "آزمونهای عملکردی از آزمون‌شونده می‌خواهند تا به انجام نوعی رفتار اقدام کند نه اینکه صرفاً سؤالهایی را جواب دهد" (ص ۱۶۷). لین و گرانلاند (۲۰۰۰) آزمون عملکردی را به عنوان "تکالیفی که از دانش‌آموزان می‌خواهند تا مسائلی را که در خارج از محدودهٔ کلاسی با اهمیت هستند حل کنند یا به گونه‌هایی عمل نمایند که به خودی خود دارای اهمیت‌اند" (ص ۴۰) تعریف کرده‌اند.

سنجش فرآیند و فرآوردهٔ یادگیری

آزمونهای عملکردی با فرایند یا شیوهٔ اجرا، فراورده یا محصول کار، و یا ترکیبی از این دو سروکار دارند؛ ماهیت عملکرد مورد سنجش مشخص می‌کند که تأکید بر چه قسمتی باید باشد. بعضی از عملکردها به محصولات یا فراورده‌های ملموس و عینی نمی‌انجامند؛ مانند کار کردن با وسایل آزمایشگاهی، سخنرانی کردن، نواختن آلات موسیقی، و انجام دادن فعالیتهای ورزشی نظیر شنا کردن یا پرتاب کردن توپ. این‌گونه فعالیتها را باید ضمن انجام دادن سنجش کنیم و لذا با شیوهٔ اجرا یا فرایند عمل سروکار داریم.

در بعضی زمینه‌های عملکردی دیگر، محصول یا فراورده مورد تأکید است و توجه چندانی به فرایند یا شیوهٔ اجرا نمی‌شود. در ارزشیابی از مقالهٔ تحقیقی، داستان، انشاء، شعر، و نقاشیِ دانش‌آموزان و دانشجویان بیشتر فراورده مورد توجه و ارزشیابی قرار می‌گیرد. در این مثالها و مثالهای مانند آنها، با شیوه‌های اجرا و روشهای مختلفی می‌توان به نتیجهٔ واحدی دست یافت و لذا آنچه تولید می‌شود بیشتر اهمیت دارد تا فرایند تولید آن. برای نمونه، معلم معمولاً به شعری که یادگیرنده سروده است بیشتر علاقه‌مند است تا به فرایند یا جریان سرودن آن شعر.

فراورده چیزی است که یادگیرنده تولید می‌کند و فرایند جریانی است که او مورد استفاده قرار می‌دهد تا به فراورده دست یابد. در بسیاری از موارد، هم فرایند و هم فراورده هر دو به عنوان جنبه‌های مهم عملکرد سنجش می‌شوند. برای مثال، مهارت در عیب‌یابی و تعمیر یک تلویزیون یا اتومبیل شامل استفاده از یک شیوهٔ اجرا و تولید یک تلویزیون یا اتومبیل سالم است. معمولاً در آغاز دورهٔ یادگیری بیشتر شیوهٔ اجرا یا فرایند کار مورد تأکید قرار می‌گیرد، و در مراحل پیشرفته‌تر یادگیری بر فراورده یا محصول کار تأکید می‌شود. برای مثال، در سنجش مهارت ماشین‌نویسی (تایپ کردن)، در آغاز دورهٔ آموزش، درست لمس کردن دکمه‌ها

ارزشیابی می‌شود، اما در مراحل پیشرفته‌تر آموزش، تمیزی، درستی، و سرعت کار تولیدی مورد ارزشیابی قرار می‌گیرد. بنابراین، آنجا که هم شیوهٔ اجرا یعنی فرایند کار و هم فراورده یا محصول کار مورد ارزشیابی قرار می‌گیرند میزان تأکیدی که به هر قسمت اختصاص می‌یابد هم به مهارت مورد سنجش و هم به زمان سنجش در طول دورهٔ آموزش وابسته است.

سنجش عملکردی یا آزمون عملکردی با مهارت[1] سروکار دارد ــ مهارت در استفاده از فرایندها و شیوه‌های اجرایی، و نیز مهارت در تولید فراورده‌ها (گرانلاند، ۱۹۸۸). برای مثال درسهای علوم با مهارتهای آزمایشگاهی، درسهای ریاضی با مهارتهای حل مسئله، درسهای فارسی و زبانهای خارجی با مهارتهای برقرار کردن ارتباط، و درسهای علوم اجتماعی با مهارتهایی چون استفاده از نقشه و نمودار و کار در گروه سروکار دارند. اینها نمونه‌هایی از مهارت در استفاده از فرایندها و شیوه‌های اجرایی هستند. علاوه بر اینها، مهارت در تولید فراورده یا محصول در درسهای هنر و موسیقی، آموزش صنعتی و تجاری، آموزش کشاورزی، تربیت بدنی، و مانند اینها نیز از هدفهای مهم آموزش و یادگیری به حساب می‌آیند. بنابراین، آزمونهای عملکردی وسایلی هستند که به عنوان مکمل آزمونهای کتبی می‌توانند معلمان را در بهتر سنجیدن میزان توفیق یادگیرندگان در رسیدن به هدفهای متنوع تحصیلی یاری دهند.

کوبیسزین و بوریچ (۲۰۰۳) گفته‌اند آزمونهای عملکردی را هم برای سنجش فرایندهای یادگیری و هم برای سنجش فراورده‌های آن می‌توان مورد استفاده قرار داد. آنان در این‌باره مثالهای زیر را ذکر کرده‌اند: در یکی از مدارس آمریکا معلمان فرایند خواندن هر یک از دانش‌آموزان را به وسیلهٔ تعیین درصد کلمه‌هایی که درست تلفظ می‌کند، تعداد جمله‌های معنی‌داری که در ضمن بیان یک داستان به زبان می‌آورد، و درصد عناصر داستانی که بعد از خواندن یک داستان می‌تواند دربارهٔ آنها صحبت کند سنجش می‌کنند. همچنین در مدرسهٔ دیگری، دانش‌آموزان کلاس چهارم دبستان کارپوشه‌ای از کارهای خود فراهم می‌آورند و برای ارزشیابی آن را در اختیار معلم می‌گذارند. این کار پوشه‌ها در برگیرندهٔ تولیدات آنان دربارهٔ شعر، مقاله، زندگی‌نامه، و تفکرات دانش‌آموزان است. باز هم در مدرسه‌ای دیگر فرایندها و فراورده‌های یادگیری در درس مطالعات اجتماعی این‌گونه سنجش می‌شود که از دانش‌آموزان خواسته می‌شود تا دربارهٔ سؤال زیر پروژه‌های گوناگونی را انجام دهند: "براساس مطالعاتی که دربارهٔ تاریخ ایالت خود انجام داده‌اید، چه مسائلی در رابطه با این ایالت مطرح‌اند که نیاز به بحث و بررسی دارند، نظر شما دربارهٔ راه‌حل این مسائل چیست، و شما برای رسیدن به راه‌حل این مشکلات چه کمکهایی می‌توانید بکنید؟" دانش‌آموزان به صورت فردی و گروهی از طریق نوشتن گزارشهای مختلف، ارائهٔ شفاهی و تدارک نمایشگاهها، به سؤال مورد نظر جواب می‌دهند.

---

1. skill

به طور کلی، هدفهای آموزش و یادگیری را از لحاظ سنجش می‌توان به دو بخش زیر تقسیم کرد: (۱) هدفهایی که به دانستن دربارهٔ امور مربوط می‌شوند (مثلاً اینکه میکروسکوپ برای مطالعهٔ جانداران کوچکی که با چشم غیرمسلح قابل مشاهده نیستند به کار می‌رود)؛ (۲) هدفهایی که به دانستن نحوهٔ انجام دادن عمل یا کاری مربوط می‌شوند (مثلاً کارکردن با میکروسکوپ). "آن دسته از وسایل اندازه‌گیری که می‌کوشند تا بسنجند که چگونه کسی می‌تواند کاری انجام دهد غالباً آزمونهای عملکردی نام دارند، و آن دسته از وسایل اندازه‌گیری که می‌کوشند تا دامنه و صحت دانش شخص را بسنجند معمولاً آزمونهای کتبی یا مداد و کاغذی نامیده می‌شوند" (بایلر[1] و اسنومن[2]، ۱۹۹۳، ص ۵۶۵).

انواع سنجشهای عملکردی

چنان‌که قبلاً اشاره کردیم، به سنجشهای عملکردی یا آزمونهای عملکردی آزمونهای واقعی یا اصیل نیز گفته می‌شود. صاحب‌نظران، براساس پژوهشهایی که دربارهٔ سنجش عملکردی انجام داده‌اند، ویژگیهای آنها را در چهاردستهٔ زیر خلاصه کرده‌اند (نیتکو، ۲۰۰۱):

۱. **تأکید بر کاربست**، یعنی سنجش اینکه آیا دانش‌آموز، علاوه بر دانستن دانش، می‌تواند آن را به کار بندد.

۲. **تأکید بر سنجش مستقیم**، یعنی سنجشِ هدف آموزشی به صورت مستقیم، به عوض غیرمستقیم.

۳. **استفاده از مسائل واقعی**، یعنی استفاده از مسائل و موقعیتهایی که در زندگی واقعی یافت می‌شوند یا شبیه به آنها هستند.

۴. **ترغیب و تشویق تفکر باز**، یعنی هدایت دانش‌آموزان به پیداکردن راه‌حلهای مختلف برای مسائل و تشویق آنان به درگیر شدن با مسائل به طور گروهی و کارکردن بر روی آنها برای مدتی طولانی (روزها، هفته‌ها، ماهها).

آزمونهای عملکردی یا سنجشهای عملکردی بسیار متنوع‌اند. ما در اینجا انواع مهم آنها را توضیح می‌دهیم.

دسته‌بندی گرانلاند از آزمونهای عملکردی

گرانلاند (۱۹۸۸) انواع روشهای سنجش یا آزمونهای عملکردی را به چهار دستهٔ زیر تقسیم کرده است:

1. Bieler 2. Snowman

۱. آزمون کتبی عملکردی
۲. آزمون شناسایی
۳. انجام عملکرد در موقعیتهای شبیه‌سازی شده
۴. نمونهٔ کار

در این تقسیم‌بندی، آزمون نوع اول، یعنی آزمون کتبی عملکردی، بیشترین فاصله را از عملکرد در زندگی واقعی دارد، ولی آخرین نوع آزمون، یعنی نمونهٔ کار، کمترین فاصله را از عملکرد در زندگی واقعی دارد. روشهای چهارگانهٔ بالا را هم می‌توان جداگانه به کار بست و هم با یکدیگر مورد استفاده قرار داد. در بعضی موارد تنها یک روش کفایت می‌کند، اما در بیشتر مواقع کاربرد ترکیب دو یا چند روش ضروری به نظر می‌رسد. ما در زیر هر یک از این روشها را توضیح می‌دهیم.

آزمون کتبی عملکردی

هرچند در آغاز این فصل گفتیم که آزمونهای کتبی یا مداد و کاغذی برای اندازه‌گیری بازده‌های یادگیری هدفهای حوزهٔ شناختی مناسب‌ترند، با این حال، در سنجش عملکرد، استفاده از آزمونهای کتبی نیز مفید است. اما باید دقت کرد که بین آزمونهای کتبی مورد استفاده برای سنجش هدفهای حوزهٔ شناختی و آزمونهای کتبی مورد استفاده برای سنجش عملکرد تفاوت وجود دارد. تفاوت عمدهٔ بین **آزمون کتبی عملکردی**[1] و آزمونهای کتبی که در فصلهای قبلی این کتاب معرفی شدند این است که در آزمون کتبی عملکردی عمدتاً بر کاربست دانش و مهارت در موقعیتهای عملی یا شبیه‌سازی شده با موقعیتهای عملی تأکید می‌شود. در این‌گونه آزمونهای عملکردی، یا بازده‌های پایانی یادگیری سنجش می‌شوند یا مراحل میانی عملکردی که برای رسیدن به بازده‌های مطلوب پایانی ضروری هستند، مانند استفادهٔ درست از ابزارها و دستگاهها.

به عنوان نمونه‌هایی از سؤالهای آزمون کتبی عملکردی که در هر یک از آنها یک بازده مهم یادگیری سنجش می‌شود به موارد زیر توجه کنید.

دانشجوی درس اندازه‌گیری و ارزشیابی آموزشی در تواناییهای زیر سنجش می‌شود:

- *ساختن یک جدول مشخصات درس برای یک واحد آموزشی*
- *ساختن تعدادی سؤال آزمون متناسب با جدول مشخصات*
- *ساختن یک فهرست وارسی برای ارزشیابی از یک آزمون پیشرفت تحصیلی*

(گرانلاند، ۱۹۸۸، ص ۸۷)

---

1. paper-and-pencil performance test

همچنین از دانش‌آموزان و دانشجویان می‌توان خواست تا یک نقشهٔ آب و هوایی، یک نمودار ستونی، طرح یک مدار برقی، طرح یک لباس، یک داستان کوتاه، یا نقشهٔ یک آزمایش علمی را بسازند. در این مثالها، محصول یادگیری که به صورت کتبی ارائه می‌شود، هم نتیجهٔ دانش فرد است و هم نتیجهٔ مهارت او، و یک مقیاس عملکردی به دست می‌دهد که به خودی خود ارزشمند است و، برخلاف آزمونهای غیرعملکردی، به عنوان وسیله‌ای برای قضاوت دربارهٔ یک صفت یا توانایی روانی غیرقابل مشاهده به کار نمی‌رود.

علاوه بر موارد بالا که در آنها محصول پایانی یا نهایی یادگیری سنجش می‌شوند، در بعضی مواقع می‌توان از آزمونهای کتبی عملکردی برای سنجش مهارتها یا فعالیتهای واسطه‌ای نیز سود جُست. برای مثال، پیش از استفاده از یک دستگاه یا یک ابزار خاص، مانند، ریزسنج[1]، ممکن است ضروری باشد که یادگیرندگان، از روی عکسهای مقیاس، موقعیتهای مختلف را مطالعه کنند. اگرچه توانایی خواندن مقیاس شرط کافی برای سنجش دقیق نیست، اما یک شرط لازم است. در اینجا آزمون کتبی برای آزمودن یادگیرندگان به صورت گروهی روش مناسبی است. علاوه بر این، آنجا که عملکرد پیچیده و دستگاه خیلی گران قیمت است، نشان دادن شایستگی انجام عمل در موقعیتهای آزمون کتبی از حوادثی که ممکن است دستگاه را ناقص کند جلوگیری به عمل می‌آورد. همچنین از آزمونهای کتبی در علوم بهداشت و پزشکی نیز می‌توان سود جست. مثلاً مهارت تشخیص و تجویز را در رابطه با بیماران فرضی می‌توان در دانشجویان پزشکی سنجش کرد تا اطمینان حاصل شود که بعدها در برخورد با بیماران واقعی با اشکال روبه‌رو نمی‌شوند. کوتاه سخن اینکه، آزمون عملکردی کتبی را می‌توان به عنوان وسیلهٔ سنجش اطلاعات مقدماتی فرد دربارهٔ کاری که انجام خواهد داد به کار بست. مثلاً اطلاعات فرد از ساختار ابزاری که مورد استفاده قرار خواهد داد، پیش از اینکه عملاً شروع به کار کردن با دستگاه بکند، اندازه گیری می‌شود. این نوع اطلاعات به صورت کتبی و در صورت لزوم به صورت گروهی قابل سنجش است.

آزمون شناسایی

منظور از **آزمون شناسایی**[2] روشی است برای سنجش توانایی یادگیرنده در تشخیص ویژگیها، محاسن، معایب و موارد استفادهٔ امور مختلف.

آزمونهای شناسایی انواع و کاربردهای مختلفی دارند. در بعضی وقتها می‌توان از یادگیرنده خواست تا یک ابزار یا دستگاه را شناسایی نماید یا کار آن را توصیف کند. در موقعیتهای دیگر می‌توان یادگیرنده را با مشکلی مواجه ساخت، مانند یک اتصالی در یک دستگاه برقی، و از او خواست تا ابزارها، وسایل، و شیوه‌های عملی رفع مشکل را شناسایی کند. مورد پیچیده‌تری از

1. micrometer 2. identification test

این نوع آزمون گوش دادن به یک دستگاه معیوب، مثلاً موتور یک اتومبیل، و بعد از روی صدا شناسایی کردن علت یا علتهای اشکال دستگاه است.

آزمونهای شناسایی هم در آموزش صنعتی کاربرد دارند، هم در علوم زیستی، و هم در سایر رشته‌های علمی. برای نمونه، از دانش‌آموزان و دانشجویان درس زیست‌شناسی می‌توان خواست تا نمونه‌های جانوران مختلفی را که در اتاق آزمایش یا کلاس درس چیده شده‌اند شناسایی کنند، یا از دانش‌آموزان و دانشجویان درس شیمی می‌توان خواست تا وسایل و مراحل انجام یک آزمایش شیمیایی را مشخص کنند. همچنین تشخیص تلفظ درست یک کلمهٔ خارجی در درس زبان خارجی و تشخیص روش درست حل یک مسئله در درس ریاضی نیز مشمول آزمونهای شناسایی هستند. علاوه بر اینها، شناسایی بهترین شیوه‌های عملی در درسهای هنر، موسیقی، تربیت بدنی، و فعالیتهای حرفه‌ای مانند کشاورزی، نجاری، مکانیکی، و غیره ضروری‌اند.

از آزمونهای شناسایی به عنوان وسایل سنجش غیرمستقیم عملکرد نیز می‌توان استفاده کرد. برای مثال، انتظار می‌رود که یک لوله‌کش ماهر از یک لوله‌کش تازه‌کار اطلاعات بیشتری دربارهٔ ابزارها و وسایل لوله‌کشی داشته باشد. بنابراین، می‌توان از یک آزمون شناسایی برای سرند کردن گروه بزرگی از داوطلبان این شغل استفاده کرد، و در مرحلهٔ اول انتخاب افراد ناوارد به کار لوله‌کشی را کنار گذاشت تا اینکه در مرحلهٔ بعد انتخاب ماهرترین افراد به وسیلهٔ آزمونهای مستقیم سنجش عملکرد سهل‌تر گردد. گرانلاند (۱۹۸۸) در رابطه با استفاده از آزمونهای شناسایی گفته است، "معمول‌تر آن است که از آزمونهای شناسایی به عنوان یک تدبیر آموزشی برای آماده کردن دانش‌آموزان و دانشجویان در اجرای عمل در موقعیتهای واقعی یا شبیه‌سازی شده استفاده شود" (ص ۸۸).

آزمون بالینی ساختارمند عینی (آسکی)[1] یکی از موارد استفادهٔ آزمون شناسایی که در آموزش پزشکی کاربرد زیادی دارد آزمون بالینی ساختارمند عینی است که به طور خلاصه آسکی خوانده می‌شود. از این نوع آزمون هم به منظور سنجش و هم با هدف آموزش استفاده می‌شود. در اتاق مخصوص سنجش ایستگاههای مختلفی وجود دارند که آزمون شونده باید به همهٔ آنها سر بزند. در یکی از آن ایستگاهها مثلاً عکسهای بالینی وجود دارند که باید تفسیر شوند، در ایستگاه بعدی مدارک دیگر بیماری موجودند که باید براساس آنها بیماری تشخیص داده شود، در ایستگاه دیگر ممکن است یک بیمار واقعی (یا کسی که نقش یک بیمار را بازی می‌کند) حاضر باشد که باید معاینه و مصاحبه شود. بنا به گفتهٔ ریس، براون، و اسمیت (۲۰۰۵)، این آزمونها دارای روایی سطح بالایی هستند، زیرا چیزی را می‌سنجند که پزشکان باید قادر به انجام آنها باشند. "معمولاً نمره‌گذاری این آزمونها آسان‌تر از امتحانهای کتبی است (نه تنها به

1. objective structured clinical exam (OSCE)

این دلیل که دست خط پزشکان از بدی زبان‌زد همگان است!). اما برنامه‌ریزی این‌گونه آزمونها وقت زیادی می‌برد" (ص ۵۳).

انجام عملکرد در موقعیتهای شبیه‌سازی شده

در آزمون شبیه سازی یا **انجام عملکرد در موقعیتهای شبیه‌سازی شده**[1] از یادگیرنده خواسته می‌شود تا در یک موقعیت شبیه‌سازی شده یا مصنوعی یا خیالی همان اعمالی را انجام دهد که در موقعیتهای واقعی ضروری هستند) برای نمونه، در تربیت بدنی، ضربه زدن به یک توپ خیالی، مشت بازی کردن به طور خیالی، یا انجام حرکات مختلف شنا در بیرون از آب نمونه‌هایی از عملکرد در موقعیتهای شبیه‌سازی شده هستند. در درسهای علوم و حرفه‌وفن کار در آزمایشگاه یا کارگاه به صورت عملکرد در شغل واقعی شبیه‌سازی شده است؛ اما در بعضی موارد، دستگاهها و وسایل بخصوصی برای آموزش و آزمودن از راه انجام عملکرد شبیه‌سازی شده فراهم می‌آیند. مثلاً در آموزش رانندگی، پیش از آنکه یادگیرنده را عهده دار راندن یک اتومبیل واقعی بکنند، او را در یک اتومبیل که مانند اتومبیلهای واقعی شبیه‌سازی شده است و همهٔ دستگاههای رانندگی از جمله فرمان، آینه، پدالها و غیره را دارد قرار می‌دهند، و از او می‌خواهند تا اعمال رانندگی را در این خودرو مصنوعی تمرین کند. به همین منوال، برای آموزش خلبانی نیز از دستگاههایی که با هواپیماهای واقعی شبیه‌سازی شده‌اند استفاده می‌کنند. این گونه امکانات شبیه‌سازی شده از صدمه رسیدن به یادگیرنده و دیگران در مراحل اولیهٔ یادگیری جلوگیری به عمل می‌آورد.

در نمونه‌های بالا، انجام عملکرد در موقعیتهای شبیه‌سازی شده بیشتر برای آموزش مهارتهای مختلف مانند رانندگی اتومبیل مورد استفاده قرار می‌گیرد. با این حال، از این روش برای ارزشیابی از تواناییها و مهارتهای یادگیرندگان در انجام کارهای مختلف (فرایند یادگیری) و تولید بازده‌های مطلوب (فراورده‌های یادگیری) نیز می‌توان سود جست. برای مثال، آنجا که عملکرد دانش‌آموز یا دانشجو در انجام آزمایشهای شیمی در آزمایشگاه سنجش می‌شود، این روش برای ارزشیابی از توانایی آموخته شدهٔ یادگیرنده مورد استفاده قرار می‌گیرد.

بعضی از صاحب‌نظران سنجشِ یادگیری (مانند هینتزه، استونر، و بول، ۲۰۰۰) به جای انجام عملکرد در موقعیتهای شبیه‌سازی شده اصطلاح سنجش قیاس‌پذیر را به کار برده‌اند. منظور آنان از سنجش قیاس‌پذیر مشاهده و ارزشیابی عملکرد یادگیرنده در موقعیتهایی است که شبیه به موقعیتهای واقعی است. هدف کسانی که سنجش قیاس‌پذیر را مورد استفاده قرار می‌دهند پیش‌بینی وقوع رفتارهای مورد سنجش در محیط زندگی واقعی است.

برای آشنایی بیشتر با آزمونهای عملکردی از نوع انجام عملکرد در موقعیتهای شبیه‌سازی

1. simulated performance

شده، یک نمونهٔ واقعی را توضیح می‌دهیم. این نمونه را از کتاب سنجش کلاسی تألیف جیمز پوفام (۲۰۰۲) نقل می‌کنیم. از این آزمون عملکردی به منظور سنجش توانایی ایجاد ارتباط کلامی در دانش‌آموزان و دانشجویان استفاده می‌شود.

موقعیت سنجش: یک حالت اضطراری (اورژانس)

فرض کنید تنها در منزل هستید و بوی دود می‌شنوید. شما [آزمون‌شونده یا دانش‌آموز] به اداره آتش‌نشانی تلفن می‌زنید و من [سنجشگر یا معلم] به تلفن شما جواب می‌دهم. با من صحبت کنید، آن‌گونه که در تلفن صحبت می‌کنید. هر چیزی را که لازم می‌دانید، برای من توضیح بدهید تا به کمک شما بیایم. (مستقیماً با من صحبت کنید، با گفتن الو شروع کنید.)

(پوفام، ۲۰۰۲، ص ۱۸۵)

نمونهٔ کار

در روش نمونهٔ کار[1] از یادگیرنده خواسته می‌شود تا اعمالی را انجام دهد که مُعرف عملکرد واقعی مورد سنجش هستند. بدین لحاظ، این روش نزدیک‌ترین روش سنجش به عملکرد واقعی یادگیرنده در محیطهای طبیعی است. در روش نمونهٔ کار، نمونه اعمالی که از یادگیرنده می‌خواهیم تا انجام دهد باید شامل عناصر مهم عملکرد کلی باشد که در شرایط کنترل شده اجرا می‌شود. برای مثال، در آزمون رانندگی اتومبیل، از یادگیرنده خواسته می‌شود تا طول فاصله‌ای را که دربرگیرندهٔ موقعیتهای رانندگی معمولی هستند و فرد در رانندگی واقعی با آنها روبه‌رو می‌شود اتومبیل رانی کند. براساس عملکرد فرد در راندن اتومبیل در طول این فاصلهٔ معین، نسبت به توانایی کلی رانندگی او قضاوت صورت می‌گیرد.

مثالی دیگر از کاربرد روش نمونهٔ کار در ارزشیابی از عملکرد یادگیرنده **آموزش دانشجویی**[2] است. در آموزش دانشجویی از دانشجوی تربیت معلم خواسته می‌شود تا به عنوان نمونه، با استفاده از روش خاصی که در کلاس درس روانشناسی تربیتی یا روشهای آموزشی یادگرفته است، درسی را به گروهی دانش‌آموز آموزش بدهد.

در سایر زمینه‌های آموزشی نیز می‌توان از روش نمونهٔ کار به خوبی استفاده کرد. مثلاً در درس ترجمه می‌توان از دانشجویان خواست تا متنی را از زبانی به زبان دیگر برگردانند؛ در درس ماشین‌نویسی می‌توان از کارآموزان خواست تا یک نامه را ماشین کنند؛ یا در درسهای حرفه‌وفن می‌توان از دانش‌آموزان خواست تا وسیله‌ای را تعمیر نمایند. روش نمونهٔ کار برای سنجش توانایی افراد در انجام کارهای مختلف از دیرباز در صنعت و مشاغل نیازمندِ فعالیتهای عملکردی متداول بوده است؛ می‌توان از آنها در محیطهای آموزشی نیز حداکثر استفاده را به عمل آورد.

---

1. work sample 2. student teaching

دسته‌بندی نیتکو از روشهای سنجش عملکردی

نیتکو (۲۰۰۱) از روشها و فنون سنجش عملکردی یک تقسیم‌بندی چهارقسمتی به صورت زیر به دست داده است:

۱. سنجش ساختارمند یا کنترل شده
۲. سنجش در موقعیتهای طبیعی
۳. پروژه‌های طولانی مدت
۴. کارپوشه

هر یک از روشهای چهارگانهٔ بالا شامل تعدادی فن سنجش است. در زیر به توضیح تقسیم‌بندی نیتکو از روشها و فنون سنجش عملکردی توجه کنید.

سنجش ساختارمند

در سنجش ساختارمند[1] یا کنترل شده معلم یا سنجش‌گر هم بر تکالیف سنجش و هم بر شرایطی که در آن دانش‌آموز باید پاسخ دهد کنترل اعمال می‌کند. به سخن دیگر، معلم تصمیم می‌گیرد که از چه موادی استفاده شود و در چه زمانی آن مواد مورد استفاده قرار گیرند. همچنین به عهدهٔ معلم است که راهنمای عملکرد دانش‌آموز و نوع بازده‌های یادگیری او را تعیین کند. از این نظر، به تکالیف مورد استفاده در این روش تکالیف کنترل شده نیز گفته می‌شود.

فنون سنجش ساختارمند با توجه به تکالیف مورد استفاده دو نوع‌اند: (۱) تکالیف مداد و کاغذی و (۲) تکالیفی که به منابع و تجهیزات فراتر از مداد و کاغذ نیازمندند. تکالیف مداد و کاغذی در واقع همان آزمونهای تشریحی یا انشایی هستند که انواع و چگونگی تهیه و تصحیح آنها را به تفصیل در فصل ۱۰ توضیح دادیم. در این نوع سنجش عملکردی از آزمون‌شونده خواسته می‌شود که نه تنها پاسخ خود را به صورت کتبی بنویسد بلکه دربارهٔ آن توضیح دهد و دلیل بیاورد. با این روش هم فرایند اندیشه سنجش می‌شود و هم فراوردهٔ آن. به نمونه‌های زیر از این‌گونه تکالیف سنجش توجه کنید:

- *این مسئلهٔ داستان‌گونه را حل کنید و راه‌حل خود را توضیح دهید.*
- *پس از مطالعهٔ نمودار زیر که نشان می‌دهد یک دانش‌آموز چگونه وقت خود را صرف می‌کند، دربارهٔ یک روز زندگی آن دانش‌آموز داستانی بنویسید.*

در بسیاری از تکالیف سنجش، مداد و کاغذ به تنهایی کفایت نمی‌کند بلکه وسایل، ابزارها، و تجهیزات دیگری مورد نیازند. در این تکالیف دانش‌آموز هم با استفاده از وسایل کاری انجام می‌دهد

1. structured assessment

و هم گاهی دربارهٔ کاری که انجام داده چیزی می‌نویسد. نمونه‌های این نوع تکالیف در زیر آمده‌اند:

- از یک مجموعه مثلث چهارتایی هر چند شکل هندسی که می‌توانید درست کنید.
- از طریق تلفن دربارهٔ یک شغل کسب اطلاع کنید و برای دریافت آن شغل تلفنی تقاضا نمایید.
- چگونگی مخلوط کردن اسید و آب را به طور عملی نشان دهید.

سنجش در موقعیتهای طبیعی

در روش سنجش در **موقعیتهای طبیعی**[1] به جای اینکه تکلیفی برای دانش‌آموز تعیین شود و موقعیت سنجش زیر کنترل قرار گیرد، معلم منتظر می‌ماند تا عملکرد در شرایط طبیعی رخ دهد و آن‌گاه به سنجش آن اقدام می‌کند. از آنجا که در این روش رفتارهای معمولی یا نوعی دانش‌آموز سنجیده می‌شوند به آن روشِ تکالیفِ **عملکرد نوعی**[2] نیز گفته می‌شود. نمونه‌هایی از این‌گونه تکالیف به قرار زیرند:

- جمع‌آوری مطالبی که دانش‌آموز در زمینه‌های مختلف نوشته و تصحیح آنها از لحاظ غلطهای دستوری، املایی، و انشایی.
- مشاهدهٔ یک دانش‌آموز در زمین بازی از لحاظ رفتاری که در ارتباط با درگیری بین دوستانش از خود نشان می‌دهد.

پروژه‌های طولانی مدت

آنچه به **پروژه‌های طولانی‌مدت**[3] شهرت یافته شامل دو قسمت است: (۱) **پروژه‌های فردی**[4] و (۲) **پروژه‌های گروهی**[5]. پروژهٔ فردی به یک فعالیت طولانی مدت دانش‌آموزی گفته می‌شود که به نوعی فراورده یا محصول مانند یک الگو، یک گزارش، یک وسیله، یا یک مجموعه (کلکسیون) می‌انجامد. ویژگی پروژهٔ دانش‌آموزی این است که در آن چند فعالیت مختلف با هم ترکیب می‌شوند تا یک فراورده به وجود آید. برای نمونه، انجام یک تحقیق کتابخانه‌ای به عنوان یک پروژهٔ دانش‌آموزی مستلزم انجام فعالیتهای زیر است: استفاده از منابع تحقیقی موجود در کتابخانه، خلاصه‌برداری از مطالب، تهیهٔ طرح یک گزارش، تدوین مطالب مختلف جمع‌آوری شده، تفسیر نتایج، سازمان‌بندی مطالب، و تهیهٔ گزارش نهایی. در زیر نمونه‌های دیگری از پروژه‌های انفرادی داده شده‌اند:

- آگهی‌های روزنامه‌ها و مجله‌ها را جمع‌آوری و طبقه‌بندی کنید.
- الگوی یک دوربین عکاسی بسازید.

---

1. naturally occuring
2. typical performance
3. long term projects
4. individual projects
5. group projects

علاوه بر پروژه‌های فردی، معلم می‌تواند عملکرد دانش‌آموزان را در پروژه‌های گروهی نیز سنجش کند. در این پروژه‌ها دو یا چند تن از دانش‌آموزان همکاری می‌کنند. هدف اصلی پروژهٔ گروهی، به عنوان یک وسیلهٔ سنجش، این است که تعیین شود آیا دانش‌آموزان می‌توانند به صورت مشارکتی با هم کار کنند و یک محصول با ارزش تولید نمایند. نمونه هایی از پروژه‌های گروهی از قرار زیرند:

- تهیهٔ یک روزنامهٔ دیواری
- نوشتن یک مقاله دربارهٔ یک موضوع مورد علاقهٔ دانش‌آموزان
- تهیهٔ یک گزارش دربارهٔ اینکه چگونه رودخانه‌ها شکل می‌گیرند

کارپوشه

منظور از کارپوشه[1] مجموعهٔ محدودی از کارهای دانش‌آموز است که برای سنجش عملکرد او مورد استفاده قرار می‌گیرد. کارپوشه هدفهای مهم و متنوعی را برآورده می‌سازد که مهم‌ترین آنها یکی معرفی بهترین کارهای دانش‌آموز در طول یک مدت معین مثل یک ترم یا سال تحصیلی و دیگری نشان دادن چگونگی رشد تحصیلی او در یک مدت معین است. ما در فصل بعدی این کتاب روش کارپوشه را به تفصیل مورد بحث قرار خواهیم داد.

قواعد تهیهٔ آزمونهای عملکردی

در تهیهٔ آزمونهای عملکردی، همان مراحل و تدابیر مورد استفاده در ساختن سایر آزمونهای معرفی شده در این کتاب مورد استفاده قرار می‌گیرند. با این حال، در تهیهٔ آزمونهای عملکردی رعایت پاره‌ای ملاحظات دیگر نیز ضروری‌اند که در این قسمت به آنها می‌پردازیم. یکی از تفاوتهای عمدهٔ آزمونهای عملکردی با سایر آزمونهایی که در فصلهای قبلی این کتاب معرفی شدند این است که تهیه و اجرای آزمونهای عملکردی وقت بیشتری لازم دارد و نمره گذاری آنها دشوارتر است. قاعده‌هایی که در زیر مورد بحث قرار می‌گیرند برای کاهش دادن دشواری تهیه، اجرا، و نمره گذاری این آزمونها از سوی صاحب‌نظران پیشنهاد شده‌اند.

۱. **بازده‌های مورد نظر عملکرد را مشخص کنید.** از میان هدفهای آموزشی درس که قبلاً تهیه شده و در بُعد هدفِ جدول مشخصات قرار داده‌اید، آنهایی را برگزینید که به سنجش عملکردی نیاز دارند؛ هدفهای قابل اندازه‌گیری با آزمونهای مرسوم را به عهدهٔ همان آزمونها بگذارید. دقت کنید آن دسته از هدفهای آموزشی که بر حسب عملکرد قابل اندازه‌گیری نوشته شده‌اند و هم شرایط عملکرد و هم ملاک عملکرد دارند با آزمونهای عملکردی به راحتی قابل

1. portfolio

سنجش‌اند. بنابراین، سعی کنید، هنگام تهیهٔ جدول مشخصات درس و نوشتن هدفهای آموزشی، هدفهای عملکردی را در قالب عملکردهای یادگیرندگان بنویسید. ملاک یا معیار عملکرد را در هدفهای عملکردی می‌توان با توجه به یکی از موارد زیر مشخص کرد.

الف ـ با توجه به دقت عملکرد، مثلاً:

با استفاده از یک دماسنج، دما را تا دهم درجه اندازه گیری کند.

ب ـ با توجه به سرعت عملکرد، مثلاً:

با در ا ختیار داشتن وسایل لازم، در پانزده دقیقه عیب یک رادیو را تشخیص دهد.

پ ـ با توجه به توالی درست مراحل، مثلاً:

در آزمایشگاه یک میکروسکوپ را با رعایت مراحل درست تنظیم کند.

ت ـ با توجه به مهارت عملکرد، مثلاً:

با در دست داشتن ارّه و چوب، یک قطعه چوب را با مهارت ارّه کند.

ث ـ با توجه به رعایت ایمنی، مثلاً:

اتومبیلی را بدون زیر پاگذاشتن هیچ یک از مقررات ایمنی رانندگی در دو چهارراه براند.

ملاکهای عملکرد بالا را می‌توان به طور انفرادی یا به صورت ترکیبی مورد استفاده قرار داد. برای نمونه، در هدف زیر دو نوع ملاک عملکرد به کار رفته‌اند:

- یادگیرنده بتواند، بر روی یک ماشین تحریر آی بی ام برقی، با سرعت پنجاه کلمه در دقیقه و حداکثر دو غلط در هر صفحه مطالبی را که به او می‌دهند ماشین کند.

در هدف بالا، هم از ملاک سرعت عملکرد (۵۰ کلمه در دقیقه) استفاده شده و هم از ملاک دقّت عملکرد (حداکثر دو غلط در هر صفحه).

اینکه کدام یک از جنبه‌های عملکرد باید مورد تأکید قرار گیرد هم به ماهیت عملکرد و هم به مرحله‌ای از آموزش که عملکرد در آن باید سنجیده شود وابسته است. برای مثال، در سنجش مهارت اندازه گیری در آزمایشگاه، در مراحل اولیهٔ آموزش دقّت مورد تأکید قرار می‌گیرد، ولی در مراحل آخر آموزش سرعت نیز اهمیت پیدا می‌کند. موقعیتی که در آن عملکرد باید انجام شود نیز باید مورد توجه قرار گیرد. مثلاً در سنجش مهارت ماشین‌نویسی در تایپ نامه‌های معمولی اداری سرعت عمل مهم است، در حالی که در ماشین کردن جدولهای آماری یا گزارشهای اقتصادی دقت عمل ضروری است.

گفتیم که از میان هدفهای درس آنهایی را برای ارزشیابی با روشهای سنجش عملکردی انتخاب کنید که با آزمونهای مداد و کاغذی به خوبی قابل سنجش نیستند. کوبیسزین و بوریچ (۲۰۰۳) هدفهای زیر را به عنوان نمونهٔ هدفهای آموزشی معلم که بهتر است به کمک روشهای سنجش عملکردی سنجش شوند ذکر کرده‌اند:

دانش‌آموز بتواند:

- برای سرزمین آمریکای شمالی، از حفظ، یک نقشهٔ جغرافیایی رسم کند و بر روی آن محل ۱۰ شهر را علامت بزند.
- با استفاده از سیم، کلید، لامپ، و باتری، یک مدار برقی درست کند.
- یک برنامهٔ کامپیوتری برای حل معادله‌های یک مجهولی بنویسد.

۲. **از تکالیف معنی‌دار استفاده کنید.** تکالیفی که در آن یادگیرنده باید عملکرد یا توانایی خود را به منظور سنجش آن توسط معلم نشان دهد باید تکالیفی معنی‌دار باشند. استفاده از تکالیف معنی‌دار سبب می‌شود که یادگیرندگان در انجام عملکرد یا حل مسائل شخصاً درگیر شوند. نیتکو (۲۰۰۱) برای خلق تکالیف معنی‌دار پیشنهادهای زیر را از بارون[1] (۱۹۹۱) نقل کرده است:

الف ـ تکالیفی را برگزینید که برای دانش‌آموزان معنی‌دار باشند.

ب ـ موقعیتهای تازه و آشنا را با هم ترکیب کنید به گونه‌ای که تکالیف سنجش برای دانش‌آموزان چالش‌انگیز باشند. یعنی تکالیف بسیار دشوار یا بیگانه را که موجب سرخوردگی دانش‌آموزان می‌شوند به کار نبرید و از تکالیف خیلی ساده هم پرهیز کنید.

پ ـ موقعیتهایی را برگزینید که بر تجارب دنیای واقعی دانش‌آموزان استوار باشند.

ت ـ از تکالیفی استفاده کنید که توانایی کاربستن دانشها و مهارتهای آموخته شدهٔ دانش‌آموز از درون کلاس به بیرون از کلاس را نشان دهد.

۳. **موقعیت آزمون را واقع‌بینانه برگزینید.** منظور از این پیشنهاد آن است که موقعیتها و شرایطی را که می‌خواهید در آن عملکرد مورد سنجش را بیازمایید به گونه‌ای انتخاب کنید که همهٔ جوانب امر از جمله ملاحظات اقتصادی، زمانی، انسانی، عملی، و غیره را منظور نمایید. برای مثال، در آزمودن مهارت کمکهای اولیه، استفاده از بیماران واقعی دارای استخوانهای شکسته، زخمهای شدید، و سایر مشکلات جدی نه میسر است و نه مطلوب. همچنین اگر استفاده از یک دستگاه گران قیمت و کمیاب برای آزمودن مهارت آزمون‌شوندگان در راه‌اندازی آنها ممکن است موجب ایجاد خسارت به دستگاه بشود، بهتر است برای این منظور از یک دستگاه شبیه‌سازی شده یا طرح دستگاه واقعی استفاده شود.

۴. **از راهنماییها و دستورالعملهایی که به روشنی موقعیت آزمون و چگونگی انجام عملکرد را مشخص می‌کنند استفاده نمایید.** منظور از این پیشنهاد آن است که راهنماییهایی باید برای آزمون فراهم شوند که در آنها عملکرد مورد نظر و شرایطی که در آن قرار است عملکرد نشان داده شود مشخص گردند. گرانلاند (۱۹۸۸) موارد زیر را برای منظور کردن در راهنمایی یا دستورالعملِ

1. Barone

آزمونِ نمونهٔ کار پیشنهاد داده است:

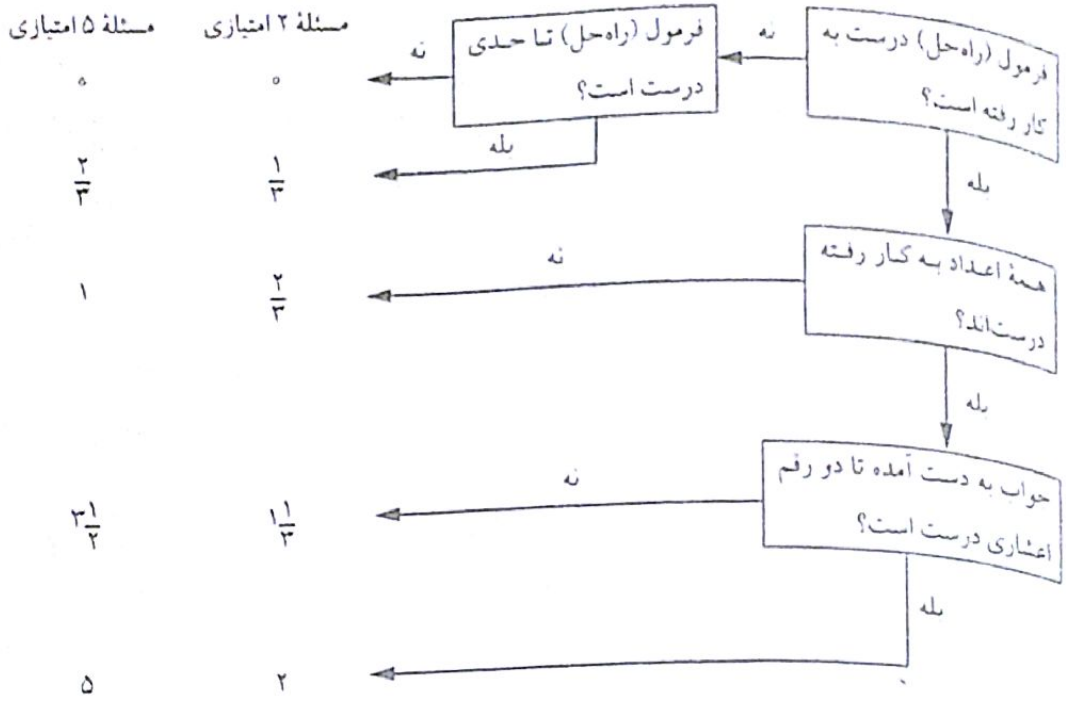
الف ـ مقصود از آزمون
ب ـ مواد و تجهیزات
پ ـ جریان آزمودن:
ـ شرایط تجهیزات
ـ عملکرد مورد انتظار
ـ محدودهٔ زمانی (در صورتی که محدودیت زمانی لازم باشد)
ت ـ روش نمره‌گذاری

دستورالعملها باید به صورت کتبی باشند تا اینکه همهٔ افراد تکالیف را یکسان دریافت کنند. در بعضی مواقع، آزمون‌شوندگان دستورالعملها را خود می‌خوانند ولی در مواقع دیگر برای آنان خوانده می‌شوند. در هر حال، دستورالعملها یا راهنماییهای چگونگی انجام عملکرد باید کاملاً واضح باشند و برای همهٔ دانش‌آموزان به طور یکنواخت به کار بروند.

۵. **در ارزشیابی از عملکرد یادگیرندگان از راهنمای نمره‌گذاری استفاده کنید.** چنانکه در فصل ۱۰ توضیح دادیم، منظور از راهنمای ارزشیابی یا راهنمای نمره‌گذاری نوعی قاعده برای سنجش عملکرد یا داوری دربارهٔ پاسخ یادگیرنده است. عموماً برای ارزشیابی از عملکرد در روشهای سنجش عملکردی از سه دسته راهنما با عنوانهای **تحلیلی، کلی (یا سراسری)**، و **کمّی** استفاده می‌شود. ما روشهای تحلیلی و کلی نمره‌گذاری را در فصل ۱۰ در ارتباط با تصحیح جواب سؤالهای تشریحی معرفی کردیم. در اینجا نیز آنها را بیشتر توضیح می‌دهیم. علاوه بر این دو روش، فهرست وارسی و مقیاس درجه‌بندی (روشهای کمّی) نیز در نمره‌گذاری پاسخهای دانش‌آموزان مورد استفاده قرار می‌گیرند. ما این روشها را در فصل بعدی (فصل ۱۴) توضیح خواهیم داد.

در روش تحلیلی، ابتدا ابعاد یا ویژگیهای پاسخ مورد سنجش را مشخص می‌کنیم و سپس برای هر یک از آنها ارزش یا نمره‌ای درنظر می‌گیریم. به عنوان نمونه، یک گزارش کتبی دربارهٔ نتایج یک آزمایش علمی را می‌توان با توجه به صحت اطلاعات، کیفیت تحلیل نتایج، و چگونگی توجیه نتایج به دست آمده مورد ارزشیابی قرار داد (لین و گرانلاند، ۲۰۰۰).

در روش تحلیلی نمره‌گذاری، پس از آنکه اجزای پاسخ نمونه (فرایند یا فراورده) نمره‌گذاری شدند، این نمره‌ها یا امتیازهای جداگانه با هم جمع می‌شوند و نمرهٔ کل دانش‌آموز در سؤال تعیین می‌گردد. شکل ۱ـ۱۲ یک نمونه پاسخ تصحیح شده با روش تحلیلی نمره‌گذاری را نشان می‌دهد. در این مثال، اجزای راه حل برای یک مسئلهٔ ریاضی با دو نوع امتیازبندی (۲ و ۵) نمره‌گذاری شده‌اند. دقت کنید که در این مثال، چون نمرهٔ کل از ۲ به ۵ تغییر کرده، امتیازهای قسمتهای مختلف راه‌حل مسئله نیز متفاوت شده است.



شکل ۱ـ۱۲ روش تحلیلی نمره دادن به اجزای راه‌حل یک مسئلهٔ ریاضی (با دو نوع امتیازبندی) (نیتکو، ۲۰۰۱، ص ۱۹۵)

علاوه بر مثالهای شکل ۱ـ۱۲، در جدول ۲ـ۱۲ نیز چند مورد راهنمای نمره گذاری به روش تحلیلی نشان داده شده. در این مثالها، جزئیات پاسخ مورد نظر به تفکیک مشخص شده‌اند. معلم در نمره گذاری خود می‌تواند برای هر قسمت امتیازهای مختلفی منظور کند.

در روش نمره گذاری کلی یا سراسری یک داوری کلی دربارهٔ پاسخ یادگیرنده به عمل می‌آید. به این منظور، تعدادی طبقه از قبل مشخص می‌شوند و قاعدهٔ نمره گذاری کلی در رابطه با آن طبقه‌ها و با توجه به میزان شایستگی پاسخها تعریف می‌شود: مثلاً، خیلی خوب، خوب، متوسط و غیره. لین و گرانلاند (۲۰۰۰)، در مقایسهٔ روش کلی ارزشیابی با روش تحلیلی، گفته‌اند "روش کلی با داوریهای کلی مورد نیاز برای نمره‌دادن به یادگیرندگان سازگارتر است، اما بازخوردهای لازم را دربارهٔ نقاط قوت و ضعف یادگیرندگان، آن‌گونه که در روش تحلیلی امکان‌پذیر است، در اختیار آنان قرار نمی‌دهد" (ص ۲۷۱).

در روش کلی یا سراسری نمره گذاری، کل فرایند یا فراورده یکباره مورد داوری قرار می‌گیرد و به اجزاء یا قسمتهای مختلف آن نمره یا امتیاز جداگانه داده نمی‌شود. جدول ۳ـ۱۲ نمونه‌ای از یک دستورالعمل یا راهنمای تصحیح با روش کلی را نشان می‌دهد. در این نمونه مهارت یادگیرنده در بیان شفاهی یک زبان خارجی ارزشیابی می‌شود.

در تعیین راهنما یا دستورالعمل برای نمره گذاری پاسخهای آزمون‌شوندگان، همچنین مشخص کنید که به آنان به صورت فردی نمره خواهید داد یا به صورت گروهی. استفاده از

جدول ۲ـ۱۲ مثالهایی از راهنمای نمره‌گذاری با روش تحلیلی (پین، ۲۰۰۳، ص ۲۹۳)

ریاضیات	نوشتن	هنر
حل مسئله	**هدفها**	**عناصر صوری**
• راهبردها:	ـ شرح دادن	ـ ساختار
ـ تشخیص مسئله	ـ ترغیب کردن	ـ ترکیب
ـ به کار بستن راهبردها، استفاده از مفاهیم، روشها، ابزارها	ـ اطلاع دادن	**عناصر فنی**
	جنبه‌ها	ـ تکنیکها
بازنمایی	ـ ترکیب	ـ مواد
ـ چارت	ـ تأکید	
ـ نمودار	ـ تفصیل	**عناصر حسی**
	ـ سازمان	• بیانی:
استدلال	ـ امور قراردادی	ـ حالت
ـ تفسیر		ـ کیفیت هیجانی
ـ تعمیم		ـ کیفیت انرژی
		عناصر هویتی
ارتباط		• انسجام عناصر
ـ آشکار		• تأثیر عناصر
ـ سازمان یافته		
ـ کامل		
ـ مفصل		
ـ اصطلاحات، نمادها، و علائم زبان ریاضی		
محتوا		
ـ مفاهیم عددی		
ـ نسبت / درصد		
ـ مفاهیم جبری		
ـ مفاهیم هنری		

جدول ۳ـ۱۲ مثالی از روش کلی سنجش عملکرد (سنجش مهارت شفاهی در زبان دوم) (نیتکو، ۲۰۰۱، ص ۲۵۹)

سطح	وصف
عالی	**سطح عالی با توانایی یادگیرنده (گوینده) در ویژگیهای زیر تعیین می‌شود:** ● شرکت مؤثر در اکثریت مکالمه‌های رسمی و غیررسمی دربارهٔ موضوعهای عملی، اجتماعی، حرفه‌ای، و انتزاعی. ● جانبداری از عقاید و فرضیه‌ها با استفاده از راهبردهای گفتنی.
پیشرفته	**سطح پیشرفته با توانایی یادگیرنده (گوینده) در ویژگیهای زیر تعیین می‌شود:** ● انجام گفتگو به شکلی روشن و مشارکتی. ● آغاز کردن، نگهداری، و به نتیجه رسانیدن انواع مختلف تکالیف ارتباطی، از جمله آنهایی که نیازمند توانایی سطح بالایی در انتقال معنی به کمک راهبردهای متنوع زبانی هستند. ● پاسخگویی به خواسته‌های مدرسه. ● بیان و توصیف مطالب در قالب پاراگرافهای منسجم.
متوسط	**سطح متوسط باتوانایی یادگیرنده (گوینده) درویژگیهای زیر تعیین می‌شود:** ● استفاده از عناصر زبانی آموخته شده در ترکیبهای تازه‌که عمدتاً حالت واکنشی داشته باشند. ● آغاز کردن، تا حدی نگهداری کردن، و به نتیجه رساندن تکالیف سادهٔ ارتباطی. ● پرسیدن سؤال و پاسخ دادن به سؤالها.
ابتدایی	**سطح ابتدایی با توانایی یادگیرنده (گوینده) در ایجاد حداقلی از ارتباط با استفاده از مطالبی که آموخته است مشخص می‌شود.**

سنجش گروهی برای ارزشیابی از عملکرد دانش‌آموزان در موقعیتهای یادگیری مشارکتی مناسب است. در این ارزشیابی هم می‌توان سهم هر یک از اعضای گروه را تعیین کرد و هم فراورده یا محصول کارگروهی را سنجید.

نکتهٔ دیگری که در اینجا لازم است به آن اشاره کنیم این است که از بعضی فنون سنجش عملکردی به این منظور که یادگیرندگان در مورد عملکرد خودشان داوری نمایند نیز می‌توان استفاده کرد. یعنی اینکه سنجش فعالیتهای یادگیرندگان به صورت خودسنجی[1] صورت پذیرد. سانتروک (۲۰۰۴) در این باره گفته است، "می‌توان از دانش‌آموزان و دانشجویان خواست تا دربارهٔ کیفیت گزارش شفاهی خودشان یا بازیهای نمایشی‌شان داوری کنند" (ص ۵۳۸).

1. self-assessment

۶. از روشها و فنون گوناگون مشاهد استفاده نمایید. چنان‌که قبلاً در آغاز این فصل توضیح دادیم، ارزشیابی از عملکرد بر فرایند، فراورده، یا ترکیبی از این دو تأکید می‌کند. فرایندها و فراورده‌های عملکرد معمولاً به وسیلهٔ روشها و فنون مختلف مشاهده، نظیر فهرست وارسی، مقیاس درجه‌بندی، و نظایر آنها سنجیده می‌شوند. بنابراین، در سنجش عملکرد یادگیرندگان از این وسایل اندازه‌گیری حداکثر استفاده را بنمایید. ما در فصل ۱۴ روشهای مختلف سنجش مشاهده‌ای را معرفی خواهیم کرد.

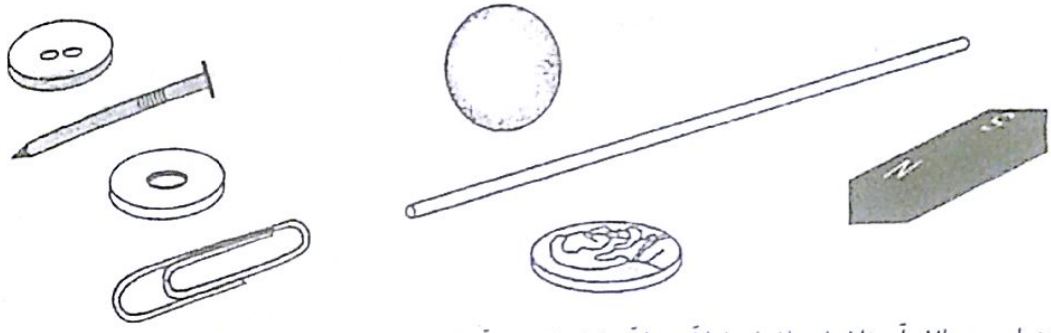
یک نمونه آزمون عملکردی همراه با راهنمای انجام عملکرد و سنجش پاسخ

در شکل ۲ـ۱۲ یک تکلیف مربوط به سنجش عملکردی در علوم داده شده. در این تکلیف دانش‌آموزان باید یک آزمایش انجام دهند و اشیاء مورد آزمایش را طبقه‌بندی نمایند. آنها همچنین باید برای طبقه‌بندی خود دلیل بیاورند. اما تکلیف استفاده از آهنربا برای آزمایش‌کردن اشیاء، طبقه‌بندی آنها به دو دستهٔ مختلف، و بیان دلایل تفاوت بین اشیاء در دو طبقهٔ مورد نظر توضیح داده شده.

آهنربا

توضیح تکلیف: استفاده از آهنربا برای تشخیص اشیاء مغناطیسی از غیرمغناطیسی و تبیین تفاوت بین آنها.

وسایل و مواد: یک آهنربا و چند شی، زیر: دکمهٔ پلاستیکی، واشر پلاستیکی یا آهنی، گیرهٔ کاغذ از جنس فلز، میخ آهنی، گلولهٔ شیشه‌ای، مبلهٔ پلاستیکی، و سکهٔ مسی.



راهنمایی به دانش‌آموزان: این اشیاء را با آهنربا آزمایش کنید و آنها را به دو دسته تقسیم نمایید. اشیاء هر دو دسته را فهرست کنید و دلیل بیاورید که چه عاملی موجب تفاوت بین آنها و دسته‌بندی متفاوت شما شده است.

طرح نمره‌گذاری: برای دسته‌بندی اشیاء به طور صحیح نمره داده می‌شود. استدلالهای زیر برای دسته‌بندی اشیاء پذیرفته می‌شوند: (۱) اشیاء یک گروه از فلز یا آهن درست شده؛ (۲) اشیاء یک گروه به وسیلهٔ آهنربا جذب می‌شوند، (۳) اشیاء یک گروه از آهن یا فلز درست شده و به وسیلهٔ آهنربا جذب می‌شود، و هرگونه استدلال درست دیگر.

شکل ۲ـ۱۲ مثالی از آزمون عملکردی در علوم (لین و گرانلاند، ۲۰۰۰، ص ۲۶۹)

امتیازها و محدودیتهای روشهای سنجش عملکردی

یکی از امتیازهای سنجش عملکردی یا آزمون عملکردی این است که به وسیلهٔ آن می‌توان یادگیریهایی را سنجید که به کمک آزمونهای کتبی یا مداد و کاغذی مرسوم میسر نیست. چنان‌که قبلاً اشاره کردیم، دانستن چگونه انجام دادن یک کار با توانایی انجام دادن آن فرق می‌کند. به عنوان مثال، "یک آزمون کتبی که می‌تواند دانش یادگیرنده را دربارهٔ ویژگیهای مؤثر ایراد سخنرانی در حضور جمع بسنجد دربارهٔ اینکه آیا او می‌تواند در حضور جمع سخنرانی کند هیچ‌گونه اطلاعی به ما نمی‌دهد" (لین و گرانلاند، ۲۰۰۰، ص ۱۱).

دومین امتیاز سنجش عملکردی که آن هم قبلاً مورد اشاره واقع شد این است که در این روش هم فراوردهٔ عمل یادگیرنده و هم فرایند یا جریان کار او سنجیده می‌شود. به عنوان نمونه، وقتی که دانش‌آموزان مشغول انجام دادن یک آزمایش علمی هستند، به کمک سنجش عملکردی، هم می‌توان نقاط قوت و ضعفشان را در استفاده از ابزارها و فنون آزمایش سنجش کرد (سنجش فرایند) و هم می‌توان توفیقشان را در کامل کردن آزمایش و مستدل ساختن نتیجه گیریهای خود مورد ارزشیابی قرار داد (سنجش فراورده).

سومین امتیازی که می‌توان برای سنجش عملکردی ذکر کرد این است که در این نوع سنجش بر نظریه‌های جدید یادگیری تأکید می‌شود. بنابه نظریه‌های جدید یادگیری، از جمله نظریه‌های سازندگی یا سازنده‌گرایی[1]، نه تنها یادگیرندگان دریافت‌کنندگان محض اطلاعات به حساب نمی‌آیند، بلکه چنین فرض می‌شود که آنان در ساختن و کشف دانش نقش مهمی برعهده دارند و فعالیتشان در جریان یادگیری تضمین‌کنندهٔ توفیق یادگیری آنان است (سیف، ۱۳۸۰). طبق این نظریه‌ها، اطلاعات جدید باید با اطلاعات قبلاً آموخته شدهٔ یادگیرندگان در هم آمیزد. "سنجش عملکردیِ دارای کیفیت سطح بالا دانش قبلی را مورد توجه قرار می‌دهد و یادگیرنده را در ساختن فعال معنی درگیر می‌سازد" (لین و گرانلاند، ۲۰۰۰، ص ۲۶۷).

در مقابل امتیازهای بالا، سنجش عملکردی دارای محدودیتهایی است که به آنها نیز اشاره می‌کنیم. مهم‌ترین محدودیتی که برای این روش گفته شد همان مشکل مربوط به آزمونهای تشریحی است که در فصل ۱۰ گفته شد، یعنی کمبود پایایی یا دقت نمره‌گذاری. بنابراین، رعایت پیشنهادهایی که در آن فصل در رابطه با آزمونهای تشریحی گفتیم می‌تواند دقت روشهای سنجش عملکردی را نیز افزایش دهد. به این منظور، معلمان باید بکوشند تا بازده‌های مورد سنجش را به روشنی تعریف کنند، تکالیفی را که در آن توانایی یادگیرنده قرار است سنجش شود انتخاب نمایند، و دستورالعمل یا راهنمای نمره‌گذاری را به طور کامل مشخص کنند و به اجرا درآورند.

محدودیت دیگر روشهای سنجش عملکردی صرف وقت زیاد برای آنهاست. از آنجاکه در

1. constructivism

سنجش عملکردی برای هر هدف یادگیری یک سنجش مستقل باید صورت پذیرد و هر نوبت سنجش به وقت قابل توجهی نیاز دارد، سنجش کامل عملکردهای یادگیرنده در ارتباط با یک موضوع درسی مستلزم صرف وقت زیادی است. برای رفع این مشکل، لازم است در طول ترم تحصیلی سنجش عملکردهای دانش‌آموزان به دفعات مختلف انجام شود، و کوشش به عمل آید تا هر تکلیف سنجش همراه با یادگیری صورت پذیرد. یعنی معلم، ضمن سنجش، آموزش هم بدهد تا از وقت کلاس به نحو بهتری استفاده نماید.

موارد استفادهٔ روشهای سنجش عملکردی

چنان‌که دیدیم، موارد استفادهٔ روشهای سنجش عملکردی بسیارند، و تقریباً همهٔ هدفهای حوزه‌های شناختی، عاطفی، و روانی ـ حرکتی را دربرمی‌گیرند. در موقعیتهایی که روشهای سنتی مداد و کاغذی کارساز نیستند، روشها و فنون سنجش عملکردی قابل استفاده‌اند ـ به ویژه در شرایطی که هم فراورده‌های رفتار و عملکرد یادگیرنده موردنظر است و هم فرایندهای آن. ما در قسمت باقیماندهٔ این فصل نمونه‌هایی از کاربردهای واقعی این روشها را معرفی می‌کنیم.

پیش از معرفی نمونه‌های سنجش عملکردی، لازم است یادآوری کنیم که استفاده از روشها و آزمونهای عملکردی برای سنجش هدفهای سطح پایین، یعنی یادآوری دانش و محفوظات، مفید نیست و به این منظور باید از آزمونها و روشهای نوع دیگری استفاده شود. علاوه بر این، هرچند که روشها و فنون سنجش عملکردی برای سنجش و ارزشیابی یادگیری دانش‌آموزان و دانشجویان در کلاس درس، به ویژه به عنوان یک روش سنجش همراه با یادگیری، بسیار مفید است، اما استفاده از آن برای مقاصد هنجاری توصیه نشده (به عنوان نمونه، پوفام (۲۰۰۲) گفته است استفادهٔ معلمان از روش سنجش عملکردی در سطح کلاس درس بسیار مؤثر است و موجب بهبود کیفیت آموزش معلمان می‌شود، اما استفاده از آن برای مقاصد پاسخگویی یا مسئولیت‌پذیری[1] ـ یعنی تعیین اینکه معلمان یک آموزشگاه یا ناحیه یا استان در کار آموزشی خود تا چه اندازه موفق‌اند ـ چندان جایز نیست.)

مثالی از سنجش عملکرد: سنجش هوشهای چندگانه

هوارد گاردنر[2] (۱۹۸۳، ۱۹۹۸)، یکی از نظریه‌پردازان هوش، نظریه‌ای وضع کرده شامل هشت نوع هوش متفاوت و مستقل از یکدیگر با نامهای زیر: زبانی[3]، منطقی ـ ریاضی[4]، فضایی[5]، بدنی ـ جنبشی[6]، موسیقایی[7]، میان فردی[8]، درون فردی[9]، و طبیعت گرایانه[10]. توصیف مختصر هر یک از این هوشها در جدول ۴ـ۱۲ آمده است.

---

۱. این مفهوم در فصل اول توضیح داده شد.

2. Howard Gardner 3. linguistic 4. logical-mathematical 5. spatial
6. bodily-kinesthetic 7. musical 8. interpersonal 9. intrapersonal 10. naturalistic

جدول ۴ـ۱۲ توصیف مختصر هوشهای هشت‌گانهٔ گاردنر (اقتباس براساس نیتکو، ۲۰۰۱، ص ۲۴۵)

نوع هوش	توصیف
هوش زبانی	توانایی استفاده از زبان خودتان و دیگر زبانها برای فهماندن مقاصد خود به دیگران و درک مقاصد آنان.
هوش منطقی ـ ریاضی	توانایی درک اصول علت و معلولی به گونه‌ای شبیه به درک دانشمندان یا منطق دانان، یا استفاده از استدلال کمّی و ریاضی به گونه‌ای شبیه به استفادهٔ ریاضی‌دانان.
هوش فضایی	توانایی بازنمایی جنبه‌های فضایی دنیا در ذهن خود، مانند خلبانان، بازیکنان شطرنج، نقاشان، مجسمه‌سازان، یا آرشیتکت‌ها.
هوش بدنی ـ جنبشی	توانایی استفاده از قسمتی یا تمامی بدن خود برای حل یک مسئله، تولید یک فراورده، یا عمل کردن شبیه به ورزشکاران یا بازیگران.
هوش موسیقایی	توانایی پردازش ذهنی موسیقی به گونه‌ای که موجب شناسایی الگوها، به یادسپردن الگوها، و دستکاری موسیقی برای حل مسائل یا بیان درک و فهم بشود.
هوش میان‌فردی یا بین افراد	توانایی درک افراد دیگر و برقرار کردن رابطهٔ معنی‌دار با آنان. این کار از راه درک اشخاص دیگر امکان‌پذیر است. یعنی چه کاری می‌توانند انجام دهند، چگونه با دنیا و دیگران رابطه برقرار می‌کنند، واکنش احتمالی آنان چیست، چه چیزی را دوست دارند، از چه چیزی اجتناب می‌کنند، و چه احساسی دارند.
هوش درون‌فردی	توانایی درک خودتان، دانستن اینکه چه کسی هستید، تواناییها و ناتوانایی‌های شما در چیست، آرزوها و هدفهایتان کدام‌اند، چه احساسی دارید، از چه چیزی اجتناب می‌کنید، و در شرایط مختلف چگونه واکنش می‌کنید.
هوش طبیعت‌گرایانه	توانایی درک طبیعت و دنیای مدرن از راه تمیز قائل شدن بین موجودات طبیعی زنده و غیرزنده و اشیاء ساخته شدهٔ دست انسان و دسته‌بندی کردن آنها.

به باور گاردنر، همهٔ افراد از این هشت نوع هوش بهره‌مندند، اما در بعضی از آنان بعضی از این هوشها قوی‌تر و بعضی ضعیف‌ترند. گاردنر در توجیه نظریهٔ خود و وجود این هوشها دلایل زیر را ذکر کرده است: (۱) این هوشها برای بقای ما ضروری‌اند؛ (۲) اینها دارای ویژگیهای تکاملی هستند (تا حدودی در جانداران غیرانسان نیز یافت می‌شوند)؛ و (۳) به وسیلهٔ فعالیتهای مغزی ویژه قابل شناسایی‌اند.

نیتکو (۲۰۰۱) گفته است در بعضی از مدارس آمریکا معلمان فعالیتهای آموزشی خود را براساس هوشهای هشت‌گانهٔ گاردنر سازمان می‌دهند. وی همچنین می‌گوید این نظریه الهام‌بخش یک فلسفهٔ تربیتی و فعالیتهای کلاسی متعددی بوده است که هدف آنها آموزش دادن موارد زیر به دانش‌آموزان است:

۱. آگاه باشند که در راههای مختلف می‌توانند هوشمند شوند.
۲. هوشهای مختلف را در خود و دیگران ارج بگذارند.
۳. هوشهای مختلف خود را به کار بندند، تمرین دهند، و تا حد توان خود در پرورش آنها بکوشند.
۴. پیشرفت خود را در هدفهای آموزشی با چندتا از هوشهای خود نشان دهند.

سنجش هوشهای چندگانه

مدافعان نظریهٔ هوشهای هشت‌گانهٔ گاردنر کوشیده‌اند تا روش سنجش عملکردی را برای اندازه‌گیری آنها به کار بندند. به باور این افراد، روشهای سنجش عملکردی می‌تواند امکاناتی فراهم آورد تا پیشرفت دانش‌آموزان به چندین راه مختلف مورد شناسایی و ارزشیابی قرار گیرد. گاردنر در رابطه با سنجش عملکردی هوشهای مختلف گفته است:

> تأکید جاری بر سنجش عملکردی با نظریهٔ هوشهای چندگانه به خوبی حمایت می‌شود.... نخست اینکه، اجازه بدهید از صافی آزمون کوته‌پاسخ به چیزها نگاه نکنیم. در عوض، اجازه بدهید تا به عملکردی که برایش ارزش قائل هستیم نگاه کنیم ــ چه آن عملکرد زبانی باشد، چه منطقی، زیبایی شناختی، یا اجتماعی. دوم اینکه، هرگز اجازه ندهید که سنجش خود را از درک و فهم دانش‌آموز صرفاً بر یک اندازهٔ خاص متمرکز کنیم، بلکه اجازه بدهید تا به دانش‌آموزان فرصت دهیم تا درک و فهم خود را در چند راه مختلف نشان دهند. (گاردنر، به نقل نیتکو، ۲۰۰۱، ص ۲۴۶)

بنا به گفتهٔ نیتکو (۲۰۰۱)، استفاده از روشهای سنجش عملکردی برای اندازه‌گیری هوشهای چندگانه مستلزم این است که استانداردهای سطح بالایی برای توفیق یادگیرنده در هدفهای یادگیری منظور شود و از او خواسته شود تا آن موفقیت را به راههای مختلفی نشان دهد. در جدول ۵ـ۱۲، برای هر یک از هوشهای گاردنر، تعدادی عملکرد نمونه و مثالهایی از سنجش عملکرد آمده است.

جدول ۵ـ۱۲ سنجش عملکردی براساس نظریهٔ هوشهای هشت‌گانهٔ گاردنر (اقتباس براساس نیتکو، ۲۰۰۱، ص ۲۴۶). به دانش‌آموزان اجازه داده می‌شود تا با انتخاب یک الگوی سنجش از میان یکی یا چند تا از طبقه‌های هوشی پیشرفت خود را نشان دهند.

هوش زبانی
- یک گفتگوی رادیویی دربارهٔ ـــــ تولید کنید.
- دربارهٔ ـــــ بحث کنید.
- دربارهٔ ـــــ یک شعر، قصه، یا مقاله بنویسید.

هوش منطقی - ریاضی
- مطلب ـــــ را به یک فرمول ریاضی تبدیل کنید.
- یک آزمایش علمی دربارهٔ ـــــ طرح و اجرا کنید.
- با استدلال منطقی ـــــ را توضیح دهید.

هوش بدنی - جنبشی
- با یک حرکت یا تعدادی حرکت ـــــ را نشان دهید.
- یک ـــــ بسازید.
- یک برنامهٔ گردش علمی به منظور ـــــ طراحی کنید و در آن شرکت نمایید.

هوش فضایی (دیداری)
- برای ـــــ یک نمودار، نقشه، یا طرح بکشید.
- یک اثر هنری تولید کنید که ـــــ را نشان دهد.
- یک بازی با چوب یا ورق ـــــ خلق کنید.

هوش موسیقایی
- آهنگی دربارهٔ ـــــ بخوانید.
- توضیح دهید که چگونه موسیقی یک آهنگ شبیه ـــــ است.
- یک ابزار موسیقی بسازید که ـــــ را نشان دهد.

هوش بین‌فردی
- به کسی دربارهٔ ـــــ آموزش دهید.
- دربارهٔ موضوع ـــــ یک سخنرانی ایراد کنید.
- از تکنولوژی به منظور ـــــ استفاده کنید.

هوش درون‌فردی
- دربارهٔ ـــــ برای خود هدفی برگزینید و برای رسیدن به آن بکوشید.
- یکی از ارزشهای شخصی خود را دربارهٔ ـــــ توضیح دهید.
- فعالیت خود را در رابطه با ـــــ بسنجید.

هوش طبیعت‌گرایانه
- تغییرات ایجاد شده در محیط زندگی خود را توضیح دهید.
- توضیح دهید که از حیوانهای خانگی، حیوانهای وحشی، باغچه، چگونه مواظبت به عمل می‌آورید؟
- اشیاء طبیعی را نقاشی کنید یا از آنها عکس بگیرید.

مثالی از سنجش عملکردی: حل مسائل زندگی واقعی

در تکلیف زیر از دانش‌آموزان خواسته می‌شود تا دربارهٔ یکی از موقعیتهای واقعی زندگی به جمع‌آوری اطلاعات لازم بپردازند و آنچه را که به صورت یک مسئله به آنان ارائه شده است حل نمایند. این سؤال عملکردی از مجموعه آزمونهای مبانی یادگیری عمومی ایالت کانکتیکت آمریکا برگزیده شده (به نقلِ وولفلک، ۲۰۰۴، ص ۵۵۵).

سؤال: بسیاری از فروشگاههای محلی ادعا می‌کنند که کالاهای خود را ارزان می‌فروشند. اما واقعاً این ادعا به معنی می‌دهد؟ آیا معنی ادعای آنان این است که همهٔ اجناس فروشگاههایشان دارای قیمت پایین‌تر است یا فقط بعضی از آنها؟ چگونه می‌توانید تعیین کنید که خرید از کدام فروشگاه بیشتر به نفع شماست؟ تکلیف شما این است که پژوهشی را طراحی و اجرا کنید که به سؤالهای بالا جواب دهد. چه اجناس و قیمتهایی را با هم مقایسه خواهید کرد و چرا؟ نمونه‌ای را که انتخاب خواهید کرد چگونه توجیه می‌کنید و اعتبار آن را چگونه می‌سنجید؟

خلاصهٔ فصل

سنجش عملکردی مجموعه روشها و فنونی را دربرمی‌گیرد که فرایندها و فراورده‌های یادگیری را می‌سنجند و، علاوه بر فعالیتهای شناختی سطح بالا، یادگیریهای حوزه‌های عاطفی و روانی ـ حرکتی را نیز شامل می‌شود. سنجش عملکردی همواره از فعالیتهای مهم معلمان به حساب می‌آمده، اما دو عامل بر گسترش شهرت و کاربرد آن در سالهای اخیر مؤثر بوده‌اند: (۱) نارضایتی پرورشکاران و متخصصان سنجش یادگیری از آزمونهای عینی، و (۲) گسترش روانشناسیهای شناختی و سازندگی.

روشهای سنجش عملکردی به صورتهای مختلفی دسته‌بندی شده‌اند. ما در این فصل دسته‌بندی‌های گرانلاند و نیتکو را از روشها و فنون سنجش عملکردی توضیح داده‌ایم. گرانلاند چهار دسته روش سنجش عملکردی را با نامهای زیر به دست داده است: (۱) آزمون کتبی عملکردی، (۲) آزمون شناسایی، (۳) انجام عملکرد در موقعیتهای شبیه‌سازی شده، و (۴) نمونهٔ کار. نیتکو نیز روشها و فنون سنجش عملکردی را در چهار دستهٔ زیر توضیح داده است: (۱) سنجش ساختارمند یا کنترل شده، (۲) سنجش در موقعیتهای طبیعی، (۳) پروژه‌های طولانی مدت، و (۴) کارپوشه.

صاحب‌نظران سنجش و اندازه‌گیری برای تهیهٔ آزمونهای عملکردی و اجرا و نمره‌گذاری درست آنها پیشنهادهایی ارائه داده‌اند. ما در این فصل آن پیشنهادها را توضیح داده‌ایم.

مانند دیگر روشهای سنجش یادگیری، روشهای سنجش عملکردی نیز دارای امتیازها و محدودیتهایی هستند که ما مهم‌ترین آنها را در این فصل مورد بحث قرار داده‌ایم. امتیازهای مهم سنجش عملکردی به قرار زیرند:

۱. با این روشها می‌توان هدفهایی را سنجید که با سایر روشهای سنجش (به ویژه روشهای مداد و کاغذی) قابل سنجش نیستند.
۲. با روشهای سنجش عملکردی هم می‌توان فراورده‌های یادگیری را سنجش کرد و هم فرایندهای آن را. این در حالی است که با سایر روشهای سنجش (مخصوصاً روشهای عینی) تنها می‌توان فراورده را سنجش کرد.
۳. روشهای سنجش عملکردی منعکس‌کنندهٔ نظریه‌های جدید روانشناسی یادگیری، عمدتاً نظریه‌های خبرپردازی و ساختن‌گرایی، هستند.

در مقابل امتیازهای بالا، برای روشهای سنجش عملکردی محدودیتهایی قائل شده‌اند. از جملهٔ آنها می‌توان به دو مورد زیر اشاره کرد:

۱. کمبود پایایی یا دقّت اندازه‌گیری.
۲. صرف وقت زیاد برای اجرا و نمره‌گذاری.

موارد استفادهٔ روشهای سنجش عملکردی فراوان‌اند. از جمله هم می‌توان به فعالیتهای شناختی سطح بالاتر و مهارتهای عملی اشاره کرد و هم به هدفهای مختلف حوزه‌های عاطفی و روانی ـ حرکتی. با این حال، باید دقت کرد که برای سنجش هدفهای سطح پایین، مانند یادآوری دانش و محفوظات، و نیز برای مقاصد ارزشیابی هنجاری از این روشها استفاده نشود.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. با روشهای سنجش عملکردی فرایند یادگیری را بهتر از فراوردهٔ آن می‌توان سنجش کرد.
الف ـ صحیح
ب ـ غلط

۲. با سنجش عملکردی هم می‌توان مهارتهای عملی یادگیرندگان را سنجش کرد و هم مهارتهای فکری آنان را.
الف ـ صحیح
ب ـ غلط

۳. نزدیک‌ترین آزمون عملکردی به عملکرد واقعی یادگیرنده در شرایط طبیعی محیط زندگی کدام است؟
الف ـ آزمون کتبی عملکردی
ب ـ نمونهٔ کار
پ ـ آزمون شناسایی

ب ـ برای مقاصد ارزشیابی هنجاری مفید نیستند.

پ ـ تنها هدفهای آموزشی حوزهٔ شناختی را می‌سنجند.

ت ـ الف و ب

ث ـ الف و ب و ج

۵. سنجش عملکردی را تعریف کنید.

۶. تفاوتهای مهم بین آزمونهای عملکردی و آزمونهای کتبی (مداد و کاغذی) را توضیح دهید.

۷. تقسیم‌بندی گرانلاند و نیتکو از روشهای سنجش عملکردی را شرح دهید و آنها را با هم مقایسه کنید.

۸. قواعد تهیهٔ آزمونهای عملکردی را توضیح دهید.

۹. روشهای مختلف نمره‌گذاری آزمونهای عملکردی را توضیح دهید.

۱۰. امتیازها و محدودیتهای روشهای سنجش عملکردی را توضیح دهید.

۱۱. موارد استفادهٔ آزمونهای عملکردی را توضیح دهید.

۱۲. برای طرح سنجش خود که به عنوان تمرین فصل ۶ تهیه کرده‌اید، یک آزمون عملکردی متناسب با هدفهایی از آن درس که با این نوع آزمونها بیشتر سنجیده می‌شوند بسازید.

۱۳

سنجش با استفاده از کارپوشه

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. روش سنجش با استفاده از کارپوشه را تعریف کند.

۲. روش کارپوشه و روش سنجش عملکردی را با هم مقایسه کند.

۳. هدفهای مختلف تهیهٔ کارپوشه را توضیح دهد.

۴. تفاوت بین کارپوشه‌های آموزشی و سنجشی را توضیح دهد.

۵. انواع کارپوشه‌های سنجشی را توضیح دهد.

۶. کارپوشهٔ الکترونیکی را تعریف کند و ویژگیهای آن را توضیح دهد.

۷. محتوای کارپوشه‌های مختلف را با توجه به هدفهای آنها توضیح دهد.

۸. قواعد تهیه و استفاده از کارپوشه را توضیح دهد.

۹. روشهای نمره گذاری کارپوشه را توضیح دهد.

۱۰. امتیازها و محدودیتهای روش سنجش به کمک کارپوشه را توضیح دهد.

۱۱. برای نشان دادن فعالیتها و موفقیتهای خود در شغل معلمی با جمع‌آوری و تهیهٔ مدارک و اسناد ضروری یک کارپوشهٔ حرفه‌ای درست کند.

ما روشهای مختلف سنجش عملکردی را در فصل ۱۲ توضیح دادیم. موضوع مورد بحث در فصل حاضر یعنی سنجش به کمک کارپوشه یکی از روشهای سنجش عملکردی است، اما از آنجا که کارپوشه یک روش مهم سنجش عملکرد است و به تازگی شهرت زیادی کسب کرده این فصل را به طور کامل به آن اختصاص داده‌ایم. لین و گرانلاند (۲۰۰۰) دربارهٔ اهمیت این روش و پذیرش روزافزون آن توسط معلمان و سنجشگران گفته‌اند:

> کارپوشه را معلمان مدارس به طور مرتب در آموزش خود و سنجش یادگیری دانش‌آموزانشان مورد استفاده قرار می‌دهند. همچنین از کارپوشه برای تعیین نمره‌های دانش‌آموزان و گزارش پیشرفت آنان به والدین استفاده می‌شود.... در بعضی موارد کارپوشه به عنوان شرایط فارغ‌التحصیل شدن از دبیرستان نیز مورد استفاده قرار می‌گیرد. همچنین گاهی از کارپوشه به عنوان بخشی از مدارک مورد نیاز پذیرش فارغ‌التحصیلان دبیرستان به دانشگاه و کسب شغل استفاده می‌شود. (ص ۲۸۹)

تعریف کارپوشه

بنابر تعریف، **کارپوشه**[1] "به مجموعه‌ای از کارهای یادگیرنده گفته می‌شود که نشان‌دهندهٔ کوشش، پیشرفت، و موفقیت تحصیلی او در یک زمینهٔ خاص است" (مرفی[2] و دیوید شوفر[3]، ۲۰۰۱، ص ۴۱۵). در روش کارپوشه، یادگیرنده برای نشان دادن تواناییها و پیشرفت خود نمونه‌هایی از کارهایش را که در طول یک مدت معین، مانند ثلث، نیمسال یا سال تحصیلی، انجام داده است با خود به کلاس می‌آورد و برای ارزشیابی در اختیار معلم می‌گذارد. اِگِن و کاوچاک (۲۰۰۱) در رابطه با این روش سنجش گفته‌اند "از آنجا که مطالب موجود در کارپوشه جنبهٔ تراکمی دارند و به هم مربوط‌اند و در طول یک دوره از زمان انجام گرفته‌اند ــ برخلاف عکسهای جدا از یکدیگری که به وسیلهٔ آزمونهای هفتگی و ترمی فراهم می‌آیند ــ می‌توانند یک 'فیلم' از چگونگی پیشرفت یادگیرنده فراهم آورند" (ص ۶۱۵). کوبیسزین و بوریچ (۲۰۰۳) در مقایسهٔ روش سنجش عملکردی با روش سنجش به کمک کارپوشه گفته‌اند:

> سنجش عملکردی نوعی نمایش است که در آن یادگیرنده درک عمیق خود را از یک زمینهٔ معین از یادگیری نشان می‌دهد. این نمایش مانند عکسی است که نشان می‌دهد یادگیرنده در یک زمان معین در سال تحصیلی چه چیزی به دست آورده است. اما نوع دیگری از سنجش عملکردی وجود دارد که چیزی بیشتر از تصویر یک مقطع زمانی خاص از یادگیریهای یادگیرنده را نشان می‌دهد. هدف اصلی

---

۱. portfolio (در اصطلاح لغوی، این کلمه به یک کیف چرمی پَهن گفته می‌شود که برای حمل برگه‌های کاغذ، صفحات پیش‌نویس یک نمایش‌نامه، یا مجموعه نقاشیهای یک فرد به کار می‌رود. این اصطلاح از آنجا در آموزش به عنوان یک روش سنجش عملکردی به کار رفته که در آن دانش‌آموز مجموعه‌ای از کارهای خود را که در طول یک سال یا نیمسال تحصیلی انجام داده است، معمولاً در داخل یک پوشه، برای نشان دادن توانایی‌اش به معلم یا ارزشیاب ارائه می‌دهد.)

2. Murphy 3. Davidshofer

این نوع سنجش آن است که دربارهٔ رشد یادگیری دانش‌آموز در طول یک دورهٔ زمانی و توفیق درازمدت او برای ما داستانی بگوید. این روش سنجش **کارپوشه** نام دارد. کارپوشه، مانند سنجش عملکردی، درک عمیق یادگیرنده را اندازه می‌گیرد، اما علاوه بر آن رشد شایستگی و درک او را در طول ترم یا سال تحصیلی نیز نشان می‌دهد. (ص ۱۷۴)

کوبیسزین و بوریچ (۲۰۰۳) در تعریفی که از این روش سنجش به دست داده گفته‌اند "کارپوشه مجموعه‌ای طرح‌ریزی شده و هدفمند از مدارک و شواهدی است که چگونگی پیشرفت یادگیرنده و گامهایی را که برای رسیدن به آن برداشته است شامل می‌شود. مجموعهٔ مورد نظر کوششهای مشترک معلم و یادگیرنده را دربارهٔ تصمیمهایی که گرفته شده، محتوای کارپوشه، و معیارهای ارزشیابی نشان می‌دهد" (ص ۱۷۴).

دقت کنید که کارپوشهٔ دانش‌آموز حاوی منتخبی از کارهایی است که او در یک مدت معین انجام داده است نه تمامی کارهای او. نیتکو (۲۰۰۱) در معرفی این روش گفته است:

> برای مقاصد سنجش، یک کارپوشه مجموعهٔ محدودی از کارهای دانش‌آموز است که به منظور معرفی بهترین کارها یا نشان دادن رشد تحصیلی او در یک مدت معین مورد استفاده قرار می‌گیرد. کارپوشه شامل همهٔ کارهای دانش‌آموز نیست. کارهایی که در کارپوشه قرار می‌گیرند به آن کارهایی که بهتر از بقیه هدف کارپوشه را برآورده می‌سازند محدود می‌شود.... موادی که در کارپوشه قرار می‌گیرند به طور دقیق و به طور عمدی به گونه‌ای انتخاب می‌شوند که مجموعهٔ کارهای انتخابی در کُل هدف مورد نظر را تحقق می‌بخشند. (ص ۲۵۴)

به طور خلاصه، کارپوشهٔ دانش‌آموز را یا مواردی از بهترین کارهای او تشکیل می‌دهد یا مثالهایی از انواع کارهای انجام شده توسط او (مانند گزارش کتاب، نامه‌ای به یک دوست، یک داستان کوتاه ابتکاری و یک مقالهٔ تحقیقی). به سخن دیگر، کارهایی که برای گنجاندن در کارپوشه انتخاب می‌شوند باید به دقت انتخاب شوند تا هدف مورد نظر از تشکیل کارپوشه را برآورده سازند.

هدفهای تهیهٔ کارپوشه

پیش از آنکه چگونگی تهیهٔ کارپوشه و استفاده از آن را مورد بحث قرار دهیم، لازم است هدفهایی را که کارپوشه برای تحقق آنها تشکیل می‌شود توضیح دهیم، زیرا هدف است که به روند تولید و نحوهٔ استفاده از همهٔ روشهای سنجش از جمله سنجش با استفاده از کارپوشه جهت می‌دهد. باتلر و مکمان (۲۰۰۶) در این‌باره می‌گویند "کارپوشه‌ها اغلب با هدف یا قصد زیربنایی مجموعه کارهایی که جمع‌آوری می‌شوند تعریف می‌شوند" (ص ۶۶). از کارپوشه برای برآوردن مقصودهای مختلفی استفاده می‌شود که ما مهم‌ترین آنها را در اینجا معرفی می‌کنیم[1].

---

۱. منابع اصلی مورد استفاده در تهیهٔ مطالب این بخش کتابهای زیر بوده‌اند: ایجاد کارپوشه برای یادگیری و سنجش، تألیف کلنوسکی، ۲۰۰۲، و اندازه‌گیری و سنجش در آموزش، تألیف لین و گرانلاند، ۲۰۰۰

استفاده از کارپوشه برای آموزش

زمانی که هدف از ایجاد کارپوشه جنبهٔ آموزشی دارد، از آن می‌توان به عنوان وسیله‌ای برای کمک به یادگیرنده در ایجاد توانایی ارزشیابی از خود یا خودسنجی سود برد. یکی از هدفهای مهم آموزش معلم این است که دانش‌آموزان یاد بگیرند کارهایشان را خودشان ارزشیابی کنند. مهارت ارزشیابی از خود یا خود ارزیابی از راهبردهای مهم فراشناختی است. فراشناخت[1] ویژگی مهم یادگیرندگانی است که می‌توانند به طور مستقل یاد بگیرند و بر پیشرفت یادگیری خود نظارت داشته باشند (سیف، ۱۳۷۶، ۱۳۸۰).

کارپوشه‌هایی که به منظور ایجاد مهارت خودارزشیابی در دانش‌آموزان تهیه می‌شوند هم شامل مثالهایی از کارهای دانش‌آموز است که خود آنها را انتخاب کرده و هم دربرگیرندهٔ اندیشه‌ها و اظهارنظرهای او دربارهٔ کارهای خودش است. "ترکیب کارهای انتخابی و اندیشه‌های دانش‌آموز دربارهٔ آن کارها هم اطلاعات مورد نیاز را دربارهٔ پیشرفت دانش‌آموز و هم شواهد لازم دربارهٔ رشد مهارتهای خودارزشیابی در او را برای معلم فراهم می‌آورد" (لین و گرانلاند، ۲۰۰۰، ص ۲۹۴).

کمپ[2] (۱۹۹۰، به نقل گیج[3] و برلاینر[4]، ۱۹۹۲) شواهدی از کسب توانایی ارزشیابی از خود را در رابطه با یک دانش‌آموز پایهٔ نهم به دست داده است. این دانش‌آموز، پس از وارسی نوشته‌های خود، مانند شکل ۱ـ۱۳ توضیح داده است.

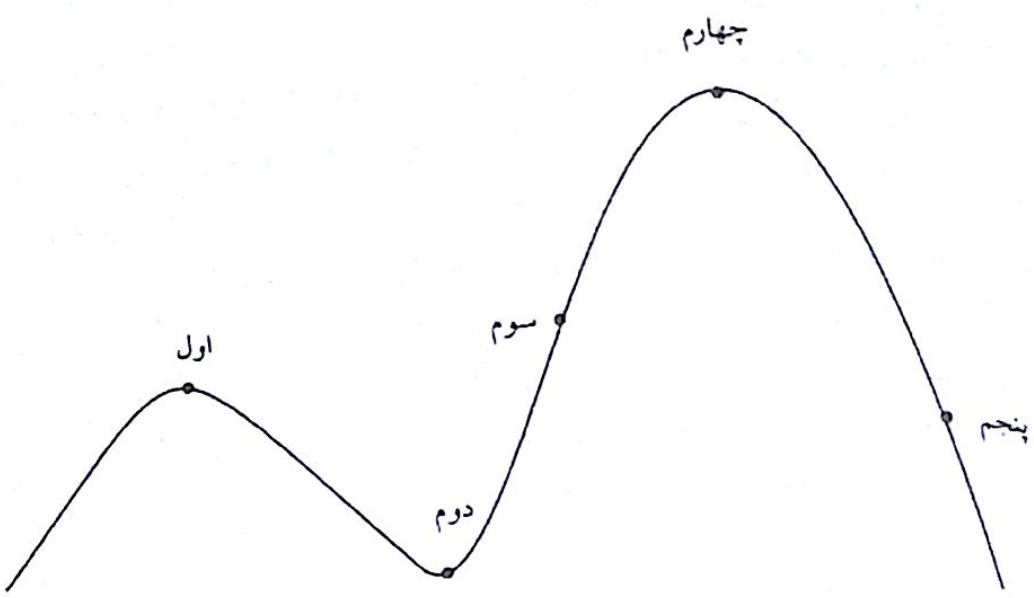
مسئولیت اصلی تهیه و تنظیم محتوای کارپوشه‌هایی که به منظور کمک به یادگیری و ایجاد مهارتهای خودارزشیابی در دانش‌آموزان به کار می‌روند به عهدهٔ خود دانش‌آموز است، اما مواردی که در این کارپوشه‌ها قرار داده می‌شوند هم توسط دانش‌آموز انتخاب می‌شوند، هم معلم، هم با همکاری معلم و دانش‌آموز، و هم با مشارکت دانش‌آموز و دوستان او (کلنوسکی[5]، ۲۰۰۲).

افزون بر ایجاد توانایی خودسنجی یا ارزشیابی از خود در یادگیرندگان، روش کارپوشه دارای هدفهای آموزشی دیگری نیز هست. کوبیسزین و بوریچ (۲۰۰۳) چند مورد از آنها را به صورت زیر فهرست کرده‌اند:

- نظارت بر پیشرفت دانش‌آموز
- گزارش به والدین دربارهٔ یادگیریهای فرزندانشان
- انتقال اطلاعات به معلمان بعدی دانش‌آموز
- پُر دادن دربارهٔ کارهای دانش‌آموز
- تعیین نمرهٔ درسی

1. metacognition 2. Camp 3. Gage 4. Berliner 5. Klenowski

فکر می‌کنم که نوشته‌های من به صورت شکل زیر دارای دو برآمدگی است.



به نظر خودم ابتدا خوب پیش می‌رفتم. بعد پیشرفتم کاهش یافت. بعد ناگهان پیشرفتم افزایش یافت، و این بار از دفعهٔ قبل بهتر بودم. ولی آخرین چیزی که نوشتم خیلی مزخرف بود.

شکل ۱ـ۱۳ توضیحات یک دانش‌آموز دربارهٔ چگونگی پیشرفت کار خود براساس محتوای کارپوشه‌اش (کمپ، ۱۹۹۰، ص ۲۷؛ به نقلِ گیج و برلاینر، ۱۹۹۲، ص ۶۶۴)

استفاده از کارپوشه برای سنجش

از روش کارپوشه هم می‌توان ارزشیابی یا سنجش تکوینی به عمل آورد و هم سنجش تراکمی. در فصل ۴ توضیح دادیم که ارزشیابی تکوینی یا سنجش تکوینی در جریان آموزش معلم و یادگیری دانش‌آموزان صورت می‌گیرد و هدف آن کمک به تکوین یا شکل‌گیری یادگیری دانش‌آموزان و رفع نواقص روش آموزشی معلم است. در مقابل، ارزشیابی تراکمی یا سنجش تراکمی (مجموعی) به عنوان وسیله‌ای برای تعیین آموخته‌های متراکم شده یا جمع شدهٔ دانش‌آموزان در طول یک ترم تحصیلی به کار می‌رود و قصد از اجرای آن تعیین صلاحیت یادگیرندگان و دادن نمره و گواهینامهٔ تحصیلی است.

کارپوشه‌هایی که در طول یک سال یا ترم تحصیلی گردآوری می‌شوند برای مقاصد سنجش تکوینی و هدایت یادگیری دانش‌آموزان مناسب‌اند. برای مثال، "مقایسهٔ ساده‌ای در بین انشاءهایی که یک دانش‌آموز در ماههای مهر، آذر، و اسفند نوشته است می‌تواند شواهد عینی لازم را دربارهٔ مثلاً پیشرفت چشم‌گیر دانش‌آموز (سازمان‌بندی مطالب) و عدم پیشرفت او (دستورزبان و علامت‌گذاری) فراهم آورد" (لین و گرانلاند، ۲۰۰۰، ص ۲۹۵).

کارپوشه‌هایی که به منظور سنجش تراکمی یا مجموعی به کار می‌روند و هدف آنها دادن نمره و صدور گواهینامهٔ تحصیلی است، مستلزم استفادهٔ همشکل و یکنواخت در میان دانش‌آموزان و کاربرد یک نظام داوری دقیق است. نتایج این کارپوشه‌ها معمولاً همراه با سایر روشهای سنجش، مانند آزمونها و امتحانات، مورد استفاده قرار می‌گیرند. لیسن و گرانلاند (۲۰۰۰) در رابطه با استفاده از کارپوشه برای مقاصد آموزش و سنجش گفته‌اند:

> زمانی که هدف اصلی کارپوشه افزایش سطح یادگیری دانش‌آموز است، و این کار از راه واداشتن او به ارزشیابی از کار خود صورت می‌پذیرد، هیچ‌گونه نگرانی دربارهٔ مقایسه‌پذیر بودن محتوای کارپوشه‌های دانش‌آموزان مختلف وجود ندارد. در اینجا تأکید عمده بر این است که به دانش‌آموزان آموزش داده شود تا در انتخاب مواد برای کارپوشه تصمیمهای منطقی بگیرند و بیاموزند که چگونه دربارهٔ کیفیت کار خود بیندیشند. از سوی دیگر، اگر هدف از کارپوشه نمره دادن به دانش‌آموز باشد، آن‌گاه مقایسه‌پذیر بودن مطالب درون کارپوشه‌های دانش‌آموزان مختلف بسیار با اهمیت است. در صورت اخیر، ممکن است معلم از همهٔ دانش‌آموزان بخواهد تا در کارپوشهٔ خود هم یک داستان کوتاه و هم یک گزارش کتاب را که خود نوشته‌اند قرار دهند. (ص ۲۹۳)

استفاده از کارپوشه برای آموزش و سنجش مهارتهای فکری و راهبردهای یادگیری

استفاده از کارپوشه توسط معلم به عنوان یک روش آموزشی و ارزشیابی نه تنها به پیشرفت تحصیلی و توانایی خودسنجی دانش‌آموزان و دانشجویان کمک می‌کند، بلکه موجب گسترش مهارتهای فکری و راهبردهای یادگیری نیز می‌شود. به طور مشخص‌تر، معلم می‌تواند با استفاده از کارپوشه به پرورش و سنجش مهارتهای فکری زیر در دانش‌آموزان کمک کند:

- تفکر
- حل مسئله
- تصمیم‌گیری
- راهبردهای شناختی: تحلیل، تفسیر، بسط معنایی، سازمان‌دهی
- راهبردهای فراشناختی: طرح‌ریزی، هدایت یادگیری، اندیشیدن دربارهٔ کارهای انجام گرفته

علاوه بر راهبردها و مهارتهای فکری سطح بالا، با استفاده از کارپوشه می‌توان آمادگیها و صفات ویژهٔ یادگیرندگان را نیز سنجش کرد. نمونه‌هایی از این گونه آمادگیها و صفات در زیر آمده‌اند:

- انعطاف‌پذیری در کارها
- توانایی سازگارشدن

- پذیرش انتقاد
- همکاری با دیگران
- اشتیاق نسبت به یادگیری

استفاده از کارپوشه برای فارغ‌التحصیل شدن

کارپوشه می‌تواند برای مقاصد فارغ‌التحصیلی نیز مورد استفاده قرار گیرد. لین و گرانلاند (۲۰۰۰) در این‌باره کارپوشه‌ای را توضیح داده‌اند که دربرگیرندهٔ فعالیتهای دانش‌آموزان در دو سال آخر دبیرستان است. موضوعهای مهمی که در این کارپوشه قرار داده می‌شوند از ۱۴ دسته تشکیل یافته‌اند. عنوانهای این ۱۴ دسته همراه با توضیح مختصری از محتوای کارپوشهٔ آنها در جدول ۱ـ۱۳ آمده است.

چنان‌که از جدول ۱ـ۱۳ پیداست، طبقه‌های کارپوشهٔ فارغ‌التحصیلی مورد بحث ما هم شامل مواد نوشتنی هستند و هم دربرگیرندهٔ شکلهای مختلف بیان شفاهی، مانند سخنرانی، بحث و مذاکره، و نمایشنامه. همچنین در این نوع کارپوشه‌ها کارهای هنری و فعالیتهای علمی قرار می‌گیرند. افزون بر اینها، نوارهای صوتی، نوارهای ویدیویی (تصویری)، و دیسکتهای کامپیوتری نیز می‌توانند بخشی از این کارپوشه باشند.

استفاده از کارپوشه برای نشان‌دادن موفقیتهای جاری و رشد تحصیلی دانش‌آموز

تقسیم‌بندی دیگر کارپوشه‌ها با توجه به موارد زیر صورت پذیرفته است: (۱) نشان دادن موفقیتهای جاری یادگیرنده و (۲) معرفی رشد پیشرفت او. وقتی که تأکید کارپوشه بر نشان دادن موفقیتهای جاری یادگیرنده است، محتوای کارپوشه معمولاً به کارهای پایان یافته محدود می‌شود و زمان نسبتاً کوتاهی را دربرمی‌گیرد. اما موقعی که هدف از تشکیل کارپوشه نشان دادن رشد تحصیلی یادگیرنده است، محتوای کارپوشه را آن دسته از فعالیتهای دانش‌آموز شامل می‌شود که در یک زمان نسبتاً گسترده‌تری انجام گرفته‌اند.

تفاوت دیگر بین دو نوع کارپوشهٔ فوق در این است که کارپوشهٔ مخصوصِ نشان دادن رشد موفقیت یا پیشرفت تحصیلی جنبه‌های مختلف یک کار واحد را دربرمی‌گیرد. "مثلاً، کارپوشهٔ یک پروژهٔ تحقیقی شامل موارد زیر است: طرح اولیه، یادداشتهایی که هنگام مطالعهٔ منابع مختلف برداشته شده‌اند، پیش‌نویس اولیهٔ گزارش تحقیقی، ارزشیابی شخصی یادگیرنده از آن پیش‌نویس و اظهارنظرهای معلم و دانش‌آموزان دیگر دربارهٔ آن، نسخهٔ اصلاح شدهٔ بعدی گزارش و بازهم اظهارنظرهای مربوط به آن، و بالاخره نسخهٔ نهایی گزارش" (لین و گرانلاند، ۲۰۰۰، ص ۲۹۶). در مقابل، چنان‌که گفتیم، کارپوشهٔ مخصوص معرفی موفقیتهای جاری تنها صورت کامل یا نسخهٔ نهایی کارهای انجام گرفته توسط دانش‌آموز را شامل می‌شود.

جدول ۱ـ۱۳ محتوای یک کارپوشهٔ فارغ‌التحصیلی، مربوط به دانش‌آموزان یکی از دبیرستانهای آمریکا (به نقل لین و گرانلاند، ۲۰۰۰، ص ۲۹۹)

۱. نقشهٔ بعد از فارغ التحصیلی	توضیحات دانش‌آموز دربارهٔ هدفش از گرفتن دیپلم دبیرستان و نقشه‌های کوتاه‌مدت و بلندمدت زندگیش، همراه با توصیه‌نامه‌هایی از معلمان که نشان‌دهندهٔ پیشرفت او باشد
۲. شرح حال شخصی	توضیحات شفاهی، کتبی، یا شکلهای دیگر (مانند عکس و نوار مصاحبه با بزرگان خانواده) دربارهٔ خانواده، همراه با اعتقادات و ارزشهای دانش‌آموز
۳. همکاری در خدمات مدرسه یا محل زندگی	یک گزارش رسمی از کارهای قبلی او که نشان‌دهندهٔ نوع کار، فعالیتهای دانش‌آموز، و پیشرفت او باشد
۴. مسائل اخلاقی و اجتماعی	شواهد نشان‌دهندهٔ توانایی دانش‌آموز در پذیرش باورها و نظرات متفاوت دیگران
۵. هنرهای زیبا و زیبایی‌شناسی	نمونه‌ای از کارها در هر یک از هنرها (مجسمه‌سازی، موسیقی، و غیره) همراه با شواهد درک یک زمینهٔ زیبایی‌شناختی
۶. رسانه‌های همگانی	شواهد مربوط به درک شکلهای مختلف رسانه‌های همگانی و تأثیر آنها بر مردم و چشم‌انداز آنها
۷. مهارتهای عملی	شواهد مربوط به کسب مهارتها در یک یا چند زمینه، مانند تکنولوژی، بهداشت، و شهروندی
۸. جغرافیا	شواهد مربوط به عملکرد دانش‌آموز در یک آزمون کلاسی و پروژهٔ دانش آموزی (مانند ساختن نقشه)
۹. زبان دوم	شواهد مربوط به توانایی خواندن، نوشتن، گوش‌دادن، و صحبت کردن در زبانی بجز زبان مادری
۱۰. علوم و تکنولوژی	شواهد معمول (مثلاً امتحانات و پروژه‌ها) و استفاده از روشهای علمی و درک نقش علم در دنیای مدرن
۱۱. ریاضیات	شواهد مربوط به عملکرد در آزمونها و پروژه‌ها که نشان‌دهندهٔ درک مفهومی و کاربرد اصول باشد
۱۲. ادبیات	فهرست مطالب خوانده شده، شامل انواع مختلف متنها، انشاءها و مقاله‌های دانش‌آموز دربارهٔ کارها و شخصیتهای ادبی
۱۳. تاریخ	شواهد مربوط به عملکرد در آزمونهای کلاسی یا ایالتی و پروژه‌های تاریخی که نشان‌دهندهٔ درک تاریخ و رابطهٔ آن با مسائل جاری باشد
۱۴. فعالیتهای بدنی	شواهد مربوط به شرکت در تیمهای ورزشی یا ورزشها و فعالیتهای انفرادی

استفاده از کارپوشه برای نمایش دادن کارها

منظور از کارپوشهٔ نمایشی[1] مجموعه کارهای انتخاب شده به وسیلهٔ دانش‌آموز است. البته در انتخاب کارها برای قراردادن در کارپوشهٔ نمایشی نیز دانش‌آموز می‌تواند از راهنمایی و نظر معلم و دوستانش بهره بگیرد. اما مهم‌ترین هدف این کارپوشه برای دانش‌آموز آن است که بیاموزد کارهایی را که به بهترین وجه مُعرف دانش و توانایی او هستند انتخاب کند و در کارپوشهٔ خود قرار دهد.

استفاده از کارپوشه برای ارائهٔ اسناد و مدارک

در مقابل کارپوشهٔ نمایشی، کارپوشهٔ اسناد و مدارک قرار دارد. این کارپوشه شواهدی را به دست می‌دهد که هم مُعرف کمیّت و هم کیفیت یادگیری دانش‌آموز است. بنابراین، محتوای این کارپوشه بسیار متنوع‌تر از محتوای کارپوشهٔ نمایشی است و صرفاً به زمینه‌های خاصی که دانش‌آموز در آن موفقیت کسب کرده است محدود نمی‌شود. به سخن دیگر، در کارپوشهٔ نمایشی تنها موردهایی گذاشته می‌شوند که به نظر دانش‌آموز بهترین کارهای او هستند، اما در کارپوشهٔ اسناد و مدارک هم کارهای شاخص دانش‌آموز و هم فعالیتهای ضعیف‌تر او درج می‌شوند.

استفاده از کارپوشه برای نشان دادن کارهای به ثمر رسیده و اقدامات در حال انجام

منظور از کارپوشهٔ کارهای به ثمر رسیده آن کارپوشه‌ای است که اقدامات کامل شدهٔ دانش‌آموز و آمادهٔ ارائه به مخاطبان موردنظر او را شامل می‌شود. به عنوان مثال، یک کارپوشهٔ تقاضای کار از نظر کارفرما باید شامل اقداماتی باشد که کامل شده و به ثمر رسیده‌اند.

در مقابلِ کارپوشهٔ کارهای کامل شده یا به ثمر رسیده، کارپوشهٔ کارهای نیمه تمام و در حال انجام قرار دارد. این کارپوشه برای نشان دادن اقدامات روزانهٔ دانش‌آموز و به قصد سنجش یا ارزشیابی تکوینی مورد استفاده قرار می‌گیرد، اما کارپوشهٔ فعالیتهای به ثمر رسیده ویژهٔ سنجش تراکمی یا ارزشیابی مجموعی است.

بهترین مورد استفادهٔ کارپوشهٔ کارهای در حال انجام تشویق دانش‌آموزان برای ابراز عقاید خود به طور آزادانه و دریافت بازخورد برای کارهایشان است. بنابراین، باید دقّت کرد که دانش‌آموزان برای بیان اندیشه‌های نادرست و مواردی که نشان‌دهندهٔ درک ناکامل آنان از مطالب درون کارپوشه است مورد تنبیه و سرزنش قرار نگیرند.

استفاده از کارپوشه برای گزینش و تأیید صلاحیت

استفاده از سنجش به کمک کارپوشه برای گزینش و تأیید صلاحیت افراد در حوزهٔ تربیت معلم شهرت یافته است. در کشور ایالات متحد آمریکا استفاده از این نوع کارپوشه هم در سطح

---

1. showcase

محلی و هم در سطح ملی متداول شده است. در سطح محلی از کارپوشه برای تربیت معلمان تازه و در سطح ملی برای تأیید صلاحیت معلمان با سابقه استفاده می‌شود. ولف[1] (۲۰۰۶) در رابطه با استفادهٔ گسترده از کارپوشه در آموزش و پرورش آمریکا گفته است "کارپوشه‌های آموزشی معلمان به سرعت در حال شهرت یافتن‌اند، و این افزایش شهرت هم در زمینهٔ ارزشیابی و هم توسعه به چشم می‌خورد. از کارپوشه در برنامه‌های تربیت معلم دانشگاهها برای ارتقاء سطح آموزش پیش از خدمت معلمان و در مدارس برای استخدام و ارزشیابی معلمان استفاده می‌شود" (ص ۱۶۸). کلنوسکی (۲۰۰۲) استانداردهای مورد استفاده در آمریکا و انگلستان را برای تعیین صلاحیت حرفهٔ معلمی به گونهٔ زیر معرفی کرده است:

- دانش حرفه‌ای (شامل دانش نظری و تعهد و شایستگی عملی)
- یادگیری تمام عمر معلم
- دانش معلمی
- ویژگیهای اخلاقی
- دانش مربوط به دانش‌آموزان و مشکلات یادگیری آنان
- دانش نظریه‌ها و راهبردهای آموزشی
- دانش منابع آموزشی
- دانش موضوعهای خاص آموزشی

کارپوشهٔ مخصوص گزینش و تأیید صلاحیت از یک مجموعه استانداردهای کاملاً مشخص و تعریف شده تشکیل می‌شود. محتوای این کارپوشه‌ها عموماً شامل کارهایی است که توانمندی‌ها و نقاط قوت صاحب کارپوشه را به بهترین وجهی نشان می‌دهند. کلنوسکی (۲۰۰۲) موردی از محتوای این نوع کارپوشه را از یکی از دانشگاههای آمریکا توضیح داده است: این کارپوشه حاوی مدارکی است حاکی از شایستگی فرد، مانند شرح فلسفهٔ آموزشی از زبان داوطلب، درسها و کارهای او به صورت کتبی و نوار ویدیوئی، و مذاکرات انجام گرفته بین معلم و همکاران یا معلم و مربی او دربارهٔ فعالیتهای آموزشی که روشنگر محتوای کارپوشه است ـ مثلاً معلم از تجربهٔ معلمی چه چیزی آموخته است، چه نکاتی از آن مهم‌اند. البته برای مقاصد سرنوشت‌ساز معمولاً از کارپوشه همراه با شواهد و اطلاعات دیگری نظیر مشاهدات انجام گرفته از کلاس درس معلم و نظرخواهی از همکاران او استفاده می‌شود. لین و گرانلاند (۲۰۰۰) دربارهٔ استفاده از کارپوشه برای ارزشیابی معلمان گفته‌اند:

به سبب علاقهٔ گسترده به کارپوشه به عنوان وسیله‌ای برای سنجش عملکرد آموزشی معلمان، تعدادی از ایالتهای کشور آمریکا به پروژه‌ای با نام کنسرسیوم بین ایالتی سنجش و حمایت از معلمان که هدف آن

1. Wolf

ایجاد یک نظام هماهنگ سنجش مبتنی بر کارپوشه است پیوسته‌اند.... کارپوشه‌هایی که به این منظور مورد استفاده قرار می‌گیرند شامل نوارهای ویدیوئی آموزش معلم، مواردی از کارهای دانش‌آموزان او، نمونه‌هایی از بازخورد دادنهای معلم به کارهای دانش‌آموزان، و اظهارنظرهای معلم دربارهٔ کارهای خودش است. (ص ۱۲)

استفاده از کارپوشه برای تشویق و ارتقاء کارکنان

از کارپوشه به عنوان وسیله‌ای برای سنجش به قصد تشویق و ارتقاء مقام و موقعیت افراد نیز استفاده می‌شود. این گونه کارپوشه‌ها دربرگیرندهٔ شواهدی است حاکی از رشد و ترقی فرد. از این روش عموماً برای ارزشیابی از کم و کیف کارکنان یک مؤسسه و تشویق و نظارت بر کار آنان استفاده می‌شود.

در روش سنجش به کمک کارپوشه برای تشویق و ارتقاء کارکنان، ابتدا کارفرما، مطابق با نیازهای مؤسسه و شرایط شغلیِ افرادی که قرار است عملکردشان سنجش شود، معیارهای مشخصی را تعیین می‌کند و در اختیار افراد مورد نظر قرار می‌دهد. آنان وظیفه دارند تا کارپوشه‌ای برای خودشان تهیه کنند و در آن شواهد و مدارکی که نشان‌دهندهٔ شایستگی‌شان در ارتباط با آن معیارهاست قرار دهند.

کلنوسکی (۲۰۰۲) تأکید محتوای کارپوشه‌هایی راکه برای ارتقاء معلمان به کار می‌روند در نشان دادن دانش و مهارتهای معلمی دانسته است. "شواهد موجود در کارپوشه در طول زمان فراهم می‌آیند و مُعرف کارهای معلم و شاگردان اوست. پاره‌ای اظهارنظرهای متفکرانه یا ارزشیابانه نیز به آن اضافه می‌شود تا شواهد و مدارک انتخابی را برای دیگران، از جمله ارزشیابان و کارفرمایان توضیح دهد" (ص ۲۰).

کارپوشهٔ الکترونیکی

از فنّاوری الکترونیکی می‌توان برای سازمان دادن، ذخیره کردن، و ارائه دادن کار پوشه‌های دانش‌آموزی استفاده کرد. اصطلاحات **کارپوشهٔ الکترونیکی**[1] و **کارپوشه براساس کامپیوتر (رایانه)**[2] به همین منظور به کار می‌روند. اِگِن و کاوچاک (۲۰۰۱) دربارهٔ امتیاز این روش سنجش می‌گویند "کارپوشه‌های الکترونیکی مشکل جا یا فضای ذخیره‌سازی محتوای کارپوشه و دسترسی‌پذیری را حل می‌کنند و قابلیت استفادهٔ از آنها را افزایش می‌دهند.... معلمانی که با مشکلات سامان‌دهی انبوه بسته‌ها و جعبه‌های مربوط به کارپوشه‌های دانش‌آموزان روبه‌رو بوده‌اند ارزش ذخیره‌سازی الکترونیکی محتوای کارپوشه‌ها را می‌دانند" (ص ۶۳۶).

کارپوشهٔ الکترونیکی می‌تواند مطالب نوشتاری، تصاویر، صدا، و ویدئو را شامل شود. بنابراین، "دانش‌آموزان و دانشجویان می‌توانند نمونهٔ نوشته‌ها، راه‌حل مسائل ریاضی، کارهای

1. electronic portfolio 2. computer-based portfolio

هنری، گزارش کارهای تحقیقی، و گزارشهای چندرسانه‌ای را در کارپوشه‌های خود قرار دهند» (سانتروک، ۲۰۰۴، ص ۵۵۲). امتیاز دیگر کارپوشهٔ الکترونیکی سهولت انتقال اطلاعات از یک معلم به معلم دیگر و از یک مدرسه به مدرسه دیگر است. انجام این کار از طریق دیسک‌های فشرده امکان‌پذیر است.

باتلر و مکمان (۲۰۰۶) کارپوشه‌های دانش‌آموزی را به پنج دسته کارپوشهٔ: (۱) بهترین کارها، (۲) مجموعه کارها، (۳) رشد پیشرفت، (۴) مهارتها، و (۵) سنجش، شایستگی، و ارتقاء تقسیم کرده‌اند. آنان کارپوشهٔ الکترونیکی را موردی از کارپوشهٔ نوع آخر یعنی کارپوشهٔ شایستگی می‌دانند، و ساختمان آن را به صورت شکل ۲ـ۱۳ نشان داده‌اند.

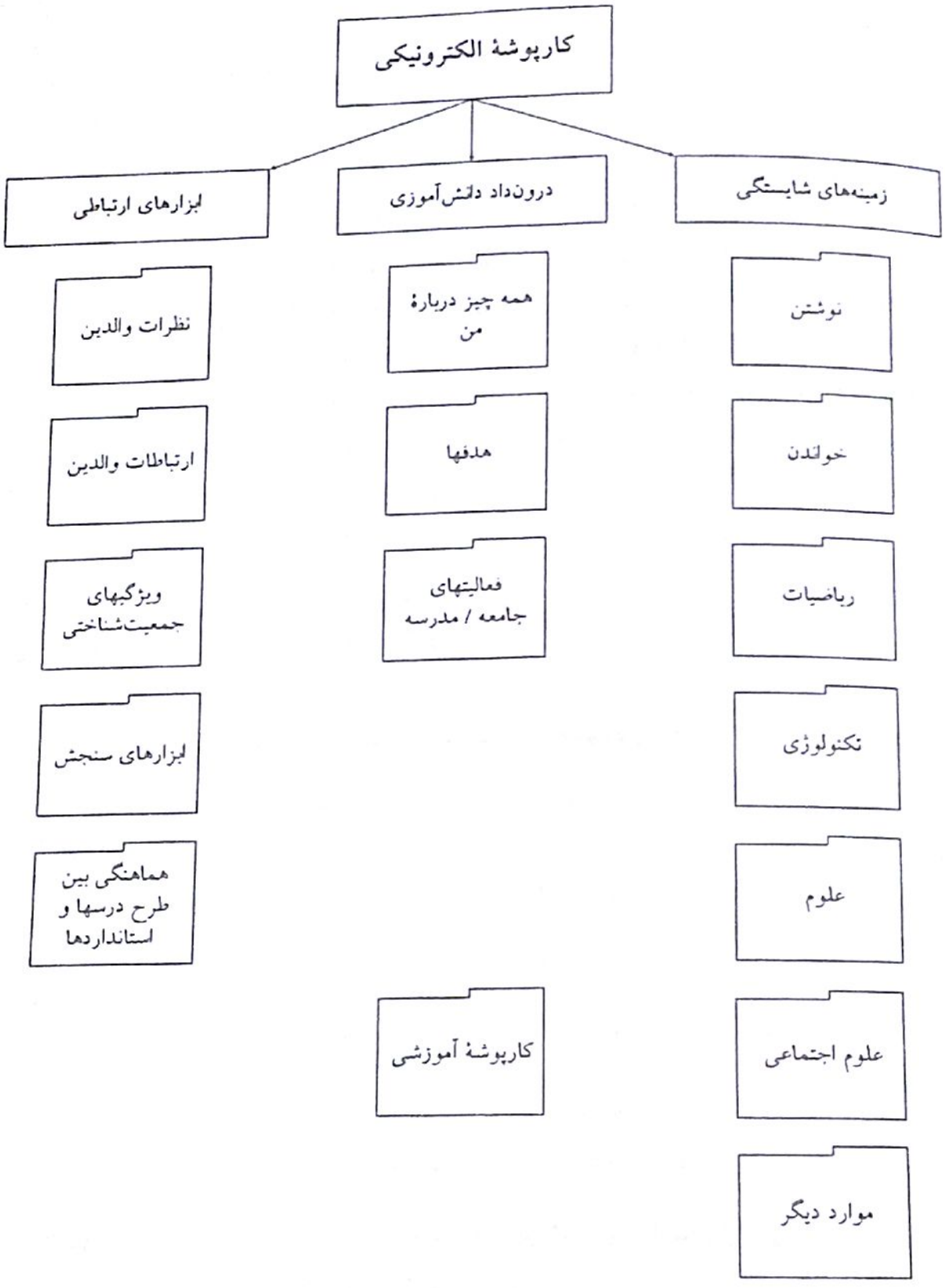
برای کارپوشه‌های الکترونیکی چندین برنامهٔ تجاری درست شده است. سانتروک (۲۰۰۴) چند مورد آنها را معرفی کرده است. معروف‌ترین برنامهٔ مربوط به کارپوشهٔ الکترونیکی با نام Aurbach's Grady portfolio شناخته شده است. در این برنامه هم دانش‌آموزان و هم معلمان می‌توانند نمونه کارهای خود را قرار دهند. این برنامه می‌تواند نمونه نوشته‌ها، نمرات آزمونهای استاندارد شده، مهارتهای ارتباطی شفاهی، و تکالیف ریاضی را شامل شود. برنامه‌های دیگری نیز با نامهای زیر وجود دارند: Hyper Studio و File Maker. این برنامه‌ها به معلمان امکان می‌دهند تا الگوی موردنظر خودشان را برای کارپوشه‌ها ابداع کنند و آنها را با نیازهای کلاس درسشان انطباق دهند.

یکی از مدارسی که از کارپوشه‌های الکترونیکی استفاده می‌کند دبیرستان سیراکیوس مینوا ـ شرقی[1] واقع در سیراکیوس ایالت نیویورک است. دانش‌آموزان این دبیرستان کارپوشه‌های الکترونیکی راکه خود خلق می‌کنند به عنوان بخشی از مدارک پذیرش به دانشگاه‌های مورد درخواست برای تحصیل می‌فرستند. همچنین این کارپوشه‌ها را به عنوان سند شایستگی خود همراه با تقاضای کار برای کارفرمایان بالقوه می‌فرستند. این کارپوشه‌ها با استفاده از برنامهٔ پیش گفتهٔ Hyper Studio ساخته می‌شوند (سانتروک، ۲۰۰۴). آنها شامل اطلاعات دربارهٔ دانش‌آموز (از جمله ریز نمرات، توصیه‌نامه‌های معلمان، و سوابق کاری) و کارهای انتخابی دانش‌آموزان (مانند نمونه نوشته‌ها، پروژه‌های چندرسانه‌ای، کارهای هنری، و ویدئو کلیپ مربوط به نمایش اجرا شده در مدرسه) هستند. دانش‌آموزان از سال دوم دبیرستان به تهیهٔ این کارپوشه‌ها اقدام می‌کنند و تا سال آخر دبیرستان مرتباً در آن تجدیدنظر می‌کنند و آن را به روز درمی‌آورند. این کارپوشه‌ها قابلیت توزیع به صورت دیسک کامپیوتری، نوار ویدئویی، و نوشتهٔ کتبی را دارند.

تشکیل کارپوشه‌های الکترونیکی در امریکا به وسیلهٔ قانون شغل از طریق آموزش و پرورش مصوب سال ۱۹۹۶[2] اجباری شده است. هدف این قانون واداشتن مدارس دولتی به تأکید بر آماده ساختن دانش‌آموزان برای محیطهای کار است. انتظار این است که سرانجام

1. East Syracuse-Minoa High School 2. 1996 Jobs Through Education Act

کارپوشه‌های الکترونیکی که دربرگیرندهٔ تاریخچهٔ کاملی از عملکردها، نمونه کارها، و فعالیتهای کلاسی دانش‌آموز است جایگزین ریز نمرات او بشوند.



شکل ۲-۱۳ چهار چوب ساختاری کارپوشهٔ الکترونیکی (باتلر و مکمان، ۲۰۰۶، ص ۶۹)

محتوای کارپوشه

ما در صفحات قبل، در ارتباط با هدفهای گوناگون استفاده از کارپوشه، دربارهٔ محتوای آن اشاره‌هایی داشتیم. در این قسمت بیشتر آن را توضیح می‌دهیم. روش سنجش به کمک کارپوشه در آغاز فقط موضوعهایی مانند هنر و ادبیات را شامل می‌شد، ولی بعداً دامنهٔ استفاده از آن گسترش یافت و اکنون در موضوعهای دیگری چون ریاضیات و علوم نیز به کار می‌رود. جدول ۲ـ۱۳ محتوای کارپوشه را در چند موضوع مختلف نشان می‌دهد.

سانتروک (۲۰۰۴) دربارهٔ مواد و مطالبی که در کارپوشه قرار می‌گیرند گفته است "کارپوشه‌ها انواع مختلف کارهای دانش‌آموز را شامل می‌شوند، مانند نمونه کارهای نوشتنی، نوارهای ویدیویی، کارهای هنری، اظهارنظرهای معلم، پوستر، مصاحبه، شعر، نتایج آزمونها، حل مسائل، خودسنجی‌ها، و هرگونه کار دیگری از دانش‌آموز که به باور معلم مهارتها و پیشرفتهای او را نشان می‌دهد" (ص ۵۴۲). در ضمن، کارپوشه‌ها را می‌توان به صورت برگه‌های کاغذ و عکسهای مختلف و بر روی نوارهای صوتی و تصویری، دیسک‌های کامپیوتری یا دیسک‌های فشرده تشکیل داد" (ص ۵۴۲).

بنا به گفتهٔ سانتروک (۲۰۰۴)، به طور کلی چهار دسته گواه و مدرک زیر را می‌توان در کارپوشهٔ دانش‌آموز قرار داد: (۱) کارهای انجام شده به وسیلهٔ دانش‌آموز[1]، (۲) کارهای تهیه شده به وسیلهٔ دانش‌آموز[2]، (۳) اظهارنظرهای دیگران دربارهٔ دانش‌آموز[3]، و (۴) تولیدات دانش‌آموز[4] در زیر به توضیح این موردها توجه کنید.

جدول ۲ـ۱۳ چند نوع کارپوشه با محتواهای مختلف در موضوعهای مختلف (اگن و کاوچاک، ۲۰۰۱، ص ۶۱۷)

موضوع	مثال
ریاضیات دبستان	تکالیف خانه، امتحانات هفتگی، امتحانات فصلی، و پروژه‌هایی که در یک مدت معین انجام گرفته‌اند.
نوشتن	نسخه‌های مراحل مختلف تولید انشاهای بیانی، توصیفی، و ترغیبی
هنر	پروژه‌های انجام شده در طول سال تحصیلی که رشد یادگیرنده را در زمینه‌های مختلف، مانند نقاشی، نشان می‌دهد.
علوم	گزارشهای آزمایشگاهی، پروژه‌ها، یادداشتهای کلاسی، امتحانهای هفتگی، و امتحانهای فصلی

1. artifacts 2. reproductions 3. attestations 4. productions

منظور از کارهای انجام شده به وسیلهٔ دانش‌آموز نتایج فعالیتهایی است که او در طول سال یا ترم تحصیلی در کلاس انجام می‌دهد، مانند تکالیف کلاسی. کارهای تهیه شده به وسیلهٔ دانش‌آموز به نتایج فعالیتهایی که او در خارج از کلاس درس انجام می‌دهد اشاره می‌کند، مانند پروژه و مصاحبه. "برای مثال، گزارش توصیفی دانش‌آموز از مصاحبه‌ای که او با یک دانشمند محلی دربارهٔ کارهای علمی آن دانشمند انجام داده است" (سانتروک، ۲۰۰۴، ص ۵۴۳). اظهارنظرهای دیگران دربارهٔ دانش‌آموز شامل گواهیها و دیگر مدارک معلم یا سایر افراد مسئول است که پیشرفت دانش‌آموز را نشان می‌دهد. برای نمونه، "ممکن است معلم دربارهٔ گزارش شفاهی دانش‌آموز یک یادداشت ارزشیابانه بنویسد و آن را در کارپوشهٔ او قرار دهد" (سانتروک، ۲۰۰۴، ص ۵۴۳). تولیدات مدارک و مطالبی هستند که دانش‌آموز اختصاصاً آنها را برای کارپوشه‌اش فراهم می‌آورد، و از سه دسته تشکیل می‌یابند: بیان هدف، اندیشه‌ها، و توضیحات. منظور آن است که دانش‌آموز هدفها و انتظارات خود را از اینکه از تهیهٔ کارپوشه چه چیزی به دست خواهد آورد توضیح دهد، اندیشه‌های خودش را دربارهٔ کارهایش بنویسد، و دربارهٔ هر یک از کارهایی که در کارپوشه قرار می‌دهد و همچنین اهمیت آنها توضیحاتی ارائه دهد.

پین (۲۰۰۳) دراین‌باره که چه چیزی را در درون یک کارپوشه می‌توان قرار داد گفته است "به معنی واقعی هر چیزی که معلم و دانش‌آموز با همدیگر مفید تشخیص بدهند در کارپوشه جای خواهد داشت" (ص ۳۳۰). او بدین منظور فهرست زیر را به دست داده است:

- پروژه‌های هنری
- مقاله و انشاء
- یادداشتهای روزانه
- اظهارنظرهای والدین
- پوستر
- مقاله‌های تحقیقی
- داستان
- داده‌های نگرشی
- مصاحبه‌ها
- نامه‌ها
- شعر
- حل مسائل
- خودسنجی‌ها
- نتایج آزمونها

پین (۲۰۰۳) پس از ارائهٔ فهرست بالا گفته است "از برکت خلاقیت معلم و دانش‌آموز، این فهرست بی‌پایان است. شرایط را ساده بگیرید و معیارهایتان را در اختیار دانش‌آموز قرار دهید." (ص ۳۳۱).

قواعد تهیه و استفاده از کارپوشه

نکته‌ای که بیش از هر چیز دیگر باید مورد توجه تهیه کنندگان کارپوشه قرار گیرد قصد و منظوری است که کارپوشه براساس آن تهیه می‌شود. ما در این باره زیر عنوان هدفهای تهیهٔ کارپوشه در صفحه‌های قبل توضیح دادیم. علاوه بر هدف کارپوشه، رعایت قواعد و رهنمودهایی که در این قسمت معرفی می‌شوند شما را در ایجاد و استفاده از کارپوشه‌ها به طور مؤثر یاری می‌دهد.

لین و گرانلاند (۲۰۰۰) گفته‌اند برای تهیهٔ کارپوشه دست کم چهار نکتهٔ زیر باید رعایت شوند:

۱. استفاده‌هایی که از کارپوشه به عمل خواهد آمد
۲. کسانی که به آن دسترسی خواهند داشت
۳. مناسبت داشتن هدف تهیهٔ کارپوشه با کارهایی که قرار است در کارپوشه قرار گیرند
۴. معیارها یا ملاکهایی که در ارزشیابی از کارپوشه به کار خواهند رفت

علاوه بر موارد فوق، باید ساخت فیزیکی یا شکل ظاهری کارپوشه نیز مشخص شود. یعنی اینکه، معلوم شود کارهای جمع‌آوری شده در درون یک پوشه[1] (مقوایی، پلاستیکی، و غیره) قرار گیرد یا به صورت دیسکت کامپیوتری باشد. همچنین باید مشخص شود که کارپوشه چه نکاتی از قبیل مقدمه یا معرفی، فهرست مطالب، فهرست منابع، بحث و نتیجه گیری، و غیره را شامل شود. آخرین نکته زمان تهیه و تحویل کارپوشه است. یعنی اینکه دانش‌آموز باید از قبل بداند که برای تهیهٔ کارپوشهٔ خود چه مقدار وقت در اختیار دارد و آخرین مهلت تحویل آن چه زمانی است و در صورت دیرکرد آیا از نمرهٔ او کسر می‌شود و به چه مقداری.

در رابطه با چگونگی تهیهٔ کارپوشه‌های دانش‌آموزی، دمبو (۱۹۹۴) رهنمودهای زیر را ارائه داده است.

۱. نمونه‌هایی از کارهای دانش‌آموزان جمع‌آوری کنید (دانش‌آموزان را در تصمیم‌گیری دخالت دهید).
۲. از پدران و مادران دانش‌آموزان بخواهید تا در فرایند تشکیل کارپوشه شرکت کنند. بدین منظور، دانش‌آموزان می‌توانند نوشته‌های خود را با والدینشان در میان بگذارند و به آنها کمک کنند تا متوجه شوند چه نوع مطالبی در کلاس بحث می‌شود.
۳. دربارهٔ ایجاد کارپوشه بحث کلاسی انجام دهید. به دانش‌آموزان توضیح دهید که آنها باید قسمتهایی از کارهای خود را که معرف علاقه، تنوع، و کوشش آنان است انتخاب کنند.

---

1. folder

۴. دربارهٔ عناصر مختلفی که دانش‌آموزان می‌توانند در کارپوشه‌های خود قرار دهند بحث کنید.
۵. به دانش‌آموزان در انتخاب کار برای درج در کارپوشه کمک کنید.
۶. از دانش‌آموزان بخواهید تا دلایل خود را برای درج هر کار در کارپوشه ذکر کنند. همچنین از آنان بخواهید تا تاریخ انجام هر قسمت از کار را بنویسند.
۷. شما (معلم) کارپوشه‌ها را به دقت مرور کنید. نظر خود را دربارهٔ نقاط قوّت و ضعف هر یک از کارها در کنار آن بنویسید. از فهرست وارسی برای ثبت مراحل پیشرفت هر دانش‌آموز استفاده کنید.
۸. به طور مرتب محتوای کارپوشه‌ها را نو کنید (به روز درآورید).
۹. به دانش‌آموزان اجازه بدهید تا در جلسات کلاس درس کارپوشه‌های خود را به دانش‌آموزان دیگر و در خانه به پدران ومادرانشان نشان دهند.
۱۰. از کارپوشهٔ دانش‌آموز برای بحث دربارهٔ جنبه‌های کار او با والدینش استفاده کنید.

افزون بر رهنمودهای بالا، به طور کلی می‌توان قواعد مربوط به تهیهٔ کارپوشه را در پنج دستهٔ زیر قرار داد: (۱) تعیین هدف یا قصد و منظور کارپوشه؛ (۲) دخالت دادن دانش‌آموز در انتخاب محتوای کارپوشه؛ (۳) مرور کردن محتوای کارپوشه با دانش‌آموز، (۴) تعیین ملاک یا معیار ارزشیابی، (۵) نمره گذاری و داوری. در زیر به توضیح هر یک از اینها توجه کنید.

۱. **هدف از تهیهٔ کارپوشه را تعیین کنید.** کارپوشه را برای مقاصد مختلفی می‌توان مورد استفاده قرار داد. سعی کنید، پیش از هرچیز، قصد و هدف خود را از تهیهٔ کارپوشه مشخص کنید. (ما در قسمتهای قبلی این فصل استفاده‌های گوناگونی را که می‌توان از کارپوشه به عمل آورد توضیح دادیم).

۲. **دانش‌آموز را در انتخاب محتوای کارپوشه دخالت دهید.** صاحب‌نظران سنجش و ارزشیابی تحصیلی پیشنهاد داده‌اند که بهتر است معلمان دانش‌آموزان خود را در تصمیمهای مربوط به انتخاب محتوای کارپوشه دخالت دهند، یا دست کم به آنها اجازه دهند تا بخشی از محتوای کارپوشه‌شان را خود از میان کارهایشان برگزینند. سانتروک (۲۰۰۴) گفته است "زمانی که به دانش‌آموزان اجازه می‌دهید تا محتوای کارپوشه‌شان را خودشان تهیه کنند، یک شیوهٔ درست این است که از آنان بخواهید تا با نوشتن توضیح مختصری دربارهٔ علت انتخاب هر بخشی از کار خود به اندیشیدن تشویق شوند" (ص ۵۴۴).

۳. **مطالب کارپوشه را با دانش‌آموز مرور کنید.** لازم است شما (معلم) در ابتدای سال تحصیلی برای دانش‌آموزان توضیح دهید که کارپوشه چیست، چگونه تهیه می‌شود، و چه مقاصدی را برآورده می‌سازد. همچنین مفید است که در طول سال تحصیلی مطالب گنجانیده شده در کارپوشه را با دانش‌آموزان مرور نمایید و به آنان کمک کنید تا برای کارهایی که در آینده در کارپوشه قرار خواهند داد برنامه‌ریزی کنند.

۴. **ملاکهای ارزشیابی را مشخص کنید.** یکی از شرایط موفقیت سنجش به کمک کارپوشه تعیین ملاکها یا معیارهای روشن برای ارزشیابی عملکرد یادگیرنده است. عامل تعیین‌کنندهٔ ملاکهای ارزشیابی کارپوشه همان هدفهای آموزشی معلم است. یعنی بیان انتظارات معلم از آنچه که قرار است در نتیجهٔ کارپوشه برآورده شوند. لین و گرانلاند (۲۰۰۰) گفته‌اند اگر هدفهای آموزشی برای معلم روشن نباشند، کوشش برای تعیین ملاکهای ارزشیابی کارپوشه یا به بازبینی و بازنویسی مجدد هدفها خواهد انجامید و یا ملاکهای ارزشیابی نامربوط و غیرمفیدی تهیه خواهند شد" (ص ۱۲).

بنابراین، معلم باید در آغاز درس و هنگام تهیهٔ طرح آموزشی خود به تهیهٔ هدفهای آموزشی‌اش اقدام نماید و این هدفها را مبنا و معیار آموزش و ارزشیابی خود قرار دهد. "از خود بپرسید می‌خواهید دانش‌آموزانتان چه دانشها و مهارتهایی کسب نمایند. در آموزش و تعیین معیار عملکرد همین‌ها باید مورد توجه شما باشند" (سانتروک، ۲۰۰۴، ص ۵۵۴).

دقت کنید که ملاکهای ارزشیابی هم برای معلمان باید روشن کنندهٔ هدفهای آموزشی باشند، هم برای دانش‌آموزان، و هم برای پدران و مادران دانش‌آموزان. بنابراین، همان‌گونه که هدفهای آموزشی معلم باید در اختیار دانش‌آموزان قرار گیرند ملاکهای ارزشیابی یا سنجش عملکرد نیز باید برای آنان روشن و در دسترس باشند.

۵. **کارپوشه را ارزشیابی کنید.** در ارزشیابی از کارپوشه روشهای مختلفی به کار می‌روند. از جملهٔ روشها و فنون نمره‌گذاری قابل استفاده برای کارپوشه می‌توان به فهرست وارسی، مقیاس درجه‌بندی، و روشهای کلی و تحلیلی، مانند آنچه در فصلهای ۱۰ و ۱۲ در رابطه با آزمونهای تشریحی و آزمونهای عملکردی توضیح دادیم اشاره کرد. ما در آن فصلها دو روش عمدهٔ تصحیح پاسخ سؤالهای تشریحی را با نامهای روش تحلیلی و روش کلی توضیح دادیم. در تصحیح محتوای یک کارپوشه نیز می توان از این روشها سود برد. روشهای سنجش مبتنی بر فهرست وارسی و مقیاس درجه‌بندی را در فصل بعد (فصل ۱۴) توضیح خواهیم داد. چنان‌که پیش‌تر در فصلهای ۱۰ و ۱۲ توضیح دادیم، در روش تحلیلی نمره‌گذاری، مجموعه‌ای از معیارهای ارزشیابی به کار می‌روند که برای هر یک از آنها نمره‌ای جداگانه در نظر گرفته می‌شود. در روش کلی، معیارهای مختلف درهم می‌آمیزند و یک نمرهٔ واحد گزارش می‌شود.

جدول ۳ـ۱۳ چگونگی نمره‌گذاری یک کارپوشه دربارهٔ یک درس دانشگاهی مربوط به روابط عمومی را نشان می‌دهد. دقت کنید که در این جدول ۱۲ اصل راهنما الگو واقع شده‌اند که هر یک از آنها با روش مقیاس درجه‌بندی نمره‌گذاری شده است؛ یعنی ترکیبی از روش تحلیلی و روش مقیاس درجه‌بندی مورد استفاده قرار گرفته است.

از دو روش اصلی مورد استفاده برای ارزشیابی کارپوشه، روش تحلیلی بیشتر برای

جدول ۳ـ۱۳ روش نمره گذاری محتوای یک کارپوشهٔ مربوط به درس روابط عمومی در سطح دانشگاه (پین، ۲۰۰۳، ص ۳۴۰)

نام دانشجو ..... نام ارزشیاب ..... تاریخ .....

راهنمایی: براساس مقیاس ۵ تا ۱ مشخص کنید که مطالب گزارش شده در این کارپوشه چگونه با معیارهای زیر مطابق است:

بالاترین ۵ ۴ ۳ ۲ ۱ پائین‌ترین

۱. نشان‌دهندهٔ خلاقیت است.
۲. نشان‌دهندهٔ مجموعه‌ای از کارهای متنوع است.
۳. دارای سبک نوشتاری درست است.
۴. دارای سازمان بندی خوب است.
۵. مطالب به روشنی بیان شده‌اند.
۶. در نوشتن مطالب اختصار رعایت شده است.
۷. مخاطبان در نظر گرفته شده‌اند.
۸. مطالب ترغیب کننده‌اند.
۹. از لحاظ دستوری جمله‌ها درست به کار رفته‌اند.
۱۰. املای کلمات درست است.
۱۱. پاراگرافهای مقدمه خوب تهیه شده‌اند.
۱۲. گزارش آراسته و حرفه‌ای جلوه می‌کند.

ارزشیابی تکوینی مناسب است و روش ترکیبی عموماً برای ارزشیابی تراکمی مفید است. همچنین، برای ارزشیابی از ماده‌های انفرادی درون کارپوشه بهتر است از روش نمره گذاری تحلیلی و برای ارزشیابی از کُل محتوای کارپوشه از روش کلی یا سراسری استفاده شود (لین و گرانلاند، ۲۰۰۰).

مطلبی راکه باید در ارتباط با سنجش عملکرد دانش‌آموزان براساس کارپوشه در نظر داشت این است که وقتی ارزشیابی جنبهٔ تراکمی دارد و هدفْ نمره دادن به دانش‌آموز است، بهتر است نام دانش‌آموز بر معلم پوشیده بماند. در تصحیح برگه‌های امتحانات انشایی یا تشریحی، علاوه بر این کار، از راه به هم زدن برگه‌ها و نمره گذاری سؤال به سؤال به جای ورقه به ورقه مخفی نگه داشتن نام دانش‌آموز را می‌توان به خوبی انجام داد، اما در مورد کارپوشه مخفی نگه‌داشتن نام دانش‌آموز قدری دشوار است. با این حال، برای انجام این منظور پیشنهادهایی مطرح شده که ما در اینجا به آنها اشاره می‌کنیم.

۱. می‌توان از دانش آموزان خواست تا نام خود را تنها بر روی جلد یا در صفحهٔ آخر کارپوشه بنویسند. البته این راه خیلی مطمئنی نیست و باید به دنبال راههای دیگری برای جلوگیری از سوگیریهای احتمالی معلم نیز بود.

۲. یکی دیگر از تدابیر مفید جلوگیری از تأثیر سوگیریهای معلم در ارزشیابی از کارپوشهٔ دانش‌آموز این است که کارپوشه را بعد از گذشت مدت زمانی از تصحیح اولیهٔ آن دوباره تصحیح کنید و نتایج دوبار ارزشیابی را با هم مقایسه نمایید.

۳. همچنین می‌توانید از یکی از معلمان همکار که همان درس شما را در همان پایه‌ای که شما تدریس می‌کنید آموزش می‌دهد بخواهید تا کارپوشه‌های دانش‌آموزان شما را تصحیح کند، و همین کار را شما برای او انجام دهید. مقایسهٔ نمره‌های شما با نمره‌های معلم همکار نه تنها در جلوگیری از اِعمال نظرهای احتمالی به شما کمک می‌کند بلکه در بهبود بخشیدن به شیوهٔ نمره گذاری کارپوشه مفید است و پایایی نمره گذاری را افزایش می‌دهد.

لین و گرانلاند (۲۰۰۰) گفته‌اند زمانی که از روش کارپوشه در مقیاس گستردهٔ منطقه‌ای یا ایالتی به عنوان بخشی از نظام ارزشیابی مدارس استفاده می‌شود یک فرایند رسمی‌تر نمره گذاری مورد استفاده قرار می‌گیرد. از جمله اینکه معلمان مورد نظر را به طور گروهی آموزش می‌دهند تا یاد بگیرند که از ملاکهای واحد نمره گذاری استفاده کنند. "اگرچه این‌گونه آموزشها به معلمان و تدارک اقدامات نمره گذاری مستلزم صرف وقت و منابع زیاد است، با این حال معلمانی که در این‌گونه برنامه‌ها شرکت جسته‌اند آنها را به عنوان تجربه‌های حرفه‌ای بسیار ارزشمند ارزیابی کرده‌اند" (لین و گرانلاند، ۲۰۰۰، ص ۳۱۰).

امتیازها و محدودیتهای روش سنجش با استفاده از کارپوشه

روش کارپوشه ابتدا در صنعت به کار بسته شد و با موفقیت مواجه گردید. بعدها که مربیان موفقیت این روش را در صنعت مشاهده کردند به استفاده از آن در آموزش تشویق شدند (مرفی و دیویدشوفر، ۲۰۰۱). اکنون در محافل تربیتی جهان، استفاده از این روش طرفداران زیادی کسب کرده است. مدافعان روش کارپوشه استدلال می‌کنند که این روش دربارهٔ چگونگی تفکر و حل مسئله در دانش‌آموزان اطلاعات بیشتری از روشهای سنتی سنجش به دست می‌دهد، دانش‌آموزان را وامی‌دارد تا نقش فعال‌تری در سنجش پیشرفت خود ایفا کنند، و به معلمان کمک می‌کند تا دانش‌آموزان خود را در مواد و مطالب معنی‌دارتری سنجش نمایند. آنان همچنین می‌گویند سنجش با روش کارپوشه به نهادهای آموزشی امکان می‌دهد تا برنامه‌های خود را بهتر ارزشیابی کنند و برای رشته‌های مختلف تحصیلی ملاکهای عملکردی مشابهی تعیین نمایند.

همچنین دربارهٔ سنجش به کمک کارپوشه گفته شده که این روش به دانش‌آموزان کمک می‌کند تا مسئولیت انتخاب هدف را برای خودشان بپذیرند و بر جریان پیشرفتشان نظارت داشته باشند. در ضمن، با استفاده از این روش از طریق نشان دادن مثالها و نمونه‌هایی از کارهای دانش‌آموزان و نمایش جریان پیشرفت آنان، معلم می‌تواند به منظور آگاه کردن والدین از

چگونگی پیشرفت فرزندانشان به طور مؤثرتری با آنان ارتباط برقرار نمایند. "کارتهای گزارش نمرات در این‌باره اطلاعاتی در اختیار می‌گذارند، اما کارپوشه، علاوه بر نمرات، شواهد حمایت‌کنندهٔ نمرات را نیز به دست می‌دهد" (کوبیسزین و بوریچ، ۲۰۰۳، ص ۱۷۵).

امتیاز دیگر استفاده از روش کارپوشه خاصیت دوگانهٔ آموزش و سنجش آن است. یعنی علاوه بر آنکه یک روش سنجش است از آن برای آموزش نیز می‌توان استفاده کرد. لین و گرانلاند (۲۰۰۰) در این‌باره گفته‌اند "به دلیل سهولت ادغام کارپوشه با آموزش، این روش مورد استقبال معلمان قرار گرفته است[1]" (ص ۲۹۰).

علاوه بر امتیازهای استفاده از کارپوشه با دانش‌آموزان، برای استفاده از این روش با معلمان نیز امتیازهایی ذکر شده است. از جمله ولف (۲۰۰۶) گفته است مدافعان این روش ادعا می‌کنند که کارپوشه‌ها در طول زمان دیدگاههای واقعی‌تری از یادگیری و آموزش ارائه می‌دهند و تصویر معتبرتری از آنچه معلمان می‌دانند و می‌توانند انجام دهند در اختیار ما می‌گذارند.

پس از آنکه روش کارپوشه با موفقیت در صنعت به کار بسته شد، مربیان امیدوار بودند که استفاده از آن در آموزش و پرورش نیز به همان نسبت موفقیت‌آمیز باشد. اما پژوهشهایی که دربارهٔ اثربخشی این روش در آموزش و پرورش انجام گرفتند نشان دادند که روش کارپوشه نتوانسته است موفقیتهای حاصل در صنعت را در زمینهٔ آموزش و پرورش به دست آورد (مرفی و دیویدشوفر، ۲۰۰۱). از جمله مشکلات وابسته به کاربرد این روش در آموزش و پرورش شاخصهای روایی و پایایی است که رضایت‌بخش نبوده‌اند[1].

جواب این سؤال را که "چرا روش کارپوشه موفقیت مورد انتظار را در زمینه‌های آموزشی به دست نیاورده در حالی که در صنعت پیشرفت قابل ملاحظه‌ای داشته است؟" باید در ملاکهای موفقیت در این دو زمینه جستجو کرد. کامارا[2] و براون[3] (۱۹۹۵) گفته‌اند که هدفهای آموزشی غالباً مبهم و تفسیرپذیرند، در حالی که در موقعیتهای صنعتی ملاکهای داوری روشن‌ترند و با دقت بیشتری تعریف می‌شوند. در واقع، شرایط استفاده از روش کارپوشه که داشتن هدفهای دقیق و ملاکهای داوری به دقت تعریف شده است در آموزش و پرورش تحقق نیافته‌اند. بنابراین، "مادام که متغیرهای ارزشیابی آموزشی به صورت بسیار روشن‌تری از آنچه فعلاً هست تعریف نشوند، به کسب نتایج دارای روایی و پایایی از کاربرد روشهای سنجش عملکرد [از جمله روش کارپوشه] در آموزش و پرورش امید زیادی نیست" (مرفی و دیویدشوفر، ۲۰۰۱، ص ۴۱۶).

پوفام (۲۰۰۲) مشکل اصلی روش کارپوشه را به شیوهٔ نمره‌گذاری آنها نسبت داده است. به باور او، بعضی از دستورعمل‌ها یا راهنماییهایی که برای نمره‌گذاری کارپوشه تهیه می‌شوند

---

۱. مباحث مربوط به روایی و پایایی در فصلهای ۲۰ و ۲۱ کتاب مطرح شده‌اند.

2. Camara 3. Brown

بسیار کلی‌اند به گونه‌ای که "بی‌شباهت به لکه‌های جوهر رورشاخ[1] نیستند و هر تصحیح‌کننده‌ای هر آنچه را که خودش می‌خواهد در آنها می‌بیند" (ص ۲۱۰). در مقابل، بعضی از دستورعمل‌های دیگر تصحیح کارپوشه آن‌قدر پیچیده و پرطول و تفصیل‌اند که مصحح را گیج می‌کنند.

تهیهٔ راهنماها یا دستورعمل‌هایی که در حد نیاز تفصیلی باشند (نه خیلی مختصر و نه خیلی طولانی) کار دشواری است. به طور کلی، غالب معلمان آن‌قدر گرفتارند که وقت تهیهٔ راهنماهای دقیق و کارآمد را ندارند. در نتیجه، بسیاری از معلمان با استفاده از ملاکهای ضعیف ارزشیابی به داوری دربارهٔ کارپوشه‌ها می‌پردازند (پوفام، ۲۰۰۲).

مشکل دیگر سنجش با کارپوشه وقت زیادی است که صرف آن می‌شود. معلم هم در کلاس و هم در خارج از کلاس باید مقدار قابل ملاحظه‌ای از وقت خود را صرف بررسی و ارزشیابی کارپوشه‌های دانش‌آموزان نماید. ولف (۲۰۰۶) می‌گوید "ساختن کارپوشه وقت‌گیر، انبار کردن آن پرزحمت، و نمره‌گذاری‌اش دشوار است" (ص ۱۶۸). البته طرفداران این روش می‌گویند این روش ارزش صرف آن همه وقت را دارد. به هر حال، معلمانی که می‌خواهند از این روش ارزشیابی استفاده نمایند باید خود را برای صرف وقت نسبتاً زیاد آماده کنند.

بازهم مشکل دیگر سنجش با کارپوشه و به طور کلی مشکل همهٔ آزمونهای عملکردی این است که آگاهی از عملکرد دانش‌آموزان در یک تکلیف خاص دربارهٔ عملکرد آنان در تکالیف دیگر اطلاعات زیادی در اختیار ما نمی‌گذارد. این مشکلِ نمونه‌گیریِ محتوایی جدی‌ترین خطر کاهش روایی و پایایی روشهای سنجش عملکردی است. بنابراین، "ممکن است لازم باشد که شش یا هفت تکلیف عملکردی مورد استفاده قرارگیرند تا ارزشیابی نسبتاً قابل اعتمادی از پیشرفت دانش‌آموز در یک واحد آموزشی به عمل آید" (نیتکو، ۲۰۰۱، ص ۲۶۲).

با توجه به اینکه هر روشی دارای محاسن و معایب خاص خودش است، هیچ روشی کامل نیست و نباید از آن برای سنجش همهٔ هدفها استفاده کرد. هر یک از روشهایی که در این کتاب معرفی شده ارزش خودش را دارد و برای سنجش هدفهای خاصی قابل استفاده است. اگر قصد معلم سنجش دانش امور واقعی و یادآوری آن توسط دانش‌آموزان است، بهتر است از آزمونهای عینی استفاده کند. اگر می‌خواهد درک و فهم یادگیرنده را در یک موضوع خاص و در یک مقطع زمانی بسنجد، بهتر است از آزمونهای تشریحی یا سنجش عملکردی استفاده کند. اگر می‌خواهد هم پیشرفت یا موفقیت و هم رشد موفقیتِ یادگیرندگان را در موقعیتهای واقعی و در یک طول زمان سنجش کند بهتر است از کارپوشه استفاده نماید. اما از آنجاکه معلمان همواره در آموزش خود با مجموعه‌ای از هدفهای یادگیری سروکار دارند، اغلب لازم می‌آید که از ترکیب چند روش استفاده کنند.

---

1. Rorschach inkblots

موارد استفادهٔ روش کارپوشه

در آغاز این فصل گفتیم که سنجش با استفاده از کارپوشه یکی از روشهای سنجش عملکردی است. بنابراین، هر نوع یادگیری که با سایر روشهای سنجش عملکردی قابل سنجش است می‌تواند با روش کارپوشه نیز مورد سنجش قرار داد. با این حال، کارپوشه نسبت به سایر روشهای عملکردی دارای امتیاز ویژه‌ای است که آن را برای مقصودهای خاصی استفاده‌پذیر می‌کند. ما به کمک روشهای مختلف سنجش عملکردی و نیز آزمونهای تشریحی (به ویژه آزمونهای تشریحی گسترده پاسخ) می‌توانیم درک و فهم یادگیرنده و فعالیتهای پیچیده‌تر شناختی را بسنجیم، اما این سنجش به مقاطع زمانی خاصی محدود می‌شود. در مقابل، با استفاده از روش کارپوشه همین کار را می‌توانیم در طول زمان انجام دهیم. یعنی اگر بخواهیم هم موفقیت یا پیشرفت یادگیرنده و هم رشد موفقیت او را در طول یک زمان معین بسنجیم بهترین روش ممکن روش کارپوشه است. در زیر به نمونه‌ای از کاربست روش کارپوشه برای سنجش فعالیتهای متنوع و پیچیده توجه کنید.

نمونه‌ای از چگونگی تشکیل و استفاده از کارپوشه

نمونه‌ای که در اینجا معرفی می‌شود فرایند شکل‌گیری و استفاده از یک سنجش واقعی مبتنی بر روش کارپوشه را نشان می‌دهد. در این نمونه، کارپوشه برای مقاصد سنجش و یادگیری در سال اول یک دورهٔ دوسالهٔ تربیت معلم در کشور انگلستان به کار رفته است[1].

در تربیت معلمان آینده فرضهایی زیربنای فعالیتهای مربیان بوده که مهم‌ترین آنها از قرار زیرند: (۱) معلمان نیاز دارند که رابطهٔ بین سنجش و حلقهٔ آموزش و یادگیری را به خوبی درک نمایند؛ (۲) معلمان نیاز دارند از آن دسته روشهای آموزش و سنجش استفاده کنند که دانش‌آموزان را وامی‌دارند تا در یادگیری خود فعالانه شرکت کنند و برای یادگیریهای خود مسئولیت بپذیرند؛ (۳) برای اینکه معلمان آینده این روشها را یاد بگیرند و بتوانند آنها را با دانش‌آموزان به کار بندند نیاز دارند که آن روشها را تجربه کنند.

ایجاد کارپوشهٔ فعالیتهای دانشجویان تربیت معلم مورد بحث بخشی از فعالیتهای آنان در دو درس زیر است: (۱) مهارتهای آموزشی کلاسی و (۲) طراحی آموزشی و راهبردهای آموزشی اثربخش. درس اول بیشتر جنبهٔ عملی ولی درس دوم عمدتاً حالت نظری دارد. هر دو درس به صورت تراکمی یا مجموعی و به کمک کارپوشه سنجش شدند. با این حال، کارپوشه‌ها در ضمن ترم تحصیلی به صورت تکوینی ساخته و سنجش شدند تا اینکه به یادگیری دانشجویان در جریان شکل‌گیری کمک شود.

---

۱. ما مطالب این قسمت را از کتاب ایجاد کارپوشه برای یادگیری و سنجش تألیف کلنوسکی (۲۰۰۲) خلاصه و نقل کرده‌ایم.

برای درس مهارتهای آموزشی کلاسی هدفهای زیر منظور شده بود:

- به کار بستن اصول ارتباطی
- نشان دادن مهارتهای تدریس
- توضیح دادن ساخت و محتوای یک طرح درس
- نوشتن طرح درسهای تفصیلی و برنامهٔ کار
- ارزشیابی طرح درسها
- توضیح دادن راهبردهای مختلف آموزشی (سخنرانی، بحث گروهی، آموزش اکتشافی، و غیره)
- بیان دلایل بدرفتاری دانش‌آموزان
- از خود نشان دادن مهارتهای کلاسداری

برای درس طراحی آموزشی و راهبردهای آموزشی اثربخش هدفهای زیر منظور شده بود:

- بیان فلسفهٔ شخصی، هدفها، و ارزشها
- ارزشیابی از آموزش خود به منظور بهبود بخشیدن به آن
- ایجاد انتظارات روشن، چالش‌انگیز، و قابل دستیابی برای دانش‌آموزان
- آموزش دادن به دانش‌آموزان متناسب با استعدادها و علاقه‌های مختلف آنان
- طراحی درس براساس جدیدترین نظریه‌های یادگیری
- ایجاد ارتباط میان هدفهای آموزشی، محتوای درس، و روشهای آموزش و سنجش
- استفادهٔ بهینه از روشهای گروهبندی دانش‌آموزان
- ایجاد انگیزه در دانش‌آموزان و تشویق آنان به یادگیری
- واکنش مناسب به موقعیتهایی که به تصادف رخ می‌دهند

به این دانشجویان دورهٔ تربیت معلم گفته شد که لازم است نمونه کارهایی را جمع‌آوری کنند که نشان‌دهندهٔ دستیابی آنان به مهارتها و شایستگیهای موجود در هدفهای فوق باشد. بدین منظور، دانشجویان راهنمایی شدند و نمونه کارهایی که می‌توانستند در کارپوشهٔ خود قرار دهند به آنها داده شد. مثلاً برای هدف "بیان فلسفهٔ شخصی، هدفها، ارزشها" موارد زیر پیشنهاد شد:

- یک نقشهٔ مفهومی که خلاصه‌ای از باورهای شما را به عنوان یک معلم نشان دهد.
- یک نوار ویدیویی که ارزشهای شما را آن‌گونه که در راهبردهای آموزشی و یادگیری مورد نظرتان منعکس است نشان دهد.
- یک مقاله که باورهای اصلی شما را به عنوان یک معلم به روشنی نشان دهد.

- یک نمایش تصویری از فعالیتهای آموزشی یا یادگیری که با توضیح هدف شما را به عنوان یک معلم نشان می‌دهد.

علاوه بر شواهد و مدارکی نظیر آنچه در بالا به آنها اشاره شد، معلمان آینده یک مطلب نوشتنی مستدل نیز باید ضمیمه می‌کردند. به آنها گفته شد که این نوشته باید چگونگی اندیشهٔ آنان را در رابطه با اینکه مطالب گنجانیده شده در کارپوشه به چه نحوی شایستگی آنان را نشان می‌دهد تحلیل و تفسیر کنند. در بخشی از نوشتهٔ یکی از آن دانشجومعلمان آمده است،

به عنوان یک معلم، من برای تفکر انتقادی ارزش قائلم و بر آن تأکید می‌کنم. این نوع آموزش مبتنی بر تفکر نه تنها به دانش‌آموزان موضوعات درسی را یاد می‌دهد بلکه به آنها می‌آموزد که در تمام زمینه‌های زندگی خود به گزینشهای عاقلانه دست بزنند.

باور من بر این است که هر دانش‌آموزی، سوای پیشرفت تحصیلی‌اش، دارای بعضی ویژگیهای خوب است. بنابراین، برای اینکه یک معلمِ مسئول و کمک‌رسان باشم باید روشهای آموزشی‌ام را با سبکهای یادگیری دانش‌آموزانم تطبیق دهم و آنان را برای یادگیری تشویق کنم و به شکوفایی استعدادهای آنان کمک نمایم (ممکن است این کار با روشهای معمول آموزشی انجام نگیرد) [به نوار آموزشی ویدیویی من مراجعه کنید. در آنجا سعی کرده‌ام که راهبردهای آموزشیم را با سبکهای یادگیری دانش‌آموزانم هماهنگ سازم.]

باور شخصی دیگر من این است که همیشه بهترین دانش‌آموز کسی نیست که بیشترین موفقیت تحصیلی را کسب می‌کند. بنابراین، من ترجیح می‌دهم که دانش‌آموزانم را به صورت انفرادی هم در زمینه‌های تحصیلی و هم در سایر زمینه‌ها سنجش کنم. برای نمونه، من برای دانش‌آموزی که از لحاظ تحصیلی خیلی ممتاز نیست اما برای موفقیت سخت می‌کوشد در مقایسه با دانش‌آموز بااستعدادی که کوشش زیادی برای موفقیت نمی‌کند و از دانش‌آموز اولی نمره‌های قدری بهتری می‌گیرد ارزش بیشتری قائلم. به اعتقاد من، اشتیاق دانش‌آموز برای یادگیری و کوششی که به خرج می‌دهد (فرایند) مهم‌تر از نمره‌ای است که می‌گیرد (بازده). چنانچه برای فرایند ارزش قائل نشویم، آنگاه آموزش و پرورش ــ از این لحاظ که دانش‌آموزان صرفاً به صورت ماشینی آنچه را از آنان خواسته می‌شود انجام خواهند داد و به فرایند یادگیری بی‌توجه خواهند بود ــ به صورت یک فعالیت بی‌معنی در خواهد آمد. (برگرفته از کارپوشهٔ یک دانشجو ـ معلم، به نقلِ کلنوسکی، ۲۰۰۲، ص ۹۲)

در آغاز درس و پس از توضیح اینکه ارزشیابی از پیشرفت دانشجویان تربیت معلم به صورت سنجش عملکردی و براساس کارپوشه‌ای که تشکیل خواهند داد صورت خواهد گرفت، چگونگی نمره گذاری فعالیتها برایشان توضیح داده شد. این ارزشیابی براساس هدفهای آموزشی از پیش تعیین شده یعنی ارزشیابی وابسته به ملاک بود و ملاکهای نمره گذاری به شرح شکل ۳ـ۱۳ در اختیار آنان گذاشته شد.

نمرهٔ الف ـ ممتاز

- شواهد موجود در کارپوشه دارای کیفیت استثنایی و فراتر از هدفهای درس است.
- توضیحات دانشجو نشان‌دهندهٔ تفکر ژرف و اصیل است. استدلال منطقی به کار رفته و بسیار مستدل است.
- کار به خوبی سازمان یافته است، با ابتکار ارائه شده، و جنبهٔ تکراری ندارد.
- نمره‌ای که به او تعلق می‌گیرد نشان‌دهندهٔ اعتماد به توانایی او و اطمینان از این است که او به عنوان یک یادگیرندهٔ مستقل پیشرفت خواهد کرد.

نمرهٔ ب ـ موفق

- شواهد انتخاب شده برای کارپوشه هدفهای درس را برآورده می‌سازد و نشان‌دهندهٔ درک صحیح دانشجو از محتوای درس است.
- توضیحات دانشجو دیدگاههای او را تأیید می‌کنند و استدلال او مستدل است.
- کار به خوبی سازمان یافته است، روان نوشته شده، و در آن منابع به طور صحیح ذکر شده است.
- نمره‌ای که به او تعلق می‌گیرد نشان‌دهندهٔ اعتماد به توانایی او و اطمینان از این است که او با قدری هدایت و کمک پیشرفت خواهد کرد.

نمرهٔ ج ـ قبول

- شواهد جمع‌آوری شده برای کارپوشه در حد هدفهای درس و نشان‌دهندهٔ یادگیری کافی محتوای درس است، اما در کمترین درجهٔ پذیرفته‌شدن قرار دارد.
- در توضیحات دانشجو تفکر مستقل دیده نمی‌شود و اندک میزانی از تحلیل و تفسیر در آن موجود است.
- کار از لحاظ سازمان‌بندی نقص دارد. زبان آن نسبتاً روان است، اما از لحاظ دستوری و انشایی خالی از غلط نیست.
- نمرهٔ دانشجو نشان‌دهندهٔ اطمینان به توانایی او در پیشرفت بیشتر است، به شرط اینکه هدایت شود و به او کمک داده شود.

نمرهٔ دال ـ مردودی مشروط

- شواهد و مدارک جمع‌آوری شده برای کارپوشه هدفهای درس را برآورده نمی‌سازند.
- توضیحات دانشجو صرفاً شامل مطالب حفظ شده است و مبتنی بر تفکر و تحقیق نیست.
- کار از سازمان‌بندی وانسجام برخوردار نیست.
- نمره حاکی از این است که دانشجو ممکن است مجدداً خود را برای ارزشیابی آماده کند اما انجام کار بهتر برای او دشوار خواهد بود. (اگر به دانشجو فرصت سنجش دوباره داده شود شایستگی نمرهٔ بالاتر از ج (قبولی) را نخواهد داشت.)

نمرهٔ هـ ـ مردودی قطعی

- شواهد و مدارک موجود در کارپوشه هدفهای درس را برآورده نمی‌سازند و درک دانشجو از محتوای مطالب را نشان نمی‌دهند.
- توضیحات دانشجو صرفاً ادعاهایی را مطرح می‌کند که پشتوانهٔ تحقیقی و استدلالی ندارند.
- کار کمترین حد پذیرفته‌شدن را دارا نیست و در بیانات دانشجو اشتباهات جدی و تکراری وجود دارد. کار انجام گرفته در جهت خواسته‌های درس نبوده و از هیچ نظر قابل پذیرش نیست.

شکل ۳ـ۱۳ راهنمای نمره‌گذاری کارپوشهٔ دانشجویان تربیت معلم (به نقلِ کلنوسکی، ۲۰۰۲، ص ۹۴)

براساس اظهارات استادان و تحلیل نتایج سنجش عملکردی دانشجویان تربیت معلم، استفاده از روش کارپوشه، علاوه بر یادگیری هدفهای درس، بر شیوهٔ یادگیری دانشجویان نیز تأثیر چشم‌گیری داشته است. تأثیرهای مثبت این روش را می‌توان به شرح زیر برشمرد:

- مطالعهٔ ژرف
- یادگیری فعال
- یادگیری مستقل
- یادگیری مبتنی بر سرعت فردی و پیشرفت گرا
- آگاهی از یادگیری خود
- انگیزش برای مشارکت
- افزایش توجه
- علاقه‌مند شدن به عملکرد و پیشرفت خود
- مشورت با استاد در رابطه با یادگیری خود
- علاقه به نحوهٔ نمره گذاری و فرایند سنجش

اظهارات دانشجویان تربیت معلم نیز که با آنها روش سنجش براساس کارپوشه مورد استفاده قرار گرفت نشان‌دهندهٔ تأثیر مثبت این روش است. از جمله نکات مثبت اظهار شده از سوی دانشجویان موارد زیر بوده‌اند:

- من دربارهٔ چگونگی تفکرم تغییرات چشم‌گیری در خودم می‌بینم. من اکنون این درس را به گونهٔ دیگری تصور می‌کنم.
- من از کارپوشه بیشتر از درس یاد گرفتم. انجام دادن این تکلیف کار دشواری بود، اما واقعاً به من کمک کرد تا یادگیری و اندیشه‌ام را سازمان دهم.

خلاصهٔ فصل

کارپوشه به مجموعه‌ای از کارهای دانش‌آموز یا دانشجو گفته می‌شود که کوشش، پیشرفت، و موفقیت تحصیلی او را در یک زمینهٔ خاص نشان می‌دهد. کارپوشه یکی از روشهای سنجش عملکردی است که به کمک آن هم می‌توان چگونگی پیشرفت و هم جریان رشد پیشرفت یادگیرنده را ارزشیابی کرد.

کارپوشه‌ها با توجه به نوع هدفی که برای تحقق آن درست می‌شوند متفاوت‌اند. ما در این فصل کارپوشه‌ها را با توجه به ۱۰ هدف زیر توضیح داده‌ایم:

۱. استفاده از کارپوشه برای آموزش
۲. استفاده از کارپوشه برای سنجش

۳. استفاده از کارپوشه برای آموزش و سنجش مهارتهای فکری و راهبردهای یادگیری
۴. استفاده از کارپوشه برای فارغ‌التحصیل شدن
۵. استفاده از کارپوشه برای نشان دادن موفقیتهای جاری و رشد تحصیلی دانش‌آموز
۶. استفاده از کارپوشه برای نمایش دادن کارها
۷. استفاده از کارپوشه برای ارائهٔ اسناد و مدارک
۸. استفاده از کارپوشه برای نشان دادن کارهای به ثمر رسیده و اقدامات در حال انجام
۹. استفاده از کارپوشه برای گزینش و تأیید صلاحیت
۱۰. استفاده از کارپوشه برای تشویق و ارتقاء کارکنان

محتوای هر کارپوشه را عمدتاً هدفهای آن تعیین می‌کند. اما کارپوشه‌ها را می‌توان با استفاده از فعالیتها و دستاوردهای مختلف یادگیرندگان نیز تشکیل داد. از جملهٔ آنها می‌توان به موارد زیر اشاره کرد: کارهای نوشتنی، نوارهای صوتی و تصویری، کارهای هنری، اظهارنظرهای معلم، پوستر، مصاحبه، شعر، نتایج آزمونها، حل مسائل، خودسنجی‌ها، و هرگونه کار دیگری از دانش‌آموز که مهارتها و پیشرفتهای او را نشان دهد. در ضمن، کارپوشه‌ها را می‌توان به صورت برگه‌های کاغذ و عکسهای مختلف و بر روی نوارهای صوتی و تصویری، دیسکت‌های کامپیوتری یا دیسک‌های فشرده تشکیل داد.

صاحب‌نظران سنجش و اندازه‌گیریِ فرایندها و فراورده‌های یادگیری دانش‌آموزان و دانشجویان برای تهیه، نمره‌گذاری، و استفادهٔ بهینه از کارپوشه‌های مختلف قواعد و رهنمودهایی به دست داده‌اند. ما در این فصل آن رهنمودها را به تفصیل توضیح داده‌ایم.

در ارتباط با امتیازها و محدودیتهای روش سنجش با استفاده از کارپوشه موردهای مختلفی را می‌توان بیان کرد. علاوه بر امتیازهای کلی سنجش عملکردی که به روش کارپوشه هم مربوط می‌شوند، امتیازهای خاص این روش به قرار زیرند:

۱. به یادگیرندگان کمک می‌کند تا مسئولیت انتخاب هدف را برای خودشان بپذیرند و بر جریان پیشرفت نظارت داشته باشند.
۲. با این روش هم می‌توان سنجش کرد و هم آموزش داد.
۳. به کمک روش کارپوشه، هم می‌توان موفقیت یادگیرندگان را سنجش کرد و هم جریان رشد موفقیت آنان را.

در مقابل امتیازهای بالا، روش کارپوشه دارای محدودیتهایی به شرح زیر است:

۱. ارزشیابی و نمره‌گذاری کارپوشه با دشواری صورت می‌پذیرد و نتایج آن از پایایی لازم برخوردار نیستند.

۲. تهیهٔ کارپوشه‌ها وقت زیادی می‌گیرد.
۳. با این روش سنجش، نمونهٔ درستی از یادگیریهای دانش‌آموزان را نمی‌شود سنجش کرد.

موارد استفادهٔ روش کارپوشه فراوان‌اند و تمام آنچه دربارهٔ روشهای سنجش عملکردی در فصل ۱۲ گفته شد دربارهٔ روش کارپوشه نیز صادق است.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. مهم‌ترین فایدهٔ کارپوشه این است که به کمک آن هم می‌توان سنجش کرد و هم آموزش داد.
الف ـ صحیح
ب ـ غلط

۲. کارپوشه‌هایی که با هدف آموزش به یادگیرندگان مورد استفاده قرار می‌گیرند بر چه چیزی تأکید می‌کنند؟
الف ـ کارهای عملی
ب ـ فرایندهای عالی حافظه
پ ـ راهبردهای یادگیری
ت ـ ارزشیابی از خود
ث ـ انتقال آموخته‌ها

۳. روش سنجش به کمک کارپوشه را تعریف کنید و آن را با سایر روشهای عملکردی مقایسه نمایید.

۴. هدفهای مختلف تهیهٔ کارپوشه‌ها را توضیح دهید.

۵. توضیح دهید که محتوای کارپوشه‌ها از چه موادی تشکیل می‌شود، کارپوشه‌ها به چه صورتی ارائه می‌شوند، و چه چیزی تعیین می‌کند که محتوای یک کارپوشه چه مواد و مطالبی را دربر بگیرد.

۶. قواعد تهیه و استفاده از کارپوشه را توضیح دهید.

۷. امتیازها و محدودیتهای روش سنجش به کمک کارپوشه را توضیح دهید.

۸. در رابطه با جدول مشخصات آزمون درسی که به عنوان تمرین برای فصل ۶ کتاب تهیه کرده‌اید، برای هدفهایی از آن درس که با روش کارپوشه قابل سنجش‌اند، راهنماییهای لازم را در مورد چگونگی تهیهٔ کارپوشه‌های مناسب بنویسید.

۱۴

سنجش مشاهده‌ای

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. فهرست وارسی را تعریف کند.
۲. ویژگیهای فهرست وارسی را توضیح دهد.
۳. چگونگی تهیهٔ فهرست وارسی را توضیح دهد.
۴. نمره گذاری فهرست وارسی را توضیح دهد.
۵. به منظور سنجش فراورده و فرایند رفتاری دانش‌آموزان، فهرستهای وارسی مورد نیاز روشهای سنجش مشاهده‌ای را تهیه کند.
۶. مقیاس درجه‌بندی را تعریف کند.
۷. ویژگیهای مقیاس درجه‌بندی را توضیح دهد.
۸. با ذکر مثال، انواع مقیاسهای درجه‌بندی را توضیح دهد.
۹. با ذکر دلیل، بهترین نوع مقیاس درجه‌بندی را توضیح دهد.
۱۰. به منظور سنجش فراورده و فرایند رفتاری دانش‌آموزان، مقیاسهای درجه‌بندی مورد نیاز روشهای سنجش مشاهده‌ای را تهیه کند.
۱۱. نمره گذاری مقیاس درجه‌بندی را توضیح دهد.
۱۲. روش واقعه‌نگاری را تعریف کند.
۱۳. چگونگی انجام واقعه‌نگاری و تهیهٔ گزارش آن را شرح دهد.
۱۴. رفتارهای تحصیلی یا اجتماعی دانش‌آموزان را در موقعیتهای مختلف مورد مشاهده و واقعه‌نگاری قرار دهد.
۱۵. امتیازها و محدودیتهای روشهای سنجش مشاهده‌ای را شرح دهد.
۱۶. موارد استفادهٔ روشهای سنجش مشاهده‌ای را توضیح دهد.

سنجش و نمره‌گذاری همهٔ عملکردهای یادگیرندگان (دانش‌آموزان و دانشجویان)، به نحوی، بر مشاهده استوار است. چه قصد معلم سنجش فرایند رفتار دانش‌آموز باشد چه فراورده‌ٔ آن، باید آن رفتار را ابتدا مشاهده کند و بعد دربارهٔ آن تصمیم بگیرد؛ یعنی آن را ارزشیابی کند. برای داوری دربارهٔ کم و کیف عملکرد یادگیرندگان، از فنون گوناگونی استفاده می‌شود که بسیاری از آنها را در فصلهای گذشتهٔ این کتاب توضیح دادیم. در این فصل نیز چند روش دیگر را که به روشهای **سنجش مشاهده‌ای**[1] شهرت یافته‌اند توضیح می‌دهیم. معلم یا سنجشگر می‌تواند به کمک آنها مشاهدات خود را نظم و ترتیب ببخشد و برای ویژگیهای رفتارهای مورد سنجش نمره یا ارزش قائل شود. معروف‌ترین روشهای سنجش مشاهده‌ای عبارت‌اند از: **فهرست وارسی**، مقیاس درجه‌بندی، و واقعه‌نگاری که در این فصل به ترتیب آنها را توضیح می‌دهیم.

فهرست وارسی

تعریف فهرست وارسی

منظور از **فهرست وارسی**[2] فهرستی از رفتارها، ویژگیها، فعالیتها، یا تولیدات یادگیرنده است همراه با جاهایی در مقابل ماده‌های فهرست برای مشخص کردن (چک کردن) اینکه هر یک از ماده‌های آن فهرست حضور دارند یا نه. کوهن و سوردلیک (۲۰۰۵) فهرست وارسی را به عنوان "مجموعه پرسشهایی که به سنجشگر یا ارزشیاب امکان می‌دهند تا حضور یا عدم حضور رفتارها، اندیشه‌ها، رویدادها، یا کیفیتها را تعیین کنند" (ص ۳۰۱) تعریف کرده‌اند. روش فهرست‌وارسی یک شیوهٔ نظامدار (سیستماتیک) برای گزارش داوریهای مشاهده‌گر یا مشاهده‌گران است. با این روش می‌توان معلوم نمود که در عملکرد شخص مورد سنجش (یا محصولی که او تولید کرده است) ویژگیهای مورد نظر موجودند یا نه. کوپر[3] (۱۹۹۴) دربارهٔ ویژگیهای مورد نظری که قرار است با روش فهرست وارسی سنجش شوند مهم بودن را مورد تأکید قرار داده و گفته است "از خود بپرسید که آیا ویژگی یا ویژگیهایی از عملکرد یا فرآوردهٔ مورد نظر آن قدر مهم هستند که لازم است بود یا نبود آنها دانسته شود؟ اگر جواب این سؤال مثبت است، آنگاه روش فهرست‌وارسی همان چیزی است که به دنبالش هستید" (ص ۳۴۵).

در روش فهرست وارسی هم شیوهٔ انجام عمل یا فرایند و هم محصول عمل یا فراورده سنجش می‌شود. فهرست وارسیِ ویژهٔ سنجش فرایند یا شیوهٔ انجام عمل مشخص می‌کند که دانش‌آموز یا دانشجو گامهای ضروری و مناسب را در انجام یک عمل یا اجرای یک تکلیف برمی‌دارد یا نه. فهرست وارسیِ ویژهٔ فراورده یا محصول تعیین می‌کند که نتیجه یا حاصل کار و فعالیت دانش‌آموز از ویژگیهای مطلوب و مورد نظر برخوردار است یا نه.

1. observational assessment 2. checklist 3. Cooper

در روش فهرست وارسی تنها بود یا نبود یا کمیت ویژگی مورد نظر ثبت می‌شود؛ بنابراین، این روش را برای سنجش درجه یا فراوانی ویژگی مورد نظر نمی‌توان مورد استفاده قرار داد. به سخن دیگر، روش فهرست وارسی تنها تعیین می‌کند که یک ویژگی حضور دارد یا نه، یا اینکه عملی انجام گرفته است یا نه. در این روش، قضاوت دربارهٔ عملکرد یا فراورده به صورت بله ـ نه مشخص می‌شود. در زیر نمونهٔ عملکردها و فراورده‌هایی که با روش فهرست وارسی قابل سنجش‌اند داده شده:

عملکردها	فرآورده‌ها
سخن گفتن	نقاشی
شرکت در بحث	رسم
هدایت یک بحث	نقشه
اجرای یک آزمایش	وسایلی که از چوب ساخته شده
حل کردن یک مسئلهٔ ریاضی	هنرهای دستی
مجسمه‌سازی	طراحی
کشیدن نقاشی	
نواختن آلات موسیقی	
آواز خواندن	

ویژگیهای فهرست وارسی

چنان‌که گفتیم، از روش فهرست وارسی هم می‌توان برای سنجش فرایند رفتار استفاده کرد و هم برای فراوردهٔ آن. هنگام تهیهٔ فهرست وارسی به منظور ارزشیابی از فرایند عملکرد، تأکید باید بر رفتارهای فرد باشد. مثلاً مراحلی که باید انجام گیرند در یک ردیف متوالی قرار داده می‌شوند و آن‌گاه مشاهده‌گر تعیین می‌کند که هر یک از آن مراحل انجام گرفته است یا نه. شکل ۱ـ۱۴ نمونه‌ای از فهرست وارسی برای سنجش فرایند عملکرد یا شیوهٔ انجام کار را نشان می‌دهد. در فهرست وارسی شکل ۱ـ۱۴ مجموعه رفتارها یا گامهای عملی مورد نیاز برای استفاده از یک دماسنج پزشکی در تعیین درجهٔ حرارت یک بیمار ارزشیابی می‌شود.

چنان‌که ملاحظه می‌شود، فهرست وارسی شکل ۱ـ۱۴ جریان کار یا فرایند عملکرد را مورد ارزیابی قرار می‌دهد. شکل ۲ـ۱۴ نمونه‌ای از فهرست وارسی برای ارزشیابی از فراورده یا محصول رفتار است.

هنگام تهیهٔ فهرست وارسی برای ارزشیابی از محصول یا فراوردهٔ رفتار بر جنبه‌ها یا ویژگیهای قابل مشاهدهٔ محصول مورد سنجش تأکید می‌شود. بنابراین، فهرست وارسیِ مخصوص فراورده شامل فهرستی از ابعادی است که یک محصول خوب را نشان می‌دهد، مانند

راهنمایی: در جای خالی سمت راست هر سطر یا هر سؤال، برای رفتار رضایت‌بخش علامت (+) و برای رفتاری که رضایت‌بخش نیست علامت (-) بگذارید.

۱. با گرفتن انتهای دماسنج، آن را از توی محفظه بیرون می‌آورد.
۲. با کشیدن تمیزکننده از انتهای دماسنج به سمت نوک آن، دماسنج را تمیز می‌کند.
۳. با گرفتن انتهای دماسنج به دست، آن را تکان می‌دهد تا پایین‌تر از ۳۶ درجه سانتی‌گراد را نشان دهد.
۴. یک لامپ‌گونهٔ دماسنج را به زیر زبان بیمار قرار می‌دهد.
۵. از بیمار می‌خواهد تا لبانش را ببندد، اما دماسنج را گاز نگیرد.
۶. دماسنج را برای ۳ دقیقه در دهان بیمار می‌گذارد.
۷. با گرفتن انتهای دماسنج در دست، آن را از دهان بیمار بیرون می‌آورد.
۸. دمای بدن را تا دو دهم درجه می‌خواند.
۹. بر روی چارت مخصوص بیمار درجهٔ بدن او را یادداشت می‌کند.
۱۰. دماسنج را تمیز می‌کند و آن را در محفظه‌اش قرار می‌دهد.

شکل ۱ـ۱۴ یک نمونه فهرست وارسی برای سنجش عملکرد: ارزشیابی استفادهٔ درست از دماسنج پزشکی (گرانلاند، ۱۹۸۸، ص ۹۴)

اندازه، رنگ، شکل، و مانند اینها. چنان که دیدیم، شکل ۲ـ۱۴ نمونه‌ای از یک فهرست وارسی برای ارزشیابی از محصول یا فراورده است. این فهرست وارسی برای ارزشیابی از آزمونهای چندگزینه‌ای درست شده است. ارزشیاب با جواب دادن به سؤالهای این فهرست وارسی تعیین می‌کند آزمونی را که یک معلم برای ارزشیابی از میزان آموخته‌های دانش‌آموزان یا دانشجویان خود تهیه کرده است، به عنوان یک محصول یا فراورده، تا چه اندازه از ویژگیهای مطلوب آزمونهای چندگزینه‌ای برخوردار است.

در فهرست وارسی تنها ویژگیهای مطلوب رفتار یا فراوردهٔ مورد نظر سنجش نمی‌شوند، بلکه می‌توان ویژگیهای نامطلوب را نیز منظور کرد. به سخن دیگر، به جای اینکه فقط معلوم نماییم که موضوع مورد سنجش دارای خصوصیات مطلوب هست یا نیست، می‌توانیم معلوم کنیم که اشتباهات یا نواقص در آن یافت می‌شوند یا نه. بدین منظور، فهرست وارسی دربرگیرندهٔ خطاها یا نواقص پدیدهٔ مورد سنجش خواهد بود. پین (۲۰۰۳) فهرست وارسی شکل ۳ـ۱۴ را به عنوان نمونه‌ای از فهرست وارسی ویژگیهای نامطلوب ذکر کرده است. این فهرست وارسی را یکی از معلمان حرفه و فن برای سنجش ایرادهای یک کار چوبی، پیش از انجام آخرین اقدامات بر روی آن، درست کرده است. پس، بنا به آنچه گفتیم، فهرست وارسی هم می‌تواند شامل ویژگیهای مطلوب موضوع مورد سنجش باشد، هم ویژگیهای نامطلوب آن، و هم ترکیبی از ویژگیهای مطلوب و نامطلوب.

در روش فهرست وارسی تنها بود یا نبود یا کمیت ویژگی مورد نظر ثبت می‌شود؛ بنابراین، این روش را برای سنجش درجه یا فراوانی ویژگی مورد نظر نمی‌توان مورد استفاده قرار داد. به سخن دیگر، روش فهرست وارسی تنها تعیین می‌کند که یک ویژگی حضور دارد یا نه، یا اینکه عملی انجام گرفته است یا نه. در این روش، قضاوت دربارهٔ عملکرد یا فراورده به صورت بله ـ نه مشخص می‌شود. در زیر نمونهٔ عملکردها و فراورده‌هایی که با روش فهرست وارسی قابل سنجش‌اند داده شده:

عملکردها	فرآورده‌ها
سخن گفتن	نقاشی
شرکت در بحث	رسم
هدایت یک بحث	نقشه
اجرای یک آزمایش	وسایلی که از چوب ساخته شده
حل کردن یک مسئلهٔ ریاضی	هنرهای دستی
مجسمه‌سازی	طراحی
کشیدن نقاشی	
نواختن آلات موسیقی	
آواز خواندن	

ویژگیهای فهرست وارسی

چنان‌که گفتیم، از روش فهرست وارسی هم می‌توان برای سنجش فرایند رفتار استفاده کرد و هم برای فراوردهٔ آن. هنگام تهیهٔ فهرست وارسی به منظور ارزشیابی از فرایند عملکرد، تأکید باید بر رفتارهای فرد باشد. مثلاً مراحلی که باید انجام گیرند در یک ردیف متوالی قرار داده می‌شوند و آن‌گاه مشاهده‌گر تعیین می‌کند که هر یک از آن مراحل انجام گرفته است یا نه. شکل ۱ـ۱۴ نمونه‌ای از فهرست وارسی برای سنجش فرایند عملکرد یا شیوهٔ انجام کار را نشان می‌دهد. در فهرست وارسی شکل ۱ـ۱۴ مجموعه رفتارها یا گامهای عملی مورد نیاز برای استفاده از یک دماسنج پزشکی در تعیین درجهٔ حرارت یک بیمار ارزشیابی می‌شود.

چنان‌که ملاحظه می‌شود، فهرست وارسی شکل ۱ـ۱۴ جریان کار یا فرایند عملکرد را مورد ارزیابی قرار می‌دهد. شکل ۲ـ۱۴ نمونه‌ای از فهرست وارسی برای ارزشیابی از فراورده یا محصول رفتار است.

هنگام تهیهٔ فهرست وارسی برای ارزشیابی از محصول یا فراوردهٔ رفتار بر جنبه‌ها یا ویژگیهای قابل مشاهدهٔ محصول مورد سنجش تأکید می‌شود. بنابراین، فهرست وارسی مخصوص فراورده شامل فهرستی از ابعادی است که یک محصول خوب را نشان می‌دهد، مانند

راهنمایی: در جای خالی سمت راست هر سطر یا هر سؤال، برای رفتار رضایت‌بخش علامت (+) و برای رفتاری که رضایت‌بخش نیست علامت (-) بگذارید.

۱. با گرفتن انتهای دماسنج، آن را از توی محفظه بیرون می‌آورد.
۲. با کشیدن تمیرکننده از انتهای دماسنج به سمت نوک آن، دماسنج را تمیز می‌کند.
۳. با گرفتن انتهای دماسنج به دست، آن را تکان می‌دهد تا پایین‌تر از ۳۶ درجه سانتی‌گراد را نشان دهد.
۴. یک لامپ‌گونهٔ دماسنج را به زیر زبان بیمار قرار می‌دهد.
۵. از بیمار می‌خواهد تا لبانش را ببندد، اما دماسنج را گاز نگیرد.
۶. دماسنج را برای ۳ دقیقه در دهان بیمار می‌گذارد.
۷. با گرفتن انتهای دماسنج در دست، آن را از دهان بیمار بیرون می‌آورد.
۸. دمای بدن را تا دو دهم درجه می‌خواند.
۹. بر روی چارت مخصوص بیمار درجهٔ بدن او را یادداشت می‌کند.
۱۰. دماسنج را تمیز می‌کند و آن را در محفظه‌اش قرار می‌دهد.

شکل ۱ـ۱۴ یک نمونه فهرست وارسی برای سنجش عملکرد: ارزشیابی استفادهٔ درست از دماسنج پزشکی (گرانلاند، ۱۹۸۸، ص ۹۴)

اندازه، رنگ، شکل، و مانند اینها. چنان که دیدیم، شکل ۲ـ۱۴ نمونه‌ای از یک فهرست وارسی برای ارزشیابی از محصول یا فراورده است. این فهرست وارسی برای ارزشیابی از آزمونهای چندگزینه‌ای درست شده است. ارزشیاب با جواب دادن به سؤالهای این فهرست وارسی تعیین می‌کند آزمونی را که یک معلم برای ارزشیابی از میزان آموخته‌های دانش‌آموزان یا دانشجویان خود تهیه کرده است، به عنوان یک محصول یا فراورده، تا چه اندازه از ویژگیهای مطلوب آزمونهای چندگزینه‌ای برخوردار است.

در فهرست وارسی تنها ویژگیهای مطلوب رفتار یا فراوردهٔ مورد نظرِ سنجش نمی‌شوند، بلکه می‌توان ویژگیهای نامطلوب را نیز منظور کرد. به سخن دیگر، به جای اینکه فقط معلوم نماییم که موضوع مورد سنجش دارای خصوصیات مطلوب هست یا نیست، می‌توانیم معلوم کنیم که اشتباهات یا نواقص در آن یافت می‌شوند یا نه. بدین منظور، فهرست وارسی دربرگیرندهٔ خطاها یا نواقص پدیدهٔ مورد سنجش خواهد بود. پین (۲۰۰۳) فهرست وارسی شکل ۳ـ۱۴ را به عنوان نمونه‌ای از فهرست وارسی ویژگیهای نامطلوب ذکر کرده است. این فهرست وارسی را یکی از معلمان حرفه و فن برای سنجش ایرادهای یک کار چوبی، پیش از انجام آخرین اقدامات بر روی آن، درست کرده است. پس، بنا به آنچه گفتیم، فهرست وارسی هم می‌تواند شامل ویژگیهای مطلوب موضوع مورد سنجش باشد، هم ویژگیهای نامطلوب آن، و هم ترکیبی از ویژگیهای مطلوب و نامطلوب.

راهنمایی: در جای خالی کنار هر جمله یا هر سؤال، با گذاشتن علامت (+) یا (-) تعیین کنید که آزمون مورد ارزشیابی از ویژگی موضوع آن سؤال برخوردار است یا نه.

۱. هر سؤال به روشنی موضوع مورد نظر را بیان می‌کند.
۲. موضوع اصلی سؤال در تنهٔ سؤال آمده است.
۳. بیشتر از یک مسئله یا یک مطلب در سؤال قرار ندارد.
۴. گزینه‌های هر سؤال از نظر دستوری و جمله‌بندی مکمل متن سؤال هستند.
۵. دو گزینهٔ متضاد که یکی از آنها درست است استفاده نشده.
۶. مطالب تکراری گزینه‌ها در تنهٔ سؤال آمده‌اند.
۷. هر سؤال یک موضوع مهم را اندازه می‌گیرد.
۸. گزینه‌های یک سؤال با هم تجانس دارند.
۹. در هر سؤال تنها یک گزینهٔ درست یا یک گزینه که از همهٔ گزینه‌ها بهتر است وجود دارد.
۱۰. در سؤالهایی که برای سنجش فرایندهای پیچیدهٔ ذهنی نوشته شده‌اند موقعیتهای تازه مطرح شده‌اند.
۱۱. گزینه‌های انحرافی آزمون‌شوندگان بی‌اطلاع از موضوع را به خود جلب می‌کنند.
۱۲. گزینه‌های هر سؤال از نظر دستوری و جمله‌بندی با تنهٔ سؤال مطابقت دارند.
۱۳. از کاربرد اشاره‌های نامربوط خودداری شده.
۱۴. در سؤالهای منفی زیر عبارات منفی خط کشیده شده.
۱۵. از کاربرد منفی مضاعف در نوشتن سؤالها خودداری شده.
۱۶. از کاربرد عبارتهایی نظیر "همهٔ آنها" و "همهٔ آنچه در بالا گفته شد" خودداری شده.
۱۷. سؤالها مستقل از یکدیگر هستند.
۱۸. از طرح سؤالهای گمراه کننده اجتناب شده.
۲۰. طول گزینهٔ درست در سؤالهای مختلف متغیر است.
۲۱. محل گزینهٔ درست در میان گزینه‌های انحرافی به‌طور تصادفی انتخاب شده.
۲۲. سؤالها کاملاً روشن و در حد درک و فهم آزمون‌شوندگان است.
۲۳. از گزینهٔ "هیچ‌یک از موارد بالا" استفاده نشده.
۲۴. جای خالی در انتهای تنهٔ سؤال قرار داده شده.

شکل ۲-۱۴ یک نمونه فهرست وارسی برای سنجش فراورده: ارزشیابی از آزمونهای چندگزینه‌ای

به طور خلاصه یک فهرست وارسی خوب از ویژگیهای زیر برخوردار است:

۱. فهرست نسبتاً کوتاه است.
۲. هر مادهٔ فهرست به طور کاملاً روشن موضوع مورد ارزشیابی را بیان می‌کند.
۳. هر ماده بر یک رفتار یا ویژگی قابل مشاهده تأکید می‌کند.
۴. تنها رفتارها یا ویژگیهای مهم منظور می‌شوند.
۵. ماده‌های فهرست به گونه‌ای آرایش می‌یابند که کل فهرست را به سادگی می‌توان مورد استفاده قرار داد.

راهنمایی: در جای خالی سمت راست هر یک از نقصهایی که در محصول مورد ارزشیابی یافت می‌شود علامت بگذارید.

ـ وجود گره
ـ پرنشدگی‌ها
ـ چسب زائد
ـ فاصلهٔ اتصالات
ـ نقص جلا
ـ لقی اتصالات
ـ اشکال پرداخت
ـ اشکال ابعاد
ـ گوشه‌های ناصاف
ـ ضربه دیدن

شکل ۳ـ۱۴ فهرست وارسی برای سنجش محصولات چوبی شامل نواقص و ایرادهای کار (پین، ۲۰۰۳، ص ۳۰۹)

قواعد تهیهٔ فهرست وارسی

تهیهٔ فهرست وارسی کار نسبتاً ساده‌ای است. کارهایی که باید انجام دهید این است که ابتدا فهرستی از رفتارها یا ویژگیهای مهم را تهیه کنید، و بعد غلطهای متداول را هم به آن بیفزایید. اخر سر اینکه ماده‌های فهرست را به گونه‌ای مرتب کنید که بتوانید به راحتی آنها را مورد استفاده قرار دهید. کوپر (۱۹۹۴) دربارهٔ افزودن غلطهای متداول به فهرست وارسی می‌گوید وجود این‌گونه غلطها در فهرست وارسی این وسیلهٔ سنجش را به صورت یک ابزار تشخیصی درمی‌آورد، و به کمک آن می‌توان اشکالهای عملکرد یا محصول فرد مورد سنجش را مشخص کرد.

برای تهیهٔ فهرست وارسیِ ویژهٔ سنجش شیوهٔ اجرای عمل یا انجام کار، لین و گرانلاند (۲۰۰۰) پیشنهادهای زیر را ارائه داده‌اند. رعایت این پیشنهادها بر کیفیت فهرستهای وارسی می‌افزاید.

۱. تمامی اعمال وصف شده در عملکرد را فهرست‌وار تعیین کنید.
۲. به فهرست خود اعمالی را که معرف غلطهای متداول هستند نیز اضافه کنید. (البته اگر در سنجش شما مفیدند، تعدادشان اندک است، و به روشنی قابل بیان هستند.)
۳. اعمال وصف شده (و غلطهای احتمالی) را به ترتیبی که انتظار وقوع آنها می‌رود مرتب کنید.
۴. یک روش ساده برای وارسی هر عمل هنگامی که به وقوع می‌پیوندد آماده کنید.

لین و گرانلاند (۲۰۰۰) گفته‌اند در حوزهٔ رشد شخصی ـ اجتماعی فهرست وارسی می‌تواند روش مناسبی برای ثبت شواهد مربوط به رشد فرد در جهت رسیدن به هدفهای یادگیری باشد. در این نوع فهرست وارسی، رفتارهایی که نشان‌دهندهٔ رفتار مورد سنجش‌اند فهرست می‌شوند. برای مثال، در زمینهٔ عادتهای کاری، یک معلم دبستان می‌تواند رفتارهای زیر را

فهرست کند و بود و نبود آنها را در دانش‌آموزان مورد سنجش قرار دهد.

- به دستورات عمل می‌کند.
- هر زمان به کمک نیاز داشته باشد کمک می‌طلبد.
- با دیگران همکاری می‌کند.
- در استفاده از مواد و وسایل نوبت را رعایت می‌کند.
- فعالیتهای تازه را امتحان می‌کند.
- کارهایی را که شروع می‌کند به انجام می‌رساند.
- وسایلی را که مورد استفاده قرار می‌دهد سر جایش برمی‌گرداند.
- مثل کارش را تمیز می‌کند.

(لین و گرانلاند، ۲۰۰۰، ص ۲۸۴)

دقت کنید که اگر صرفاً انجام دادن یا ندادن رفتارهای فوق مورد نظر باشد فهرست وارسی روش ساده و مفیدی است، اما اگر درجه یا فراوانی آنها را بخواهیم سنجش کنیم باید از مقیاس درجه‌بندی استفاده نماییم. مقیاس درجه‌بندی بعداً در این فصل توضیح داده می‌شود.

نمره‌گذاری فهرست وارسی

فهرست وارسی را می‌توان به چند صورت نمره گذاری کرد: ۰-۱، بود ـ نبود و بله ـ نه. در فهرستهای وارسی که ویژگیهای منفی یا اشتباهات را نیز شامل می‌شوند می‌توان برای ویژگیهای مثبت ۱+ و برای ویژگیهای منفی ۱- را منظور کرد.

در فهرست وارسی معمولاً برای هر ویژگی فهرست که از نظر معلم یا سنجشگر در رفتار یا محصول موردنظر به صورت مطلوب وجود دارد یک نمره (۱+) منظور می‌شود و نمرهٔ کل دانش‌آموز از مجموع این نمرات به دست می‌آید. با این حال، اگر بعضی عناصر یا ویژگیهای فهرست از ویژگیهای دیگر مهم‌ترند، می‌توان برای آنها نمرهٔ بیشتری در نظر گرفت. "احتمالاً دامنه‌ای از نمره‌های ۱ تا ۳ کفایت می‌کند" (پین، ۲۰۰۳، ص ۳۰۸).

مقیاس درجه‌بندی

تعریف مقیاس درجه‌بندی

در تعریف مقیاس درجه‌بندی[1] گفته شده که آن یک روش جمع‌آوری اطلاعات است که در آن اعداد یا عبارتهایی در ارتباط با نقاطی بر روی یک پیوستار مورد استفاده قرار می‌گیرند و به

1. rating scale

وسیلهٔ آن می‌توان دربارهٔ مقدار یا درجهٔ ویژگیها، نگرشها، عواطف، و سایر متغیرها داوری کرد (پین، ۲۰۰۳؛ کوهن و سوردلیک، ۲۰۰۵).

به خلاف روش فهرست وارسی که در آن تنها بود یا نبود ویژگی مورد نظر یا وقوع یا عدم وقوع رفتار مورد سنجش ثبت می‌شود، در مقیاس درجه‌بندی (درجه، مقدار، یا فراوانی رفتار یا ویژگی تعیین می‌گردد. مانند فهرست وارسی، مقیاس درجه‌بندی وسیله‌ای فراهم می‌آورد که به کمک آن می‌توان تمامی دانش‌آموزان یک کلاس یا گروهی از افراد را در ابعاد معینی مورد قضاوت قرار داد، و در آن داوریهای مربوط به تک‌تک افراد را ثبت نمود. گرانلاند و لین (۱۹۹۰) سه فایدهٔ زیر را برای روش مقیاس درجه‌بندی ذکر کرده‌اند:

۱. مشاهدات را به سوی جنبه‌های مشخص رفتار هدایت می‌کند.
۲. یک چهارچوب داوری برای مقایسهٔ همهٔ دانش‌آموزان در یک مجموعه ویژگی فراهم می‌آورد.
۳. روش مناسبی را برای ثبت داوریهای مشاهده‌گران فراهم می‌کند.

چنان‌که دربارهٔ روش فهرست وارسی صادق است، روش مقیاس درجه‌بندی نیز هم برای ارزشیابی از فرایند رفتار یا شیوهٔ اجرا مورد استفاده قرار می‌گیرد و هم برای ارزشیابی از محصول یا فراوردهٔ رفتار به کار می‌رود. شکل ۴ـ۱۴ نمونه‌ای از کاربرد مقیاس درجه‌بندی برای هر دو منظور را نشان می‌دهد. چنان‌که می‌بینید، در شکل ۴ـ۱۴ در قسمت **الف** عملکرد یا شیوهٔ اجرا سنجش می‌شود ولی در قسمت ب محصول یا فراوردهٔ رفتار سنجش می‌شود.

ویژگیهای مقیاس درجه‌بندی

چنان‌که در شکل ۴ـ۱۴ دیده می‌شود، مقیاسهای درجه بندی ساده از مجموعه‌ای رفتار یا ویژگی و توضیحاتی برای مشاهده‌گر تشکیل می‌یابند که از او می‌خواهند تا با مشخص کردن شماره‌های داده شده دامنه یا کیفیت رفتارهای مورد ارزشیابی را مشخص نماید. مقیاس درجه‌بندی به خودی خود یک تدبیر گزارش نتایج مشاهدات است. ارزش آن در سنجش یادگیری و رشد دانش‌آموزان عمدتاً وابسته به دقّتِ تهیهٔ آن است. مانند سایر روشهای سنجش، مقیاس درجه‌بندی نیز باید براساس هدفهای یادگیری ساخته شود، و استفاده از آن به زمینه‌هایی محدود گردد که برای آنها فرصت کافی مشاهده وجود دارد.

مقیاسهای درجه‌بندی معمولی از دو بخش تشکیل می‌شوند: (۱) مجموعه‌ای از متغیرهای محرک (چیزهایی که باید درجه‌بندی شوند)، و (۲) الگوی گزینه‌های پاسخ (درجه‌بندی که قرار است انجام پذیرد) (ثرندایک و هیگن، ۱۹۷۹). در ساده‌ترین و معمول‌ترین شکل مقیاس درجه‌بندی، متغیرهای محرک از نامهای ویژگیها یا صفات و گزینه‌های پاسخ از طبقه‌های

راهنمایی: با کشیدن دایره‌ای به دور یکی از شماره‌های سمت راست هر یک از جملات زیر، آن را درجه‌بندی نمایید. شماره‌ها به صورت زیر ارزش‌گذاری می‌شوند:
۵ = ممتاز ۴ = بالاتر از متوسط ۳ = متوسط ۲ = پایین‌تر از متوسط ۱ = ضعیف

الف: درجه‌بندی فرایند عملکرد یا شیوهٔ اجرا

فرایند عملکرد دانش‌آموز در هر یک از زمینه‌های زیر چگونه بود؟
۵ ۴ ۳ ۲ ۱ (الف) تهیهٔ یک طرح کامل برای پروژه
۵ ۴ ۳ ۲ ۱ (ب) تعیین مقدار مواد مورد نیاز
۵ ۴ ۳ ۲ ۱ (پ) انتخاب ابزارها و وسایل مناسب
۵ ۴ ۳ ۲ ۱ (ت) رعایت دستورالعملهای درست برای هر یک از عملیات
۵ ۴ ۳ ۲ ۱ (ث) استفاده از ابزارها به طور درست و ماهرانه
۵ ۴ ۳ ۲ ۱ (ج) استفاده از مواد بدون ریخت و پاش غیرضروری
۵ ۴ ۳ ۲ ۱ (چ) به پایان رسانیدن کار در یک زمان معقول

ب: درجه بندی محصول یا فراوردهٔ عملکرد

محصول یا فراوردهٔ مورد نظر تا چه میزانی با ملاکهای زیر مطابق است؟
۵ ۴ ۳ ۲ ۱ (الف) محصول درست و کامل به نظر می‌رسد
۵ ۴ ۳ ۲ ۱ (ب) ابعاد محصول با طرح اصلی مطابقت دارد
۵ ۴ ۳ ۲ ۱ (پ) کار به خوبی صیقل یافته است
۵ ۴ ۳ ۲ ۱ (ت) اتصالها و اجزاء به خوبی به هم وصل شده‌اند
۵ ۴ ۳ ۲ ۱ (ث) از مواد به خوبی استفاده شده است

شکل ۴ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی برای ارزشیابی از فرایند رفتار و محصول رفتار: پروژهٔ کارهای چوبی (گرانلاند، ۱۹۸۸، ص ۹۵)

عددی تشکیل می‌یابند. نمونه‌هایی از آن را در شکل ۴ـ۱۴ دیدید. علاوه بر این نوع مقیاس درجه‌بندی، انواع دیگری وجود دارند که در زیر آنها را توضیح می‌دهیم.

انواع مقیاسهای درجه‌بندی

چنان‌که گفتیم، مقیاسهای درجه‌بندی شکلهای مختلفی دارند. ما در اینجا معروف‌ترین آنها را معرفی می‌کنیم.

مقیاس درجه‌بندی عددی

ساده‌ترین نوع مقیاس درجه‌بندی، **مقیاس درجه‌بندی عددی**[1] است. در این مقیاس درجه‌بندی، ارزشیاب یا مشاهده‌گر، با کشیدن دایره‌ای به دور یک عدد یا علامت‌گذاری (چک کردن) آن، میزان یا درجهٔ موجود بودن رفتار یا ویژگی مورد ارزشیابی را مشخص می‌کند. شکل ۴ـ۱۴ دو

1. numerical rating scale

راهنمایی: با کشیدن دایره‌ای به دور یکی از اعداد مقابل هر جمله میزان شرکت دانش‌آموز مورد نظر را در بحث کلاسی تعیین کنید.
عالی = ۵ بالای متوسط = ۴ متوسط = ۳ زیر متوسط = ۲ ضعیف = ۱

۱. تا چه اندازه این دانش‌آموز در بحث شرکت می‌کند؟
۱ ۲ ۳ ۴ ۵

۲. اظهار نظرهای این دانش‌آموز تا چه اندازه به موضوع بحث مربوط اند؟
۱ ۲ ۳ ۴ ۵

شکل ۵ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی عددی

نمونه از نوع این مقیاس درجه‌بندی را نشان می‌دهد. به طور معمول، هر یک از اعداد مقیاس درجه‌بندی معنی خاصی دارد که در تمامی طول مقیاس یکسان است. معمولاً عدد ۱ کمترین ارزش را دارد و هرچه بر این عدد افزوده می‌شود بر ارزش ویژگی مورد ارزشیابی نیز افزوده می‌شود. تعریف اعداد مورد استفاده در مقیاس درجه‌بندی عددی می‌تواند درجه یا **میزان مناسب بودن، خوشایند بودن، فراوانی**، یا **توافق با یک بیان** را نشان دهد.

مقیاس درجه‌بندی عددی زمانی قابل استفاده است که ویژگیها یا کیفیتهای مورد درجه‌بندی قابل طبقه‌بندی کردن در طبقات محدودی باشند و نیز در رابطه با طبقه‌ای که مُعرف هر عدد است توافق وجود داشته باشد. بااین‌حال، معمولاً اعداد خیلی دقیق تعریف نمی‌شوند، و بنابراین تفسیرها و موارد استفاده‌های این نوع مقیاس نیز متفاوت است. در شکل ۵-۱۴ مثال دیگری از مقیاس درجه‌بندی عددی آمده است.

در مقیاس درجه‌بندی عددی، تصمیم در مورد تعداد طبقه‌ها یا درجه‌ها یک امر دلبخواهی است. با این حال، صاحب‌نظران بیشتر از ۱۰ طبقه را توصیه نمی‌کنند ـ به ویژه ارقام ۳ و ۷ بیشترین طرفدار را دارد. همچنین گفته شده است که مقیاسهای دارای تعداد طبقه‌های فرد (۳ یا ۷) بهتر از مقیاسهای دارای تعداد طبقه‌های زوج است (مثلاً ۴ یا ۶) (کوپر، ۱۹۹۴).

مقیاس درجه‌بندی نگاره‌ای

در **مقیاس درجه‌بندی نگاره‌ای**[1]، به جای استفاده از اعداد، از خطوط افقی استفاده می‌شود. در این مقیاس، از سنجشگر یا مشاهده‌گر خواسته می‌شود تا داوری خود را دربارهٔ ویژگی مورد مشاهده در طول یک خط مستقیم، با گذاشتن نوعی علامت، مشخص کند. در طول خط، مجموعه‌ای از مقوله‌ها، مکانها، یا نقاط مشخص می‌شوند، اما سنجشگر آزاد است که بین این نقاط نیز علامت بگذارد، در شکل ۶ـ۱۴ مثالی از این مقیاس درجه‌بندی آمده است.

در مقیاس درجه‌بندی شکل ۶ـ۱۴، برای همهٔ ویژگیهای مورد ارزشیابی طبقات یکسانی به کار رفته‌اند. به همین سبب، به این نوع مقیاسِ درجه‌بندیِ دارای طبقات یکسان **مقیاس**

1. graphic rating scale

راهنمایی: با گذاشتن علامت (×) در روی خط افقی مربوط به هر سؤال میزان شرکت دانش‌آموز را در بحث کلاسی تعیین کنید.

۱. تا چه اندازه این دانش‌آموز در بحث شرکت می‌کند؟

خیلی کم — کم — متوسط — زیاد — خیلی زیاد

۲. اظهار نظرهای این دانش‌آموز تا چه اندازه به موضوع بحث مربوط‌اند؟

خیلی کم — کم — متوسط — زیاد — خیلی زیاد

شکل ۶ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی نگاره‌ای یکنواخت

درجه‌بندی یکنواخت[1] گفته می‌شود. در بعضی از مقیاسها، برای ویژگیهای مختلف، طبقات مختلفی به کار می‌رود. به این گونه مقیاس، **مقیاس درجه‌بندی متغیر**[2] می‌گویند. شکل ۷ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی متغیر را نشان می‌دهد.

اشکال مهم مقیاسهای درجه‌بندی عددی و نگاره‌ای که در آنها واژه‌های خیلی زیاد، زیاد، کم، همیشه، غالباً، گاهی، به ندرت، و غیره (یا اعداد مُعرف آنها) به کار می‌روند این است که در میان افراد مختلف در معنای این واژه‌ها اتفاق نظر وجود ندارد. پین (۲۰۰۳) پژوهشی را نقل می‌کند

راهنمایی: با گذاشتن علامت (×) در روی خط افقی مربوط به هر سؤال نظر خود را نسبت به استاد این درس (نام درس) مشخص کنید.

۱. میزان علاقهٔ استاد به تدریس؟

بسیار علاقه‌مند — نسبتاً علاقه‌مند — دارای علاقهٔ متوسط — نسبتاً بی‌علاقه — بسیار بی‌علاقه

۲. مقدار مطالبی که در یک جلسهٔ درس آموزش می‌دهد؟

خیلی زیاد — زیاد — متوسط — کم — خیلی کم

۳. میزان توانایی در ایجاد انگیزه و علاقه در دانشجویان؟

بسیار توانا — نسبتاً توانا — دارای توانایی متوسط — نسبتاً ناتوان — بسیار ناتوان

شکل ۷ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی نگاره‌ای متغیر

1. constant-alternative scale 2. changing-alternative scale

جدول ۱-۱۴ معانی نسبت داده شده به اصطلاحات مورد استفاده در مقیاسهای درجه‌بندی (پین، ۲۰۰۳، ص ۳۰۷)

اصطلاح	میانگین نمره‌ای که برای اصطلاح پیشنهاد شده است	دامنهٔ متوسط ۵۰٪ پیشنهادها
همیشه	۹۹	۱۰۰-۹۸
غالباً	۸۸	۹۳-۸۰
معمولاً	۸۵	۹۰-۷۰
عموماً	۷۸	۸۵-۶۳
نسبتاً	۶۵	۸۰-۴۵
بعضی وقتها	۲۰	۳۵-۱۳
گاهی از اوقات	۲۰	۳۳-۱۰
به ندرت	۱۰	۱۸-۱۶
به طور اتفاقی	۵	۱۰-۳

که در آن از دانش‌آموزان دبیرستانی و دانشجویان دانشگاهی دربارهٔ معنی کلمه‌هایی مانند کلمه‌های پیش‌گفته پرسش به عمل آمد. از جمله سؤالهایی که از آنها پرسیده شد این بود که مثلاً کلمهٔ غالباً به معنی ۸۵ از ۱۰۰ است یا ۶۵ از ۱۰۰. بخشی از یافته‌های این پژوهش در جدول ۱-۱۴ آمده است. چنان‌که در این جدول می‌بینید، افراد مختلف برای کلمه‌های مختلف معانی گوناگونی قائل‌اند. "این اختلافها بدون تردید هم روایی و هم پایایی روش مقیاس درجه‌بندی را کاهش می‌دهد" (پین، ۲۰۰۳، ص ۲۰۶).

یکی از راههای غلبه بر مشکل فوق این است که، به کمک اعداد، فراوانی هر یک از درجه‌های مقیاس را تعریف کنید. پین (۲۰۰۳) برای این منظور پیشنهادهای زیر را داده است:

۵ = همیشه (۸۶ تا ۱۰۰ درصد موارد)
۴ = تقریباً همیشه (۶۵ تا ۸۵ درصد موارد)
۳ = غالباً (۳۵ تا ۶۵ درصد موارد)
۲ = گاهی از اوقات (۱۶ تا ۳۵ درصد موارد)
۱ = به ندرت (۰ تا ۱۵ درصد موارد)

راه دیگر جلوگیری از تفسیرهای مختلف سنجشگران یا ارزشیابان استفاده از مقیاس درجه‌بندی نگاره‌ای توصیفی است. در زیر به توضیح این مقیاس توجه کنید.

مقیاس درجه‌بندی نگاره‌ای توصیفی

در مقیاس درجه‌بندی نگاره‌ای توصیفی[1] نقاط روی مقیاس توصیف می‌شوند. توصیفهایی که

1. descriptive graphic rating scale

راهنمایی: با قرار دادن علامت (×) در روی خط مربوط به هر سؤال ویژگی مورد نظر را ارزشیابی کنید.

۱. کیفیت کار این دانش‌آموز را در انجام تکالیف درسی چگونه ارزیابی می‌کنید؟

کار خیلی خوب است، و هیچ نیازی به نظارت ندارد.	کارخوب است، به‌ندرت احتیاج به نظارت دارد.	کارخوب است، گاه به گاه نیاز به نظارت دارد.	کار قابل قبول است، اما احتیاج به نظارت فراوان دارد.	کار ضعیف است، غلطهای زیادی دارد.

شکل ۸ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی نگاره‌ای توصیفی

برای این منظور به کار می‌روند عبارتهایی هستند که به صورت رفتاری نشان می‌دهند که دانش‌آموز یا فرد مورد مشاهده و مورد ارزشیابی چگونه عمل می‌نماید. شکل ۸-۱۴ نمونه‌ای از مقیاسهای درجه‌بندی نگاره‌ای را نشان می‌دهد.

بعضی وقتها در طول مقیاس، علاوه بر توصیفهای کلامی، از اعداد نیز استفاده می‌شود. شکل ۹-۱۴

راهنمایی: با قرار دادن علامت (×) در روی خط مربوط به هر سؤال ویژگی خواسته شده را ارزشیابی نمایید.

۱. گزارش دانش‌آموز دربرگیرندهٔ مطالب مربوط به هدف آن است؟

۴	۳	۲	۱
همهٔ مطالب گفته شده توسط دانش‌آموز مربوط و معنی‌دار است.	غالب مطالب گفته شده مربوط و معنی‌دار است.	تنها حدود ۵۰ درصد مطالب گفته شده مربوط و معنی‌دار است.	بیشتر مطالب گفته شده واقعاً بی‌معنی و نامربوط بود.

۲. مطالب گزارش شده از لحاظ دستوری درست است؟

۴	۳	۲	۱
از لحاظ دستوری بی‌غلط است.	حدود ۱ تا ۳ غلط دیده می‌شود.	در جمله‌ها حدود ۵۰ درصد غلط وجود دارد.	بیشتر جمله‌ها از لحاظ دستوری غلط هستند.

۳. آیا دانش‌آموز هنگام صحبت کردن به دانش‌آموزان کلاس نگاه می‌کرد؟

۴	۳	۲	۱
مرتباً نگاه می‌کرد.	اکثر اوقات نگاه می‌کرد.	تنها ۵۰ درصد اوقات نگاه می‌کرد.	بیشتر اوقات به سمت دیگری نگاه می‌کرد.

شکل ۹ـ۱۴ نمونه‌ای از مقیاس درجه‌بندی نگاره‌ای توصیفی همراه با عدد برای سنجش گزارش شفاهی دانش‌آموز در کلاس (نیتکو، ۲۰۰۱، ص ۲۶۰)

نمونه‌ای از این نوع مقیاسها را نشان می‌دهد. این مقیاس درجه‌بندی برای سنجش کیفیت گزارش شفاهی یک دانش‌آموز درست شده است.

نیتکو (۲۰۰۱)، در ارتباط با مقیاس درجه‌بندی نگاره‌ای، گفته است بهترین شکل مقیاس درجه‌بندی همین نوع است. دلیلی که او در این مورد ذکر کرده این است که "توصیفهای نقاط روی مقیاس، با استفاده از توصیفهای رفتاری، همسانی درجه‌بندی را در بین دانش‌آموزان و ارزیابان افزایش می‌دهد" (ص ۲۸۲).

قواعد تهیهٔ مقیاس درجه‌بندی

برای اینکه مقیاس درجه‌بندی کارآمدتری تهیه کنید و نتایج معتبرتری به دست آورید، نکاتی را که در این قسمت توضیح می‌دهیم رعایت نمایید. رعایت این نکات هم موقعی که ویژگیهای مورد ارزشیابی را برمی‌گزینید مفید است، هم وقتی که فرم درجه‌بندی را طراحی می‌کنید، و هم زمانی که شرایطی را که قرار است در آن درجه‌بندی‌ها به دست آیند تعیین می‌نمایید. از آنجاکه مقیاس درجه‌بندی نگاره‌ای توصیفی قابل استفاده‌ترین فرم درجه‌بندی برای مقاصد تحصیلی است، نکاتی که در اینجا پیشنهاد می‌شوند عمدتاً به ساخت و استفاده از این نوع مقیاس درجه‌بندی اشاره دارند[1].

۱. سعی کنید ویژگیهایی را که از لحاظ آموزشی مهم‌اند برگزینید. قبلاً در بحث مربوط به فهرست وارسی نیز گفتیم که فقط ویژگیها یا رویدادهای مهم را مورد ارزیابی قرار دهید. در اینجا نیز همین قاعده را توصیه می‌کنیم. روش مقیاس درجه‌بندی و سایر روشهای مورد بحث در این فصل، مانند تمامی روشهای ارزشیابی معرفی شده در این کتاب، باید با هدفهای آموزش و یادگیری همخوانی داشته باشند. بنابراین، به هنگام تهیهٔ مقیاس درجه‌بندی، بهترین راهنمای انتخاب ویژگیها، همان هدفهای آموزشی و بازده‌های یادگیری از پیش تعیین شده هستند.

۲. ویژگیهای مورد ارزشیابی را به گونه‌ای بیان کنید که مستقیماً قابل مشاهده باشند. چنان‌که ویژگی مورد نظر قابل مشاهده باشد قابل درجه‌بندی و ارزشیابی نیز خواهد بود. برای این منظور، ابتدا باید توجه داشت ویژگیهایی برگزیده شوند که در موقعیتهای آموزشگاه از دانش‌آموزان سر می‌زنند و لذا برای معلم امکان مشاهدهٔ آنها وجود دارد. علاوه بر این، ویژگیهای انتخابی باید به روشنی قابل مشاهده باشند. رفتارهایی مانند شرکت در بحث کلاسی، تلفظ کلمات، و مهارت در برقراری روابط دوستانه از جمله رفتارهای قابل مشاهده‌اند. در مقابل، علاقه به درس، احساس

---

۱. پیشنهادهای این بخش عمدتاً از کتاب اندازه‌گیری و ارزشیابی در آموزش، اثر گرانلاند و لین (۱۹۹۰)، و کتاب اندازه‌گیری و سنجش در آموزش، اثر لین و گرانلاند، (۲۰۰۰)، اقتباس شده‌اند.

حقارت، و نگرش نسبت به همکلاسیها در ردیف رفتارهایی قرار دارند که به خوبی قابل مشاهده نیستند. تا حد امکان، باید درجه‌بندی را به رفتارهای قابل مشاهده اختصاص دهید.

۳. **هم ویژگیهای مورد ارزشیابی و هم نقاط روی مقیاس را به روشنی تعریف کنید.** بزرگ‌ترین مشکل مقیاس درجه‌بندی کاربرد اصطلاحات و صفات کلی و مبهم در آنهاست. این ابهام هم در بیان ویژگیهای مورد ارزشیابی ممکن است پیش بیاید و هم در خود مقیاس درجه‌بندی. برای کاستن از این مشکل، سعی کنید هم ویژگی مورد ارزشیابی را به دقت تعریف کنید و هم، با افزودن توضیحاتی به مقیاس، خودِ مقیاس را برای مشاهده‌گر روشن سازید. برای نمونه، توضیحات مختصری که در مقیاس درجه‌بندی توصیفی به کار می‌روند از این لحاظ مفیدند.

۴. **شمارهٔ درجات مقیاس را بین ۳ تا ۷ برگزینید و به ارزشیابان اجازه بدهید تا در بین این نقاط نیز درجه‌بندی انجام دهند.** تعداد دقیق نقاط روی یک مقیاس درجه‌بندی تنها به وسیلهٔ داوریهایی که انجام می‌گیرد تعیین می‌شود. در زمینه‌هایی که فقط داوریهای کلی میسر است، تعداد کمتری از این نقاط لازم است. معمولاً داشتن بیش از ۷ نقطه امتیازی به حساب نمی‌آید، و به ندرت ما به تمایزهای دقیق‌تر از ۷ مورد نیاز داریم. البته در این‌گونه موارد می‌توان از ارزشیاب خواست تا در بین نقاط نیز علامت بگذارید.

۵. **از ارزشیابان بخواهید هر زمان به سؤالی برمی‌خورند که قادر به قضاوت دربارهٔ ویژگی مطرح شده در آن نیستند به آن سؤال جواب ندهند.** بارها اتفاق می‌افتد که مشاهده‌گر یا ارزشیاب فرصت ندارد تا برخی از ویژگیهای مورد ارزشیابی را به دقت مشاهده کند یا به هر دلیل دیگر شایسته نیست تا آن را درجه‌بندی نماید. در چنین حالتی باید از او خواست تا از قضاوت دربارهٔ این‌گونه موارد خودداری کند. به همین منظور، می‌توان در مقیاس درجه‌بندی در هر سؤال محلی برای علامت زدن "نمی‌توانم قضاوت کنم" یا "اطلاعات کافی ندارم" پیش بینی کرد. یا به جای درنظرگرفتن چنین محلی می‌توان از ارزشیاب خواست تا در قسمتی که برای اظهارنظر مربوط به هر سؤال منظور شده است این کار را انجام دهد.

۶. **در صورت امکان از چند نفر بخواهید تا رفتارها یا محصولات مورد نظر را مشاهده و ارزشیابی کنند.** اگر به جای یک نفر از دو یا چند نفر بخواهید تا به ارزشیابی از ویژگیهای مورد نظر بپردازند و بعد میانگین نتایج حاصل را مورد استفاده قرار دهید، نمرات پایاتر و قابل اعتمادتری در اختیار خواهید داشت. به ویژه زمانی که نتایج سنجش عملکرد پیامدهای درازمدتی خواهند داشت انجام این کار ضروری است.

۷. **ابتدا عملکرد همهٔ دانش‌آموزان را در یک تکلیف درجه‌بندی کنید و آنگاه به تکلیف بعدی بروید.** امتیاز درجه‌بندی عملکرد همهٔ دانش‌آموزان در یک تکلیف پیش از رفتن به تکلیف بعدی شبیه

به امتیاز تصحیح جوابهای همهٔ دانش‌آموزان به یک سؤال تشریحی پیش از رفتن به سؤال بعدی است. ما فواید این کار را در فصل مربوط به آزمونهای تشریحی (فصل ۱۰) توضیح داده‌ایم.

۸. **در صورت امکان درجه‌بندی خود را بدون اطلاع از نام دانش‌آموز انجام دهید.** امتیاز مربوط به این پیشنهاد نیز در فصل مربوط به آزمونهای تشریحی توضیح داده شده. البته برای سنجش همه نوع عملکردی ناشناس ماندن دانش‌آموز میسر نیست (مثلاً در سنجش گزارش شفاهی)، اما آنجا که این کار عملی است بی‌خبر بودن ارزشیاب از نام فردی که عملکردش مورد ارزشیابی قرار دارد مفید است.

نمره‌گذاری مقیاس درجه‌بندی

برای نمره‌گذاری مقیاس درجه‌بندی عددی می‌توان از اعداد روی مقیاس (مثلاً همان ۱ تا ۵) استفاده کرد. بنابراین، اگر برای یکی از ویژگیهای رفتاری یا محصول رفتاری یک دانش‌آموز در یک مقیاس درجه‌بندی عدد ۵ و برای دانش‌آموز دیگر عدد ۱ انتخاب شد نمرهٔ آنان در ویژگی به ترتیب ۵ و ۱ است. اما در مقیاس درجه‌بندی نگاره‌ای، از آنجا که ارزشیاب می‌تواند در روی خطوط مقیاس و در بین اعداد داوری خود را نشان دهد، نمرهٔ دانش‌آموز در یک ویژگی الزاماً عدد صحیح نخواهد بود و می‌تواند اعداد اعشاری را شامل باشد. بنابراین، گزارش کردن نمره‌های دارای ارقام اعشاری برای دانش‌آموزان با روش مقیاس درجه‌بندی غیرمعمول نیست.

واقعه‌نگاری

تعریف واقعه‌نگاری

تعریف واقعه‌نگاری

گذا به تعریف، روش **واقعه‌نگاری**[1] یا روش ثبت رویداد به "توصیفهای واقعی از رویدادها و اتفاقات معنی‌داری که معلم [مشاهده‌گر] در نتیجهٔ مشاهده از زندگی دانش‌آموزان به دست می‌آورد" (گرانلاند و لین، ۱۹۹۰، ص ۳۷۷) گفته می‌شود. در این روش معلم یا مشاهده‌گر باید هر اتفاق یا رویدادی را که در زندگی دانش‌آموز رخ می‌دهد و آن را مهم می‌داند، بلافاصله بعد از وقوع، ثبت نماید. بنابراین، این روش بیشتر جنبهٔ تحلیلی دارد تا ارزشیابی.

توصیف رویدادها را یا می‌توان در دفتر یادداشت نوشت یا اینکه در کارتهایی نظیر آنچه در شکل ۱۰ـ۱۴ دیده می‌شود وارد کرد. چنان‌که در این شکل می‌بینید، بهتر است توصیف عینی رویداد یا رفتار دانش‌آموز از تفسیری که می‌توان از آن رویداد یا رفتار به عمل آورد مجزا باشد.

---

1. anecdotal record

کلاس: چهارم
تاریخ: ۱۹۹۴/۲/۴
مشاهده‌گر: م.ز.

دانش‌آموز: ماری جانسون
محل: کلاس درس

رویداد

چند لحظه پیش از آنکه کلاس شروع شود، ماری شعری را که دربارهٔ "بهار" سروده بود به من نشان داد. شعر بسیار زیبایی بود. به او گفتم اگر دوست دارد می‌تواند آن را برای کلاس بخواند. او شعرش را با صدای آرام خواند؛ مرتب به روی کاغذ نگاه می‌کرد، پای راستش را به جلو و عقب تکان می‌داد، و با یقهٔ بلوزش ور می‌رفت. وقتی که ماری شعرخوانی‌اش را تمام کرد، استیو (از ردیف آخر کلاس) گفت "من صدایت را خوب نشنیدم. ممکن است یک بار دیگر آن را با صدای بلند بخوانی؟" ماری گفت "نه" و در صندلی خود نشست.

تفسیر

ماری از نوشتن داستان و شعر لذت می‌برد و توانایی خلاق خود را به نمایش می‌گذارد. با این حال، در مقابل جمع خجول و دستپاچه به نظر می‌رسد. امتناع او از خواندن دوبارهٔ شعر حاکی از دستپاچگی است.

شکل ۱۰ـ۱۴ فرم مورد استفاده در روش واقعه‌نگاری (اقتباس از گرانلاند و لین، ۲۰۰۰، ص ۳۱۸)

ویژگیهای واقعه‌نگاری

روش واقعه‌نگاری یک روش مشاهدهٔ مستقیم رفتار است، و هرچند از این روش در زمینه‌های تحصیلی می‌توان استفاده کرد اما بیشتر برای سنجش هدفهای آموزشی حوزه‌های عاطفی و روانی ـ حرکتی قابل استفاده است. در گذشته این روش عمدتاً برای سنجش سازگاری اجتماعی به کار می‌رفت، اما به تازگی در سایر زمینه‌های تربیتی نیز از آن استفاده می‌کنند. مهرنز و لهمان (۱۹۸۴) در این باره گفته‌اند که استفاده از روش واقعه‌نگاری در یک مدت معین می‌تواند تصویر نسبتاً کاملی از تغییراتی که در یک دانش‌آموز خاص صورت می‌پذیرد در اختیار معلم قرار دهد. پین (۲۰۰۳) نیز در این‌باره گفته است "داده‌های حاصل از روش واقعه‌نگاری در طول یک دورهٔ زمانی می‌تواند یک دیدگاه طولی از رشد و الگوهای تغییر دانش‌آموز به ما بدهد" (ص ۳۱۲).

پین (۲۰۰۳) نیز در ارتباط با ویژگیهای روش واقعه‌نگاری نکاتی را مطرح ساخته است. ما خلاصهٔ آنها را در زیر می‌آوریم:

۱. روش واقعه‌نگاری برای تکمیل اطلاعات جمع‌آوری شده به وسیلهٔ روشهای دیگر مفید است.
۲. روش واقعه‌نگاری می‌تواند بینش معلم را نسبت به نقاط قوت و ضعف دانش‌آموزان افزایش دهد.
۳. بسیاری از رویدادهایی که در گزارش واقعه‌نگاری می‌آیند می‌توانند برای مشاوران مدرسه مفید باشند. بنابراین، معلمان باید بکوشند تا این اطلاعات را در اختیار مشاوران و سایر معلمانی که این اطلاعات را نیاز دارند قرار دهد.

۴. پاره‌ای از نتایج واقعه‌نگاری ممکن است برای هدایت شغلی و ارتباط با پرسنل آموزشی و حرفه‌ای مفید باشند. باید بکوشید تا محرمانه بودن این اطلاعات حفظ شوند.

۵. واقعه‌نگاری ممکن است جنبه‌هایی از برنامهٔ درسی را که برای دانش‌آموزان خاص نیاز به تجدیدنظر دارد مشخص نماید.

از آنجا که روش واقعه‌نگاری روشی وقت‌گیر است و در یک زمان با تعداد زیادی دانش‌آموز قابل کاربستن نیست، بهتر است استفاده از آن را به مقاصد ویژه محدود کنیم. آن دسته از بازده‌های یادگیری را که به خوبی می‌توان با آزمونهای کتبی، آزمونهای عملکردی، یا حتی با روشهای فهرست وارسی و مقیاس درجه‌بندی سنجش کرد نباید با روش واقعه‌نگاری مورد سنجش قرار داد. تنها آن دسته از رفتارهایی که لازم است در شرایط کاملاً طبیعی محیط مشاهده و ارزشیابی شوند باید با این روش مورد ارزشیابی قرار گیرند. از جملهٔ این رفتارها می‌توان موارد زیر را ذکر کرد: چگونگی برخورد دانش‌آموز با مسائل، پشتکار دانش‌آموز در ادامهٔ کار، اشتیاق به گوش دادن به عقاید دیگران، نحوهٔ همکاری در فعالیتهای کلاسی، نحوهٔ اظهارنظر در موقعیتهای مختلف اجتماعی، نحوهٔ رفتار کردن در موقعیتهای مختلف اجتماعی، و مانند اینها.

چنان‌که گفته شد، بهترین کاربرد این روش با رفتارها و ویژگیهایی از دانش‌آموزان است که با وسایل اندازه‌گیری و سنجش متداول، یعنی آزمونها و امتحانات، قابل ارزشیابی نیستند. برای نمونه، معلمی که می‌خواهد علاقه‌مندی یک دانش‌آموز ابتدایی را به خواندن بسنجد می‌تواند او را در یک یا چند جلسهٔ خواندنِ آزاد مورد مشاهده قرار دهد و نتایج را به روش واقعه نگاری ثبت کند. و اگر می‌خواهد روابط اجتماعی او را بسنجد می‌تواند رفتار او را در زنگ تنفس مطالعه نماید. همچنین معلمی که می‌خواهد نگرش یک دانش‌آموز دبیرستانی را نسبت به علوم بسنجد می‌تواند اظهارنظرهای او را در بحثهای علمی ثبت نماید یا عادات کاری و مهارتهای آزمایشگاهی او را در جلسات آزمایشگاهی مورد مشاهده قرار دهد.

هرچند که روش واقعه نگاری را می‌توان با تمامی دانش‌آموزان کلاس مورد استفاده قرار داد، اما این روش با برخی دانش‌آموزانِ ویژه کاربرد بیشتری دارد. مثلاً کودکی که عقب‌مانده است، کودکی که از نظر اجتماعی طرد شده است، و نیز کودک معلول نیاز به توجه بیشتری دارند. مشاهدات دقیق از این افراد به تشخیص و رفع مشکلاتشان کمک می‌کند.

آخرین نکته‌ای را که در رابطه با استفاده از روش واقعه‌نگاری باید توضیح داد این است که مشاهدات را باید به افراد معین و رفتارهای معین محدود کرد. تلاش برای مشاهدهٔ تعداد زیادی رفتار و به مقدار زیاد بی‌فایده و مأیوس‌کننده خواهد بود. محدود کردن مشاهدات و گزارشها به رفتارهای معین و دانش‌آموزان خاص، برای استفادهٔ بهینه از این روش، ضروری است. بنابراین، باید سعی شود که پیش از اقدام به استفاده از روش واقعه‌نگاری یک نقشهٔ عینی و عملی از فعالیتها طرح‌ریزی گردد.

قواعد انجام واقعه‌نگاری و تهیهٔ گزارش آن

برای اینکه روش واقعه‌نگاری یا روش ثبت رویداد بهترین نتیجه را به دست بدهد باید نکات زیر در آن رعایت شوند:

۱. یک توصیف واقعی از آنچه اتفاق افتاده است، زمانی که اتفاق افتاده است، و در وضعیت یا شرایطی که اتفاق افتاده است به دست دهد.
۲. تفسیر از واقعه و اعمال پیشنهادی باید از توصیف واقعه جدا باشد.
۳. هر یک از موارد ثبت واقعه باید تنها یک واقعه را شامل باشد.
۴. آنچه ثبت می‌شود باید از لحاظ رشد و تحول دانش‌آموز یا بازده‌های یادگیری و هدفهای آموزشی رویداد مهمی باشد.
۵. تنها آن قسمت از ویژگیها یا حوزه‌های رفتاری مورد مشاهده و واقعه‌نگاری قرار گیرد که با روشهای دیگر ارزشیابی قابل ارزشیابی نباشد.

علاوه بر موارد بالا، گرانلاند و لین (۱۹۹۰) و لین و گرانلاند (۲۰۰۰) در جهت بهبود روش واقعه‌نگاری رهنمودهایی را مطرح ساخته‌اند. ما در زیر خلاصهٔ آنها را می‌آوریم.

۱. **از پیش تعیین کنید که چه چیزی را مورد مشاهده قرار دهید، اما آمادهٔ ثبت رفتارهای غیرمعمول نیز باشید.** طرح ریزی برای برنامهٔ مشاهده و تعیین نوع رفتارهایی که باید مشاهده شوند برای روش واقعه‌نگاری ضروری است، اما آمادگی داشتن برای ثبت وقایع غیرمترقبه نیز لازم است. بنابراین، روش کار مشاهده کننده باید انعطاف‌پذیر باشد و خود او همواره مترصد ثبت رویدادهای پیش بینی نشده اما معنی‌دار باشد.

۲. **نتایج مشاهدات را برای منابع احتمالی سوگیری تحلیل کنید.** دقت کنید که تعصبات و سوگیریهای احتمالی خود را که احتمالاً ناشی از جنس، نژاد، قومیت، و سایر ویژگیهای فرهنگی است در گزارش واقعه‌نگاری دخالت ندهید. جدا ساختن آنچه عملاً مشاهده می‌شود و تفسیر خودتان از آن می‌تواند مفید باشد. "برای مثال، در نظر بیاورید که اگر شخص دیگری مورد نظر شما بود، آیا همان مشاهدات ثبت شده به نتیجه گیری و تفسیر متفاوتی می‌انجامید؟" (لین و گرانلاند، ۲۰۰۰، ص ۳۲۱). البته تفسیرهای مختلف از مشاهدات ثبت شدهٔ واحد اگر بر اساس دانش مربوط به الگوهای متفاوت رفتار دو دانش‌آموز باشد ممکن است درست باشند اما تفاوت تفسیرها نباید در نتیجهٔ سوگیریهای جنسیتی، نژادی، قومی، و غیرهٔ مشاهده‌گر یا ارزشیاب باشد.

۳. **آن مقدار از موقعیت را مشاهده کنید که رفتار را برای شما معنی‌دار سازد.** غالباً لازم می‌آید که، علاوه بر رفتار دانش‌آموز مورد نظر، رفتار سایر دانش‌آموزان یا عوامل دیگر موجود در محیط را نیز مورد مشاهده قرار داده و گزارش نمایید. برای نمونه، یک عمل پرخاشگرانه، مانند هُل

دادن دانش‌آموز دیگر، ممکن است به قصد بازی و تفریح، جلب توجه، تلافی کردن، یا دشمنی ورزیدن انجام گیرد. برای درک معنی چنین رفتاری، توجه به رفتار دانش‌آموزان دیگر و موقعیت ویژه‌ای که رفتار در آن اتفاق می‌افتد ضروری است. بنابراین، ثبت رویداد هم باید شامل رفتار دانش‌آموز مورد مشاهده و هم شامل شرایط موقعیتی لازم برای درک آن رفتار باشد.

۴. **بلافاصله پس از مشاهدهٔ واقعه یا رویداد آن را ثبت کنید.** از آنجا که جزئیات مشاهدات به سرعت فراموش می‌شوند، بهتر است بلافاصله پس از آنکه رفتاری را مورد مشاهده قرار دادید آن را ثبت نمایید. اگر ثبت کامل رویداد بلافاصله پس از مشاهده امکان‌پذیر نیست، نکات مهم آن را به طور خلاصه یادداشت کنید، تا بعد سر فرصت شرح کامل آن را بنویسید.

۵. **هر یک مورد ثبت واقعه را به توصیف مختصری از یک رویداد محدود نمایید.** توصیفهای مختصر وقت کمتری برای نوشتن و خواندن لازم دارند و در عوض به سادگی خلاصه می‌شوند. بنابراین، توصیفات خود را به حداقل کلمات لازم محدود کنید. برای این منظور در هر بار یک واقعه را مشاهده و ثبت نمایید. همچنین در توصیف خود صرفاً به اعمال و گفتار دانش‌آموز و موقعیتی که در آن این اعمال و رفتار رخ داده‌اند بسنده نمایید.

۶. **توصیف واقعیتها و تفسیر خودتان از آنها را از هم جدا کنید.** این نکته را قبلاً هم توضیح دادیم و در اینجا نیز بر آن تأکید می‌کنیم. توصیف یک واقعه باید تا حد امکان دقیق و عینی باشد. یعنی اینکه دقیقاً باید آنچه را که اتفاق افتاده است وصف کنید و از عبارتهای ذهنی و کلی بپرهیزید. از کاربرد واژه‌های ناخشنود، خجول، غمگین، خودخواه، و مانند اینها پرهیز کنید. اگر لازم شد، این واژه‌ها و تعابیر را در بخش مجزایی که دربرگیرندهٔ تفسیرهای شما از رویداد است قرار دهید.

گرانلاند و لین (۱۹۹۰) دربارهٔ توصیف عینی واقعیتها می‌گویند چیزهایی مانند جملات زیر بنویسید: "امروز در زمین بازی ماری و هلن در بازی سافت‌بال همبازی بودند. بتی گفت، من می‌خواهم در تیم ماری باشم، و اول بازی کنم، در غیر این صورت بازی نخواهم کرد." (ص ۳۸۲). اما، از نوشتن مطالبی مانند آنچه در زیر می‌آید خودداری کنید:

- امروز در زمین بازی بتی بار دیگر نشان داد که همیشه می‌خواهد حرف خودش را به کرسی بنشاند.
- بتی در زمین بازی خودخواه و خرابکارانه عمل کرد.
- بتی نمی‌تواند با دیگران بازی کند زیرا او تنها فرزند خانواده است و خودخواه بار آمده است.

۷. **هم موارد مثبت و هم موارد منفی رفتار را ثبت کنید.** به سبب اینکه، در مشاهدهٔ رفتار دانش‌آموزان، غالباً مشکلات رفتاری یا رفتارهای منفی خودنمایی می‌کنند، بنابراین گزارشهای واقعه‌نگاری غالباً شامل موارد منفی رفتار هستند. اما در ارزشیابی درست از رفتار کودکان و به ویژه در گزارش از جریان رشد کودکان ثبت موارد مثبت رفتار به اندازهٔ مشاهده و ثبت موارد منفی ضرورت دارد.

۸. **پیش از قضاوت و نتیجه‌گیری دربارهٔ رفتار یک دانش‌آموز، تعداد کافی گزارش واقعه‌نگاری تهیه نمایید.** هرگز به یک بار مشاهده و یک مورد واقعه‌نگاری برای قضاوت دربارهٔ دانش‌آموزان بسنده نکنید، زیرا هر کسی ممکن است به تصادف یک مورد رفتار استثنایی (بسیار مثبت یا بسیار منفی) از خود بروز دهد. برای اینکه قضاوت درستی دربارهٔ یک دانش‌آموز داشته باشید جمع‌آوری تعدادی مورد واقعه‌نگاری ضروری است.

۹. **در نوشتن گزارش واقعه‌نگاری تمرین کنید.** آخرین نکته اینکه، نوشتن گزارش مشاهدات کار نسبتاً پیچیده‌ای است که مانند هر کار جدی دیگر نیاز به تمرین دارد. تمرین کردن نباید صرفاً به نوشتن نتایج مشاهدات محدود گردد، بلکه هم در انتخاب رویدادهای مهم، هم در مشاهدهٔ دقیق آنها، و هم در توصیف عینی آنچه مشاهده می‌کنید تمرین نمایید. گرانلاند و لین (۱۹۹۰) می‌گویند اگر تمامی کارکنان آموزشگاه قرار است به کار واقعه‌نگاری بپردازند، لازم است یک برنامهٔ آموزش ضمن خدمت منظم برای این منظور فراهم شود. اگر تنها یک نفر معلم می‌خواهد این روش را بیاموزد، از راهنمای آموزشی یا یک معلم همکار که در این کار مهارت دارد می‌توان خواست تا به معلم مورد نظر کمک کند. گرانلاند و لین همچنین پیشنهاد می‌کنند که بهترین موقعیت برای شروع مشاهدهٔ رفتار دانش‌آموزان مشاهدهٔ عادتهای مطالعهٔ آنان است، زیرا این موقعیت به مشاهده‌گر فرصت کافی می‌دهد تا رفتار مهمی را مشاهده و ثبت نماید.

امتیازها و محدودیتهای روشهای سنجش مشاهده‌ای

هرچند درموقعیتهای آموزشگاهی از روشهای سنجش مشاهده‌ای به اندازهٔ سایر روشهای سنجش استفاده نمی‌شود، با این حال استفادهٔ درست از آنها در بسیاری مواقع می‌تواند ارزشمند باشد. در زیر به امتیازهای استفاده از روشهای سنجش مشاهده‌ای توجه کنید:

۱. روشهای سنجش مشاهده‌ای امکاناتی را فراهم می‌آورند تا دربارهٔ رفتارها و به ویژه دربارهٔ رفتارهای سازگاری، هیجانی، و اجتماعی دانش‌آموزان اطلاعاتی جمع‌آوری شود که سایر روشهای سنجش از عهدهٔ آن برنمی‌آیند.

۲. روشهای سنجش مشاهده‌ای به ما امکان می‌دهند تا توانایی فرد را در کاربست اطلاعات در موقعیتهای زندگی سنجش کنیم.

۳. به سبب شباهت میان موقعیتهای سنجش و موقعیتهایی که در آن دانش و مهارتهای یادگیرنده به کار خواهند رفت، نتایج حاصل از روشهای مشاهده، در قیاس با سایر روشهای سنجش، از قدرت پیش‌بینی بیشتری برخوردارند.

۴. روشهای سنجش مشاهده‌ای به سادگی قابل استفاده در موقعیتهای مختلف، با تکالیف

گوناگون، افراد متفاوت، و همهٔ سنها و تمامی سطوح تحصیلی هستند.

۵. اطلاعات حاصل از روشهای سنجش مشاهده‌ای می‌تواند به صورت منابع کمکی ارزشمندی برای اطلاعات مربوط به پیشرفت و توانایی حاصل از سایر روشهای سنجش مورد استفاده قرار گیرد.

۶. روشهای سنجش مشاهده‌ای هم داده‌های کمّی به دست می‌دهند و هم اطلاعات کیفی.

۷. چون که مشاهدات در موقعیتهای طبیعی صورت می‌پذیرند ادغام آنها با برنامه‌های کلی آموزشی بیشتر میسر است و به معلم فرصت می‌دهد تا از مشاهده به عنوان بخشی از فرایند آموزش استفاده کند.

به رغم امتیازهای فوق، روشهای سنجش مشاهده‌ای محدودیتها و اشکالهایی نیز دارند. در زیر پاره‌ای از اشکالات مهم این روشها را ذکر می‌کنیم.

۱. احتمال تأثیر تعصبات و سوگیریهای مشاهده‌گر یا سنجشگر در گزارش و تفسیر نتایج مشاهدات خود وجود دارد.

۲. نیاز به صرف وقت زیاد و استفاده از وسایل و تجهیزاتی که برای همهٔ معلمان و در بسیاری از مدارس امکان‌پذیر نیست کاربرد این روشها را با دشواری مواجه می‌سازد.

۳. استنتاج از رویدادهای واحد یا محدود و در دسترس نداشتن داده‌های ضد داده‌های موجود می‌تواند به نتیجه گیریهای غلط بینجامد.

۴. مشاهدات محدود همیشه نمی‌تواند علتهای رفتار را نشان دهد، زیرا یک رفتار واحد می‌تواند علتهای پیش‌آیندی مختلفی داشته باشد. برای رفع این مشکل، باید از رفتار مورد نظر در موقعیتهای گوناگون نمونه گیری شود که کاری بسیار دشوار است.

۵. امکان مشاهدهٔ رفتارهای مختلف در کلاسهای درس و متمایز ساختن رفتارهای مربوط و نامربوط که همزمان رخ می‌دهند دشوار است.

۶. ثبت مشاهده، همزمان با انجام مشاهده، کار بسیار دشواری است. ثبت مشاهده پس از گذشت زمان نیز خیلی دقیق نیست، زیرا تمایل ما آدمیان بر این است که چیزهایی را که با باورمان سازگار نیستند یا دوست نداریم به سرعت فراموش کنیم.

۷. قطعیت غیرموجه یکی از اشکالات دیگر مشاهدات غیرمحتاطانه است. در نتیجه استنتاجهای ناشی از مشاهدات را باید بسیار احتمالی و فرضی تلقی کرد تا اینکه شواهد تأییدکنندهٔ بیشتری به دست آیند.

موارد استفادهٔ روشهای سنجش مشاهده‌ای

روشهای سنجش مشاهده‌ای در آموزشگاه موارد استفادهٔ فراوانی دارند. چنان‌که در آغاز این فصل گفتیم، همهٔ عملکردهای یادگیرندگان بر مشاهده استوارند، و روشهای ثبت و ضبط نتایج

مشاهدات بر روشهای سنجش مشاهده‌ای مورد بحث در این فصل متکی‌اند. با این روشها می‌توان هم جنبه‌های شناختی دانش‌آموزان و دانشجویان را ثبت و سنجش کرد، هم جنبه‌های عاطفی، و هم جنبه‌های روانی - حرکتی را. افزون بر آن، با روشهای سنجش مشاهده‌ای هم می‌توان فراورده‌های رفتار را مورد ارزشیابی قرار داد و هم فرایندهای آن را.

نکتهٔ مهم دیگری را که در ارتباط با کاربرد روشهای سنجش مشاهده‌ای باید مورد تأکید قرار داد این است که این روشها را با سایر روشهای سنجش یادگیری، مانند سنجش عملکردی و کارپوشه، نیز می‌توان مورد استفاده قرار داد. یعنی روشهای سنجش مشاهده‌ای می‌توانند تکمیل‌کنندهٔ سایر روشهای سنجش باشند. بنابراین، از آنها می‌توان هم به صورت انفرادی و هم به صورت تکمیلی، یعنی با روشهای دیگر و با یکدیگر، استفاده کرد. در زیر به نمونه‌هایی از کاربرد ترکیبی این روشها توجه کنید.

استفاده از روشهای مختلف سنجش مشاهده‌ای به صورت ترکیبی

هرچند که ما روشهای سنجش مشاهده‌ای را جداگانه بحث و بررسی کردیم، اما در عمل معمولاً آنها را به صورت مکمل یکدیگر مورد استفاده قرار می‌دهند. لین و گرانلاند (۲۰۰۰) در این‌باره سنجش فعالیتهای دانش‌آموزان در آزمایشگاه علوم را مثال زده‌اند. چگونگی استفاده از روشهای مختلف برای این منظور در جدول ۲ـ۱۴ نشان داده شده است.

جدول ۲ـ۱۴ استفاده از فنون مختلف برای سنجش فعالیتهای دانش‌آموز در یک آزمایشگاه علوم (لین و گرانلاند، ۲۰۰۰، ص ۲۸۵)

یادگیری موردنظر	نمونه‌هایی از عملکرد مورد سنجش	فنون سنجش
دانش روشهای آزمایشی	روشهای مناسب را شرح می‌دهد. وسایل و موارد استفادهٔ آنها را مشخص می‌کند. آزمایشهای ناقص را نقد می‌کند.	آزمونهای مداد و کاغذی (کتبی) آزمونهای تشخیصی آزمایشگاهی
مهارت در طراحی آزمایش علمی	یک آزمایش علمی طراحی می‌کند.	سنجش عملکردی با تأکید بر فراورده (فهرست وارسی)
مهارت در انجام آزمایش علمی	وسایل لازم را انتخاب می‌کند. وسایل را آماده می‌کند. آزمایش را انجام می‌دهد.	سنجش عملکردی با تأکید بر برفرایند (مقیاس درجه‌بندی)
مهارت در تفسیر نتایج آزمایش	روابط معنی‌دار را تشخیص می‌دهد. نقاط ضعف داده‌ها را تشخیص می‌دهد. نتیجه گیریهای درست انجام می‌دهد.	سنجش عملکردی و پرسش شفاهی
عادتهای کاری	از وسایل درست استفاده می‌کند. کار را به موقع تمام می‌کند. جایی را که کار کرده تمیز می‌کند.	سنجش عملکردی با تأکید بر فرایند (فهرست وارسی)

خلاصهٔ فصل

روشهای سنجش مشاهده‌ای ابزارهایی هستند که معلم یا سنجشگر می‌تواند به کمک آنها نتایج مشاهدات خود از فرایندها و فراورده‌های رفتاری یادگیرندگان را ثبت و ضبط نماید. روشهای سنجش مشاهده‌ای که در این فصل مورد بحث قرار گرفته عبارت‌اند از: فهرست وارسی، مقیاس درجه‌بندی، و واقعه‌نگاری.

فهرست وارسی شامل فهرستی است از رفتارها، ویژگیها، فعالیتها، یا تولیدات دانش‌آموز همراه با جایی در مقابل هر مادهٔ فهرست برای تعیین (یا چک کردن) بود یا نبود آن ماده. در فهرست ماده‌ها هم می‌توان ویژگیهای مثبت فراورده یا فرایند رفتار مورد نظر را قرار داد، هم ویژگیهای منفی آن، و هم ترکیبی از آندو را.

در مقیاس درجه‌بندی درجه یا مقدار ویژگی مورد نظر سنجش می‌شود. با این روش نیز هم می‌توان فرایند رفتار و هم فراوردهٔ آن را سنجش کرد. مقیاس درجه‌بندی انواع مختلفی دارد که مهم‌ترین آنها به قرار زیرند: مقیاس درجه‌بندی عددی، مقیاس درجه‌بندی نگاره‌ای، و مقیاس درجه‌بندی نگاره‌ای توصیفی.

واقعه‌نگاری به توصیفهای واقعی از رویدادها و اتفاقات معنی‌داری که معلم یا سنجشگر در نتیجهٔ مشاهده از رفتار دانش‌آموزان به دست می‌آورد گفته می‌شود. در این روش، معلم هر اتفاق یا رویدادی را که در زندگی دانش‌آموز رخ می‌دهد و آن را مهم می‌داند بلافاصله پس از وقوع ثبت می‌نماید.

فهرست وارسی را به راههای مختلفِ ۰ـ۱، بود ـ نبود، و بله ـ نه می‌توان نمره‌گذاری کرد. در فهرستهای وارسی که ویژگیهای منفی یا اشتباهات را نیز شامل می‌شوند، می‌توان برای ویژگیهای مثبت ۱+ و برای ویژگیهای منفی ۱ـ را منظور کرد.

برای نمره‌گذاری مقیاس درجه‌بندی عددی می‌توان از اعداد روی مقیاس (مثلاً ۱ تا ۵) استفاده کرد. در مقیاسهای درجه‌بندی نگاره‌ای چون که ارزشیاب می‌تواند در روی خطوط مقیاس و در بین اعداد داوری خود را نشان دهد، نمرهٔ دانش‌آموز در ویژگی مورد سنجش می‌تواند عدد اعشاری هم باشد.

متخصصان سنجش، برای هر سه روش مشاهده‌ای مورد بحث در این فصل، رهنمودهایی را ارائه داده‌اند. ما آن رهنمودها را زیرعنوان قواعد تهیهٔ هر یک از روشها به تفصیل بحث کرده‌ایم.

روشهای سنجش مشاهده‌ای، مانند سایر روشهای سنجش مورد بحث در این کتاب، دارای امتیازها و محدودیتهایی هستند. امتیازهای این روشها را می‌توان به نحو زیر خلاصه کرد:

۱. فراهم آوردن امکانات برای جمع‌آوری اطلاعات دربارهٔ رفتارها، به ویژه رفتارهای سازگاری، هیجانی، و اجتماعی.

۲. فراهم‌آوردن امکان سنجش کاربست اطلاعات در موقعیتهای زندگی.

۳. بالا بردن قدرت پیش‌بینی نتایج سنجش مشاهده‌ای.
۴. سهولت استفاده با تکالیف و موقعیتهای گوناگون.
۵. به دست دادن نتایج قابل استفاده به عنوان منابع کمکی برای سایر روشهای سنجش.
۶. به دست دادن داده‌های کمّی و اطلاعات کیفی.
۷. امکان ادغام این روشها با برنامه‌های کلی آموزشی و دادن فرصت مشاهده به معلم به عنوان بخشی از فرایند آموزش.

برای روشهای سنجش مشاهده‌ای محدودیتهایی نیز قائل شده‌اند. این محدودیتها را به گونهٔ زیر می‌توان خلاصه کرد:

۱. وجود احتمال تأثیر تعصبات و سوگیریهای مشاهده‌گر درگزارش و تفسیر نتایج مشاهدات.
۲. نیاز به صرف وقت زیاد و استفاده از وسایل و تجهیزاتی که غالباً در مدارس در دسترس نیستند.
۳. امکان نتیجه گیریهای غلط.
۴. نشان ندادن علتهای رفتار.
۵. مشکل بودن مشاهدهٔ رفتارهای گوناگون در کلاس و متمایز ساختن رفتارهای مربوط از نامربوط.
۶. دشواری ثبت مشاهدات همزمان.
۷. قطعیت غیرموجه.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. در تهیهٔ فهرست وارسی می‌توان به جای فهرستی از ویژگیهای مثبت رفتار یا محصول رفتاری مورد نظر فهرستی از ویژگیهای منفی آن را قرار داد.
الف ـ صحیح
ب ـ غلط

۲. فهرست وارسی را به کدام صورت زیر می‌توان نمره گذاری کرد؟
الف ـ ۰ـ ۱
ب ـ بله ـ نه
پ ـ ۱ـ ۱+
ت ـ الف و پ
ث ـ الف و ب و پ

۳. شمارهٔ درجات مقیاس درجه‌بندی را از بین چه ارقامی برمی‌گزینند؟

الف ـ ۳ و ۵

ب ـ ۳ و ۷

پ ـ ۵ و ۷

ت ـ ۳ و ۹

ث ـ ۵ و ۹

۴. ویژگیهای فهرست وارسی و مقیاس درجه‌بندی را با هم مقایسه کنید.

۵. روش واقعه‌نگاری را تعریف کنید.

۶. انواع مقیاسهای درجه‌بندی را با ذکر مثال با هم مقایسه کنید.

۷. نمره‌گذاری روشهای فهرست وارسی و روش مقیاس درجه‌بندی را با هم مقایسه کنید.

۸. امتیازها و محدودیتهای روشهای سنجش مشاهده‌ای را توضیح دهید.

۹. موارد استفادهٔ روشهای سنجش مشاهده‌ای را توضیح دهید.

۱۰. در رابطه با جدول مشخصات آزمون درسی که به عنوان تمرین فصل ۶ این کتاب تهیه کرده‌اید، برای هدفهایی از آن درس که با روش فهرست وارسی قابل سنجش است فهرست وارسی بنویسید، و برای هدفهایی که با روش مقیاس درجه‌بندی قابل سنجش است مقیاس درجه‌بندی بنویسید. همچنین هدفهایی را که سنجش آنها نیازمند ثبت و تفسیر با روش واقعه‌نگاری است مشخص کنید.

۱۵

سنجش رفتاری

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. سنجش رفتاری را تعریف کند.

۲. ویژگیهای سنجش رفتاری را با توجه به نظریهٔ روانشناسی یادگیری زیربنای آن توضیح دهد.

۳. مفاهیم سنجش مستقیم، سنجش غیرمستقیم، سنجش مشاهده‌ای، و سنجش رفتاری را با هم مقایسه کند.

۴. با ذکر مثال، بُعدهای رفتار را تعریف کند.

۵. روشهای مختلف سنجش رفتاری را با ذکر مثالی برای هریک از آنها تعریف کند.

۶. تفاوت بین سنجش موقعیتی و سنجش در شرایط طبیعی را توضیح دهد.

۷. مشاهدهٔ نهان‌یابی و آموزش موقعیتی را با هم مقایسه کند.

۸. توضیح دهد چرا روش نظارت شخصی و گزارش شخصی را از روشهای سنجش مستقیم به حساب نمی‌آورند.

۹. روش سنجش شناختی - رفتاری را با روشهای سنجش رفتاری مقایسه کند.

۱۰. ویژگیهای سنجش روانی - فیزیولوژیکی را توضیح دهد.

۱۱. امتیازها و محدودیتهای روشهای سنجش رفتاری را شرح دهد.

۱۲. موارد استفادهٔ روشهای سنجش رفتاری را توضیح دهد.

۱۳. برای هدفهای آموزشی مختلف، به ویژه آن دسته از هدفها که با روشهای مرسوم قابل سنجش نیستند، روشهای مناسب سنجش رفتاری را به کار بندد.

ما در این فصل روشهایی را معرفی می‌کنیم که به روشهای سنجش رفتاری شهرت دارند. این روشها نیز مانند روشهای توضیح داده شده در فصل قبل بر مشاهده استوارند اما، چنان که در توضیحاتی که به دنبال می‌آیند خواهید دید، این روشها به یک دسته نظریهٔ یادگیری خاص وابسته‌اند که در آنها از مشاهده صرفاً با هدف شناسایی ویژگیهای رفتار مورد مشاهده استفاده می‌شود نه صفات درونی زیربنای آنها. روشهای سنجش رفتاری با غالب روشهای معرفی شده در فصلهای قبلی این کتاب هم از لحاظ هدف و هم از لحاظ فرضهای نظری زیربنای آنها متفاوت‌اند. اگرچه روشهای بحث شده در فصلهای قبلی به هیچ رویکرد یا نظریهٔ ویژهٔ روانشناسی وابستگی ندارند، اما روشهای این فصل منحصراً به دیدگاه روانشناسی رفتاری یا اصطلاحاً رفتارگرایی[1] وابسته‌اند. در زیر دربارهٔ این موضوع بیشتر توضیح می‌دهیم.

تعریف و ویژگیهای سنجش رفتاری

هرچند که همهٔ روشهای بحث شده در فصلهای قبلی این کتاب رفتار انسانها را می‌سنجند و از این رو می‌توان آنها را روشهای سنجش رفتاری به شمار آورد، اما از آنجا که غالب آنها از سنجش رفتاری به قصد شناسایی نوعی توانایی یا ویژگی درونی افراد استفاده می‌کنند، به معنی واقعی، سنجش رفتاری نیستند. به عنوان نمونه، درست است که ما به کمک یک پرسشنامهٔ نگرش سنج یا یک آزمون استعداد نمونه‌هایی از رفتارهای پاسخ‌دهندگان را سنجش می‌کنیم، اما هدف ما از این کار شناسایی آن رفتارها نیست بلکه پی‌بردن به یک ویژگی درونی افراد، یعنی نگرش یا استعداد آنان، است. از این رو، به روشهایی که به این منظور به کار می‌روند روشهای سنجش غیرمستقیم می‌گویند. در مقابل، به روشهای **سنجش رفتاری**[2] که موضوع فصل حاضرند **روشهای سنجش مستقیم** گفته می‌شود زیرا به طور بی‌واسطه و مستقیم رفتار را سنجش می‌کنند، و در استفاده از آنها سنجشگر صرفاً می‌خواهد رفتارها یا عملکردهای آشکار افراد را بسنجد و به دنبال هیچ‌گونه صفت، ویژگی، یا توانایی مکنون یا پنهان نیست. بنابراین، می‌توان روشهای سنجش رفتاری را به عنوان روشهایی که به طور مستقیم رفتارها یا عملکردهای فرد را می‌سنجند و هدف آنها اندازه‌گیری کمیت یا کیفیت رفتار است، نه هیچ‌گونه صفت یا ویژگی درونی تعریف کرد. بنا به گفتهٔ کوهن و سوردلیک (۲۰۰۵)،

> [طبق نظریه‌های رفتاری] صفات، حالات، انگیزه‌ها، سائقها، نیازها، و سایر سازه‌های روانی وجود خارجی ندارند. آنها سازه‌هایی هستند که وجودشان را باید از رفتار استنباط کرد.... رویکرد سنتی به سنجش روانی را می‌توان رویکرد **علامتی**[3] نامید، زیرا در این رویکرد از پاسخ آزمون به عنوان علامت برای شخصیت یا توانایی زیربنایی استفاده می‌شود.... در مقابل، در سنجش رفتاری تأکید بر رفتاری است که فرد انجام می‌دهد نه بر استنباطهایی دربارهٔ اینکه چه صفات یا ویژگیهایی را داراست. (ص ۴۰۲)

---

1. behaviorism 2. behavioral assessment 3. sign

روانشناسان رفتاری یا **رفتارگرایان**[1] بر این باورند که، به عوض استفاده از اصطلاحات مبهمی که به وجود خصیصه‌هایی در درون افراد اشاره می‌کنند، بهتر است خودِ رفتار مورد توجه و رسیدگی قرار گیرد. در نتیجه، به جای استفاده از آزمون و سایر ابزارهای سنجش غیرمستقیم، رفتارهای آشکار افراد را مستقیماً ثبت و ضبط می‌کنند. بنا به گفتهٔ هوگان (۲۰۰۳)،

> طرفداران سنجش رفتاری این گونه استدلال می‌کنند که اگر ما به اطلاعاتی دربارهٔ افسردگی، هراس (فوبی)، یا اضطراب نیاز داریم، چرا باید یک رشته سؤال تستی بپرسیم که صرفاً اطلاعات غیرمستقیمی دربارهٔ این ویژگیها به ما می‌دهند؟ چرا نباید آن ویژگیها را مستقیماً مشاهده و سنجش کنیم؟ مثلاً، چرا موردهایِ رفتار افسرده، رفتار هراس، یا رفتار اضطراب را مشاهده و ثبت ننماییم؟ (ص ۶۰۱)

بنا به توضیحات بالا، در رویکرد رفتاری، شخصی را که با صفت عقب‌ماندهٔ ذهنی معرفی می‌شود با ویژگیهای رفتاری مانند نمونه‌های زیر تعریف می‌کنند و به مشاهده و سنجش آنها می‌پردازند:

- *نمی‌تواند بند کفشش را ببندد.*
- *برای قضای حاجت از توالت استفاده نمی‌کند.*
- *نمی‌تواند با قاشق و چنگال غذا بخورد.*

همچنین دانش‌آموز مبتلا به ناتوانی یادگیری با رفتارهایی نظیر رفتارهای زیر تعریف می‌شود و همین رفتارها سنجش می‌شوند:

- *بیشتر از چند دقیقه به درس توجه نمی‌کند.*
- *کلمات مشابه را به جای هم به کار می‌برد.*
- *هنگام خواندن کلمات را وارونه می‌خواند.*

چنان که در صفحات بعدی این فصل خواهیم دید، سنجش رفتاری از فنون مختلفی برای جمع‌آوری اطلاعات برخوردار است، اما مهم‌ترین آنها مشاهدهٔ مستقیم رفتار در موقعیتهای طبیعی است. اسپیگلر[2] و گوریمونت[3] (۱۹۹۸)، در معرفی روش مشاهدهٔ طبیعی و سنجش مستقیم رفتار، به عنوان مثالی از تعریف دقیق رفتار، مورد زیر را در ارتباط با ایجاد رابطهٔ اجتماعی صمیمانه ذکر کرده‌اند. دقت کنید که در این مثال هم ویژگیهای رفتار موردنظر و هم تمایز آن با رفتارهای نامربوط به رفتار موردنظر مشخص شده‌اند. سنجشگرِ رفتاری باید به هنگام مشاهده و ثبت و ضبط این‌گونه رفتارها همهٔ ویژگیهای گفته شده را درنظر بگیرد.

---

1. behaviorists 2. Spiegler 3. Guevremont

رابطهٔ اجتماعی صمیمانه

تعریف: (۱) فرد موردنظر از طریق سلام و احوالپرسی یا آغاز کردن یک تعامل با شخص دیگر در ایجاد ارتباط اجتماعی پیش‌قدم می‌شود، (۲) هنگام صحبت کردن از یک لحن دوستانه استفاده می‌کند، (۳) ضمن صحبت کردن با شخص مورد خطاب مستقیماً به او نگاه می‌کند، و (۴) در شروع ارتباط در فاصلهٔ چهار یا پنج متری از او قرار می‌گیرد.

مثالها: معرفی خود به شخص دیگر، پرسیدن یک سؤال از شخص دیگر (مثلاً، "لطفاً بفرمایید درِ خروجی کجاست؟")، صدا زدن شخص دیگر با اسم کوچک، یا شروع یک گفتگو با شخص دیگر با یک ابرازنظر (مثلاً، "امروز بازی فوتبال خوبی انجام شد").

غیرمثالها: داد زدن یا از لحن غیردوستانه استفاده کردن، صحبت کردن با شخص دیگر تنها پس از آنکه آن شخص سر صحبت را باز کرد، صحبت کردن با یک شخص بدون نگاه کردن به او، یا ایجاد ارتباط اجتماعی از فاصله‌ای بیشتر از ۵ متر.

(اسپیگلر و گوریمونت، ۱۹۹۸، ص ۹۵)

پیش از ادامهٔ بحث دربارهٔ روشهای سنجش رفتاری، لازم است بگوییم که این روشها رقیب روشهای معرفی شده در دیگر فصلهای این کتاب نیستند. ما روشهای فصل حاضر را به عنوان مکمل روشهای بحث شدهٔ قبلی می‌دانیم و به معلمان سفارش می‌کنیم که هرجا از هر روشی که صلاح دیدند استفاده کنند و در صورت لزوم از ترکیبی از روشهای مختلف سود ببرند. گتینگر[1] و سیبرت[2] (۲۰۰۰) در این باره گفته‌اند:

> نظریه‌های رفتاری و شناختی به ایجاد چندین جهت تازه برای سنجش مسائل مربوط به مهارتهای تحصیلی کمک کرده‌اند. برای مثال، نظریه‌پردازان شناختی استفاده از تکالیف سنجشی را که بر تفکر سطح بالا و حل مسئله تأکید می‌کنند مورد حمایت قرار داده‌اند. نظریه‌پردازان رفتاری بر کاربرد سنجش به عنوان تابعی که در آن روابط میان رویدادهای محیطی (آموزش) و رفتارهای موردنظر (مهارتهای تحصیلی) مشخص می‌شوند تأکید کرده‌اند. (ص ۱۴۰)

بُعدهای رفتار

رفتار هم دارای ویژگیهای ظاهری و هم ویژگیهای زمانی است. به مجموعهٔ این ویژگیها **بعدهای رفتاری**[3] می‌گوییم. ما در این قسمت هریک از این بعدها را به کمک مثال توضیح می‌دهیم. در سنجش رفتاری هم می‌توان این بعدها را توصیف و اندازه‌گیری کرد و هم شرایط و موقعیتهای محیطی را که رفتارها در آن ظاهر می‌شوند سنجش کرد.

شکل ظاهری[4] **رفتار**، یعنی اینکه رفتار در چه شکل یا صورتی رخ می‌دهد.

---

1. Gettinger 2. Seibert 3. behavioral dimensions 4. topography

مثال

- پروین هنگام قشقرق جیغ می‌کشد، پاهایش را به زمین می‌کوبد، و موهایش را می‌کند.
- احمد س‌ها را وارونه می‌نویسد.

✓ **شدت یا نیروی[1] رفتار**، یعنی اینکه رفتار با چه شدتی رخ می‌دهد.

مثال

- زهرا هنگام نوشتن تکالیف قلمش را آن‌قدر بر روی کاغذ فشار می‌دهد که آن را سوراخ می‌کند.
- جمشید طوری آهسته حرف می‌زند که صدایش توسط همکلاسی‌هایش به خوبی شنیده نمی‌شود.

✓ **نرخ[2] یا فراوانی[3] رفتار**، یعنی تعداد دفعات وقوع رفتار. اگر صرفاً تعداد دفعات وقوع رفتار تعیین شود به آن فراوانی رفتار می‌گویند. اما اگر تعداد دفعات وقوع رفتار در یک واحد زمانی تعیین شود به آن نرخ رفتار گفته می‌شود.

مثال

- حسین ۸ مسئلهٔ ریاضی را حل کرد (فراوانی).
- بیژن در یک جلسهٔ درسِ یک ساعتی ۵ بار از کلاس خارج شد (نرخ).

✓ **طول مدت[4] رفتار**، یعنی زمانی که انجام دادن یک کار یا یک پاسخ به طول می‌انجامد.

مثال

- فاطمه یک ساعت صرف انشاء نوشتن کرد.
- قشقرق پروین ۲۵ دقیقه ادامه داشت.

✓ **دورهٔ نهفتگی[5] رفتار**، یعنی مدتی که به طول می‌انجامد تا شخص انجام کار یا رفتاری را که از او خواسته شده است شروع کند.

مثال

- پس از آنکه به عباس گفته شد که از توالت بیرون بیاید، پنج دقیقه طول کشید تا از توالت خارج شد.
- بعد از آنکه پروین را برای قشقرق راه‌انداختن در اتاقِ محرومیت قرار دادند، ۱۰ دقیقه طول کشید تا دست از قشقرق برداشت.

✓ **مکان یا موقعیت[6] انجام رفتار**، یعنی جایی که رفتار در آن رخ می‌دهد.

---

1. force 2. rate 3. frequency 4. duration 5. latency 6. locus

مثال

- مرتضی معمولاً در شلوارش ادرار می‌کند.
- پروین بیشتر در اتاق غذاخوری قشقرق راه می‌اندازد.

ما در فصل حاضر به معرفی روشهایی می‌پردازیم که به کمک آنها ویژگیها یا بُعدهای مختلف رفتاری فوق سنجش می‌شوند. از آنجاکه روشهای سنجش رفتاری متنوع‌اند، ابتدا آنها را دسته‌بندی می‌کنیم و بعد به توضیح هریک از آنها می‌پردازیم.

انواع روشهای سنجش رفتاری

روشهای سنجش رفتاری بسیار متنوع‌اند. ما در این فصل انواع مهم آنها را با ذکر مثال توضیح می‌دهیم.

سنجش محصولات ماندنی رفتار

بعضی از رفتارهای انسان، مانند فعالیتهای آموزشگاهی، از خود محصولاتی پایدار به جا می‌گذارند به گونه‌ای که سنجشگر یا معلم می‌تواند سر فرصت به مشاهده و ارزشیابی آنها بپردازد. از جملهٔ این محصولات می‌توان به نتایج انجام دادن تکالیف درسی، حل کردن مسائل ریاضی، نوشتن دیکته و انشاء از سوی دانش‌آموزان، و سایر فراورده‌های مشابه اشاره کرد.

روش سنجش این محصولات رفتاری پس از حصول و حتی مدتی پس از آن امکان‌پذیر است. از این رو، به این روشهای سنجش نام بعد از واقعه[1] نیز داده شده. با این حال، بهتر است در اولین فرصت به سنجش این‌گونه محصولات اقدام شود و دانش‌آموز یا دانشجو از نتیجهٔ کار خود آگاه گردد. به منظور ارزشیابی و نمره‌گذاریِ محصولات ماندنی رفتار باید از همان روشهایی که در فصلهای قبلی این کتاب، به ویژه در فصلهای مربوط به آزمونهای انشایی و عملکردی (۱۰ و ۱۲) معرفی شدند، استفاده کرد.

از آنجاکه روشهای سنجش فراوردهٔ رفتار روشهای نسبتاً ساده و شناخته شده‌ای هستند، بهتر است معلمان بکوشند تا حد امکان رفتارها را بر حسب فراورده یا محصولشان تعریف و سنجش کنند. تعریفی که معلم یا سنجشگر برای رفتار یا فراوردهٔ رفتاری به دست می‌دهد باید دقیق و عینی باشد به گونه‌ای که بتواند آن را به خوبی مشاهده و سنجش کند. آلبرتو[2] و تروتمن[3] (۱۹۹۱) در این باره مثالهای زیر را ذکر کرده‌اند:

---

1. expost facto 2. Alberto 3. Troutman

اضطراب آزمون، بر حسب تعداد خط خوردگیها یا موارد پاک شده در برگهٔ آزمون دانش‌آموز یا دانشجو.

شلختگی، بر حسب تعداد خُرده کاغذهایی که در شعاع یک‌متری صندلی دانش‌آموز ریخته شده.

سنجش رویدادهای گذرای رفتار (ثبت مشاهده‌ای)

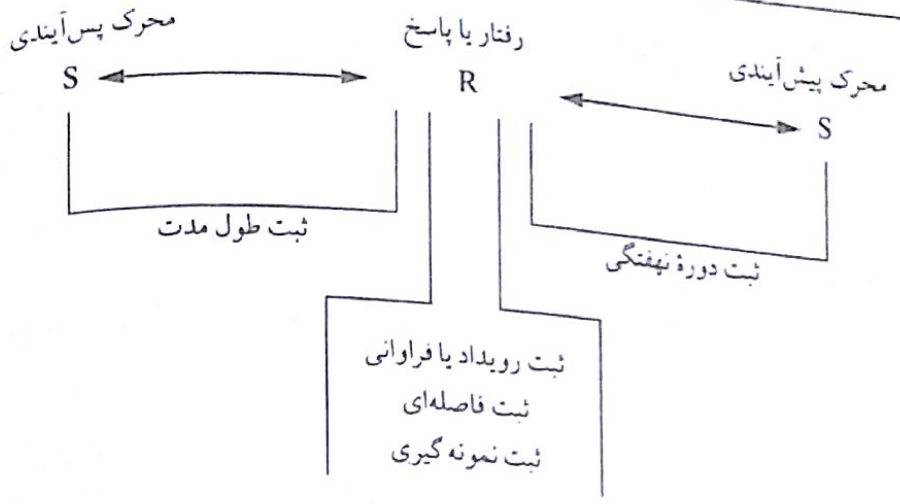
بسیاری از رفتارهای دانش‌آموزان و دانشجویان در محیطهای طبیعی کلاس و آموزشگاه رخ می‌دهند و هیچ‌گونه اثر یا محصول پایداری که بتوان آنها را سنجش کرد از خود به جای نمی‌گذارند. به عنوان نمونه می‌توان شرکت در بحث گروهی، لبخندزدن، اخم کردن، خارج از نوبت صحبت کردن، دعوا راه‌انداختن، گریستن، انجام رفتار شیطنت‌آمیز، یا تبلی را نام برد. این رفتارها را باید در ضمن وقوع و از طریق مشاهدهٔ طبیعی سنجش کرد. از این رو، به روشهای سنجش این رفتارها نام روشهای **ثبت مشاهده‌ای**[1] نیز داده شده.

به عنوان موردی از کاربرد روش ثبت مشاهده‌ای، دانش‌آموزی را درنظر بگیرید که در آموزشگاه رفتارهای پرخاشگرانه انجام می‌دهد. می‌توانیم از یک مشاهده‌گر بخواهیم تا فراوانی، شکل ظاهری، طول مدت، و شدت این رفتار را مشاهده و ثبت نماید. اینها بُعدهای رفتار پرخاشگرانه هستند که ثبت و سنجش آنها برای شناسایی و درمان پرخاشگری ضروری‌اند. روانشناس آموزشگاه که این مشکل به او ارجاع خواهد شد ممکن است از این اطلاعات به عنوان خط پایه‌ای برای مقایسهٔ رفتارهای بعدی دانش‌آموز و رسیدگی به جریان پیشرفت درمان در او استفاده نماید.

روشهای ثبت مشاهده‌ای انواع مختلفی دارند. اگر معلم یا سنجشگر بخواهد تعداد دفعات وقوع رفتار را ثبت کند، از روش **ثبت فراوانی** بهره می‌گیرد؛ اگر بخواهد طول مدتی را که فرد صرف انجام رفتار می‌کند ثبت نماید، از روش ثبت **طول مدت رفتار** استفاده می‌کند؛ اگر بخواهد مدتی را که طول می‌کشد تا فرد کاری را شروع کند اندازه بگیرد، روش **ثبت دورهٔ نهفتگی** را به کار می‌گیرد؛ و اگر قصد او تعیین نسبتی از زمانی است که در آن رفتار رخ می‌دهد، از روش ثبت فاصله‌ای یا روش ثبت **نمونه گیری زمان** سود می‌جوید.

در سنجش رفتاری، علاوه بر رفتار موردنظر، معمولاً محرکهای **پیش‌آیندی** و محرکهای **پس‌آیندی** رفتار نیز مورد مشاهده قرار گرفته و ثبت می‌شوند، زیرا در توصیف و تحلیل رفتار برای مقاصد آموزشی و بالینی مفیدند. رابطهٔ بین روشهای ثبت مشاهده‌ای و اجزای سه‌گانهٔ محرک پیش‌آیندی، رفتار، و محرک پس‌آیندی در شکل ۱-۱۵ نشان داده شده. منظور از محرک پیش‌آیندی در این شکل محرکهای ایجادکنندهٔ رفتار (مثلاً درخواست معلم از دانش‌آموز برای

1. observational recording



شکل ۱-۱۵ روشهای سنجش مشاهده‌ای یا سنجش فرایند رفتار (آلبرتو و تروتمن، ۱۹۹۱، ص ۱۰۵)

انجام دادن کاری) است. منظور از رفتار یا پاسخ، عمل یا کاری است که دانش‌آموز انجام می‌دهد. و منظور از محرک پس‌آیندی نتیجه‌ای است که رفتار برای دانش‌آموز به بار می‌آورد (مثلاً تقویت یا پاداشی که از سوی معلم دریافت می‌کند)[1].

ثبت فراوانی یا ثبت رویداد رفتار

اگر دفعات وقوع یک رفتار یا یک پاسخ را در یک فاصلهٔ زمانی معین شمارش و ثبت کنیم از روش **ثبت رویداد**[2] استفاده کرده‌ایم. گاهی به روش ثبت رویداد روش **ثبت فراوانی**[3] نیز گفته می‌شود. اصطلاح دیگری که غالباً به جای فراوانی به کار می‌رود **نرخ** است. بنابراین، به روش ثبت رویداد یا ثبت فراوانی رفتار روش **ثبت نرخ رفتار**[4] نیز گفته می‌شود. روش ثبت فراوانی یا ثبت نرخ رفتار متداول‌ترین روش ثبت مشاهده‌ای است. در این روش فاصله‌های زمانی مختلفی مورد استفاده قرار می‌گیرند، مثلاً ده دقیقه، یک جلسه، یک ساعت، یک روز، طول مدت صرف یک وعده غذا، طول مدت یک برنامهٔ تلویزیونی، و مانند اینها. بعد تعیین می‌شود که رفتار در هر یک از این فاصله‌های زمانی چند بار رخ می‌دهد. دقت شود که در این روش سنجش رفتاری، در هر فاصلهٔ زمانی، تمامی موارد بروز رفتار ثبت می‌شوند.

روش ثبت فراوانی رفتار برای اندازه‌گیری رفتارهایی که مجزا و مشخص‌اند بسیار مفید است ــ یعنی رفتارهایی که آغاز و انجام تعریف شده‌ای دارند. از جملهٔ این گونه رفتارها می‌توان به تعداد صفحاتی که دانش‌آموز مطالعه می‌کند، تعداد جلساتی که در کلاس درس

---

۱. برای کسب اطلاعات بیشتر در این‌باره، به اثر دیگر مؤلف با عنوان تغییر رفتار و رفتاردرمانی: **نظریه‌ها و روشها** (ویرایش دوم). تهران: دوران (۱۳۷۷) مراجعه کنید.

2. event recording 3. frequency recording 4. behavioral rate recording

حاضر می‌شود، تعداد جوابهای درستی که می‌دهد، تعداد دقایقی که دیر به کلاس می‌رسد، تعداد دفعاتی که بغل دستی‌اش را می‌زند یا او را گاز می‌گیرد، تعداد حرفهای زشتی که به زبان می‌آورد، و مانند اینها اشاره کرد.

معمولاً داده‌های خام حاصل از مشاهدات رفتاری را در برگه‌ای که به صورت جدول تنظیم شده و به **برگهٔ ثبت داده**[1] معروف است وارد می‌کنند. به عنوان موردی از ثبت فراوانی رفتار به مثال زیر که از کتاب تغییر رفتار (مارتین[2] و پیبر[3]، ۱۹۹۲) نقل شده است توجه کنید. فرض کنید تغییر دهندهٔ رفتار می‌خواهد مشکلات غذا خوردن یک کودک که ریختن غذا و با دست غذا خوردن است را برطرف کند و به او رفتار درست غذا خوردن را آموزش دهد. بنابراین، یکی از اقدامات او این است که ابتدا فراوانی ریختن غذا و فراوانی با دست غذا خوردن را در ضمن صرف غذا توسط کودک ثبت کند. اولین گام این خواهد بود که یک تعریف دقیق رفتاری از ریختن غذا و با دست غذا خوردن به عمل آورد که هر کسی که کودک را در حال انجام این رفتارها می‌بیند بتواند موارد رفتار نامطلوب را به سهولت تشخیص دهد. فرض کنید رفتارهای فوق به نحو زیر تعریف شده‌اند:

ریختن غذا: ریختن غذا به خارج از دهان یا بشقاب غذا، در نتیجهٔ یکی از علتهای زیر:
حرکت قاشق توسط کودک از بشقاب به دهان یا برعکس
پر کردن قاشق با غذا توسط کودک یا بریدن غذا با کارد
حرکت بشقاب غذا توسط کودک
ریختن غذا از دهان کودک

هر یک از رویدادهای بالا به عنوان موردی از ریختن غذا ثبت خواهد شد، بدون توجه به اینکه مقدار غذای ریخته شده چقدر است.

با دست غذا خوردن: هر زمان که کودک غذا را برای گذاشتن در دهان با دستش لمس کند.

اگر کودک هنگام غذا خوردن هم غذا را بریزد و هم با دستش غذا بخورد هر دو مورد ریختن و با دست غذا خوردن ثبت می‌شوند.

حال مشاهده‌گر می‌داند که به دنبال چه رفتاری باید باشد. برای ثبت فراوانی رویدادهای بالا، برگهٔ ثبت داده که در شکل ۲-۱۵ نشان داده شده مورد استفاده قرار می‌گیرد.

چنان که گفتیم، با روش ثبت رویداد عمدتاً فراوانی یا دفعه‌های وقوع رفتار موردنظر مشاهده و ثبت می‌شوند. داده‌های حاصل از این روش به سادگی قابل تبدیل به نرخ یا درصد هستند. به عنوان مثالی برای انجام این کار، فرض کنید در یک جلسهٔ پرسش و پاسخ، معلم از

1. data sheet 2. Martin 3. Pear

مشاهده‌گر: جان اچ

مشاهدات

تاریخ: اول ژانویه

دانش‌آموز: کورین

	موارد	جمع موارد	زمان	توضیحات اضافی
ریختن غذا	//// //// //	۱۲	۲۰ دقیقه	هنگام صرف شام، سه دانش‌آموز دیگر در سر میز غذا حضور دارند. شام از موارد زیر تشکیل می‌شود: سوپ، پورهٔ سیب‌زمینی، همبرگر، سبزی، ژله، و شیر
با دست غذا خوردن	///	۳	۲۰ دقیقه	هنگام صرف شام، سه دانش‌آموز دیگر در سر میز غذا حضور دارند. شام از موارد زیر تشکیل می‌شود: مانند بالا

شکل ۲-۱۵ نمونه‌ای از برگهٔ ثبت داده برای رفتار ریختن غذا و با دست غذا خوردن
(مارتین و پیبر، ۱۹۹۲، ص ۲۵۸)

کلاس ۲۰ سؤال می‌پرسد و یکی از دانش‌آموزان ۵ بار دستش را برای جواب دادن به سؤال بلند می‌کند. این دانش‌آموز در ۲۵٪ موارد کوشیده است تا به سؤالهای معلم جواب دهد. در ضمن اگر جلسهٔ پرسش و پاسخ را یک ساعت فرض کنیم، نرخ دست‌بلندکردنهای او نیز ۲۵ بوده است. تبدیل داده‌ها از فراوانی به نرخ یا درصد کار مقایسه را آسان می‌کند. مثلاً اگر در یک جلسهٔ ۴۰ دقیقه‌ای یک دانش‌آموز ۵ تمرین ریاضی حل کند و در یک جلسهٔ ۲۰ دقیقه‌ای دانش‌آموز دیگری نیز ۵ تمرین حل نماید، فراوانی حل کردن مسائل هر دو دانش‌آموز ۵ است، ولی اگر ندانیم که دانش‌آموز اول زمانی دوبرابر دانش‌آموز دوم در اختیار داشته است تصور خواهیم کرد که میزان حل کردن مسائل در هر دو نفر برابر بوده. وقتی که فراوانی‌های این دو دانش‌آموز را به درصد یا به نرخ تبدیل کنیم معلوم می‌شود که دانش‌آموز دوم دوبرابر دانش‌آموز اول کار کرده است.

ثبت طول مدت رفتار

معلم یا سنجشگر در برخورد با بعضی از رفتارها، نظیر قشقرق[1] راه انداختن کودکان، بیشتر از اینکه به فراوانی (دفعات) آنها علاقه‌مند باشد به طول مدت آنها نظر دارد و روش موردنیاز او **ثبت طول مدت**[2] است. رفتارهایی که طول مدت وقوع آنها مهم‌تر از تعداد آنهاست عبارت‌اند از: زمان توجه کردن کودک، ترک کردن صندلی در کلاس درس، تراشیدن مداد، رؤیای روز، نوشتن تکالیف درسی، و مانند اینها. برای سنجش این گونه رفتارها از روش ثبت طول مدت

1. temper tantrum 2. duration recording

رفتار استفاده می‌شود. برای نمونه، معلمی که می‌خواهد طول مدت حل کردن یک مسئلهٔ ریاضی را توسط یک دانش‌آموز ثبت کند، زمانی را که کودک شروع به حل کردن مسئله می‌کند و زمانی را که حل آن به پایان می‌رسد از روی ساعت خود تعیین می‌نماید و تفاوت بین این دو زمان را محاسبه می‌کند و نتیجهٔ حاصل طول مدتی است که دانش‌آموز مشغول حل کردن مسئله بوده است. یا مثلاً اگر معلم بخواهد مدت تأخیر ورود دانش‌آموزی را به کلاس تعیین کند، زمانی را که کودک قرار بوده است در کلاس حاضر شود (ساعت ۸) از زمانی که وارد کلاس می‌شود (۸:۱۵) کم می‌کند و نتیجهٔ حاصل (۱۵ دقیقه) مقدار تأخیر ورود او را به کلاس نشان خواهد داد.

ثبت دورهٔ نهفتگی یا درنگ رفتار

به کمک روش **ثبت دورهٔ نهفتگی**[1] می‌توان تعیین کرد که چه زمانی طول می‌کشد تا شخص رفتاری را آغاز کند. یعنی با این روش مدت زمان بین لحظه‌ای که از شخص انجام رفتاری خواسته می‌شود تا لحظه‌ای که او عملاً شروع به انجام آن رفتار می‌کند تعیین می‌گردد. برای مثال، دانش‌آموزی را درنظر بگیرید که پس از شروع به کار خوب پیش می‌رود، اما مشکل او این است که خیلی دیر دست به کار می‌شود؛ یعنی پس از آنکه معلم از او می‌خواهد کاری را شروع کند، مرتب این دست آن دست می‌کند و مدتی از وقتش را برای آماده شدن به هدر می‌دهد. روش ثبت دورهٔ نهفتگی این گونه تأخیرها یا درنگها را در آغاز کردن کار یا رفتار می‌سنجد. روش ثبت دورهٔ نهفتگی شبیه به روش ثبت طول مدت رفتار است.

به منظور روشن ساختن بُعدهای رفتاری بالا (فراوانی، طول مدت، و دورهٔ نهفتگی)، برای هریک مثالی دیگر می‌آوریم. اگر دانش‌آموزی برای جواب دادن به سؤالهای محاسباتی در کلاس درس ریاضی وقت زیادی صرف می‌کند، دورهٔ نهفتگی پاسخدهی به سؤالهای معلم بُعد رفتاری موردنظر است. اگر دانش‌آموز زمان اندکی را صرف نگاه کردن به سوی معلم، وقتی که مشغول خواندن داستان در کلاس داستان‌خوانی است، می‌کند، طول مدت رفتار بُعد موردنظر است. و اگر دانش‌آموز در جلسهٔ درس نوشتن، تعداد کلمات اندکی را در کتابچهٔ خود می‌نویسد، بُعد رفتاری موردنظر فراوانی یا نرخ است (دالی[2] و مورداک[3]، ۲۰۰۰).

ثبت فاصلهٔ زمان یا ثبت فاصله‌ای

بهترین روش سنجش یک رفتار آن است که تمامی موارد آن را به طور مداوم مورد مشاهده قرار دهیم و جزئیات وقوع آن را ثبت نماییم. به روش ثبت تمامی موارد یک رفتار در یک مدت زمان مشخص روش **ثبت مداوم**[4] می‌گویند. ولی این روش به دلیل وقت‌گیر بودن و پر هزینه بودن چندان مورد استفاده قرار نمی‌گیرد و به جای آن معمولاً از روش **ثبت فاصلهٔ زمان**[5] یا روش **ثبت فاصله‌ای** استفاده می‌کنند. روش ثبت فاصله‌ای و روش ثبت نمونه گیری زمان که بعداً

1. latency recarding 2. Daly 3. Murdoch 4. continuous recording 5. interval recording

معرفی خواهند شد برای رفتارهایی که آغاز و انجام قابل تمایزی ندارند مناسب‌اند. مثلاً فرض کنید کودکی سر و صداهای مزاحمی از خود تولید می‌کند، مانند داد زدن، روی میز کوبیدن، و جا به جا کردن صندلی. شمردن تعداد دفعات این گونه رفتارها یا اندازه‌گیری طول مدت وقوع آنها کار دشواری است، اما می‌توان بود یا نبود چنین رفتارهایی را در فاصله‌های زمانی معین تعیین کرد.

در روش ثبت فاصلهٔ زمان، یک مقطع زمانی معین، مثلاً یک زمان ۶۰ دقیقه‌ای انتخاب می‌شود. بعد این زمان به فاصله‌های مساوی کوچک‌تر (که معمولاً بیشتر از ۳۰ ثانیه نیستند) تقسیم می‌شود. در ضمنِ مشاهده، چنانچه رفتار مورد نظر در هر یک از فاصله‌های مثلاً ۱۰ ثانیه‌ای رخ بکند، وقوع آن ثبت می‌شود. رفتار در هر یک از فاصله‌های زمانی، تنها یک بار ثبت می‌شود حتی اگر بیشتر از یک بار هم رخ دهد. در ضمن طول مدت وقوع رفتار در هر یک از فاصله‌های زمانی هر اندازه که باشد، چه کم چه زیاد، باز هم رفتار در هر فاصلهٔ زمانی تنها یک بار ثبت می‌شود. همچنین رفتار در هر وقت که در طول فاصلهٔ زمانی رخ دهد، چه در آغاز چه در وسط چه در آخر، فرقی نمی‌کند.

شکل ۳-۱۵ برگهٔ ثبت داده‌های رفتارِ ترک کردن صندلی یک دانش‌آموز را نشان می‌دهد. داده‌های این شکل برای مدت ۱۵ دقیقه به دست آمده و به فاصله‌های ۱۰ ثانیه‌ای تقسیم شده است. در این شکل علامت (+) مُعرف انجام دادن رفتار مورد سنجش (ترک کردن صندلی) و علامت (-) نشانگر انجام ندادن رفتار ترک کردن صندلی (نشستن در صندلی) است. مشاهده‌گر (معلم) دانش‌آموز موردنظر را در فاصله‌های ۱۰ ثانیه‌ای مشاهده می‌کند. اگر دانش‌آموز تمام فاصلهٔ ۱۰ ثانیه‌ای را در صندلی خود نشسته باشد، این فاصله را با علامت (-) مشخص می‌کند. اما اگر هر زمان در خلال این مدت ۱۰ ثانیه‌ای دانش‌آموز صندلی خود را بدون اجازه ترک کند برای آن فاصله علامت (+) را قرار می‌دهد.

در شکل ۳-۱۵ داده‌های مربوط به ردیف اول بالای شکل (یک دقیقهٔ اول) نشان‌دهندهٔ این است که دانش‌آموز در ۳۰ ثانیهٔ اول سرجایش نشسته بوده، اما در ۳۰ ثانیهٔ دوم در صندلی‌اش قرار نداشته است.

با مشاهدهٔ علائم مربوط به وقوع و عدم وقوع رفتار در یک فاصلهٔ زمانی، می‌توان از داده‌های رفتاری اطلاعات زیر را استنباط کرد:

- تعداد تقریبی رویدادهای رفتاری
- طول مدت تقریبی رفتار در درون دورهٔ مشاهده
- توزیع رفتار در طول دورهٔ مشاهده

دانش‌آموز : مالکوین
تاریخ : ۸۲/۲/۲
مشاهده‌گر : آقای رایلی
شروع : ۹:۱۵
پایان : ۹:۳۰
رفتار : ترک کردن صندلی

+ = وقوع رفتار
− = عدم وقوع رفتار

خلاصهٔ داده‌ها
تعداد فاصله‌هایی که در آنها رفتار اتفاق افتاده: ۳۸
درصد فاصله‌هایی که در آنها رفتار اتفاق افتاده: ۴۲٪
تعداد فاصله‌هایی که در آنها رفتار اتفاق نیفتاده: ۵۲
درصد فاصله‌هایی که رفتار در آنها اتفاق نیفتاده: ۵۸٪

دقیقه‌ها	فاصله‌های ۱۰ ثانیه‌ای ۱	۲	۳	۴	۵	۶
۱	−	−	−	+	+	+
۲	+	+	+	+	+	+
۳	+	+	+	+	+	+
۴	+	+	−	−	−	−
۵	−	−	−	−	+	−
۶	−	−	+	−	−	−
۷	−	−	−	−	+	−
۸	+	+	+	+	+	+
۹	+	+	+	−	+	+
۱۰	+	+	+	+	+	−
۱۱	−	−	−	−	−	−
۱۲	−	−	−	−	−	−
۱۳	−	−	−	+	−	−
۱۴	−	−	−	−	−	+
۱۵	−	−	−	−	−	−

شکل ۳-۱۵ سنجش رفتار ترک کردن صندلی با استفاده از روش ثبت فاصله‌ای (آلبرتو و تروتمن، ۱۹۹۱، ص ۱۱۶)

ثبت نمونه‌گیری زمان

در روش **ثبت نمونه‌گیریِ زمان**[1] وقوع یا عدم وقوع یک رفتار در فاصله‌های کوتاه مشاهده و ثبت می‌شود. بنا به تعریف، "ثبت حضور یا عدم حضور رفتار در طول فاصله‌های زمانی کوتاه و همانند را روش ثبت نمونه گیری زمان می‌نامند" (سولزر ـ آزاروف[2]، و مایر[3]، ۱۹۷۷، ص ۲۶۴). در روش ثبت نمونه گیری زمان، مانند روش ثبت فاصله‌ای، رفتار در خلال فاصله‌های زمانی

1. time sampling recording 2. Sulzer-Azaroff 3. Mayer

معین مورد مشاهده قرار می‌گیرد و وقوع یا عدم وقوع آن ثبت می‌شود. تنها تفاوت بین این دو روش آن است که در روش ثبت فاصله‌ای رفتار مورد نظر در یک مدت زمان معین، مثلاً ۶۰ دقیقه، به طور پیوسته اما در فاصله‌های کوتاه مثلاً ۱۰ ثانیه مشاهده می‌شود، در حالی که در روش ثبت نمونه گیری زمان رفتار مورد نظر باز هم در فاصله‌های کوتاه، مثلاً همان ۱۰ ثانیه، مشاهده می‌شود اما نه به طور پیوسته بلکه با فاصله‌های زمانی بین مقاطع ۱۰ ثانیه‌ای.

به عنوان مثالی برای استفاده از روش ثبت نمونه گیری زمان، فرض کنید پدر یا مادر کودکی می‌خواهد رفتارهای نشستن، ایستادن، یا به جلو و عقب تکان خوردن کودک خود را مشاهده و ثبت کند. البته بهترین کار آن است که در تمام طول شبانه‌روز که کودک بیدار است او را مشاهده کند و هر زمان که رفتارهای مورد نظر اتفاق می‌افتند آنها را ثبت نماید. اما این کار عملی نیست. در عوض پدر یا مادر تصمیم می‌گیرد، در آغاز هر ساعت از اوقات بیداری کودک، او را به مدت ۱۵ ثانیه مورد مشاهده قرار دهد و وقوع یا عدم وقوع هر یک از رفتارهای فوق را ثبت کند. پس از این فاصلهٔ مشاهده، تا آغاز ساعت دیگر به کارهای خود می‌پردازد و بدین طریق در هر ساعت به مدت ۱۵ ثانیه نمونه‌ای از رفتارهای کودک را به دست می‌آورد. شکل ۴-۱۵ برگهٔ ثبت داده‌های این مورد ثبت نمونه گیری زمان را نشان می‌دهد.

اظهارنظر	موقعیت			رفتار		
	اتاق‌خواب	اتاق‌نشیمن	آشپزخانه	تکان‌خوردن	ایستادن	نشستن
۸:۰۰ پیش‌ازظهر	+					+
۹:۰۰					+	
۱۰:۰۰		+		+		
۱۱:۰۰						
۱۲:۰۰ بعدازظهر						
۱:۰۰						
۲:۰۰						
۳:۰۰						
۴:۰۰						
۵:۰۰						
۶:۰۰						
۷:۰۰						
۸:۰۰						
۹:۰۰ بعدازظهر						

شکل ۴-۱۵ سنجش رفتارهای نشستن، ایستادن، و به جلو و عقب تکان خوردن با استفاده از روش ثبت نمونه گیری زمان (مارتین و پیپر، ۱۹۹۲، ص ۲۶۸)

طبق شکل ۴-۱۵، رفتارهای مختلف کودک در اول سه ساعت صبح (۸ و ۹ و ۱۰) مورد مشاهده قرار گرفته و نتایج در سه ردیف نخست شکل وارد شده‌اند. این نتایج نشان می‌دهند که در ۱۵ ثانیه بعد از ساعت هشت، کودک در اتاق خواب در حال نشستن بوده است؛ در ۱۵ ثانیه بعد از ساعت ۹، در آشپزخانه در حال ایستادن بوده؛ و در ۱۵ ثانیه بعد از ساعت ۱۰ در اتاق نشیمن مشغول تکان دادن خود بوده است. والدین می‌توانند این مشاهدات را ادامه بدهند و برای تمام ساعتهای روز داده‌ها را در شکل وارد کنند.

معمولاً از روش ثبت فاصله‌ای و روش نمونه گیری زمان با هم استفاده می‌شود، زیرا امکان دارد که در ضمن مشاهده و ثبت رفتار با روش ثبت فاصله‌ای قسمتی از رفتار مورد مشاهده از نظر بیفتد. برای مثال، فرض کنید مشاهده گری می‌کوشد تا وقوع یا عدم وقوع رفتاری را در فواصل ۱۰ ثانیه‌ای ثبت کند. ممکن است رفتار در آغاز فاصلهٔ ۱۰ ثانیه‌ای بعدی اتفاق بیفتد، در حالی که مشاهده گر مشغول ثبت نتایج فاصلهٔ ۱۰ ثانیهٔ قبلی است، و بنابراین وقوع رفتار نادیده گرفته شود. پس، روش معمول این است که مشاهده گر دانش‌آموز را برای یک مدت مشخص (مثلاً ۱۰ ثانیه) مورد مشاهده قرار دهد و بعد در طول فاصلهٔ زمانی بعد (۱۰ ثانیه بعدی) به ثبت رفتار بپردازد و این کار را به مدت نیم ساعت ادامه دهد.

مورد استفادهٔ دیگر ترکیب دو روش ثبت فاصله‌ای و ثبت نمونه گیری زمان هنگامی است که مشاهده گر بخواهد رفتار چند دانش‌آموز را همزمان ثبت کند. در چنین حالتی، مشاهده گر می‌تواند یک دانش‌آموز را برای مدت ۱۰ ثانیه مشاهده کند و وقوع یا عدم وقوع رفتار او را ثبت نماید، بعد دانش‌آموز دیگری را برای ۱۰ ثانیه مشاهده کند و وقوع یا عدم وقوع رفتار او را نیز ثبت نماید، و الی آخر، تا اینکه تمام دانش‌آموزان مورد مشاهده قرار بگیرند. بار دوم، باز هم به همین نحو تمام دانش‌آموزان را مورد مشاهده قرار می‌دهد. "به بیان دقیق، چنین نظام مشاهده‌ای را می‌توان به عنوان روش نمونه گیری زمان همراه با فاصلهٔ زمانی بسیار مختصر بین فواصل مشاهده نیز توصیف کرد" (مارتین و پیپر، ۱۹۹۲، ص ۲۶۸).

به عنوان نمونه، فرض کنید معلمی بخواهد رفتار توجه کردن چهار دانش‌آموز را در یک جلسهٔ درسِ مطالعه از روی کتاب به طور انفرادی سنجش کند. به این منظور یک دقیقهٔ اول کلاس را به فاصله‌های ۱۵ ثانیه‌ای تقسیم می‌کند و آن را به یکی از دانش‌آموزان اختصاص می‌دهد. یک دقیقهٔ دوم را نیز به ۴ فاصلهٔ ۱۵ ثانیه‌ای تقسیم می‌کند و آن را به دانش‌آموز دوم اختصاص می‌دهد. به همین ترتیب برای دانش‌آموزان دیگر نیز عمل می‌نماید (مطابق شکل ۵-۱۵). چنان‌که در شکل ۵-۱۵ دیده می‌شود در ردیف بالای برگ ثبت داده‌ها نام دانش‌آموزان نوشته می‌شود و برای وقوع رفتاری که به دقت توصیف شده است در خانه‌های زیر نام هر دانش‌آموز علامت گذاشته می‌شود.

امتیاز روش سنجش نمونه گیری زمان این است که مشاهده گر، مثلاً معلم، لازم نیست تمام مدت شاهد رفتار شخص مورد مشاهده باشد. بلکه تنها زمانی که می‌خواهد رفتار او را

	اولین فاصله‌های ۱۵ ثانیه‌ای	دومین فاصله‌های ۱۵ ثانیه‌ای	سومین فاصله‌های ۱۵ ثانیه‌ای	چهارمین فاصله‌های ۱۵ ثانیه‌ای
	احمد	پرویز	محسن	جمشید
۱				
۲				
۳				
۴				

شکل ۵-۱۵ سنجش رفتار توجه کردن به درس در چهار دانش‌آموز به طور همزمان با استفاده از روش ثبت نمونه‌گیری زمان

اندازه‌گیری کند به آن توجه می‌نماید و در اوقات دیگر می‌تواند به وظایف عادی کلاسداری و آموزش خود بپردازد.

آلبرتو و تروتمَن (۱۹۹۱) پیشنهاد داده‌اند که بهتر است در روش نمونه‌گیری زمان، به جای ثانیه‌ها، از دقیقه‌ها استفاده کرد، زیرا "چنین کاری فرصت مشاهدهٔ رفتار را برای مدت طولانی‌تر فراهم می‌آورد" (ص ۱۱۸). همچنین موقعی که فاصله‌های زمانی به دقیقه تقسیم می‌شوند، فاصلهٔ بین مشاهدات بیشتر خواهد بود، و بنابراین امکان آموزش دادن همراه با جمع‌آوری اطلاعات برای معلم فراهم می‌آید. با این حال، باید دقت کرد که اگر طول فاصله‌های زمانی خیلی زیاد باشد، شباهت بین داده‌های به دست آمده و رفتارهایی که به واقع رخ می‌دهند کاهش می‌یابد.

☆ ☆ ☆ ☆ ☆

روشهای سنجش رفتاری بالا، چنان که دیدیم، عمدتاً روشهای **مشاهدهٔ رفتاری**[1] هستند. در این روشها فعالیتها یا رفتارهای فرد مورد مشاهده قرار می‌گیرد و ویژگیها یا بُعدهای آن ثبت می‌شوند (کوهن و سوردلیک، ۲۰۰۵). در سنجش به کمک مشاهدهٔ رفتاری، هم سنجشگر (معلم، مشاور، یا درمانگر) شخصاً به مشاهده و ثبت رفتار می‌پردازد و هم می‌تواند از مشاهده‌گران تربیت‌شده بدین منظور استفاده نماید. حتی در بعضی مواقع می‌توان از شخص مورد سنجش (دانش‌آموز، دانشجو، یا مددجو) خواست که خودش به مشاهده و ثبت رفتارهایش اقدام کند. در این صورت به روش مورد استفاده **مشاهدهٔ شخصی یا خودمشاهده‌گری**[2] گفته می‌شود.

در پاره‌ای مواقع، مشاهدهٔ رفتار به کمک وسایل و ابزارهایی، مانند ضبط صوت یا ضبط تصویر، امکان‌پذیر است. در این صورت مشاهده‌گر یا سنجشگر نیازی به حضور فیزیکی در

1. behavioral observation 2. self-observation

موقعیتی که رفتار موردنظر اتفاق می‌افتد ندارد، و می‌تواند سر فرصت به مشاهدهٔ فیلم تهیه شده از رفتار بپردازد و به تحلیل آن اقدام کند.

صرف‌نظر از اینکه رفتار موردنظر به وسیلهٔ یک شخص مشاهده و ثبت می‌شود یا از طریق یک دستگاه، عواملی که در مشاهده و تحلیل رفتار باید مورد توجه قرار گیرند عبارت‌اند از حضور یا عدم حضور رفتار، فزونی، نقصان، و سایر ویژگیها یا بُعدهای آن رفتار و نیز عوامل به‌وجودآورنده و پیامدهای رفتار موردنظر (نگاه کنید به شکل ۱-۱۵ صفحه‌های پیشین این فصل). در سنجش رفتاری، علاوه بر روشهای مشاهدهٔ رفتاری فوق، از روشها و فنون دیگری استفاده می‌شود. ما در دنبالهٔ مطالب این فصل به معرفی دیگر روشهای شناخته شدهٔ سنجش رفتاری می‌پردازیم.

سنجش قیاس‌پذیر

در روش سنجش قیاس‌پذیر[1] که به آن **مشاهدهٔ رفتاری قیاس‌پذیر**[2] نیز می‌گویند، شخص موردنظر در محیطی که به منظور افزایش احتمال بروز رفتار آماج یا رفتار مورد سنجش طراحی شده است قرار داده می‌شود و رفتارش در آن محیط مشاهده و سنجش می‌شود. کوهن و سوردلیک (۲۰۰۵) سنجش قیاس‌پذیر را با **مطالعهٔ (پژوهش) قیاس‌پذیر**[3] شبیه دانسته‌اند.

> مطالعهٔ قیاس‌پذیر نوعی مطالعه (پژوهش) است که در آن یک یا چند متغیر شبیه به یا قابل قیاس با متغیرهای واقعی موردنظر پژوهشگر هستند... از جمله، مطالعهٔ قیاس‌پذیر برای توصیف پژوهشهایی به کار رفته است که با دانشجویان تمام‌وقت انجام می‌گیرند و هدف آنها در واقع کسب اطلاع دربارهٔ افرادی است که در موقعیتهای شغلی به صورت تمام‌وقت کار می‌کنند. همچنین از مطالعهٔ قیاس‌پذیر برای توصیف پژوهشهایی که دربارهٔ پرخاشگری انجام می‌گیرند استفاده می‌شود و در آنها پرخاشگری ایجاد شده در آزمایشگاه براثر وارد آوردن شوک برقی به آزمودنیها تعریف می‌شود؛ در حالی که هدف پژوهشگر به دست آوردن اطلاعاتی دربارهٔ پرخاشگری در زندگی واقعی خارج از آزمایشگاه است. (ص ۴۱۰)

در سنجش قیاس‌پذیر، یا مشاهدهٔ رفتاری قیاس‌پذیر، مانند مطالعهٔ قیاس‌پذیر، فرد در موقعیتی شبیه به موقعیت زندگی واقعی مورد مشاهده و سنجش قرار می‌گیرد، زیرا مشاهده و سنجش رفتار در شرایط واقعی یا دشوار و یا نامیسر است. برای نمونه، در ارتباط با شخصی که به سبب ترس از مار قدم به دشت و صحرا نمی‌گذارد، به منظور مشاهدهٔ رفتار ترس‌آمیز او، به جای فرستادن او به دشت و صحرا که کاری آسان نیست، سنجشگر یا درمانگر با نشان دادن عکس یا فیلم مار یا یک مار در قفس به مشاهده و سنجش (و احتمالاً درمان) او می‌پردازد.

---

1. analogue assessment 2. analogue behavioral observation 3. analogue study

سنجش موقعیتی

چنان که گفتیم، مشاهدهٔ رفتار در شرایط طبیعی، یعنی همان مشاهدهٔ رفتاری، بهترین روش سنجش رفتاری است، اما بعضی وقتها این کار عملاً غیرممکن است، زیرا سنجشگر از زمان و چگونگی وقوع رفتارها در موقعیتهای مختلف اطلاع کافی نمی‌یابد. هوگان (۲۰۰۳) در این باره مثالهای زیر را ذکر کرده است:

> شخصی [دانش‌آموزی] ممکن است هر دو هفته یک بار دچار حمله‌های هراس[1] یا وحشتزدگی بشود. این اتفاق می‌تواند زندگی او را نابود کند، اما ما نمی‌دانیم که حملهٔ بعدی دقیقاً چه زمانی اتفاق خواهد افتاد. یک زن و مرد ممکن است ماهی یک بار دعوای خانوادگی داشته باشند، اما ما نمی‌توانیم پیش‌بینی کنیم که چه وقت منتظر دعوای بعدی بین آنها باشیم. (ص ۶۰۱)

برای موقعیتهایی مانند آنچه در بالا به آنها اشاره شد، روش سنجش براساس مشاهدات طبیعی امکان‌پذیر نیست، و باید روشهای سنجش رفتاری دیگری را مورد استفاده قرار داد. یکی از روشهای سنجش رفتاری که برای این منظور به کار رفته **سنجش قیاس‌پذیر** است که در بالا آن را توضیح دادیم. روش **سنجش موقعیتی**[2] یا **آزمون موقعیتی**[3] (موضوع بحث این قسمت) و روش نقش‌گزاری (موضوع بحث بعدی) نمونه‌های دیگری از سنجش قیاس‌پذیرند.

در سنجش موقعیتی یا آزمون موقعیتی آزمون‌شونده در موقعیتی شبیه به یک موقعیت واقعی که ما می‌خواهیم رفتارش را در آن موقعیت پیش‌بینی کنیم قرار داده می‌شود و در آن موقعیت رفتار او مشاهده و ثبت می‌گردد. به عنوان نمونه، اگر بخواهیم رفتار یک فرد را در یک جمع بدونِ مدیر یا رهبر مشاهده و سنجش کنیم و قصدمان مثلاً این باشد که قدرت رهبری یا مدیریت او را بسنجیم، وی را در جمعِ افرادی که هیچ‌کس ادارهٔ آن را به عهده ندارد قرار می‌دهیم و رفتار او را مشاهده می‌کنیم. این موقعیت به ما اجازه می‌دهد که واکنشهای فرد موردنظر را در گروه و حتی قدرت ادارهٔ گروه را توسط او مشاهده و سنجش نماییم.

ثرندایک و هیگن (۱۹۷۹) سنجش موقعیتی یا آزمون موقعیتی را سنجش در موقعیتهای شبیه‌سازی شده با موقعیت زندگی واقعی دانسته و گفته‌اند معمولاً عملکرد فرد در این گونه موقعیتها، در ضمن وقوع، مشاهده و درجه‌بندی (سنجش) می‌شود. آنان روش آزمون موقعیتی را ترکیبی از مشاهده و آزمون به حساب آورده و در این باره مثال زیر را ارائه داده‌اند.

آزمون رودخانه[4]

آزمون‌شوندگان به گروههای ۵ یا ۶ نفری تقسیم می‌شوند و هر گروه به کنار نهری که حدود سه

---

1. panic attacks 2. situational assessment 3. situational test

۴. این آزمون نمونه‌ای از آزمونهای موقعیتی است که طی جنگ جهانی دوم به وسیلهٔ ادارهٔ خدمات استراتژیک ارتش آمریکا برای انتخاب افراد به منظور خدمت در مأموریتهای ویژه مورد استفاده قرار گرفته‌اند.

متر عرض دارد برده می‌شود. در کنار نهر، یک قطعه چوب، چند تکه سنگ، تعدادی تخته چوب به اندازه‌های مختلف، مقداری طناب، یک قرقرهٔ بزرگ، و بعضی اشیاء دیگر وجود دارند. بعد از آنکه افراد به کنار نهر برده می‌شوند و چیزهای بالا در اختیار آنان گذاشته می‌شود، دستورات زیر را دریافت می‌کنند:

برای حل مشکلی که در پیش روی شماست باید از قدرت تخیل خود استفاده کنید. در مقابل شما رودخانهٔ عمیقی با جریانی تند قرار دارد به گونه‌ای که هیچ چیز را نمی‌توان بر روی آن قرار داد. دیوارهای نهر نیز شیب عمودی دارند. پس فقط می‌توان در کنار نهر ایستاد و کار کرد....

مأموریت شما در این منطقه آن است که مسافت‌یاب ظریفی را که با مهارت کامل به شکل کنده چوبی استاد شده به آن سوی رودخانه برسانید.... برای انجام این کار می‌توانید از هر شیئی که در این حوالی یافت می‌شود استفاده کنید....

رفتار اعضای گروه هنگام درگیر شدن با مسئله و کوشش برای حل کردن آن مورد مشاهده قرار می‌گیرد و "در آنان ویژگیهایی چون مشارکت، قدرت رهبری، ابتکار، و نیروی بدنی، ارزشیابی می‌شود"، (کوهن و سوردلیک، ۲۰۰۲، ص ۳۹۷).

در آزمون موقعیتی بالا، آنچه مورد توجه ارزشیابان قرار دارد چیزی نیست که به آنان گفته شده، بلکه هدف از آزمون سنجش تواناییهای ذهنی، جسمی، و ارتباطی داوطلبان است. بنابراین، "در اکثریت این‌گونه آزمونها شخص مورد مشاهده ممکن است نفهمد که برای چه منظوری مورد مشاهده قرار می‌گیرد (یا اصلاً مورد مشاهده قرار می‌گیرد)" (ثرندایک و هیگن، ۱۹۷۹، ص ۲۳).

آزمونهای موقعیتی نخستین‌بار به وسیلهٔ روانشناسان آلمانی و انگلیسی در جنگ جهانی دوم مورد استفاده قرار گرفت. اما این روش به طور گسترده، توسط ادارهٔ خدمات استراتژیک ارتش آمریکا[1] به کار بسته شد (کوهن و سوردلیک، ۲۰۰۲). هرچند که این روش در آغاز برای مقاصد نظامی ابداع و به کار گرفته شد، با این حال، اکنون سنجش موقعیتی عمدتاً برای استخدام نیروی انسانی موردنیاز در صنعت و تجارت مورد استفاده قرار می‌گیرد. بنا به گفتهٔ ثرندایک و هیگن (۱۹۷۹) "با این روش بیش از ۱۰۰,۰۰۰ نفر در سیستم بل[2] و قدری کمتر از این تعداد نیز در سایر شرکتها سنجش شده‌اند" (ص ۵۰۲).

هرچند که سنجش موقعیتی عمدتاً در خارج از آموزش و پرورش و برای تعیین تواناییهای گوناگون بزرگسالان مورد استفاده قرار گرفته، اما از این روش در آموزش و پرورش و برای سنجش ویژگیهای کودکان و نوجوانان نیز استفاده شده است. به عنوان نمونه، به آزمون زیر که برای سنجش یکی از ویژگیهای اخلاقی دانش‌آموزان، یعنی درستکاری، به کار رفته است توجه کنید.

---

1. Office of Strategic Services 2. Bell system

آزمون درست‌کاری

از دانش‌آموزان خواسته شد تا تعدادی وزنه را که تفاوتهای بسیار اندکی با هم داشتند از سنگین‌ترین به سبک‌ترین آنها ردیف کنند. تنها راه انجام دادن این کار به طور صحیح این بود که به ته وزنه‌ها که مقدار دقیق آن نوشته شده بود نگاه شود. پس از آنکه دانش‌آموزان کار خود را انجام دادند و نتیجه را بر روی کاغذ ثبت نمودند، از آنها نسخه‌برداری شد و اصل برگه‌ها به دانش‌آموزان بازگردانده شد تا خودشان کارشان را نمره‌گزاری کنند. پس از مقایسهٔ نمرات واقعی هر دانش‌آموز با نمره‌ای که به خودش داده بود میزان درست‌کاری در او سنجش می‌شد.

کوهن و سوردلیک، (۲۰۰۵)، به جای آزمون موقعیتی اصطلاح اندازه گیری عملکرد موقعیتی[1] را به کار برده‌اند. مثالی که آنان برای اندازه‌گیری عملکرد موقعیتی به دست داده‌اند به قرار زیر است: داوطلبان پروژه‌های فضانوردی در موقعیتهای شبیه‌سازی شده با شرایط فضایی قرار داده می‌شوند و رفتار آنان در رابطه با متغیرهایی چون تحمل بی‌وزنی، تحریک‌پذیری؛ توانایی خوردن، خوابیدن، و انجام کارهای روزمره؛ و توانایی کنار آمدن با دیگران تحت چنان شرایطی مورد مشاهده و ارزیابی قرار می‌گیرد.

نقش‌گزاری

در روش سنجش بر اساس **نقش‌گزاری** یا **ایفای نقش**[2] از شخص مورد سنجش یا آزمون شونده خواسته می‌شود تا در موقعیتی به ایفای یک نقش بپردازد که با موقعیت واقعی زندگی که رفتار شخص در آن قرار است پیش‌بینی شود شبیه است. به عنوان نمونه، به منظور سنجش اعتماد به نفس آزمون شونده از او خواسته می‌شود تا نقش متقاضی کار را ایفا کند که برای مصاحبه به پیش کارفرما رفته و سنجشگر نقش کارفرما را بازی می‌کند. یا آزمون‌شونده نقش کارمندی را ایفا می‌کند که به قضاوت ناحق رئیس اعتراض می‌نماید و سنجشگر نقش رئیس را ایفا می‌کند. در روش نقش‌گزاری، هنگامی که آزمون‌شونده در موقعیت مورد نظر ایفای نقش می‌کند، سنجشگر به مشاهده و ارزشیابی رفتار او می‌پردازد. "دقت کنید که در این روش ما به دنبال ویژگیهای درونی یا ناهشیار فرد نیستیم بلکه مستقیماً رفتار او را مشاهده می‌کنیم تا در صورت لزوم آن را تغییر دهیم و اصلاح نماییم" (هوگان، ۲۰۰۳، ص ۶۰۲).

هینتزه، استونر، و بول (۲۰۰۰) مثال زیر را برای سنجش به کمک ایفای نقش به دست داده‌اند. در این مثال رابطهٔ بین والدین و کودک مشاهده و سنجش می‌شود.

بازی والدین و کودک

در این نقش‌گزاری یک کودک و پدر یا مادر او در یک اتاق بازی مشاهده می‌شوند. در این اتاق اسباب‌بازیهای مختلف ویژهٔ سن کودک، مانند بلوکهای ساختمانی، کامیونهای اسباب‌بازی، انواع اتومبیلهای

1. situational performance measure 2. role playing

بازی، عروسکهای مختلف، پازلها، انواع مداد رنگی و کاغذ موجودند. در **بازی کودک**، از پدر یا مادر خواسته می‌شود تا در هر فعالیتی که کودک انتخاب می‌کند شرکت کند و به کودک اجازه دهد تا چگونگی و قواعد بازی را تعیین نماید. بنابراین، بازی کودک یک موقعیت **بازی آزاد** را به وجود می‌آورد. در **بازی پدر یا مادر**، از پدر یا مادر خواسته می‌شود تا کودک را به فعالیتهایی که چگونگی و قواعد آن را پدر یا مادر تعیین می‌کند شرکت جوید. در این وضعیت، بازی پدر یا مادر یک موقعیت **دستوری** را به وجود می‌آورد. هر یک از دو حالت به مدت پنج دقیقه مشاهده می‌شود و تعامل یا کنش متقابل بین دو نفر به طور منظم ثبت می‌شود. رفتارهای مورد نظر پدر یا مادر عبارت‌اند از: بیانات تأییدکننده و توجه بدنی، توجه اجتماعی، سؤال کردن، دستور دادن، افطار کردن، و قهر کردن. رفتارهای مورد نظر در کودک به قرار زیرند: رفتارهای اطاعت، رفتارهای عدم اطاعت و نامناسب، مانند گریستن، داد زدن، قشقرق راه انداختن، پرخاشگری کردن، و حرفهای زشت.

(هینتزه، استونر، و بول، ۲۰۰۰، ص ۱۱۷)

سنجش نهان‌یابی

یکی دیگر از روشهای سنجش رفتاری که آن نیز بر مشاهده استوار است، **سنجش نهان‌یابی (نامرئی)**[1] نام گرفته. در این روش آزمون‌شوندگان یا افراد مورد سنجش نمی‌دانند که سنجشگر به جمع‌آوری اطلاعات دربارهٔ رفتار آنان می‌پردازد. "فرض زیربنای این روش آن است که اگر مردم ندانند که به وسیلهٔ دیگران مورد مشاهده قرار دارند رفتارشان طبیعی‌تر و معمولی‌تر است" (آندرسون و بورک، ۲۰۰۰، ص ۵۵). به مثالهای زیر از آندرسون و بورک توجه کنید.

هدف یک پژوهش تعیین علاقهٔ دیدارکنندگان از قسمتهای مختلف موزهٔ علم و صنعت شهر شیکاگو بود. بدین منظور، پژوهشگر به مسؤلان ادارهٔ موزه مراجعه نمود و از آنان پرسید که هرچند وقت یکبار کاشی‌های مقابل هریک از قسمتهای مختلف موزه را تعویض می‌کنند. فراوانی دفعات تعویض کاشیها به عنوان شاخصی از علاقه‌مندیِ بازدیدکنندگان به قسمتهای مختلف موزه تعیین شد. یعنی هرچه کاشیهای یک قسمت موزه زودتر خراب می‌شد و در نتیجه زودتر تعویض می‌شد، آن قسمت دیدارکنندگان بیشتری را به خود جلب می‌کرد و لذا بیشتر مورد علاقهٔ آنان بود.

در پژوهشی دیگر، هدف تعیین محبوبیت ایستگاه‌های رادیویی محلی بود. بدین منظور، پژوهشگر در یکی از پمپ‌بنزین‌های آن محل برای مدتی نقش فروشندهٔ بنزین را ایفا کرد و ضمن پر کردن باک بنزین مشتریان، پاک کردن شیشهٔ اتومبیل آنان، و انجام سایر خدمات معمول در پمپ بنزینها[2] به رادیوهای آنان گوش می‌داد و تعداد ایستگاههای رادیویی مختلف را که صدایشان از اتومبیلها پخش می‌شد به عنوان شاخص محبوبیت ایستگاهها تعیین می‌کرد. یعنی

---

1. unobtrusive assessment

۲. در پمپ بنزینهای آمریکا، مسئول فروش بنزین، علاوه بر اینکه شخصاً اتومبیل را بنزین‌گیری می‌کند، معمولاً هم شیشهٔ اتومبیل را پاک می‌کند و هم روغن و آب موتور را وارسی می‌نماید.

هرچه تعداد رادیوهای اتومبیلها یک ایستگاه رادیویی را بیشتر پخش می‌کردند آن ایستگاه محبوب‌تر ارزیابی می‌شد.

هر دو نمونهٔ فوق مواردی از مشاهدهٔ نهان‌یابی یا نامرئی هستند، زیرا در هیچ‌کدام از آنها شخص یا اشخاصی که دربارهٔ آنان اطلاعات جمع‌آوری شده از اینکه به وسیلهٔ پژوهشگر یا مشاهده‌گر مورد مشاهده قرار داشتند آگاه نبودند.

نظارت شخصی

روش دیگر سنجش رفتاری این است که از شخص بخواهیم تا رفتار خودش را مورد مشاهده قرار داده آن را ثبت و به سنجشگر گزارش نماید. به این روش **نظارت شخصی**[1] گفته می‌شود. در این روش به شخص گفته می‌شود تا به‌طور منظم رفتار خودش و شرایط درونی و بیرونی را که موجب بروز آن می‌شود مشاهده کرده گزارش نماید.

ثبت نتایج را می‌توان با هر یک از روشهایی که در بخش سنجش رویدادهای گذرای رفتار در صفحه‌های قبلی این فصل توضیح دادیم انجام داد. مثلاً می‌توان از روش ثبت فراوانی رفتار به این صورت استفاده کرد که از آزمون‌شونده خواسته شود تا تعداد سیگارهایی را که روزانه دود می‌کند ثبت نماید. همچنین از او می‌توان خواست تا شرایط یا عواملی را که میل به سیگار کشیدن را در او ایجاد می‌کنند نیز گزارش کند. همچنین از روش ثبت نمونه گیری زمان می‌توان این‌گونه استفاده کرد که از دانش‌آموز یا دانشجو خواسته شود که سر هر ساعت در طول روز کاری را که انجام می‌دهد ثبت نماید. با این روش می‌توان مقدار زمانی را که او در طول روز صرف رفتار مورد نظر مثلاً مطالعه می‌کند تعیین نمود.

هر چندکه روش نظارت شخصی یکی از روشهای سنجش رفتاری به حساب آمده (کوهن و سوردلیک، ۲۰۰۵)، با این حال این روش از دقّت روشهای دیگر رفتاری برخوردار نیست. به همین سبب، گرچه ما عمدتاً روشهای سنجش رفتاری را در این فصل در ردیف روشهای سنجش مستقیم به حساب آورده‌ایم، اما روش نظارت شخصی را در جای دیگر (سیف، ۱۳۷۷) جزو روشهای غیرمستقیم دانسته‌ایم. علت قرار دادن این روش در ردیف روشهای سنجش غیرمستقیم آن است که در آن خودِ سنجشگر یا یک مشاهده‌گر آموزش دیده به مشاهده و ثبت رفتار آزمون‌شونده نمی‌پردازد و در نتیجه نمی‌توان به نتایج آن اعتماد کامل داشت.

کوهن و سوردلیک (۲۰۰۵) مشکل عمدهٔ روش نظارت شخصی را **واکنش‌پذیری**[2] آن دانسته‌اند. واکنش‌پذیری را می‌توان به صورت "تأثیر بالقوهٔ حضور مشاهده‌گر بر رفتار مورد مشاهده" (سیف، ۱۳۷۷، ص ۱۰۲) تعریف کرد. کوهن و سوردلیک در این باره گفته‌اند، "واکنش‌پذیری به امکان ایجاد تغییر در رفتار، تفکر، یا عملکرد مشاهده شونده، در نتیجهٔ اینکه

---

1. self monitoring 2. reactivity

او مورد مشاهده، سنجش، یا ارزشیابی قرار گرفته است گفته می‌شود" (ص ۴۱۰). بنابراین، داده‌های حاصل از روش نظارت شخصی و گزارش شخصی را باید با احتیاط مورد استفاده قرار داد و در صورت امکان روشهای دیگر سنجش را با آن همراه کرد.

سنجش شناختی - رفتاری

چنان‌که دیدیم، در روشهای سنجش رفتاری صرفاً رفتارهای آشکار فرد مورد مشاهده و سنجش قرار می‌گیرند. در روش **سنجش شناختی - رفتاری**[1] افکار و اندیشه‌های فرد نیز رفتار تلقی می‌شوند و هم فرایند و هم آثار و تظاهرات آشکار آنها مورد مشاهده قرار می‌گیرند و سنجش می‌شوند.

در سنجش اندیشه‌ها با روش شناختی - رفتاری، مانند سنجش رفتار آشکار، باید کاملاً دقیق و عینی عمل کرد. در این روش، موضوع مورد سنجش یا اندیشه‌های فرد باید به روشنی توصیف و به صورت مشاهده‌پذیر تعریف شوند.

یکی از روشهایی که به این منظور ابداع شده روش **بلند حرف زدن**[2] است. در این روش، فرد خود را در یک موقعیت مثلاً هراسناک تصور می‌کند و اندیشه‌هایش را به زبان می‌آورد. در واقع، از او خواسته می‌شود تا افکارش را به زبان آورد، لذا به آن روش **بلند فکر کردن**[3] نیز گفته شده.

در مقایسه با استفاده کنندگانِ روشهای سنتی سنجش (عموماً روشهای عینی، مانند آزمونهای صحیح - غلط، جور کردنی، و چند گزینه‌ای) که غالباً فراورده‌های تفکر و اندیشهٔ دانش‌آموزان و دانشجویان را می‌سنجند، چنان‌که پیشتر نیز گفتیم، طرفداران روشهای جدید سنجش از جمله سنجش شناختی - رفتاری بیشتر علاقه‌مند به جریان یا فرایند تفکر هستند. همه می‌دانیم که دانش‌آموزان و دانشجویانی که در یک آزمون عینی مربوط به حل مسائل (مثلاً آزمون مسائل ریاضی) نمره‌های مشابهی می‌گیرند ممکن است از راههای مختلفی به نتیجهٔ واحدی رسیده باشند که این امر در این نوع آزمونها نشان داده نمی‌شود. روش بلند فکر کردن به ما کمک می‌کند تا به چگونگی اندیشیدن دانش‌آموزان و رسیدن به راه‌حل مسئله در آنان پی‌ببریم و آن را سنجش کنیم.

برای کسب اطلاع دربارهٔ افکار دانش‌آموزان در ضمن اینکه آنها مشغول انجام تکالیف تحصیلی هستند، سه فن بلند فکر کردن ابداع شده است. یکی از آنها **بلند فکر کردن همزمان**[4] است. در این فن سنجش، معلم یا سنجشگر از آزمون‌شونده می‌خواهد تا مرتباً تمام افکاری را که به هنگام انجام یک فعالیت، مثل حل کردن یک مسئله، به ذهنش می‌رسد بلند بلند به زبان آورد. "اگر دانش‌آموزان قادر به انجام چنین کاری باشند، فن بلند فکر کردن همزمان کامل‌ترین نمونه را از فرایندهای تفکر آنان به دست می‌دهد" (گتینگر و سیبرت، ۲۰۰۰، ص ۱۵۴). این

1. cognitive-behavioral assessment 2. talk-aloud 3. think-aloud 4. concurrent thinking aloud

روش برای دانش‌آموزان کمتر از ۸ سال مناسب است، زیرا بسیاری از دانش‌آموزان بالای ۸ سال یاد گرفته‌اند که به زبان آوردن افکارشان را بازداری نمایند یا اینکه حرف زدن به همراه انجام کار برای آنان چنان عادی شده که ممکن است افکار زیربنای فعالیت جاری خود را نتوانند گزارش کنند.

فن دیگر بلند فکر کردن که مناسب دانش‌آموزان بزرگ‌تر از ۸ سال است **شیوهٔ مصاحبهٔ همزمان**[1] نام دارد. در این فنِ سنجش، معلم یا سنجشگر متناوباً آزمون‌شونده را از کار باز می‌دارد و سؤالهای خاصی از او می‌پرسد تا به راهبردها (استراتژیها) و افکار او در آن لحظات پی ببرد. "هر چند که این روش محدودیتهایی دارد (از جمله اینکه ممکن است برای فعالیت فکری آزمون‌شونده مزاحمت ایجاد کند)، اما برای بسیاری از یادگیرندگان یک فن سازنده است" (کتینگر و سیبرت، ۲۰۰۰، ص ۱۵۵).

فن سوم سنجش مبتنی بر بلند فکر کردن **شیوهٔ بازسازی**[2] نامیده شده است. در این فن سنجش، فعالیت درک مطلب (خواندن و فهمیدن متن) یا حل مسئلهٔ دانش‌آموز ضبط ویدیوئی می‌شود، و بعد از او خواسته می‌شود تا نوار فعالیت خود را مشاهده و در مراحل مختلف تفکرات خود را بازگو کند. "در هر سه فن پیش گفته، نشان داده شده که سنجش بلند فکر کردن دربارهٔ راهبردهای شناختی و فراشناختی[3] یادگیرندگان [دانش‌آموزان] اطلاعات با ارزشی در اختیار می‌گذارند" (کتینگر و سیبرت، ۲۰۰۰، ص ۱۵۵).

سنجش روانی ـ فیزیولوژیکی

یکی دیگر از روشهایی که روانشناسان رفتاری فراوان مورد استفاده قرار می‌دهند **سنجش روانی ـ فیزیولوژیکی**[4] است. هوگان (۲۰۰۳) در این باره به گونهٔ زیر نظر داده است:

> اندازه‌های فیزیولوژیکی یکی از بهترین مثالهای مربوط به سنجش رفتاری هستند. رفتارگرایان با خوشحالی از آنها استقبال می‌کنند. بنا به باور رفتارگرایان اصول‌گرا، ترس تنها مجموعه‌ای از واکنشهای فیزیولوژیکی است. همین گونه است اضطراب و عشق! بنابراین، اگر می‌خواهید اضطراب یا ترس را اندازه بگیرید، واکنشهای فیزیولوژیکی را اندازه‌گیری کنید. به عنوان مثال، ترس از صحبت کردن در حضور جمع را در نظر بگیرید. اندازه‌های فیزیولوژیکی این ترس عبارت‌اند از افزایش ضربان قلب، افزایش پاسخ برقی پوست، تنگ شدن مردمک چشم، و مانند اینها.

یکی از موارد جالب سنجش روانی ـ فیزیولوژیکی **بازخورد زیستی**[5] است. اضطراب، ترس، وتنش اغلب همبسته‌های فیزیولوژیکی دارند. وقتی که شخصی مضطرب می‌شود، ممکن

---

1. concurrent interview procedure 2. reconstruction procedure

۳. congnitive and metacognitive strategies (برای کسب اطلاعات گسترده دربارهٔ این راهبردها و چگونگی آموزش دادن آنها به دانش‌آموزان اثر دیگر مؤلف [روانشناسی پرورشی: روانشناسی یادگیری و آموزش، ویرایش ششم، فصل ۱۱، ۱۳۸۶] را ببینید.)

4. psychophysiological assessment 5. biofeedback

است اضطرابش را در ماهیچه‌ها یا نحوهٔ تنفس یا تعریق احساس کند. روش بازخورد زیستی به فرد امکان می‌دهد تا متوجه تغییرات روانی - فیزیولوژیکی خود و نیز حالات روانی که موجب آنها می‌شوند گردد. اُلسون[1] و هرگنهان[2] (۲۰۰۹، ترجمهٔ سیف، ۱۳۸۸) در این باره گفته‌اند "پژوهشگران گزارش کرده‌اند که می‌توان به بیماران قلبی کنترل نابهنجاریهای عروقی، به بیماران صرعی واپس زدن فعالیتهای نابهنجار مغزی، و به افراد مبتلا به سردردهای میگرنی کنترل انبساط رگهای خونی اطراف مغز را که منجر به سردرد میگرنی می‌شوند آموزش داد" (ص ۱۷۹).

برای کمک به شخص در انجام کارهای فوق از راه سنجش روانی - فیزیولوژیکی، با استفاده از یک وسیله، تغییرات ایجاد شده در رویدادهای درونی فرد را که سعی در کنترل آنها دارد، مثلاً فشار خون یا بی‌نظمی‌های قلب، به او نشان می‌دهند. به این نشان دادنِ تغییراتِ ایجاد شده بازخورد زیستی می‌گویند زیرا به فرد دربارهٔ رویدادهای درونی زیستی اطلاعاتی می‌دهد. معمولاً در این روش (که نوعی یادگیری است) از تقویت‌کننده‌های بیرونی استفاده نمی‌شود، بلکه اطلاعات حاصل از تدابیر بازخورد زیستی تنها چیزی است که برای ایجاد این نوع یادگیری لازم است. در واقع، اطلاعات به دست آمده نقش تقویت‌کننده را ایفا می‌کنند. معمولاً پس از مشاهدهٔ نتایج سنجش حاصل از بازخورد زیستی، آزمودنیها (بیماران) از حالت درونی خود آگاه می‌شوند و می‌توانند متناسب با آن پاسخ دهند، یعنی مثلاً فشار خون خود را بالا یا پایین ببرند، و این کار را بدون کمک بازخورد زیستی ادامه دهند.

مصاحبهٔ رفتاری

از روش مصاحبه، به طور عمده، برای سنجش ویژگیهای عاطفی استفاده می‌شود. به همین سبب، ما این روش را، به تفصیل، در فصل بعد (سنجش عاطفی) توضیح خواهیم داد. اما در اینجا نیز، به عنوان یکی از فنون سنجش رفتاری به آن اشاره‌ای می‌کنیم. در مصاحبهٔ رفتاری، مصاحبه کننده تمام توجهش را به رفتارهای خاصی از مصاحبه‌شونده معطوف می‌کند. در این روش، هم شناسایی و سنجش ویژگیهای رفتارهای خاص مورد نظر است، و هم کوشش به عمل می‌آید تا محرکهای پیش آیندی و محرکهای پس آیندی رفتارها (طبق شکل ۱-۱۵ صفحات قبلی این فصل) شناسایی و سنجش شوند. هوگان (۲۰۰۳) در این رابطه مثال زیر را برای مصاحبهٔ رفتاری ذکر کرده است:

مورد شخصی را در نظر بگیرید که نمی‌تواند عصبانیتش را کنترل کند. مصاحبه‌کننده موارد خاص رفتار عصبانیت [و محرکهای پیش‌آیندی و پس‌آیندی آن] را پرس‌وجو می‌کند. خوب، پس به سر هم‌اتاقی‌ات داد زدی؟ چه وقت این اتفاق افتاد؟ کس دیگری هم حضور داشت؟ قبل از اینکه عصبانی شوی چه اتفاقی افتاد؟ بعدش هم‌اتاقی‌ات چه کار کرد؟ عصبانیت شما چه مدت طول کشید؟

(هوگان، ۲۰۰۳، ص ۶۰۲)

1. Olson 2. Hergenhahn

امتیازها و محدودیتهای روشهای سنجش رفتاری

سنجش رفتاری یکی از رویکردهای با سابقه در روانشناسی است که به تازگی در آموزش و پرورش نیز شهرت یافته است. عمده‌ترین امتیاز این روش آن است که به وسیلهٔ آن می‌توان رفتارها و عملکردهای تحصیلی و غیرتحصیلی را در شرایط کلاس درس، آزمایشگاه، زمین بازی، و سایر موقعیتهای طبیعی به‌طور مستقیم و بی‌واسطه مشاهده و ثبت و ضبط کرد. به طور کلی امتیازهای روشهای سنجش رفتاری را می‌توان به صورت زیر بیان کرد:

۱. با روش سنجش رفتاری هم می‌توان ویژگیهای رفتار مورد نظر را شناسایی و سنجش کرد و هم عوامل ایجادکننده و نگهدارندهٔ آن را.

۲. به کمک سنجش رفتاری نه تنها می‌توان اطلاعات مورد نیاز اندازه‌گیری و ارزشیابی موضوعهای مورد نظر را به دست آورد، بلکه با استفاده از روشهای سنجش رفتاری می‌توان داده‌های مورد نیاز پژوهشهای آموزشی را نیز گردآوری کرد و تأثیرهای مختلف متغیرهای آموزشی مورد نظر را آزمون نمود. دالی و مورداک (۲۰۰۰) به نقل از نلسون[1] و هیز[2] گفته‌اند "تازگی حوزهٔ سنجش رفتاری نه تنها به هدفها یا راهبردهای آن وابسته است بلکه به کوشش عمدی آن برای بهبود شناسایی و اندازه‌گیری متغیرهای وابسته، به منظور افزایش احتمال انتخاب متغیرهای مستقل (مثلاً فنون آموزشی) و پالایش ارزشیابی آن متغیرها، مربوط می‌شود" (ص ۴۷).

۳. از آنجا که در سنجش رفتاری، خودِ رفتار مستقیماً مشاهده و سنجش می‌شود، مشکل روایی چندان مطرح نیست، زیرا روایی منعکس‌کنندهٔ نگرانی دربارهٔ صحتِ عملِ ابزارهای سنجش است، و به همین سبب یکی از مشکلات درگیر با سنجش غیرمستقیم تعیین روایی آن است؛ یعنی اثبات اینکه چیزی را که وسیلهٔ سنجش یا آزمون قرار است بسنجد می‌سنجد.

۴. از روشهای سنجش مشاهده‌ای رفتاری می‌توان به جای روشهای دیگر سنجش، از جمله پرسشنامه، برای سنجش ویژگیهای دانش‌آموزان سود برد. آندرسون و بورک (۲۰۰۰) برای اثبات این مطلب پژوهشی را از پورکی[3]، کیج[4]، و گریوز[5] (۱۹۷۳) نقل کرده‌اند. پژوهشگران، با مراجعه به پژوهشهای انجام شده دربارهٔ عزت نفس دانش‌آموزان رفتارهای مختلفی را که معرف عزت نفس مثبت و عزت نفس واقعی بودند شناسایی کردند. بعد از گروهی معلم ابتدایی که به تصادف برگزیده شدند خواستند تا در مورد آن رفتارها نظر بدهند. از میان آنها ۱۸ رفتار کلاسی دانش‌آموزان که مُعرف عزت نفس شناخته شد انتخاب گردیدند. سپس از سایر معلمان دورهٔ ابتدایی خواسته شد، از راه مشاهده، دانش‌آموزان خود

1. Nelson 2. Hayes 3. Purkey 4. Cage 5. Graves

را با توجه به فراوانی ۱۸ رفتار انتخابی شناسایی نمایند. یافتهٔ مهم این پژوهش که به بحث فعلی ما مربوط است این است که بین نتایج سنجش رفتاری معلمان این پژوهش و پرسشنامهٔ عزت نفس کوپر اسمیت[1] همبستگی بالایی یافت شد. یعنی اینکه، در این پژوهش، مشاهدهٔ مستقیم و سنجش رفتاری توانست نتایجی با اعتماد نتایج یک پرسشنامهٔ معتبر استاندارد شده به دست دهد.

۵. از روشهای سنجش رفتاری، علاوه بر شناسایی رفتارهای آشکار و ویژگیهای عاطفی ـ شخصیتی، می‌توان برای سنجش ویژگیهای شناختی نیز استفاده کرد. هوگان (۲۰۰۳) گفته است می‌خواهید بدانید که پیشرفت **جانی** در سال آینده در درس ریاضیات چگونه خواهد بود؟ از او یک امتحان ریاضی به عمل بیاورید، آزمون توانایی ذهنی لازم نیست. اگر این دانش‌آموز در امتحان شما موفقیتی کسب نکرد، فرایندهای شناختی او را تجزیه و تحلیل نکنید. ریاضیات به او یاد بدهید" (ص ۶۰۰).

روشهای سنجش رفتاری، مانند سایر روشهای مورد بحث در این کتاب، دارای محدودیتهایی نیز هستند که در ضمن استفاده از این روشها باید به آنها توجه نمود. در زیر به موردهای مهم این محدودیتها اشاره می‌کنیم.

۱. اگر بخواهیم رفتارهای نوعی یا معمول افراد را که معرف ویژگیهای نسبتاً ثابت در آنهاست ـ مانند نگرش و سایر ویژگیهای عاطفی ـ سنجش کنیم در آن صورت باید فرد را در موقعیتهای گوناگون مورد مشاهده قرار دهیم. این کار مستلزم صرف وقت و نیروی زیادی است. آندرسون و بورک (۲۰۰۰) در این باره گفته‌اند "کوشش برای استنباط ویژگیهای عاطفی از یک مشاهدهٔ واحد از رفتار یا از مشاهدهٔ رفتار در یک موقعیت واحد کار بیهوده‌ای است" (ص ۵۶).

۲. در رابطه با امتیاز نبود مشکل عدم روایی در سنجش رفتاری، گاهی وقتها ممکن است سنجشگر رفتاری نیز نگران کمبود روایی روش سنجش خود باشد. زمانی این نگرانی به وجود می‌آید که پدیدهٔ واکنش‌پذیری یا واکنش‌گری اتفاق بیفتد. یعنی اینکه حضور مشاهده‌گر ممکن است باعث شود که شخص یا اشخاص مورد مشاهده رفتاری به غیر از رفتار معمول خود انجام دهند. اسکینر[2]، دیتمر[3]، و هاول[4] (۲۰۰۰) در ارتباط با واکنش‌پذیری در آموزشگاه گفته‌اند "واکنش‌پذیری زمانی رخ می‌دهد که فرایند جمع‌آوری داده‌ها محیط کلی کلاس و رفتارهایی را که مورد مشاهده قرار می‌گیرند تغییر دهد" (ص ۴۰). آنان برای کاهش واکنش‌پذیری پیشنهادهای زیر را ارائه داده‌اند:

الف ـ دربارهٔ حضور مشاهده‌گر اطلاعات روشنی به دانش‌آموزان داده نشود.

---

1. Cooper Smith Selt-Esteem Inventory 2. Skinner 3. Dittmer 4. Howell

ب ـ از مشاهده‌گر خواسته شود با دانش‌آموزانی که رفتارشان مورد مشاهده قرار دارد تماس چشمی برقرار نکند.

پ ـ از امکاناتی نظیر آئینه‌های یک‌طرفه و ضبط ویدیویی استفاده شود.

ت ـ از مشاهده‌گر خواسته شود تا در جایی از اتاق که خیلی جلوی چشم نباشد بنشیند.

ث ـ از مشاهده‌گر خواسته شود تا پیش از دانش‌آموزان وارد کلاس شود.

۳. در رابطه با فنون خاص سنجش رفتاری، مانند سنجش روانی ـ فیزیولوژیکی، نظارت شخصی و گزارش شخصی نیز محدودیتهایی ذکر شده است. مشکل عمدهٔ روشهای نظارت شخصی و گزارش شخصی امکان نبود صداقت در گزارش افراد است. در رابطه با سنجش فیزیولوژیکی و روشهای دیگر سنجش رفتاری، هوگان (۲۰۰۳) مشکل را به نحو زیر بیان کرده است:

> در اختیار داشتن داده‌های ثبت شدهٔ فیزیولوژیکی از واکنشهای اضطرابی یک شخص در طول روز بسیار خوب است. اما چه کسی دلش می‌خواهد دستگاه مورد نظر را به خودش ببندد و این سو و آن سو برود؟ در استفاده از روش نمونه‌گیری زمان، یک شخص در طول روز چند بار صدای زنگ دستگاه خبردهنده را می‌تواند تحمل کند؟ چه کسی حاضر است در تمام روز در ته کلاس درس بنشیند تا رفتار پرخاشگرانهٔ یک دانش‌آموز را ثبت نماید؟ (۶۰۴)

موارد استفادهٔ روشهای سنجش رفتاری

چنان‌که دیدیم، روشهای سنجش رفتاری بسیار متنوع‌اند و در موقعیتهای مختلف آموزشگاهی، اجتماعی، بیمارستانی، و محیط کار مورد استفاده قرار می‌گیرند. ما در اینجا نمونه‌هایی از آنها را می‌آوریم.

نمونهٔ ساده‌ای از سنجش فراوانی رفتار در یک موقعیت آموزشگاهی

خانم برگمن بدگویی‌ها را شمارش می‌کند

دانش‌آموزان کلاس سوم خانم برگمن بیشتر وقتشان را صرف بدگویی دربارهٔ همدیگر می‌کردند. این کار به دو دلیل باعث نگرانی خانم برگمن (معلم کلاس) شده بود، یکی اینکه آنها به کارهایشان نمی‌رسیدند، و دیگر اینکه با این کارشان خانم معلم را عصبانی می‌کردند. او از این بابت برای کمک به پیش خانم باربی رفت و از او شنید که اولین اقدام برای حل مشکل این است که تعداد بدگویی‌ها را شمارش کند.

خانم برگمن تصمیم گرفت هر بار که دانش‌آموزی نام دانش‌آموز دیگری را به زبان می‌آورد و از او بدگویی می‌کند آن را به عنوان یک رفتار ناشایست شمارش کند. نمونه رفتارهای

ناشایستی که او تصمیم به شمارششان گرفت عبارت بودند از "جانی تکالیفش را انجام نمی‌دهد و مرا اذیت می‌کند،" "هارولد و پیتر با هم حرف می‌زنند."

خانم برگمن برای انجام این کار دوباره به سراغ خانم باربی رفت تا از او کمک بخواهد. "چگونه همهٔ رفتارهای ناشایست را ثبت و ضبط کنم، در حالی که باید مرتب به این سو و آن سوی کلاس بروم و به کار بچه‌ها رسیدگی کنم؟ خانم باربی خندید." لازم به استفاده از قلم و کاغذ و ثبت و ضبط نیست.» بعد ادامه داد، "شاید باور کردنش برای شما دشوار باشد، اما من در انجام این گونه کارها از یک تکه قیطان (نخ کلفت) استفاده می‌کنم. یک سر آن را به کمربندم می‌بندم و سر دیگر آن را آزاد رها می‌کنم و هر وقت می‌خواهم رفتاری را از دانش‌آموزی ثبت کنم یک گره به قیطان می‌زنم و آخر ساعت همهٔ گره‌ها را می‌شمارم. کار بسیار آسان و دقیقی است."

(آلبرتو و تروتمن، ۱۹۹۱، صص ۱۱۴-۱۱۳)

نمونه‌ای از آزمون موقعیتی در یک موقعیت شغلی

ما در ضمن توضیح روشهای سنجش رفتاری در این فصل موردی از آزمون موقعیتی گروهی را توضیح دادیم. در اینجا مثالی دیگر از این نوع آزمون راکه کاربرد فردی دارد توضیح می‌دهیم.

در این آزمون سناریو این گونه است که ناگهان به آزمون‌شونده گفته می‌شود که باید مدیریت شعبه‌ای از یک فروشگاه زنجیره‌ای محلی را بپذیرد. او در روز تعطیل آخر هفته از راه رسیده و بر روی میز خود انبوهی از مکاتبات و یادداشتها را می‌بیند. به آزمون‌شونده دربارهٔ منطقه و کارکنان فروشگاه اطلاعاتی داده می‌شود. آزمون‌شونده یک زمان محدود (۳ ساعت) در اختیار دارد تا مطالب را جذب کند، ترتیبی برای رسیدگی به امور بر اساس اولویت‌بندی تعیین نماید، و دستورالعمل‌های لازم را برای زیردستان خود مشخص کند تا در طول هفتهٔ بعد که او در مسافرت خواهد بود آنها کارها را انجام دهند. او همچنین در این باره یادداشت‌های لازم را درباره فعالیتهایی که در غیاب او باید انجام بگیرند آماده می‌کند.

پس از پایان یافتن زمان مشخص شده برای آزمون، آزمون‌شونده توسط یکی از اعضای گروهِ سنجش مصاحبه می‌شود تا معلوم گردد که روش کار او چگونه بوده و دلایل او در تعیین اولویت‌ها و دادن دستورالعمل‌ها چه هستند. سنجشگر گزارشی دربارهٔ آزمون‌شونده تهیه می‌کند، و در آن موارد زیر را می‌گنجاند: مهارتهای نوشتن، مهارتهای روابط انسانی، خلاقیت، نیاز به تأیید سرپرستان، استانداردهای کاری، تحمل بی‌اطمینانی‌ها، مقاومت در مقابل استرس یا تنش روانی، انرژی، سازمان‌دهی و طرح‌ریزی، و تصمیم‌سازی.

(ثرندایک و هیگن، ۱۹۷۹، ص ۵۰۳)

خلاصهٔ فصل

در روشهای سنجش رفتاری، بر خلاف آزمون و سایر روشهای غیرمستقیم سنجش، رفتار به‌طور مستقیم و بی‌واسطه مورد مشاهده و سنجش قرار می‌گیرد. این روشها وابسته به نظریه‌های رفتارگرایی روانشناسی هستند که در آنها اعتقاد بر این است که، به عوض مراجعه به مفاهیم و سازه‌های غیرقابل مشاهدهٔ روانی به عنوان منشاءهای رفتار، خود رفتار باید مورد توجه و رسیدگی قرار گیرد. لذا در سنجش نیز ویژگیهای رفتاری اندازه گیری می‌شوند.

برای سنجش رفتاری، ابتدا باید رفتار به دقت و به صورت عملیاتی تعریف شود و ویژگیها یا بُعدهای آن شناسایی گردند. بُعدهای عمدهٔ رفتاری مورد بحث در این فصل عبارت‌اند از شکل ظاهری، شدت، فراوانی، نرخ، طول مدت، دورهٔ نهفتگی، و موقعیت.

روشهای مختلف سنجش رفتاری شامل دو دسته‌اند: (۱) سنجش محصولات ماندنی رفتار، و (۲) سنجش رویدادهای گذرای رفتار. سنجش محصولات ماندنی رفتار به روشهایی اشاره می‌کند که توسط آنها آثاری که از رفتار باقی می‌مانند و پس از پایان یافتن رفتار قابل سنجش هستند مشاهده و سنجش می‌شوند. سنجش رویدادهای گذرای رفتار روشهایی را در بر می‌گیرد که برای اندازه‌گیری رفتارهایی ابداع شده‌اند که آثار ماندنی از خود به جای نمی‌گذارند و لذا در حین وقوع باید مورد مشاهده و سنجش قرار گیرند. روشهای عمدهٔ سنجش رویدادهای گذرای رفتار که در این فصل بحث شدند عبارت‌اند از: ثبت فراوانی یا ثبت رویداد رفتار، ثبت طول مدت رفتار، ثبت دورهٔ نهفتگی یا درنگ رفتار، ثبت فاصلهٔ زمان یا ثبت فاصله‌ای، و ثبت نمونه گیری زمان.

افزون بر روشهای فوق، ما در این فصل روشهای سنجش قیاس‌پذیر، سنجش موقعیتی، نقش‌گزاری، مشاهدهٔ نهان‌یابی، نظارت شخصی، سنجش شناختی - رفتاری، سنجش روانی - فیزیولوژیکی، و مصاحبهٔ رفتاری را به عنوان فنون سنجش رفتاری توضیح دادیم. در سنجش قیاس‌پذیر، شخص موردنظر در محیطی که به منظور افزایش احتمال بروز رفتار مورد سنجش طراحی شده است قرار داده می‌شود و در آن محیط رفتارش مشاهده و سنجش می‌شود. در سنجش موقعیتی یا آزمون موقعیتی، آزمون‌شونده در موقعیتی شبیه به یک موقعیت واقعی گذاشته می‌شود و در آن موقعیت رفتارش مشاهده و ثبت می‌گردد. در نقش‌گزاری یا ایفای نقش، از آزمون‌شونده خواسته می‌شود تا در موقعیتی که با موقعیت واقعی زندگی که رفتار شخص در آن قرار است پیش‌بینی شود شبیه است به ایفای یک نقش بپردازد و در آن هنگام رفتار او سنجش می‌شود. در مشاهدهٔ نهان‌یابی، سنجشگر در شرایطی به مشاهده و ثبت آزمون‌شونده می‌پردازد که او متوجه این کار نیست و به طور طبیعی به انجام رفتار مورد سنجش ادامه می‌دهد. در روش نظارت شخصی یا گزارش شخصی، از آزمون‌شونده یا فرد مورد سنجش خواسته می‌شود تا رفتار خودش را مورد مشاهده قرار داده آن را ثبت کند و به سنجشگر گزارش نماید.

در سنجش شناختی ـ رفتاری، فرایندهای شناختی مانند تفکر و اندیشیدن به عنوان رفتارهای نهان فرد تلقی شده و به کمک تدابیری مانند بلند فکر کردن و بلند حرف زدن آنها را سنجش می‌کنند. در سنجش روانی ـ فیزیولوژیکی، به کمک فنون ویژه‌ای نظیر بازخورد زیستی واکنشهای فیزیولوژیکی را به عنوان شاخصی برای رفتارهای فرد سنجش می‌کنند. و بالاخره در مصاحبهٔ رفتاری، تأکید سنجشگر در گفتگو با آزمون‌شونده یا شخص مورد سنجش بر شناسایی و ثبت ویژگیهای رفتار آشکار او و نیز کشف عوامل و رویدادهای ایجادکننده و نگهدارندهٔ رفتار است.

روشهای سنجش رفتاری، مانند سایر روشهای مورد بحث در این کتاب، دارای امتیازها و محدودیتهایی هستند. از امتیازهای این روش می‌توان موارد زیر را نام برد: (۱) با روش سنجش رفتاری هم می‌توان ویژگیهای رفتار مورد نظر را شناسایی و سنجش کرد و هم عوامل ایجادکننده و نگهدارندهٔ آن را. (۲) به کمک روشهای سنجش رفتاری هم می‌توان اطلاعات مورد نیاز ارزشیابی از موضوعها را به دست آورد و هم داده‌های مورد نیاز پژوهشهای آموزشی را گردآوری کرد. (۳) روشهای سنجش مستقیم مشکل روایی را که یکی از دردسرهای سنجش غیرمستقیم است ندارند. (۴) از روشهای سنجش رفتاری در بسیاری موارد می‌توان به جای روشهای دیگر سنجش از جمله پرسشنامه استفاده کرد. (۵) از روشهای سنجش رفتاری هم می‌توان ویژگیهای رفتار آشکار را شناسایی و سنجش کرد و هم فرایندهای شناختی را.

برای روشهای سنجش رفتاری محدودیتهایی نیز مطرح شده است. مهم‌ترین این محدودیتها به قرار زیرند: (۱) استفاده از این روشها غالباً وقت‌گیر و مستلزم صرف نیروی زیاد است. (۲) در پاره‌ای مواقع، به علت وجود پدیدهٔ واکنش‌پذیری، این روشها نیز از روایی کامل برخوردار نیستند. (۳) پاره‌ای از فنون سنجش رفتاری، مانند سنجش فیزیولوژیکی و نظارت شخصی و گزارش شخصی، دارای مشکلاتی هستند. مشکل روشهای روانی ـ فیزیولوژیکی دشواری اجرا و مشکل روش نظارت شخصی و گزارش شخصی احتمال نبود صداقت کامل در گزارشهای تهیه شده است.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. روشهای سنجش رفتاری و روشهای سنجش وابسته به آزمون از نظریه‌های روانشناسی متفاوتی مشتق شده‌اند.

الف ـ صحیح

ب ـ غلط

۲. از روش ثبت فاصله‌ای برای سنجش کدام بعد رفتار می‌توان استفاده کرد؟

الف ـ فراوانی

ب ـ کیفیت

پ ـ نرخ
ت ـ شدت

۳. مشکل عمدهٔ روش نظارت شخصی این است که:
الف ـ ضریب پایایی آن رضایت‌بخش نیست.
ب ـ واکنش‌پذیری آن زیاد است.
پ ـ در آن سنجشگر و شخص مورد سنجش یک نفر است.
ت ـ با سایر روشهای سنجش رفتاری همبستگی زیادی ندارد.

۴. سنجش رفتاری را تعریف کنید و ویژگیهای آن را توضیح دهید.

۵. روشهای مختلف سنجش رفتاری را با ذکر نمونه‌ای برای هر یک از آنها توضیح دهید.

۶. ویژگیهای روشهای زیر را با هم مقایسه کنید:
الف ـ سنجش مبتنی بر مشاهدهٔ رفتار در موقعیتهای طبیعی
ب ـ سنجش موقعیتی
پ ـ نقش‌گزاری
ت ـ سنجش نهان‌یابی
ث ـ نظارت شخصی و گزارش شخصی
ج ـ سنجش شناختی ـ رفتاری
چ ـ سنجش روانی ـ فیزیولوژیکی
ح ـ مصاحبهٔ رفتاری

۷. واکنش‌پذیری را تعریف کنید و راههای کاهش دادن آن را توضیح دهید.

۸. بُعدهای مختلف رفتاری را نام ببرید و با ذکر مثال آنها را توضیح دهید.

۹. امتیازها و محدودیتهای روشهای سنجش رفتاری را توضیح دهید.

۱۰. از جدول مشخصات آزمونی که به عنوان تمرین برای فصل ۶ کتاب تهیه کرده‌اید هدفهایی را که به صورت سنجش رفتاری قابل سنجش هستند انتخاب کنید، روش یا روشهای مناسب سنجش آنها را مشخص کنید، و مراحل لازم برای سنجش آن رفتارها را توضیح دهید.

۱۶

سنجش عاطفی

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. ویژگیهای عاطفی را تعریف کند.

۲. اهمیت ویژگیهای عاطفی و پرورش و سنجش آنها را توضیح دهد.

۳. ویژگیهای روشهای سنجش عاطفی را در مقایسه با ویژگیهای سنجش شناختی و روانی-حرکتی توضیح دهد.

۴. روشهای مختلف سنجش ویژگیهای عاطفی را نام ببرد.

۵. با ذکر مثال، روشهای پرسشنامه‌ای را برای سنجش عواطف توضیح دهد.

۶. قواعد تهیه و اجرای پرسشنامه‌های سنجش عواطف را توضیح دهد.

۷. برای سنجش ویژگیهای عاطفی دانش‌آموزان پرسشنامه‌های لازم را تهیه کند.

۸. روش تفکیک معنایی را شرح دهد، و قواعد تهیه و اجرا و چگونگی نمره گذاری آن را توضیح دهد.

۹. برای یکی از ویژگیهای عاطفی دانش‌آموزان از روش تفکیک معنایی استفاده کند.

۱۰. مقیاس لیکرت را شرح دهد.

۱۱. قواعد تهیه و اجرا و چگونگی نمره گذاری مقیاس لیکرت را توضیح دهد.

۱۲. برای سنجش ویژگیهای عاطفی دانش‌آموزان یک پرسشنامه مبتنی بر مقیاس لیکرت تهیه و اجرا کند.

۱۳. گروه‌سنجی را تعریف کند و دو روش معروف آن را توضیح دهد.

۱۴. امتیازها و محدودیتهای روشهای سنجش عاطفی را توضیح دهد.

۱۵. موارد استفادهٔ روشهای سنجش عاطفی را توضیح دهد.

۱۶. برای هدفهای عاطفی مختلف روشهای سنجش عاطفی متناسب با آنها را به کار بندد.

توجه به ویژگیهای عاطفی دانش‌آموزان و پرورش عاطفه‌های مثبت در آنان از وظایف مهم آموزش و پرورش است، زیرا ویژگیهای عاطفی هم در رشد شخصیت و هم در پیشرفت تحصیلی یادگیرندگان نقش بسیار مهمی ایفا می‌کنند. به عنوان نمونه، نگرش نسبت به درس و مدرسه بر فعالیت یادگیری، و اعتماد به نفس و مفهوم خود (خودپنداره) بر چگونگی انجام کارهای دانش‌آموزان تأثیر می‌گذارند، و ارزشی که آنان برای راستی و درستی قائل می‌شوند رفتار و کردارشان را شکل می‌دهد.

معلمان همواره شاهد این واقعیت بوده‌اند دانش‌آموزانی که از لحاظ توانایی سرآمدند امابه انجام کارها چندان علاقه‌مند نیستند از کسانی که توانایی کمتری دارند ولی از انگیزهٔ سطح بالایی برای انجام کارها برخوردارند کمتر موفق می‌شوند. همچنین پژوهشهای متعدد تربیتی نشان داده‌اند که افزایش جنبه‌های مثبت عاطفی دانش‌آموزان به ویژه در رابطه با درس و تحصیل نه تنها بر موفقیت تحصیلی بلکه بر سلامت روانی آنان نیز تأثیر می‌گذارد (بلوم، ۱۹۸۲، ترجمهٔ سیف، ۱۳۶۸).

تعریف ویژگیهای عاطفی

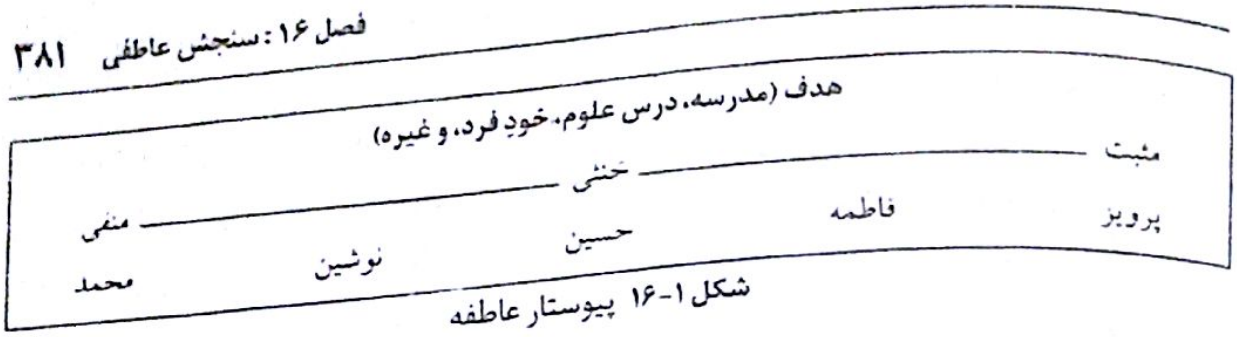
آندرسون و بورک (۲۰۰۰)، در اثر تخصصی خود با عنوان سنجش ویژگیهای عاطفی در آموزشگاه، در تعریفی که برای ویژگی عاطفی به دست داده گفته‌اند یک ویژگی آدمی در صورتی که دارای دو ملاک زیر باشد ویژگی عاطفی نامیده می‌شود: (۱) دربرگیرندهٔ احساس[1] یا هیجان[2] باشد، و (۲) ویژگی معمولی یا نوعی عاطفی فرد باشد. افزون بر اینها، هر ویژگی عاطفی باید دارای سه ملاکِ خاص شدت، جهت، و هدف نیز باشد.

منظور از شدت[3] درجه یا نیروی احساس یا عاطفه است. بعضی از احساسها از دیگر احساسهای ما از شدت بیشتری برخوردارند. مثلاً "عشق" از "دوستی" شدیدتر است. جهت[4] به جنبهٔ مثبت یا منفی احساس گفته می‌شود. به سخن دیگر، جهت مشخص می‌کند که یک احساس خوب است یا بد. بعضی از احساسها ذاتاً خوب یا بد هستند. مثلاً درد بد است، اما لذت خوب است. بعضی احساسهای دیگر، بنا به تعریفی که در فرهنگهای مختلف از آنها می‌شود، خوب یا بد هستند. مثلاً، لذت بردن از مدرسه، در بیشتر فرهنگها، یک احساس خوب است. در مقابل، اضطراب بد به حساب می‌آید. اکثر عواطف مثبت یک جنبهٔ منفی هم دارند. مثلاً، در مقابل لذت بردن از مدرسه که یک عاطفهٔ مثبت است، تنفر از مدرسه یک عاطفهٔ منفی تلقی می‌شود. همچنین آرامش به عنوان یک عاطفهٔ مثبت در مقابل تنش و اضطراب که منفی هستند قرار دارد.

بهتر است احساسها یا عواطف را، به جای اینکه در دو قطب متضاد و در مقابل هم قرار دهیم، در یک پیوستار تصور کنیم که در یک سوی آن قوی‌ترین عاطفهٔ مثبت و در سوی دیگرش

---

1. feeling 2. emotion 3. intensity 4. direction



شکل ۱-۱۶ پیوستار عاطفه

قوی‌ترین عاطفهٔ منفی قرار می‌گیرد، و در میانهٔ آنها درجات مختلفی از عاطفه دیده می‌شوند. در این پیوستار، نقطهٔ وسطْ عاطفهٔ مثبت را از عاطفهٔ منفی جدا می‌کند، و هرچه از این نقطه به طرف دو قطب فاصله بیشتر باشد به همان نسبت شدت مثبت یا منفی بودن عاطفه بیشتر خواهد بود. شکل ۱-۱۶ این پیوستار عاطفی را نشان می‌دهد.

چنان که در شکل ۱-۱۶ می‌بینید، پنج نفر دانش‌آموزی که نامشان در شکل آمده از لحاظ شدت عاطفه، جهت عاطفه، یا هردو با هم فرق دارند. پرویز و محمد هر دو از لحاظ شدت عاطفه یکسان‌اند، اما جهت عاطفه‌شان با هم متفاوت است. همین حالت در رابطه با فاطمه و نوشین نیز صادق است. احساسهای پرویز و فاطمه از لحاظ جهت هم‌سو هستند، اما از لحاظ شدت متفاوت‌اند. همین وضع در مورد نوشین و محمد نیز صادق است. عاطفهٔ حسین نه جهت دارد و نه شدت.

سومین ملاک عواطف یعنی هدف یا آماج[1] به موضوعی که عاطفه یا احساس متوجه آن است اشاره می‌کند. به عنوان مثال، اگر اضطراب ویژگی عاطفی دانش‌آموز باشد می‌تواند آماجهای مختلفی را شامل شود. اضطرابِ دانش‌آموز ممکن است متوجه درس ریاضی، معلمان مدرسه، امتحانات، و غیره باشد. همین‌ها هستند که در بالای پیوستار اضطراب (شکل ۱-۱۶) قرار داده شده‌اند. آندرسون و بورک (۲۰۰۰)، در ادامهٔ تعریف و تفسیر خود از ویژگیهای عاطفی، مطلب را به نحو زیر خلاصه کرده‌اند:

> سخن کوتاه، ویژگیهای عاطفی دارای پنج جنبهٔ تعریف‌کننده هستند که دوتای آنها کلی و سه تای دیگر اختصاصی‌اند. نخست اینکه، ویژگیهای عاطفی احساس یا هیجان هستند. این جنبه عاطفه را از سایر ویژگیهای آدمی متمایز می‌سازد. دوم اینکه ویژگیهای عاطفی احساسهای معمولی یا نوعی فرد هستند. این جنبه از ویژگیهای عاطفی آنها را از واکنشهای هیجانی که در موقعیتهای خاص ظاهر می‌شوند متمایز می‌سازد. سوم اینکه، ویژگیهای عاطفی از درجه‌ای از شدت برخوردارند. چهارم اینکه، دارای جهت هستند. و پنجم اینکه، برای ویژگیهای عاطفی نوعی هدف یا آماج وجود دارد که سمت و سوی آنها را مشخص می‌کند. (ص ۶)

اهمیت ویژگیهای عاطفی

در ارتباط با آموزش و پرورش، ویژگیهای عاطفی دانش‌آموزان هم وسیله به حساب می‌آیند و هم هدف. آندرسون و بورک (۲۰۰۰)، برای ویژگیهای عاطفی دانش‌آموزان سه نقش مهم زیر را قائل شده‌اند:

---

1. target

۱. به عنوان ویژگیهای ورودی عاطفی (در کنار ویژگیهای ورودی شناختی و روانی - حرکتی) به صورت پیش‌نیازهای یادگیری[1] درسهای مختلف در آموزشگاه.
۲. به عنوان صافی یا فیلتری که یادگیرندگان از طریق آن فعالیتها و رویدادهای مدرسه را می‌بینند. "ویژگیهای عاطفی بخش مهمی از لنزهایی را تشکیل می‌دهند که یادگیرندگان از درون آنها رویدادها و فعالیتهایی را که در آموزشگاه و کلاس درس رخ می‌دهند ادراک می‌کنند و به آنها واکنش نشان می‌دهند" (آندرسون و بورک، ۲۰۰۰، ص ۸).
۳. به عنوان هدفهای مهم آموزشی که به خودی خود دارای اهمیت‌اند. اینها بخشی از نتایج تحصیل در آموزشگاه را تشکیل می‌دهند که به بازده‌های عاطفی یادگیری شهرت دارند (بلوم، ۱۳۸۲، ترجمهٔ سیف، ۱۳۶۳). این بازده‌های عاطفی (از جمله علاقه به یادگیری تمام عمر و نگرش نسبت به خود یا خودپنداره) تأثیر درازمدتی بر شخصیت و سلامت روانی فرد به جای می‌گذارد.

به رغم اهمیت فراوانی که پرورش و سنجش عواطف دانش‌آموزان دارد، برای معلمان و سایر پرورشکاران وظایف درخور توجهی از این بابت پیش‌بینی نشده است. بنا به گفتهٔ پوفام (۲۰۰۲)، "چندبار شنیده‌اید که معلمی کوشیده باشد تا شواهد سیستماتیک یا نظامداری دربارهٔ نگرشها و ارزشهای دانش‌آموزان خود جمع‌آوری نماید؟ متأسفانه، سنجشِ نظامدارِ عواطف غیرمتداول است" (ص ۲۱۷).

اگر بپذیریم که پرورش ویژگیهای عاطفی دانش‌آموزان، همراه با پرورش ویژگیهای شناختی و روانی - حرکتی آنان از هدفهای مهم فعالیتهای آموزشی معلمان است، باید بپذیریم، همچنان که سنجش ویژگیهای شناختی و روانی - حرکتی دارای اهمیت است، سنجش ویژگیهای عاطفی نیز باید مورد توجه قرار گیرد. آندرسون و بورک (۲۰۰۰) در ارتباط با اهمیت سنجش عاطفی گفته‌اند:

> به عبارت ساده، ما نیاز داریم بدانیم دانش‌آموزان (الف) وقتی که وارد مدرسه و کلاس درس می‌شوند از چه ویژگیهای عاطفی‌ای برخوردارند، (ب) کلاسها و مدرسه‌شان را چگونه تصور می‌کنند، و (پ) به چه میزانی به هدفهای مهم عاطفی دست می‌یابند. در مقیاس گروهی، این اطلاعات در طراحی محیطهای آموزشی (کلاس و مدرسه) به ما کمک می‌کنند. در مقیاس فردی، این اطلاعات ما را در شناسایی دانش‌آموزانی یاری می‌دهند که از ویژگیهای عاطفی مطلوب برخوردار نیستند، یا این ویژگیها در آنان رشد نکرده، یا اینکه ویژگیهای عاطفی‌شان مانع موفقیت تحصیلی یا شرکت فعال آنان در کارهای آموزشگاهی می‌شود. شناسایی به موقع این دانش‌آموزان به ما امکان می‌دهد تا برای افزایش احتمال موفقیت آنان در تحصیل و کاهش افت تحصیلی کمکشان کنیم. (صص ۱۷-۱۶)

---

۱. برای کسب اطلاع از ویژگیهای ورودی سه‌گانه و تأثیر آنها بر پیشرفت تحصیلی دانش‌آموزان و دانشجویان به سیف (۱۳۸۶) مراجعه کنید.

ویژگیهای روشهای سنجش عاطفی

پیش از معرفی روشها و فنون سنجش عاطفی، این نکته قابل ذکر است که هدفهای عاطفی را نمی‌توان با استفاده از آزمونهای توانایی سنجید. چنان‌که در فصل ۵ گفتیم، آزمونهای توانایی از آزمون‌شونده می‌خواهند بکوشد تا حداکثر توان خود را در پاسخ به سؤالهای آزمون نشان دهد، اما ابزارهای سنجش هدفهای عاطفی از آزمون‌شونده می‌خواهند هر آنچه را که به طور معمول احساس می‌کند بگوید. این همان تمایزگذاری بین عملکرد بیشینه و عملکرد نوعی است که در فصل ۱ به آن اشاره کردیم. بنابراین، مشکلِ اساسیِ روشهای سنجش هدفهای عاطفی این است که باید موقعیتهایی را به وجود آورند که در آنها آزمون‌شونده رفتار عاطفی مورد سنجش را بروز دهد. به سخن دیگر، به خلاف سنجش شناختی که در آن کوشش می‌شود تا اطلاعات مربوط به ویژگیهای شناختی فرد از طریق سنجش رفتارهایی به دست آیند که قابل تقسیم شدن به درست یا غلط هستند یا می‌توان آنها را در یک پیوستار از درست یا غلط یا از کیفیت خوب تا کیفیت بد قرار داد، در سنجش عاطفی سعی می‌شود تا اطلاعات مربوط به ویژگیهای عاطفی فرد از طریق سنجش رفتارهایی به دست آیند که جزو رفتارهای نوعی او هستند و به طور معمول از او سر می‌زنند.

به سبب ویژگیهای پیش‌گفته، اگر آزمون شونده (دانش‌آموز یا دانشجو) احساس کند که نتایج سنجش عاطفی همانند نتایج آزمونهای توانایی برای نمره دادن و قضاوت دربارهٔ او به کار خواهند رفت، احتمال زیادی دارد که در پاسخ به سؤالهای وسایل سنجش جواب درست ندهد و لذا روایی نتایج حاصل لطمه ببیند. البته آنجا که اطلاعات گروهی کفایت می‌کنند و نیازی به این نیست که فرد فرد دانش‌آموزان شناخته شوند، مانند زمانی که ارزشیابی از یک برنامهٔ درسی یا روش آموزشی یک معلم مورد نظر است، می‌توان از آنان خواست تا نظرهای خود را به طور ناشناس بیان کنند. در چنین شرایطی، از بابت اینکه پاسخ دهندگان به سؤالهای ابزار سنجش پاسخ درست بدهند مشکل زیادی به وجود نمی‌آید. اما بعضی وقتها لازم است دانش‌آموزان در سنجش رفتارهای عاطفی شناسایی شوند، مثل زمانی که از نتایج سنجش برای ارزشیابی تکوینی یا راهنمایی تحصیلی و شغلی دانش‌آموز استفاده می‌شود. در چنین مواردی باید به دانش‌آموز اطمینان داد که از نتایج ارزشیابی عاطفی برای نمره دادن یا قضاوت دربارهٔ او استفاده نخواهد شد. بنابراین، در سنجش ویژگیهای عاطفی دانش‌آموزان و دانشجویان باید اطمینان حاصل نمایند که از نتایج حاصل به ضررشان استفاده نخواهد شد تا اینکه در پاسخ به سؤالها صادقانه عمل کنند.

روشهای مختلف سنجش ویژگیهای عاطفی

هرچند از روشهایی که تاکنون در این کتاب، به ویژه سه فصل اخیر، مورد بحث قرار دادیم می‌توان برای سنجش عواطف نیز سود برد، با این حال، به خاطر اهمیت روزافزون هدفهای عاطفی در آموزش و پرورش، ما فصل حاضر را به طور کامل به سنجش این هدفها اختصاص داده‌ایم. به طور کلی، می‌توان روشهای سنجش عاطفی را به دو دستهٔ اصلی تقسیم کرد: (۱) مشاهده[1] و (۲) گزارش شخصی[2]. آندرسون و بورک (۲۰۰۰) این دو روش را با این جمله که ما "هم نگاه می‌کنیم و هم گوش می‌دهیم" (ص ۴۹) معرفی کرده‌اند. آنان، در توصیف این دو روش، کار معلم یا سنجشگر را با کار پزشک در سنجش و تشخیص علتهای مشکلات بیماران مقایسه کرده‌اند. وقتی که بیماری به پزشک مراجعه می‌کند و از او برای درمان بیماری‌اش کمک می‌طلبد، پزشک ابتدا می‌کوشد تا علتهای مشکل او را تشخیص دهد. بدین منظور، به جمع‌آوری اطلاعات موردنیاز خود برای انجام این کار اقدام می‌نماید. به چشم، گوش، یا گلوی بیمار نگاه می‌کند، درجهٔ تب او را می‌بیند، وزنش را اندازه می‌گیرد، و نتایج آزمایشهای او را بررسی می‌کند. افزون بر اینها، پزشک شکایتها و توضیحات بیمار را می‌شنود و به ضربان قلب و سرفه کردن او گوش می‌دهد. آنگاه، براساس اطلاعات حاصل از این دو نوع منبع (مشاهده و گزارش شخصی بیمار)، در مورد علتهای بیماری او نتیجه گیری می‌کند.

معلم یا سنجشگر، مانند پزشک، دو روش مهم جمع‌آوری اطلاعات دربارهٔ ویژگیهای عاطفی دانش‌آموزان و دانشجویان یعنی مشاهده و گزارش شخصی را در اختیار دارد. روش مشاهده یعنی جمع‌آوری اطلاعات از راه دیدن اعمال و رفتارهای شخص مورد سنجش، و روش گزارش شخصی یعنی جمع‌آوری اطلاعات از راه پاسخهایی که شخص مورد سنجش یا آزمون‌شونده به سؤالهایی که از او پرسیده می‌شود می‌دهد. ما روشهای سنجش مشاهده‌ای را به تفصیل در دو فصل پیش توضیح دادیم. روشهای گزارش شخصی را در دنبالهٔ مطالب فصل حاضر توضیح می‌دهیم.

راههای مختلفی وجود دارند که در آنها می‌توان از دانش‌آموز یا دانشجو خواست تا به صورت گزارش شخصی اطلاعات موردنیاز سنجشگر را در اختیار او قرار دهد. ما در دنبالهٔ مطالب این فصل مهم‌ترین این راه‌ها را توضیح می‌دهیم.

استفاده از پرسشنامه برای سنجش عواطف

یکی از روشهای بسیار معمول سنجش متغیرهای عاطفی پرسشنامه است. البته درست‌تر آن است که در سنجش عواطف به جای پرسشنامه از اصطلاح عقیده‌سنج[3] استفاده کنیم. تفاوت این دو اصطلاح در آن است که عقیده‌سنج بر احساس تأکید می‌کند و برای سنجش متغیرهای عاطفی مناسب‌تر است، در حالی که پرسشنامه بر دانش واقعیتها تأکید می‌کند و برای سنجش متغیرهای

1. observation 2. self-report 3. opinionnaire

شناختی مفیدتر است.) با این حال، چون اصطلاح پرسشنامه متداول‌تر از عقیده‌سنج است، ما نیز در این کتاب آن را ترجیح داده‌ایم.

به طور کلی، دو نوع عقیده‌سنج یا پرسشنامه برای سنجش متغیرهای عاطفی وجود دارند: بسته‌پاسخ و بازپاسخ. در زیر به توضیح آنها توجه کنید.

پرسشنامه‌های بسته پاسخ

روش استفاده از **پرسشنامه‌های بسته‌پاسخ یاگزینه بایست**[1] بسیار شبیه به آزمونهای عینی مورد استفاده در سنجش رفتارهای شناختی است، زیرا در هر دو روش اندازه‌گیری آزمون‌شونده باید از میان پاسخهای از پیش تعیین‌شده برای هر سؤال یکی را برگزیند و از خود چیزی به آن نیفزاید. در زیر به نمونه‌هایی از سؤالهای این‌گونه پرسشنامه‌ها که برای سنجش هدفهای مختلف طبقه‌های حوزهٔ عاطفی تهیه شده‌اند توجه کنید. این سؤالها برای سنجش هدفهای عاطفی مربوط به خواندن درست شده‌اند. آزمون‌شونده به سؤالها با بله یا نه پاسخ می‌دهد.

دریافت کردن

- دوست دارید در یک انجمن کتاب که بیشتر فعالیت اعضای آن بحث و بررسی کتابهای مختلف است عضو شوید؟

بله نه

پاسخ دادن

- معمولاً برای شما غیرممکن است که یک ساعت مطالعه کنید و حوصله‌تان سر نرود؟

بله نه

ارزش گذاردن

- آیا اغلب چنان غرق خواندن می‌شوید که از اتفاقاتی که دور و برتان می‌افتد بی‌خبر می‌مانید؟

بله نه

سازمان دادن

- آیا خواندن کتابها یا مقاله‌ها موجب شده است که شما یک سرگرمی تازه برای خود برگزینید؟

بله نه

(بلوم، هستینگز[2]، و مادوس[3]، ۱۹۷۱، ص ۲۳۲)

در زیر نمونه‌های دیگری از سؤالهای گزینه بایست را که برای سنجش هدفهای عاطفی در درس آموزش موسیقی تهیه شده‌اند می‌آوریم. دانش‌آموز به سؤالها با صحیح یا غلط پاسخ می‌دهد.

دریافت کردن

- اگر یکی از اعضای خانوادهٔ من به موسیقی رادیو گوش کند من ترجیح می‌دهم که اتاق را ترک کنم.

صحیح غلط

---

1. forced-choice questionnaires 2. Hastings 3. Madaus

• من دوست دارم مقداری از وقتم را صرف گوش دادن به موسیقی بکنم.
صحیح غلط

پاسخ دادن

• معمولاً نمی‌توانم برای مدت دو ساعت به طور متوالی به موسیقی گوش کنم.
صحیح غلط

• دست کم سالی یک بار به کنسرت موسیقی می‌روم.
صحیح غلط

ارزش گذاردن

• غیرمعمول نیست که دربارهٔ موضوعهای مربوط به موسیقی صحبت کنم.
صحیح غلط

• بیشتر اوقات مقاله‌هایی را که دربارهٔ موسیقی نوشته شده‌اند با اشتیاق می‌خوانم.
صحیح غلط

سازمان دادن

• معمولاً وقتی که به موسیقی گوش می‌دهم می‌کوشم تا دربارهٔ کیفیت آن به داوری بپردازم.
صحیح غلط

• دوست دارم نظرم را دربارهٔ یک قطعه موسیقی که می‌شنوم با دیگران در میان بگذارم.
صحیح غلط

(بلوم، هستینگز، و مادوس، ۱۹۷۱، ص ۲۳۴)

پرسشنامه‌های بسته‌پاسخ یا گزینه بایست را، علاوه بر نمونه‌های بالا، می‌توان به شکلهای مختلف دیگری تهیه کرد. در زیر به نمونه‌های دیگری از این‌گونه پرسشنامه‌ها توجه کنید.

• در جلسهٔ بعدی درس علوم‌اجتماعی،
الف ـ می‌خواهم تنها کار کنم.
ب ـ می‌خواهم با یک یا دو نفر از دوستانم کار کنم.
پ ـ می‌خواهم با یک گروه کار کنم.

• برای درس علوم، زمان چهل دقیقه‌ای _____ .
الف ـ خیلی زیاد است.
ب ـ کافی است.
پ ـ کم است.

	همیشه	گاهی‌وقتها	هرگز
• آیا معلم به شما وقت کافی می‌دهد تا آزمونهایتان را تمام کنید؟	ــــ	ــــ	ــــ
• وقتی که با مشکلی روبه‌رو می‌شوی معلم به هرفعالیت گوش می‌کند؟	ــــ	ــــ	ــــ
• به نظر شما وقت این کلاس زود تمام می‌شود؟	ــــ	ــــ	ــــ

• برای هر یک از فعالیتهای زیر نظر خود را باکشیدن دایره‌ای به دور شماره‌های ۵ تا ۱ نسبت به آن فعالیت مشخص کنید. شمارهٔ ۵ یعنی آن فعالیت را خیلی دوست دارید، ۴ یعنی آن را دوست دارید، ۳ یعنی نسبت به آن بی‌تفاوت هستید، ۲ یعنی آن را دوست ندارید، و ۱ یعنی اصلاً آن را دوست ندارید.

الف - صرف وقت در کتابخانه

ب - گوش دادن به موسیقی کلاسیک

۵	۴	۳	۲	۱
۵	۴	۳	۲	۱

(دمبو، ۱۹۹۴، ص ۵۹۰)

باز هم در زیر به نمونه‌های دیگری از سؤالهای گزینه بایست یا پرسشنامه‌های بسته‌پاسخ توجه کنید.

• من زبان را ــــ از ریاضی دوست دارم.

الف - خیلی بیشتر

ب - کمی بیشتر

پ - خیلی کمتر

ت - کمی کمتر

• من این مسئله را "؟ = ۸۷۴۸۳ + ۹۷۴۸ + ۴۶۹۳۷ - ۹۹۹ + ۶۸۴۰ - ۵۷۴ - ۳۵۹" از این مسئله "قد علی نصف قد جمشید است. قد محمد نصف قد علی است. قد حسن نصف قد جمشید است. قد جمشید ۱۳۰ سانتیمتر است. قد محمد چقدر است؟" ــــ دوست دارم.

الف - خیلی بیشتر

ب - کمی بیشتر

پ - خیلی کمتر

ت - کمی کمتر

• اگر در آزمون عملکرد خوبی داشتم علتش این است:

الف ـ آزمون ساده بود.

ب ـ من خوب درس خواندم.

• اگر دانش‌آموزان ما چیزی از درس معلم یاد نمی‌گیرند علتش این است که:

الف ـ معلم خوب درس نمی‌دهد.

ب ـ دانش‌آموزان توانایی یادگیری مطالب را ندارند.

(بلوم، هستینگز، و مادوس، ۱۹۷۱، ص ۲۳۹)

پرسشنامه‌های بازپاسخ

در سنجش هدفهای حوزهٔ عاطفی، روش **پرسشنامه‌های بازپاسخ**[1] با روش آزمونهای تشریحی و کوته‌پاسخ در سنجش هدفهای حوزهٔ شناختی شبیه است. در اینجا نیز، مانند آزمونهای تشریحی، آزمون‌شونده سؤالی یا بیانی را دریافت می‌کند و در پاسخ به آن نظر خود را توضیح می‌دهد. در روش سؤالهای بازپاسخ، غالباً سؤالها به صورت جملات ناتمام در اختیار آزمون‌شونده قرار می‌گیرند و او باید با نوشتن نظر خود در دنبالهٔ جملات ناتمام آنها را کامل کند. به نمونه‌هایی از سؤالهای بازپاسخ که در زیر آمده‌اند و برای سنجش نگرشها، باورها، فعالیتها، و ارزشها طرح شده‌اند توجه کنید.

۱. با یک هدیهٔ ۱۰۰۰۰۰ تومانی...

۲. اگر آخر این هفته سه روز تعطیل بود...

۳. می‌توانم روی بهترین دوستم حساب کنم که...

۴. بدترین روزهای من آنهایی هستند که...

۵. کودکان من مجبور نخواهند بود که...

۶. وقتی که می‌بینم مردم... خیلی ناراحت می‌شوم.

۷. اگر اتومبیلی از خودم داشتم...

۸. بالاخره تصمیمم را گرفته‌ام که... یاد بگیرم.

۹. اگر حق عضویت مجانی عضو شدن در دو مجله به من داده شود... و... را انتخاب خواهم کرد.

۱۰. من بر این باورم که...

۱۱. من آرزو می‌کنم که...

۱۲. پیشنهاد من به جوانان این است که...

(بلوم، هستینگز، و مادوس، ۱۹۷۱، ص ۲۳۷)

در زیر به نمونه‌های دیگری از سؤالهای بازپاسخ که برای سنجش هدفهای عاطفی تهیه شده‌اند توجه کنید. دقت کنید که این سؤالها برای سنجش نگرش دانش‌آموزان نسبت به کلاس درس و عوامل وابسته به آن قابل استفاده‌اند.

---

1. open-ended questions

۱. معلم من...
۲. آنچه بیشتر از هر چیز دیگر دربارهٔ این کلاس دوست دارم...
۳. آنچه بیشتر از هر چیز دیگر دربارهٔ این کلاس من را ناراحت می‌کند...
۴. بچه‌های دیگر در این کلاس...

(دمبو، ۱۹۹۴، ص ۵۹۰)

پرسشنامهٔ بازپاسخ از پرسشنامهٔ بسته پاسخ برای استفاده در کلاس درس مفیدتر است، زیرا سؤالهای این‌گونه پرسشنامه‌ها به معلم اجازه می‌دهند تا دامنهٔ گسترده‌ای از موضوعهای گوناگون را به صورتی کارآمد مورد استفاده قرار دهد (پین، ۲۰۰۳). با این حال، تهیهٔ سؤالهای خوب برای پرسشنامه‌ها و تحلیل درست نتایج دشوار و وقت‌گیر است. در زیر به رهنمودهایی برای کاستن از این‌گونه مشکلات توجه کنید.

قواعد تهیه و اجرای پرسشنامه‌های سنجش عواطف

پین (۲۰۰۳) برای سهولت تهیه و اجرای روشها و ابزارهای سنجش عاطفه‌ها قواعدی را در اختیار معلمان قرار داده است. ما آنها را به طور خلاصه در زیر نقل می‌کنیم.

۱. پیش از تهیهٔ ابزار یا وسیلهٔ سنجش، منظور و مشخصات چیزی را که مورد نیازتان است مشخص کنید.
۲. پس از آنکه تصمیم به تهیهٔ پرسشنامه گرفتید، جمله‌های ساده، واضح، و مختصر بنویسید.
۳. از نوشتن جمله‌هایی که هیچ کس آنها را تأیید نمی‌کند پرهیز کنید.
۴. جمله‌ها را در زمان حال بنویسید نه در زبان گذشته یا آینده.
۵. از نوشتن جمله‌های منفی، به ویژه منفی مضاعف، مثلاً "در این مدرسه یک معلم یافت نمی‌شود که به دانش‌آموزان احترام نگذارد"، پرهیز کنید.
۶. جمله‌هایی بنویسید که مُعرف عواطف باشند نه واقعیتها.
۷. در هر جمله یا سؤال پرسشنامه بیشتر از یک اندیشه قرار ندهید.
۸. تعداد سؤالها را زیاد نکنید. اگر دانش‌آموزان از پاسخ دادن خسته شوند به سؤالها جواب دقیق نخواهند داد.
۹. پیش از تهیهٔ فرم نهایی، پرسشنامهٔ تهیه شده را با معلمان همکار و تعدادی از دانش‌آموزان مرور کنید تا سؤالهای مبهم، اصطلاحات پیچیده، و معانی نارسا را پیدا و اصلاح نمایید.
۱۰. هنگام اجرا، از اینکه دانش‌آموزان هدف پرسشنامه را درک می‌کنند مطمئن شوید و آنان را به اهمیت پاسخ دادن به طور کامل و صادقانه متقاعد سازید.
۱۱. سعی کنید دانش‌آموزان را وادار نمایید تا با علاقه و دقّت به سؤالها جواب دهند.

۱۲. جلسهٔ اجرای پرسشنامه را به دقت کنترل نمایید و از اینکه دانش‌آموزان دربارهٔ سؤالها با هم مشورت کنند جلوگیری نمایید.
۱۳. دانش‌آموزان را تشویق کنید تا آن‌گونه که خودشان فکر می‌کنند پاسخ دهند نه به صورتی که مورد علاقهٔ معلم است.

روش تفکیک معنایی

یکی دیگر از روشهای مهم سنجش نگرش روش **تفکیک معنایی**[1] یا افتراق معنایی چارلز آزگود[2] (آزگود، سوسی[3]، و تَننبام[4]، ۱۹۶۷) است. این روش شامل تعدادی صفت دوقطبی (مثلاً خوب-بد، زیبا-زشت، عاقل-ابله، و غیره) است که در آن هر جفت صفت یک پیوستار هفت بخشی می‌سازد و به صورت مقیاس درجه‌بندی تنظیم می‌گردد. در این روش، موضوع نگرش در بالای صفحه قرار داده می‌شود و از پاسخ دهندگان خواسته می‌شود تا با گذاشتن علامت (مثلاً x) در روی پیوستار موضوع نگرش را درجه‌بندی کنند. در شکل ۲-۱۶ به نمونه‌ای از روش تفکیک معنایی توجه کنید.

در استفادهٔ گسترده، تفکیک معنایی روشی است برای اندازه‌گیری معنای ضمنی مفاهیم. مفهومهایی مانند خانه، پدر، آزادی، مدرسه، و غیره به دو گونه معنی می‌دهند: (۱) **معنی صریح**[5] که معنی واقعی یا واژگانی آن است؛ و (۲) **معنی ضمنی**[6] که نشان‌دهندهٔ نگرش فرد نسبت به آن مفهوم است. مدرسه به عنوان یک مفهوم، به طور صریح یا واژگانی، فضایی تشکیل یافته از یک حیاط و تعدادی کلاس درس معنی می‌دهد، اما به طور ضمنی معنی‌های دیگری دارد که در شکل ۳-۱۶ آمده‌اند. هر مفهوم یا محرک را می‌توان در طول تعدادی صفت یا ویژگیِ دو قطبی درجه‌بندی کرد، و این درجه‌بندی مُعرف معنی اصطلاح مورد درجه‌بندی است.

سازمان ملل متحد

بی‌فایده	۱ : ۲ : ۳ : ۴ : ۵ : ۶ : ۷	مفید
ضعیف	۱ : ۲ : ۳ : ۴ : ۵ : ۶ : ۷	قوی
غیرمؤثر	۱ : ۲ : ۳ : ۴ : ۵ : ۶ : ۷	مؤثر

شکل ۲-۱۶ یک نمونه تفکیک معنایی

1. semantic differential
2. Charles Osgood
3. Suci
4. Tannenbaum
5. connotative meaning
6. denotative meaning

	مدرسه							
خوب	+۳	+۲	+۱	۰	-۱	-۲	-۳	بد
فعال	+۳	+۲	+۱	۰	-۱	-۲	-۳	منفعل
متنوع	+۳	+۲	+۱	۰	-۱	-۲	-۳	تکراری
مفید	+۳	+۲	+۱	۰	-۱	-۲	-۳	غیرمفید

شکل ۳-۱۶ نمونه‌ای از تفکیک معنایی

برای تعیین نمرهٔ پرسشنامهٔ تفکیک معنایی، تمام ارزشهای عددی مکانهایی که برای آنها علامت (x) گذاشته شده است با هم جمع می‌شوند. حاصل این جمع نمرهٔ کل پرسشنامه را به دست می‌دهد. علاوه بر روش علامت‌گذاری به کار رفته در شکل ۲-۱۶، می‌توان در آخرین نقطهٔ قطب مثبت مقیاس نمرهٔ ۳+، در آخرین نقطهٔ قطب منفی مقیاس نمرهٔ ۳-، در وسط مقیاس نمرهٔ صفر (۰)، و در بقیهٔ نقاط مقیاس نمره‌های بین ۳+ و ۳- را قرار داد. در شکل ۳-۱۶ شیوهٔ نمره‌گذاری این نوع مقیاس مشخص شده است.

آزگود و همکاران (۱۹۶۷) روش تفکیک معنایی را با تعداد زیادی از افراد به کار بستند و نتایج حاصل را با روش آماری تحلیل عاملی[1] تحلیل کردند. آنها در نتایج بررسی تحلیل عاملی سه عامل عمدهٔ ارزشیابی[2]، توانمندی[3]، و فعالیت[4] را به دست آوردند. عامل ارزشیابی بیشتر از بقیهٔ عوامل پراکندگی میان درجه‌بندیها را توجیه کرد. در نتیجه، آزگود و همکاران این عامل را با نگرش معادل دانستند. از این‌رو، مقیاسهای صفات دو قطبی که در بُعد ارزشیابی قرار دارند برای سنجش نگرش به کار می‌روند. بلوم، هستینگز، و مادوس (۱۹۷۱، ص ۲۴۳) نمونه‌های زیر را برای عامل ارزشیابی به دست داده‌اند: "خوب ـ بد"، "پاکیزه ـ کثیف"، "باارزش ـ بی‌ارزش"، "مفید ـ مضر" و "مطبوع ـ نامطبوع".

قواعد تهیهٔ پرسشنامهٔ تفکیک معنایی

آندرسون و بورک (۲۰۰۰) قواعد تهیه و اجرای پرسشنامه‌های مبتنی بر روش تفکیک معنایی را در ۵ مرحله توضیح داده‌اند. ما در زیر این مراحل را شرح می‌دهیم.

۱. جفتهای مناسب صفتهای دو قطبی را برگزینید. جفتهایی که انتخاب می‌کنید باید شامل صفتهایی باشند که در رابطه با مفهوم موردنظر به طور آشکار خاصیت ارزشیابانه داشته باشند. (مثالهایی از این‌گونه صفتها در شکلهای ۲-۱۶، ۳-۱۶، و ۴-۱۶ آمده‌اند).

۲. یک پاسخنامه شبیه به نمونهٔ زیر (و شکل ۳-۱۶) درست کنید:

با انصاف ـ: ـ: ـ: ـ: ـ: ـ: ـ: ـ: ـ بی‌انصاف

1. factor analysis 2. evaluation 3. potency 4. activity

همکلاسیهای من

خوب	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	بد
مهربان	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	کینه‌توز
درست‌کار	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	خلاف‌کار
باارزش	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	بی‌ارزش
آرام	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	عصبی
با ادب	ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ / ـ	بی‌ادب

شکل ۴-۱۶ نمونه پاسخنامه‌های روش تفکیک معنایی (آندرسون و بورک، ۲۰۰۰، ص ۱۹۹)

۳. مقیاس خود را با گروهی از افرادی که مقیاستان را به قصد سنجش ویژگیهای عاطفی آنان ساخته‌اید اجرا کنید. (دقت کنید که دست کم تعداد آزمون‌شوندگان شما باید ۱۰ برابر تعداد سؤالهایتان باشد ـ یعنی ۱۰ برابر تعداد صفتهای دوقطبی.)

۴. همبستگی بین هر جفت صفت دوقطبی را با نمرهٔ کل مقیاس حساب کنید.

۵. صفتهایی را که همبستگی معنی‌دار با کل مقیاس به دست نداده‌اند حذف کنید.

شکل ۴-۱۶ نمونهٔ دیگری از مقیاس تفکیک معنایی است. در این مقیاس نمونه، نظر دانش‌آموز نسبت به همکلاسیهایش سنجش می‌شود و در آن از صفات دوقطبی خوب ـ بد، مهربان ـ کینه‌توز، درست‌کار ـ خلاف‌کار، باارزش ـ بی‌ارزش، آرام ـ عصبی، و با ادب ـ بی‌ادب استفاده شده است.

مقیاس لیکرت

یکی دیگر از روشهای متداول سنجش عاطفی روشی است که به **مقیاس لیکرت**[1] شهرت یافته است. این مقیاس که عمدتاً برای سنجش نگرش به کار می‌رود اولین بار در سال ۱۹۳۲ به وسیلهٔ رنسیس لیکرت[2] ساخته شد و به همین منظور مورد استفاده قرار گرفت. این مقیاس از یک رشته بیان یا گویهٔ نگرشی دربارهٔ اشخاص، گروهها، یا اشیاء درست شده است. از پاسخ‌دهندگان خواسته می‌شود تا میزان یا درجهٔ موافقت یا مخالفت خود را با هر یک از آن بیانات مشخص کنند. براساس نمرهٔ کلی حاصل از مقیاس، معلوم می‌شود که نگرش فرد مثبت است یا منفی. شکل ۵-۱۶ یک مقیاس لیکرت بسیار مختصر را نشان می‌دهد. از این مقیاس می‌توان برای سنجش نگرش نسبت به آزمون و اندازه‌گیری استفاده کرد.

1. Likert Scale 2. Rensis Likert

راهنمایی: با گذاشتن علامت (×) در مقابل هر یک از عبارتهای زیر میزان موافقت یا عدم موافقت خود را با آن عبارت نشان دهید.

	کاملاً موافق	موافق	بی‌نظر	مخالف	کاملاً مخالف
۱. به سختی می‌توانم در جلسات این درس در کلاس بیدار بمانم.					
۲. این درس باید برای معلمان یک درس الزامی باشد.					
۳. دوست دارم یاد بگیرم سؤال چندگزینه‌ای بنویسم.					
۴. اکثر اوقات در کلاس حواسم جای دیگری است.					
۵. با علاقه وارد این کلاس می‌شوم.					

شکل ۵ـ۱۶ یک مقیاس لیکرت پنج ماده‌ای برای سنجش نگرش نسبت به آزمون و اندازه‌گیری (کوبیسزین و بوریچ، ۲۰۰۳، ص ۴۴۱)

نمره‌گذاری مقیاس لیکرت

پس از آنکه پاسخ‌دهندگان با گذاشتن علامت در فرم مربوط به مقیاس لیکرت به سؤالهای آن جواب دادند، به منظور نمره‌گذاری، باید به هر سؤال وزن خاصی داده شود. وزنهایی که به هر سؤال داده می‌شود وابسته به این است که به صورت مثبت نوشته شده باشد یا به صورت منفی. در مقیاس شکل ۵ـ۱۶ وزنهای گزینه‌های سؤال ۱ که به صورت منفی نوشته شده به صورت زیر است:

کاملاً موافق	۱
موافق	۲
بی‌نظر	۳
مخالف	۴
کاملاً مخالف	۵

وزنهای سؤال ۲ که به صورت مثبت نوشته شده درست به عکس سؤال ۱ و به صورت زیر است:

کاملاً موافق	۵
موافق	۴
بی‌نظر	۳
مخالف	۲
کاملاً مخالف	۱

نمرهٔ هر فرد برای هر سؤال برابر است با ارزش گزینهٔ انتخاب شده. برای مثال، اگر فردی برای سؤال ۱ (سؤال منفی) گزینهٔ "کاملاً موافق" را علامت زده باشد نمره‌اش ۱ است. اگر همان فرد برای سؤال ۲ (سؤال مثبت) گزینهٔ "کاملاً موافق" را علامت زده باشد نمره‌اش ۵ است.

راهنمایی: با گذاشتن علامت (x) در مقابل هر یک از عبارتهای زیر میزان موافقت یا عدم موافقت خود را با آن عبارت نشان دهید.

	کاملاً موافق	موافق	بی‌نظر	مخالف	کاملاً مخالف
۱. به سختی می‌توانم در جلسات این درس در کلاس بیدار بمانم.	۱	۲×	۳	۴	۵
۲. این درس باید برای معلمان یک درس الزامی باشد.	۵	۴	۳	۲×	۱
۳. دوست دارم یاد بگیرم سؤال چندگزینه‌ای بنویسم.	۵	۴	۳	۲	۱×
۴. اکثر اوقات در کلاس حواسم جای دیگری است.	۵	۲	۳	۴	۱×
۵. با علاقه وارد این کلاس می‌شوم.	۵	۴×	۳	۲	۱

شکل ۶ـ۱۶ نمونه‌ای از مقیاس لیکرت با وزنهای مربوط به هر یک از سؤالهای مثبت و منفی (کوبیسزین و بوریچ، ۲۰۰۳، ص ۴۴۲)

شکل ۶ـ۱۶ همان مقیاس شکل ۵ـ۱۶ را با وزنهای مختلفی که به سؤالهای مثبت و منفی تعلق می‌گیرد نشان می‌دهد.

برای تعیین نمرهٔ هر پاسخ‌دهنده، مجموع وزنهای انتخاب شدهٔ او را بر تعداد کل سؤالها تقسیم می‌کنیم. نتیجهٔ حاصل نمرهٔ میانگین نگرش او خواهد بود. نمرهٔ میانگین دانشجوی فرضی ما که به مقیاس شکل ۵ـ۱۶، به صورتی که در شکل ۶ـ۱۶ علامت زده، پاسخ داده است به صورت زیر محاسبه می‌شود:

$$۲ + ۲ + ۱ + ۱ + ۴ = ۱۰$$

$$۱۰ \div ۵ = ۲$$

حال باید تعیین کنیم که نگرش این دانشجو نسبت به آزمون و اندازه‌گیری مثبت است یا منفی. اگر نمرهٔ او کمتر از ۳ (نمرهٔ متوسط) باشد که در مورد دانشجوی فرضی ما چنین است، نگرش او منفی است. به طور کلی، پاسخ‌دهندگانی که در مقیاس لیکرت نمرهٔ بالای میانگین می‌گیرند نگرششان مثبت است و کسانی که نمرهٔ پایین‌تر از میانگین می‌گیرند نگرششان منفی است.

تعداد گزینه‌های مقیاس لیکرت

معمولاً سؤالهای مقیاس لیکرت با پنج گزینه یا پنج طبقه پاسخ با عنوانهایی مانند کاملاً موافق، موافق، بی‌نظر، مخالف، و کاملاً مخالف دنبال می‌شوند که به ترتیب نمرهٔ از ۵ تا ۱ به آنها داده می‌شود. آندرسون و بورک (۲۰۰۰) در کتاب معروف خود با عنوان سنجش ویژگیهای عاطفی در آموزشگاه، به جای تعداد گزینه‌های فرد مثلاً ۳ یا ۵، تعداد گزینه‌های زوج ۴ یا ۶ را پیشنهاد کرده‌اند. بنابراین، آنها به جای پنج طبقه پاسخ فوق، طبقهٔ خنثای وسط را حذف می‌کنند و چهار طبقهٔ کاملاً موافق، موافق، مخالف، و کاملاً مخالف را که از ۴ تا ۱ نمره به آنها تعلق می‌گیرد مورد استفاده قرار

می‌دهند. آنان به استناد پژوهشهای انجام گرفته دلایل خود را به نحو زیر بیان داشته‌اند:

داشتن چهار یا شش طبقهٔ پاسخ بهتر از سه یا پنج طبقه است. دلیلش این است که طبقهٔ وسط که معمولاً به صورت بی‌نظرم، مطمئن نیستم، یا نمی‌دانم مشخص می‌شود تنها به وسیلهٔ کسانی که واقعاً نظری ندارند، مطمئن نیستند، یا نمی‌دانند انتخاب نمی‌شود، بلکه کسانی هم که عقیدهٔ مشخصی دارند ولی نمی‌خواهند آن را ابراز کنند انتخاب می‌شود. همچنین کسانی که دربارهٔ سؤال نظری دارند اما نظرشان خیلی قطعی نیست و اگر طبقهٔ خنثای وسط نباشد یکی از گزینه‌های دیگر را انتخاب خواهند کرد جذب گزینهٔ وسطی می‌شوند. لذا این گزینه از تمایز گذاری دقیق می‌کاهد و، از طریق کاستن از پراکندگی، پایایی یا دقت اندازه گیری را کاهش می‌دهد. (آندرسون و بورک، ۲۰۰۰، ص ۹۴)

اما برای حل مشکل کسانی که واقعاً در مورد سؤال بی‌نظر، نامطمئن یا بی‌اطلاع‌اند، آندرسون و بورک (۲۰۰۰) دو راه را پیشنهاد می‌کنند. یکی از آن دو راه این است که در کنار پرسشنامه و دور از گزینه‌ها جایی با عنوان نمی‌توانم تصمیم بگیرم یا نظری ندارم تعیین کنید تا پاسخ‌دهندگان نامطمئن دربارهٔ سؤال آن را علامت بزنند. راه دوم این است که از پاسخ‌دهندگان بخواهید وقتی به سؤالی می‌رسند که دربارهٔ جواب دادن به آن نامطمئن‌اند آن را بی‌جواب بگذارند.

شکلهای مختلف مقیاس لیکرت

تاکنون، ما در بحث دربارهٔ متداول‌ترین شکل مقیاس لیکرت گفتیم که آن شامل تعدادی سؤال است و هر سؤال از چند طبقه پاسخ یا چند گزینه تشکیل شده که دربرگیرندهٔ چهارطبقهٔ از ۴ تا ۱ نمره‌ای یا پنج طبقهٔ از ۵ تا ۱ نمره‌ای است. افزون بر این شکل، مقیاس لیکرت می‌تواند شکلهای دیگری نیز به خود بگیرد که در این قسمت معروف‌ترین آنها را معرفی می‌کنیم.

✓ **استفاده از تنهٔ مشترک برای سؤالها** این نوع مقیاس لیکرت شبیه به آزمونهای چندگزینه‌ای است که دارای یک تنه یا ساقهٔ مشترک برای همهٔ گزینه‌هاست. این شکل از مقیاس در وقت پاسخ دهنده برای خواندن سؤالها صرفه‌جویی به عمل می‌آورد. به عنوان نمونه، یک مقیاس ۳۰ سؤالی دربارهٔ کیفیت مدرسه، به جای ۳۰ سؤال جداگانه، می‌تواند از یک سؤال که در آغاز پرسشنامه می‌آید و به دنبال آن پاسخها قرار داده می‌شوند تشکیل شود. شکل ۷-۱۶ نمونهٔ ساده‌ای از این نوع مقیاس را نشان می‌دهد.

مدرسهٔ من جایی است که:	کاملاً موافقم	موافقم	مخالفم	کاملاً مخالفم
۱. واقعاً دوست دارم هر روز به آنجا بروم.	□	□	□	□
۲. دانش‌آموزان آن خیلی صمیمی‌اند.	□	□	□	□
۳. معلمانش مهربان‌اند.	□	□	□	□

شکل ۷-۱۶ نمونه‌هایی از مقیاس لیکرت با تنهٔ مشترک سؤالها (آندرسون و بورک، ۲۰۰۰، ص ۹۶)

استفاده از طبقه‌های یکسان پاسخها به خلاف گزینه‌های دوقطبی متداول، در این شکل از مقیاس لیکرت برای گزینه‌ها از ابعاد یک قطبی مانند اهمیت یا فراوانی استفاده می‌شود. به عنوان نمونه، می‌توان از دانش‌آموزان دربارهٔ یکی از ویژگیهای عاطفی پرسش به عمل آورد و از آنان خواست تا بر حسب اهمیت به آن سؤال به صورت بسیار با اهمیت، با اهمیت، کم اهمیت، بسیار کم‌اهمیت جواب دهند. به همین قیاس، می‌توان از دانش‌آموزان سؤال کرد که به چه میزانی (فراوانی) در یک فعالیت شرکت می‌کنند و از آنان خواست تا به صورت همیشه، تقریباً همیشه، گاهی، به ندرت، و هرگز جواب دهند. شکل ۸-۱۶ نمونه‌ای از این گونه مقیاسها را نشان می‌دهد. در این پرسشنامه تنش روانی[1] دانش‌آموزان دبیرستانی سنجیده می‌شود و در آن هم از فراوانی تجارب تنش‌زا و هم از شدت تنش به عنوان طبقه‌های پاسخ استفاده شده است.

پرسشنامهٔ تنش روانی دانش‌آموزان دبیرستانی

راهنمایی: ۶۰ سؤال یا بیان زیر از میان پاسخهای دانش‌آموزان دبیرستانی به این سؤال انتخاب شده‌اند: "تنش روانی چه تأثیراتی روی شما دارد؟" لطفاً سؤالها را به دقت بخوانید و به نحو زیر به آنها پاسخ دهید:

اگر در یکی دو ماه گذشته هریک از موارد ذکر شده را تجربه کرده‌اید، لطفاً مشخص کنید که: (۱) به چه میزانی آن مورد رخ داده (فراوانی)، و (۲) نیرومندی آن چه قدر بوده (شدت) است. پاسخ خود را با کشیدن دایره به دور عدد موردنظرتان، به گونه‌ای که در زیر تعریف شده است، مشخص کنید.

فراوانی	شدت
۰ـ هرگز	۰ـ بی‌تأثیر
۱ـ کمتر از یک بار در هفته	۱ـ با شدت کم
۲ـ یک یا دوبار در هفته	۲ـ با شدت متوسط
۳ـ اکثر روزها	۳ـ با شدت زیاد
۴ـ دست‌کم روزی یک بار	۴ـ تقریباً با شدت غیرقابل تحمل

سؤال نمونه

احساس دلشوره

فراوانی	شدت
۰ ۱ ۲ ۳ ۴	۰ ۱ ۲ ۳ ۴

شکل ۸-۱۶ راهنمای پرسشنامهٔ تنش روانی ویژهٔ دانش‌آموزان دبیرستانی (آندرسون، و بورک، ۲۰۰۰، ص ۹۷)

1. stress

استفاده از بیانات زوج در این شکل از مقیاس لیکرت برای هر مادهٔ پرسشنامه دو بیان یا گفتهٔ متضاد به کار می‌روند و از پاسخ‌دهنده خواسته می‌شود تا یکی از آنها را انتخاب کند. به نمونهٔ زیر توجه کنید.

من ریاضیات را دوست دارم....
(کاملاً موافق، موافق / موافق، کاملاً موافق)

... من از ریاضیات متنفرم

امتیاز مهم این شکل از مقیاس لیکرت آن است که پاسخ دهنده را در موقعیتی قرار می‌دهد که حتماً یکی از گزینه‌ها را انتخاب کند. در مقابل، این نوع مقیاس دارای اشکالهایی نیز هست. از جمله اینکه خواندن سؤالها وقت زیادی می‌گیرد. مشکل دیگر این است که گرچه دوست داشتن و تنفر داشتن به طور آشکار ضد یکدیگرند، اما بعضی وقتها یافتن زوج کلماتی که به این روشنی ضد یکدیگر باشند دشوار است.

استفاده از پرسشنامه‌های چندمقیاسی هریک از پرسشنامه‌های سنجش عاطفی معمولاً از تعدادی مقیاس تشکیل می‌شوند. منظور از **مقیاس**[1] در اینجا مجموعه‌ای سؤال به هم وابسته است که بر روی هم یک ترکیب عاطفه ـ هدف را می‌سنجند (آندرسون و بورک، ۲۰۰۰). به عنوان نمونه، سه سؤال شکل ۷-۱۶ یک مقیاس را تشکیل می‌دهند که به وسیلهٔ آن یک عاطفه (نگرش) و یک هدف (مدرسه) سنجیده می‌شود. لازم به ذکر است که اصطلاح "مقیاس" به معناهای دیگری نیز به کار می‌رود. از جمله اینکه، در رابطه با پرسشنامهٔ لیکرت، مقیاس به تعداد گزینه‌ها یا طبقه‌های پاسخ گفته می‌شود. از آنجا که آندرسون و بورک (۲۰۰۰) در کتاب تخصصی خود در این زمینه اصطلاح مقیاس را برای مجموعه‌ای سؤال و طبقه را برای گزینه‌های سؤال به کار برده‌اند، ما نیز در اینجا همین تمایز را رعایت کرده‌ایم.

دلیل اینکه چرا بهتر است در پرسشنامه‌های سنجش عاطفی، به جای یک مقیاس، از چند مقیاس استفاده شود این است که به ندرت یک ویژگی عاطفی واحد یا یک هدف عاطفی واحد اطلاعات لازم را دربارهٔ عاطفهٔ یک شخص به دست می‌دهد. نگرش نسبت به مدرسه را که در شکل ۷-۱۶ مثال آوردیم در نظر بگیرید. مسلماً همهٔ دانش‌آموزان نسبت به مدرسه یک عاطفهٔ واحد و مشخص ندارند، بلکه از جنبه‌های مختلف به آن می‌نگرند و در رابطه با آن جنبه‌ها عواطف مختلفی دارند. بنابراین، اگر ما به کمک مقیاسهای مختلف که هریک از چند سؤال تشکیل یافته‌اند ویژگیهای دانش‌آموزان را نسبت به مدرسه بسنجیم تصویر بهتری از نگرش آنان دربارهٔ این موضوع به دست می‌آوریم.

1. scale

در تهیهٔ پرسشنامه‌های چندمقیاسی، دقت کنید که سؤالهای هر مقیاس را در یک گروه و به دنبال هم نیاورید، بلکه آنها را با یکدیگر مخلوط سازید. این کار از اینکه پاسخ دهندگان، با جواب دادن به یک سؤال از هر مقیاس، بقیه را هم بدون دقت کافی و متناسب با همان سؤال اول جواب دهند جلوگیری به عمل می‌آورد.

قواعد تهیهٔ پرسشنامه‌های عاطفی مبتنی بر مقیاس لیکرت

هرچند که قواعد مربوط به تهیه و اجرای پرسشنامه‌های سنجش عاطفی که در صفحات قبلی این فصل توضیح دادیم دربارهٔ مقیاس لیکرت نیز قابل استفاده است، با این حال صاحب‌نظران سنجش در رابطه با مقیاس لیکرت پیشنهادهایی اختصاصی مطرح کرده‌اند که ما در اینجا خلاصهٔ آنها را نیز توضیح می‌دهیم.

۱. ابتدا ویژگی عاطفی موردنظر خود را انتخاب کنید و آن را به دقت تعریف نمایید. برای متغیرهای روانی عموماً دو نوع تعریف وجود دارد: **مفهومی**[1] و **عملیاتی**[2]. تعریف مفهومی به معنی انتزاعی یک متغیر مربوط است؛ تعریف عملیاتی به مجموعه‌ای رفتار (مانند بیان کلمات و انجام اعمال) اشاره می‌کند که می‌توان بر اساس آنها به استنباط دربارهٔ متغیر موردنظر پرداخت. تعریفهای مفهومی بیشتر در تدوین نظریه‌های روانشناختی و بیان روابط انتزاعی میان متغیرها مورد استفاده قرار می‌گیرند، ولی برای مقاصد سنجش و اندازه‌گیری، متغیرها باید به صورت عملیاتی تعریف شوند.

۲. تعدادی بیان، جمله، یا سؤال در رابطه با ویژگی عاطفی مورد سنجش انتخاب کنید یا بنویسید به طوری که آشکارا مثبت یا منفی باشند. (سعی کنید تعداد بیانات مثبت و منفی تقریباً هم‌اندازه باشند.)

۳. از چند نفر داور بخواهید تا سؤالهای شما را وارسی کنند و دربارهٔ مثبت یا منفی بودن آنها نظر بدهند.

۴. هریک از سؤالها را که توسط همهٔ داوران مثبت یا منفی ارزیابی نشده کنار بگذارید. سؤالها یا جمله‌های خنثی برای گنجانیدن در مقیاس لیکرت مفید نیستند.

۵. دربارهٔ تعداد گزینه‌ها یا طبقه‌های پاسخ و عنوان آنها تصمیم بگیرید. چنان که پیشتر گفتیم، مقیاس اولیهٔ لیکرت شامل ۵ گزینهٔ زیر بود: کاملاً موافق، موافق، بی‌نظر، مخالف، کاملاً مخالف. می‌توان این تعداد را کم و زیاد کرد. معمولاً با دانش‌آموزان کم سن‌تر از تعداد گزینه‌های کمتر و عنوانهای ساده‌تری استفاده می‌شود. آندرسون و بورک (۲۰۰۰)، با توجه به دلایلی که قبلاً توضیح دادیم، چهار گزینهٔ کاملاً موافق، موافق، مخالف، کاملاً مخالف را ترجیح می‌دهند.

---

1. conceptual 2. operational

۶. با مخلوط کردن سؤالها، فرم مقدماتی پرسشنامه را آماده کنید و دربارهٔ چگونگی پاسخ دادن به سؤالها راهنمایی لازم را بنویسید.

۷. فرم مقدماتی پرسشنامه را با تعدادی از افرادی که شبیه به کسانی هستند که پرسشنامه را برایشان تهیه می‌کنید اجرا نمایید تا مشکلات احتمالی کشف و برطرف شوند. دقت کنید که برای هر سؤال باید دست کم ۱۰ آزمون شونده داشته باشید. این قاعده را هم در رابطه با اجرای مقدماتی و هم اجرای اصلی پرسشنامه رعایت کنید. یعنی اگر پرسشنامهٔ شما ۴۵ سؤال دارد، آزمودنیهای شما باید دست کم ۴۵۰ نفر باشند (آندرسون و بورک، ۲۰۰۰).

۸. جوابها را نمره گذاری کنید و همبستگی بین پاسخهای هر سؤال و کل پرسشنامه را تعیین نمایید.

۹. در سؤالهایی که همبستگی بین آنها و نمرهٔ کل آزمون معنی‌دار نیست تجدیدنظر کنید.

استفاده از مصاحبه برای سنجش عواطف

روش مصاحبه[1]، مانند روش استفاده از پرسشنامه، جزو روشهای گزارش شخصی به حساب می‌آید. هر چند که این دو روش از جهاتی با هم فرق دارند، اما اساساً به هم بسیار شبیه‌اند. بنا به گفتهٔ آندرسون و بورک (۲۰۰۰)،

> در هر دو روش، از شخص خواسته می‌شود تا به تعدادی جمله یا سؤال پاسخ دهد؛ و این کار را یا از راه علامت گذاشتن در یک مربع یا دایره (در پرسشنامه‌های بسته پاسخ)، یا از راه نوشتن یک توضیح مختصر (در پرسشنامه‌های باز پاسخ)، و یا به صورت پاسخ شفاهی (در مصاحبه) انجام می‌دهد. هم پرسشنامه و هم مصاحبه می‌توانند کم و بیش ساختارمند باشند.... اما به‌طور کلی، مصاحبه بازتر از پرسشنامه است.... با این حال، اصولاً هیچ چیز نمی‌تواند سبب شود که پرسشنامه‌ها بازتر و مصاحبه‌ها کاملاً بسته باشند. به عنوان نمونه، مصاحبه‌های تلفنی اغلب بسته‌اند.... به مصاحبه‌های بسته مصاحبه‌های استاندارد شده گفته می‌شود. (صص ۶۰-۵۹)

در تعریف مصاحبه گفته شده که "مصاحبه یک ابزار سنجش است که به کمک آن اطلاعات مورد نیاز از راه ارتباط مستقیم و دو جانبه گردآوری می‌شود" (کوهن و سوردلیک، ۲۰۰۵، ص ۴۲۳). هر چند که مصاحبه عمدتاً به صورت کلامی صورت می‌پذیرد، اما در مصاحبه‌های رودررو، علاوه بر کلام، ارتباط غیرکلامی نیز تأثیرگذار است. به عنوان مثال، مصاحبه‌گر می‌تواند، در ضمن مصاحبه، از طرز لباس پوشیدن، لحن صدا، ادب و احترام، و واکنشهای مصاحبه‌شونده در مقابل سؤالها و اظهارات مصاحبه کننده سنجش به عمل آورد.

---

1. interview

اگرچه ما در اینجا مصاحبه را به عنوان روشی برای سنجش ویژگیهای عاطفی به حساب آورده‌ایم، اما این روش کاربردهای گسترده‌تری دارد. از جمله اینکه "داده‌های به دست آمده از روش مصاحبه می‌تواند اطلاعات تکمیلی بسیار مفیدی برای نتایج آزمونها باشد." (کاپلان[1] و ساکوزو[2]، ۲۰۰۱، ص ۱۲). بنا به گفتهٔ کوهن و سوردلیک (۲۰۰۵)،

از مصاحبه در موقعیتهای بالینی، مشاوره‌ای، پزشکی قانونی، و روانی - عصبی به منظور کمک به تشخیص و درمان استفاده می‌شود. روانشناسان آموزشگاهی و دیگر متخصصان در آموزش و پرورش از مصاحبه برای تصمیم‌گیری دربارهٔ مناسب بودن روشهای مختلف آموزشی یا گروهبندی کلاسی استفاده می‌کنند. در حوزهٔ منابع انسانی، مصاحبه به روانشناسان کمک می‌کند تا پیشنهادهای مؤثرتری در رابطه با استخدام، اخراج، و ترفیع کارکنان ارائه دهند.... همچنین پژوهشگران در زمینه‌های روانشناسی، آموزش و پرورش، و سایر زمینه‌های وابسته از مصاحبه برای شناسایی متغیرهای مختلف سود می‌برند. (ص ۸)

بنا به آنچه گفته شد، مصاحبه یکی از روشهای مفید تشخیص و سنجش ویژگیهای مختلف افراد، از جمله ویژگیهای عاطفی، است. صاحب‌نظران در ارتباط با استفادهٔ بهینه از روش مصاحبه پیشنهادهایی ارائه داده‌اند که ما در دنبالهٔ این مطلب آنها را توضیح می‌دهیم.

قواعد انجام روش مصاحبه

از آنجاکه روش مصاحبه، علاوه بر سنجش ویژگیهای عاطفی، برای مقاصد مهم دیگری نیز به کار می‌رود، ما در اینجا قواعدی راکه توضیح می‌دهیم برای استفاده از این روش در موقعیتهای مختلف، از جمله سنجش ویژگیهای عاطفی دانش‌آموزان، است.

۱. در شروع مصاحبه سعی کنید بین خود و مصاحبه‌شونده (دانش‌آموز) یک جو اعتماد ایجاد نمایید و کاری کنید که دانش‌آموز احساس راحتی کند تا اینکه اطلاعات لازم را دربارهٔ آنچه مورد نظر شماست در اختیارتان بگذارد. بدین منظور، در آغاز مصاحبه خودتان سر صحبت را باز کنید و سؤالهای ساده‌ای دربارهٔ وضع شخصی او بپرسید و نوعی معارفه بین خود و او انجام دهید.

۲. به گفتار و حرکات مصاحبه‌شونده کاملاً توجه کنید. هم گفتار و هم اعمال و رفتار مصاحبه‌شونده منابع اطلاعاتی مهمی برای شناسایی و سنجش ویژگیهای عاطفی و شخصیتی او هستند. به اینها به دقت توجه نمایید و مصاحبه‌شونده را به ادامهٔ صحبت تشویق کنید. خودتان زیاد صحبت نکنید.

۳. دربارهٔ عقاید و باورهای مصاحبه‌شونده داوری نکنید. واکنشهای شما نباید بر ارزشها و

1. Kaplan 2. Saccuzzo

اعتقادات مصاحبه‌شونده تأثیر بگذارد. دربارهٔ گفته‌های او به داوری نپردازید. اگر او بفهمد که شما از اظهاراتش خوشتان نمی‌آید، ادامه سخن نخواهد داد. وقتی که از کلمه‌هایی مانند خوب، بد، عالی، ضعیف، تأسف‌بار، امیدوارکننده، و مانند اینها استفاده می‌کنید به ارزش داوری می‌پردازید. "بجز زمانی که هدف شما از مصاحبه سنجش واکنش مصاحبه‌شونده به ارزشیابی از اوست، از اظهارات مبتنی بر ارزشیابی خودداری کنید" (کاپلان و ساکوزو، ۲۰۰۱، ص ۲۳۰)

۴. به مصاحبه‌شونده اطمینان بدهید که اطلاعاتی که او در اختیارتان می‌گذارد محرمانه خواهند ماند. اگر دانش‌آموز به شما اطمینان نکند و احساس نماید که اطلاعاتی را که در اختیار شما می‌گذارد در جای دیگر بازگو خواهد شد اطلاعات لازم را به شما نخواهد داد.

۵. هم از سؤالهای باز پاسخ استفاده کنید و هم از سؤالهای بسته پاسخ. بنا به نوع اطلاعاتی که می‌خواهید از مصاحبه‌شونده دریافت نمایید، شکل سؤالهای خود را تغییر دهید. سؤالهای بازپاسخ شبیه به سؤالهای زیر هستند: • *قدری دربارهٔ خودت صحبت کن*، • *توضیح بده که چه چیزهایی را بیشتر دوست داری* و • *دلت می‌خواهد دربارهٔ چه چیزی برای من حرف بزنی؟* نمونه‌های سؤالهای بسته پاسخ عبارت‌اند از: • *مدرسه را دوست داری؟* • *چند سال داری؟* • *در کلاس چندم هستی؟*

در آغاز مصاحبه از سؤالهای باز پاسخ شروع کنید تا مصاحبه‌شونده به صحبت کردن دربارهٔ خودش در جلسهٔ مصاحبه عادت کند.

۶. در واکنش به گفته‌های مصاحبه‌شونده، از بیاناتی که نشان دهد مشکل او را درک می‌کنید و او را مورد تأیید قرار می‌دهید استفاده کنید. کاپلان و ساکوزو (۲۰۰۱)، با استناد به پژوهشهای انجام گرفته دربارهٔ تأثیر این گونه اظهارات مصاحبه کننده در پاسخ به گفته‌های مصاحبه‌شونده، گفته‌اند "زمانی‌که به دیگران نشان می‌دهید که آنها را درک می‌کنید، دربارهٔ مشکل خود بیشتر صحبت می‌کنند و خود را عمیق تر می‌کاوند" (ص ۲۳۵). به نمونهٔ زیر از یک مصاحبه بین روانشناس مدرسه و یک دانش‌آموز توجه کنید:

روانشناس:	*بیل، امروز چه اتفاقی افتاده است؟ (سؤال باز پاسخ)*
دانش‌آموز (بیل):	*معلم فیزیک در حضور کلاس به سرم داد زد.*
روانشناس:	*ناراحت‌کننده است (نشان دادن اینکه احساس او را درک می‌کنید.)*
بیل:	*نه تنها آن، او همیشه به من گیر می‌دهد.*
روانشناس:	*حتماً این کار تو را عصبانی می‌کند (درک کردن مصاحبه‌شونده).*
بیل:	*آره. همیشه یک اشکالی تو کارهای من پیدا می‌کند. هر کاری بکنم، او راضی نمی‌شود.*
روانشناس:	*واقعاً ناراحت‌کننده است، بیل. مثل اینکه هیچ جوری نمی‌توانی او را راضی کنی (درک کردن مصاحبه‌شونده).*

بیل: در امتحانی که چند روز پیش از ما گرفت، من نمرهٔ ردّی (ه) گرفتم. بعد از امتحان، جواب سؤالهایم را با جواب سؤالهای دوستم تام مقایسه کردم، و هر دو مثل هم بودند. آن وقت به من داده بود (ه) ولی به تام داده بود (ب). عادلانه نیست (روشن ساختن و درک کردن).

روانشناس: البته که عادلانه نیست. وقتی که سعی کردم با او دربارهٔ این موضوع صحبت کنم، حاضر نشد به حرفهایم گوش بدهد.

بیل: خیلی ناراحت‌کننده است (درک کردن).

روانشناس: واقعاً همین‌طور است. اگر یک نمرهٔ ردّی (ه) دیگر بگیرم از مدرسه اخراج می‌شوم.

بیل: این واقعاً یک مشکل جدی است (روشن ساختن).

روانشناس: آره، اگر از مدرسه اخراج بشوم، روی دیدن پدر و مادر و دوستانم را نخواهم داشت.

بیل: و اینها واقعاً شما را ناراحت کرده است (درک کردن).

روانشناس: (کاپلان و ساکوزو، ۲۰۰۱، صص ۲۳۵ و ۲۳۴)

دقت کنید که اظهارات روانشناس در مصاحبهٔ بالا با دانش‌آموز به منظور تأیید نگرانیهای بجای او صورت می‌گیرد و به او نشان می‌دهد که مشکلاتش را درک می‌کند. این اظهارات با آنچه در بند ۳ بالا در رابطه با عدم قضاوت دربارهٔ عقاید مصاحبه‌شونده گفتیم فرق دارد. در آنجا گفتیم که عقاید و باورهای مصاحبه‌شونده را به چالش نگیرید و دربارهٔ آنها داوری ارزشی نکنید، زیرا عقیدهٔ هر کس برای خودش محترم است. اما در اینجا می‌گوییم که با احساسهای منطقی و نگرانیهای مصاحبه‌شونده همدلی کنید و نشان دهید که او را درک می‌نمایید.

گروه‌سنجی

یکی دیگر از روشهای سنجش عاطفی **گروه‌سنجی**[1] است. در گروه‌سنجی روابط عاطفی میان اعضای گروهها تعیین می‌شود. کوبیسزین و بوریچ (۲۰۰۳) گروه‌سنجی را به عنوان "اندازه‌گیری الگوهای تعامل در گروهها" (ص ۴۳۹) تعریف کرده‌اند. لین و گرانلاند (۲۰۰۰) دربارهٔ این روش گفته‌اند "با گروه‌سنجی می‌توان پذیرش اجتماعی فردفرد دانش‌آموزان و ساختار اجتماعی گروه را سنجش کرد" (ص ۳۲۶).

برای سنجش روابط اجتماعی میان اعضای گروهها تعدادی ابزار گروه‌سنجی وجود دارد. ما در اینجا دو مورد معروف آنها را توضیح می‌دهیم.

نگارهٔ گروه[2] نگارهٔ گروه یک بازنمایی نموداری از روابط اجتماعی درون گروه یا ساختار

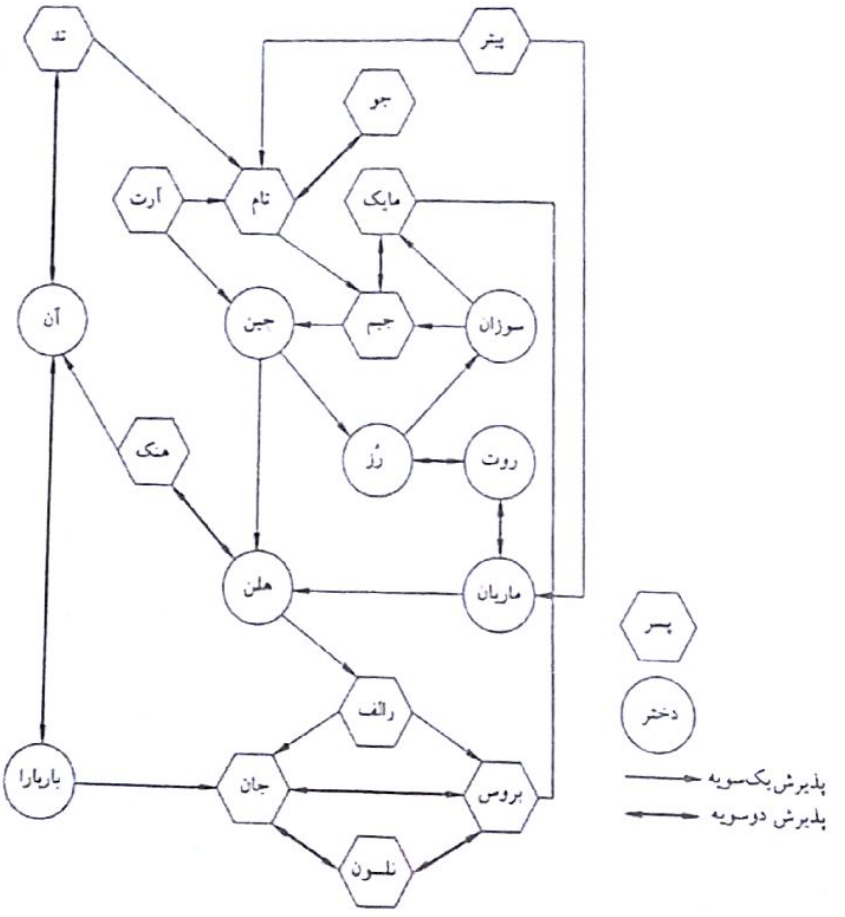
1. sociometry 2. sociogram

گروه است. با بررسی نگارهٔ گروه، می‌توان در مواردی چون رهبران گروه، افراد منزوی، دوستداران یکدیگر، دار و دسته‌های درون گروه، و مانند اینها را شناسایی کرد.

برای تهیهٔ نگارهٔ گروه لازم است گامهایی برداشته شوند. ابتدا از هریک از دانش‌آموزان بخواهید تا به طور محرمانه اطلاعات زیر را در اختیار شما بگذارند.

- کسانی را که دوست دارید در یک گزارش شفاهی با آنها همکاری کنید نام ببرید.
- اگر اجازه داشته باشید، دوست دارید کنار چه کسانی در کلاس بنشینید؟
- دوست دارید در یک کمیتهٔ برنامه‌ریزی کلاسی با چه کسانی کار کنید؟
- با چه کسانی دوست ندارید کار کنید؟

بعد از جمع‌آوری اطلاعات، انتخابهای هریک از افراد را مشخص کنید و نگاره یا نمودار آن را رسم نمایید. سمت انتخاب افراد را با یک پیکان یک‌سویه نشان دهید. برای نشان دادن انتخابهای دوجانبه از پیکانهای دوسویه استفاده کنید. برای نشان دادن دوست نداشتن یا بی‌علاقگی خط نقطه‌چین رسم کنید. شکل ۹-۱۶ یک نمونه نگارهٔ گروه است که نگرش دانش‌آموزان یک کلاس را نسبت به یکدیگر نشان می‌دهد.



شکل ۹-۱۶ یک نمونه نگارهٔ گروه (کالاهان[1] و کلارک[2]، ۱۹۹۰، ص ۱۳۴)

1. Callahan
2. Clark

در ارتباط با این سؤال که "آیا واداشتن دانش‌آموزان به نام بردن از کسانی که مورد علاقه‌شان نیست کاری درست است یا نه؟" بین متخصصان روانسنجی اختلاف‌نظر وجود دارد. بعضیها مخالف آن هستند و می‌گویند این کار دانش‌آموزان را نسبت به احساس بی‌علاقگی و نفرتشان از بعضی دانش‌آموزان آگاه‌تر می‌سازد و همین امر موجب تضعیف روحیهٔ گروهی و کاستن از رشد هیجانی دانش‌آموزان می‌شود. لین و گرانلاند (۲۰۰۰) برای رفع اختلاف‌نظر راه میانه‌ای را برگزیده و می‌گویند،

> بهترین روش این است که از انتخابهای منفی جلوگیری شود مگر اینکه برای منظوری که گروه‌سنجی به کار خواهد رفت ضروری باشد. اگر چنین است، باید به دانش‌آموزان اجازهٔ انتخابهای منفی داده شود نه اینکه به انجام این کار مجبور شوند. چیزی شبیه به جملهٔ زیر قاعدتاً باید کفایت کند: "اگر دانش‌آموزانی وجود دارند که دوست ندارید با آنان هم‌گروه باشید، می‌توانید نام آنها را نیز بنویسید." (ص ۳۲۸)

و اما در این‌باره که از دانش‌آموزان خواسته شود تا چند نفر از همکلاسیهای خود را نام ببرند، صاحب‌نظران و متخصصان روانسنجی ۲ تا ۵ انتخاب را پیشنهاد داده‌اند. در اکثر موارد، پنج انتخاب برای هر فعالیت به عنوان یک رقم مناسب پیشنهاد شده است. لین و گرانلاند (۲۰۰۰) می‌گویند مطالعات انجام گرفته دربارهٔ روشهای گروه‌سنجی نشان داده‌اند که پنج انتخاب بیشترین پایایی را به دست می‌دهد و افزودن بر تعداد انتخابها پایایی روش را بالا نمی‌برد. البته در ارتباط با دانش‌آموزان سالهای نخست دبستان ۲ تا ۳ انتخاب مناسب است.

روش حدس بزن چه کسی

روش ساده‌تری که آن نیز برای سنجش روابط عاطفی میان اعضای یک گروه یا یک کلاس درس به کار می‌رود روش **حدس بزن چه کسی**[1] نام دارد. از این روش برای یافتن سرنخ‌هایی از عادتها، علاقه‌ها، دوستیها، سرگرمیها، و مشکلات دانش‌آموزان استفاده می‌شود.

در این روش به دانش‌آموزان تعدادی جملهٔ توصیف‌کنندهٔ رفتار داده می‌شود و از آنان خواسته می‌شود تا دانش‌آموزانی را که دارای آن رفتارها یا ویژگیها هستند نام ببرند. نمونه‌های این‌گونه جمله‌ها در زیر آمده‌اند.

- این شخص خواندن را دوست دارد.
- این شخص همیشه نگران است.
- این شخص دوستان زیادی دارد.

(کالاهان و کلارک، ۱۹۹۰، ص ۱۳۴)

---

1. guess-who

تصمیم‌گیری در این‌باره که، افزون بر جمله‌های مثبت (مانند، این شخص مهربان است)، جمله‌های منفی (مانند، این شخص کینه‌توز است) نیز داده شود یا نه تصمیمی است به عهدهٔ خود معلمان، زیرا آنها می‌دانند که چنین کاری تأثیر منفی بر جو کلاسی دارد یا نه. "زمانی که بین دانش‌آموزان و بین دانش‌آموزان و معلم روابط خوبی برقرار باشد، استفاده از جملات منفی مشکلی ایجاد نمی‌کند. اما اگر گمان برود که این کار مشکلاتی به بار می‌آورد، بهتر است از آن صرف‌نظر کنید تا اینکه روحیهٔ کلاس خراب نشود" (لین و گرانلاند، ۲۰۰۰، ص ۳۲۴).

در روش حدس بزن چه کسی، پاسخ‌دهندگان آزادند تا در پاسخ به هریک از جمله‌ها هر تعداد از افراد موردنظر خود را که می‌خواهند نام ببرند. در نمره‌گذاری، صرفاً تعداد کسانی که از یک شخص نام برده‌اند شمارش می‌شوند و همان نمرهٔ آن شخص در ویژگی موردنظر به حساب می‌آید. اگر، علاوه بر جمله‌های مثبت، جمله‌های منفی نیز به کار رفته باشند، تعداد انتخابهای منفی فرد باید از تعداد انتخابهای مثبت او کم شود. به عنوان نمونه اگر دانش‌آموزی در ۸ مورد دوستانه و در ۳ مورد کینه‌توزانه ذکر شده است، نمرهٔ نهایی او ۵ = ۳-۸ خواهد بود.

استفاده از یادداشتهای روزانه برای سنجش عواطف

روش دیگری که برای سنجش ویژگیهای عاطفی دانش‌آموزان پیشنهاد شده **یادداشتهای روزانه**[1] است. در این روش، معلم از دانش‌آموز می‌خواهد تا احساسهای خود را دربارهٔ تجربه‌های شخصی و فعالیتهای کلاسی به صورت یادداشتهایی ثبت کند و آنها را با معلم در میان بگذارد. از این راه، معلم می‌تواند با احساسها و عواطف دانش آموزان آشنا شود و آنها را یاری دهد. باتلر و مکمان (۲۰۰۶) گفته‌اند "روش یادداشتهای روزانه می‌تواند جریان پیشرفت امور را نشان دهد. در این روش، دانش‌آموزان می‌توانند اندیشه‌ها، مشاهده‌ها، و سؤالهای خود را ثبت کنند، و به یک گفتگوی نوشتاری با دیگران بپردازند" (ص ۶۴).

از روش یادداشتهای روزانه برای شناسایی روشها و راهبردها (استراتژیها)ی یادگیری دانش‌آموزان نیز می‌توان استفاده کرد. بدین منظور، می‌توان از دانش‌آموزان خواست تا، ضمن مطالعه و یادگیری مطالب مختلف درسی خود، روشها و فنونی را که به کار می‌گیرند یادداشت نمایند، و بعد با معلم خود دربارهٔ آنها گفتگو کنند. معلم از این راه می‌تواند دانش‌آموزان را در کاربرد راهبردهای درست یاری دهد.

در استفادهٔ بهینه از روش یادداشتهای روزانه، دمبو (۱۹۹۴) به معلمان سفارش می‌کند که به دانش‌آموزان خود اطمینان بدهید که یادداشتهای خصوصی آنان را در رابطه با تجارب و احساسهای شخصی‌شان به صورت محرمانه نگه خواهید داشت و با کس دیگری در میان نخواهید گذاشت. در چنین شرایطی، دانش‌آموزان با اطمینان خاطر نگرش و احساسهای خود را نسبت به معلم و مدرسه در اختیار معلم خواهند گذاشت.

---

1. journals

امتیازها و محدودیتهای روشهای سنجش عاطفی

روشهای سنجش عاطفی، به ویژه پرسشنامه‌ها یا عقیده‌سنج‌ها، دارای امتیازهایی هستند که از جمله می‌توان به موردهای زیر اشاره کرد:

۱. اجرای این روشها، در مقایسه با بعضی روشهای دیگر سنجش، نیاز به صرف وقت زیادی ندارد.
۲. این روشها با گروههای کوچک و بزرگ قابل اجرا هستند.
۳. استفاده از این روشها نسبتاً کم هزینه و آسان است.

در مقابل امتیازهای بالا، روشهای سنجش عاطفی، مخصوصاً پرسشنامه‌ها (چه بازپاسخ و چه بسته‌پاسخ) با محدودیتهایی روبه‌رو هستند. از جمله اینکه نتایج آنها چندان قابل اعتماد نیست، یعنی داده‌های حاصل از عینیت لازم برخوردار نیستند (پین، ۲۰۰۳). دیگر اینکه در تفسیر نتایج (مخصوصاً پرسشنامه‌های بازپاسخ) ذهنیت زیادی دخالت می‌کند. همچنین گفته شده است که ساختن پرسشهای مفید و مؤثر دشوار است و از عهدهٔ همه کس برنمی‌آید. با این حال، پین (۲۰۰۳) اکثر اشکالات روشهای پرسشنامه‌ای را به استفادهٔ نادرست از آنها ربط می‌دهد تا خود آنها. "پرسشنامه‌ها اگر به خوبی ساخته و تحلیل شوند می‌توانند اطلاعات بسیار ارزشمندی دربارهٔ متغیرهای عاطفی فراهم آورند" (ص ۳۵۹).

موارد استفادهٔ روشهای سنجش عاطفی

روشهای سنجش عاطفی، از پرسشنامه‌های بازپاسخ و بسته پاسخ گرفته تا روش افتراق معنایی آزگود و مقیاس لیکرت عمدتاً برای سنجش ویژگیهای عاطفی، از جمله علاقه و نگرش، مفیدند. از آنجا که پرسشنامه‌های مبتنی بر مقیاس لیکرت مشهورترین ابزارهای سنجش عاطفی هستند، در این قسمت دو پرسشنامهٔ معروف ویژگیهای عاطفی مخصوص دانش‌آموزان را توضیح می‌دهیم. پرسشنامهٔ اول با نام پرسشنامهٔ کیفیت زندگی در مدرسه[1] ویژهٔ دانش‌آموزان دبستانی و پرسشنامهٔ دوم با نام پرسشنامهٔ سنجش نگرش نسبت به مدرسه (فرم تجدیدنظر شده)[2] ویژهٔ دانش‌آموزان دبیرستانی است.

پرسشنامهٔ کیفیت زندگی در مدرسه (اینلی[3] و بورک، ۱۹۹۲، نقل شده در آندرسون و بورک، ۲۰۰۰، صص ۲۰۴-۲۰۱)

پرسشنامهٔ کیفیت زندگی در مدرسه از ۳۹ ماده (سؤال) تشکیل شده است. این ۳۹ ماده دربرگیرندهٔ هفت مقیاس است که دو تا از آنها کلی (رضایت کلی و عاطفهٔ منفی) و پنج تای دیگر

1. The Quality of School Life Questionnaire (QSL)
2. School Attitude Assessment Survey-Revised
3. Ainley

اختصاصی‌اند (همبستگی اجتماعی، فرصت، معلم، ماجرا، و پیشرفت یا موفقیت).

سؤالهای پرسشنامهٔ کیفیت زندگی در مدرسه دارای چهار گزینه یا چهار طبقهٔ پاسخ کاملاً موافق، موافق، مخالف، و کاملاً مخالف‌اند که برای نمره گذاری به ترتیب ۴، ۳، ۲، و ۱ نمره به آنها تعلق می‌گیرد. اگر دانش‌آموزی به سؤال جواب ندهد یا به آن بیشتر از یک جواب بدهد در آن سؤال برای دانش‌آموز نمره‌ای منظور نمی‌شود.

راهنمایی:

می‌خواهیم به کمک سؤالهایی که در زیر جواب می‌دهید چگونگی احساس شما را دربارهٔ دبستانتان بدانیم. **این یک آزمون نیست**، و برای سؤالها جواب درست یا غلط وجود ندارد. چیزی که ما می‌خواهیم نظر شماست. پس سعی کنید نظر خودتان را دربارهٔ مدرسه‌تان به ما نشان دهید. جواب شما در اختیار هیچ کس دیگری گذاشته نخواهد شد.

لطفاً، اول از همه، سؤالهای زیر را جواب دهید:

نام مدرسه‌تان چیست؟ .....

کلاس چندم هستید؟ .....

پسرید یا دختر؟ .....

دقت کنید که هریک از سؤالهای زیر با این عبارت: **مدرسهٔ من جایی است که** شروع می‌شود و به دنبال آن اتفاق خاصی که برای شما در رابطه با مدرسه‌تان افتاده یا احساس خاصی که در رابطه با مدرسه دارید می‌آید. شما باید با گذاشتن علامت (×) در یکی از چهارخانه‌های مقابل هر سؤال نظر خود را به صورت **کاملاً موافق، موافق، مخالف، و کاملاً مخالف** در مورد آن اتفاق یا احساس تعیین کنید.

سعی کنید به همهٔ سؤالها جواب دهید. اما اگر برای جواب دادن به بعضی سؤالها نتوانستید تصمیم بگیرید که موافقید یا مخالف، به آنها جواب ندهید.

فراموش نکنید که پیش از جواب دادن به هر سؤال باید به این موضوع فکر کنید که **مدرسهٔ من جایی است که ...** مثلاً، **مدرسهٔ من جایی است که .... در آنجا احساس می‌کنم آدم مهمی هستم.**

در هر سطر یکی از مربعها را علامت بزنید

مدرسهٔ من جایی است که...	کاملاً موافقم ۴	موافقم ۳	مخالفم ۲	کاملاً مخالفم ۱
۱. واقعاً دوست دارم هر روز به آنجا بروم.	☐	☐	☐	☐
۲. معلمانم در آنجا با من مهربان هستند.	☐	☐	☐	☐
۳. در آنجا با دیگران کنار می‌آیم.	☐	☐	☐	☐
۴. در آنجا احساس موفقیت می‌کنم.	☐	☐	☐	☐
۵. در آنجا احساس خشنودی می‌کنم.	☐	☐	☐	☐
۶. در آنجا دانش‌آموزان دیگر مرا دوست دارند.	☐	☐	☐	☐
۷. در آنجا می‌دانم چگونه کارهایم را به خوبی انجام دهم.	☐	☐	☐	☐

	کاملاً موافقم ۴	موافقم ۳	مخالفم ۲	کاملاً مخالفم ۱
مدرسهٔ من جایی است که...				
۸. دوست دارم آنجا باشم.	☐	☐	☐	☐
۹. انجام تکالیفش مرا برای آینده‌ام آماده می‌کند.	☐	☐	☐	☐
۱۰. دوست دارم در آنجا بیشتر از تکالیفم کار انجام دهم.	☐	☐	☐	☐
۱۱. در آنجا احساس شادمانی می‌کنم.	☐	☐	☐	☐
۱۲. چیزهایی که در آنجا یاد می‌گیرم برایم مهم هستند.	☐	☐	☐	☐
۱۳. از یادگیری در آنجا لذت می‌برم.	☐	☐	☐	☐
۱۴. در آنجا احساس تنهایی می‌کنم.	☐	☐	☐	☐
۱۵. چیزهایی که در آنجا یاد می‌گیرم مرا برای تحصیلات بعدی آماده می‌کند.	☐	☐	☐	☐
۱۶. در انجام کارهایش موفقم.	☐	☐	☐	☐
۱۷. در آنجا دانش‌آموز بودن برایم افتخارآمیز است.	☐	☐	☐	☐
۱۸. در آنجا احساس نگرانی به من دست می‌دهد.	☐	☐	☐	☐
۱۹. در آنجا معلمانم در کارهایم به من کمک می‌کنند.	☐	☐	☐	☐
۲۰. در آنجا دیگران به من اعتماد می‌کنند.	☐	☐	☐	☐
۲۱. در آنجا خیلی به من خوش می‌گذرد.	☐	☐	☐	☐
۲۲. در آنجا معلمانم به آنچه من می‌گویم گوش می‌دهند.	☐	☐	☐	☐
۲۳. از کارهایی که در آنجا انجام می‌دهم لذت می‌برم.	☐	☐	☐	☐
۲۴. در آنجا احساس ناخشنودی می‌کنم.	☐	☐	☐	☐
۲۵. در آنجا آنچه راکه نیاز دارم یاد می‌گیرم.	☐	☐	☐	☐
۲۶. در آنجا می‌دانم که از عهدهٔ کارهایم برمی‌آیم.	☐	☐	☐	☐
۲۷. با انجام کارهای مربوط به آن سر شوق می‌آیم.	☐	☐	☐	☐
۲۸. در آنجا عصبانی می‌شوم.	☐	☐	☐	☐
۲۹. آدمهایش خیلی به فکر من هستند.	☐	☐	☐	☐
۳۰. با سایر دانش‌آموزان همکلاسی‌ام به‌خوبی کنار می‌آیم.	☐	☐	☐	☐
۳۱. در آنجا هر چه یاد می‌گیرم مفید است.	☐	☐	☐	☐
۳۲. در آنجا کارهایی که انجام می‌دهم جالب‌اند.	☐	☐	☐	☐
۳۳. از بودن در آنجا لذت می‌برم.	☐	☐	☐	☐
۳۴. در آنجا معلمانم کمکم می‌کنند تا کارهایم را درست انجام دهم.	☐	☐	☐	☐
۳۵. در آنجا دیگران می‌توانند به من اعتماد کنند.	☐	☐	☐	☐
۳۶. در آنجا دانش‌آموزان دیگر مهربان هستند.	☐	☐	☐	☐
۳۷. در آنجا احساس بی‌قراری می‌کنم.	☐	☐	☐	☐
۳۸. معلمانم در کلاس با من دوستانه رفتار می‌کنند.	☐	☐	☐	☐
۳۹. درآنجا هرچه یادمی‌گیرم بعد از اینکه مدرسه را ترک کنم به دردم خواهد خورد.	☐	☐	☐	☐

اگر دوست دارید چیز دیگری دربارهٔ مدرسه‌تان به ما بگویید، آن را در اینجا بنویسید.

توصیف هفت مقیاس پرسشنامهٔ کیفیت زندگی در مدرسه

نام مقیاس	توصیف مختصر مقیاس و یک نمونه سؤال برای آن	شمارهٔ سؤالهای مقیاس
رضایت کلی	احساس کلیِ احساس آرامش در مدرسه "من احساس شادمانی می‌کنم."	۱، ۸، ۱۱، ۱۷، ۲۱، ۳۳ (۶ سؤال)
عاطفهٔ منفی	احساس کلیِ منفی دربارهٔ مدرسه "عصبانی می‌شوم."	۵، ۱۴، ۱۸، ۲۸، ۳۷ (۵ سؤال)
معلمان	احساس مربوط به درستی رابطهٔ متقابل بین دانش‌آموزان و معلمان "معلمانم کمکم می‌کنند تا کارهایم را به بهترین وجه انجام دهم."	۲، ۱۹، ۲۲، ۳۴، ۳۸ (۵ سؤال)
همبستگی اجتماعی	احساس تعلق داشتن، احساس ارزشمندی و توانایی کنار آمدن با دیگران "من با سایر دانش‌آموزان همکلاسی‌ام به خوبی کنار می‌آیم."	۳، ۶، ۲۰، ۲۴، ۲۹، ۳۰، ۳۵، ۳۶ (۸ سؤال)
فرصت	مقیاس مربوط به آینده دربارهٔ باور فرد نسبت به مفید بودن و ربط داشتن مدرسه "انجام کارها مرا برای آینده‌ام آماده می‌کند."	۹، ۱۲، ۱۵، ۲۵، ۳۱، ۳۹ (۶ سؤال)
موفقیت	احساس اطمینان به توانایی برای کسب توفیق در انجام کارهای مدرسه "احساس موفقیت می‌کنم."	۴، ۷، ۱۶، ۲۶ (۴ سؤال)
ماجرا	احساس خودانگیختگی در یادگیری و این باور که یادگیری به خودی خود لذت‌بخش است "از کارهایی که انجام می‌دهم لذت می‌برم"	۱۰، ۱۳، ۲۳، ۲۷، ۳۲ (۵ سؤال)

توجه: به یاد داشته باشید که هر یک از سؤالهای نقل شده در بالا با مقدمهٔ "مدرسهٔ من جایی است که در آنجا..." شروع می‌شود.

پرسشنامهٔ سنجش نگرش نسبت به مدرسه ـ فرم تجدیدنظرشده (مککوچ[1] و سیگل[2]، ۲۰۰۳)

پرسشنامهٔ سنجش نگرش نسبت به مدرسه (فرم تجدیدنظرشده) شامل ۳۵ ماده (سؤال) و ۵ عامل یا مقیاس است. هدف این پرسشنامه شناسایی دانش‌آموزانی است که توانایی یادگیری موضوعهای درسی را دارند اما در عمل دانش‌آموزان ناموفق یا کم‌اندوزی[3] هستند. منظور از دانش‌آموزان کم‌اندوز کسانی هستند که بین پیشرفتی که از آنان انتظار می‌رود و پیشرفت واقعی آنان اختلاف محسوسی وجود دارد. البته این تفاوت نباید به علت ناتواناییهای یادگیری باشد (مککوچ و سیگل، ۲۰۰۵).

1. McCoach 2. Siegll 3. underachiver

سؤالهای پرسشنامهٔ سنجش نگرش نسبت به مدرسه (فرم تجدیدنظرشده) دارای هفت گزینهٔ کاملاً موافقم، موافقم، کمی موافقم، نظری ندارم، کمی مخالفم، مخالفم، و کاملاً مخالفم است که برای نمره‌گذاری به ترتیب ۷، ۶، ۵، ۴، ۳، ۲، و ۱ نمره به آنها تعلق می‌گیرد.

راهنمایی:
پر کردن این پرسشنامه ۵ تا ۱۰ دقیقه از وقت شما را می‌گیرد. لطفاً هر سؤال را به دقت بخوانید و نظر خود را با انتخاب یکی از پاسخهای کاملاً موافقم (۷)، موافقم (۶)، کمی موافقم (۵)، نظری ندارم (۴)، کمی مخالفم (۳)، مخالفم (۲)، و کاملاً مخالفم (۱) با گذاشتن علامت (x) در جلوی هر سؤال مشخص کنید. توجه داشته باشید که برای هر سؤال فقط یک پاسخ انتخاب کنید.

با تشکر از همکاری صمیمانهٔ شما

	کاملاً موافقم	موافقم	کمی موافقم	نظری ندارم	کمی مخالفم	مخالفم	کاملاً مخالفم
	۷	۶	۵	۴	۳	۲	۱
۱. کلاسهای من جالب هستند.	☐	☐	☐	☐	☐	☐	☐
۲. من باهوشم.	☐	☐	☐	☐	☐	☐	☐
۳. می‌توانم موضوعهای تازه را به سرعت در مدرسه یاد بگیرم.	☐	☐	☐	☐	☐	☐	☐
۴. از این که به این مدرسه می‌روم خوشحالم.	☐	☐	☐	☐	☐	☐	☐
۵. مدرسهٔ من مدرسهٔ خوبی است.	☐	☐	☐	☐	☐	☐	☐
۶. من در مدرسه تلاش زیادی می‌کنم.	☐	☐	☐	☐	☐	☐	☐
۷. با معلمانم روابط خوبی دارم.	☐	☐	☐	☐	☐	☐	☐
۸. خودم دوست دارم تکالیف درسی‌ام را انجام دهم.	☐	☐	☐	☐	☐	☐	☐
۹. این مدرسه برای من بهترین مدرسه است.	☐	☐	☐	☐	☐	☐	☐
۱۰. انجام دادن کارهای این مدرسه برای من راحت است.	☐	☐	☐	☐	☐	☐	☐
۱۱. معلمانم را دوست دارم.	☐	☐	☐	☐	☐	☐	☐
۱۲. معلمان یادگیری را برای من جالب می‌سازند.	☐	☐	☐	☐	☐	☐	☐
۱۳. معلمانم به من اهمیت می‌دهند.	☐	☐	☐	☐	☐	☐	☐
۱۴. موفق شدن در مدرسه برای هدفهای شغلی آینده من مهم است.	☐	☐	☐	☐	☐	☐	☐
۱۵. من این مدرسه را دوست دارم.	☐	☐	☐	☐	☐	☐	☐
۱۶. می‌توانم مفاهیم پیچیده را در مدرسه بفهمم.	☐	☐	☐	☐	☐	☐	☐
۱۷. موفق شدن در مدرسه یکی از هدفهای من است.	☐	☐	☐	☐	☐	☐	☐
۱۸. تکالیف درسی‌ام را به طور منظم انجام می‌دهم.	☐	☐	☐	☐	☐	☐	☐
۱۹. برای من مهم است که در مدرسه نمره‌های خوبی بگیرم.	☐	☐	☐	☐	☐	☐	☐
۲۰. در انجام دادن تکالیف درسی‌ام آدم منظمی هستم.	☐	☐	☐	☐	☐	☐	☐
۲۱. برای یادگرفتن مطالب جدید از روشهای مختلفی استفاده می‌کنم.	☐	☐	☐	☐	☐	☐	☐
۲۲. می‌خواهم در مدرسه تا آنجاکه می‌توانم بکوشم.	☐	☐	☐	☐	☐	☐	☐

	کاملاً موافقم ۷	موافقم ۶	کمی موافقم ۵	نظری ندارم ۴	کمی مخالفم ۳	مخالفم ۲	کاملاً مخالفم ۱
۲۳. کسب موفقیت در مدرسه برای من مهم است.	□	□	□	□	□	□	□
۲۴. برای انجام تکالیف درسی‌ام وقت زیادی صرف می‌کنم.	□	□	□	□	□	□	□
۲۵. بیشتر معلمان این مدرسه معلمان خوبی هستند.	□	□	□	□	□	□	□
۲۶. من یک دانش‌آموز وظیفه‌شناس هستم.	□	□	□	□	□	□	□
۲۷. برای انجام دادن تکالیف درسی‌ام تلاش زیادی می‌کنم.	□	□	□	□	□	□	□
۲۸. کلاسهایم را دوست دارم.	□	□	□	□	□	□	□
۲۹. برای انجام دادن تکالیف درسی‌ام تمام حواسم را جمع می‌کنم.	□	□	□	□	□	□	□
۳۰. قبل از آن که تکالیفم را تحویل دهم آنها را وارسی می‌کنم.	□	□	□	□	□	□	□
۳۱. می‌توانم در همهٔ درسهایم نمره‌های عالی بگیرم.	□	□	□	□	□	□	□
۳۲. می‌خواهم در درسهایم نمره‌های خوبی بگیرم.	□	□	□	□	□	□	□
۳۳. در یادگیری موضوعهای تازه در مدرسه موفق هستم.	□	□	□	□	□	□	□
۳۴. در مدرسه دانش‌آموز زرنگی هستم.	□	□	□	□	□	□	□
۳۵. به بودن در این مدرسه افتخار می‌کنم.	□	□	□	□	□	□	□

توصیف پنج عامل یا مقیاس پرسشنامهٔ سنجش نگرش نسبت به مدرسه ـ فرم تجدیدنظرشده

نام عامل	توصیف مختصر عامل و یک نمونه سؤال برای آن	شمارهٔ سؤالهای عامل
ادراک از خود نسبت به تحصیل	توصیف و ارزیابی فرد از تواناییهای تحصیلی خودش "من باهوشم."	۲، ۳، ۱۰، ۱۶، ۳۱، ۳۳، ۳۴
نگرش نسبت به معلمان و کلاسها	طرز فکر و علاقه‌مندی یا بی‌علاقگی دانش‌آموز نسبت به معلمان و آنچه در کلاس می‌گذرد "با معلمانم روابط خوبی دارم."	۱، ۷، ۱۱، ۱۲، ۱۳، ۲۱، ۲۵، ۲۸
نگرش نسبت به مدرسه	طرز فکر و علاقه‌مندی یا بی‌علاقگی دانش‌آموز نسبت به کل مدرسه "من این مدرسه را دوست دارم."	۴، ۵، ۹، ۱۵، ۳۵
ارزش قائل شدن برای هدف	قدر و ارزشی که پیشرفت در مدرسه و کسب موفقیت برای دانش‌آموز دارد "برای من مهم است که در مدرسه نمره‌های خوبی بگیرم."	۱۴، ۱۷، ۱۹، ۲۲، ۲۳، ۳۲
انگیزش و خودسامانی یا خودتنظیم‌گری	میزان علاقه‌مندی دانش‌آموز نسبت به یادگیری و اندیشه‌ها، احساسها، و اعمال خودانگیختهٔ او "من یک دانش‌آموز وظیفه‌شناس هستم."	۶، ۸، ۱۸، ۲۰، ۲۱، ۲۴، ۲۶، ۲۷، ۲۹، ۳۰

خلاصهٔ فصل

توجه به ویژگیهای عاطفی دانش‌آموزان و پرورش عاطفه‌های مثبت در آنان از وظایف مهم آموزش و پرورش است. ویژگیهای عاطفی هم به خودی خود از هدفهای مهم آموزشی هستند و هم در رساندن دانش‌آموزان و دانشجویان به سایر هدفهای آموزشی وسیلهٔ مهمی به حساب می‌آیند.

عاطفه‌ها را با توجه به ۵ ملاک تعریف می‌کنند. یکی اینکه دربرگیرندهٔ احساس یا هیجان است. دیگری اینکه ویژگی معمولی یا نوعی فرد است. همچنین هر عاطفه دارای شدت، جهت، و هدف است. منظور از شدت درجه یا نیروی عاطفه است. منظور از جهت جنبهٔ مثبت یا منفی آن است، و منظور از هدف یا آماج موضوعی است که عاطفه متوجه آن است.

سنجش عواطف با سنجش تواناییها یک فرق عمده دارد و آن هم این است که آزمونهای توانایی از آزمون‌شونده می‌خواهند تا حداکثر توان خود را در پاسخ به سؤالهای آزمون نشان دهد، اما وسایل سنجش عواطف از آزمون‌شونده می‌خواهند هرآنچه را به طور معمول احساس می‌کند بگوید یا بنویسد. بنابراین، مشکل اساسی روشهای سنجش هدفهای عاطفی این است که اگر آزمون‌شونده (دانش‌آموز یا دانشجو) احساس کند که نتایج سنجش برای نمره‌دادن و قضاوت دربارهٔ او به کار خواهند رفت احتمال زیادی دارد که در پاسخ به سؤالهای وسیلهٔ سنجش جواب درست ندهد.

یکی از روشهای معروف سنجش عاطفی مورد بحث در این کتاب روش پرسشنامه است. اصطلاح نزدیک به پرسشنامه عقیده‌سنج است. پرسشنامه‌ها یا عقیده‌سنجها دو دسته‌اند: بازپاسخ و بسته‌پاسخ. پرسشنامه‌های بازپاسخ شبیه به آزمونهای تشریحی و پرسشنامه‌های بسته‌پاسخ شبیه به آزمونهای عینی هستند.

روش تفکیک معنایی یا افتراق معنایی آزگود وسیلهٔ دیگری برای سنجش عواطف است. این روش شامل تعدادی صفت دوقطبی است که در آن هر جفت صفت یک پیوستار را شامل می‌شود و به صورت مقیاس درجه‌بندی تنظیم می‌گردد. از پاسخ‌دهندگان خواسته می‌شود تا با گذاشتن علامت در روی پیوستار موضوع نگرش را درجه‌بندی کنند.

روش دیگر سنجش عاطفی مقیاس لیکرت است که عمدتاً برای سنجش نگرشها از آن استفاده می‌شود. این مقیاس از یک رشته بیانات دربارهٔ اشخاص، گروهها، یا اشیاء درست شده است. از پاسخ‌دهندگان خواسته می‌شود تا میزان یا درجهٔ موافقت یا مخالفت خود را با هر یک از آن بیانات مشخص کنند.

در نمره‌گذاری مقیاس لیکرت به هر سؤال وزن خاصی داده می‌شود. بسته به اینکه سؤال به صورت مثبت نوشته شده باشد یا به صورت منفی وزنهای آن فرق می‌کند.

افزون بر روشهای فوق، از مصاحبه، گروه‌سنجی، و یادداشتهای روزانه نیز می‌توان برای

سنجش عواطف دانش‌آموزان سود برد. مصاحبه یک ابزار سنجش است که به کمک آن اطلاعات موردنیاز از راه ارتباط مستقیم و دوجانبه بین سنجشگر و فرد مورد سنجش به دست می‌آید. در گروه‌سنجی روابط عاطفی میان اعضای یک گروه تعیین می‌شود. در روش یادداشتهای روزانه معلم دانش‌آموز را تشویق می‌کند تا مرتباً احساسهای خود را نسبت به معلم و مدرسه یادداشت کند و آن را با معلم در میان بگذارد.

امتیازهای روشهای سنجش عاطفی به قرار زیرند. (۱) نیاز به صرف وقت زیاد ندارند، (۲) با گروههای کوچک و بزرگ قابل اجرا هستند، و (۳) استفاده از آنها ساده و کم‌هزینه است. محدودیتهایی که برای روشهای سنجش عاطفی ذکر کرده‌اند یکی این است که از عینیت کامل برخوردار نیستند؛ دیگر اینکه در تفسیر نتایج ذهنیت سنجشگر دخالت می‌کند. سوم اینکه ساختن پرسشهای مفید و مؤثر دشوار است.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. علت استفادهٔ زیاد از اصطلاح پرسشنامه در مقایسه با اصطلاح عقیده‌سنج این است که پرسشنامه بیشتر از عقیده‌سنج بر احساس و عاطفه تأکید می‌کند تا متغیرهای شناختی.

الف ـ صحیح

ب ـ غلط

۲. بیشترین مورد استفادهٔ مقیاس لیکرت برای سنجش _____ است.

الف ـ نگرش

ب ـ علاقه

پ ـ احساس

ت ـ شخصیت

ث ـ درون‌گرایی ـ برون‌گرایی

۳. ویژگیهای عاطفی را تعریف کنید، و با ذکر مثال ملاکهای تشخیص ویژگیهای عاطفی را توضیح دهید.

۴. اهمیت سنجش عواطف را در آموزشگاه توضیح دهید.

۵. ویژگیهای روشهای سنجش عاطفی را توضیح دهید.

۶. روش سنجش به کمک پرسشنامه را توضیح دهید و با ذکر مثال انواع آن را مشخص کنید.

۷. تفاوت بین پرسشنامه و عقیده‌سنج را توضیح دهید.

۸. قواعد تهیه و اجرای پرسشنامه‌های سنجش عواطف را توضیح دهید.
۹. روش تفکیک معنایی را توضیح دهید و تفاوت آن را با پرسشنامه بنویسید.
۱۰. مقیاس لیکرت را تعریف کنید و چگونگی تهیه و نمره‌گذاری آن را شرح دهید.
۱۱. قواعد تهیه و اجرای مقیاس لیکرت را توضیح دهید.
۱۲. گروه‌سنجی را تعریف کنید و روشهای معروف آن را توضیح دهید.
۱۳. روش یادداشتهای روزانه را تعریف کنید و چگونگی استفادهٔ بهینه از آن را توضیح دهید.
۱۴. امتیازها و محدودیتهای روشهای سنجش عاطفی را توضیح دهید.
۱۵. در ارتباط با جدول مشخصات آزمون خود که به عنوان تمرین فصل ۶ تهیه کرده‌اید، برای هدفهای حوزهٔ عاطفی، از طریق پرسشنامه‌های بازپاسخ و بسته‌پاسخ، روش تفکیک معنایی، و مقیاس لیکرت، یک برنامهٔ سنجش عاطفی تهیه کنید.

۱۷

اجرا، نمره‌گذاری، و تحلیل آزمون

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. روشهای مختلف مرتب کردن سؤالهای گوناگون یک آزمون را توضیح دهد.

۲. دستورالعمل یا راهنمای آزمون را تعریف کند و قسمتهای مهم آن را نام ببرد.

۳. برای آزمونی که خود ساخته است نکات توضیح داده شده در بارهٔ نحوهٔ مرتب‌کردن سؤالها را به کاربندد و یک دستورالعمل یا راهنمای مناسب بنویسد.

۴. شرایط مطلوب اجرای آزمون را توضیح دهد.

۵. پس از تهیه و اجرای یک آزمون و جمع‌آوری پاسخنامه‌ها اقدامات زیر را انجام دهد:

الف ـ پاسخنامه‌ها را تصحیح کند.

ب ـ ضریب دشواری و ضریب تمیز کلیهٔ سؤالها را حساب کند.

پ ـ گزینه‌های انحرافی هر سؤال را وارسی کند و گزینه‌های ضعیف را تشخیص دهد.

ت ـ با استفاده از اطلاعات حاصل از مراحل بالا، آزمون خود را مورد تجدید نظر قرار دهد.

۶. روشهای مختلف محاسبهٔ ضریبهای دشواری و تمیز را توضیح دهد و آنها را با هم مقایسه نماید.

۷. روش تحلیل سؤال آزمونهای تشریحی را توضیح دهد.

۸. روش تحلیل آزمونهای وابسته به ملاک را توضیح دهد.

۹. تفاوتهای مهم بین تحلیل سؤال آزمونهای وابسته به هنجار و تحلیل سؤال آزمونهای وابسته به ملاک را توضیح دهد.

۱۰. روش محاسبهٔ ضریب تمیز پیش و پس از آزمون را توضیح دهد.

۱۱. روش محاسبهٔ ضریب تمیز بین حد تسلط و غیر حد تسلط را توضیح دهد.

۱۲. تفاوت بین ضریب تمیز پیش و پس از آموزش را با ضریب تمیز بین حد تسلط و غیرحد تسلط توضیح دهد.

۱۳. برای هر نوع آزمون عینی وابسته به هنجار یا وابسته به ملاک و یا آزمون تشریحی روش مناسب تحلیل سؤال آن آزمون را مورد استفاده قرار دهد.

۱۴. نظریهٔ صفت مکنون و نظریهٔ سؤال ـ پاسخ را توضیح دهد و رابطهٔ بین آنها را بیان کند.

۱۵. منحنی ویژگی سؤال را توضیح دهد و چگونگی تهیهٔ آن را با ذکر مثال نشان دهد.

۱۶. نحوهٔ تحلیل سؤالهای آزمون را با روش منحنی ویژگی سؤال توضیح دهد.

فصل حاضر مکمل فصلهای قبلی کتاب است. ما در فصلهای قبلی با روش تهیهٔ آزمونهای مختلف و چگونگی نوشتن سؤالهای گوناگون آشنا شدیم. در این فصل روش اجرای آزمونهای کتبی، طرز نمره‌گذاری پاسخها، و چگونگی تحلیل سؤالهای این آزمونها را یاد خواهیم گرفت. همچنین در این فصل خواهیم آموخت که با استفاده از نتایج حاصل از آزمون خود چگونه به تجدید نظر در آن و رفع نواقص و اشکالات آن اقدام کنیم.

شرایط اجرای آزمونهای کتبی

گام نخست در اجرای آزمونهای پیشرفت تحصیلی آماده‌سازی و فراهم آوردن شرایط مناسب برای اجرای آنهاست. تهیهٔ دفترچهٔ آزمون و پاسخنامه، ترتیب پشت سرهم قراردادن سؤالهای مختلف، و درنظر گرفتن زمان و مکان مناسب برای اجرای آزمون نیاز به دقت فراوان دارند. مجموعهٔ این عوامل باید به گونه‌ای فراهم شوند که برای هریک از آزمون‌شوندگان حداکثر امکانات را جهت نشان دادن بهترین عملکرد او فراهم آورند. نکاتی که در زیر توضیح می‌دهیم ناظر بر این هدف هستند.

ترتیب قرار دادن سؤالهای آزمون به دنبال یکدیگر

قبلاً در رابطه با چگونگی تهیهٔ سؤالهای مختلف آزمونهای پیشرفت تحصیلی گفتیم که سؤالهایی باید تهیه شوند که تکالیفی را که در اختیار آزمون‌شوندگان قرار می‌دهند تا حد امکان برای آنان آشکار و قابل فهم باشند. یکی از راههای کمک به این امر آن است که سؤالهای مختلف یک آزمون را با رعایت اصول صحیح به دنبال هم قرار دهیم. این کار به آزمون‌شونده نظم فکری می‌دهد و به او کمک می‌کند تا راحت‌تر به سؤالها پاسخ بدهد.

گرانلاند و لین (۱۹۹۰) در رابطه با قرار دادن سؤالهای آزمون به دنبال یکدیگر چند پیشنهاد زیر را ارائه داده‌اند:

۱. سؤالها را با توجه به نوع آنها دسته‌بندی کنید و پشت سرهم قرار دهید. در این باره می‌توان ترتیب زیر را برای ردیف کردن سؤالهای یک آزمون به دنبال هم مورد استفاده قرار داد:

الف ـ صحیح ـ غلط

ب ـ جورکردنی

پ ـ چندگزینه‌ای

ت ـ کوته‌پاسخ

ث ـ تشریحی

علاوه بر دسته‌بندی سؤالها بر حسب نوع، می‌توان سؤالهای مختلف هر نوع را نیز بر حسب طبقه‌بندی هدفهای آموزشی دسته‌بندی کرد. برای نمونه، سؤالهای مربوط به بخش چندگزینه‌ای را می‌توان به ترتیب (۱) به یادآوردن دانش، (۲) فهمیدن دانش، (۳) به کار بستن دانش، (۴) تحلیل دانش، و مانند اینها مرتب کرد. "کنار هم قرار دادن سؤالهایی که بازده‌های یادگیری مشابهی را می‌سنجند در تعیین نوع بازده‌های یادگیری که برای یادگیرندگان مشکل‌زا هستند معلم را یاری می‌دهند" (گرانلاند و لین، ۱۹۹۰، ص ۲۳۴).

۲. سؤالها را از ساده به دشوار مرتب کنید. اگر مرتب کردن سؤالهای یک آزمون با توجه به بازده‌های یادگیری امکان‌پذیر نباشد، می‌توانید آنها را، با توجه به ساده و دشوار بودن، پشت سرهم قرار دهید. یعنی ابتدا سؤالهای ساده را قرار دهید و بعد به ترتیب سؤالهایی را که از لحاظ درجهٔ دشواری در سطح بالاتری قرار دارند به دنبال هم بیاورید.

وقتی که یک آزمون با سؤالهای ساده آغاز می‌شود، آزمون‌شوندگان با پاسخ دادن به این سؤالها احساس اعتماد به نفس می‌کنند و این امر کارآیی آنان را در عملکرد روی سؤالهای پیچیدهٔ بعدی نیز افزایش می‌دهد. در این حالت، حتی ضعیف‌ترین دانش‌آموزان نیز خواهند کوشید تا حداکثر توان خود را در پاسخدهی به سؤالها به کار گیرند.

دقت کنید که در این روش مرتب کردن سؤالها نیز ابتدا سؤالها بر حسب نوع دسته‌بندی می‌شوند؛ آن گاه، در درون هر نوع، سؤالها به ترتیب از ساده به دشوار چیده می‌شوند. در واقع تنها فرق این روش با روش قبلی در این است که در روش قبلی سؤالهای هر دسته یا هر نوع به ترتیب طبقات طبقه‌بندی هدفهای آموزشی، یعنی دانش، فهمیدن، کاربستن، و غیره مرتب می‌شوند، اما در روش دوم سؤالهای داخل هر نوع به ترتیب از ساده به دشوار مرتب می‌شوند.

۳. سؤالها را به ترتیب سازمان اصلی مطالب به دنبال هم مرتب کنید. منظور این است که ترتیب سؤالها باید با ترتیب مطالب درس هماهنگ باشد، یعنی به ترتیب فصلهای کتاب یا بخشهای درس که آموزش داده شده‌اند مرتب شوند. رعایت این نکته می‌تواند به ایجاد نظم فکری در آزمون‌شوندگان کمک کند. با این حال، گرانلاند و لین (۱۹۹۰) در این باره اظهار داشته‌اند که، "در ساختن آزمونهای پیشرفت تحصیلی کلاسی، مرتب کردن سؤالها با توجه به ساختمان محتوای درس کمک چندانی نمی‌کند. اگر چنین کاری مطلوب جلوه کرد، مثلاً موقعی که می‌خواهید دوره‌های تاریخی را از هم جدا کنید، بکوشید تا این قسمت‌بندی را به حداقل برسانید" (ص ۲۳۴).

علاوه بر نکات بالا، سعی کنید سؤالها را خوش‌خط و خوانا بنویسید و در بین سؤالها فاصله‌های لازم را برای سهولت مطالعه رعایت کنید. آزمونهایی که سؤالهای آنها درهم‌اند و خوانا نیستند مقدار زیادی از وقت و نیروی آزمون‌شونده را به‌هدر می‌دهند و باعث دلسردی او می‌شوند.

در ضمن اگر سؤالی دارای شکل یا نمودار است آن را در بالای تنهٔ سؤال قرار دهید نه در پایین گزینه‌ها. اگر شکلها و نمودارها، یا هر چیز دیگری که متن سؤال به آن اشاره دارد، پس از سؤال قرار داده شود، آزمون شونده ناچار خواهد بود بعد از خواندن هر گزینه به آن مراجعه نماید و این کار موجب اتلاف وقت او خواهد شد. این امر در رابطه با قاعده‌ای که در فصل ۹ گفتیم، یعنی اینکه تمام مطالب مربوط به یک سؤال باید در تنهٔ سؤال قرار داده شوند نه در گزینه‌ها، توضیح داده شد. همچنین به یاد داشته باشید که جای گزینه‌های درست در میان دیگر گزینه‌ها باید به طور تصادفی انتخاب شود و از هیچ الگوی مشخصی پیروی نکند.

نوشتن دستورالعمل یا راهنمای آزمون

راهنمای آزمون باید مختصر و مفید باشد و به‌طور آشکار به آزمون شونده بگوید که چه کاری را انجام دهد. مهرنز و لهمان (۱۹۸۴) گفته‌اند که راهنمای آزمون باید اطلاعات زیر را در اختیار آزمون‌شوندگان قرار دهد:

۱. زمان لازم برای هر قسمت
۲. ارزش هر سؤال
۳. مجاز یا غیرمجاز بودن حدس زدن

بجز برای دانش‌آموزان خردسال، راهنمای آزمون باید آن قدر آشکار باشد که آزمون شونده با خواندن آن و بدون نیاز به هیچ‌گونه توضیح شفاهی کار خود را آغاز کند. اگر آزمون‌شوندگان نیاز به توضیح شفاهی داشته باشند، معلوم می‌شود که راهنمای آزمون به حد کافی روشن نبوده است و باید بازنویسی شود. اما هنگام اجرای آزمون با کودکان سالهای پایین ممکن است، علاوه بر توضیحات کتبی، راهنماییهای شفاهی هم ضرورت داشته باشند.

افزون بر یک راهنمایی کلی برای تمامی آزمون، برای هر بخش آزمون نیز یک راهنمایی جداگانه بنویسید (نگاه کنید به شکل ۱ـ۱۷). برای مسائل عددی، میزان دقتی را که انتظار دارید مشخص کنید (مثلاً چند عدد اعشاری باید محاسبه شود). همچنین واحدهای محاسبه چون ولت، اهم، سانتی‌متر، یا میلی‌متر را مشخص کنید. اگر در تصحیح پاسخها به مواردی چون خوش‌خطی، املای کلمه‌ها و جمله‌بندی نمره می‌دهید، این موارد را نیز در راهنمای سؤال بنویسید. نحوهٔ نمره‌گذاری پاسخها را به اطلاع آزمون‌شوندگان برسانید. مثلاً در آزمون مربوط به مسائل ریاضی توضیح دهید که تنها جواب مسئله را می‌خواهید یا مراحل رسیدن به جواب نیز ضروری هستند. همچنین مشخص کنید که به هر بخش چه مقدار نمره تعلق می‌گیرد.

در راهنمای آزمونهای عینی نحوهٔ نمره‌گذاری پاسخها را در رابطه با حدس زدن یا حدس نزدن جوابها مشخص کنید. اگر برای حدس‌زدن از نمرهٔ دانش‌آموزان کسر می‌کنید، مقدار دقیق آن

راهنمایی کلی: پیش از آنکه دفترچهٔ آزمون را باز کنید، مطالب زیر را به دقت بخوانید. از آنجا که پاسخهای شما با ماشین تصحیح می‌شوند، لازم است محل پاسخهای خود را با مداد کاملاً سیاه کنید تا نمره‌ای را از دست ندهید.

راهنماییهای اختصاصی: این آزمون شامل ۷۵ سؤال چهارگزینه‌ای است که ارزش هر یک از آنها یک نمره است. هر سؤال را به دقت بخوانید و ببینید کدام گزینه بهترین جواب سؤال یا کامل کنندهٔ جملهٔ سؤال است.

۱. در پاسخنامه با خط خوانا نام خود، نام آزمون، و تاریخ آزمون را بنویسید.
۲. دقت کنید که پاسخ خود را برای هر سؤال در محل مربوط به آن در پاسخنامه علامت بزنید.
۳. از کاغذ اضافی که به شما داده شده است به عنوان پیش‌نویس استفاده کنید.
۴. برای هر سؤال یک پاسخ را علامت بزنید، اگر بیش از یک پاسخ برای هر سؤال انتخاب کنید به آن نمره‌ای تعلق نخواهد گرفت.
۵. نمرهٔ کل شما در این آزمون برابر با تعداد کل پاسخهای درست شما خواهد بود. سعی کنید به همهٔ سؤالها جواب بدهید، اما برای یک سؤال وقت زیادی صرف نکنید.
۶. اگر جواب سؤالی را نمی‌دانید حدس بزنید، نمرهٔ منفی وجود ندارد.
۷. حال می‌توانید شروع کنید. موفق باشید.

این آزمون شامل ۷۵ سؤال در ۸ صفحه است. دقت کنید سؤالی از نظر نیفتد.

شکل ۱-۱۷ راهنمایی یک آزمون چندگزینه‌ای (اقتباس از اثر مهرنز و لهمان، ۱۹۸۴)

را معین کنید (مثلاً توضیح دهید که در آزمونهای چهارگزینه‌ای برای هر سه پاسخ غلط ۱ نمره کم می‌کنید). روش نمره گذاری برای جبران حدس زدن در دنبالهٔ مطالب این فصل معرفی خواهد شد. مهرنز و لهمان (۱۹۸۴، ص ۱۸۲) می‌گویند: "ما بر اساس شواهد پژوهشی موجود متقاعد شده‌ایم که [در آزمونهای پیشرفت تحصیلی و استعداد تحصیلی] باید به آزمون‌شوندگان گفته شود که همهٔ سؤالها را جواب بدهند و در تصحیح پاسخها هیچ‌گونه نمره‌ای از آنها برای حدس زدن نباید کسر بشود."

اجرای آزمون

شرایط اجرای آزمونهای پیشرفت تحصیلی باید به گونه‌ای باشد که از هر لحاظ آسایش جسمی و فکری آزمون‌شوندگان را تأمین نماید. این نکته را در رابطه با آزمون‌شوندگان خردسال با دقت بیشتری رعایت کنید. علاوه بر شرایط فیزیکی، شرایط روانی آزمون‌شوندگان نیز بر عملکرد آنان تأثیر دارد. مهم‌ترین این عوامل اضطراب زیاد است. پژوهشهای انجام شده نشان داده‌اند که اندکی اضطراب سبب افزایش دقت و کوشش آزمون‌شوندگان می‌شود، اما اضطراب زیاد بر عملکرد آنان تأثیر منفی دارد. بنابراین، سعی کنید که در جلسهٔ امتحان از عوامل تنش‌زا و اضطراب‌برانگیز جلوگیری به عمل آورید.

نمره‌گذاری

پس از تهیه و اجرای آزمونهای پیشرفت تحصیلی، نوبت به تصحیح برگه‌های آزمون و نمره‌گذاری آنها می‌رسد. در فصل ۱۰ روش تصحیح آزمونهای تشریحی را به تفصیل توضیح دادیم، در این فصل روش تصحیح آزمونهای عینی را شرح می‌دهیم.

تصحیح برگه‌های آزمونهای عینی، به ویژه اگر پاسخنامه‌های جداگانه‌ای که در آن پاسخ سؤالها به وسیلهٔ آزمون‌شوندگان علامت‌گذاری می‌شود به کار بروند، به سادگی امکان‌پذیر است. برای تصحیح این نوع پاسخنامه‌ها، ابتدا مصحح می‌تواند یکی از آنها را خود علامت‌گذاری کند و با سوراخ کردن محل پاسخهای درست پاسخنامهٔ کلید را بسازد. بعد با قراردادن این کلید بر روی پاسخنامه‌های دانش‌آموزان و شمارش تعداد علامتهای درست، آنها را نمره گذاری کند. پیش از قرار دادن پاسخنامهٔ کلید بر روی پاسخنامه‌ای که باید نمره گذاری شود، دقت کنید که برای هر سؤال بیش از یک علامت وجود نداشته باشد. هر سؤالی که بیش از یک علامت برای آن منظور شده است باید حذف شود.

درنمره گذاری آزمونهای عینی دوروش عمده وجود دارد: (۱) محاسبهٔ کلیهٔ پاسخهای‌درست بدون کسر نمره برای حدس‌زدن؛ (۲) کسر مقداری از نمره برای جبران حدس زدن. روش اول روش متداول‌تر و ساده‌تری است. اما اگر معلم بخواهد روش دوم را مورد استفاده قرار دهد باید، هنگام اجرای آزمون، آزمون‌شوندگان را آگاه سازد که در ازای پاسخهای غلطی که می‌دهند مقداری از نمرهٔ آنان کم خواهد شد. برای تعیین این مقدار چندین روش وجود دارد که معروف‌ترین آنها استفاده از فرمول زیر است.

$$\text{نمرهٔ اصلاح شده} = R - \frac{W}{N-1}$$

در این فرمول:

R = تعداد پاسخهای درست

W = تعداد پاسخهای غلط

N = تعداد گزینه‌های هر سؤال

فرض کنید که در یک آزمون چهار گزینه‌ای ($N = ۴$) که دارای ۵۰ سؤال است، دانش‌آموزی به ۴۵ سؤال پاسخ داده است و از این تعداد ۶ سؤال غلط ($W = ۶$) و ۳۹ سؤال درست است ($R = ۳۹$). نمرهٔ اصلاح شدهٔ این دانش‌آموز برابر است با:

$$\text{نمرهٔ اصلاح شده} = R - \frac{W}{N-1}$$

$$\text{نمرهٔ اصلاح شده} = ۳۹ - \frac{۶}{۴-۱}$$

$$= ۳۹ - \frac{۶}{۳}$$

$$= ۳۹ - ۲ = ۳۷$$

تحلیل سؤالهای آزمون

هدف از تحلیل سؤالهای آزمون وارسی تک تک سؤالها و تعیین میزان دقت و نارساییهای آنهاست. در تحلیل سؤالهای آزمون، نقاط قوت و ضعف یک آزمون و کیفیت همهٔ سؤالهای آن تعیین می‌شود. بنابراین، لازم است که معلمان، پس از اجرای هر آزمون، سؤالهای آن را تحلیل کنند و با استفاده از نتایج حاصل به تجدید نظر در آزمون و بهبود کیفیت سؤالها برای استفاده‌های بعدی اقدام نمایند. برای تهیهٔ یک آزمون خوب، علاوه بر رعایت پیشنهادهایی که در فصلهای پیش داده شد، تحلیل پاسخهای آزمون‌شوندگان و بررسی تجربی آنها ضروری است. تحلیل پاسخهای آزمون‌شوندگان همچنین اطلاعات تشخیصی لازم را برای بررسی کیفیت یادگیری دانش‌آموزان و مشکلات آموزشی معلمان فراهم می‌آورد.

مراحل تحلیل سؤال

اطلاعات موردنیاز برای تحلیل سؤالهای یک آزمون پاسخهایی هستند که آزمون‌شوندگان به هر سؤال داده‌اند. بنابراین، باید تعیین شود که در هر سؤال چند نفر گزینهٔ درست را انتخاب کرده‌اند، هریک از گزینه‌های انحرافی چند نفر را به خود جلب کرده است، و چند نفر آن را بی‌جواب گذاشته‌اند. برای این منظور، باید هر سؤال را در کارتی مانند شکل ۲-۱۷ وارد کنید. این کارت را

عنوان آزمون: سنجش و اندازه گیری

موضوع سؤال: پایایی

هدف یادگیری: دانشجو بتواند موارد استفادهٔ پایایی‌های مختلف را تشخیص دهد.

سؤال

یک آزمون ۱۰۰ سؤالی پیشرفت تحصیلی از ۵ خرده آزمون ۲۰ سؤالی مربوط به موضوعهای درسی مختلف تشکیل شده است. اگر بخواهیم ضریب پایایی این آزمون را حساب کنیم بهترین روش کدام است؟

الف ـ اسپیرمن ـ براون

* ب ـ آلفای کرانباخ

ج ـ KR۲۰

د ـ KR۲۱

گروهها	الف	ب	پ	ت	بدون پاسخ	
۲۵٪ بالا	۰	۵	۳	۰	۲	۱۰
۲۵٪ پایین	۵	۲	۳	۰	۰	۱۰

ضریب دشواری: $P = \frac{۵+۲}{۱۰+۱۰} \times ۱۰۰ = \frac{۷}{۲۰} \times ۱۰۰ = ۳۵$

ضریب تمیز یا تشخیص: $D = \frac{۵-۲}{۱۰} = \frac{۳}{۱۰} = ۰/۳$

شکل ۲-۱۷ کارت تحلیل سؤال آزمون چندگزینه‌ای

می‌توان برای هر سؤال در روی یک صفحهٔ کاغذ یا مقوا درست کرد و نتایج تحلیل آن سؤال را در یک طرف و صورت اصلاح شدهٔ سؤال را در طرف دیگر کارت نوشت.

اطلاعات موجود در شکل ۲-۱۷ که برای تحلیل یکی از سؤالهای آزمون و محاسبهٔ شاخصهای آماری آن سؤال ضروری هستند، از طریق تعیین دو گروه قوی (بالا) و ضعیف (پایین) آزمون‌شوندگان و تعیین نحوهٔ پاسخدهی آنان به گزینه‌های مختلف سؤال به دست آمده‌اند. گروه بالا از ۲۵٪ تا ۳۳٪ آزمون‌شوندگان دارای نمره‌های بالا و گروه پایین از ۲۵٪ تا ۳۳٪ آزمون‌شوندگان دارای نمره‌های پایین تشکیل می‌یابد (کلاین[1]، ۲۰۰۵). این اطلاعات برای تک تک سؤالهای هر آزمون ضروری هستند. برای تعیین گروههای بالا و پایین، برگه‌های آزمون را به ترتیب نمره‌ای که گرفته‌اند از کوچک به بزرگ مرتب کنید. بعد از بالاترین نمره شروع کنید و تعداد برگه‌هایی را که می‌خواهید در گروه بالا قرار دهید انتخاب کنید. و به همین ترتیب از پایین‌ترین نمره شروع کنید و تعداد برگه‌هایی را که می‌خواهید در گروه پایین قرار دهید نیز برگزینید. متخصصان آزمون‌سازی (از جمله ویتنی[2] و سیبرز[3]، ۱۹۷۰؛ و نیتکو، ۱۹۸۳) پیشنهاد کرده‌اند که اگر تعداد کل دانش‌آموزانی که در آزمون شرکت کرده‌اند (یعنی تعداد کل برگه‌های آزمون) بیشتر از ۲۰ و تا ۴۰ نفر هستند، ۱۰ برگهٔ بالا و ۱۰ برگهٔ پایین را انتخاب کنید و کاری به بقیهٔ برگه‌ها نداشته باشید. اگر تعداد دانش‌آموزان ۲۰ نفر یا کمتر است، برگه‌های آزمون را کلاً به دو دستهٔ بالا و پایین تقسیم کنید. در شرایطی که تعداد کل آزمون‌شوندگان بیش از ۴۰ نفر است، بهترین رقم برای گروه بالا و پایین ۲۷٪ کل برگه‌هاست. این رقم برای شرایطی که نمرات دارای توزیع بهنجار باشند پیشنهاد شده است، اما برای آزمونهای کلاسی هر یک از ارقام ۲۵ تا ۳۳ درصد به عنوان تعداد برگه‌های گروه بالا یا پایین مناسب است (نیتکو، ۱۹۸۳). بعد از تعیین گروههای بالا و پایین، لازم است برای هر سؤال اطلاعات زیر را تهیه کنید.

۱. تعداد افراد گروه بالا که هریک از گزینه‌های سؤال را انتخاب کرده یا آن را بی‌جواب گذاشته‌اند.
۲. تعداد افراد گروه پایین که هریک از گزینه‌های سؤال را انتخاب کرده یا آن را بی‌جواب گذاشته‌اند.

پس از انجام مراحل بالا، اطلاعات مورد نیاز برای تعیین شاخصهای آماری و تحلیل سؤالها در دست هستند. این شاخصها به شرح زیر محاسبه می‌شوند.

محاسبهٔ ضریب دشواری سؤال

بنا به تعریف، درصد کل آزمون‌شوندگانی که به یک سؤال جواب درست می‌دهند **ضریب دشواری**[4] آن سؤال است که با حرف P نشان داده می‌شود. اگر در تحلیل یک سؤال کلیهٔ افراد یا کلیهٔ برگه‌های

1. Kline
2. Whitney
3. Sabers
4. difficulty index

امتحانی دخالت داشته باشند برای محاسبهٔ ضریب دشواری آن سؤال کافی است که تعداد کل افرادی را که به آن سؤال جواب درست داده‌اند بر تعداد کل آزمون‌شوندگان تقسیم کنیم و نتیجه را در ۱۰۰ ضرب نماییم. رقم حاصل ضریب دشواری سؤال است. مطابق فرمول زیر:

$$P = ۱۰۰ \left(\frac{R}{T}\right)$$

در این فرمول R معرف تعداد کسانی است که به سؤال جواب درست داده و T نشان‌دهندهٔ تعداد کل آزمون‌شوندگان است. به عنوان مثال، اگر تعداد پاسخ‌دهندگان به یک سؤال ۸۰ نفر باشند و از این تعداد ۶۵ نفر به آن سؤال جواب درست داده باشند ضریب دشواری این سؤال برابر خواهد بود با:

$$P = ۱۰۰ \frac{۶۵}{۸۰} = ۸۱$$

در مواردی که تعداد آزمون‌شوندگان (تعداد برگه‌های آزمون) زیاد است و اطلاعات ما به نحوهٔ پاسخدهی افراد گروه بالا و گروه پایین محدود می‌شود، لازم است از فرمول زیر استفاده کنیم.

$$\text{ضریب دشواری سؤال} = P = \frac{\text{انتخابهای درست گروه پایین} + \text{انتخابهای درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} \times ۱۰۰$$

برای محاسبهٔ ضریب دشواری سؤال شکل ۲-۱۷ به نحو زیر عمل می‌کنیم:

$$P = \text{ضریب دشواری سؤال} = \frac{۵ + ۲}{۱۰ + ۱۰} \times ۱۰۰$$

$$= \frac{۷}{۲۰} \times ۱۰۰$$

$$= ۳۵$$

رقم ۳۵ نشان می‌دهد که درصد نسبتاً کمی از دو گروه بالا و پایین پاسخ درست سؤال را برگزیده‌اند. بر اساس اطلاعات مربوط به سؤال می‌توان نتیجه گرفت که این سؤال نسبتاً دشوار است، زیرا حتی نصف گروه انتخابی (بالا و پایین) نتوانسته است به آن جواب درست بدهد (نگاه کنید به شکل ۲-۱۷).

براساس توضیحات بالا، هر اندازه ضریب دشواری یک سؤال بزرگ‌تر (به ۱۰۰ نزدیک‌تر) باشد، آن سؤال آسان‌تر است و هر اندازه که این ضریب کوچک‌تر (به صفر نزدیک‌تر) باشد سؤال دشوارتر است. بنابراین، به جای ضریب دشواری می‌توان از **ضریب آسانی یا سهولت**[1] نام برد. اما معمول این است که به آن ضریب دشواری بگویند.

---

1. facility index

چنان‌که ملاحظه شد، فرمول بالا ضریب دشواری را بر حسب یک عدد صحیح دورقمی نشان می‌دهد. برخی از متخصصان اندازه‌گیری و ارزشیابی ترجیح می‌دهند که این ضریب، مانند ضریب تمیز، که در قسمت بعد توضیح داده خواهد شد نیز بر حسب اعداد اعشاری باشد که تفسیر آن با تفسیر ضریب تمیز همانند صورت گیرد. در این حالت، فرمول محاسبهٔ ضریب دشواری به صورت زیر است:

$$\text{ضریب دشواری} = \frac{\text{نسبت افراد گروه پایین که به سؤال جواب درست داده‌اند} + \text{نسبت افراد گروه بالا که به سؤال جواب درست داده‌اند}}{۲}$$

برای محاسبهٔ ضریب دشواری سؤال شکل ۲-۱۷ با این فرمول، به نحوهٔ زیر عمل می‌کنیم:

$$\text{ضریب دشواری} = \frac{۰/۵ + ۰/۲}{۲}$$

$$= \frac{۰/۷}{۲}$$

$$= ۰/۳۵$$

ملاحظه می‌کنید که ارقام حاصل از دو فرمول برابرند، اما در فرمول اول ضریب حاصل به صورت عدد صحیح نشان داده می‌شود در حالی که در فرمول دوم به صورت عدد اعشاری است.

تفسیر ضریب دشواری

اگر در تفسیر ضریب دشواری یا ضریب سهولت سؤال با آزمونهای وابسته به هنجار سروکار داشته باشیم، می‌توانیم این ضریب را از دیدگاه آماری مورد بررسی قرار دهیم. یکی از انتظارات ما از آزمونهای وابسته به هنجار این است که برای آزمون‌شوندگان مختلف نمراتی به دست بدهد که در طول یک پیوستار پراکنده باشند، و هرچه این پراکندگی بیشتر باشد بهتر است. به سخن دیگر، هر چه واریانس نمرات حاصل از یک آزمون وابسته به هنجار بزرگ‌تر باشد آن آزمون آزمون بهتری است. بنابراین، یکی از راههای قضاوت در بارهٔ مفید بودن سؤالهای یک آزمون آن است که ببینیم سؤال تا چه اندازه به پراکندگی یا واریانس نمرات کمک می‌کند.

واریانس نمرات گروهی پاسخ دهنده به یک آزمون از دو قسمت تشکیل می‌یابد: (۱) واریانس سؤالها و (۲) همبستگیهای بین سؤالها. بنابراین، هر سؤالی که دارای واریانس بزرگ‌تری است و با سایر سؤالهای آزمون همبستگی بیشتری دارد به واریانس کُل آزمون کمک بیشتری می‌کند. در مقابل، سؤالی که با سایر سؤالها همبستگی ندارد و دارای واریانس کوچکی است کمک زیادی به واریانس کل آزمون نمی‌کند. بنابراین، اگر در تحلیل آزمون به انتخاب سؤالهایی بپردازیم که همبستگی زیادی با یکدیگر دارند و در ضمن دارای واریانسهای بزرگ‌تری هستند، در مجموع آزمون بهتری را تدارک دیده‌ایم. اگر همهٔ سؤالهای یک آزمون با یکدیگر همبستگی کامل داشته

باشند و ضریب دشواری همهٔ آنها ۰/۵ باشد، آن گاه نیمی از آزمون‌شوندگان در آزمون نمرهٔ صفر و نیم دیگر نمرهٔ کامل خواهند گرفت.

برای محاسبهٔ واریانس یک سؤال از فرمول زیر استفاده می‌شود:

$$\text{واریانس سؤال} = P \times (1 - P)$$

در فرمول بالا، P برابر است با درصد کسانی که به سؤال جواب درست داده‌اند، یعنی همان ضریب دشواری. یک سؤال زمانی دارای حداقل واریانس است که ضریب دشواری آن ۰ یا ۱ باشد، یعنی هیچ یک از آزمون‌شوندگان به آن سؤال جواب درست نداده باشد یا اینکه همهٔ آزمون‌شوندگان به آن جواب درست داده باشند.

$$\text{واریانس سؤال} = P \times (1 - P)$$
$$= ۰ \times (1 - ۰) = ۰$$
$$= ۱ \times (1 - ۱) = ۰$$

اما زمانی که ضریب دشواری سؤال $P = ۰/۵$ باشد، واریانس سؤال حداکثر خواهد بود.

$$\text{واریانس سؤال} = P \times (1 - P)$$
$$= ۰/۵ \times (1 - ۰/۵)$$
$$= ۰/۵ \times ۰/۵ = ۰/۲۵$$

در نتیجه، از لحاظ انتخاب برای گنجانیدن در فرم نهایی آزمون، سؤالهایی بهتر هستند که ضریب دشواری آنها از ۱ کمتر و از صفر بیشتر و به ۰/۵ نزدیک باشد. البته انتخاب سؤالهایی با ضرایب دشواری مناسب به نوع سؤالهای آزمون مربوط است. در آزمونهای صحیح ـ غلط ضریب دشواری $P = ۰/۵$ زمانی به دست می‌آید که همهٔ آزمودنیها با حدس زدن به سؤال جواب بدهند. آلن[1] و ین[2] (۱۹۷۹) گفته‌اند برای آزمونهای چند گزینه‌ای سطح بهینهٔ دشواری اندکی کمتر از وسط فاصلهٔ بین ۱ و *سطح موفقیت با حدس زدن* است. منظور از سطح موفقیت با حدس زدن عبارت است از نسبت پاسخهای درستی که با حدس زدن به دست می‌آید. بنابراین، برای سؤالهای چهارگزینه‌ای سطح موفقیت با حدس زدن حدود ۰/۲۵ است، و لذا سطح بهینهٔ دشواری در وسط ۰/۲۵ و ۱ یعنی در حدود ۰/۶ خواهد بود.

برای تعیین رقم دقیق معرف ضریب دشواری بهینهٔ سؤال، ابتدا سطح موفقیت ۱۰۰٪ را از سطح عملکرد با حدس زدن (مثلاً ۲۵٪) کم می‌کنیم و نتیجه را به ۲ تقسیم می‌نماییم و نقطهٔ میانی را مشخص می‌کنیم.

$$\text{نقطهٔ میانی} = \frac{۱/۰۰ - ۰/۲۵}{۲} = \frac{۰/۷۵}{۲} = ۰/۳۷۵$$

1. Allen 2. Yen

پس از آن، رقم حاصل از محاسبات بالا (یعنی نقطهٔ میانی) را با سطح مُعرف عملکرد از طریق حدس زدن جمع می‌کنیم:

ضریب دشواری بهینه = عملکرد با حدس زدن + نقطهٔ میانی

۰/۶۲۵ = ۰/۲۵ + ۰/۳۷۵

راه ساده‌تر برای به دست آوردن همان نتیجه این است که عدد ۱ را با عملکرد مبتنی بر حدس‌زدن جمع کنیم و نتیجه را بر ۲ تقسیم نماییم:

$$\frac{1+0/25}{2} = 0/625$$

به طور کلی، ضریبهای دشواری بین ۰/۳ تا ۰/۷ حداکثر اطلاع را در بارهٔ تفاوت بین آزمودنیها به دست می‌دهند. با این حال، در شرایط ویژه، ضریب دشواری فرق می‌کند. مثلاً آزمونی که به صورت آزمون ورودی فوق لیسانس (کارشناسی ارشد) یا دکتری برای انتخاب بهترین داوطلبان به کار می‌رود و حدوداً ۱۰٪ بالای داوطلبان را برمی‌گزیند باید شامل سؤالهای بسیار دشوار باشد. در مقابل، آزمونی که برای انتخاب کودکان برای یک دورهٔ آموزش ترمیمی یا اصلاحی مورد استفاده قرار می‌گیرد باید شامل سؤالهای بسیار آسان باشد. حتی در آزمونهای دشوار نیز آزمون‌ساز ممکن است ترجیح دهد که در آغاز آزمون چند سؤال آسان برای تشویق آزمون‌شوندگان قرار دهد.

محاسبهٔ ضریب تمیز سؤال

برخلاف ضریب دشواری که میزان آسان بودن یا دشوار بودن یک سؤال را برای گروه آزمون‌شوندگان نشان می‌دهد، **ضریب تمیز**[1] که با D نشان داده می‌شود قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون‌شوندگان مشخص می‌کند؛ یعنی معلوم می‌نماید که سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبهٔ ضریب تمیز یک سؤال از فرمول زیر استفاده می‌شود:

$$\text{ضریب تمیز سؤال} = D = \frac{\text{انتخابهای درست گروه بالا} - \text{انتخابهای درست گروه پایین}}{\text{تعداد افراد یک گروه (بالا یا پایین)}}$$

ضریب تمیز سؤال شکل ۲-۱۷ به شرح زیر محاسبه می‌شود:

$$D = \text{ضریب تمیز سؤال} = \frac{۵-۲}{۱۰} = \frac{۳}{۱۰} = ۰/۳$$

1. discriminative index

رقم ۰/۳ نشان می‌دهد که سؤال تا اندازه‌ای توانسته است گروه بالا و گروه پایین را از یکدیگر متمایز کند. یعنی از ۱۰ نفر گروه بالا ۵ نفر به سؤال جواب درست داده‌اند، در حالی که تنها ۲ نفر از ۱۰ نفر گروه پایین به این سؤال جواب درست داده‌اند.

تفسیر ضریب تمیز

هر قدر ضریب تمیز بزرگ‌تر باشد، قوهٔ تمیز آن سؤال بیشتر و هر قدر این ضریب کوچک‌تر باشد قوهٔ تمیز آن کمتر است. مثلاً اگر قوهٔ تمیز سؤالی ۰/۹۰ باشد، آن سؤال آزمون‌شوندگان قوی و ضعیف را خیلی خوب از هم جدا خواهد کرد؛ اما اگر ضریب تمیز سؤالی ۰/۱۰ باشد، آن سؤال از عهدهٔ جداسازی دانش‌آموزان قوی و ضعیف به خوبی بر نخواهد آمد. ضریب تمیز صفر حاکی از این است که آن سؤال به هیچ‌وجه نتوانسته بین گروه قوی و ضعیف تمایز قائل شود.

گاه اتفاق می‌افتد که ضریب تمیز سؤالی منفی است. این نوع ضریب تمیز نشان می‌دهد که در آن سؤال گروه قوی بدتر از گروه ضعیف عمل کرده است. این گونه سؤالها دارای معایب اساسی هستند که یا باید به کلی کنار گذاشته شوند یا در آنها تجدید نظر اساسی صورت پذیرد. اگر سؤال دارای اشکال فنی نباشد، دلیل چنین ضریبی می‌تواند این باشد که افراد گروه بالا یا آن سؤال را به طور کامل یاد نگرفته‌اند یا آن را به غلط آموخته‌اند. بنابراین، تمامی ضریبهای تمیز سؤالهای آزمون باید مثبت باشند.

لازم به گفتن است که وجود ضریب تمیز منفی در آزمونهای توانایی نامطلوب است، اما "در ابزارهای سنجشِ شخصیت، علاقه، و نگرش ضریب تمیز منفی مشکل‌زا نیست. در این گونه آزمونها، تمیز دادن بین سنخها یا گروههای مختلف کاری مطلوب است و سؤالهای دارای ضریب تمیز بالا (چه مثبت و چه منفی) در تمایز قائل شدن بین آنها کمک می‌کنند" (کلاین، ۲۰۰۵، ص ۹۸).

همبستگی سؤال با کُل آزمون

بر اساس توضیحات بالا، می‌توان دید که ضریب تمیز سؤال نوعی ضریب توافق سؤال با کُل آزمون است. سؤالی را در نظر بگیرید که در آن نمرهای افرادی که در کل آزمون نمره‌های بالایی گرفته‌اند بالا است و کسانی که در کُل آزمون نمره‌های پایینی گرفته‌اند در آن سؤال نمره‌های‌شان پایین است. این سؤال با کل آزمون همخوانی دارد و ضریب تمیز آن مثبت و بالاست. در مقابل سؤال دیگری را در نظر بگیرید که در آن افراد قوی (یعنی کسانی که در کل آزمون نمرات بالا گرفته‌اند) نمرات ضعیفی کسب کرده‌اند و افراد ضعیف (یعنی کسانی که در کل آزمون نمره‌های بالایی نگرفته‌اند) نمره‌شان بالاست. این سؤال دارای ضریب تمیز منفی است. بنا بر آنچه گفته شد، سؤالهای دارای قوهٔ تمیز خوب با کل آزمون همبسته‌اند، اما سؤالهای دارای قوهٔ تمیز صفر یا منفی یا با کل آزمون همبسته نیستند یا با آن همبستگی منفی دارند.

پس یکی دیگر از راههای محاسبهٔ ضریب تمیز سؤالهای یک آزمون محاسبهٔ ضریب همبستگی بین نمرات آزمون‌شوندگان در آن سؤالها و نمرات آنان در کل آزمون است. این ضریب همبستگی که معمولاً به صورت **همبستگی دورشته‌ای نقطه‌ای**[1] محاسبه می‌شود در واقع ضریب توافق سؤال با کل آزمون است[2]. هرچه این ضریب همبستگی بیشتر باشد، بدان معنی است که سؤال با مجموعه سؤالهای دیگر همخوانی بیشتری دارد، و لذا بین افراد قوی و ضعیف بیشتر تمیز قائل می‌شود؛ و هرچه این ضریب کمتر باشد مقدار این همخوانی کمتر است، و لذا آن سؤال بین افراد قوی و ضعیف کمتر تمیز قائل می‌شود. اگر ضریب همبستگی سؤالی با کل آزمون منفی باشد، بدان معنی است که سؤال با کل آزمون ناهمخوانی دارد، یعنی چیزی را می‌سنجد که با آنچه بقیهٔ سؤالهای آزمون می‌سنجند متفاوت است، و لذا بین افراد قوی و ضعیف به طور منفی تمیز قائل می‌شود. و بالاخره اگر ضریب همبستگی سؤالی با کل آزمون صفر باشد، بدان معنی است که این سؤال نمی‌تواند بین افراد قوی و ضعیف تمیز قائل شود.

از امتیازهای استفاده از ضریب همبستگی بین سؤال و کل آزمون به عنوان ضریب تمیز، یکی این است که می‌توان این ضریب را از لحاظ معنی‌داری آماری آزمون کرد. دوم اینکه، از لحاظ عملی می‌توان معنی‌داری این ضریب را مورد قضاوت قرار داد. برای مثال، اگر ضریب همبستگی بین یک سؤال و کل آزمون ۰/۴۰+ باشد، می‌دانیم که این سؤال ۱۶ درصد ($۱۶ = ۴^۲$) واریانس نمرات را به خود اختصاص می‌دهد. در حالی که ضریب تمیز به‌دست آمده از فرمول محاسبهٔ ضریب تمیز با توجه به گروه‌های بالا و پایین را نمی‌توان این‌گونه تفسیر کرد. از اینها گذشته، چنان‌که در فصل ۲۱ خواهیم دید، همبستگی بین سؤال و آزمون مستقیماً به پایایی آزمون مربوط است.

همبستگی میان سؤالهای آزمون

بهتر است به هنگام تحلیل آزمون، علاوه بر محاسبهٔ ضریب همبستگی بین هر سؤال و کل آزمون، ضریب همبستگی تک تک سؤالها را با یکدیگر نیز تعیین کنیم. این کار چند فایده دارد: نخست اینکه، با در دست داشتن تعداد سؤالهای آزمون و میانگین ضرایب همبستگی بین سؤالها می‌توان ضریب پایایی آزمون را محاسبه کرد. دوم اینکه، با استفاده از این نوع همبستگی می‌توان به مفهوم قوهٔ تمیز سؤال دست یافت. به سخن دیگر، با استفاده از همبستگی بین سؤالهای یک آزمون می‌توان فهمید که چرا بعضی از سؤالها توانایی متمایز ساختن بین آزمون شوندگان قوی و ضعیف را ندارند.

اگر همبستگی بین یک سؤال و کل آزمون کم باشد، دو دلیل برای آن وجود دارد: یکی اینکه ممکن است سؤال مورد نظر با هیچ‌کدام از دیگر سؤالهای آزمون همبسته نباشد. در این حالت، یا

---

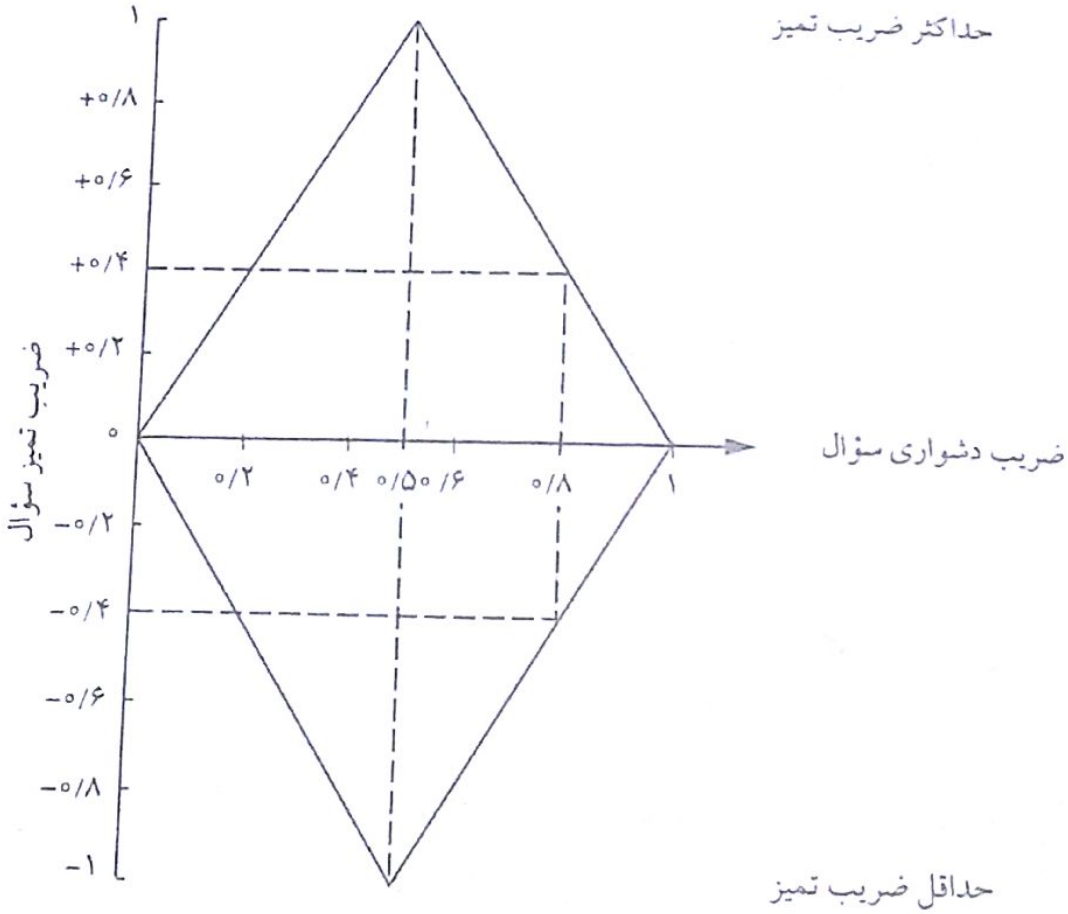
1. point biserial correlation

۲. برای چگونگی محاسبهٔ این ضریب همبستگی، نگاه کنید به فصل ۱۸ کتاب.

باید سؤال را بازنویسی کرد یا آن را کلاً به کناری نهاد. امکان دوم این است که سؤال با بعضی از سؤالهای آزمون همبستگی قابل ملاحظه‌ای نشان می‌دهد، اما با دیگر سؤالها همبستگی نزدیک به صفر یا حتی منفی دارد. مثلاً در جایی که آزمون دو ویژگی کاملاً متمایز را می‌سنجد چنین وضعی ممکن است اتفاق افتد. نمونه‌ای برای این آزمون یک آزمون ریاضی است که شامل تعدادی مسئله است که با عبارات کلامی نسبتاً پیچیده‌ای طرح شده‌اند و لذا آزمون هم درک کلامی و هم توانایی ریاضی را می‌سنجد. نظریهٔ پایایی می‌گوید این‌گونه آزمونها که سؤالهای آنها چیزهای مختلفی را اندازه می‌گیرند (و بنابراین با هم همبسته نیستند) اندازه‌های باثباتی را به دست نمی‌دهند (مرفی و دیویدشوفر، ۱۹۹۴).

رابطهٔ بین ضریب دشواری و ضریب تمیز

شکل ۳-۱۷ رابطهٔ بین ضریب دشواری و ضریب تمیز سؤال را نشان می‌دهد. حدود چهارضلعی این شکل حداکثر و حداقل ضرایب دشواری و تمیز را نشان می‌دهد. برای مثال، اگر ضریب دشواری برابر ۰/۸ باشد، ضریب تمیز بین ۰/۴+ و ۰/۴- خواهد بود. دقت کنید که تنها در صورتی ضریب تمیز به حداکثر یا ۱+ می‌رسد که ضریب دشواری برابر ۰/۵ یعنی در حد وسط باشد. در نتیجه سؤالهای خوب یک آزمون سؤالهایی هستند با ضریب دشواری متوسط و ضریب تمیز بزرگ.



شکل ۳-۱۷ رابطهٔ ضریب تمیز با ضریب دشواری

تحلیل گزینه‌های انحرافی

در تحلیل سؤالهای آزمون، علاوه بر تعیین ضریبهای دشواری و تمیز برای هر سؤال، بررسی نحوهٔ پراکندگی پاسخهای مربوط به گزینه‌های انحرافی هر سؤال نیز ضروری است. پیشتر در این کتاب گفتیم که هدف از قرار دادن گزینه‌های انحرافی در سؤالها منحرف کردن آزمون‌شوندگانی است که جواب درست سؤال را نمی‌دانند. بنابراین، گزینه‌های انحرافی خوب باید عملاً از عهدهٔ این کار برآیند. برای تعیین این مطلب باید به گزینه‌هایی که توسط آزمون‌شوندگان انتخاب شده‌اند مراجعه کنیم.

به منظور بررسی گزینه‌های انحرافی سؤال شکل ۲ـ۱۷، با مراجعه به اطلاعات داده شده معلوم می‌شود که گزینهٔ انحرافی **الف** گزینهٔ انحرافی بسیار خوبی است، زیرا از گروه قوی هیچ‌کس آن را انتخاب نکرده است اما از گروه ضعیف ۵ نفر آن را برگزیده‌اند. گزینه‌های انحرافی **پ** و **ت** کارشان را به خوبی انجام نداده‌اند. گزینهٔ **ت** به کلی بی‌فایده است و بود و نبود آن فرقی نمی‌کند، زیرا هیچ یک از افراد دو گروه را به خود جلب نکرده است. گزینهٔ **پ** نیز برای هر دو گروه یکسان عمل کرده است. در تجدید نظر و اصلاح این سؤال، به منظور استفاده‌های بعدی، لازم است گزینهٔ **د** با گزینهٔ تازه‌ای عوض شود و گزینهٔ **پ** نیز اصلاح گردد.

در تحلیل گزینه‌های انحرافی قاعدهٔ کلی به شرح زیر است: هر گزینهٔ انحرافی دست‌کم باید یک نفر از افراد گروه ضعیف را به خود جلب کند و اگر هر دو گروه قوی و ضعیف را به خود جلب می‌کند تعداد افراد گروه ضعیف باید بیشتر از تعداد افراد گروه قوی باشد. دلیل این امر را این گونه می‌توان توضیح داد که افراد گروه ضعیف در مجموع اطلاعات کمتری از مطالب آزمون دارند. بنابراین، در صورتی یک سؤال به خوبی عمل می‌کند که افراد ضعیف بیشتر از افراد گروه قوی گزینه‌های انحرافی آن سؤال را انتخاب نمایند.

تحلیل سؤالهای آزمونهای تشریحی و آزمونهای عملکردی

هرچند که تحلیل سؤال بیشتر برای آزمونهای عینی متداول است، با این حال سؤالهای غیرعینی را نیز پس از اجرا باید مورد تجزیه و تحلیل قرار داد و اشکالهای آنها را برطرف کرد. ما در اینجا دو روش معروف تحلیل آزمونهای غیرعینی را توضیح می‌دهیم.

یکی از روشهای تحلیل سؤالهای آزمونهای تشریحی و عملکردی روش ویتنی و سیبرز (۱۹۷۰) است. این روش را همراه با یک مثال در زیر توضیح می‌دهیم.

۱. دو گروه ۲۵ درصدی بالا و پایین از کل دانش‌آموزان را به شرحی که قبلاً گذشت تعیین کنید.
۲. برای هر سؤال مجموع نمرات گروه بالا و گروه پایین را حساب کنید.
۳. برای محاسبهٔ ضریبهای دشواری و تمیز نتایج را در فرمولهای زیر قرار دهید.

جدول ۱-۱۷ اطلاعات مورد نیاز برای تحلیل سؤال تشریحی

نمرهٔ سؤال	گروه بالا		گروه پایین	
s	f	fs	f	fs
۵	۸	۴۰	۳	۱۵
۴	۵	۲۰	۷	۲۸
۳	۶	۱۸	۸	۲۴
۲	۴	۸	۶	۱۲
۱	۷	۷	۲	۲
۰	۰	۰	۴	۰
	۳۰	۹۳	۳۰	۸۱

ضریب دشواری $P = \frac{\Sigma_H + \Sigma_L - (۲N\ Score_{min.})}{۲N\ (Score_{max.} - Score_{min.})}$

ضریب تمیز $D = \frac{\Sigma_H - \Sigma_L}{N\ (Score_{max.} - Score_{min.})}$

در این فرمولها:

Σ_H = مجموع نمره‌های افراد یا برگه‌های گروه بالا (۲۵ درصد)

Σ_L = مجموع نمره‌های افراد یا برگه‌های گروه پایین (۲۵ درصد)

N = تعداد افراد یا برگه‌های ۲۵ درصد گروه موردنظر

$Score_{max.}$ = بالاترین نمرهٔ ممکن در سؤال

$Score_{min.}$ = پایین‌ترین نمرهٔ ممکن در سؤال

مثال:

برای یکی از سؤالهای یک آزمون تشریحی اطلاعات جدول ۱-۱۷ به دست آمده است. با استفاده از اطلاعات بالا می‌توان ضریبهای دشواری و تمیز سؤال فوق را به روش زیر محاسبه کرد.

$$\text{ضریب دشواری} = P = \frac{۹۳ + ۸۱ - (۲ \times ۳۰ \times ۰)}{۲ \times ۳۰ (۵ - ۰)} = \frac{۱۷۴}{۳۰۰} = ۰/۵۸$$

$$\text{ضریب تمیز} = D = \frac{۹۳ - ۸۱}{۳۰ (۵ - ۰)} = \frac{۱۲}{۱۵۰} = ۰/۰۸$$

علاوه بر روش بالا، نیتکو (۲۰۰۱) نیز، برای تحلیل سؤال آزمونهای تشریحی و عملکردی، روش دیگری را معرفی کرده که ما در اینجا به توضیح آن می‌پردازیم. در این روش، ضریب دشواری یک سؤال تشریحی یا عملکردی به صورت نمرهٔ میانگین گروه آزمون‌شوندگان برای آن سؤال تعریف شده است. دقت کنید که نمرهٔ میانگین سؤالی که به صورت ۰ و ۱ نمره گذاری می‌شود برابر

است با ارزش P آن که از فرمول $P = \frac{R}{T}$ محاسبه می‌شود. از این رو، تعجب‌آور نیست که ضریب دشواری سؤال آزمونهای تشریحی و عملکردی هردو به صورت نمرهٔ میانگین سؤال تعریف شده است.

برای آشنا شدن با چگونگی محاسبهٔ ضریب دشواری سؤالهای تشریحی و عملکردی با روش نیتکو (۲۰۰۱) فرض کنید سؤال شمارهٔ ۱ یک آزمون دارای ارزش از ۱ تا ۶ نمره است. اگر نمرهٔ میانگین این سؤال ۴/۲ باشد، آن‌گاه ضریب دشواری این سؤال برابر با ۴/۲ است.

برای اینکه ضریب دشواری به دست آمده از روش بالا را به مقیاس ضریب دشواری سؤالهای دارای ارزش ۰ و ۱، یعنی P، تبدیل نماییم باید این ضریب را به دامنهٔ ممکن نمرات آن سؤال تقسیم کنیم. حاصل این عملیات رقمی بین ۰ و ۱ است که دارای دامنهٔ ضریبهای دشواری حاصل از فرمول $P = \frac{R}{T}$ خواهد بود. فرمول این محاسبات به صورت زیر نشان داده می‌شود:

$$\text{ضریب دشواری } P = \frac{\text{نمرهٔ میانگین سؤال}}{\text{دامنهٔ ممکن نمرات سؤال}} = \frac{\text{نمرهٔ میانگین سؤال}}{\text{پایین‌ترین نمرهٔ ممکن - بالاترین نمرهٔ ممکن}}$$

با استفاده از این فرمول، ضریب دشواری سؤالی که دارای دامنهٔ نمرات ۱ تا ۶ و نمرهٔ میانگین ۴/۲ است به صورت زیر محاسبه خواهد شد:

$$P = \frac{۴/۲}{۶-۱} = \frac{۴/۲}{۵} = ۰/۸۴$$

در تفسیر ضریب دشواری $P = ۰/۸۴$، می‌گوییم که به طور متوسط پاسخ‌دهندگان ۸۴٪ بالاترین نمرهٔ ممکن را در سؤال به دست آورده‌اند (که البته حاکی از یک سؤال نسبتاً ساده است). چنانچه دامنهٔ نمرات ممکن برای این سؤال ۰ تا ۱۰ بود، آن‌گاه ضریب دشواری به شرح زیر به دست می‌آمد.

$$P = \frac{۴/۲}{۱۰-۰} = \frac{۴/۲}{۱۰} = ۰/۴۲$$

برای محاسبهٔ ضریب تمیز سؤالهای آزمونهای عملکردی و تشریحی با روش نیتکو (۲۰۰۱)، تفاوت بین نمرهٔ میانگین گروه بالا و نمرهٔ میانگین گروه پایین را حساب می‌کنیم. گروه بالا و گروه پایین همان گروههایی هستند که قبلاً معرفی شدند، یعنی کسانی که در آزمون بیشترین نمرات (گروه بالا) و کسانی که کمترین نمرات (گروه پایین) را به دست آورده‌اند.

برای آشنا شدن با چگونگی محاسبهٔ ضریب تمیز سؤال، فرض کنید میانگین نمرات گروه بالا برای یک سؤال ۵/۳ و میانگین نمرات گروه پایین برای همین سؤال ۲/۸ است. همچنین فرض کنید که دامنهٔ نمرات ممکن برای این سؤال از ۱ تا ۶ است. ضریب تمیز این سؤال برابر است با:

$$D = ۵/۳ - ۲/۸ = ۲/۵$$

باز هم برای اینکه این ضریب تمیز را به مقیاس ضریب تمیز سؤالهای دارای ارزش ۰ و ۱ تبدیل نماییم باید این ضریب را به دامنهٔ ممکن سؤال تقسیم کنیم. نتیجه رقمی بین ۱- تا ۱+ خواهد

بود. چگونگی انجام این محاسبات در زیر می‌آید:

$$\text{ضریب تمیز } D = \frac{\text{تفاوت بین نمرات میانگین گروههای بالا و پایین برای سؤال}}{\text{دامنهٔ نمرات ممکن برای سؤال}}$$

$$= \frac{\text{نمرهٔ میانگین سؤال برای گروه پایین} - \text{نمرهٔ میانگین سؤال برای گروه بالا}}{\text{پایین‌ترین نمرهٔ ممکن سؤال} - \text{بالاترین نمرهٔ ممکن سؤال}}$$

$$= \frac{۵/۳ - ۲/۸}{۶ - ۱} = \frac{۲/۵}{۵} = ۰/۵$$

از آنجا که ما تفاوت بین نمرات میانگین گروههای بالا و پایین را به دامنهٔ ممکن نمرات سؤال تقسیم می‌کنیم، می‌توانیم نتیجهٔ حاصل را این‌گونه تفسیر نماییم که تفاوت بین نمرات میانگین گروههای بالا و پایین برای سؤال فوق ۵۰٪ دامنهٔ ممکن سؤال است. این رقم ضریب تمیز نسبتاً خوبی است. ضریب تمیز می‌تواند مثبت، منفی، یا صفر باشد. چنان‌که قبلاً دیدیم، اگر ضریب تمیز منفی باشد، بدین معنی است که گروه پایین به طور متوسط در سؤال موردنظر بهتر از گروه بالا در این سؤال عمل کرده است، که نشان‌دهندهٔ ضعف سؤال است.

تجدید نظر و اصلاح آزمون

پس از تعیین ضریبهای دشواری و تمیز همهٔ سؤالهای یک آزمون و بررسی گزینه‌های هر سؤال، باید به اصلاح سؤالها و بازنویسی آنهایی که نیاز به تغییر دارند اقدام کرد. سؤالهایی که ضریب دشواری آنها بسیار بالا یا بسیار پایین است یا ضریب تمیز آنها خیلی کم است باید مورد تجدید نظر قرار گیرند. همچنین گزینه‌های انحرافی معیوب نیز باید اصلاح یا عوض شوند.

تحلیل سؤالهای آزمونهای وابسته به ملاک

لازم به ذکر است که استفاده از ضریبهای دشواری و تمیز به شرحی که گذشت صرفاً برای آزمونهای وابسته به هنجار مناسب است، اما در آزمونهای وابسته به ملاک این ضریبها مورد استفاده‌ای ندارند. چنان‌که قبلاً در این کتاب گفته شد، نتیجهٔ آرمانی آزمونهای وابسته به ملاک این است که در آنها اکثر آزمون‌شوندگان اکثر سؤالها را درست جواب دهند. در چنین حالتی ضریب دشواری سؤالها بسیار بزرگ و ضریب تمیز آنها بسیار کوچک خواهد بود، یعنی برای گروه آزمودنی مورد نظر اکثر سؤالها آسان و قدرت تمیز آنها بسیار کم خواهد بود.

تمیز بین پیش و پس از آموزش

بعضی از متخصصان ارزشیابی پیشرفت تحصیلی پیشنهاد می‌کنند که آزمونهای وابسته به ملاک

مورد استفاده در سنجش **یادگیری در حد تسلط**[1] تنها باید شامل سؤالهایی باشند که پس از آموزش مطالب درسی آسان‌تر از پیش از آموزش آن مطالب جلوه کنند. یعنی اینکه نشان دهند سؤالهای امتحان تا چه اندازه تأثیر آموزش را می‌سنجند. برای تحقق این امر انجام مراحل زیر پیشنهاد شده است:

۱. اجرای آزمون پیش از آموزش و پس از آن
۲. محاسبهٔ ضریب دشواری هر سؤال برای پیش‌آزمون و پس‌آزمون
۳. محاسبهٔ تفاوت ضرایب دشواری سؤالهای پیش‌آزمون و پس‌آزمون

این ضریب تمیز را ضریب **تمیز پیش و پس از آموزش**[2] می‌نامند و آن را با D نشان می‌دهند (ویرسما و جورس، ۱۹۹۰). نام دیگر این ضریب، ضریب **حساسیت نسبت به آثار آموزش**[3] است و آن را با S معرفی می‌کنند. این ضریب با فرمول زیر محاسبه می‌شود:

$$D \text{ یا } S = \frac{R_A - R_B}{T}$$

در این فرمول:

$R_A = PA$ تعداد کسانی که پس از آموزش به سؤال جواب درست داده‌اند

$R_B = PB$ تعداد کسانی که پیش از آموزش به سؤال جواب درست داده‌اند

$T =$ تعداد کسانی که هم پیش و هم پس از آموزش به سؤال جواب داده‌اند، یعنی تعداد کل افراد

برای اینکه با چگونگی استفاده از این فرمول و محاسبهٔ ضریب D یا S آشنا شوید، از داده‌های جدول ۲-۱۷ استفاده می‌کنیم. داده‌های این جدول نمرات ۶ دانش‌آموز در یک آزمون ۵ سؤالی هم پیش از آموزش (پیش‌آزمون) و هم پس از آموزش (پس آزمون) است. عدد ۱ نشان دهندهٔ این است که دانش‌آموز سؤال را درست جواب داده است و عدد صفر معرف این است که دانش‌آموز سؤال را درست جواب نداده است. حرف B معرف نمرهٔ پیش‌آزمون (پیش از آموزش) و حرف A نشان دهندهٔ نمرهٔ پس‌آزمون (پس از آموزش) است.

با استفاده از داده‌های جدول ۲-۱۷ و قرار دادن آنها در فرمول ضریب پیش و پس از آموزش، برای سؤالهای این جدول ضرایب D یا S به نحو زیر به دست می‌آیند.

---

۱. mastery learning [یادگیری مطالب آموزش داده شده تا یک حد معین؛ مثلاً اینکه معلم از پیش تعیین می‌کند که یادگیرندگان دست کم ۸۵ یا ۹۰ درصد مطالب آموزش داده شده را یاد بگیرند. برای کسب اطلاعات لازم در این باره به کتاب ویژگیهای آدمی و یادگیری آموزشگاهی، تألیف بنجامین اس. بلوم (۱۹۸۲، ترجمهٔ سیف، ۱۳۶۳)، و کتاب روانشناسی پرورشی نوین: روانشناسی یادگیری و آموزش، تألیف سیف (۱۳۸۷)، مراجعه کنید.]

2. pre and post discrimination

3. sensivity to instructional effects

جدول ۲-۱۷ عملکرد ۶ دانش‌آموز در یک آزمون هم پیش از آموزش (B) و هم پس از آموزش (A) (گرانلاند و لین، ۱۹۹۰)

	سؤالها←	۱		۲		۳		۴		۵	
	پیش‌آزمون (B)										
	پس آزمون (A)	A	B	A	B	A	B	A	B	A	B
دانش‌آموزان	الف	۱	۰	۱	۱	۰	۰	۰	۱	۱	۰
	ب	۱	۰	۱	۱	۰	۰	۰	۱	۱	۱
	پ	۱	۰	۱	۱	۰	۰	۰	۱	۱	۰
	ت	۱	۰	۱	۱	۰	۰	۰	۱	۱	۰
	ث	۱	۰	۱	۱	۰	۰	۰	۱	۱	۱
	ج	۱	۰	۱	۱	۰	۰	۰	۱	۰	۰

سؤال ۱ D یا $S = \frac{۶-۰}{۶} = ۱$

سؤال ۲ D یا $S = \frac{۶-۶}{۶} = ۰$

سؤال ۳ D یا $S = \frac{۰-۰}{۶} = ۰$

سؤال ۴ D یا $S = \frac{۰-۶}{۶} = -۱$

سؤال ۵ D یا $S = \frac{۵-۲}{۶} = ۰/۵$

با توجه به داده‌های جدول ۲-۱۷ و ضرایب تمیز محاسبه شده برای سؤالها، می‌توان هر سؤال را در ارتباط با اثربخشی آن به عنوان وسیله‌ای برای قضاوت در بارهٔ آثار آموزش به نحو زیر موردبحث قرار داد (گرانلاند و لین، ۱۹۹۰):

سؤال ۱. سؤال بسیار خوبی برای یک آزمون وابسته به ملاک مربوط به یادگیری در حد تسلط است، زیرا همهٔ دانش‌آموزان پیش از آموزش آن را غلط جواب دادند اما پس از آموزش به آن جواب درست دادند.

سؤال ۲. این سؤال بسیار *آسان* است و نتوانسته آثار آموزش را بسنجد، زیرا همهٔ دانش‌آموزان پیش و پس از آموزش به آن جواب درست داده‌اند.

سؤال ۳. این سؤال خیلی *دشوار* است و نتوانسته است آثار آموزش را اندازه بگیرد، یا اینکه آموزش در رابطه با موضوع این سؤال ناکافی بوده است.

سؤال ۴. این یک سؤال بسیار معیوب است یا یک سؤال آسان است که با آموزش غلط دنبال شده است.

سؤال ۵. این یک سؤال کارآمد است با یک الگوی نسبتاً طبیعی، زیرا بعضی از دانش‌آموزان پیش از آموزش به آن جواب درست داده‌اند، اما تعداد بیشتری از دانش‌آموزان پس از آموزش به آن جواب درست داده‌اند.

مانند ضریب تمیز آزمونهای وابسته به هنجار، در اینجا نیز ضریب تمیز از ۱- تا ۱+ در تغییر است. سؤالهای دارای ضریبهای تمیز مثبت به کیفیت خوب کل آزمون کمک می‌کنند. وقتی که ضریب تمیز سؤالی صفر یا نزدیک به صفر است نشان می‌دهد که سؤال به قدرت تمیز کل آزمون هیچ کمکی نمی‌کند و آن سؤالی است که همهٔ آزمون‌شوندگان در پیش‌آزمون و پس‌آزمون به آن جواب درست یا غلط داده‌اند.

شرایط آرمانی برای هر سؤال آن است که تفاوت ضریبهای دشواری قبل و بعد از آموزش آن (ضریب حساسیت نسبت به آثار آموزش) ۱+ باشد، یعنی ضریب دشواری سؤال در پیش‌آزمون صفر و ضریب دشواری آن در پس‌آزمون ۱+ باشد، مانند سؤال شمارهٔ ۱ جدول ۲-۱۷. این بدان معنی است که پیش از آموزش هیچ یک از دانش‌آموزان جواب سؤال را نمی‌دانسته است، اما در نتیجهٔ آموزش همهٔ دانش‌آموزان جواب آن سؤال را یاد گرفته‌اند. بنابراین، برای استفاده‌های بعدی سؤالهایی که دارای تفاوت ضریبهای دشواری قبلی و بعدی زیاد یا نزدیک به رقم ۱+ هستند باید انتخاب گردند.

با وجود این، نیتکو (۱۹۸۳) معتقد است که وقتی هدف از تهیهٔ یک آزمون تفسیر پیشرفت مطلق آزمون‌شوندگان است، استفاده از شاخص بالا (تفاوت بین ضریبهای دشواری قبل و بعد از آموزش) کار مفیدی نیست. به اعتقاد او، از آنجاکه محتوای سؤالها در حوزهٔ مورد نظر مهم است، انتخاب تعدادی از سؤالها و کنار گذاشتن تعدادی دیگر مجموعه سؤالهایی را که معرف محتوای مورد نظر باشد به دست نمی‌دهد. همچنین تنها انتخاب سؤالهایی که از نظر تفاوت ضریب دشواری قبل و بعد از آموزش مثبت هستند سبب انتخاب آن دسته از سؤالهایی می‌شود که آموزش برای آنها اثربخش بوده است. در چنین حالتی، هیچ راهی برای ما وجود ندارد که تعیین کنیم: (۱) چه بخشی از محتوای درسی را دانش‌آموزان نیاموخته‌اند (یعنی موضوع سؤالهایی که تفاوت ضرایب دشواری قبلی و بعدی آنها صفر است)، یا (۲) آموزش چه بخشی از محتوای درس را به حالتی در آورده است که دانش‌آموزان نتوانند به سؤالهای آن جواب درست بدهند (یعنی موضوع سؤالهایی که تفاوت ضرایب دشواری قبل و بعد از آموزش آنها کمتر از صفر است).

تمیز بین حد تسلط و غیر حد تسلط

از آنجاکه آزمونهای وابسته به ملاک دارای یک نمرهٔ مرزی برای حد تسلط هستند، مثلاً نمرهٔ ۸۵ از ۱۰۰، بنابراین لازم است معلوم شود که آیا سؤالهای آزمون در آن نقطه بین آزمون‌شوندگان قوی و

ضعیف تمیز قائل می‌شوند یا نه. برنان[1] (۱۹۷۲) یک ضریب حد تسلط ـ غیر حد تسلط به دست داده که برای منظور فعلی ما مفید است. برای محاسبهٔ این ضریب یک شکل چهارخانه‌ای شبیه شکل ۴-۱۷ درست می‌شود که در خانه‌های عمودی آن حد تسلط و غیر حد تسلط و در خانه‌های افقی آن درست و غلط قرار داده می‌شوند.

در شکل ۴-۱۷، حرف a معرف تعداد آزمون‌شوندگانی است که سؤال را درست جواب داده‌اند اما در کل آزمون نمره‌ای کمتر از حد تسلط گرفته‌اند. همچنین b تعداد آزمون‌شوندگانی است که سؤال را درست جواب داده‌اند و در نمرهٔ کل آزمون نیز به حد تسلط رسیده‌اند. حرف c نشان دهندهٔ تعداد آزمون‌شوندگانی است که سؤال را غلط جواب داده‌اند و در کل به حد تسلط نرسیده‌اند. و بالاخره حرف d معرف تعداد کسانی است که سؤال مورد نظر را درست جواب نداده‌اند اما در نمرهٔ کل آزمون به حد تسلط رسیده‌اند. تعداد آزمون‌شوندگانی که به حد تسلط رسیده‌اند برابر است با b+d و تعداد کسانی که به حد تسلط نرسیده‌اند مساوی است با a+c.

ضریب تمیز، یعنی D، برابر است با تفاوت بین سطوح دشواری سؤال برای کسانی که به‌حد تسلط رسیده و آنهایی که به حد تسلط نرسیده‌اند، مطابق با فرمول زیر:

$$\text{تمیز بین حد تسلط و غیرحد تسلط} = D = \frac{b}{b+d} - \frac{a}{a+c}$$

این ضریب تمیز از ۱- تا ۱+ متغیر است. مقدار مثبت D نشان می‌دهد که سؤال بین کسانی که به حد تسلط رسیده و کسانی که به حد تسلط نرسیده‌اند تمیز قائل شده است. وقتی که ضریب تمیز سؤالی منفی است، نشان دهندهٔ آن است که سؤال در جهت خلاف سؤالهای دیگر تمیز قائل می‌شود. چنین سؤالی نیاز به بازبینی دارد. ممکن است معلم، پس از بازبینی، آن را به کلی کنار بگذارد. ممکن است لازم باشد سؤال را با افراد کلاس در میان بگذارد تا در بارهٔ آن بحث کنند و علت مشکل را بیابند. ممکن است کلید تصحیح اشتباه بوده باشد. در هر صورت، ضریب تمیز منفی نشان دهندهٔ مشکلی در سؤال است که باید مورد توجه و رسیدگی قرار گیرد.

سؤال	آزمون: a+c (غیرحدتسلط)	آزمون: b+d (حدتسلط)
درست	a	b
غلط	c	d

شکل ۴-۱۷ اطلاعات موردنیاز برای محاسبهٔ ضریب تمیز حد تسلط ـ غیر حد تسلط

1. Brennan

مثالی برای تحلیل سؤال یک آزمون وابسته به ملاک

برای روشن‌تر ساختن نکات بالا، مثال کاملی برای تحلیل یک آزمون ملاکی توضیح می‌دهیم. فرض کنید اطلاعات جدول ۳-۱۷ نمرات حاصل از یک آزمون ۸ سؤالی وابسته به ملاک است که با ۱۰ نفر دانش‌آموز اجرا شده است. در این جدول، نمرهٔ ۱ نشان دهندهٔ این است که دانش‌آموز سؤال را درست جواب داده است و نمرهٔ صفر معرف جواب غلط دانش‌آموز است. نمرهٔ کل دانش‌آموز برابر است با مجموع جوابهای درست.

در پایین جدول ۳-۱۷ ضریبهای دشواری و تمیز همهٔ سؤالها محاسبه شده‌اند. چنان‌که قبلاً گفتیم، ضریب دشواری عبارت است از درصد کُل دانش‌آموزانی که سؤال را درست جواب داده‌اند. در مثال بالا، همهٔ دانش‌آموزان سؤال شمارهٔ ۱ را درست جواب داده‌اند، بنابراین ضریب P آن ۱ است. از ۱۰ نفر دانش‌آموز، ۹ نفر سؤال شمارهٔ ۲ را درست جواب داده‌اند؛ بنابراین P آن برابر است با ۰/۹. بقیهٔ ضریبهای دشواری نیز به همین ترتیب حساب شده‌اند. دقت کنید که ضریب دشواری سؤال شمارهٔ ۶ بسیار کوچک‌تر از بقیهٔ سؤالهاست. این سؤال باید مورد بازبینی قرار گیرد تا علت دشواری بیش از اندازهٔ آن معلوم گردد. شاید محتوای آن با محتوای بقیهٔ سؤالها قدری متفاوت است یا شاید اشکال دیگری دارد. می‌توان از دانش‌آموزان سؤال کرد که چرا با آن سؤال بیشتر از بقیه مشکل داشته‌اند.

جدول ۳-۱۷ نمرات مربوط به سؤالهای یک آزمون وابسته به ملاک (ویرسما و جورس، ۱۹۹۰، ص ۲۴۷)

دانش‌آموزان	سؤالها ۱	۲	۳	۴	۵	۶	۷	۸	نمره کل
الف	۱	۱	۱	۱	۱	۰	۱	۱	۷
ب	۱	۱	۰	۱	۱	۰	۰	۰	۴
پ	۱	۱	۱	۱	۱	۱	۱	۱	۸
ت	۱	۱	۱	۱	۰	۱	۱	۱	۷
ث	۱	۰	۱	۱	۱	۰	۰	۰	۴
ج	۱	۱	۰	۱	۰	۰	۱	۱	۵
چ	۱	۱	۱	۱	۱	۱	۱	۱	۸
ح	۱	۱	۰	۰	۱	۰	۰	۰	۳
خ	۱	۱	۱	۱	۱	۱	۱	۱	۸
د	۱	۱	۱	۱	۱	۰	۱	۱	۷
p = ضریب دشواری	۱/۰	۰/۹	۰/۷	۰/۹	۰/۸	۰/۴	۰/۷	۰/۷	
D = ضریب تمیز	۰/۰	۰/۲۵	۰/۷۵	۰/۲۵	۰/۰۸	۰/۶۷	۰/۷۵	۰/۷۵	

حال فرض کنید که این آزمون دارای یک حد تسلط ۹۰ از صد بوده است. یعنی از پیش به دانش‌آموزان گفته شده است که باید در آزمون دست‌کم ۹۰ نمره از ۱۰۰ نمره را دریافت کنند یا ۹۰ درصد ۸ سؤال را درست جواب دهند. بنابر این مرز حد تسلط، دانش‌آموزانی که دارای نمرهٔ ۷ یا بیشتر بوده‌اند به حد تسلط رسیده‌اند. بر اساس اطلاعات موجود در جدول ۳-۱۷، از ۱۰ نفر دانش‌آموز ۶ نفر به این حد از تسلط دست یافته‌اند.

حال به منظور اینکه ضریب تمیز بین حد تسلط و غیر حد تسلط را برای سؤالها محاسبه کنیم، اطلاعات موردنیاز برای این محاسبه را طبق شکل ۴-۱۷ برای هریک از سؤالها در فرمول زیر قرار می‌دهیم.

$$D = \frac{b}{b+d} - \frac{a}{a+c}$$

برای نمونه ضریب تمیز مربوط به سؤال شمارهٔ ۳ را حساب می‌کنیم. اطلاعات موردنیاز این سؤال در شکل ۵-۱۷ داده شده است.

$$D = \frac{۶}{۶+۰} - \frac{۱}{۱+۳}$$

$$= ۱ - ۰/۲۵$$

$$= ۰/۷۵$$

چنان‌که از شکل ۵-۱۷ پیداست، از ۶ نفر دانش‌آموزی که به حد تسلط رسیده‌اند، هر ۶ نفر سؤال را درست جواب داده و هیچ کس آن را غلط جواب نداده است. همچنین از ۴ نفر دانش‌آموزی که به حد تسلط نرسیده‌اند ۱ نفر سؤال را درست و ۳ نفر آن را غلط جواب داده‌اند. ضریب محاسبه شده برای قدرت تمیز سؤال ۳ برابر است با ۰/۷۵+ که نشان می‌دهد این سؤال به‌خوبی و به صورت مثبت بین گروهی که به حد تسلط رسیده و گروهی که به این حد نرسیده‌اند تمیز قائل شده است. یعنی در جداسازی دانش‌آموزان دو گروه موفق بوده است.

برای محاسبهٔ ضریبهای تمیز سایر سؤالهای آزمون نیز می‌توان از همان روش مورد استفاده با سؤال شمارهٔ ۳ سود جست. این ضریبها برای همهٔ سؤالها در پایین جدول ۳-۱۷ وارد شده‌اند.

بازبینی ضریبهای تمیز سؤالهای جدول ۳-۱۷ نشان می‌دهد که همهٔ آنها بجز سؤال ۱ بالاتر از صفرند. سؤالهای ۳، ۷ و ۸ دارای حداکثر ضریب تمیز هستند. ضریب تمیز سؤال ۱ به این دلیل صفر است

سؤال	آزمون: حد تسلط ۶	آزمون: غیر حد تسلط ۴
درست	۶	۱
غلط	۰	۳

شکل ۵-۱۷ اطلاعات موردنیاز برای محاسبهٔ ضریب تمیز حد تسلط ـ غیر حد تسلط سؤال شمارهٔ ۳ جدول ۳-۱۷

که همهٔ دانش‌آموزان، چه آنهایی که به حد تسلط رسیده و چه آنهایی که به حد تسلط نرسیده‌اند، آن را درست جواب داده‌اند. این موضوع نباید تهیه کنندهٔ سؤال را نگران سازد، زیرا هدف آزمون این بوده است که همهٔ دانش‌آموزان به آن جواب صحیح بدهند. با این حال، از لحاظ تمایزگذاری بین گروه حد تسلط و گروه غیر حد تسلط، این سؤال درست کار نکرده است. ضریب تمیز سؤال ۵ نیز نزدیک به صفر است و دربارهٔ این سؤال نیز مانند سؤال شمارهٔ ۱ می‌توان داوری کرد.

تحلیل آزمون به کمک نظریهٔ سؤال ـ پاسخ

روشهایی که در قسمت پیشین این فصل در رابطه با تحلیل آزمون توضیح داده شدند **روشهای کلاسیک تحلیل آزمون**[1] نام دارند. روشهای کلاسیک یا قدیمی که به روشهای سادهٔ تحلیل آزمون نیز شهرت دارند روشهای کمابیش سرراستی هستند که به‌وسیلهٔ معلمان قابل استفاده‌اند. روشهای پیچیده‌تری برای این منظور به وجود آمده‌اند که با دست قابل استفاده نیستند و کاربرد آنها نیاز به کامپیوتر دارد. این روشهای نسبتاً پیچیدهٔ تحلیل آزمون به **نظریهٔ سؤال ـ پاسخ**[2] (آی آر تی) یا **نظریهٔ نوین آزمون‌سازی**[3] وابسته‌اند. ما در اینجا خوانندگان را با اصول کلی این روشها آشنا می‌کنیم، اما برای کسب اطلاعات جامع در بارهٔ آنها لازم است به منابع و مراجع تخصصی که در ضمن مباحث این قسمت معرفی می‌شوند مراجعه نمایند.

نظریهٔ سؤال ـ پاسخ بر نظریهٔ **صفت مکنون**[4] استوار است (راست[5] و گولومبوک[6]، ۱۹۹۲). نظریهٔ صفت مکنون این فرض را می‌پذیرد که یک صفت یا ویژگی زیربنایی وجود دارد که به شخص امکان می‌دهد تا در یک تکلیف معین موفقیت کسب نماید. از جملهٔ این‌گونه صفتها می‌توان "دانش معنی‌لغات"، "استدلال ریاضی"، یا "تجسم فضایی" را نام برد (ثرندایک، کانینگهام، ثرندایک، و هیگن، ۱۹۹۱).

اندیشهٔ زیربنای آی آر تی بسیار روشن و منطقی است. این نظریه همان منطق زیربنای اندیشهٔ آلفرد بینه[7] بنیان‌گذار آزمونهای هوشی را به کار می‌گیرد (گرانلاند و لین، ۱۹۹۰). بینه می‌گفت نسبت کودکانی که پاسخ یک سؤال را می‌دانند با سن افزایش می‌یابد. در نظریهٔ سؤال ـ پاسخ، به جای سن، توانایی به کار می‌رود و گفته می‌شود که احتمال پاسخ درست به سؤال با افزایش توانایی افزایش می‌یابد. به سخن دیگر، در نظریهٔ سؤال ـ پاسخ فرض بر این است که مثلاً احتمال پاسخ درست دادن به یک سؤال جبر با افزایش دانش جبر آزمون‌شونده افزایش می‌یابد، و این بالقوه یک بیان منطقی است.

در نظریهٔ سؤال ـ پاسخ (آی آر تی) یا نظریهٔ صفت مکنون، میزان احتمال پاسخ درست دادن یک فرد (آزمون‌شونده) به یک سؤالِ آزمون با سطح معینی از توانایی مورد اندازه‌گیری مشخص

---

1. classical test analysis
2. item-response theory (IRT)
3. modern test theory
4. latent trait theory
5. Rust
6. Golombok
7. Alfred Binet

می‌شود. به سخن دیگر، آی آر تی مجموعه‌ای الگوی احتمالاتی به دست می‌دهد که به کمک آنها می‌توان رابطهٔ بین سطح توانایی یک آزمون شونده [که معمولاً با θ (تتا) نشان داده می‌شود] و احتمال پاسخ درست دادن او به هر سؤال آزمون را تعیین کرد.

شولتز[1] و ویتنی[2] (۲۰۰۵) می‌گویند الگوهای اولیهٔ آی آر تی (در دهه‌های ۱۹۷۰-۱۹۴۰ میلادی) برای کار با داده‌های دو ارزشی (صفر برای جواب غلط و ۱ برای جواب درست) پدید آمدند که بیشتر با آزمونهای تواناییهای ذهنی قابل استفاده بودند، اما پژوهشگران بعدها متوجه شدند که آن الگوها را می‌توان با داده‌های حاصل از ابزارهای سنجش شخصیت و نگرش که در آنها پاسخ سؤالها به صورت موافق ـ مخالف یا بله ـ نه تصحیح می‌شوند نیز مورد استفاده قرار داد. بعدها (تا دههٔ ۱۹۸۰) الگوهایی ابداع شدند که با داده‌های چندارزشی، مانند داده‌های حاصل از مقیاسهای ۱ تا ۵ ارزشی (کاملاً موافق تا کاملاً مخالف)، قابل کاربست‌اند.

مقایسهٔ نظریهٔ کلاسیک با نظریهٔ سؤال ـ پاسخ آزمون‌سازی

چنان که دیدیم، در روش کلاسیک تحلیل آزمون، سؤالها یا ماده‌های یک آزمون با توجه به شاخصهای *دشواری* (P) و تمیز (D) حفظ یا حذف می‌شوند. در این روش، مقدار P از طریق محاسبهٔ درصد پاسخ‌دهندگانی که به یک سؤال پاسخ درست می‌دهند و مقدار D از راه محاسبهٔ ضریب همبستگی بین سؤال و کل آزمون تعیین می‌گردند. افزون بر این، نمرهٔ آزمون‌شونده در کل آزمون (یا سطح توانایی او در موضوع مورد آزمون) از مجموع سؤالهایی که درست جواب داده است ــ صرف‌نظر از اینکه کدام سؤالها را درست جواب داده ــ به دست می‌آید.

به خلاف روش کلاسیک که در آن نمرهٔ آزمون‌شونده از مجموع پاسخهای درست او به سؤالهای آزمون به دست می‌آید، در روش آی آر تی نمرهٔ آزمون‌شونده برابر با مجموع پاسخهای درست او نیست. نمرهٔ آی آر تی که معمولاً با θ نشان داده می‌شود هم تابع تعداد پاسخهای درست آزمون شونده و هم تابع ویژگیهای سؤالهای آزمون است (هوگان، ۲۰۰۷). به عنوان مثال، اگر در یک آزمون دو آزمون‌شونده به تعداد مساوی سؤال درست جواب داده باشند اما یکی از آنها به سؤالهای آسان ولی دیگری به همان تعداد سؤال دشوار پاسخ داده باشد نمرهٔ آندو مساوی نخواهد بود. به سخن دیگر، هرچند که این دو آزمون‌شونده به تعداد مساوی سؤال جواب داده‌اند، اما آنکه به سؤالهای سخت‌تر جواب داده است نمرهٔ θ بیشتری خواهد گرفت.

مورد دیگر اختلاف بین این دو روش در آن است که در نظریهٔ کلاسیک اندازه‌گیری اطلاعات مربوط به پارامترها یا شاخصهای سؤال، مانند ضریبهای دشواری و تمیز، در نمونه‌های مختلف آزمون‌شوندگان متفاوت به دست می‌آیند. به عنوان مثال، سؤالی که از یک آزمون ریاضیات دانشگاهی (سطح کارشناسی) انتخاب شده باشد برای دانش‌آموزان دبیرستانی سؤالی سخت، برای دانشجویان

1. Schultz
2. Whitney

دورهٔ کارشناسی سؤالی معمولی، و برای دانشجویان کارشناسی ارشد سؤالی آسان به نظر خواهد رسید.

به خلاف نظریهٔ کلاسیک، در نظریهٔ سؤال ـ پاسخ، اطلاعات به دست آمده دربارهٔ پارامترهای دشواری و تمیز سؤال از نمونه‌ای از آزمون‌شوندگان به نمونه‌ای دیگر فرق نمی‌کند. دلیل این امر آن است که "الگوی ریاضی مورد استفاده برای محاسبهٔ پارامترهای سؤالها (از جمله الگوی لوجستیک[1] یا الگوی بهنجار اُجایو[2]) براساس صفت مکنون به دست می‌آید نه از روی نمرهٔ کل آزمون‌شونده" (شولتز و ویتنی، ۲۰۰۵، ص ۳۲۹). بنابراین، در استفاده از آی آر تی، اطلاعات حاصل از یک نمونه از آزمون‌شوندگان برابر خواهد بود با اطلاعات حاصل از نمونهٔ دیگری از آزمون‌شوندگان ـ صرف‌نظر از میانگین توانایی گروههای آزمون‌شونده.

نظریهٔ کلاسیک آزمون‌سازی در طول سالها ابزار بسیار مفیدی در دست آزمون‌سازان برای ساختن آزمونهای مختلف بوده است، و هنوز هم روشهای وابسته به این نظریه برای تهیه و بازبینی آزمونهای مختصر روشهای قابل استفاده‌ای هستند و به طور گسترده مورد استفاده قرار می‌گیرند (شولتز و ویتنی، ۲۰۰۵). روشهای برخاسته از نظریهٔ صفت مکنون یا نظریهٔ سؤال ـ پاسخ روشهای نوتر و پیشرفته‌تری هستند که اطلاعات جامع‌تر و تعمیم‌پذیرتری به دست می‌دهند. با این حال، هرچند که این روشهای نو از روشهای کلاسیک پیشرفته‌ترند اما بیشتر برای استفاده با آزمونهای جامع و استاندارد شدهٔ ملی مفیدند، ولی برای آزمونهای مختصر معلم‌ساخته همان روش کلاسیک بیشتر قابل استفاده است. کوهن و سوردلیک (۲۰۰۵) علت کم مورد استفاده قرار گرفتن روشهای جدید برای آزمونهای معلم‌ساخته را به "مسائل فنی و پیچیده‌ای که باید حل و فصل شوند تا معلوم گردد که داده‌های به‌دست آمده با ریاضیات الگوی موردنظر سازگارند یا نه" (ص ۲۱۹) نسبت داده‌اند.

نکته‌ای را که در اینجا باید یادآور شویم این است که روشهای کلاسیک و آی آر تی ضد هم نیستند، بلکه هر دو را می‌توان برای اکثریت آزمونها به کاربست. با این حال، بنا به گفتهٔ کوهن (۲۰۰۷)،

> روشهای آی آر تی به نمونه‌های بزرگ نیاز دارند؛ در غیراین‌صورت، روشهای ریاضی آنها درست از کار درنمی‌آیند. از این‌رو، روشهای آی آر تی برای آزمونهای معلم‌ساخته کارآیی ندارند. برای آزمونهای کلاسی معلم‌ساخته روشهای کلاسیک روانسنجی قابل استفاده‌ترند. در عوض، روشهای آی آر تی برای آزمونهای جامع و ملی استاندارد شده مفیدند. (ص ۱۴)

منحنی ویژگی سؤال و استفاده از آن برای تحلیل سؤال آزمون

یکی از روشهای تحلیل آزمون برخاسته از نظریهٔ سؤال ـ پاسخ استفاده از **منحنی ویژگی سؤال** (آی سی سی)[3] یا **تابع سؤال ـ پاسخ** (آی آر اف)[4] است. منحنی ویژگی سؤال یک بازنمایی نموداری

---

1. Logistic
2. normal ogive
3. item characteristic curve (ICC)
4. item characteristic function (IRF)

از رابطهٔ بین احتمال پاسخ درست دادن به یک سؤال و موقعیت آزمون‌شونده در صفت مورد اندازه‌گیری است. در رابطه با سنجش توانایی، این منحنی راهی برای تعیین اینکه شخصی با توانایی x بتواند عملکردی در سطح y داشته باشد به دست می‌دهد. در رابطه با سنجش شخصیت، این احتمال راکه شخصی با x مقدار از یک ویژگی شخصیتی در ابزاری که برای سنجش آن ویژگی درست شده است عملکردی در سطح y دارد تعیین می‌کند (کوهن و سوردلیک، ۲۰۰۵).

تهیهٔ منحنی ویژگی سؤال و تعیین مقدار شاخصهای دشواری و تمیز عموماً با استفاده از کامپیوتر عملی است. با این حال، برای نشان دادن چگونگی رسم‌کردن این منحنی می‌توان تعداد اندکی نمره و محاسبهٔ با دست را مورد استفاده قرار داد. ما در زیر، به منظور آگاه ساختن خوانندگان از چگونگی رسم این منحنی و نحوهٔ استفاده از آن، مثالهای ساده‌ای می‌آوریم.

برای رسم منحنی ویژگی یک سؤال، ابتدا نسبت‌ها یا درصدهای آزمون‌شوندگانی را که در آزمون نمرات مختلف گرفته‌اند و به آن سؤال جواب درست داده‌اند تعیین و این نسبتها را بر روی محور عمودی محور مختصات مشخص می‌کنیم. این نسبتها از یک تابع ریاضی به نام معادلهٔ لوجستیک[1] استخراج می‌شوند (نگاه کنید به بیکر[2]، ترجمهٔ هومن و عسگری، ۱۳۸۱). پس از آن، بر روی محور افقی محور مختصات نمره‌های کل آزمون [در قالب نمره‌های خام یا نمره‌های استاندارد به صورت θ (تتا)] را مشخص می‌کنیم. آنگاه، نقاط تقاطع خطهای عمودی برخاسته از هریک از نمره‌های آزمون (بر روی محور افقی) و خطهای عمودی برخاسته از هریک از نسبتها (بر روی محور عمودی) را تعیین می‌کنیم. سرانجام، از طریق وصل این نقاط به یکدیگر منحنی ویژگی سؤال را رسم می‌کنیم. (هرچند انجام محاسبات موردنیاز و رسم منحنی ویژگی سؤال به کمک کامپیوتر انجام می‌گیرد، با این حال ما در اینجا از اطلاعات جدول ۴-۱۷ و محاسبات دستی استفاده می‌کنیم.)

جدول ۴-۱۷ نتایج یک آزمون ۱۳ سؤالی را نشان می‌دهد. در این آزمون ارزش هر سؤال ۱ نمره است، و نمرهٔ کل هر آزمون‌شونده از مجموع نمرات درست او به دست آمده است. در جدول ۴-۱۷ داده‌های لازم برای رسم منحنی‌های ویژگی سؤالهای ۱ و ۲ و ۳ و ۴ این آزمون داده شده‌اند. به عنوان نمونه، در این جدول ۷۷ درصد کسانی که در آزمون نمرهٔ ۸ گرفته‌اند به سؤال ۱ جواب درست داده‌اند. با توجه به بقیهٔ نسبتهای درست سؤال ۱ منحنی ویژگی آن در شکل ۶-۱۷ رسم شده است.

دقت کنید که در منحنی ویژگی سؤال ۱ (شکل ۶-۱۷) نسبتهای درست برای این سؤال همراه با افزایش نمرهٔ کل آزمون افزایش می‌یابد. این مطلب نشان می‌دهد که هر چه نمرهٔ کل آزمون یک فردبزرگ‌تر باشد با احتمال بیشتری به این سؤال پاسخ درست خواهد داد. یعنی، براساس منطق آی آر تی، هرچه توان آزمون‌شونده در ارتباط با ویژگی مورد اندازه‌گیری بیشتر بوده در این آزمون نمرهٔ بهتری به دست آورده است.

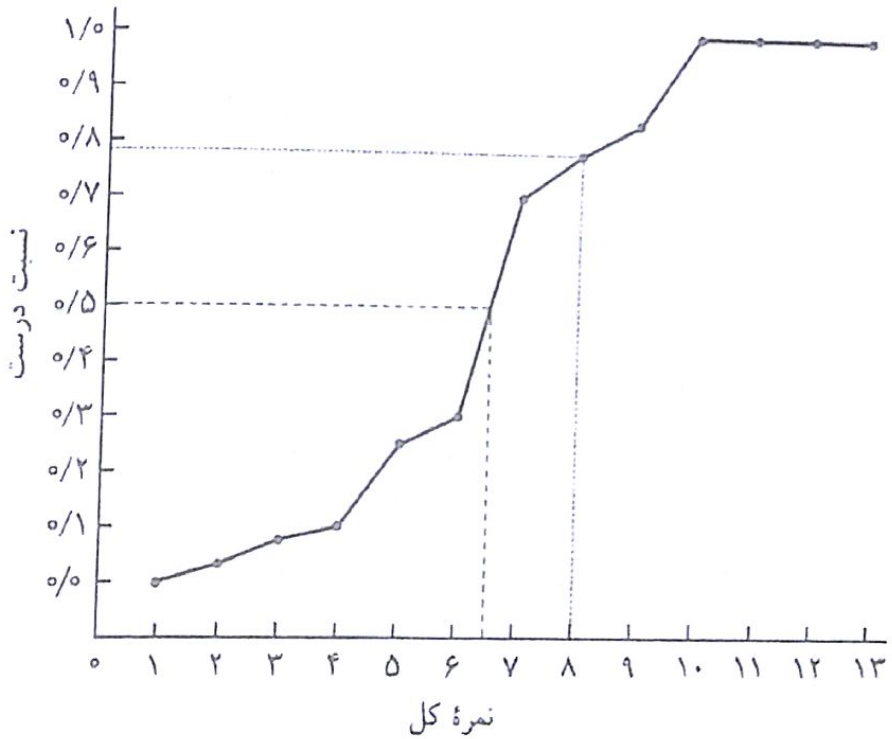
---

1. logistic function
2. Baker

جدول ۴-۱۷ نسبت یا درصد درست پاسخهای ۴ سؤال از یک آزمون ۱۳ سؤالی

نمرهٔ کل	شمارهٔ سؤال			
	۱	۲	۳	۴
۰	.۰۰	.۰۰	.۰۰	.۰۰
۱	.۰۰	.۵۰	.۰۰	.۱۷
۲	.۰۳	.۶۱	.۰۰	.۱۸
۳	.۰۶	.۱۹	.۰۰	.۲۲
۴	.۰۸	.۳۰	.۰۳	.۳۱
۵	.۲۳	.۸۹	.۰۶	.۴۰
۶	.۳۱	.۳۱	.۰۸	.۴۳
۷	.۷۰	.۷۹	.۲۳	.۵۷
۸	.۷۷	.۲۰	.۳۱	.۶۱
۹	.۸۲	.۵۲	.۷۰	.۷۲
۱۰	.۹۹	.۵۹	.۷۷	.۸۲
۱۱	۱.۰۰	.۳۱	.۸۲	.۸۳
۱۲	۱.۰۰	.۷۰	.۹۹	.۸۵
۱۳	۱.۰۰	۱.۰۰	۱.۰۰	۱.۰۰

از روی منحنی ویژگی سؤال می‌توان ضریبهای دشواری و تمیز سؤال را تعیین کرد. ضریب دشواری یا پارامتر دشواری با علامت b نشان داده می‌شود و مُعرف نقطهٔ عطف منحنی ویژگی سؤال است. در سؤالهایی که عامل حدس در پاسخگویی به سؤالها تأثیر ندارد، نقطهٔ عطف منحنی ویژگی سؤال ۵۰٪ (یا نسبت ۰/۵) است. یعنی ضریب دشواری یا سطح دشواری سؤال عبارت است از نمرهٔ معیاری که در آن ۵۰ درصد آزمون‌شوندگان به سؤال جواب درست داده‌اند. به سخن دیگر، "سطح دشواری سؤال نمره‌ای است که در آن پنجاه‌درصد آزمون‌شوندگان جواب درست

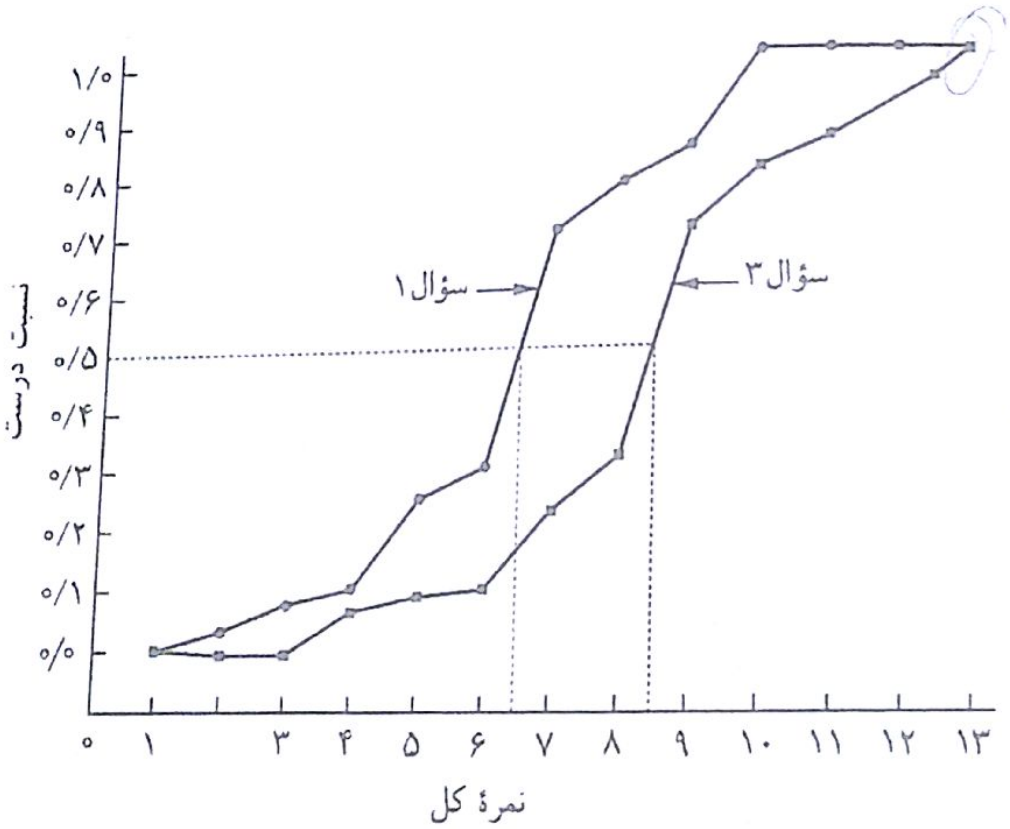


شکل ۶-۱۷ منحنی ویژگی سؤال ۱ مربوط به جدول ۴-۱۷

می‌دهند»(کوالی[1]، ۲۰۱۰، ص ۴۲). (از آنجاکه در مثالهای فرضی ما عامل حدس منظور نشده است، ضریبهای دشواری سؤالها از روی همان ۵۰ درصد محاسبه شده‌اند.) دقت کنید که محور نمرات کل آزمون یا اندازه‌های توانایی و سطح دشورای یکی است؛ از این‌رو، می‌توان دشواری سؤال را مستقیماً به سطح توانایی مربوط به آن تبدیل کرد. در منحنی ویژگی سؤال شکل ۶-۱۷ ضریب دشواری سؤال با نمرهٔ ۶/۵ نشان داده شده است. این ضریب نشان می‌دهد کسانی که در آزمون فرضی ۱۳ نمره‌ای ما نمرهٔ ۶/۵ و بالاتر گرفته باشند می‌توانند به این سؤال پاسخ درست بدهند.

اگر همهٔ آزمون‌شوندگان به یک سؤال جواب درست بدهند، منحنی ویژگی آن سؤال یک خط افقی در بالای محور عمودی خواهد بود که از نسبت ۱ می‌گذرد. در این‌صورت ضریب دشواری سؤال b=۱ است. به عکس، اگر همهٔ آزمون‌شوندگان به یک سؤال جواب غلط بدهند، منحنی ویژگی آن سؤال نیز یک خط افقی خواهد بود که این‌بار از پایین محور عمودی و از نسبت صفر می‌گذرد. ضریب دشواری این سؤال b=۰ است. با توجه به این توضیحات، می‌توان ضریب دشواری سؤال را مستقیماً به سطح زیر منحنی ویژگی سؤال ربط داد؛ هرچه سطح زیر منحنی ویژگی سؤال بیشتر باشد، ضریب دشواری سؤال بزرگ‌تر است (راجرز[2]، ۱۹۹۵). شکل ۷-۱۷ منحنیهای ویژگی سؤالهای ۱ و ۳ جدول ۴-۱۷ و ضریبهای دشواری آنها را نشان می‌دهد. ضریب دشواری سؤال ۱ برابر با ۶/۵ و ضریب دشواری سؤال ۳ برابر با ۸/۵ است.

نقطهٔ تقاطع خط افقی برآمده از نسبت ۰/۵ و منحنی ویژگی سؤال را که نقطهٔ عطف منحنی است **آستانه**[3] می‌گویند، زیرا در این نقطه جهت منحنی تغییر می‌کند. افرادی که نمرهٔ کل آزمون آنها



شکل ۷-۱۷ منحنیهای ویژگی سؤالهای ۱ و ۳ جدول ۴-۱۷ با ضریبهای دشواری متفاوت و ضرایب تمیز برابر

1. Coaly 2. Rogers 3. threshold

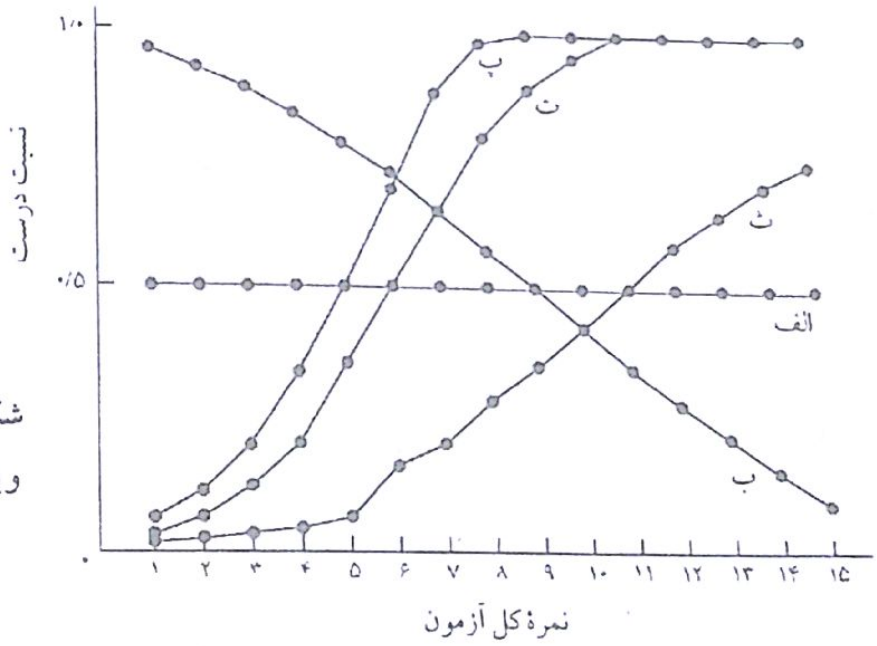
پایین این نقطه است احتمال جواب درست دادن به سؤال برای آنها در حد تصادف است، اما کسانی که نمرهٔ کل آزمون آنها بالاتر از این نقطه است با احتمال زیاد به سؤال جواب درست خواهند داد. به این سبب گفته می‌شود که عملکرد آزمون‌شونده در سؤال به توانایی او وابسته است.

ضریب تمیز یا پارامترِ قدرتِ تشخیص سؤال با علامت a نشان داده می‌شود و برابر است با ضریب زاویهٔ شیبِ منحنی ویژگی سؤال. در آزمونهای استاندارد شدهٔ جامع که در آنها عامل حدس بی‌تأثیر است پارامتر تشخیص یا قدرت تمیز با همبستگی دو رشته‌ای ارتباط دارد به گونه‌ای که می‌توان آن را با استفاده از ضریب همبستگی دورشته‌ای با فرمول زیر تعیین کرد (لُرد[1]، ۱۹۷۵).

$$a = \frac{rbis}{\sqrt{1 - r^2 bis}}$$

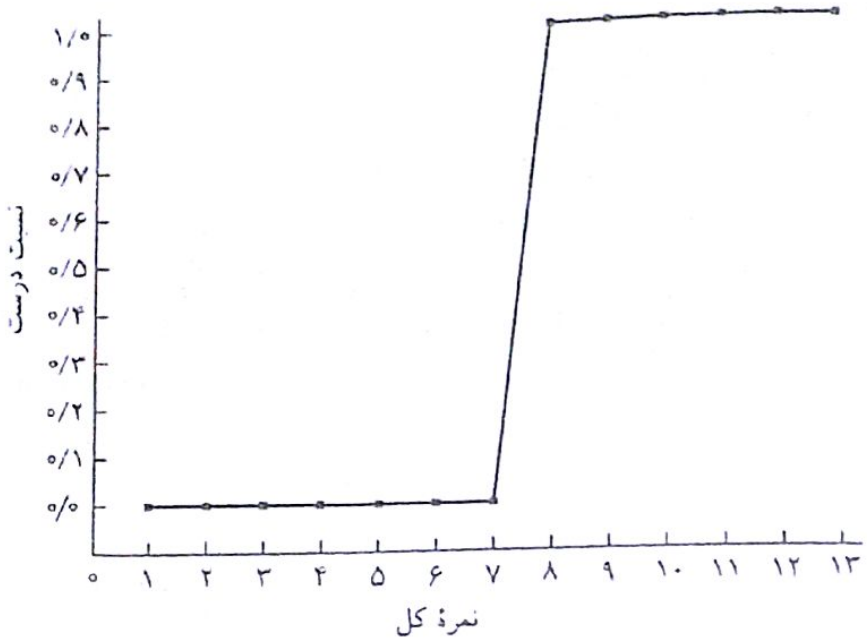
در استفاده از فرمول بالا برای محاسبهٔ ضریب تمیز سوال (a)، علاوه بر اینکه توانایی موردنظر (θ) باید دارای توزیع نرمال (طبیعی) باشد، لازم است چند شرط آماری دیگر نیز برآورده شوند. برای اطلاعات بیشتر، به منابع تخصصی روانسنجی، از جمله لرد (۱۹۷۵)، لرد و نویک[2] (۱۹۶۸)، و امبرتسون[3] و رایز[4] (۲۰۰۰) مراجعه کنید.

در شکل ۸-۱۷ منحنیهای ویژگی ۵ سؤال با هم مقایسه شده‌اند. چنان که دیده می‌شود، سؤال الف یک سؤال ضعیف است، زیرا بین آزمون شوندگان دارای نمرات متفاوت در آزمون (یا آزمون‌شوندگان قوی و ضعیف) تمیز قائل نمی‌شود؛ احتمال جواب درست دادن به این سؤال در تمام سطح نمرات کل آزمون یکسان است. سؤال ب نیز یک سؤال معیوب است، زیرا این سؤال با کل آزمون رابطهٔ معکوس دارد. به سخن دیگر، این سؤال دارای ضریب تمیز منفی است. سؤالهای



شکل ۸-۱۷ مقایسهٔ منحنیهای ویژگی ۵ سؤال مختلف

1. Lord
2. Novic
3. Embretson
4. Reise



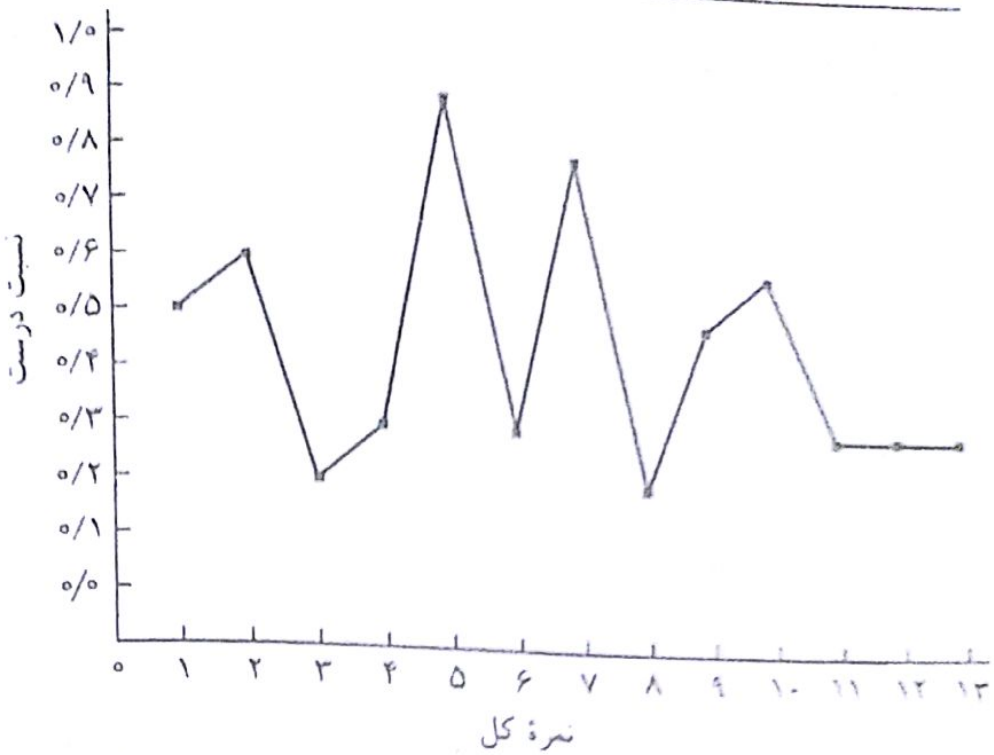
شکل ۹-۱۷
منحنی ویژگی سؤال با قدرت تمیز بسیار خوب

پ و ت سؤالهای نسبتاً خوبی هستند. منحنی ویژگی سؤال ث که از منحنی‌های ویژگی سؤالهای پ و ت خوابیده‌تر است حاکی از این است که به خوبیِ دو سؤال قبلی بین گروههای قوی و ضعیف آزمون‌شوندگان تمیز قائل نمی‌شود.

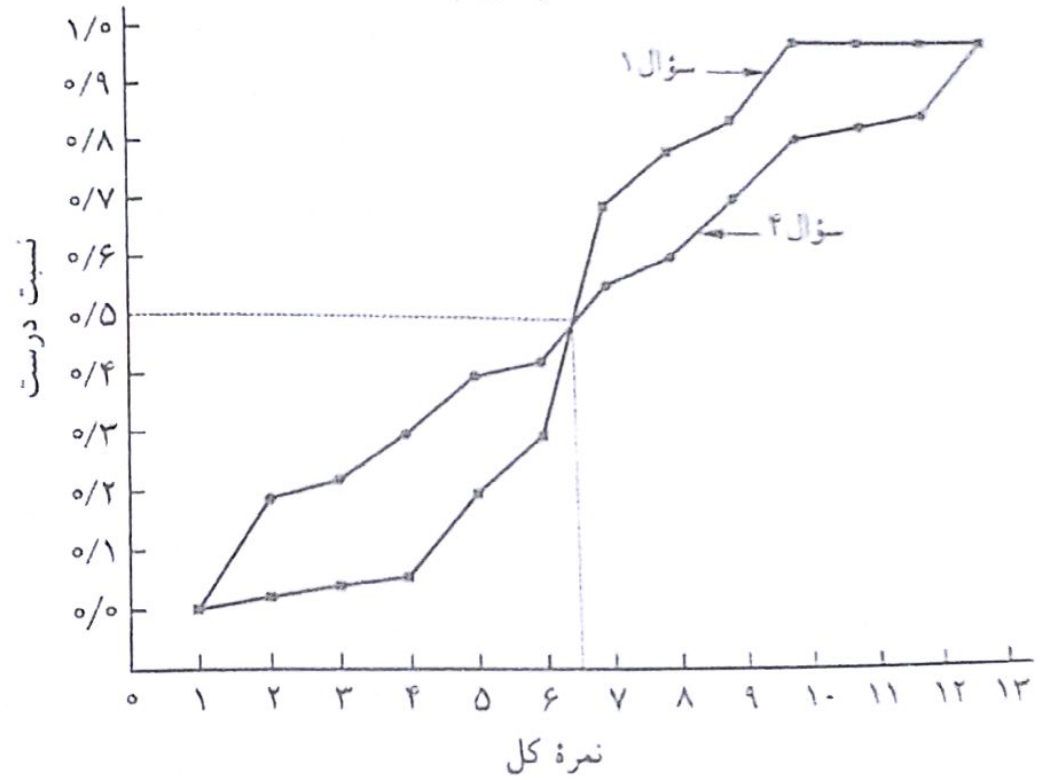
حال به منحنی ویژگی سؤال شکل ۹-۱۷ نگاه کنید. این منحنی در فاصلهٔ بین نمرات ۷ و ۸ دارای شیب بسیار تند است. سؤالی که این منحنی مُعرف آن است آزمون‌شوندگان دارای نمرات پایین ۸ و آزمون‌شوندگان دارای نمرات بالای ۸ را به خوبی از یکدیگر متمایز می‌کند. یعنی اینکه، آزمون‌شوندگانی که نمرهٔ کل آنها در آزمون پایین‌تر از ۸ است به این سؤال پاسخ درست نداده‌اند، ولی آزمون‌شوندگانی که در آزمون نمرهٔ ۸ یا بالاتر گرفته‌اند به سؤال پاسخ درست داده‌اند. هر چه منحنی ویژگی یک سؤال به شکل ۹-۱۷ نزدیک‌تر باشد، سؤال از لحاظ ضریب تمیز بهتر است، یعنی بهتر می‌تواند نمرهٔ کل آزمون‌شونده را پیش‌بینی کند. به سخن دیگر، هرچه منحنی ویژگی یک سؤال حالت پلکانی بیشتری داشته باشد همبستگی بین آن سؤال و کل آزمون بیشتر است.

شکل ۱۰-۱۷ منحنی ویژگی سؤال ۲ مربوط به جدول ۴-۱۷ را نشان می‌دهد. این سؤال به هیچ وجه بین آزمون‌شوندگان مختلف تمیز قائل نمی‌شود. این سؤال هیچ گونه رابطهٔ منظمی بین نسبتهای درست و نمرات کل آزمون نشان نمی‌دهد و لذا به خوبی عمل نمی‌نماید.

ویژگیهای دشواری و تمیز سؤالهای یک آزمون گرچه به هم وابسته‌اند، اما تا حدودی از یکدیگر مستقل‌اند. برای روشن‌شدن مطلب به منحنیهای ویژگی سؤالهای ۱ و ۴ مربوط به جدول ۴-۱۷ در شکل ۱۱-۱۷ نگاه کنید. در این شکل نشان داده شده است که سؤال ۱ دارای قدرت تمیز بیشتری از سؤال ۴ است، زیرا شیب آن تندتر و به حالت پلکانی نزدیک‌تر است. این در حالی است که آستانه یا ضریب دشواری هر دو سؤال برابر است.



شکل ۱۰-۱۷ منحنی ویژگی یک سؤال ضعیف

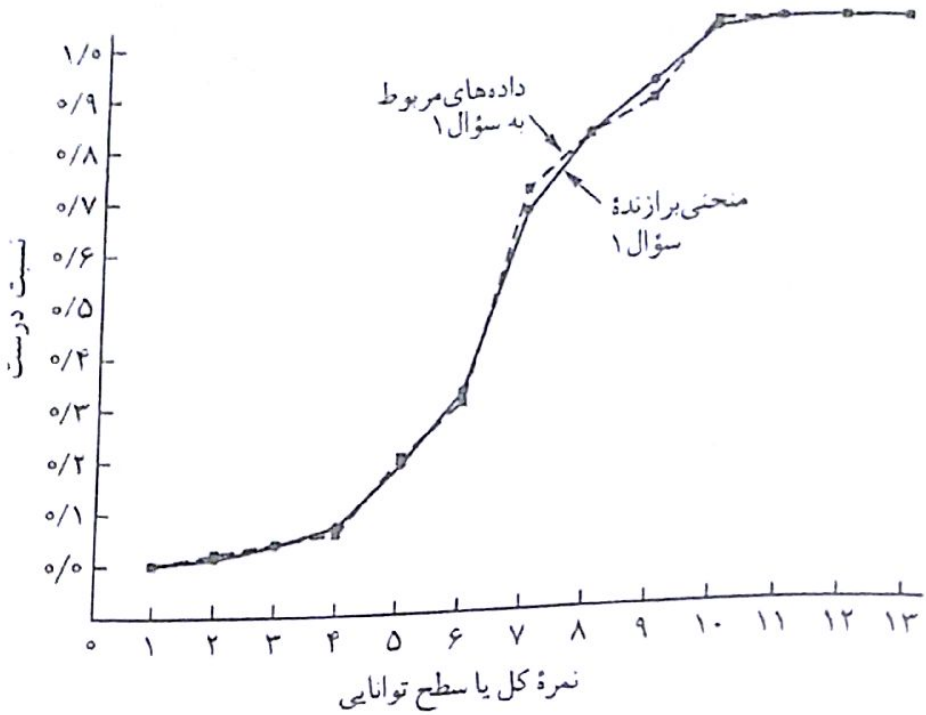


شکل ۱۱-۱۷ منحنیهای ویژگی سؤالهای ۱ و ۴ جدول ۴-۱۷ با ضریبهای دشواری برابر و ضریبهای تمیز متفاوت

برازنده کردن منحنی ویژگی سؤال

برای استفادهٔ بیشتر از منحنی ویژگی سؤال، کوشش به عمل می‌آید تا مناسب‌ترین یا برازنده‌ترین[1] منحنی ریاضی آن سؤال رسم شود. معمولاً برازنده‌ترین منحنی ریاضی یک سؤال به صورت منحنی‌ای که توان دوم فواصل بین آن و داده‌های مشاهده شدهٔ مربوط به سؤال حداقل

1. fittest



شکل ۱۲-۱۷
برازنده‌ترین منحنی ریاضی سؤال شمارهٔ ۱ جدول ۳-۱۷

باشد تعریف می‌شود (راجرز، ۱۹۹۵). شکل ۱۲-۱۷ برازنده‌ترین منحنی ویژگی سؤال ۱ جدول ۴-۱۷ را نشان می‌دهد.

پس از آنکه منحنی ویژگی سؤال به صورت برازنده رسم شد، از روی آن می‌توان چند پارامتر مهم سؤال را تعیین کرد. سه پارامتر یا سه ارزش لازم‌اند تا به کمک آنها منحنی ویژگی سؤال توصیف شود. این پارامترها به قرار زیرند:

۱. پارامتر a که نشان‌دهندهٔ شکل یا شیب منحنی ویژگی سؤال است. چنان که قبلاً دیدیم، شیب صعودی یا منحنی دارای شکل پلکانی نشان‌دهندهٔ قدرت تمیز زیاد سؤال است.

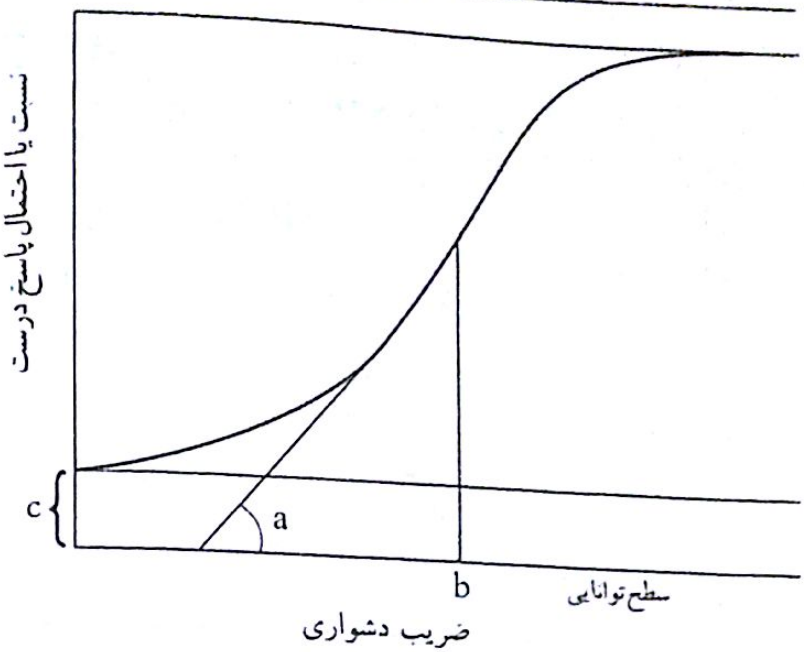
۲. پارامتر b معرف آستانه یا همان ضریب دشواری سؤال است. ضریب دشواری سؤال برابر است با نمره‌ای از آزمون که مقابل نقطهٔ تغییر یا نقطهٔ عطف در منحنی ویژگی سؤال است.

۳. پارامتر c یا پارامتر تصادف میزان حدس‌پذیری سؤال را مشخص می‌کند. هر چه امکان کسب نمره از راه حدس زدن در یک سؤال بیشتر باشد مقدار c برای آن سؤال نیز بزرگ‌تر است.

مجموع پارامترهای فوق را در منحنی ویژگی سؤال برازنده شده یا هموار شده که **تابع ویژگی سؤال**[1] نیز نام دارد می‌توان به صورت شکل ۱۳-۱۷ نشان داد. دقت کنید که شکل این منحنی یا این تابع به صورت تابع تراکمی توزیع طبیعی اجایو است.

پارامتر c یا میزان حدس‌پذیری در آزمونهایی مانند آزمونهای چندگزینه‌ای که در آنها امکان

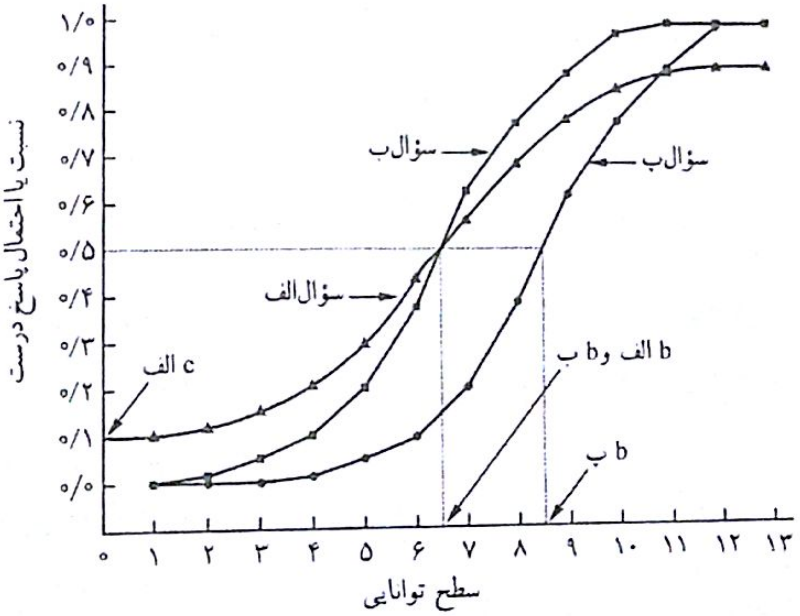
1. item characteristic function



شکل ۱۳-۱۷ شکل کلی منحنی یا تابع ویژگی سؤال
a = tagα = ضریب تمیز

درست جواب دادن یک سؤال از راه حدس زدن کورکورانه یا هشیارانه وجود دارد مطرح است. هرچه این امکان بیشتر باشد مقدار c نیز بیشتر است. اگر عامل حدس در سؤالی بالا باشد، تعداد بیشتری از آزمون‌شوندگان دارای توانایی کم، احتمالاً، به آن جواب درست خواهند داد، و در این صورت منحنی از محلی در بالای نقطهٔ صفر در روی محور عمودی شروع خواهد شد. حال به منحنی شکل ۱۱-۱۷ (ص ۴۴۸) نگاه کنید. می‌بینید که سؤال ۴ دارای عامل حدس بالاست، اما در سؤال ۱ عامل حدس صفر است؛ و این در حالی است که هر دو سؤال دارای ضریب دشواری متوسط‌اند.

چنان که دیدیم، a و b و c پارامترهای ریاضی حاصل از منحنی ویژگی سؤال برازنده شده هستند. شکل ۱۴-۱۷ منحنیهای هموارشده یا برازنده شدهٔ سه سؤال مختلف را نشان می‌دهند.



شکل ۱۴-۱۷ منحنیهای ویژگی سه سؤال دارای پارامترهای مختلف

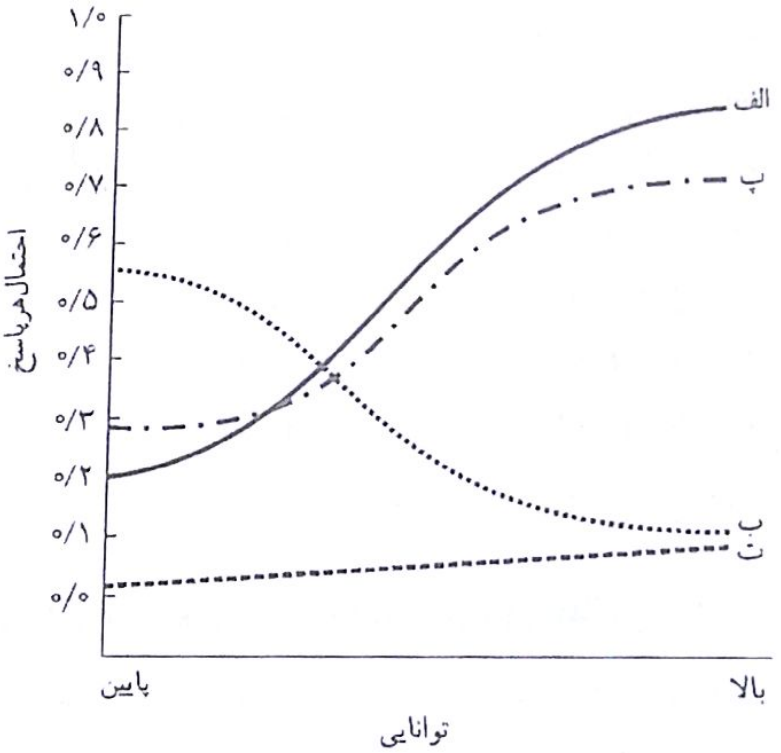
سؤالهای ب و پ از لحاظ پارامتر شکل یا قدرت تمیز مساوی‌اند، اما از لحاظ پارامتر آستانه یا ضریب دشواری با هم فرق دارند. سؤالهای الف و ب از لحاظ پارامتر ضریب دشواری برابرند، اما از لحاظ پارامتر ضریب تمیز با هم فرق دارند. سؤال الف دارای ضریب حدس‌پذیری است، زیرا نسبت کوچکی از آزمون‌شوندگان دارای توانایی صفر سؤال را به طور صحیح جواب داده‌اند.

تحلیل گزینه‌های انحرافی به کمک منحنی ویژگی سؤال

در نظریهٔ سؤال ـ پاسخ هریک از گزینه‌های یک سؤال چندگزینه‌ای (چه گزینهٔ درست و چه گزینه‌های انحرافی) می‌تواند اطلاعاتی دربارهٔ دانش یا مهارت پاسخ‌دهنده به دست دهد. برای این منظور، لازم است، علاوه بر منحنی ویژگی گزینهٔ درست، برای گزینه‌های غلط نیز منحنی ویژگی رسم کنیم. به عنوان نمونه، شکل ۱۵-۱۷ هم منحنی ویژگی گزینهٔ درست (**الف**) و هم منحنیهای ویژگی گزینه‌های انحرافی (**ب**، **پ**، **ت**) را نشان می‌دهد.

چنان‌که در شکل ۱۵-۱۷ دیده می‌شود، گزینهٔ درست یا گزینهٔ **الف** دارای قدرت تمیز مثبت و بالاست. گزینهٔ انحرافی **پ** نیز دارای قدرت تمیز مثبت و نسبتاً بالاست. اما گزینهٔ **د** را کمتر کسی انتخاب کرده و قدرت تمیز ندارد. گزینهٔ **ب** دارای قدرت تمیز منفی است، یعنی پاسخ‌دهندگان ضعیف بیشتر از پاسخ‌دهندگان قوی آن را انتخاب کرده‌اند.

در نمره گذاری سؤال به روش کلاسیک، پاسخ به یک سؤال یا درست ارزیابی می‌شود یا غلط. اگر پاسخ‌دهنده گزینهٔ درست را انتخاب کند نمرهٔ ۱ می‌گیرد و اگر هریک از گزینه‌های انحرافی را انتخاب کند نمرهٔ صفر می‌گیرد. اما در نظریهٔ سؤال ـ پاسخ انتخاب گزینه‌های انحرافی از سوی پاسخ‌دهنده یکسان تلقی نمی‌شود، زیرا در سؤالهای چندگزینه‌ای بعضی گزینه‌های غلط به گزینهٔ



شکل ۱۵-۱۷ منحنی ویژگی گزینه‌های یک سؤال چندگزینه‌ای

درست نزدیک‌تر و بعضی از آنها از گزینهٔ درست بسیار فاصله دارند. به عنوان مثال، در سؤال شکل ۱۵-۱۷ گزینهٔ **پ** بسیار نزدیک به گزینهٔ **الف** است، اما گزینه‌های **ب** و **ت** با گزینهٔ **الف** بسیار متفاوت‌اند. بنابراین، در نمره‌گذاری پاسخ‌دهندگان به این سؤال، می‌توانیم به کسانی که گزینهٔ **الف** را انتخاب کرده‌اند نمرهٔ کامل بدهیم. همچنین برای کسانی که گزینهٔ **پ** را برگزیده‌اند نمرهٔ نزدیک به نمرهٔ کامل منظور کنیم. ولی کمترین نمره را برای گزینه‌های **ت** و **ب** درنظر بگیریم.

خلاصهٔ فصل

شرایط اجرای آزمون باید به گونه‌ای باشد که برای هر یک از آزمون‌شوندگان حداکثر امکانات را فراهم آورد تا آنان بتوانند بهترین عملکرد ممکن را از خود نشان دهند. برای فراهم آوردن بهترین شرایط اجرای آزمون رعایت نکات زیر ضروری است:

۱. سؤالهای مختلف هر آزمون را با توجه به نوع سؤالها، انواع یادگیریها، یا ترتیب مطالب آموزش داده شده در آزمون دسته‌بندی کنید.
۲. برای هر آزمون راهنماییهای لازم را در ابتدای دفترچهٔ آزمون بنویسید.
۳. آزمون را در شرایط مناسب فیزیکی و روانی اجرا کنید.

ما در این فصل علاوه بر نکات بالا روش نمره‌گذاری آزمون و چگونگی تحلیل سؤالهای آزمون را معرفی کرده‌ایم. در نمره‌گذاری آزمون، اگر حدس زدن مجاز باشد، مجموع جوابهای درست هر آزمون‌شونده نمرهٔ او برای آن آزمون به حساب می‌آید. اگر از پیش به آزمون‌شوندگان گفته شود که حدس کورکورانهٔ آنان منجر به کسر شدن مقداری نمره خواهد شد، در ازای هر سه جواب غلطی که هر آزمون شونده در یک آزمون چهارگزینه‌ای می‌دهد یک نمره از نمرات درست او کسر خواهد شد.

پس از تصحیح برگه‌های امتحان، باید به تحلیل سؤالها پرداخت. برای تحلیل سؤالهای یک آزمون، لازم است که ضرایب دشواری و تمیز هر سؤال را تعیین کرد و به تحلیل گزینه‌های انحرافی سؤالها پرداخت. آزمون‌ساز باید پس از تحلیل سؤالهای آزمون، از نتایج این تحلیل برای تجدید نظر و اصلاح آزمون خود استفاده نماید. ما در این فصل هم روش تحلیل آزمونهای وابسته به هنجار را معرفی کرده‌ایم و هم روش تحلیل آزمونهای وابسته به ملاک را.

ما همچنین در این فصل نظریهٔ صفت مکنون و الگوی سؤال ـ پاسخ را توضیح دادیم. در این نظریه‌ها با استفاده از منحنی ویژگی سؤال، ضرایب دشواری و تمیز سؤال تعیین می‌شوند.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. یکی از سؤالهای آزمون دارای ضریب تمیز ۰/۸۲ است. با استفاده از این ضریب می‌توان گفت

دانش‌آموزانی که نمرهٔ بالایی در آزمون گرفته‌اند در این سؤال نیز نمرهٔ خوبی گرفته‌اند.

الف ـ صحیح

ب ـ غلط

۲. اگر بخواهیم تعیین کنیم که دانش‌آموزان گروه بالا یک سؤال را درست جواب می‌دهند یا دانش‌آموزان گروه پایین، از ضریب دشواری سؤال استفاده می‌کنیم.

الف ـ صحیح

ب ـ غلط

برای تحلیل یک سؤال اطلاعات زیر را در اختیار داریم:

گزینه‌ها		۱۰ برگهٔ گروه قوی	۱۰ برگهٔ گروه ضعیف
الف		۰	۰
ب	گزینهٔ درست	۵	۳
پ		۱	۴
ت		۳	۲
ث		۱	۱

۳. ضریب دشواری سؤال فوق چند است؟

الف – ۵۰

ب – ۴۰

پ – ۳۰

ت – ۲۰

ث – ۰

۴. ضریب تمیز سؤال فوق چند است؟

الف – ۰/۱۰+

ب – ۰/۱۰–

پ – ۰/۲۰+

ت – ۰/۲۰–

۵. کدام یک از گزینه‌های انحرافی سؤال فوق بهترین گزینهٔ انحرافی است؟

الف – الف

ب – پ

پ – ت

ت – ث

۶. اگر سؤالی ضریب تمیز بزرگی داشته باشد، می‌توان گفت که نمرات آزمون‌شوندگان در آن سؤال:
الف - هیچ رابطه‌ای با نمرات کل آزمون ندارند.
ب - همبستگی زیادی با نمرات کل آزمون دارند.
پ - همبستگی ضعیفی با نمرات کل آزمون دارند.
ت - همبستگی منفی با نمرات کل آزمون دارند.

۷. تمیز پیش و پس از آموزش به این معنی است که آیا سؤال،
الف - در پیش‌آزمون دشوارتر است یا در پس‌آزمون
ب - به‌خوبی بین آزمون‌شوندگان ضعیف و قوی تمیز قائل می‌شود
پ - دانش‌آموزان حد تسلط و غیر حد تسلط را از هم مجزا می‌سازد
ت - باید پیش از آموزش به اجرا درآید یا پس از آن

۸. اگر به یک کلاس ۳۰ نفری آزمونی بدهیم و ۲۰ نفر از آنها سؤال شمارهٔ ۱۰ را درست جواب بدهند، ضریب دشواری سؤال شمارهٔ ۱۰ چند خواهد بود؟
الف - ۱۰
ب - ۱۰-
پ - ۰/۳۳
ت - ۰/۶۷

۹. اگر گروههای بالا و پایین در یک آزمون هریک شامل ۱۵ دانش‌آموز باشد و ۹ نفر از گروه بالا و ۱۲ نفر از گروه پایین به یک سؤال جواب درست داده باشند، ضریب دشواری آن سؤال چقدر است؟
الف - ۷۰
ب - ۸۰
پ - ۳۰
ت - ۵۰

۱۰. فرض کنید سؤالی دارای ضریب تمیز منفی باشد. بر اساس این ضریب، می‌توان پیش‌بینی کرد دانش‌آموزی که به این سؤال جواب درست داده است در آزمون نمرهٔ ــ گرفته است.
الف - بالایی
ب - پایینی
پ - متوسطی
ت - صفر

۱۱. کدام‌یک از ضریبهای دشواری زیر مُعرف این است که سؤال دشوارتر است؟

الف - ۰/۹۰

ب - ۰/۵۰

پ - ۰/۳۵

ت - ۰/۱۱

۱۲. بهترین ضریب دشواری برای آزمونهای وابسته به هنجار کدام است؟

الف - ۰

ب - ۰/۵۰

پ - ۰/۸۰

ت - ۱

۱۳. ضریب همبستگی یک سؤال با کل آزمون مُعرف چیست؟

الف - پایایی آزمون

ب - روایی آزمون

پ - ضریب دشواری سؤال

ت - ضریب تمیز سؤال

۱۴. ترتیب قرار گرفتن سؤالهای مختلف یک آزمون را به دنبال یکدیگر توضیح دهید.

۱۵. اگر در یک آزمون ۵ گزینه‌ای که دارای ۷۰ سؤال است، دانش‌آموزی به ۶۰ سؤال پاسخ داده باشد و از این تعداد ۸ سؤال غلط و ۵۲ سؤال درست باشد نمرهٔ او را با کسر مقدار لازم برای حدس زدن حساب کنید.

۱۶. روش محاسبهٔ ضریبهای تمیز و دشواری سؤالهای تشریحی را توضیح دهید.

۱۷. نحوهٔ استفاده از ضریبهای تمیز و دشواری در آزمونهای ملاکی را توضیح دهید.

۱۸. نتایج حاصل از تحلیل سؤالهای یک آزمون و ضریبهای دشواری و تمیز را برای تجدید نظر و اصلاح یک آزمون چگونه مورد استفاده قرار می‌دهید؟

۱۹. برای تجدید نظر و اصلاح یک آزمون که خود ساخته‌اید، نتایج حاصل از تحلیل سؤالهای آن و ضریبهای دشواری و تمیز را چگونه مورد استفاده قرار می‌دهید؟ روش مورد استفادهٔ خود را توضیح دهید.

۲۰. اگر ضریب دشواری سؤالی ۱ باشد ضریب تمیز آن چقدر است؟ دلایل خود را بیان کنید.

۲۱. از لحاظ ضرایب دشواری و تمیز بهترین شرایط برای یک آزمون چیست؟ نظر خود را بیان کنید.

۲۲. نظریهٔ صفت مکنون و نظریهٔ سؤال ـ پاسخ را تعریف کنید و رابطهٔ آنها را با یکدیگر توضیح دهید.

۲۳. منحنی ویژگی سؤال را تعریف کنید و نحوهٔ رسم آن را با ذکر مثال بیان کنید.

۲۴. روش تحلیل سؤال را با استفاده از منحنی ویژگی سؤال توضیح دهید.

۱۸

استفاده از روشهای آماری برای تفسیر نمره‌های آزمون

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. متغیر، فراوانی، و توزیع فراوانی را تعریف کند.

۲. جدول توزیع فراوانی نمراتی را که از اجرای یک آزمون به دست آمده است، یک بار به صورت اعداد طبقه‌بندی نشده و یک بار هم به صورت اعداد طبقه‌بندی شده، تهیه کند.

۳. نمودار ستونی، چند ضلعی، و تراکمی نمرات یک آزمون را رسم کند.

۴. نما، میانگین، و میانهٔ نمرات یک آزمون را (هم به صورت نمرات خام و هم به صورت نمرات طبقه‌بندی شده) تعیین کند.

۵. دامنهٔ تغییر، واریانس، و انحراف معیار نمرات یک آزمون را تعیین کند.

۶. ضریب همبستگی را تعریف کند، مفهوم آن را توضیح دهد، و روشهای محاسبهٔ ضریب همبستگی زیر را در مورد نمراتی که در اختیار او گذاشته می‌شوند به دست آورد:

الف ـ همبستگی گشتاوری پیرسون

ب ـ همبستگی رتبه‌ای اسپیرمن

پ ـ همبستگی دورشته‌ای نقطه‌ای

ت ـ همبستگی دورشته‌ای

ث ـ همبستگی فی

ج ـ همبستگی تتراکوریک

بعد از تهیه و اجرای آزمونهای پیشرفت تحصیلی، برای تفسیر و معنی‌دادن به نتایج و مورد استفاده قراردادن آنها، معمولاً لازم است که در بارهٔ نمرات اقداماتی انجام دهیم. از جمله لازم می‌آید که چگونگی تمرکز نمرات را تعیین کنیم، پراکندگی آنها را مشخص نماییم، آنها را به صورت نمودار نشان دهیم، رابطهٔ نمرات یک درس را با درسی دیگر معین کنیم، یا عملیات دیگری در مورد آنها انجام دهیم. برای اینکه بتوانیم این کارها را به‌دقت انجام دهیم، لازم است با پاره‌ای از مفاهیم و روشهای آماری آشنا شویم. در این فصل، روشهای سادهٔ آماری مورد نیاز برای تفسیر نمرات امتحانی را معرفی می‌کنیم.

تعریف متغیر، فراوانی، و توزیع فراوانی

به هرگونه ویژگی یا صفتی از اشیاء، اشخاص، یا رویدادها که بتواند مقادیر مختلفی به خود بگیرد **متغیر**[1] گفته می‌شود. قد، وزن، هوش، انگیزش، یادگیری، پیشرفت تحصیلی، و مانند اینها نمونه‌های متغیر هستند. دفعات وقوع یک متغیر را **فراوانی**[2] آن متغیر می‌گویند. اگر ۱۲ دانش‌آموز در یک آزمون نمرهٔ ۱۵ بگیرند، گفته می‌شود که نمرهٔ ۱۵ دارای فراوانی ۱۲ است. به فهرستی از تعدادی نمره که به ترتیب بزرگی مرتب شده باشند و فراوانی آن نمرات نیز مشخص شده باشد **توزیع فراوانی**[3] گفته می‌شود. بنا به تعریف، توزیع فراوانی جدولی است که نشان می‌دهد هر یک از نمرات در گروه نمرات چند بار تکرار شده است.

تهیهٔ توزیع فراوانی نمرات

پس از آنکه آزمونی را اجرا کردیم و نتایج آن را به‌دست آوردیم، با دسته‌ای از اعداد به نام **نمرات خام**[4] روبه‌رو هستیم که به خودی خود چندان قابل استفاده نیستند. یکی از کاربردهای روشهای آماری دسته‌بندی و تلخیص نمرات خام و تفسیرپذیر کردن آنهاست. برای این منظور، روش تهیهٔ توزیع فراوانی نمرات را مورد استفاده قرار می‌دهیم. برای تهیهٔ جدول توزیع فراوانی، باید نمره‌ها را به ترتیب بزرگی یا کوچکی پشت سر هم قرار دهیم و در مقابل هر نمره فراوانی آن، یعنی تعداد دفعاتی را که آن نمره تکرار شده است، نیز بنویسیم. جدول حاصل جدول توزیع فراوانی نمرات نام دارد. جدول ۱ـ۱۸ نمونه‌ای از توزیع فراوانی تعدادی نمره است.

هرگاه تعداد نمرات زیاد باشد، پشت سر هم قراردادن آنها، برای تهیهٔ جدول توزیع فراوانی، یک ستون طولانی از نمرات به دست می‌دهد که خلاصهٔ خوبی از آنها نیست. در ضمن، محاسبهٔ شاخصهای آماری برای چنین فهرستی طولانی از نمرات کار وقت‌گیری است. برای رفع این

1. variable
2. frequency
3. frequency distribution
4. raw scores

جدول ۱ـ۱۸ توزیع فراوانی نمرات طبقه‌بندی نشده

نمره X	فراوانی f
۲۰	۱
۱۹	۲
۱۸	۲
۱۷	۵
۱۶	۳
۱۵	۴
۱۴	۶
۱۳	۳
۱۲	۲
۱۱	۱
۱۰	۱

N = ۳۰

نواقص، نمرات را طبقه‌بندی می‌کنند و توزیع فراوانی نمرات طبقه‌بندی شده را برای انجام محاسبات آماری مورد استفاده قرار می‌دهند.

برای اینکه نمرات را طبقه‌بندی کنیم، ابتدا باید یک **فاصلهٔ طبقه**[1] مناسب انتخاب نماییم و بعد طبقه‌بندی را بر مبنای آن فاصلهٔ طبقه انجام دهیم. روش معمول برای انتخاب فاصلهٔ طبقه این است که فاصله‌ای انتخاب کنیم که در مجموع اعداد را به حدود ۱۵ طبقه تقسیم کند. برای انجام این کار، راه ساده این است که تفاضل بزرگ‌ترین و کوچک‌ترین نمره یا دامنهٔ تغییر نمرات را حساب کنیم و عدد حاصل را بر ۱۵ تقسیم نماییم و خارج قسمت این تقسیم را اندازهٔ فاصلهٔ طبقه قرار دهیم. چنانچه این خارج‌قسمت عددی اعشاری باشد، نزدیک‌ترین عدد صحیح به آن را به عنوان فاصلهٔ طبقه برمی‌گزینیم. برای تعیین فاصلهٔ طبقهٔ نمرات جدول ۲ـ۱۸ به طریق زیر عمل شده است:

بزرگترین نمره − کوچکترین نمره = دامنهٔ نمرات

۷۱ − ۳ = ۶۸

۶۸ ÷ ۱۵ = ۴/۵۳

فاصلهٔ طبقاتی (نزدیکترین نمره به ۴/۵۳) = ۵

جدول ۲ـ۱۸ توزیع فراوانی نمرات خام ۴۹ دانش‌آموز را در درس علوم به شرح زیر نشان می‌دهد: ۴۳، ۶۰، ۳۷، ۴۳، ۲۱، ۷۱، ۵۲، ۴۳، ۶۱، ۴۷، ۴۲، ۳۳، ۳۸، ۵۸، ۴۹، ۴۷، ۴۵، ۳، ۶۷، ۴۰، ۳۸، ۵۰، ۱۳، ۶۸، ۵۴، ۲۲، ۳۹، ۴۱، ۳۶، ۴۴، ۲۷، ۲۸، ۴۰، ۴۴، ۳۸، ۱۵، ۵۵، ۴۴، ۳۶، ۵۸، ۴۶، ۲۵، ۳۳، ۸، ۳۲، ۲۴، ۴۱، ۱۸، ۳۶.

1. class interval (CI)

جدول ۲ـ۱۸ توزیع فراوانی نمرات ۴۹دانش‌آموز دردرس علوم (نمرات‌طبقه‌بندی شده)

طبقات	خط‌نشان	فراوانی
۷۰ـ۷۴	/	۱
۶۵ـ۶۹	//	۲
۶۰ـ۶۴	//	۲
۵۵ـ۵۹	///	۳
۵۰ـ۵۴	///	۳
۴۵ـ۴۹	𝍸	۵
۴۰ـ۴۴	𝍸 𝍸 /	۱۱
۳۵ـ۳۹	𝍸 ///	۸
۳۰ـ۳۴	///	۳
۲۵ـ۲۹	///	۳
۲۰ـ۲۴	///	۳
۱۵ـ۱۹	//	۲
۱۰ـ۱۴	/	۱
۵ـ۹	/	۱
۰ـ۴	/	۱

N = ۴۹

ستون فراوانیهای جدول ۲ـ۱۸ نشان می‌دهد که در هر طبقه چند نمره موجود است. برای نمونه، در طبقهٔ ۴۹ـ۴۵، پنج نمره یا پنج دانش‌آموز قرار دارند. از روی این جدول نمی‌توان مشخص کرد که چند نفر از این ۵ نفر دانش‌آموز نمرهٔ ۴۵ گرفته‌اند، چند نفر نمرهٔ ۴۶، و الی آخر، زیرا ما هنگام طبقه‌بندی کردن اعداد، این اطلاعات را از دست داده‌ایم. فرض ما این است که این افراد به طور مساوی در این طبقه‌بندی جای دارند.

لازم به یادآوری است آنجا که دامنهٔ نمرات خیلی زیاد نیست، طبقه‌بندی اعداد ضرورتی ندارد. در نظام نمره‌گذاری آموزشگاههای کشور ایران که نمرات کلاً ۲۰ طبقهٔ از صفر تا ۲۰ را تشکیل می‌دهند طبقه‌بندی نمرات چندان فایده‌ای ندارد و بهتر است با نمرات طبقه‌بندی نشده جدول توزیع فراوانی را تهیه نماییم (مانند جدول ۱ـ۱۸)، و محاسبات آماری را با چنین جدولی انجام دهیم که بسیار آسانتر از انجام محاسبات با نمرات طبقه‌بندی شده است.

نمایش نمرات به صورت نمودار

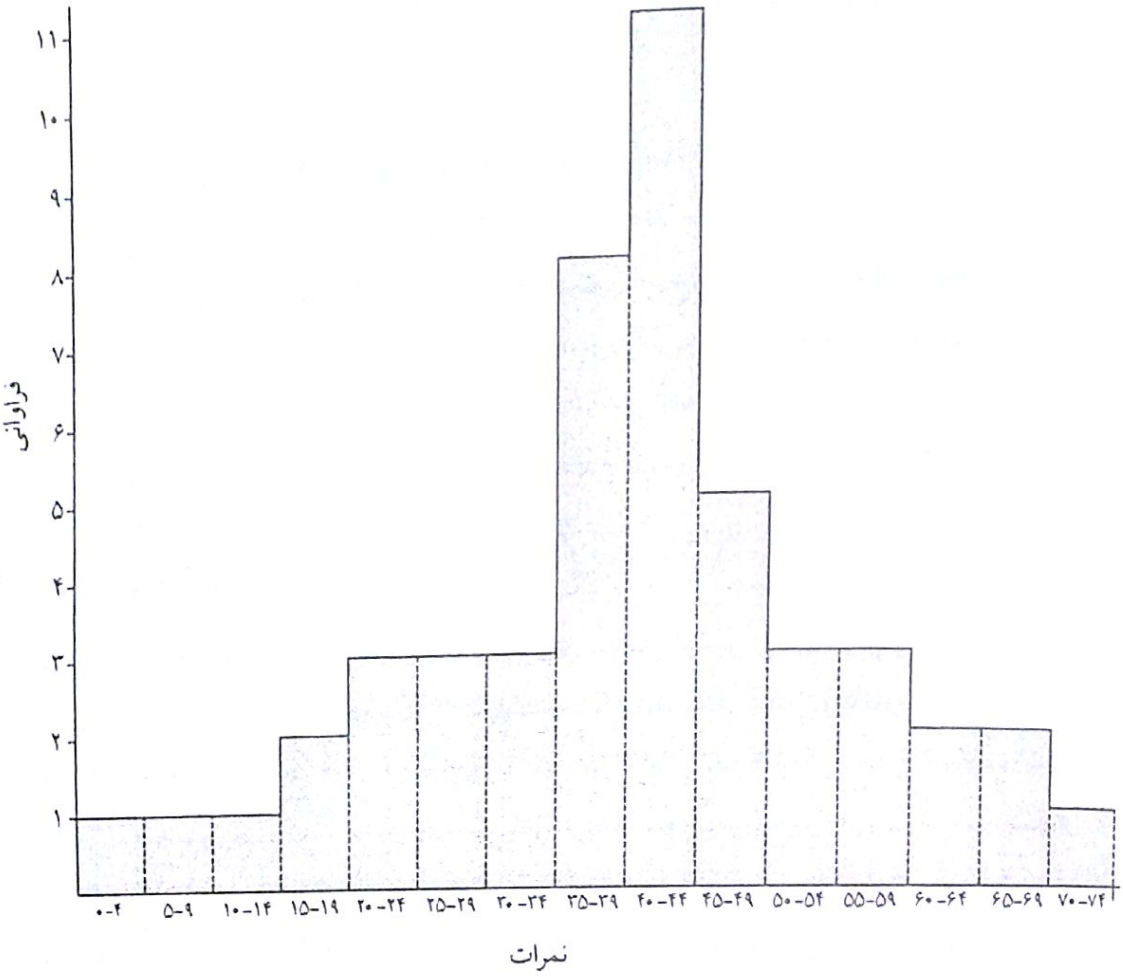
اطلاعات مندرج در توزیع فراوانی را می‌توان به صورت **نمودار**[1] نشان داد. ارائهٔ توزیع فراوانی به صورت نمودار کار تحلیل و تفسیر نمرات را آسان می‌سازد. به طور کلی، "نمودار به یک رسم[2] یا شکل[3] گفته می‌شود که برای تبدیل داده‌های آماری مورد استفاده قرار می‌گیرد" (ساکس، ۱۹۹۷). نمودارها

1. graph 2. drawing 3. figure

غالباً برای سهولت ارتباط به کار می‌روند زیرا آنها شکلهای دیداری قابل درکی را به دست می‌دهند. ما در زیر نمودارهای معروف را معرفی می‌کنیم.

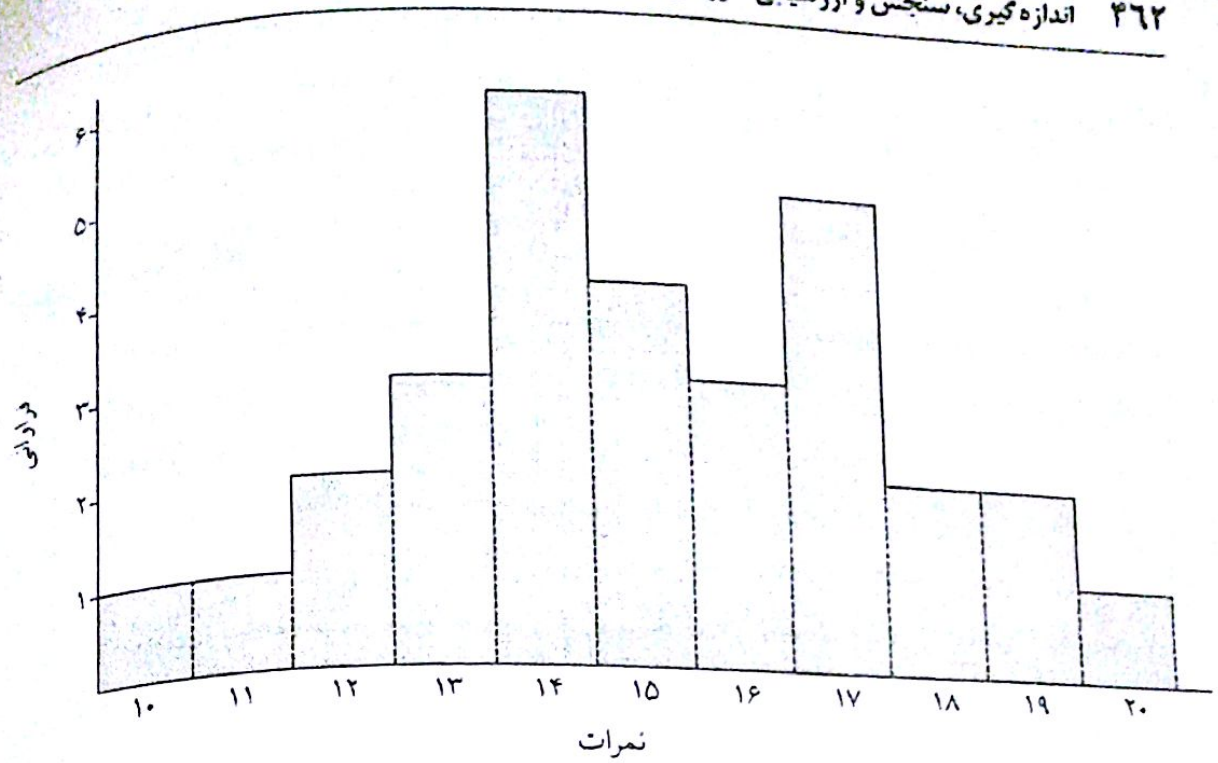
نمودار هیستوگرام

در همهٔ نمودارها از جمله هیستوگرام[1]، توزیع فراوانی نمرات با استفاده از محورهای مختصات نشان داده می‌شود. برای ترسیم نمودار هیستوگرام، فاصلهٔ طبقات را بر روی محور افقی و فراوانیها را بر روی محور عمودی مشخص می‌کنیم. بعد بر روی هر طبقه ستون یا مربع مستطیلی می‌سازیم به گونه‌ای که عرض آن برابر با فاصلهٔ طبقه و ارتفاعش به اندازهٔ فراوانی همان طبقه باشد. شکل حاصل نمودار هیستوگرام توزیع نمرات است. شکل ۱ـ۱۸ نمودار هیستوگرام نمرات توزیع فراوانی جدول ۲ـ۱۸ را نشان می‌دهد. این نمودار نشان می‌دهد که در فاصلهٔ ۴ـ۰ یک نمره، در فاصلهٔ ۹ـ۵ یک نمره،...، در فاصلهٔ ۱۹ـ۱۵ دو نمره، و الی آخر وجود دارند.



شکل ۱ـ۱۸ نمودار هیستوگرام نمرات طبقه‌بندی شده (مربوط به جدول ۲ـ۱۸)

1. histogram



شکل ۲ـ۱۸ نمودار هیستوگرام نمرات طبقه‌بندی نشده (مربوط به جدول ۱ـ۱۸)

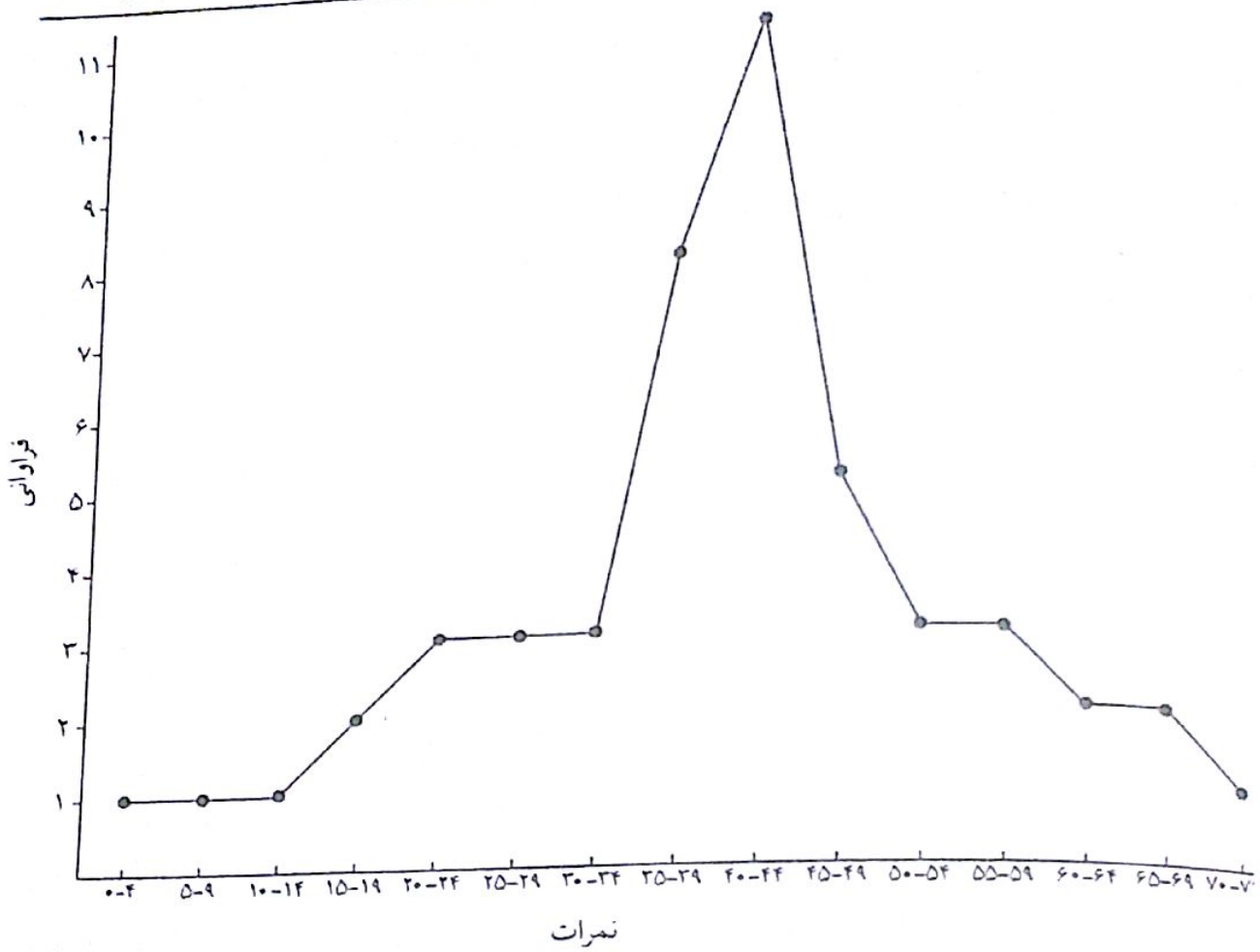
برای ترسیم هیستوگرام نمرات طبقه‌بندی نشده، فاصلهٔ طبقه را ۱ فرض می‌کنیم که از نیم نمره قبل از هر نمره تا نیم نمره بعد از آن ادامه می‌یابد. مثلاً فاصلهٔ طبقهٔ ۱۰ از ۹/۵ است تا ۱۰/۵، فاصلهٔ طبقهٔ ۱۱ از ۱۰/۵ است تا ۱۱/۵، و الی آخر. برای ترسیم نمودار هیستوگرام این نمرات نیز، به تعداد طبقات (در اینجا به تعداد نمرات) بر روی محور افقی ستونها یا مربع مستطیلهایی درست می‌کنیم که عرض هر یک از آنها برابر ۱ واحد و بلندی هر یک از آنها به اندازهٔ فراوانی نمرهٔ مربوط به آن باشد. شکل ۲ـ۱۸ نمودار هیستوگرام نمرات طبقه‌بندی نشدهٔ جدول ۱ـ۱۸ را نشان می‌دهد.

نمودار چندضلعی

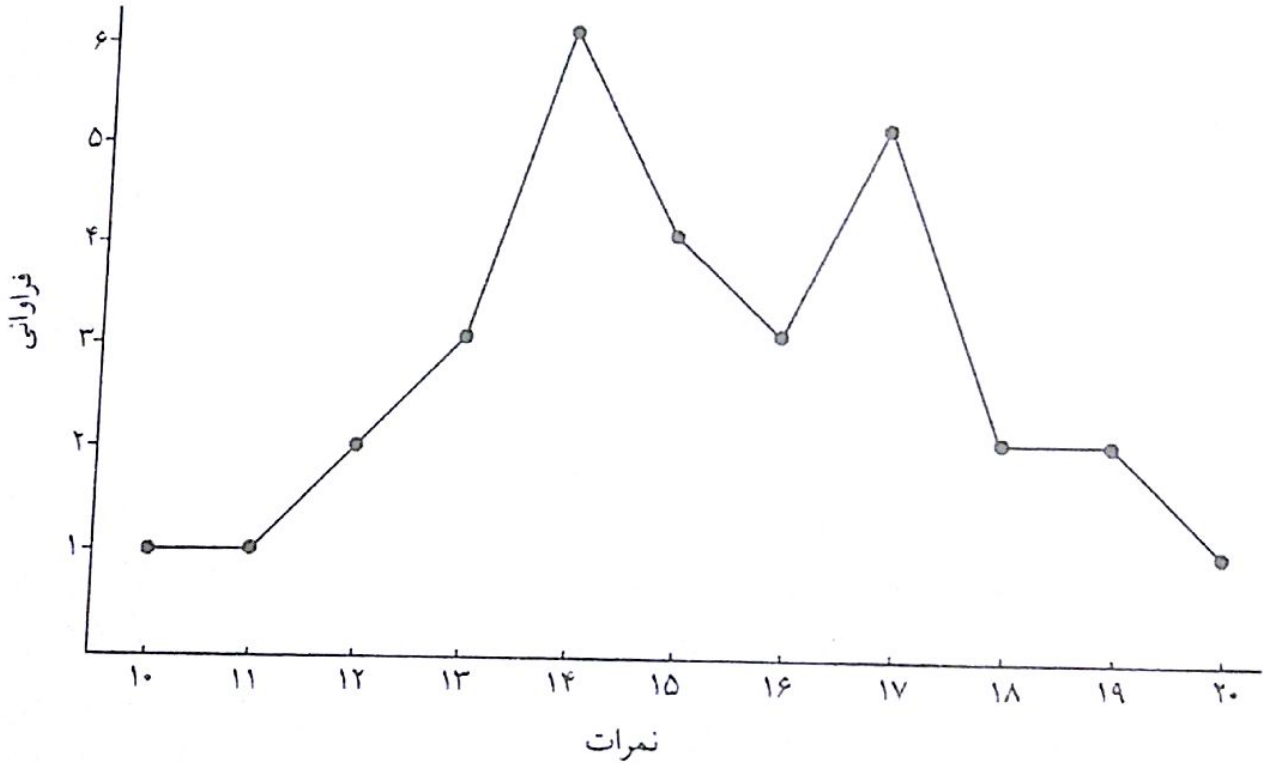
اگر وسط قاعدهٔ فوقانی هر یک از مربع‌مستطیلهای نمودار هیستوگرام را به هم وصل کنیم، یا به عوض کشیدن مربع مستطیل، در بالای **نقطهٔ وسط** هر یک از فواصل طبقاتی نقطه‌ای به اندازهٔ فراوانی آن فاصلهٔ طبقاتی در روی صفحه مشخص کنیم و این نقاط را به یکدیگر وصل نماییم، شکل حاصل که به صورت یک چندضلعی است، نمودار **چندضلعی**[1] توزیع فراوانی نام دارد. شکلهای ۳ـ۱۸ و ۴ـ۱۸ به ترتیب نمودار چندضلعی توزیع فراوانی نمرات جدولهای ۲ـ۱۸ و ۱ـ۱۸ هستند. نمودار هیستوگرام و نمودار چندضلعی فرق عمده‌ای با یکدیگر ندارند، و انتخاب میان آن‌دو به سلیقهٔ شخصی معلم وابسته است.

---

1. polygon



شکل ۳ـ۱۸ نمودار چندضلعی نمرات طبقه‌بندی شده (مربوط به جدول ۲ـ۱۸)



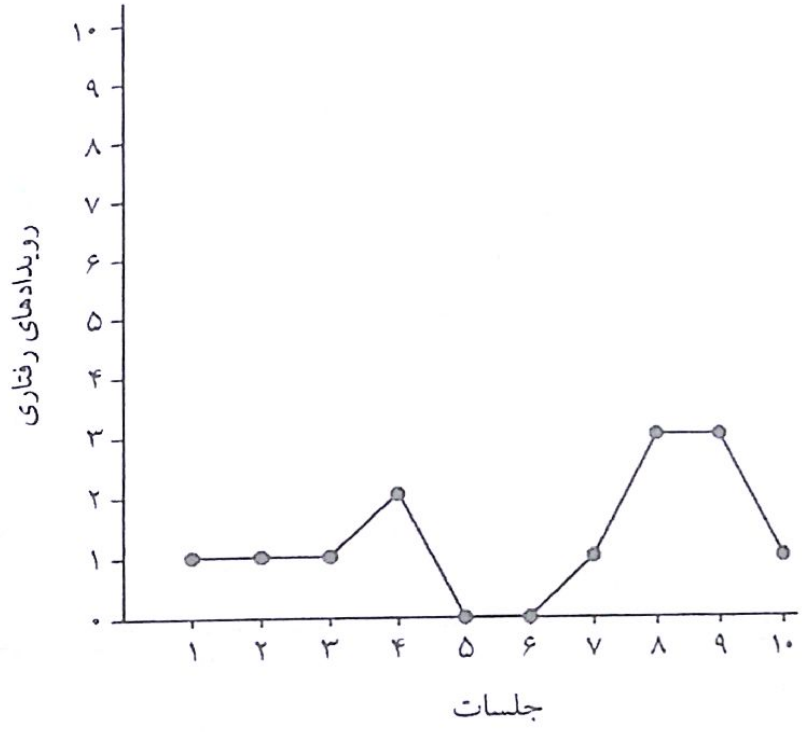
شکل ۴ـ۱۸ نمودار چندضلعی نمرات طبقه‌بندی نشده (مربوط به جدول ۱ـ۱۸)

جدول ۳ـ۱۸ داده‌های خام جریان پیشرفت کار یک دانش‌آموز در سوارکردن قطعات یک ماشین

جلسات	رویدادهای رفتاری
۱	۱
۲	۱
۳	۱
۴	۲
۵	۰
۶	۰
۷	۱
۸	۳
۹	۳
۱۰	۱

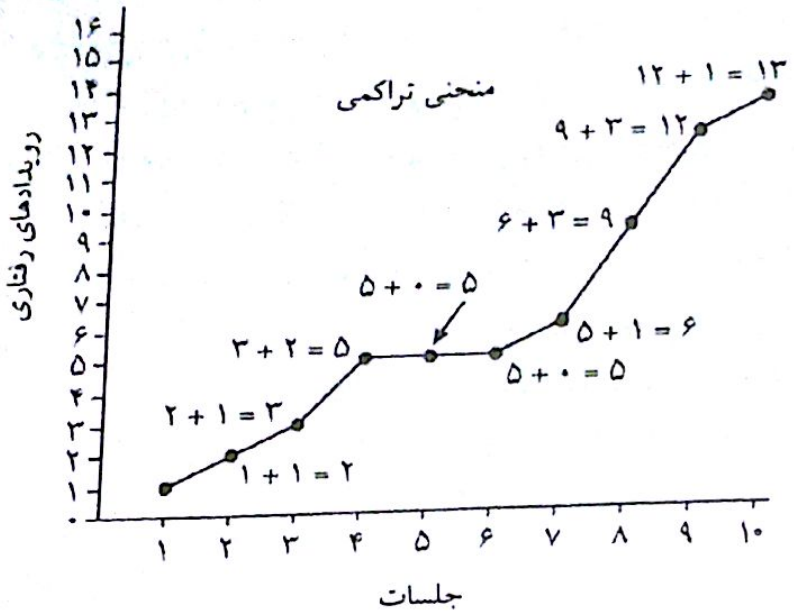
نمودار تراکمی

در نمودار چندضلعی، چنان‌که دیدیم، در روی منحنی، هر نقطهٔ معرفِ یک طبقه از نمرات، بدون توجه به نقاط دیگر، تعیین می‌شود. اما در **نمودار تراکمی یا ثبت تراکمی**[1]، برای تعیین هر نقطهٔ معرف



شکل ۵ـ۱۸ نمودار چندضلعی داده‌های جدول ۳ـ۱۸

1. cumulative record



شکل ۶ـ۱۸ نمودار تراکمی داده‌های جدول ۳ـ۱۸

هر طبقه، فراوانی طبقهٔ مورد نظر را به فراوانیهای طبقات قبل از آن اضافه می‌کنیم. بنابراین، نمودار تراکمی هرگز به پایین نمی‌رود. خط نمودار یا رو به بالا می‌رود یا با محور افقی موازی باقی می‌ماند. به منظور اینکه تفاوت بین نمودار چندضلعی و نمودار تراکمی را روشن کنیم، برای داده‌های جدول ۳ـ۱۸ این دو نوع نمودار را رسم کرده‌ایم. داده‌های جدول ۳ـ۱۸ نتایج پیشرفت کار یک دانش‌آموز در سوارکردن قطعات یک ماشین هستند.

شکل ۵ـ۱۸ نمودار چندضلعی داده‌های جدول ۳ـ۱۸ و شکل ۶ـ۱۸ نمودار تراکمی همان داده‌ها را نشان می‌دهند. دقت کنید که در نمودار چندضلعی شکل ۵ـ۱۸ خط نمودار بالا و پایین می‌رود در حالی که در نمودار شکل ۶ـ۱۸ خط نمودار همواره رو به بالا می‌رود.

نمودارهای هیستوگرام و چندضلعی مخصوص داده‌های پیوستهٔ حاصل از مقیاسهای فاصله‌ای و نسبتی هستند. برای داده‌هایی که با مقیاسهای اسمی و ترتیبی به دست می‌آیند و عموماً جنبهٔ گسسته دارند نمودارهای دیگری مورد استفاده قرار می‌گیرند که در زیر آنها را معرفی می‌کنیم.

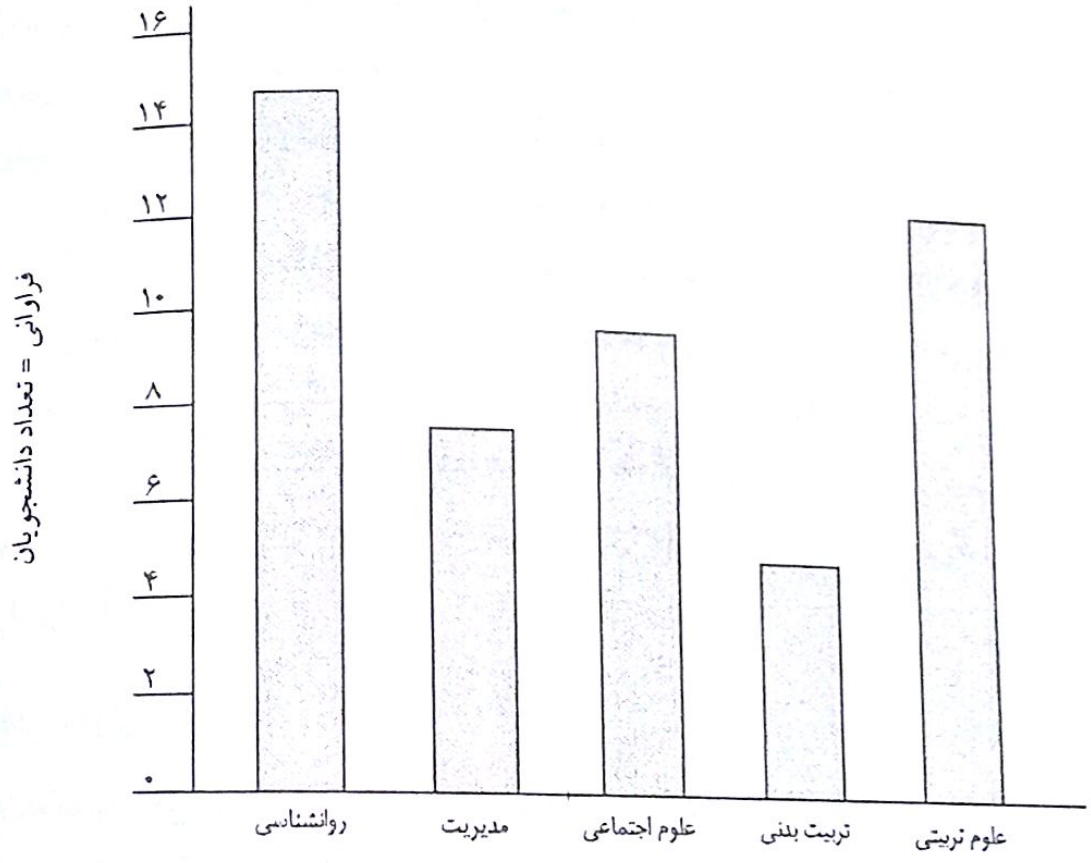
نمودار ستونی

چنان‌که گفتیم، هر زمان که با داده‌های مربوط به مقیاسهای اسمی یا ترتیبی سروکار داریم، می‌توانیم از نمودارهای دیگری به غیر از آنچه در بالا توضیح دادیم استفاده کنیم. یک نوع نمودار که به این منظور به کار می‌رود **نمودار ستونی یا میله‌ای**[1] است. برای رسم نمودار ستونی، بسیار شبیه

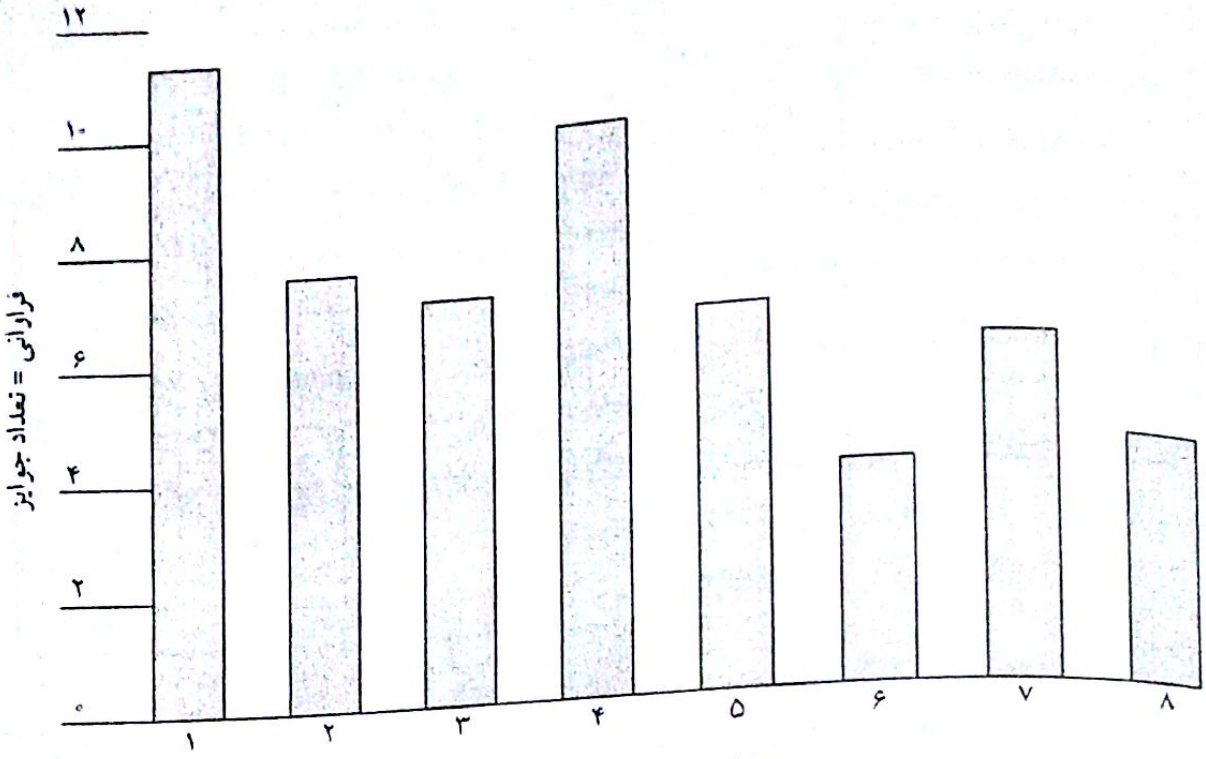
1. bargraph

به رسم هیستوگرام عمل می‌کنیم، یعنی بر بالای هر طبقه از اعداد یک ستون یا یک میلهٔ عمودی به گونه‌ای می‌سازیم که عرض آن معرف فاصلهٔ طبقه و ارتفاع آن نشان‌دهندهٔ فراوانی طبقه باشد. با این حال، به هنگام رسم نمودار ستونی برای داده‌های حاصل از مقیاسهای اسمی دو نکتهٔ زیر را باید در نظر گرفت: نخست اینکه در رسم این نوع نمودارها هیچ‌گونه ترتیبی در نظر گرفته نمی‌شود؛ یعنی طبقات مختلف را می‌توان بر روی محور افقی با هر ترتیب دلبخواهی مثلاً به ترتیب حروف الفبا به دنبال هم قرار داد. نکتهٔ دوم این است که در رسم این نوع نمودارها، برخلاف هیستوگرام، ستونها یا میله‌ها را باید جدا از یکدیگر رسم کرد تا نشان دهد که هیچ‌گونه پیوستگی بین داده‌ها وجود ندارد. شکل ۷ـ۱۸ یک نمونه نمودار ستونی برای داده‌های اسمی ناپیوسته است. این نمودار نشان‌دهندهٔ تعداد دانشجویان رشته‌های مختلف دانشگاهی است که در یک درس روانشناسی مقدماتی ثبت‌نام کرده‌اند.

برای داده‌های مربوط به مقیاس ترتیبی نیز از نمودار ستونی می‌توان استفاده کرد. تفاوت این نمودار با نمودار ستونی مربوط به مقیاس اسمی این است که در نمودار ستونی مربوط به مقیاس اسمی هیچ‌گونه ترتیبی در نظر گرفته نمی‌شود مگر یک ترتیب دلبخواهی، نظیر ترتیب حروف الفبا، اما داده‌های حاصل از مقیاس ترتیبی دارای نوعی ترتیب است که باید به هنگام رسم نمودار ستونی در نظر گرفته شود (مثلاً بزرگ‌تر یا کوچک‌تر). شکل ۸ـ۱۸ نمونه‌ای از یک نمودار ستونی



شکل ۷ـ۱۸ نمودار ستونی تعداد دانشجویان رشته‌های مختلف که در یک درس روانشناسی مقدماتی ثبت‌نام کرده‌اند

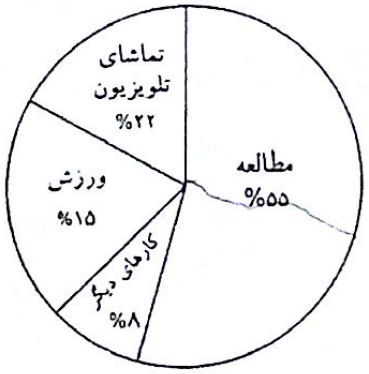


شکل ۸ـ۱۸ نمودار ستونی تعداد جوایزی که دانش‌آموزان رتبهٔ اول تا هشتم یک کلاس در یک مسابقهٔ علمی دریافت کرده‌اند.

برای داده‌های مقیاس ترتیبی است. این نمودار تعداد جوایزی را که هشت دانش‌آموز رتبهٔ اول تا هشتم یک کلاس در یک رشته مسابقهٔ علمی دریافت کرده‌اند نشان می‌دهد.

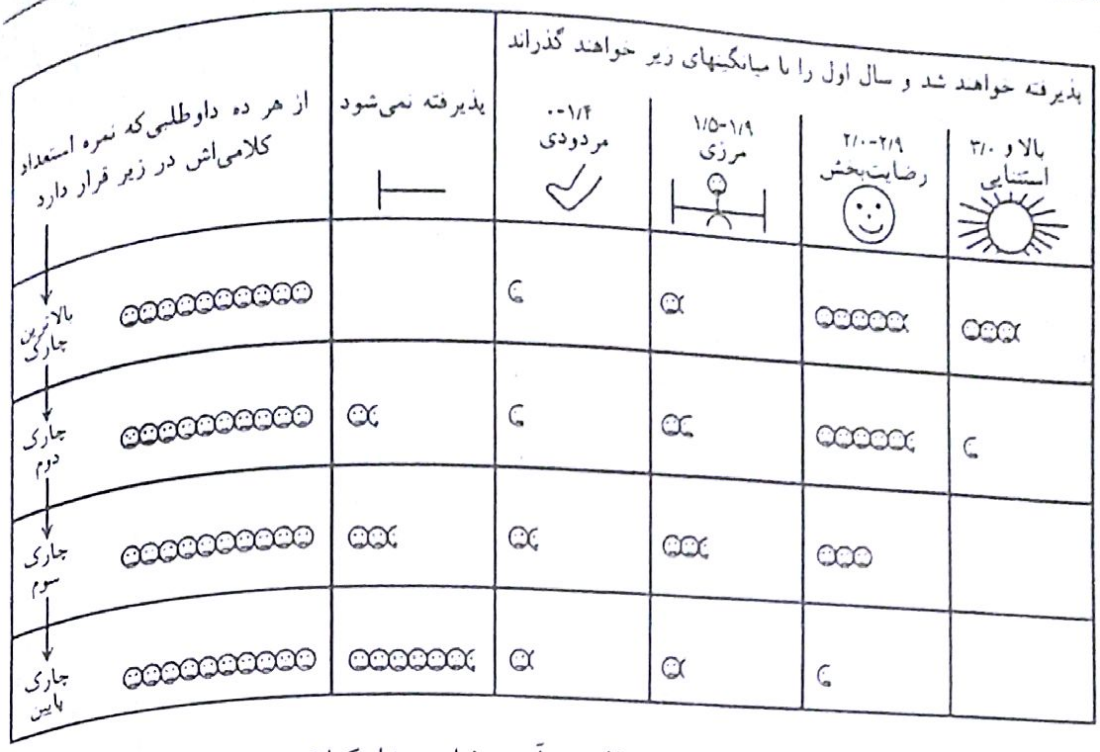
نمودار دایره‌ای

منظور از **نمودار دایره‌ای**[1] یک منحنی دایره‌شکل است که نسبتهای مختلف یک کل را نشان می‌دهد. به عنوان مثال، نمودار شکل ۹–۱۸ نسبتهای ساعتهایی از روز را نشان می‌دهد که دانشجویان یک دانشکده براساس اطلاعاتی که از طریق اجرای یک پرسشنامه به دست آمده صرف مطالعه، تماشای تلویزیون، و ورزش می‌کنند.



شکل ۹–۱۸ ساعتهایی که دانشجویان روزانه صرف مطالعهٔ درسهای خود، تماشای تلویزیون، و ورزش می‌کنند.

1. piegraph



شکل ۱۰-۱۸ نمودار تصویری که رابطهٔ بین آزمون استعداد کلامی مجموعه آزمون استعداد تحصیلی و موفقیت دانشگاهی را در دانشگاه هاوایی نشان می‌دهد (به نقل ساکس، ۱۹۹۷، ص ۱۸۵)

نمودار تصویری

در **نمودار تصویری**[1] تعدادی تصویر یا شکل مورد استفاده قرار می‌گیرند به گونه‌ای که هرکدام مُعرف فراوانی یا درصد وقوع یک رویداد است. نمودار تصویری شکل ۱۰-۱۸ درصد دانشجویان موفق و ناموفق سال اول دانشگاه هاوایی را براساس پیش‌بینی خرده‌آزمون استدلال کلامی آزمونهای استعداد تحلیلی اجرا شده در کلاس یازدهم دبیرستان نشان می‌دهد. نمادهایی که در نمودار تصویری به‌کار می‌روند باید ساده اما معنی‌دار باشند. در ضمن در هر ردیف بیشتر از حدود ۱۵ تصویر نباید استفاده شود زیرا موجب سردرگمی خواهد بود (ساکس، ۱۹۹۷).

اندازه‌های گرایش مرکزی

اگر معلم بخواهد، با استناد به یک نمرهٔ واحد، چگونگی تمرکز نمرات یک کلاس را مشخص کند، از **اندازه‌های گرایش مرکزی**[2] استفاده می‌کند. این اندازه‌ها عبارت‌اند از **نما**[3]، **میانگین**[4] و **میانه**[5].

---

1. pitograph
2. measures of central tendency
3. mode
4. mean
5. median

نما

نمای هر دسته از نمرات عبارت است از *نمره‌ای که بیشترین فراوانی را دارد*. مورد استفادهٔ نما زمانی است که معلم بخواهد، با یک نگاه سطحی، نمره‌ای را به عنوان نمایندهٔ نمرات گروه نشان دهد. با مراجعه به جدول ۱ـ۱۸، خواهیم دید که نمره‌ای که بیش از همه تکرار شده، نمرهٔ ۱۴ است که فراوانی آن ۶ است. پس نمای آن نمرات همین نمرهٔ ۱۴ است. در نمرات طبقه‌بندی شده، نما نقطهٔ وسط طبقه‌ای است که دارای بزرگترین فراوانی است. با این حساب، نمای نمرات طبقه‌بندی شدهٔ جدول ۲ـ۱۸، *نقطهٔ وسط طبقهٔ* ۴۴ـ۴۰، یعنی ۴۲ است.

نما، علاوه بر اینکه به صورت یکی از اندازه‌های گرایش مرکزی به کار می‌رود، برای توصیف شکل بعضی از نمودارها نیز مفید است. اگر نمودار هیستوگرام یا چند ضلعی دسته‌ای از نمرات دارای یک برآمدگی باشد، به توزیع آن نمرات توزیع یک‌نمایی می‌گویند (نگاه کنید به شکلهای ۱ـ۱۸ و ۳ـ۱۸). اگر نمودار دو برآمدگی داشته باشد، به توزیع آن، توزیع دونمایی گویند، مانند شکلهای ۲ـ۱۸ و ۴ـ۱۸. در توزیع نمرات شکلهای ۲ـ۱۸ و ۴ـ۱۸ هم نمرهٔ ۱۴ نمرهٔ نماست و هم نمرهٔ ۱۷، زیرا این دو نمره دو اوج در منحنی ایجاد کرده‌اند، و شکل خاصی به آن داده‌اند. باز هم اگر نمودار توزیعی دارای سه برآمدگی باشد به آن سه‌نمایی گویند، و الی آخر.

میانگین

میانگین یا معدل دسته‌ای از نمرات، بنا به تعریف، عبارت است از *حاصل جمع همهٔ آن نمرات بخش بر تعداد آنها*. فرمول محاسبهٔ میانگین اعداد طبقه‌بندی نشده به قرار زیر است:

$$\bar{X} = \frac{\Sigma X}{N}$$

در این فرمول:

$\bar{X}$ = میانگین

ΣX = مجموع نمرات

N = تعداد نمرات

برای مثال، میانگین نمرات جدول ۱ـ۱۸ به طریق زیر محاسبه می‌شود:

$$\bar{X} = \frac{۲۰\times۱ + ۱۹\times۲ + ۱۸\times۲ + ۱۷\times۵ + ۱۶\times۳ + ۱۵\times۴ + ۱۴\times۶ + ۱۳\times۳}{۳۰}$$

$$+ \frac{۱۲\times۲ + ۱۱\times۱ + ۱۰\times۱}{۳۰}$$

$$= \frac{۴۵۵}{۳۰}$$

$$\bar{X} = ۱۵/۱۶$$

همچنین برای محاسبهٔ نمرات خام ۴۹ دانش‌آموز در درس علوم (صفحهٔ ۴۵۹) باید این نمرات را جمع کرد و حاصل را بر تعداد آنها تقسیم نمود. حاصل جمع این نمرات برابر است با ۱۹۵۳، که وقتی بر تعداد آنها، ۴۹، تقسیم شود میانگین به دست آمده ۳۹/۸۶ خواهد بود.

علاوه بر روش فوق، آنجا که تعداد نمرات زیاد است و لازم است آنها را به صورت نمرات طبقه‌بندی شده ارائه دهیم، روش دیگری را مورد استفاده قرار می‌دهیم. در این روش مراحلی را به شرح زیر طی می‌کنیم:

۱. یکی از طبقات را که ظاهراً نزدیک‌ترین طبقه به میانگین به نظر می‌رسد به عنوان نقطهٔ شروع انتخاب می‌کنیم. این طبقه را طبقهٔ صفر می‌نامیم. برای این منظور، هر یک از طبقات را می‌توان برگزید، اما اگر یکی از طبقات نزدیک به وسط توزیع فراوانی را انتخاب کنیم کار محاسبهٔ میانگین کمتر خواهد بود. در اینجا، برای محاسبهٔ میانگین نمرات جدول ۲ـ۱۸، طبقهٔ ۴۴ـ۴۰ را انتخاب کرده‌ایم. برای انجام محاسبات موردنیاز از جدول ۴ـ۱۸ که همان نمرات جدول ۲ـ۱۸ را نشان می‌دهد استفاده می‌کنیم. عدد میانی طبقهٔ انتخاب شده، یعنی ۴۲، **میانگین فرضی**[1] نام دارد.

۲. مرحلهٔ بعدی ایجاد ستون d است. ارقام این ستون این‌گونه به‌دست می‌آیند که به طبقهٔ انتخابی ارزش صفر می‌دهیم و به اولین طبقهٔ بالاتر از این طبقه ارزش ۱ + و به دومین طبقهٔ بالایی ۲+، و به همین ترتیب ارزشهایی به سایر طبقات می‌دهیم. برای طبقات پایین‌تر از طبقهٔ دارای ارزش صفر نیز به ترتیب ارزشهای ۱ـ، ۲ـ، ... الی آخر را قائل می‌شویم.

۳. ستون بعدی fd است. برای محاسبهٔ ارقام این ستون اعداد دو ستون قبلی یعنی f (فراوانیها) و d (انحرافها) را درهم ضرب می‌کنیم و حاصل‌ضربها را در ستون fd وارد می‌کنیم.

۴. حال برای محاسبهٔ میانگین از فرمول زیر استفاده می‌کنیم:

$$\text{میانگین} = \bar{X} = \text{A. M.} + \left(\frac{\Sigma fd}{N}\right) \text{c.i.}$$

در این فرمول:

$\bar{X}$ = میانگین

A. M. = میانگین فرضی (عدد وسط طبقهٔ انتخابی)

$\Sigma fd = d$ و f حاصل جمع حاصل‌ضربهای دو ستون

c. i. = فاصلهٔ طبقاتی

N = تعداد نمرات

---

1. assumed mean

جدول ۴ـ۱۸ توزیع نمرات طبقه‌بندی شده، برای محاسبهٔ میانگین

فاصلهٔ طبقه	f	d	fd
۷۰ـ۷۴	۱	۶	۶
۶۵ـ۶۹	۲	۵	۱۰
۶۰ـ۶۴	۲	۴	۸
۵۵ـ۵۹	۳	۳	۹
۵۰ـ۵۴	۳	۲	۶
۴۵ـ۴۹	۵	۱	۵
			+ ۴۴
۴۰ـ۴۴	۱۱	۰	۰
۳۵ـ۳۹	۸	-۱	-۸
۳۰ـ۳۴	۳	-۲	-۶
۲۵ـ۲۹	۳	-۳	-۹
۲۰ـ۲۴	۳	-۴	-۱۲
۱۵ـ۱۹	۲	-۵	-۱۰
۱۰ـ۱۴	۱	-۶	-۶
۵ـ۹	۱	-۷	-۷
۰ـ۴	۱	-۸	-۸
	N = ۴۹		Σfd = + ۴۴ ـ ۶۶ = ـ۲۲

برای محاسبهٔ میانگین اعداد جدول ۴ـ۱۸، به طریق زیر عمل می‌کنیم:

$$\bar{X} = ۴۲ + \left(\frac{۴۴ - ۶۶}{۴۹}\right) ۵$$

$$= ۴۲ + \left(\frac{-۲۲}{۴۹}\right) ۵$$

$$= ۴۲ - ۲/۲۴$$

$$= ۳۹/۷۶$$

رقم ۳۹/۷۶ خیلی نزدیک به رقم ۳۹/۸۶ است که برای نمرات خام (طبقه‌بندی نشده) به دست آمد. این دو رقم اکثراً با هم تفاوت دارند، و دلیل آن هم این است که ما، در دسته کردن نمرات به طبقات، مقداری از دقت عمل خود را از دست می‌دهیم.

ویژگیهای میانگین

میانگین به صورت بهترین اندازهٔ گرایش مرکزی معرفی شده است، و معمولاً این اندازه به تنهایی بهترین معرف گروه نمرات است. با وجود این، بزرگ‌ترین عیب میانگین این است که نسبت به

نمرات خیلی بزرگ یا خیلی کوچک زیاد حساس است، یعنی یک نمرهٔ خیلی زیاد یا خیلی کم ممکن است میانگین را به مقدار چشم‌گیری بالا یا پایین ببرد. در چنین حالتی به جای میانگین معمولاً از میانه استفاده می‌شود.

از ویژگیهای دیگر میانگین یکی این است که اگر عدد ثابتی به کلیهٔ نمرات گروه اضافه شود یا از آنها کم شود، به میانگین هم عددی به همان اندازه اضافه یا از آن کم می‌شود. همچنین اگر همهٔ نمرات در عدد ثابتی ضرب یا بر آن تقسیم شوند، میانگین این اعداد نیز در آن عدد ثابت ضرب یا بر آن تقسیم می‌شود.

میانه

بنا به تعریف، میانه نقطه‌ای است در روی مقیاس نمرات یا توزیع نمرات که نصف بالا و نصف پایین گروه نمرات را از یکدیگر جدا می‌کند. این نقطه‌ای است که ۵۰ درصد نمرات در بالای آن و ۵۰ درصد دیگر نمرات در پایین آن قرار دارند. این نقطه گاه یک نمرهٔ واقعی و گاه عددی به غیر از نمرات است. برای مثال، میانهٔ نمرات زیر:

۲۸ ۲۴ ۲۲ ۲۱ ۲۰ ۱۸ ۱۷

برابر است با ۲۱، زیرا این عدد درست در وسط توزیع نمرات قرار دارد، و در بالای آن ۳ نمره و در پایین آن نیز ۳ نمره وجود دارند. از سوی دیگر، میانهٔ نمرات زیر:

۲۹ ۲۸ ۲۳ ۲۲ ۲۱ ۲۰ ۱۸ ۱۷

۲۱/۵ است، زیرا این عدد نقطه‌ای را مشخص می‌کند که ۵۰ درصد نمرات (۴ نمره) در بالای آن و ۵۰ درصد دیگر نمرات (۴ نمره) در پایین آن قرار دارند. پس برای پیداکردن میانهٔ نمرات طبقه‌بندی نشده، می‌توان از یک سر توزیع نمرات شروع کرد و به شمارش اعداد پرداخت تا وقتی که به نقطهٔ میانی توزیع برسیم.

با توجه به مثالهای بالا، در مواردی که نمرات اعداد یک توزیع فرد است نمرهٔ وسطی توزیع به عنوان میانه انتخاب می‌شود، اما در توزیعاتی که تعداد نمرات آنها زوج است دو نمرهٔ میانی را باید با هم جمع و حاصل را بر ۲ تقسیم کنیم تا میانهٔ آن اعداد به دست آید، یعنی:

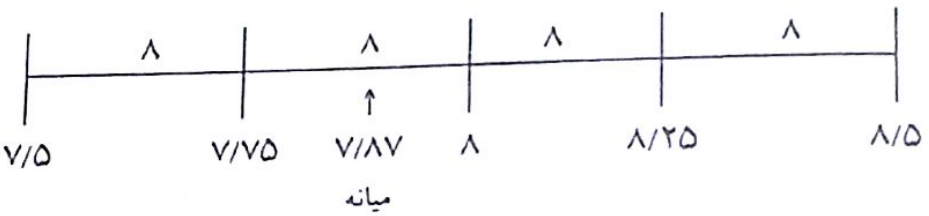
$$\text{میانه} = Md = \frac{\text{مجموع دو نمرهٔ میانی}}{۲}$$

در مواردی که نمرهٔ میانی دارای فراوانی بیش از یک است، مثل نمرات زیر، محاسبهٔ میانه نیاز به کوشش بیشتری دارد.

۵ ۶ ۶ ۷ ۸ ۸ ۸ ۸ ۱۱ ۱۳ ۱۵

برای این منظور باید مفهوم طبقهٔ اعداد را در مورد اعداد طبقه‌بندی نشده نیز مورد استفاده قرار دهیم. قبلاً گفتیم که فاصلهٔ طبقه عبارت از حدودی است که به طور قراردادی برای آن طبقه مشخص می‌شود. در اعداد طبقه‌بندی نشده نیز فاصلهٔ طبقه را می‌توان در نظر گرفت. در این اعداد هر عدد یک فاصلهٔ طبقه را شامل می‌شود که حدود آن از نیم نمره پایین‌تر از حد عدد تا نیم نمره بالاتر از آن گسترش می‌یابد. برای مثال، فاصلهٔ طبقهٔ مربوط به عدد ۸ از ۷/۵ است تا ۸/۵.

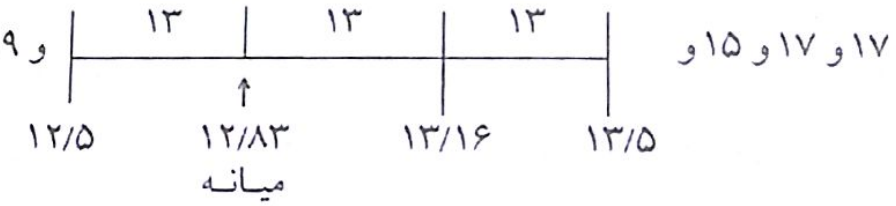
در توزیع نمرات بالا، نمرهٔ ۸ دارای فراوانی ۴ است. چون این نمره دارای فاصلهٔ طبقهٔ ۸/۵ـ۷/۵ است، فرض می‌شود که ۴ نمرهٔ ۸ به طور مساوی در این طبقه پراکنده شده‌اند و هریک از این نمرات مطابق شکل زیر ۰/۲۵ واحد (۰/۲۵ = ۴ ÷ ۱) را در بر دارد.



از آنجا که نمرهٔ میانی نمرات پیش‌گفته در محدودهٔ دومین نمرهٔ ۸ از شکل فوق واقع است، باید نقطهٔ وسط فاصلهٔ مربوط به ۸ دوم یعنی ۸ـ۷/۷۵ را حساب کنیم که می‌شود $\frac{۸ + ۷/۷۵}{۲} = ۷/۸۷$، و این میانهٔ نمرات بالاست.

به نمونهٔ دیگری از محاسبهٔ میانه برای نمرات زیر توجه کنید: ۱۷ و ۱۷ و ۱۵ و ۱۳ و ۱۳ و ۱۳ و ۹ و ۷ و ۵ و ۲.

تعداد نمرات بالا زوج است و میانهٔ آنها نقطه‌ای است بین نمرهٔ پنجم و نمرهٔ ششم، یعنی بین ۱۳ اول و ۱۳ دوم. با توجه به شکل زیر فاصلهٔ طبقهٔ نمرهٔ ۱۳ از ۱۲/۵ است تا ۱۳/۵، و از آنجا که در این فاصله سه نمرهٔ ۱۳ موجود است، سهم هر نمرهٔ ۱۳ برابر است با ۰/۳۳ = ۳ ÷ ۱ و چون میانهٔ نمرات درست در بین دو نمرهٔ ۱۳ سمت چپ شکل زیر قرار دارد، این نقطه دارای ارزش عددی ۱۲/۸۳ است.



برای محاسبهٔ میانهٔ نمرات طبقه‌بندی شده، از فراوانی تراکمی نمرات استفاده می‌کنیم (جدول ۵ـ۱۸). در ستون فراوانیهای تراکمی جدول ۵-۱۸، هر یک از اعداد مقابل طبقات، نشان‌دهندهٔ تعداد نمراتی است که از حد بالایی آن طبقه کوچک‌تر یا با آن مساوی است. مثلاً از تمام نمرات جدول ۵ـ۱۸، پنج نمره وجود دارند که از ۱۹ کوچک‌تر یا با آن برابر هستند؛ یا ۸ نمره مساوی با ۲۴ و یا کوچک‌تر از آن وجود دارند، و الی آخر. گفته شد که، در تعیین میانه، به دنبال

جدول ۵ـ۱۸ توزیع فراوانی به همراه فراوانی تراکمی برای محاسبهٔ میانه

فاصلهٔ طبقه	فراوانی	فراوانی تراکمی	
C. I.	f	cum.f	
۷۰ـ۷۴	۱	۴۹	
۶۵ـ۶۹	۲	۴۸	
۶۰ـ۶۴	۲	۴۶	
۵۵ـ۵۹	۳	۴۴	
۵۰ـ۵۴	۳	۴۱	
۴۵ـ۴۹	۵	۳۸	
۴۰ـ۴۴	۱۱	۳۳	طبقهٔ میانه
۳۵ـ۳۹	۸	۲۲	
۳۰ـ۳۴	۳	۱۴	
۲۵ـ۲۹	۳	۱۱	
۲۰ـ۲۴	۳	۸	
۱۵ـ۱۹	۲	۵	
۱۰ـ۱۴	۱	۳	
۵ـ۹	۱	۲	
۰ـ۴	۱	۱	

N = ۴۹

نقطه‌ای هستیم که در بالا و پایین آن ۵۰ درصد نمرات قرار دارند. چون ۵۰ درصد ۴۹ برابر است با ۲۴/۵، باید در جدول ۵ـ۱۸ طبقه‌ای را بیابیم که نقطهٔ دارای فراوانی تراکمی ۲۴/۵ (نقطهٔ ۵۰ درصدی) در آن واقع است.

چون، طبق جدول ۵ـ۱۸، ۲۲ نمره پایین‌تر از طبقهٔ ۴۴ـ۴۰ قرار دارند، و تعداد نمرات پایین‌تر از طبقهٔ بعدی، یعنی ۴۹ـ۴۵، برابر با ۳۳ است، معلوم می‌شود که نقطهٔ مربوط به ۲۴/۵ جایی در طبقهٔ ۴۴ـ۴۰ قرار دارد. دقت کنید که در طبقهٔ ۴۴ـ۴۰ یازده نمره قرار دارد، و ما تنها به $\frac{۲/۵}{۱۱}$ نمره (۲/۵ = ۲۲ ـ ۲۴/۵) از این نمرات نیاز داریم. چنان‌که پیشتر گفته شد، فرض طبقه‌بندی نمرات این است که تمام نمرات یک طبقه به طور مساوی در آن طبقه پراکنده شده‌اند. بنابراین، برای افزودن $\frac{۲/۵}{۱۱}$ نمره از نمرات طبقهٔ ۴۴ـ۴۰ به حد پایینی این طبقه، لازم است در این طبقه از نقطهٔ پایین آن آغاز کنیم و به اندازهٔ $\frac{۲/۵}{۱۱}$ نمره از نمرات آن طبقه بالا برویم تا نقطهٔ میانه را پیدا کنیم، یعنی در واقع به $\frac{۲/۵}{۱۱}$ نمره از ۵ نمرهٔ این طبقه نیاز داریم (۵ فاصلهٔ طبقه است).

در اینجا لازم است معنی نمرهٔ ۴۰ را توضیح دهیم. نمرات حاصل از یک آزمون که به صورت

اعداد ۴۰، ۴۱، ۴۲، و غیره نشان داده می‌شوند، نشان‌دهندهٔ یک توانایی هستند که دارای توزیعی پیوسته است و کلیهٔ اعداد حد فاصل بین واحدهای موجود را نیز شامل می‌شود. می‌توان این وضعیت را به ارقام روی یک صفحهٔ ساعت برقی تشبیه کرد. اگر چه زمان پیوسته است، دقیقه شمار ساعت به طور جهشی حرکت می‌کند و در هر جهش تمام واحد دقیقه را طی می‌نماید. بنابراین، ۴۰ فاصله‌ای از ۳۹/۵ تا ۴۰/۵ را شامل می‌شود. همچنین طبقهٔ ۴۴ ـ ۴۰ در واقع از ۳۹/۵ تا ۴۴/۵ را دربر می‌گیرد.

با این توصیف، حدود واقعی طبقه‌ای که میانه در آن قرار دارد از ۳۹/۵ تا ۴۴/۵ است، و ما باید از ۳۹/۵ به بالا پیش برویم و $\frac{۲/۵}{۱۱} \times ۵$ نمره جدا کنیم و به حد پایین طبقه بیفزاییم. برای انجام این کار و محاسبهٔ میانه از فرمول زیر استفاده می‌کنیم.

$$\text{میانه} = Md = \text{حد پایین طبقهٔ میانه} + \left(\frac{\frac{N}{۲} - \text{فراوانی تراکمی پیش از طبقهٔ میانه}}{\text{فراوانی طبقهٔ میانه}} \right) (\text{فاصلهٔ طبقات})$$

$$Md = ۳۹/۵ + \left[\frac{\frac{۴۹}{۲} - ۲۲}{۱۱} \right] \times ۵$$

$$= ۳۹/۵ + \left[\frac{۲۴/۵ - ۲۲}{۱۱} \right] \times ۵$$

$$= ۳۹/۵ + ۰/۲۳ \times ۵$$

$$= ۳۹/۵ + ۱/۱۵$$

$$= ۴۰/۶۵$$

اندازه‌های تغییرپذیری

اگر دو دسته نمره با میانگین مساوی داشته باشیم و تنها اطلاع ما در بارهٔ نمرات میانگین آنها باشد تنها مطلبی که در مورد این دو دسته نمره می‌توانیم اظهار کنیم این است که این دو دسته نمره دارای میانگین مساوی هستند و چیز دیگری در بارهٔ آنها نمی‌توانیم بگوییم. ممکن است یکی از این دو توزیع دارای بزرگ‌ترین نمرهٔ ۲۰ و کوچک‌ترین نمرهٔ ۳، ولی توزیع دیگر دارای بزرگ‌ترین نمرهٔ ۱۵ و کوچکترین نمرهٔ ۱۰ باشند، و میانگین این دو توزیع هر دو ۱۳ باشد. بنابراین، برای معرفی یک توزیع، نه تنها به یکی از اندازه‌های گرایش مرکزی نیاز داریم، بلکه در دست داشتن یکی از **اندازه‌های تغییرپذیری**[1] نیز ضرورت دارد. اندازه‌های معروف تغییرپذیری عبارت اند از **دامنهٔ تغییر**[2]، **واریانس یا پراکنش**[3]، و **انحراف معیار**[4].

---

1. measures of variabilty
2. range
3. variance
4. standard deviation

دامنهٔ تغییر

ساده‌ترین اندازهٔ تغییرپذیری دامنهٔ تغییر نمرات است. دامنهٔ تغییر گروهی نمره برابر است با فاصلهٔ بین بزرگ‌ترین و کوچک‌ترین نمرهٔ آن گروه. برای مثال، دامنهٔ تغییر نمرات درس علوم، صفحهٔ ۴۵۹، برابر است با ۶۸ = ۳ ـ ۷۱. این اندازه یک شاخص بسیار ابتدایی و بی‌ثبات است، زیرا در تعیین آن تنها دو نمره دخالت دارند و یک نمرهٔ بسیار بزرگ یا کوچک مقدار آن را تغییر می‌دهد.

واریانس و انحراف معیار

واریانس (پراکنش) و انحراف معیار بهترین اندازه‌های تغییرپذیری هستند. برای محاسبهٔ این دو اندازه در نمرات طبقه‌بندی نشده مراحل زیر را طی می‌کنیم:

۱. ابتدا میانگین نمرات را تعیین می‌کنیم. میانگین این نمرات ۱۵، ۱۲، ۹، ۱۰، ۱۴ برابر است با:

$$\text{میانگین} = \bar{X} = \frac{۱۵ + ۱۲ + ۹ + ۱۰ + ۱۴}{۵} = \frac{۶۰}{۵}$$

$$\bar{X} = ۱۲$$

۲. بعد تفاضل هر یک از نمرات را از میانگین حساب می‌کنیم. اعداد حاصل را نمرات انحراف[1] نامند.

نمرات	میانگین	نمرات انحراف
X	$\bar{X}$	x
۱۵	۱۲ =	۳
۱۴	۱۲ =	۲
۱۲	۱۲ =	۰
۱۰	۱۲ =	۲ -
۹	۱۲ =	۳ -

۳. اگر می‌شد میانگین نمرات انحراف را حساب کنیم، این میانگین به عنوان شاخص بسیار خوبی برای پراکندگی اعداد به حساب می‌آمد. اما چون تعدادی از این نمرات انحراف علامت منفی دارند، و جمع آنها عدد صفر را به دست می‌دهد، این نمرات را مجذور می‌کنیم تا علامتهای منفی آنها به مثبت تبدیل شود.

X	$\bar{X}$		x	$x^۲$
۱۵	۱۲	=	۳	۹
۱۴	۱۲	=	۲	۴
۱۲	۱۲	=	۰	۰
۱۰	۱۲	=	۲ -	۴
۹	۱۲	=	۳ -	۹
				$\Sigma x^۲ = ۲۶$

1. deviation scores

۴. اگر حاصل جمع مجذورات نمرات انحراف (ΣX^2) را بر تعداد نمرات تقسیم کنیم، عدد حاصل، یعنی میانگین مجذورات نمرات انحراف، واریانس یا پراکنش نمرات است.

$$\text{واریانس} = \frac{\Sigma x^2}{N}$$
$$= \frac{۲۶}{۵}$$
$$= ۵/۲$$

۵. به خاطر اینکه برای از بین بردن علامتهای منفی نمرات انحراف آنها را مجذور کردیم، یعنی مقیاس نمرات را به هم زدیم، برای برگرداندن عدد حاصل به مقیاس قبلی، یعنی همان مقیاس اعداد اصلی، لازم است که از واریانس جذر بگیریم. جذر واریانس انحراف معیار نام دارد. بنابراین، انحراف معیار نمرات طبقه‌بندی نشده را با استفاده از فرمول زیر می‌توان حساب کرد.

$$\sigma = s = SD = \sqrt{\frac{\Sigma x^2}{N}}$$

در این فرمول:

$\sigma = s = SD$ = انحراف معیار

Σx^2 = مجموع مجذورات نمرات انحراف

N = تعداد نمرات

$$SD = \sqrt{\frac{۲۶}{۵}}$$

$$SD = ۲/۳$$

اگر مثلاً بخواهیم مستقیماً از نمرات خام برای محاسبهٔ انحراف معیار استفاده کنیم، بدون اینکه نیازی به تعیین Σx^2 داشته باشیم، از فرمول زیر استفاده می‌کنیم:

$$SD = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

در این فرمول:

ΣX^2 = مجموع مجذورات نمرات خام

$(\Sigma X)^2$ = مجذور مجموع نمرات خام

برای محاسبهٔ انحراف معیار پنج نمرهٔ بالا، با استفاده از فرمول فوق، به شرح زیر عمل می‌کنیم:

X	X^2
۱۵	۲۲۵
۱۴	۱۹۶
۱۲	۱۴۴
۱۰	۱۰۰
۹	۸۱
$\Sigma X = ۶۰$	$\Sigma X^2 = ۷۴۶$

$$SD = \sqrt{\frac{۷۴۶ - \frac{(۶۰)^2}{۵}}{۵}}$$

$$= \sqrt{\frac{۷۴۶ - ۷۲۰}{۵}}$$

$$= \sqrt{\frac{۲۶}{۵}}$$

$$= ۲/۳$$

محاسبهٔ انحراف معیار نمره‌های طبقه‌بندی شده

برای محاسبهٔ انحراف معیار نمرات طبقه‌بندی شده، باید از جدولی شبیه به جدول ۶ـ۱۸ استفاده کنیم. این جدول همان جدول ۴ـ۱۸ است که ستون fd^2 به آن اضافه شده است. ارقام این ستون هریک معرف تعداد موارد یک طبقه یا فراوانی آن طبقه (f) ضرب در مجذور نمرهٔ انحراف آن طبقه از مبدأ انتخابی (d^2) است. مجموع ارقام این ستون معرف حاصل جمع مجذورات نمرات انحراف است که به صورت واحدهای فاصلهٔ طبقاتی معرفی شده است. به منظور اینکه این انحرافها را به واحد نمرات تبدیل کنیم و آنها را نسبت به میانگین حقیقی بسنجیم، باید عملیاتی را به شرح زیر انجام دهید.

۱. میانگین نمرات را با توجه به توضیحاتی که قبلاً در بخش میانگین داده شد حساب کنید.

۲. ستون fd^2 را کامل کنید. برای محاسبهٔ ارقام این ستون فراوانی طبقه (f) را در مجذور انحراف آن طبقه از مبدأ انتخابی (d^2) ضرب کنید. برای انجام این کار، دو ستون ماقبل آخر جدول ۶ـ۱۸، یعنی d و fd را در هم ضرب کنید.

۳. همهٔ ارقام ستون fd^2 را با هم جمع کنید تا Σfd^2 به دست آید.

۴. اعداد لازم را در جای مناسبشان در فرمول زیر قرار دهید تا انحراف معیار نمرات به دست آید.

در این فرمول:

$$SD = C.I. \sqrt{\frac{\Sigma fd^2}{N} - \left[\frac{\Sigma fd}{N}\right]^2}$$

SD = انحراف معیار

C. I. = فاصلهٔ طبقه

Σfd = حاصل جمع نمرات انحراف از میانگین

Σfd^2 = حاصل جمع مجذورات نمرات انحراف

انحراف معیار نمرات جدول ۶ـ۱۸، با استفاده از فرمول بالا، به صورت زیر محاسبه می‌شود:

$$SD = ۵\sqrt{\frac{۴۵۶}{۴۹} - \left[\frac{۲۲}{۴۹}\right]^2}$$

$$= ۵\sqrt{۹/۳۰۶۱ - (۰/۴۵)^2}$$

$$= ۵\sqrt{۹/۱۰۳۶}$$

$$= ۱۵/۱۰$$

جدول ۶ـ۱۸ توزیع نمرات طبقه‌بندی شده، برای محاسبهٔ واریانس و انحراف معیار

طبقات	f	d	fd	fd^2
۷۰_۷۴	۱	۶	۶	۳۶
۶۵_۶۹	۲	۵	۱۰	۵۰
۶۰_۶۴	۲	۴	۸	۳۲
۵۵_۵۹	۳	۳	۹	۲۷
۵۰_۵۴	۳	۲	۶	۱۲
۴۵_۴۹	۵	۱	۵	۵
			+ ۴۴	
۴۰_۴۴	۱۱	۰	۰	۰
۳۵_۳۹	۸	-۱	-۸	۸
۳۰_۳۴	۳	-۲	-۶	۱۲
۲۵_۲۹	۳	-۳	-۹	۲۷
۲۰_۲۴	۳	-۴	-۱۲	۴۸
۱۵_۱۹	۲	-۵	-۱۰	۵۰
۱۰_۱۴	۱	-۶	-۶	۳۶
۵_۹	۱	-۷	-۷	۴۹
۰_۴	۱	-۸	-۸	۶۴
	N = ۴۹		Σfd = - ۲۲	Σfd^2 = ۴۵۶

$$SD = C.I. \sqrt{\frac{\Sigma fd^2}{N} - \left[\frac{\Sigma fd}{N}\right]^2}$$

در این فرمول:

SD = انحراف معیار

C. I. = فاصلهٔ طبقه

Σfd = حاصل جمع نمرات انحراف از میانگین

Σfd^2 = حاصل جمع مجذورات نمرات انحراف

انحراف معیار نمرات جدول ۶ـ۱۸، با استفاده از فرمول بالا، به صورت زیر محاسبه می‌شود:

$$SD = ۵\sqrt{\frac{۴۵۶}{۴۹} - \left[\frac{۲۲}{۴۹}\right]^2}$$

$$= ۵\sqrt{۹/۳۰۶۱ - (۰/۴۵)^2}$$

$$= ۵\sqrt{۹/۱۰۳۶}$$

$$= ۱۵/۱۰$$

جدول ۶ـ۱۸ توزیع نمرات طبقه‌بندی شده، برای محاسبهٔ واریانس و انحراف معیار

طبقات	f	d	fd	fd^2
۷۰ـ۷۴	۱	۶	۶	۳۶
۶۵ـ۶۹	۲	۵	۱۰	۵۰
۶۰ـ۶۴	۲	۴	۸	۳۲
۵۵ـ۵۹	۳	۳	۹	۲۷
۵۰ـ۵۴	۳	۲	۶	۱۲
۴۵ـ۴۹	۵	۱	۵	۵
			+۴۴	
۴۰ـ۴۴	۱۱	۰	۰	۰
۳۵ـ۳۹	۸	-۱	-۸	۸
۳۰ـ۳۴	۳	-۲	-۶	۱۲
۲۵ـ۲۹	۳	-۳	-۹	۲۷
۲۰ـ۲۴	۳	-۴	-۱۲	۴۸
۱۵ـ۱۹	۲	-۵	-۱۰	۵۰
۱۰ـ۱۴	۱	-۶	-۶	۳۶
۵ـ۹	۱	-۷	-۷	۴۹
۰ـ۴	۱	-۸	-۸	۶۴
	N = ۴۹		Σfd = -۲۲	Σfd^2 = ۴۵۶

لازم به تذکر است که مجذور انحراف معیار واریانس نمرات را به دست می‌دهد. همچنان که انحراف معیار از جذرگرفتن واریانس به دست می‌آید. به طور خلاصه واریانس و انحراف معیار را این‌گونه می‌توان تفسیر کرد که هرچه واریانس و انحراف معیار بزرگ‌تر باشند، بدین علت است که پراکندگی نمرات زیادتر است، و هرچه این دو شاخص آماری کوچکتر باشند، دلیلش آن است که پراکندگی نمرات کمتر است. به سخن دیگر، ما از بزرگی و کوچکی مقدار واریانس و انحراف معیار به زیادی و کمی پراکندگی نمرات پی می‌بریم.

همبستگی بین نمرات

در تعبیر و تفسیر نمرات دانش‌آموزان و دانشجویان، گاه لازم می‌آید که رابطهٔ بین دو یا چند دسته نمره را تعیین کنیم. برای مثال، اغلب سؤال می‌شود که نمرات هوش و نمرات پیشرفت تحصیلی چه رابطه‌ای با هم دارند، و بسیاری از معلمان علاقه‌مندند که بدانند بین نمرات پیشرفت تحصیلی سالهای قبل و بعد رابطه‌ای وجود دارد یا نه. میزان رابطهٔ بین دو دسته نمرهٔ متعلق به گروه واحدی از دانش‌آموزان یا دانشجویان را با محاسبهٔ **ضریب همبستگی**[1] مشخص می‌کنند. متداول‌ترین روش تعیین همبستگی، روش محاسبهٔ ضریب **همبستگی گشتاوری پیرسون**[2] است. این ضریب همبستگی با r نشان داده می‌شود.

ضریب همبستگی مقادیر از ۱+ تا ۱- را شامل می‌شود که از این مقادیر ۱+ نشان‌دهندهٔ رابطهٔ کامل مثبت بین دو متغیر، ۱- نشان‌دهندهٔ رابطهٔ کامل منفی بین دو متغیر، و صفر حاکی از نبود رابطهٔ بین دو متغیر است. برای مثال، همبستگی بین قطرهای تعدادی دایره و محیطهای آنها همبستگی کامل مثبت است، زیرا دایره‌ای که قطرش از همه بیشتر است اندازهٔ محیطش نیز از سایر دایره‌ها بیشتر است، و آن دایره‌ای که قطرش از همه کمتر است محیطش نیز از همه کمتر است، و الی آخر. از سوی دیگر، همبستگی بین قد و وزن تعدادی دانش‌آموز، هرچند مثبت است، اما همبستگی کامل نیست، زیرا گرچه معمولاً کسانی که قد بلندتری دارند وزنشان هم بیشتر است، این رابطه، آن‌چنان که در مورد قطر و محیط دایره‌ها صادق است، در اینجا به طور کامل صدق نمی‌کند، یعنی همیشه دانش‌آموزانی که قدشان از بقیهٔ دانش‌آموزان بلندتر است، دارای بیشترین وزن نیستند.

وضعیت فوق در مورد رابطهٔ بین نمرات هوش و نمرات پیشرفت تحصیلی در درسهای مختلف نیز صادق است. همچنین ضریب همبستگی محاسبه شده بین نمرات پیشرفت تحصیلی قبلی و بعدی و نمرات درسهای مختلف نشان داده‌اند که، گرچه بین این متغیرها ضریبهای بزرگ‌تر از صفر محاسبه شده است، اما این ضریبهای همبستگی نشان‌دهندهٔ رابطهٔ کامل نیستند.

افزون بر موارد بالا، متغیرها یا ویژگیهایی وجود دارند که بین آنها هیچ‌گونه رابطه یا

1. correlation coefficient 2. Pearson product moment correlation

جدول ۷ـ۱۸ ضریبهای همبستگی بین تعدادی متغیر

متغیرها	ضریب همبستگی
قد دوقلوهای همسان	۰/۹۵
نمرات هوش دوقلوهای همسان	۰/۸۸
نمرات آزمون خواندن کلاس سوم و ششم	۰/۸۰
قد و وزن کودکان ده‌ساله	۰/۶۰
نمرات آزمون محاسبات ریاضی و آزمون هوش غیرکلامی (کلاس ۸)	۰/۵۴
قد برادران (اصلاح‌شده از لحاظ سنی)	۰/۵۰
نمرات آزمون هوش کودکان و سطح شغلی پدران و مادران	۰/۳۰
اندازهٔ قد و هوشبهر (آزمون بینه)	۰/۶
نسبت طول و عرض سر و هوش	۰/۰۱
علاقهٔ هنرمندان و علاقهٔ بانکداران	ـ۰/۶۴
نمرات شایستگی افراد اعزامی به ارتش و تعداد سالهای مردودی در مدرسه	ـ۰/۲۷

همبستگی وجود ندارد. برای مثال، بین نمرات هوش گروهی دانش‌آموز و اندازهٔ قد آنها هیچ‌گونه رابطه‌ای دیده نشده است. در این‌گونه موارد، گفته می‌شود که بین متغیرها، مثلاً هوش و قد، همبستگی وجود ندارد.

باز هم ممکن است بین دو ویژگی یا دو متغیر رابطه‌ای به غیر از آنچه در مثالهای فوق گفته شد وجود داشته باشد. مشاهده شده است که دانش‌آموزان مسن دبستانی نسبت به سایر همکلاسیهای کم‌سن‌تر خود، از لحاظ پیشرفت تحصیلی، ضعیف‌تر هستند. در این‌گونه موارد، گفته می‌شود که متغیر سن و پیشرفت تحصیلی با یکدیگر رابطهٔ عکس دارند، یا به اصطلاح آماری با هم همبستگی منفی نشان می‌دهند. راننده‌ای را تصور کنید که به سمت مقصدی در حال حرکت است. هرچه این راننده سریعتر حرکت کند مدت زمان کمتری برای رسیدن به مقصد لازم خواهد داشت. بنابراین، بین سرعت رانندگی و مدت زمان لازم برای رسیدن به مقصد، ضریب همبستگی منفی وجود دارد. ثرندایک و هیگن (۱۹۷۹، ص ۵۱) ضرایب همبستگی بین تعدادی متغیر را به دست داده‌اند که برخی از آنها در جدول ۷ـ۱۸ نقل می‌شود.

روش محاسبهٔ ضریب همبستگی گشتاوری پیرسون

همان‌طور که پیش‌تر گفته شد، متداول‌ترین روش همبستگی، همبستگی گشتاوری پیرسون است. برای نشان دادن مراحل محاسبهٔ این ضریب همبستگی نمرات جدول ۸ـ۱۸ را مورد استفاده قرار می‌دهیم.

دو دسته نمرهٔ X و Y جدول ۸ـ۱۸ را نمرات جبر (X) و هندسهٔ (Y) یک گروه پنج‌نفری از دانش‌آموزان یک کلاس دبیرستان فرض کنید. برای محاسبهٔ ضریب همبستگی بین این دو دسته

جدول ۷ـ۱۸ ضریبهای همبستگی بین تعدادی متغیر

متغیرها	ضریب همبستگی
قد دوقلوهای همسان	۰/۹۵
نمرات هوش دوقلوهای همسان	۰/۸۸
نمرات آزمون خواندن کلاس سوم و ششم	۰/۸۰
قد و وزن کودکان ده‌ساله	۰/۶۰
نمرات آزمون محاسبات ریاضی و آزمون هوش غیرکلامی (کلاس ۸)	۰/۵۴
قد برادران (اصلاح‌شده از لحاظ سنی)	۰/۵۰
نمرات آزمون هوش کودکان و سطح شغلی پدران و مادران	۰/۳۰
اندازهٔ قد و هوشبهر (آزمون بینه)	۰/۶
نسبت طول و عرض سر و هوش	۰/۰۱
علاقهٔ هنرمندان و علاقهٔ بانکداران	-۰/۶۴
نمرات شایستگی افراد اعزامی به ارتش و تعداد سالهای مردودی در مدرسه	-۰/۲۷

همبستگی وجود ندارد. برای مثال، بین نمرات هوش گروهی دانش‌آموز و اندازهٔ قد آنها هیچ‌گونه رابطه‌ای دیده نشده است. در این‌گونه موارد، گفته می‌شود که بین متغیرها، مثلاً هوش و قد، همبستگی وجود ندارد.

باز هم ممکن است بین دو ویژگی یا دو متغیر رابطه‌ای به غیر از آنچه در مثالهای فوق گفته شد وجود داشته باشد. مشاهده شده است که دانش‌آموزان مسن دبستانی نسبت به سایر همکلاسیهای کم‌سن‌تر خود، از لحاظ پیشرفت تحصیلی، ضعیف‌تر هستند. در این‌گونه موارد، گفته می‌شود که متغیر سن و پیشرفت تحصیلی با یکدیگر رابطهٔ عکس دارند، یا به اصطلاح آماری با هم همبستگی منفی نشان می‌دهند. راننده‌ای را تصور کنید که به سمت مقصدی در حال حرکت است. هرچه این راننده سریعتر حرکت کند مدت زمان کمتری برای رسیدن به مقصد لازم خواهد داشت. بنابراین، بین سرعت رانندگی و مدت زمان لازم برای رسیدن به مقصد، ضریب همبستگی منفی وجود دارد. ثرندایک و هیگن (۱۹۷۹، ص ۵۱) ضرایب همبستگی بین تعدادی متغیر را به دست داده‌اند که برخی از آنها در جدول ۷ـ۱۸ نقل می‌شود.

روش محاسبهٔ ضریب همبستگی گشتاوری پیرسون

همان‌طور که پیش‌تر گفته شد، متداول‌ترین روش همبستگی، همبستگی گشتاوری پیرسون است. برای نشان دادن مراحل محاسبهٔ این ضریب همبستگی نمرات جدول ۸ـ۱۸ را مورد استفاده قرار می‌دهیم.

دو دسته نمرهٔ X و Y جدول ۸ـ۱۸ را نمرات جبر (X) و هندسهٔ (Y) یک گروه پنج‌نفری از دانش‌آموزان یک کلاس دبیرستان فرض کنید. برای محاسبهٔ ضریب همبستگی بین این دو دسته

جدول ۸ـ۱۸ محاسبهٔ ضریب همبستگی گشتاوری پیرسون با استفاده از نمرات انحراف

نام دانش‌آموز	نمرهٔ خبر X	نمرهٔ هندسه Y	x	y	x^2	y^2	xy
پرویز	۸	۶	۱۷ـ۸ = ۹	۸ـ۶ = ۲	۸۱	۴	۱۸
حسن	۲۰	۱۴	۱۷ـ۲۰ = ـ۳	۸ـ۱۴ = ـ۶	۹	۳۶	۱۸
بیژن	۲۶	۱۰	۱۷ـ۲۶ = ـ۹	۸ـ۱۰ = ـ۲	۸۱	۴	۱۸
علی	۱۷	۸	۱۷ـ۱۷ = ۰	۸ـ۸ = ۰	۰	۰	۰
سهراب	۱۴	۲	۱۷ـ۱۴ = ۳	۸ـ۲ = ۶	۹	۳۶	۱۸
	$\bar{X} = ۱۷$	$\bar{Y} = ۸$			$\Sigma = ۱۸۰$	۸۰	۷۲

نمره، میانگین نمرات هر دو دسته را تعیین می‌کنیم و فاصلهٔ هر نمره را از میانگین (نمرات انحراف) نیز مشخص می‌کنیم، و در ستونهای x و y وارد می‌نماییم. برای تعیین ارقام ستونهای x^2 و y^2، نمرات انحراف را مجذور می‌کنیم. در ضمن با ضرب‌کردن دو ستون x و y در یکدیگر، ستون xy را درست می‌کنیم. (مشاهده می‌کنید که اعداد ستون xy جدول ۸ـ۱۸ همه با هم مساوی هستند. این یک امر تصادفی است که در مورد نمرات مثال ما رخ داده است؛ مشاهدهٔ چنین پدیده‌ای در مورد نمرات دیگر ضروری نیست.)

بعد از کامل کردن جدول ۸ـ۱۸، با استفاده از فرمول زیر به محاسبهٔ ضریب همبستگی گشتاوری پیرسون می‌پردازیم.

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \qquad r_{xy} = \frac{۷۲}{\sqrt{۱۸۰ \times ۸۰}} = \frac{۷۲}{۱۲۰} \qquad r_{xy} = +\,۰/۶۰$$

علاوه بر فرمول بالا، ضریب همبستگی را می‌توان مستقیماً با استفاده از نمرات خام نیز محاسبه کرد. معروف‌ترین فرمول ضریب همبستگی که در آن نمرات خام به کار می‌روند فرمول زیر است.

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

در این فرمول:

ΣX = مجموع نمرات خام گروه X

ΣY = مجموع نمرات خام گروه Y

ΣXY = مجموع حاصل ضربهای X در Y

ΣX^2 = مجموع مجذورات Xها

ΣY^2 = مجموع مجذورات Yها

$(\Sigma X)^2$ = مجذور حاصل جمع Xها

$(\Sigma Y)^2$ = مجذور حاصل جمع Yها

N = تعداد Xها یا Yها

جدول ۸ـ۱۸ محاسبهٔ ضریب همبستگی گشتاوری پیرسون با استفاده از نمرات انحراف

نام دانش‌آموز	نمرهٔ جبر X	نمرهٔ هندسه Y	x	y	x^2	y^2	xy
پرویز	۸	۶	۱۷ـ۸ = ۹	۸ـ۶ = ۲	۸۱	۴	۱۸
حسن	۲۰	۱۴	۱۷ـ۲۰ = ـ۳	۸ـ۱۴ = ـ۶	۹	۳۶	۱۸
بیژن	۲۶	۱۰	۱۷ـ۲۶ = ـ۹	۸ـ۱۰ = ـ۲	۸۱	۴	۱۸
علی	۱۷	۸	۱۷ـ۱۷ = ۰	۸ـ۸ = ۰	۰	۰	۰
سهراب	۱۴	۲	۱۷ـ۱۴ = ۳	۸ـ۲ = ۶	۹	۳۶	۱۸
	$\bar{X} = ۱۷$	$\bar{Y} = ۸$			$\Sigma = ۱۸۰$	۸۰	۷۲

نمره، میانگین نمرات هر دو دسته را تعیین می‌کنیم و فاصلهٔ هر نمره را از میانگین (نمرات انحراف) نیز مشخص می‌کنیم، و در ستونهای x و y وارد می‌نماییم. برای تعیین ارقام ستونهای x^2 و y^2، نمرات انحراف را مجذور می‌کنیم. در ضمن با ضرب‌کردن دو ستون x و y در یکدیگر، ستون xy را درست می‌کنیم. (مشاهده می‌کنید که اعداد ستون xy جدول ۸ـ۱۸ همه با هم مساوی هستند. این یک امر تصادفی است که در مورد نمرات مثال ما رخ داده است؛ مشاهدهٔ چنین پدیده‌ای در مورد نمرات دیگر ضروری نیست.)

بعد از کامل کردن جدول ۸ـ۱۸، با استفاده از فرمول زیر به محاسبهٔ ضریب همبستگی گشتاوری پیرسون می‌پردازیم.

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \qquad r_{xy} = \frac{۷۲}{\sqrt{۱۸۰ \times ۸۰}} = \frac{۷۲}{۱۲۰} \qquad r_{xy} = +\,۰/۶۰$$

علاوه برفرمول بالا، ضریب همبستگی را می‌توان مستقیماً با استفاده از نمرات خام نیز محاسبه کرد. معروف‌ترین فرمول ضریب همبستگی که در آن نمرات خام به کار می‌روند فرمول زیر است.

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

در این فرمول:

ΣX = مجموع نمرات خام گروه X

ΣY = مجموع نمرات خام گروه Y

ΣXY = مجموع حاصل ضربهای X در Y

ΣX^2 = مجموع مجذورات هاX

ΣY^2 = مجموع مجذورات هاY

$(\Sigma X)^2$ = مجذور حاصل جمع هاX

$(\Sigma Y)^2$ = مجذور حاصل جمع هاY

N = تعداد هاX یا هاY

جدول ۹ـ۱۸ محاسبهٔ ضریب همبستگی گشتاوری پیرسون بااستفاده ازنمرات خام

X	Y	XY	X^2	Y^2
۸	۶	۴۸	۶۴	۳۶
۲۰	۱۴	۲۸۰	۴۰۰	۱۹۶
۲۶	۱۰	۲۶۰	۶۷۶	۱۰۰
۱۷	۸	۱۳۶	۲۸۹	۶۴
۱۴	۲	۲۸	۱۹۶	۴
$\Sigma = ۸۵$	۴۰	۷۵۲	۱۶۲۵	۴۰۰

برای استفاده از فرمول بالا، باید جدولی شبیه به جدول ۹ـ۱۸ درست کنیم. نمرات جدول ۹ـ۱۸ همان نمرات جدول ۸ـ۱۸ هستند، منتها در جدول ۹ـ۱۸ به جای ستونهای x، y، x^2، y^2 و xy که از نمرات انحراف حاصل شده‌اند، ستون‌های XY، X^2، و Y^2 را ایجاد کرده‌ایم که ارقام آنها مستقیماً از نمرات خام به دست آمده‌اند. حال با قراردادن ارقام جدول ۹ـ۱۸ در فرمول ضریب همبستگی، ضریب همبستگی نمرات جبر و هندسه را، با استفاده از نمرات خام، به شرح زیر تعیین می‌کنیم.

$$r_{xy} = \frac{۵ \times ۷۵۲ - (۸۵ \times ۴۰)}{\sqrt{[۵ \times ۱۶۲۵ - ۷۲۲۵][۵ \times ۴۰۰ - ۱۶۰۰]}}$$

$$= \frac{۳۷۶۰ - ۳۴۰۰}{\sqrt{۹۰۰ \times ۴۰۰}}$$

$$= \frac{۳۶۰}{\sqrt{۳۶۰۰۰۰}} = \frac{۳۶۰}{۶۰۰}$$

$$r_{xy} = +۰/۶۰$$

روش محاسبهٔ ضریب همبستگی رتبه‌ای اسپیرمن

جایی که تعداد نمرات کم است و نمرات به صورت رتبه‌ای هستند یا می‌توان آنها را به صورت رتبه‌ای در آورد، می‌توان ضریب **همبستگی رتبه‌ای اسپیرمن**[1] را برای آنها حساب کرد. رتبهٔ افراد در یک گروه بر حسب ترتیبی که آنها در یک ویژگی مورد نظر به دنبال هم قرار می‌گیرند نشان داده می‌شود. مثلاً گفته می‌شود که در درس جبر بیژن در بالاترین مرتبه است، و بعد از او حسن قرار دارد، و نفر سوم علی است، و الی آخر. (به مبحث مقیاسهای اندازه گیری، مقیاس ترتیبی، در فصل

1. Spearman rank order correlation

دوم کتاب حاضر، مراجعه کنید.) همچنان که در مورد اعداد ترتیبی در فصل دوم کتاب گفته شد، در رابطه با محاسبهٔ ضریب همبستگی اسپیرمن، هیچ فرضی مبتنی بر فواصل مساوی بین رتبه‌ها وجود ندارد. برای تمایز دادن این ضریب همبستگی با ضریب همبستگی پیرسون، ضریب اسپیرمن را با علامت r_{ho} (رُ) نشان می‌دهند. فرمول محاسبهٔ r_{ho} به قرار زیر است:

$$r_{ho} = ۱ - \frac{۶\Sigma d^{۲}}{N(N^{۲} - ۱)}$$

در این فرمول:

d = تفاوت بین رتبه‌های هر جفت نمرات

$d^{۲}$ = مجذور تفاوت بین رتبه‌های هر جفت نمرات

$\Sigma d^{۲}$ = مجموع مجذورات تفاوتهای بین رتبه‌های هر جفت نمرات

N = تعداد نمرات هر یک از دو گروه

به عنوان مثال، به طرز محاسبهٔ ضریب r_{ho} برای دو دسته نمرهٔ رتبه‌بندی شده که در جدول ۱۰ـ۱۸ داده شده‌اند توجه کنید.

$$r_{ho} = ۱ - \frac{۶ \times ۱۴}{۵ \times ۲۴}$$

$$= ۱ - \frac{۸۴}{۱۲۰}$$

$$= ۱ - ۰/۷$$

$$= + ۰/۳$$

تفسیر ضریب همبستگی

در تفسیر ضریب همبستگی نکات زیر را که از ویژگیهای مهم همبستگی بین متغیرهاست باید در نظر گرفت:

جدول ۱۰ـ۱۸ روش محاسبهٔ ضریب همبستگی اسپیرمن (r_{ho})

X	Y	d	$d^{۲}$
۱	۱	۰	۰
۳	۲	۱	۱
۵	۳	۲	۴
۴	۴	۰	۰
۲	۵	-۳	۹
			$\Sigma = ۱۴$

۱. رابطهٔ بین متغیرهای مورد نظر ممکن است از جمعیتی به جمعیت دیگر فرق کند. برای مثال، در انسانها، بین سالهای ۱۰ و ۱۶، نیروی بدنی و سن ضریب همبستگی بسیار بالایی دارند، ولی این همبستگی بین سالهای ۲۰ و ۲۶ به صفر می‌رسد. سن ذهنی و سن تقویمی، در میان کودکان، دارای ضریب همبستگی مثبت است؛ این رابطه در میانسالی از بین می‌رود، و در سنین بالا معکوس است (ضریب همبستگی منفی وجود دارد).

۲. هرچه افراد مورد نظر در متغیرهایی که ضریب همبستگی بین آنها محاسبه می‌شود نامتجانس‌تر باشند ضریب همبستگی حاصل بزرگ‌تر است. برای مثال، ضریب همبستگی بین هوش و پیشرفت تحصیلی در یک جمعیت بزرگ، مانند دانش‌آموزان دبیرستانها، بیشتر از مقدار این ضریب در دانشجویان دانشگاههاست، زیرا دانشجویان دانشگاه در متغیر هوش و پیشرفت تحصیلی از دانش‌آموزان دبیرستانی متجانس‌تر هستند.

۳. ممکن است وجود همبستگی بین دو متغیر به علت یک رابطهٔ واقعی بین آن دو نباشد، بلکه دلیل آن، ارتباط بین این دو متغیر با یک متغیر ثالث باشد. برای مثال، در روزهای برفی و بارانی زمستان، بین پوشیدن پالتو و استفاده از چتر، ضریب همبستگی مثبت وجود دارد. اما براساس این همبستگی نه می‌توان گفت که پالتو پوشیدن دلیل استفاده از چتر است، نه استفاده از چتر دلیل پالتو پوشیدن. بلکه دلیل هر دو اینها سردی هوا و بارندگی است. بنابراین، صرفاً براساس ضریب همبستگی نمی‌توان بین دو متغیر رابطهٔ علیت برقرار کرد.

۴. ضریب همبستگی را نباید به صورت درصد تعبیر کرد. اما مجذور ضریب همبستگی نشان‌دهندهٔ درصد وجه مشترک بین دو متغیر است. اگر بین دو آزمون پیشرفت تحصیلی برای یک موضوع واحد درسی ضریب همبستگی ۰/۸۰+ وجود داشته باشد، می‌توان گفت که ۶۴ درصد [۶۴ = $(۸)^۲$] مطالب این دو آزمون مشترک‌اند.

۵. هرچه قدر مطلق ضریب همبستگی بین دو متغیر بیشتر باشد، مقدار رابطهٔ بین آنها بیشتر است. بنابراین، ضریب همبستگی ۰/۶۰+ بیشتر از ضریب همبستگی ۰/۵۵+ است و ضریب همبستگی ۰/۷۲- از هر دو آنها بزرگ‌تر است. تنها تفاوت بین ضرایب مثبت و منفی در جهت رابطهٔ بین متغیرهاست. ضریب همبستگی مثبت نشان‌دهندهٔ رابطهٔ مستقیم، اما ضریب همبستگی منفی نشان‌دهندهٔ رابطهٔ معکوس است.

سایر همبستگیها

علاوه بر روشهای مهم محاسبهٔ ضریب همبستگی که در صفحات پیش بحث شدند، روشهای دیگری وجود دارند که در دنبالهٔ این فصل معرفی می‌شوند. دو روشی که قبلاً معرفی کردیم، یعنی روش همبستگی گشتاوری پیرسون و روش همبستگی اسپیرمن، ضریب همبستگی بین جفت

متغیرهایی را به دست می‌دهند که هر دو متغیر از یک جنس باشند، یعنی اگر متغیر X نسبتی یا فاصله‌ای یا رتبه‌ای است متغیر Y هم به همین‌گونه است. اما آنجا که متغیرهای X و Y یکسان نباشند از روشهای همبستگی نوع دیگری استفاده می‌شود.

همبستگی دو رشته‌ای نقطه‌ای

زمانی که از دو متغیر مورد نظر که می‌خواهیم بین آنها ضریب همبستگی به دست آوریم یکی پیوسته و دیگری دو مقولگی یا دو ارزشی[1] باشد از روش **همبستگی دو رشته‌ای نقطه‌ای**[2] استفاده می‌کنیم. منظور از متغیر دو ارزشی آن است که تنها یکی از دو مقدار یک یا صفر را شامل می‌شود، مانند بله ـ نه، زن ـ مرد، و قبول ـ رد. معمولاً به منظور انجام محاسبات به یکی از این دو مقوله نمرهٔ ۱ و به دیگری نمرهٔ ۰ (صفر) داده می‌شود. نمونهٔ روشن این نوع نمره گذاری، نمره‌دادن به سؤالهای آزمونهای عینی، مانند آزمونهای چندگزینه‌ای است. در این آزمونها به هر سؤال درست نمرهٔ ۱ و به هر سؤال غلط نمرهٔ ۰ داده می‌شود.

فرمول محاسبهٔ ضریب همبستگی دورشته‌ای نقطه‌ای که مورد خاصی از ضریب همبستگی گشتاوری است به شرح زیر است:

$$r_{pbis} = \frac{\overline{X}p - \overline{X}t}{\sigma_t}\sqrt{\frac{p}{q}}$$

در این فرمول:

$\overline{X}_p$ = میانگین نمرات متغیر پیوسته (متغیر Y) برای همهٔ آزمون‌شوندگانی که در آزمون دوارزشی (متغیر X) نمرهٔ ۱ گرفته‌اند، یا:

$$\frac{\Sigma XY}{N_p} \quad N_p = \Sigma X$$

$\overline{X}_t$ = میانگین نمرات برای همهٔ آزمون‌شوندگان در متغیر پیوسته

σ_t = انحراف معیار همهٔ نمرات در متغیر پیوسته

p = نسبت همهٔ آزمون شوندگانی که در متغیر دوارزشی نمرهٔ ۱ گرفته‌اند، یا: $\Sigma\frac{X}{N}$

$q = ۱ - p$ نسبت آزمون‌شوندگانی که در متغیر دوارزشی نمرهٔ ۰ گرفته‌اند.

1. dichotomous
2. point biserial correlation

برای آشنایی با چگونگی استفاده از فرمول بالا، داده‌های جدول ۱۱-۱۸ را مورد استفاده قرار می‌دهیم. داده‌های این جدول این‌گونه به دست آمده‌اند که پژوهشگری فرض کرده است تعداد برگهای جریمه‌ای که یک راننده دریافت می‌کند از روی عملکرد او در نخستین آزمایش آیین‌نامهٔ رانندگی که به صورت قبول (۱) یا رد (۰) نمره گذاری می‌شود قابل پیش‌بینی است.

براساس داده‌های موجود در جدول ۱۱ـ۱۸،

$$p = \frac{۱۰}{۱۵} = ۰/۶۷$$

$$\overline{X}_t = \frac{۵۰}{۱۵} = ۳/۳۳$$

$$\overline{X}_P = \frac{۵+۱+۴+۷+۱+۶+۲+۶+۴+۲}{۱۰} = \frac{۳۸}{۱۰} = ۳/۸۰$$

جدول ۱۱ـ۱۸ عملکرد راننده‌ها در آزمون آیین‌نامهٔ رانندگی و برگهای جریمه‌ای که پس از گرفتن گواهینامهٔ رانندگی دریافت کرده‌اند (دیک[1] و هگرتی[2]، ۱۹۷۱)

شمارهٔ راننده	X آزمون آیین‌نامهٔ رانندگی	Y تعداد برگهای جریمهٔ رانندگی
۰۱	۱	۵
۰۲	۰	۳
۰۳	۱	۱
۰۴	۱	۴
۰۵	۱	۷
۰۶	۰	۰
۰۷	۱	۱
۰۸	۱	۶
۰۹	۰	۵
۱۰	۱	۲
۱۱	۱	۶
۱۲	۰	۳
۱۳	۰	۱
۱۴	۱	۴
۱۵	۱	۲
	$\Sigma = ۱۰$	$\Sigma = ۵۰$

1. Dick 2. Hagerty

$$\sigma_t = \sqrt{\frac{(۵-۳/۳۳)^۲ + (۳-۳/۳۳)^۲ + \cdots + (۲-۳/۳۳)^۲}{N}}$$

$$= \sqrt{\frac{۲۳۲}{۱۵} - (۳/۳۳)^۲}$$

$$= \sqrt{۱۵/۴۷ - ۱۱/۰۹} = ۴/۳۸ = ۲/۱$$

$$r_{pbis} = \frac{۳/۸۰ - ۳/۳۳}{۲/۱} \sqrt{\frac{۰/۶۷}{۰/۳۳}}$$

$$= ۰/۲۲\sqrt{۲/۰۳} = (۰/۲۲)(۱/۴۲)$$

$$= ۰/۳۱$$

براساس ضریب به‌دست آمده، یعنی ۰/۳۱، می‌گوییم کسانی که در نخستین آزمایش آیین‌نامهٔ رانندگی خود موفق بوده‌اند از کسانی که در این آزمون موفق نبوده‌اند بیشتر جرائم رانندگی مرتکب شده‌اند. با این حال، ضریب همبستگی آن قدر زیاد نیست که بتوان با اطمینان آن را مورد استفاده قرار داد.

ما در فصل ۱۷ گفتیم که یکی از شاخصهای قدرت تمیز یا تشخیص سؤال آزمون ضریب همبستگی دورشته‌ای نقطه‌ای بین سؤال و آزمون است. برای این منظور نیز مثالی می‌آوریم.

فرض کنید ۸۰ درصد دانشجویان یک درس سنجش و اندازه‌گیری به سؤال شمارهٔ ۱۰ آزمون درسی درست جواب داده‌اند و میانگین نمرات آنها در آزمون ۱۸ از ۲۰ است. همچنین فرض کنید میانگین همهٔ دانشجویان در آزمون این درس ۱۶ و انحراف معیار آزمون ۵ است. ضریب تمیز سؤال شمارهٔ ۱۰ برحسب ضریب همبستگی دورشته‌ای نقطه‌ای بین آن سؤال و کل آزمون به گونهٔ زیر محاسبه می‌شود:

$$r_{pbis} = \frac{\bar{X}_p - \bar{X}_t}{\delta_t} \sqrt{\frac{p}{q}} = \frac{۱۸ - ۱۶}{۵} \sqrt{\frac{۸۰}{۲۰}}$$

$$= \frac{۲}{۵}\sqrt{۴} = \frac{۲}{۵} \times ۲ = \frac{۴}{۵}$$

$$= ۰/۸$$

نکته‌ای را که باید دربارهٔ ضریب همبستگی دورشته‌ای نقطه‌ای یادآور شویم این است که در آزمونهای مختصر (دارای تعداد کمی سؤال) عملکرد پاسخ‌دهندگان به یک سؤال بر کل آزمون تأثیر می‌گذارد. مثلاً در یک آزمون ۸ سؤالی، نمرهٔ هر سؤال $\frac{۱}{۸}$ نمرهٔ کل آزمون را تشکیل می‌دهد و

این خود عاملی برای افزایش همبستگی بین سؤال و آزمون است. برای جبران این افزایش، می‌توان نمرهٔ سؤال را از نمرهٔ کل حذف کرد و به محاسبهٔ ضریب همبستگی بین هر سؤال و مابقی سؤالها پرداخت.

همبستگی دورشته‌ای

زمانی که یکی از دو متغیر مورد نظر پیوسته است، اما متغیر دوم یک متغیر دوارزشی غیرحقیقی یا ساختگی است، یعنی نمرات به عمد به دو مقوله تقسیم شده‌اند، از روش **همبستگی دورشته‌ای**[1] استفاده می‌شود. مثالی برای این مورد است که ما نمرات از ۱۰ تا ۲۰ را قبول (۱) و از ۱۰ به پایین را مردود (۰) به حساب می‌آوریم. در این مورد، اعداد پیوستهٔ از صفر تا ۲۰ را عمداً به یک متغیر دوارزشی تبدیل کرده‌ایم.

فرمول محاسبهٔ ضریب همبستگی دورشته‌ای به شرح زیر است:

$$r_{bis} = \frac{\overline{X}_p - \overline{X}_t}{\sigma_t}\left(\frac{p}{z}\right)$$

در این فرمول:

$Z =$ آن مقدار از منحنی بهنجار دارای مساحت ۱ که با p مطابق است

بقیهٔ علامتها همانهایی هستند که در فرمول ضریب همبستگی دو رشته‌ای نقطه‌ای معرفی شدند.

جدول ۱۲ـ۱۸ نمرات دو آزمون چندگزینه‌ای و تشریحی برای گروهی دانش‌آموز

شمارهٔ دانش‌آموز	X نمرهٔ آزمون چندگزینه‌ای	Y نمرهٔ آزمون تشریحی*
۱	۲۵	۰
۲	۳۰	۱
۳	۲۰	۱
۴	۲۵	۰
۵	۴۰	۱
۶	۳۰	۰
۷	۲۰	۱
۸	۳۵	۰
۹	۲۵	۱
۱۰	۴۵	۱
	$\Sigma = ۲۹۵$	$\Sigma = ۶$

* قبول = ۱ مردود = ۰

1. biserial correlation

برای آشنایی با چگونگی استفاده از فرمول بالا، داده‌های جدول ۱۲ـ۱۸ را مورد استفاده قرار می‌دهیم. داده‌های این جدول نتایج دو آزمون چندگزینه‌ای و تشریحی هستند که با گروه واحدی از دانش‌آموزان اجرا شده‌اند. نمرات متغیر X یعنی آزمون چندگزینه‌ای نمرات پیوسته هستند، اما نمرات آزمون تشریحی به صورت دوارزشی ۱ و ۰ درآمده‌اند.

بر اساس داده‌های موجود در جدول ۱۲ـ۱۸،

$$N_p = \Sigma y = ۶ \qquad N = ۱۰$$

$$\bar{X}_p = \frac{۳۰ + ۲۰ + ۴۰ + ۲۰ + ۲۵ + ۴۵}{N_p} = \frac{۱۸۰}{۶} = ۳۰$$

$$\bar{X}_t = \frac{۲۵ + ۳۰ + ۲۰ + ۲۵ + ۴۰ + ۳۰ + ۲۰ + ۳۵ + ۲۵ + ۴۵}{N}$$

$$= \frac{۲۹۵}{۱۰} = ۲۹/۵$$

$$\sigma_t = \sqrt{\frac{(۲۵ - ۲۹/۵)^۲ + (۳۰ - ۲۹/۵)^۲ + \cdots + (۴۵ - ۲۹/۵)^۲}{N}}$$

$$= \sqrt{\frac{(۴/۵)^۲ + (۰/۵)^۲ + \cdots + (۱۵/۵)^۲}{۱۰}}$$

$$= \sqrt{\frac{۶۶۱/۵}{۱۰}} = \sqrt{۶۶/۱۵} = ۸/۱۲$$

$$p = \frac{۶}{۱۰} = ۰/۶$$

$$Z = ۰/۳۸۶$$

$$r_{bis} = \frac{۳۰ - ۲۹/۵}{۸/۱۲} \times \frac{۰/۶}{۰/۳۸۶}$$

$$= \left[\frac{۰/۵}{۸/۱۲}\right](۱/۵۵)$$

$$= (۰/۰۶)(۱/۵۵) = ۰/۰۹$$

ضریب ۰/۰۹ نشان می‌دهد که بین دو آزمون فوق تقریباً هیچ نوع همبستگی وجود ندارد، زیرا این رقم خیلی به صفر نزدیک است.

ضریب همبستگی دورشته‌ای مستقل از مقادیر p و q است. در ضمن مقدار این ضریب همبستگی گاه از ۱ + بزرگ‌تر و زمانی از ۱ ـ کوچک‌تر است. علاوه بر این، وقتی که مبنای محاسبهٔ ضریب همبستگی داده‌های واحدی هستند، مقادیر حاصل از ضریب همبستگی دورشته‌ای نقطه‌ای همیشه کوچک‌تر از مقادیر حاصل از ضریب همبستگی گشتاوری و دورشته‌ای خواهند بود (مگنوسون[1]، ۱۹۶۷، ترجمهٔ براهنی، ۱۳۵۶).

همبستگی فی

زمانی که هر دو متغیر مورد نظر متغیرهای دوارزشی واقعی هستند از ضریب همبستگی فی (فای)[2] استفاده می‌شود. برای مثال، می‌توان همبستگی بین متغیر جنسیت (مرد ـ زن) را با متغیر موفقیت ـ شکست در یک فعالیت تعیین کرد. مورد دیگری از کاربرد همبستگی فی تعیین همبستگی بین پرسشهای یک آزمون است که به صورت صحیح یا غلط دسته‌بندی شده‌اند.

فرمول محاسبهٔ ضریب همبستگی فی به شرح زیر است.

$$\phi = \frac{p_{ij} - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}} \quad \text{یا} \quad \phi = \frac{p_{ij} - p_i p_j}{\sqrt{(p_i q_i)(p_j\ q_j)}}$$

در این فرمول:

p_{ij} = نسبت آزمون‌شوندگانی که در هر دو متغیر نمرهٔ قبولی گرفته‌اند

p_i = نسبت آزمون‌شوندگانی که در متغیر X نمرهٔ قبولی گرفته‌اند

p_j = نسبت آزمون‌شوندگانی که در متغیر Y نمرهٔ قبولی گرفته‌اند

برای آشنایی با چگونگی استفاده از فرمول بالا از داده‌های جدول ۱۳ ـ ۱۸ استفاده می‌کنیم. داده‌های این جدول حاصل دو متغیر جنسیت و آزمون گواهینامهٔ رانندگی هستند که هر دو به صورت دوارزشی محاسبه شده‌اند. متغیر جنسیت به صورت مرد (۱) و زن (۰) و آزمون گواهینامهٔ رانندگی به صورت قبول (۱) و مردود (۰) نمره گذاری شده‌اند.

جدول ۱۳ ـ ۱۸ نتایج آزمون گواهینامهٔ رانندگی و جنسیت افراد

متغیرها											
جنسیت	۱	۰	۱	۱	۰	۰	۰	۰	۱	۱	
آزمون رانندگی	۰	۰	۰	۱	۰	۰	۱	۰	۱	۱	

1. Magnusen
2. Phi coefficient

بر اساس داده‌های جدول ۱۳ـ۱۸،

$$p_i = \frac{۵}{۱۰} = ۰/۵$$

$$p_j = \frac{۴}{۱۰} = ۰/۴۰$$

$$p_{ij} = \frac{۳}{۱۰} = ۰/۳$$

$$q_i = ۰/۵$$

$$q_j = ۰/۶۰$$

$$\phi = \frac{۰/۳۰ - (۰/۵۰)(۰/۴۰)}{\sqrt{[(۰/۵۰)(۰/۵۰)][(۰/۴۰)(۰/۶۰)]}}$$

$$= \frac{۰/۳۰ - ۰/۲۰}{\sqrt{(۰/۲۵۰)(۰/۲۴۰)}} = \frac{۰/۱۰}{\sqrt{۰/۰۶۰}} = \frac{۰/۱۰}{۰/۲۴} = ۰/۴۲$$

بر اساس داده‌های فرضی جدول ۱۳ـ۱۸ و ضریب فی محاسبه شده بین متغیر جنسیت و آزمون رانندگی اتومبیل ($\phi = ۰/۴۲$) تا حد ضعیفی می‌توان عملکرد داوطلبان دریافت گواهینامهٔ رانندگی را از روی جنس (یا زن و مرد بودن) آنان پیش‌بینی کرد. با این حال، این ضریب همبستگی در حدی بالا نیست که بتوان به آن اطمینان کرد. ضریب همبستگی فی نیز مورد خاصی از ضریب همبستگی گشتاوری پیرسون است.

همبستگی چهارخانه‌ای یا تتراکوریک

اگر با دو دسته عدد یا نمره سروکار داشته باشیم که در اصل هیچ کدام از آنها دوارزشی واقعی نباشد اما هر دوی آنها را اجباراً و از روی ضرورت به صورت دوارزشی درآورده باشیم، در آن صورت از روش **همبستگی تتراکوریک یا چهارخانه‌ای**[1] استفاده می‌کنیم. فرمول ضریب همبستگی تتراکوریک همبستگی بین دو متغیر مورد نظر را به فرض اینکه پیوسته هستند به دست می‌دهد.

فرض کنید یک آزمون استعداد تحصیلی را با گروهی دانش‌آموز اجرا کرده‌ایم و نمرات حاصل از ۰ تا ۱۰۰ را شامل بوده‌اند و همهٔ دانش‌آموزان پایین ۵۰ را مردود (۰) و همهٔ دانش‌آموزان بالای ۵۰ را قبول (۱) نمره‌گذاری کرده‌ایم. همچنین این دانش‌آموزان را بر اساس نمرات کلاسی به دو دستهٔ پذیرفته شده (از ۱۵ تا ۲۰) و پذیرفته نشده (از ۱۵ به پایین) تقسیم کرده‌ایم. چنان‌که ملاحظه می‌کنید هر دو متغیر بالا، یعنی استعداد تحصیلی و نمرهٔ کلاسی متغیرهای پیوسته‌ای هستند که عمداً به صورت دوارزشی درآمده‌اند. نتایج در جدول ۱۴ـ۱۸ نشان داده شده‌اند.

1. Tetrachoric correlation

جدول ۱۴ـ۱۸ ارائهٔ دو متغیر به صورت چهارگانه (دیک و هگرتی، ۱۹۷۱، ص ۳۴۳)

نمرات آزمون استعداد	نمرات کلاسی: پذیرفته‌نشده	پذیرفته شده	کل
۰ـ۵۰	۲۰ (a)	۳۰ (b)	۵۰
۵۱ـ۱۰۰	۱۰ (c)	۴۰ (d)	۵۰
کل	۳۰	۷۰	

N = ۱۰۰

با استفاده ازداده‌های جدول ۱۴ـ۱۸ همبستگی چهارخانه‌ای را به‌نحو زیر محاسبه می‌کنیم:

$$r_{tet} = \frac{(a)(d)}{(b)(c)} = \frac{(۲۰)(۴۰)}{(۳۰)(۱۰)} = \frac{۸۰۰}{۳۰۰} = ۲/۶۷$$

حال مقدار حاصل از فرمول $\frac{ad}{bc}$ فوق، یعنی ۲/۶۷، را در جدول مربوط به آن که در غالب کتابهای آمار یافت می‌شود وارد می‌کنیم و به رقم ۰/۳۶ می‌رسیم. علت نیاز به مراجعه به جدول مقادیر $\frac{ad}{bc}$ این است که رقم حاصل از کسر $\frac{ad}{bc}$ مستقیماً ضریب همبستگی تتراکوریک را به ما نمی‌دهد. کسر $\frac{ad}{bc}$ یک برآورد از ضریب همبستگی تتراکوریک است. روش مستقیم محاسبهٔ ضریب همبستگی تتراکوریک بسیار پیچیده و وقت‌گیر است، لذا به جای آن از فرمول $\frac{ad}{bc}$ استفاده می‌شود، اما رقم حاصل را باید در جدول مربوط به مقادیر $\frac{ad}{bc}$ قرار داد تا ضریب همبستگی بین متغیرهای مورد نظر به دست آید.

فرمول $\frac{ad}{bc}$ باید همواره رقمی بزرگ‌تر از ۱ را به دست بدهد. بنابراین، فرق نمی‌کند که جدول چهارخانه را به کدام صورت از دو صورت زیر ترتیب دهیم.

	بله	نه	
بله	a	b	a+b
نه	c	d	d+c
کل	b+d	b+c	a+b+c+d

یا

	بله	نه	
بله	b	a	a+b
نه	d	c	d+c
کل	a+c	b+d	a+b+c+d

در فرمول $\frac{ad}{bc}$، همواره رقم بزرگ‌تر در صورت کسر و رقم کوچک‌تر در مخرج گذاشته می‌شود. بنابراین،

$$\frac{ad}{bc} = \frac{(۳۰)(۲۰)}{(۴۰)(۲۰)} = \frac{۶۰۰}{۸۰۰}$$

را باید معکوس کرد.

$$\frac{۸۰۰}{۶۰۰} = ۱/۳۳$$

با مراجعه به جدول ارقام $\frac{ad}{bc}$، از روی رقم ۱/۳۳ برآورد ضریب همبستگی تتراکوریک برابر است با $r_{tet} = ۰/۱۱$.

با استفاده از روشهای همبستگی دورشته‌ای و دورشته‌ای نقطه‌ای، می‌توانیم برآورد کنیم که سؤال بخصوصی تا چه اندازه همان متغیری را می‌سنجد که سؤالهای دیگر می‌سنجند، یعنی همخوانی یک سؤال را با سؤالهای دیگر آزمون مشخص می‌کنیم. برای این کار بین توزیع نمره‌های یک سؤال که به صورت دوارزشی ۱ و صفر نشان داده شده و توزیع نمرات سؤالهای دیگر آزمون، که به صورت پیوسته است، همبستگی حساب می‌کنیم. حال اگر بخواهیم تعیین کنیم که سؤالهای مختلف تا چه اندازه متغیر واحدی را نشان می‌دهند باید جفت‌جفت روابط بین سؤالهای یک آزمون را محاسبه کنیم. برای محاسبهٔ رابطهٔ همبستگی بین دو سؤال از یک آزمون می‌توانیم از روش همبستگی تتراکوریک استفاده نماییم.

از روش همبستگی تتراکوریک می‌توان در مواردی نیز که توزیعهای مورد نظر بیشتر از دو طبقه دارند، مانند همبستگی بین توزیعهایی که از دو آزمون کامل به دست می‌آیند استفاده کرد (فرگوسون[1]، ۱۹۶۶). فایدهٔ چنین کاری این است که یک برآورد سریع از همبستگی گشتاوری به‌دست می‌آید. برای این منظور لازم است که توزیعهای مورد نظر را به دوارزشی تبدیل کنیم، اما باید سعی کنیم که این تقسیم به دو بخش در حوالی نقطهٔ میانه صورت بگیرد.

از میان روشهای همبستگی معرفی شده در این فصل، روش همبستگی گشتاوری پیرسون، روش همبستگی اسپیرمن، همبستگی دو رشته‌ای نقطه‌ای، و همبستگی فی، از یک خانواده هستند و تنها تفاوت آنها در نوع مقیاس اندازه‌گیری به کار رفته در آنهاست. چنان‌که دیدیم، از ضریب همبستگی گشتاوری پیرسون زمانی استفاده می‌شود که متغیرهای موردنظر با مقیاس فاصله‌ای یا نسبتی اندازه‌گیری شده باشند. ضریب اسپیرمن در جایی مورد استفاده قرار می‌گیرد که متغیرها با مقیاس رتبه‌ای اندازه گرفته شده باشند. و از همبستگی دو رشته‌ای نقطه‌ای و ضریب فی در موقعیتهایی استفاده می‌شود که یکی از دو متغیر موردنظر یا هر دو آنها جنبهٔ دو مقولگی داشته یعنی مقیاس اندازه‌گیری آنها شبیه به مقیاس اسمی باشد (مانند زن ـ مرد).

دو روش همبستگی باقیمانده، یعنی همبستگی دو رشته‌ای و همبستگی چهارخانه‌ای یا تتراکوریک از خانوادهٔ ضرایب همبستگی پیرسون نیستند. چنان‌که در بحث از این دو روش دیدیم، هرگاه هر دو متغیر مورد نظر یا یکی از آن دو به طور ساختگی به صورت دومقولگی در آید این روشها را به کار می‌بندیم. از آنجاکه ضریب همبستگی دو رشته‌ای و ضریب همبستگی تتراکوریک از خانوادهٔ ضریب گشتاوری پیرسون نیستند، روش محاسبهٔ آنها با روش محاسبهٔ ضریبهای همبستگی خانوادهٔ پیرسون متفاوت است.

حال که با انواع ضریبهای همبستگی آشنا شدیم، می‌توانیم ویژگیهای مختلف آنها را با هم مقایسه کنیم. این مقایسه در جدول ۱۵-۱۸ انجام گرفته است.

---

1. Ferguson

جدول ۱۵-۱۸ مقایسهٔ انواع همبستگیهای بین متغیرها

نوع همبستگی	متغیر اول X	متغیر دوم Y	علامت	فرمول	موارد استفاده و ویژگیها
گشتاوری پیرسون	پیوسته	پیوسته	r	$r=\frac{N\Sigma xy-(\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$	بهترین و دقیق‌ترین روش محاسبهٔ همبستگی بین دو متغیر.
رتبه‌ای اسپیرمن	رتبه‌ای	رتبه‌ای	r_{ho}	$r_{ho}=1-\frac{6\Sigma d^2}{N(N^2-1)}$	آنجاکه تعداد نمرات کم یا به صورت رتبه‌ای هستند.
دو رشته‌ای نقطه‌ای	پیوسته	دوارزشی واقعی	r_{pbis}	$r_{pbis}=\frac{\overline{X}p-\overline{X}t}{\sigma_t}\sqrt{\frac{p}{q}}$	وقتی که یک متغیر پیوسته و متغیر دیگر دوارزشی واقعی است مقدار این همبستگی غالباً کمتر از ۱ است.
دورشته‌ای	پیوسته	دوارزشی ساختگی	r_{bis}	$r_{bis}=\frac{\overline{X}p-\overline{X}t}{\sigma_t}\left[\frac{p}{z}\right]$	وقتی که یک متغیر پیوسته و متغیر دیگر به عمد یا به‌طور ساختگی به دوارزشی تبدیل شده است. مقدار این همبستگی گاه از ۱+ بیشتر و زمانی از ۱- کمتر است.
فی (فای)	دوارزشی واقعی	دوارزشی واقعی	ϕ	$\phi=\frac{p_{ij}-p_ip_j}{\sqrt{(p_iq_i)(p_jq_j)}}$	آنجاکه هر دو متغیر دوارزشی واقعی هستند به کار می‌رود. بهترین مورد استفادهٔ آن محاسبهٔ همبستگی بین سؤالهای یک آزمون است.
چهارخانه‌ای یا تترا کوریک	دوارزشی ساختگی	دوارزشی ساختگی	r_{tet}	$r_{tet}=\frac{(a)(d)}{(b)(c)}$	وقتی که دو دسته نمره داریم که هیچ‌کدام از آنها دوارزشی واقعی نیستند اما آنها را اجباراً به صورت دوارزشی درمی‌آوریم به کار می‌رود. یک برآورد سریع از همبستگی گشتاوری به دست می‌دهد. بهتر است دو متغیر را در حوالی نقطهٔ میانه به دو بخش تقسیم کنیم.

خلاصهٔ فصل

ما در این فصل مفاهیم و روشهای آماری موردنیاز معلمان برای تفسیر نمره‌های کلاسی‌شان را معرفی کردیم. به‌طور اجمال در این فصل نکات زیر مورد بحث قرار گرفتند.

۱. متغیر، فراوانی، و توزیع فراوانی

۲. نمودار:
الف ـ هیستوگرام
ب ـ چندضلعی
پ ـ تراکمی
ت ـ ستونی

۳. اندازه‌های گرایش مرکزی:
الف ـ نما
ب ـ میانگین
پ ـ میانه

۴. اندازه‌های تغییرپذیری:
الف ـ دامنهٔ تغییر
ب ـ واریانس
پ ـ انحراف معیار

۵. همبستگی
الف ـ همبستگی گشتاوری پیرسون
ب ـ همبستگی رتبه‌ای اسپیرمن
پ ـ همبستگی دورشته‌ای نقطه‌ای
ت ـ همبستگی دورشته‌ای
ث ـ همبستگی فی
ج ـ همبستگی چهارخانه‌ای یا تتراکوریک

علاوه بر روش محاسبهٔ ضرایب همبستگی بالا، ویژگیهای مهم و موارد استفادهٔ هر یک از روشهای آماری مورد بحث را در این فصل نیز توضیح داده‌ایم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. در توزیعهای طبیعی مقادیر میانگین، میانه و، نما برابرند.
الف ـ صحیح
ب ـ غلط

۲. معلمی برای یک آزمون اجرا شدهٔ خود با دانش‌آموزان یک کلاس دبیرستانی میانگین ۱۵ و انحراف معیار ۶- را به دست آورده است. در رابطه با نتایج حاصل چه نظری می‌توان داد؟
الف ـ نمرات این دانش‌آموزان بسیار از هم فاصله دارند.
ب ـ سطح نمرات کل کلاس متوسط است.
پ ـ تعداد معدودی از این دانش‌آموزان به حد ملاک تسلط رسیده‌اند.
ت ـ معلم در محاسبات خود دچار اشتباه شده است.
ث ـ نمرات دانش‌آموزان از همدیگر فاصلهٔ زیادی دارند.

۳. واریانس توزیع یک دسته نمره معرف کدام‌یک از شاخصهای زیر است؟
الف ـ پراکندگی
ب ـ گرایش مرکزی
پ ـ رابطه
ت ـ موقعیت

۴. برای محاسبهٔ همبستگی بین هوش و پیشرفت تحصیلی چه چیزهایی لازم هستند؟
الف ـ نمرات پیشرفت گروهی از افراد و نمرات هوش گروهی دیگر از افراد
ب ـ نمرات پیشرفت گروهی از افراد و نمرات هوش همان گروه از افراد
پ ـ دو نمرهٔ پیشرفت تحصیلی و دو نمرهٔ هوش برای دو گروه از افراد
ت ـ دو نمرهٔ پیشرفت تحصیلی و دو نمرهٔ هوش گروه واحدی از افراد

۵. کدام‌یک از ارقام زیر معرف ضریب همبستگی بیشتری است؟
الف ـ ۰/۵۵ ـ
ب ـ ۰/۶۱ ـ
پ ـ ۰/۴۹+
ت ـ ۰/۲۳+

۶. کدام‌یک از موارد زیر جزو خواص میانگین است؟
الف ـ اگر عددی بر همهٔ اعداد بیفزاییم همین عدد بر میانگین آنها نیز افزوده می‌شود.
ب ـ میانگین تحت تأثیر اعداد خیلی کوچک یا خیلی بزرگ قرار دارد.
پ ـ در توزیع بهنجار میانگین با میانه برابر است.
ت ـ الف و ب
ث ـ الف و ب و پ

۷. به فهرستی از تعدادی نمره که به ترتیب بزرگی مرتب شده باشند چه می‌گویند؟

الف ـ توزیع

ب ـ نمودار

پ ـ واریانس

ت ـ فراوانی

۸. کدام‌یک از شاخصهای آماری زیر اطلاعاتی کلی در بارهٔ توزیع نمرات یک آزمون به ما می‌دهد؟

الف ـ انحراف معیار

ب ـ ضریب تمیز سؤالها

پ ـ ضریب دشواری سؤالها

ت ـ ضریب همبستگی بین سؤالها

۹. کدام روش یا روشهای همبستگی زیر جزو خانوادهٔ روش همبستگی پیرسون است؟

الف ـ اسپیرمن

ب ـ دو رشته‌ای نقطه‌ای

پ ـ همبستگی فی

ت ـ الف و ب

ث ـ الف و ب و پ

۱۰. برای محاسبهٔ ضریب همبستگی بین نمرات یک سؤال چندگزینه‌ای و نمرات کل آزمون کدام روش تعیین همبستگی بهترین است؟

الف ـ رتبه‌ای اسپیرمن

ب ـ دورشته‌ای نقطه‌ای

پ ـ دورشته‌ای

ت ـ گشتاوری پیرسون

ث ـ فی

۱۱. متغیر، فراوانی، و توزیع فراوانی را تعریف کنید.

۱۲. ویژگیهای شاخص نمودارهای هیستوگرام، چندضلعی، تراکمی، و ستونی را با هم مقایسه کنید.

۱۳. نما، میانگین، و میانه را با هم مقایسه کنید و کاربردهای مختلف آنها را توضیح دهید.

۱۴. موارد استفادهٔ واریانس و انحراف معیار را توضیح دهید.

۱۵. ضریب همبستگی را تعریف کنید.

۱۶. انواع ضریب همبستگی و موارد استفادهٔ آنها را توضیح دهید.

۱۷. در یک توزیع فراوانی که بزرگ‌ترین عدد ۸۱ و کوچک‌ترین آن ۲۳ است، موارد زیر را حساب کنید:

الف ـ دامنهٔ تغییر نمرات

ب ـ فاصلهٔ طبقات

پ ـ پایین‌ترین طبقه و حدود واقعی آن

۱۸. میانگین، میانه، واریانس، و انحراف معیار اعداد زیر را حساب کنید:

۵، ۶، ۸، ۱۰، ۱۰، ۱۱، ۱۲، ۱۸

۱۹. برای دو دسته نمرهٔ زیر محاسبات خواسته شده را انجام دهید.

X	Y
۲	۴
۳	۲
۵	۸
۶	۶

الف ـ ضریب همبستگی پیرسون را حساب کنید.

ب ـ با تبدیل نمرات به رتبه ضریب همبستگی رتبه‌ای اسپیرمن را حساب کنید.

پ ـ با تبدیل نمرات y به دوارزشی، هم ضریب همبستگی دورشته‌ای نقطه‌ای و هم ضریب همبستگی دورشته‌ای را حساب کنید.

ت ـ با تبدیل هر دو متغیر x و y به دوبخشی ضرایب همبستگی فی و چهارخانه‌ای را حساب کنید.

۱۹

هنجارها و نیمرخها

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. هنجار را تعریف کند و فایدهٔ آن را بگوید.

۲. انواع هنجارها را نام ببرد و موارد استفادهٔ آنها را بگوید.

۳. هنجارهای زیر را تعریف کند و نحوهٔ به دست آوردن آنها را توضیح دهد.

الف ـ هنجار سنی

ب ـ هنجار کلاسی

پ ـ هنجار سنی نمایی

ت ـ هنجار سن ذهنی

ث ـ هوشبهر

ج ـ هنجار درصدی

خ ـ نمرهٔ Z

د ـ نمرهٔ T

ذ ـ نمرات ۹ بخشی

ر ـ هوشبهر انحرافی

س ـ نمرهٔ سنی معیار

ز ـ نمرهٔ معادل منحنی بهنجار یا نمرهٔ NCE

۴. نظامهای مختلف نمرات را با هم مقایسه کند.

۵. توزیع بهنجار و منحنی بهنجار را تعریف کند و ویژگیهای آنها را توضیح دهد.

۶. نیمرخ را تعریف کند.

۷. انواع نیمرخ را توضیح دهد.

۸ نیمرخ عملکرد یک دانش‌آموز را در یک مجموعه آزمون رسم کند و برای آن تفسیر بنویسد.

پس از تصحیح برگه‌های آزمون، برای هریک از دانش‌آموزان نمره‌ای به نام **نمرهٔ خام**[1] به دست می‌آید. نمرات خام به‌خودی‌خود قابل تفسیر و استفاده نیستند و باید در رابطه با آنها اقداماتی انجام داد تا قابل استفاده شوند. از اقداماتی که در این باره می‌توان انجام داد تعیین وضع تمرکز یا حالت پراکندگی نمرات یعنی محاسبهٔ شاخصهای گرایش مرکزی و شاخصهای پراکندگی است. همچنین می‌توان نمره‌های یک آزمون را به صورت توزیع فراوانی و به شکل نمودارها و منحنیهای مختلف نشان داد. ما در فصل پیش این روشها را توضیح دادیم. موضوع مورد بحث در این فصل **هنجارها**[2] و **نیمرخها**[3] هستند. هنجارها و نیمرخها روشهایی هستند که به ما کمک می‌کنند تا سطح موفقیت دانش‌آموزان را در درسهای مختلف با هم مقایسه کنیم و میزان توفیق آنان را نسبت به یکدیگر بسنجیم.

پیش از معرفی مفهوم هنجار، لازم است یادآوری کنیم که مقایسهٔ نمرات دانش‌آموزان با یکدیگر مخصوص آزمونهای وابسته به هنجار است، نه آزمونهای وابسته به ملاک. بنابراین، مطالبی که در این فصل در رابطه با هنجار و هنجاریابی گفته می‌شوند ویژهٔ آزمونهای وابسته به هنجار هستند. در تفسیر نمرات آزمونهای وابسته به ملاک، چنان که قبلاً در فصل چهارم توضیح دادیم، عملکرد هریک از دانش‌آموزان در آزمون با یک ملاک از پیش تعیین شده که همان ملاک عملکرد ملحوظ در هدفهای آموزشی است مقایسه می‌شود. برای نمونه، می‌توانیم عملکرد دانش‌آموزی را برحسب سرعت انجام عملکرد (مثلاً، ماشین کردن ۵۰ کلمه در دقیقه)، میزان دقت عملکرد (مثلاً، اندازه‌گیری طول یک میز با اشتباه حداکثر نیم‌سانتی‌متر)، درصد ماده‌های یک تکلیف یادگیری (مثلاً، معنی کردن ۹۰ درصد کلمات جدید) و مانند اینها بسنجیم.

تعریف و فایدهٔ هنجار

معنی لغوی هنجار (خلاصه‌شدهٔ **داده‌های هنجاری**[4]) حد متوسط است و در ابزارسازی نمرهٔ متوسط (میانگین یا میانه) گروه معینی از افراد معنی می‌دهد. این گروه معین **گروه هنجار**[5] نام دارد. گروه هنجار گروه مرجعی است که آزمون یا پرسشنامه بر روی آن اجرا می‌شود و عملکرد متوسط افراد آن گروه به صورت معیار برای مقایسهٔ عملکرد آزمون‌شوندگانی که در آینده به آزمون یا پرسشنامه جواب خواهند داد مورد استفاده قرار می‌گیرد. امتیاز بزرگ ابزارهای اندازه‌گیری میزان‌شده، مانند آزمونها و پرسشنامه‌های میزان شده، وجود همین هنجارها یا داده‌های هنجاری آنهاست که تعبیر و تفسیر نمرات حاصل از اجرای آنها با افراد و گروههای مختلف را امکان‌پذیر می‌سازد.

با وجود هنجارهای مناسب، می‌توان نمرات آزمون‌شوندگان به یک آزمون یا پاسخ‌دهندگان به یک پرسشنامه را تفسیر کرد و تعیین نمود که عملکرد آنان در چه وضعی است. برای مثال،

---

1. raw score
2. norms
3. profiles
4. normative data
5. norm group

می‌توان مشخص کرد که عملکرد یک کودک در یک آزمون توانایی (مثلاً آزمون هوش) نسبت به عملکرد کودکان دیگر هم‌سن او بالاتر یا پایین‌تر است. همچنین می‌توان تعیین کرد که در یک پرسشنامهٔ روانی (مثلاً پرسشنامهٔ اضطراب) فردی که مشکوک به داشتن نوعی اختلال هیجانی (اضطراب) است نسبت به کسانی که به عنوان افراد دارای آن اختلال شناخته شده‌اند چه وضعی دارد.

انواع هنجارها

هنجارهای ابزارهای اندازه‌گیری متنوع‌اند، اما معروف‌ترین آنها عبارت‌اند از: هنجار سنی، هنجار کلاسی، و هنجار صدکی. در هنجار سنی و هنجار کلاسی عملکرد فرد با عملکرد گروه مرجعی که فرد منطقاً عضوی از آن به حساب می‌آید مقایسه می‌شود، اما در هنجار صدکی یا هنجار درصدی موقعیت نسبی فرد نسبت به گروهی که او خود عضوی از آن است نشان داده می‌شود. در زیر به توضیح بیشتر این هنجارها توجه کنید.

هنجار سنی

هنجار سنی[1] یا معادل سنی[2] بر میانگین یا میانهٔ نمراتی مبتنی است که کودکان و نوجوانان در سنهای مختلف کسب می‌کنند و به صورت معادلهای سنی نشان داده می‌شود. برای مثال، اگر دانش‌آموزانی که ۱۰ سال و دو ماه سن دارند در یک آزمون هوش یا پیشرفت تحصیلی نمرهٔ X بگیرند، به این نمره معادل سنی ۱۰/۲ تعلق می‌گیرد. یا اگر قد پسربچه‌های ۹ ساله را اندازه‌گیری کنیم و میانگین یا میانهٔ قد آنها را به دست آوریم، رقم حاصل هنجار یا معادل سنی قد پسربچه‌های ۹ ساله خواهد بود. بنابراین، هنجار سنی مربوط به سنهای مختلف در یک ویژگی خاص را، از راه دادن آزمون مربوط به آن ویژگی به افراد نمونهٔ سنهای مختلف و محاسبهٔ میانگین یا میانهٔ نمرات آنان، تعیین می‌کنند.

بسیاری از آزمونهای میزان‌شدهٔ پیشرفت تحصیلی دارای هنجارهای سنی هستند. در این آزمونها، معادل‌های سنی با عنوان موضوعهای درسی نام‌گذاری می‌شوند. برای مثال، سن خواندن، سن حساب، سن املا، و جز اینها. تمام هنجارهای سنی آزمونهای مختلف به نحو یکسان تفسیر می‌شوند. مثلاً دانش‌آموزی که دارای سن خواندن ۱۲/۶ و سن حساب ۱۳/۰ است، نمره‌اش در آزمون خواندن برابر است با نمرهٔ کودک متوسط دوازده سال و نیم (دوازده سال و شش ماه) و نمره‌اش در آزمون حساب معادل نمرهٔ کودکی است که در آغاز ۱۳ سالگی قرار دارد.

هنجار کلاسی

آنچه به هنجار کلاسی[3] شهرت یافته، شبیه به هنجار سنی است. تنها تفاوت عمدهٔ بین این دو نوع هنجار این است که در هنجار سنی گروههای مرجع گروههای سنی هستند، اما در هنجار کلاسی از

1. age norm 2. age equivalent 3. grade norm

جدول ۱ـ۱۹ نمایش چگونگی ایجاد نمرات معادل کلاسی (گرانلاند و لین، ۱۹۹۰)

زمان اجرای آزمون		میانهٔ نمرات خام*	معادل کلاسی
کلاس یا پایه	ماه		
۴	مهر (۰/۱)	۲۸	۴/۱
۴	اردیبهشت (۰/۸)	۳۵	۴/۸
۵	مهر (۰/۱)	۳۷	۵/۱
۵	اردیبهشت (۰/۸)	۴۱	۵/۸

* پنجاه درصد دانش‌آموزان در گروه هنجار در بالای میانهٔ نمرهٔ خام و ۵۰ درصد در پایین میانه واقع‌اند.

گروههای کلاسی به عنوان گروههای مرجع استفاده می‌کنند؛ یعنی ابزار اندازه‌گیری را با گروههای نمونهٔ پایه‌ها یا کلاسهای مختلف اجرا می‌کنند و میانگین یا میانهٔ نمرات کلاسهای مختلف را به صورت معادلهای کلاسی یا هنجارهای کلاسی به کار می‌گیرند. برای مثال، اگر دانش‌آموزان گروه نمونه‌ای که کلاس یا پایهٔ پنجم را شروع کرده‌اند در یک آزمون پیشرفت تحصیلی نمرهٔ خام ۱۹ بگیرند، به این نمرهٔ خام معادل کلاسی ۵/۰ داده می‌شود.

معادلهای کلاسی را به صورت دو رقمی نشان می‌دهند. رقم اول نشان‌دهندهٔ سال و رقم دوم نشان‌دهندهٔ ماه است که از صفر تا ۹ را شامل می‌شود. علت اینکه رقم دوم معادلهای کلاسی فقط تا ۹ را شامل می‌شود این است که سال تحصیلی از ۹ ماه تشکیل می‌یابد. بنابراین، هنجار یا معادل کلاسی ۸/۳ نشان‌دهندهٔ پیشرفت در سطح ماه سوم از کلاس هشتم و معادل کلاسی ۹/۰ معرف پیشرفت در آغاز کلاس نهم است.

برای آگاه شدن از نحوهٔ تبدیل نمرات خام به معادل‌های کلاسی نگاه کنید به جدول ۱-۱۹. در این جدول نحوهٔ به‌وجود آمدن نمرات معادل کلاسی نشان داده شده است. چنان‌که می‌بینید، در اولین ستون سمت راست این جدول کلاس یا پایهٔ تحصیلی قرار دارد. در ستون دوم از سمت راست ماه سال تحصیلی داده شده است. در ستون سوم میانهٔ نمرات خام آمده‌اند. بالاخره در آخرین ستون از سمت راست یا اولین ستون از سمت چپ معادلهای کلاسی ذکر شده‌اند.

هنجار سنی‌نمایی

معمولاً دانش‌آموزان یک کلاس معین که بر اساس نمرات آنان هنجارهای کلاسی تعیین می‌شود دامنهٔ سنی بزرگی را تشکیل می‌دهند. لذا، نمرات بعضی از دانش‌آموزان که از لحاظ سنی خیلی بزرگ‌تر یا خیلی کوچک‌تر از دانش‌آموزان متوسط هستند نیز در گروه هنجاریابی شرکت داده می‌شوند. برای به دست آوردن شاخص درست‌تری از نمرهٔ متوسط دانش‌آموزان در یک کلاس معین، نمرات دانش‌آموزان خیلی بزرگ یا خیلی کوچک را کنار می‌گذارند و نمرهٔ هنجار را تنها

برای دانش‌آموزان سن مورد نظر برای آن کلاس محاسبه می‌کنند. این نوع هنجار کلاسی محدودشده را **هنجار سنی‌نمایی**[1] می‌نامند.

هنجار سن ذهنی

هنجار سن ذهنی یا **سن عقلی**[2] که اولین‌بار به وسیلهٔ آلفرد بینه معرفی شد نوعی هنجار سنی است که در آزمونهای هوش به کار می‌رود. روش محاسبهٔ سن ذهنی یک دانش‌آموز شبیه به محاسبهٔ معادلهای سنی است. بنابراین، اگر نمرهٔ خام یک دانش‌آموز در یک آزمون هوش برابر با میانهٔ نمرات خام کودکان ۹ ساله باشد، سن ذهنی او ۹ خواهد بود.

هوشبهر

هوشبهر[3] یا بهرهوشی (IQ) یک خارج قسمت است که از فرمول زیر به دست می‌آید:

$$\text{هوشبهر} = IQ = \frac{\text{سن ذهنی}}{\text{سن تقویمی}} \times ۱۰۰$$

بنابراین فرمول، اگر سن ذهنی یک دانش‌آموز را بر سن تقویمی او تقسیم کنیم و حاصل را در ۱۰۰ ضرب نماییم رقم حاصل معرف هوشبهر او خواهد بود.

هنجار درصدی

چنان‌که دیدیم، در هنجارهای سنی و کلاسی، از راه تعیین یک گروه سنی یا کلاسی که نمرهٔ فرد معادل نمرات متوسط آن است، به نمرهٔ او معنی داده می‌شود. در **هنجار درصدی (هنجار صدکی)**[4]، فرد با گروه سنی یا کلاسی خودش یعنی گروهی که منطقاً می‌توان او را عضوی از آن دانست مقایسه می‌شود. در هنجارهای درصدی از **رتبه‌های درصدی**[5] یا **صدکها**[6] استفاده می‌شود.

رتبهٔ درصدی یا **نمرهٔ درصدی**[7] وضعیت نسبی فرد در گروه را بر حسب کسانی که نمرهٔ پایین‌تر از او گرفته‌اند مشخص می‌کند. هر نمرهٔ خام دارای یک رتبهٔ درصدی است. مثلاً نمرهٔ خام ۱۵ یک دانش‌آموز دارای رتبهٔ درصدی ۶۵ خواهد بود، به شرط اینکه ۶۵ درصد دانش‌آموزان گروه نمرهٔ خام کمتر از ۱۵ گرفته باشند. به بیان دیگر، نمرهٔ این دانش‌آموز (۱۵) از نمرهٔ ۶۵ درصد گروه هنجار بیشتر است. بنابراین، در هنجار درصدی نیز، مانند هنجارهای دیگر، هر نمرهٔ خام یک معادل هنجاری دارد که رتبهٔ درصدی آن نمره است.

رتبهٔ درصدی و صدک دو اصطلاح مکمل یکدیگرند. رتبه‌های درصدی عملکرد نسبی

---

1. modal age norm
2. mental age norm
3. intellignce quotient (IQ)
4. percentile norm
5. percentile ranks
6. percentiles
7. percentile score

آزمون‌شوندگان را نشان می‌دهند، در حالی که صدکها به نقاطی بر روی توزیع نمرات اشاره می‌کنند. اگر نمرهٔ یک دانش‌آموز دارای رتبهٔ درصدی X باشد، آن نمره محلی را بر روی توزیع نمرات نشان می‌دهد که صدک X به حساب می‌آید. در مثال بالا، گفته شد که نمرهٔ خام دانش‌آموز ۱۵ و رتبهٔ درصدی او ۶۵ است. بنا به توضیحی که داده شد، نمرهٔ ۱۵ او صدک شصت‌وپنجم به حساب می‌آید. به عبارت دیگر، می‌گوییم این دانش‌آموز دارای رتبهٔ درصدی (رتبهٔ صدکی) ۶۵ است، یا نمرهٔ او در صدک شصت‌وپنجم قرار دارد.

می‌توان صدک را نوعی رتبه در یک گروه ۱۰۰ نفری تلقی کرد، با این تفاوت که در رتبه‌بندی معمولاً شمارش از بالا شروع می‌شود و به بهترین فرد گروه رتبهٔ ۱ تعلق می‌گیرد، اما در مورد صدکها شمارش از پایین شروع می‌شود و هر اندازه صدک کوچک‌تر باشد به همان نسبت جایگاه فردی که صدک مُعرف نمرهٔ خام اوست در گروه پایین‌تر است.

صدک را معمولاً با علامت P و رتبهٔ درصدی یا رتبهٔ صدکی را با PR نشان می‌دهند. صدک ۵۰ یا $P_{۵۰}$ همان میانه است که عملکرد متوسط دانش‌آموزان را نشان می‌دهد. صدکهای کوچک‌تر از ۵۰ نشان‌دهندهٔ عملکرد پایین‌تر از متوسط و صدکهای بزرگ‌تر از ۵۰ مُعرف عملکرد بالاتر از متوسط هستند. صدک ۲۵ که به نقطهٔ چارکی اول و صدک ۷۵ که به نقطهٔ چارکی سوم معروف‌اند پایین‌ترین و بالاترین ربع توزیع را مشخص می‌کنند. این دو صدک نیز مرزهای مشخصی برای توزیع نمره‌ها به حساب می‌آیند.

روش محاسبهٔ صدکها شبیه به روش محاسبهٔ میانه است، زیرا چنان‌که گفته شد میانه معادل صدک ۵۰ ($P_{۵۰}$) است. طرز محاسبهٔ صدک ۵۰ (میانه) در فصل پیش توضیح داده شد. در اینجا به عنوان مثال به محاسبهٔ صدک ۲۵ ($P_{۲۵}$) می‌پردازیم. برای این منظور از نمرات جدول ۵ـ۱۸ (فصل پیش) استفاده می‌کنیم. برای پیدا کردن صدک ۲۵، باید نمره‌ای را پیدا کنیم که ۲۵ درصد نمرات پایین‌تر از آن قرار دارند. ۲۵ درصد ۴۹ نفر (جدول ۵ـ۱۸) برابر است با: ۱۲/۲۵ = ۴۹ × ۰/۲۵ با طی همان مراحلی که در مورد محاسبهٔ ($P_{۵۰}$) انجام دادیم، باید از ستون فراوانی تراکمی بالا برویم تا به طبقهٔ ۲۹ـ۲۵ برسیم که تا آنجا فراوانیهای تراکمی ۱۱ است، و در نتیجه از طبقهٔ بعدی (۳۴ـ۳۰) نیز ۱/۲۵ نمره به این ۱۱ نمره اضافه کنیم. با استفاده از فرمولی که در مورد محاسبهٔ میانه به کار بردیم، صدک ۲۵ به شرح زیر محاسبه می‌شود:

$$P_{۲۵} = ۲۹/۵ + \left(\frac{۰/۲۵ \times ۴۹ - ۱۱}{۳}\right) ۵$$

$$= ۲۹/۵ + \frac{۱/۲۵}{۳} \times ۵ = ۲۹/۵ + ۲/۰۸$$

$$= ۳۱/۵۸$$

به عنوان مثالی دیگر، به طرز محاسبهٔ صدک ۷۰ توجه کنید. ۷۰ درصد ۴۹ نفر برابر است با ۰/۷۰ × ۴۹ = ۳۴/۳۰ پس ما باید رتبهٔ درصدی ۳۴/۳۰ را حساب کنیم. باز هم اگر به ستون فراوانی تراکمی جدول ۵-۱۸ مراجعه کنیم عدد ۳۴/۳۰ ما را به طبقهٔ ۴۴ ـ ۴۰ می‌برد و از طبقهٔ ۴۹ ـ ۴۵ که دارای فراوانی ۵ است ۱/۳۰ نمره (۱/۳۰ = ۳۳ - ۳۴/۳۰) نیاز داریم. با استفاده از فرمول بالا، صدک ۷۰ می‌شود:

$$P_{۷۰} = ۴۴/۵ + \left(\frac{۰/۷۰ \times ۴۹ - ۳۳}{۵}\right) \times ۵$$

$$= ۴۴/۵ + \left(\frac{۳۴/۳ - ۳۳}{۵}\right) \times ۵$$

$$= ۴۵/۸$$

یک نمونه استفاده از هنجارهای درصدی در جدول ۲ ـ ۱۹ نشان داده شده است. این هنجارها به **آزمونهای تشخیص استعداد**[1] تعلق دارند. چنان‌که دیده می‌شود در این جدول نمرات خام برای هر یک از خرده‌آزمونها (یا آزمونهای فرعی) و مجموعهٔ خرده‌آزمونها در ردیفهای افقی جدول و معادلهای درصدی یا صدکهای آنها در ستونهای عمودی کناره‌های جدول داده شده‌اند. روش به دست آوردن رتبهٔ درصدی هر نمرهٔ خام این است که نمره را در جای مخصوص آن در جدول مشخص کنیم و در کنار جدول رتبهٔ درصدی معادل آن را بیابیم. برای نمونه، نمرهٔ خام ۳۷ در خرده‌آزمون استدلال انتزاعی معادل رتبهٔ درصدی یا صدک ۸۰، و نمرهٔ خام ۱۹ در خرده‌آزمون توانایی عددی با صدک ۴۵ برابر است. دقت کنید که در این جدول هنجاری، تنها رتبه‌های درصدی یا صدکهای انتخابی داده شده‌اند و بنابراین، باید صدکها را نقاط وسط تعدادی صدک تصور کرد. یعنی مثلاً صدک ۸۰ را نقطهٔ میانی دامنه‌ای از صدکها تصور کرد که از ۷۸ شروع و به ۸۲ ختم می‌شوند.

یکی از معایب هنجارهای درصدی، مانند هنجارهای کلاسی و سنی، این است که واحدهای مربوط به صدکها در طول مقیاس مساوی نیستند. تفاوت صدکهای اواسط مقیاس (مثلاً تفاوت صدکهای ۴۵ تا ۵۵) خیلی کمتر از تفاوت صدکهای دو انتهای مقیاس (مثلاً ۸۵ تا ۹۵) است. علت آن است که تعداد زیادی از افراد هر گروه نمراتی به دست می‌آورند که در حوالی وسط مقیاس قرار دارند، در حالی که تعداد نسبتاً کمی از افراد نمره‌هایشان خیلی زیاد یا خیلی کم است. در نتیجه با افزایشی اندک در نمرهٔ خام دانش‌آموزی که نمره‌اش نزدیک میانگین است ممکن است مقدار زیادی بر رتبه‌اش اضافه شود، اما دانش‌آموزی که نمرهٔ بالایی دارد، برای اینکه همان مقدار به رتبه‌اش اضافه شود باید بر نمرهٔ خام او مقدار زیادی افزوده گردد. بنابراین، هنگام استفاده از رتبه‌های درصدی لازم است که تفاوت دو سه رتبه را در دو سر مقیاس بیش از حوالی وسط مقیاس با اهمیت بدانیم. در واقع تفاوتهای اندک مربوط به وسط توزیع نمرات را می‌توان نادیده گرفت.

برای مثال، دانش‌آموزی که دارای رتبهٔ درصدی ۴۴ است با دانش‌آموز دیگری که رتبهٔ

1. Differential Aptitude Test

جدول ۲ ـ ۱۹ هنجارهای آزمونهای تشخیص استعداد (شرکت روانشناسی، ۱۹۸۲، به نقل از گرانلاند و لین، ۱۹۹۰، ص ۳۸۴)

پایهٔ ۱۰، فصل پاییز	هنجارهای درصدی برای فرمهای ۷ و W								
	نمرات خام								مذکر ($N = 6215$)
صدکها	کاربرد زبان	هجی کردن	روابط فضایی	استدلال مکانیکی	منشی‌گری	استدلال انتزاعی	VR + NA	توانایی عددی NA	استدلال کلامی VR	صدکها
۹۹	۴۴_۵۰	۸۷_۹۰	۵۷_۶۰	۶۶_۷۰	۸۳_۱۰۰	۴۴_۴۵	۸۴_۹۰	۳۹_۴۰	۴۷_۵۰	۹۹
۹۷	۴۱_۴۳	۸۳_۸۶	۵۴_۵۶	۶۴_۶۵	۷۱_۸۲	۴۲_۴۳	۸۰_۸۳	۳۷_۳۸	۴۴_۴۶	۹۷
۹۵	۳۸_۴۰	۸۰_۸۲	۵۰_۵۳	۶۲_۶۳	۶۳_۷۰	۴۱	۷۴_۷۹	۳۵_۳۶	۴۱_۴۳	۹۵
۹۰	۳۵_۳۷	۷۶_۷۹	۴۷_۴۹	۶۱	۵۸_۶۲	۳۹_۴۰	۶۹_۷۳	۳۳_۳۴	۳۸_۴۰	۹۰
۸۵	۳۳_۳۴	۷۲_۷۵	۴۴_۴۶	۵۹_۶۰	۵۴_۵۷	۳۸	۶۴_۶۸	۳۱_۳۲	۳۵_۳۷	۸۵
۸۰	۳۱_۳۲	۶۹_۷۱	۴۱_۴۳	۵۸	۵۲_۵۳	۳۷	۶۱_۶۳	۲۹_۳۰	۳۳_۳۴	۸۰
۷۵	۳۰	۶۷_۶۸	۳۸_۴۰	۵۶_۵۷	۵۰_۵۱	۳۶	۵۷_۶۰	۲۸	۳۰_۳۲	۷۵
۷۰	۲۸_۲۹	۶۴_۶۶	۳۶_۳۷	۵۵	۴۸_۴۹	۳۵	۵۴_۵۶	۲۶_۲۷	۲۸_۲۹	۷۰
۶۵	۲۷	۶۲_۶۳	۳۴_۳۵	۵۴	۴۶_۴۷	۳۴	۵۱_۵۳	۲۴_۲۵	۲۷	۶۵
۶۰	۲۵_۲۶	۵۹_۶۱	۳۱_۳۳	۵۳	۴۵	۳۳	۴۸_۵۰	۲۳	۲۵_۲۶	۶۰
۵۵	۲۴	۵۷_۵۸	۲۹_۳۰	۵۱_۵۲	۴۳_۴۴	۳۲	۴۵_۴۷	۲۲	۲۳_۲۴	۵۵
۵۰	۲۲_۲۳	۵۴_۵۶	۲۷_۲۸	۵۰	۴۲	۳۱	۴۲_۴۴	۲۰_۲۱	۲۱_۲۲	۵۰
۴۵	۲۱	۵۲_۵۳	۲۵_۲۶	۴۸_۴۹	۴۰_۴۱	۳۰	۳۹_۴۱	۱۹	۲۰	۴۵
۴۰	۲۰	۴۹_۵۱	۲۳_۲۴	۴۷	۳۹	۲۹	۳۶_۳۸	۱۷_۱۸	۱۸_۱۹	۴۰
۳۵	۱۸_۱۹	۴۷_۴۸	۲۲	۴۵_۴۶	۳۸	۲۷_۲۸	۳۴_۳۵	۱۶	۱۷	۳۵
۳۰	۱۷	۴۵_۴۶	۲۰_۲۱	۴۳_۴۴	۳۶_۳۷	۲۶	۳۱_۳۳	۱۵	۱۵_۱۶	۳۰
۲۵	۱۵_۱۶	۴۳_۴۴	۱۹	۴۱_۴۲	۳۴_۳۵	۲۴_۲۵	۲۹_۳۰	۱۳_۱۴	۱۴	۲۵
۲۰	۱۴	۴۱_۴۲	۱۷_۱۸	۳۹_۴۰	۳۳	۲۱_۲۳	۲۶_۲۸	۱۲	۱۲_۱۳	۲۰
۱۵	۱۲_۱۳	۳۸_۴۰	۱۶	۳۶_۳۸	۳۰_۳۲	۱۷_۲۰	۲۳_۲۵	۱۰_۱۱	۱۱	۱۵
۱۰	۱۰_۱۱	۳۵_۳۷	۱۴_۱۵	۳۱_۳۵	۲۷_۲۹	۱۲_۱۶	۲۰_۲۲	۸_۹	۹_۱۰	۱۰
۵	۹	۳۱_۳۴	۱۲_۱۳	۲۷_۳۰	۲۳_۲۶	۸_۱۱	۱۷_۱۹	۷	۸	۵
۳	۷_۸	۲۷_۳۰	۱۱	۲۴_۲۶	۱۷_۲۲	۶_۷	۱۵_۱۶	۵_۶	۷	۳
۱	۰_۶	۰_۲۶	۰_۱۰	۰_۲۳	۰_۱۶	۰_۵	۰_۱۴	۰_۴	۰_۶	۱
میانه	۲۳/۴	۵۵/۸	۲۹/۸	۴۸/۷	۴۳/۳	۲۹/۴	۴۴/۶	۲۱/۱	۲۳/۴	میانه
انحراف معیار	۹/۵	۱۵/۵	۱۲/۵	۱۰/۸	۱۳/۷	۹/۶	۱۸/۳	۸/۹	۱۰/۸	انحراف معیار

درصدی او ۴۶ است تفاوت زیادی ندارد. چنان‌که در دنبالهٔ این فصل خواهید دید، در گزارش نمرات دانش‌آموزان به صورت نیمرخ پیشرفت، به جای نقاط درصدی از یک نوار که چندین درصد را شامل است استفاده می‌شود. این نوارهای مُعرف صدکها مبتنی بر خطای معیار اندازه‌گیری هستند که در فصل بیست و یکم معرفی خواهند شد.

نکتهٔ دیگری را که در ارتباط با رتبهٔ درصدی باید در نظر گرفت این است که برای رتبه‌های درصدی نمی‌توان میانگین محاسبه کرد. مناسب‌ترین میانگین شاخص برای رتبه‌های درصدی رتبهٔ درصدی ۵۰ است. این رتبه معرف نقطهٔ وسط توزیع نمرات است و همان نقطه‌ای است که میانه نام دارد.

هنجارهای مربوط به نمرات معیار

گفتیم که واحدهای نمره‌های درصدی یا رتبه‌های درصدی نامساوی هستند، و این عدم تساوی مقایسهٔ افراد مختلف را از لحاظ عملکرد با یکدیگر با دشواری مواجه می‌سازد. برای رفع این مشکل، متخصصان روانسنجی از روش دیگری که این عیب را ندارد استفاده می‌کنند. در این روش که به روش **نمرات معیار**[1] معروف است، موقعیت نسبی یک دانش‌آموز در یک گروه با نشان دادن فاصلهٔ نمرهٔ او از میانگین نمرات معلوم می‌شود. نمرات معیار، عملکرد هر دانش‌آموز را در یک آزمون، بنا بر اختلاف عملکرد او از میانگین گروه، بر حسب واحد انحراف معیار نشان می‌دهد. انواع مختلف نمرات معیار وجود دارند. ما در اینجا تعدادی از معروف‌ترین آنها را معرفی می‌کنیم.

نمرهٔ Z

آن دسته از نمرات معیار که دارای توزیعی با میانگین صفر و انحراف معیار *واحد* (۱) هستند به **نمرات Z**[2] (نمرات زی) شهرت دارند. برای تبدیل نمرات خام به نمرات Z از فرمول زیر استفاده می‌شود.

$$\text{نمرهٔ Z معادل نمرهٔ خام X} = \frac{X - \bar{X}}{SD}$$

در این فرمول:

X= نمرهٔ خام

$\bar{X}$= میانگین نمرات

SD = انحراف معیار نمرات

برای مثال، اگر نمرهٔ خام دانش‌آموزی در یک آزمون که دارای میانگین ۵۶ ($\bar{X}$) و انحراف معیار (SD) ۴ است، ۵۸ باشد نمرهٔ Z او برابر خواهد بود با:

1. standard scores
2. Z – scores

$$Z=\frac{۵۸-۵۶}{۴}=۰/۵$$

همچنین نمرهٔ Z دانش‌آموز دیگری که دارای نمرهٔ خام ۵۰ است برابر خواهد بود با:

$$Z=\frac{۵۰-۵۶}{۴}=-۱/۵$$

نمرهٔ Z به ما نشان می‌دهد که یک نمرهٔ خام به اندازهٔ چه تعداد از واحدهای انحراف معیار بالاتر یا پایین‌تر از میانگین است. بنابراین، دانش‌آموزی که نمرهٔ خام او به اندازهٔ یک انحراف معیار بالای میانگین باشد دارای نمرهٔ Z برابر با ۱+ و دانش‌آموزی که نمرهٔ خام او به اندازهٔ دو انحراف معیار پایین‌تر از میانگین باشد دارای نمرهٔ Z برابر با ۲- خواهد بود، و الی آخر.

همهٔ نمره‌های بالاتر از میانگین دارای نمرهٔ Z مثبت و تمام نمرات پایین‌تر از میانگین دارای نمرهٔ Z منفی هستند. نمرات Z به سهولت قابل محاسبه‌اند و به همین دلیل کاربرد زیادی دارند. با وجود این، از آنجاکه این نمرات گاه به صورت منفی و زمانی به صورت اعداد اعشاری نشان داده می‌شوند، در عمل مشکلاتی به وجود می‌آورند. به همین سبب بعضی وقتها لازم می‌آید که به جای آنها از نمرات معیار دیگری استفاده شود.

نمرهٔ T

مجموعه نمرات معیار دیگری که به **نمرات T**[1] (نمرات تی) شهرت دارند برای رفع مشکلات نمرات Z ابداع شده‌اند. مقیاس نمرات T با ضرب نمرات Z در ۱۰ و جمع کردن نمرات حاصل با ۵۰ به دست می‌آید. بنابراین، آن دسته از نمرات معیار که دارای توزیعی با میانگین ۵۰ و انحراف معیار ۱۰ است توزیع نمرات T نام دارد. می‌توان فرمول محاسبهٔ نمرات T را به صورت زیر نوشت:

$$T=۱۰(Z)+۵۰$$

برای نمونه، اگر نمرات دو دانش‌آموز در مقیاس Z برابر ۰/۵ و ۱/۵- باشند نمرات آنها در مقیاس T به شرح زیر خواهد بود:

$$T=۱۰(۰/۵)+۵۰$$
$$=۵+۵۰=۵۵$$
$$T=۱۰(-۱/۵)+۵۰$$
$$=-۱۵+۵۰=۳۵$$

از آنجاکه همهٔ نمرات T دارای میانگین ۵۰ و انحراف معیار ۱۰ هستند، هر نمرهٔ T به

1. T-scores

تنهایی و به‌طور مستقیم قابل تفسیر و استفاده است. برای مثال، نمرهٔ T = ۵۵ حاکی از این است که آن نمره به اندازهٔ نصف انحراف معیار بالای میانگین است، و نمرهٔ T = ۳۵ نشان می‌دهد که آن نمره به اندازهٔ یک‌ونیم انحراف معیار پایین میانگین قرار دارد، و الی آخر. بعضی از سازمانها و مؤسسه‌های آزمون‌سازی از نمرات معیار دارای میانگین ۵۰۰ و انحراف معیار ۱۰۰ استفاده می‌کنند. در این نمرات، نمرهٔ ۵۵۰ به اندازهٔ نصف انحراف معیار بالای میانگین قرار دارد و نمرهٔ ۳۰۰ به اندازهٔ دو انحراف معیار پایین‌تر از میانگین قرار دارد، و الی آخر. بنابراین، نمرات معیار نمراتی با میانگین و انحراف معیار دلبخواهی هستند.

نمرات SAT

یک دستهٔ دیگر از نمرات استاندارد به نمرات SAT شهرت دارند. کلمات SAT به **آزمون استعداد تحصیلی**[1] کشور آمریکا اشاره می‌کرد. از سال ۱۹۹۴ میلادی به بعد، آزمون استعداد تحصیلی به **آزمون سنجش تحصیلی**[2] با همان علامت SAT قبلی تغییر نام داد. مقیاس نمرات SAT که معرف یک توزیع بهنجار است از نتایج عملکرد ۱,۰۵۲,۰۰۰ دانش‌آموز سالهای آخر دبیرستان به دست آمده است (نیتکو، ۲۰۰۱). نمرات این گروه مرجع به صورت بهنجار با میانگین ۵۰۰ و انحراف معیار ۱۰۰ تغییر یافتند. بنابراین، نمرهٔ SAT به نحو زیر محاسبه می‌شود:

$$\text{نمرهٔ SAT} = ۱۰۰Z + ۵۰۰$$

نمرات ۹ بخشی

نوع دیگری از نمرات هنجار به نام نمرات ۹ بخشی[3] وجود دارد که به صورت نمرات معیار یک رقمی نشان داده می‌شود. علت نام‌گذاری این مقیاس نمرات به ۹ بخشی آن است که توزیع این نمرات به ۹ بخش محدود می‌شود. این مقیاس دارای میانگین ۵ و انحراف معیار ۲ است.

نیتکو (۱۹۸۳) نمرات ۹ بخشی را برای نمره‌گذاری آزمونهای تشریحی که ارزشیابی آنها جنبهٔ کیفی دارد، به ویژه آزمونهای خواندن و فهمیدن، به شرح زیر پیشنهاد داده است:

۹ـ بسیار عالی (۴٪ بالا)
۸ـ عالی (۷٪ بعدی)
۷ـ خیلی خوب (۱۲٪ بعدی)
۶ـ خوب (۱۷٪ بعدی)
۵ـ متوسط (۲۰٪ وسط)

1. Scholastic Aptitude Test
2. Scholastic Assessment Test
3. stanines (standard nine)

۴ـ نسبتاً ضعیف (۱۷٪ بعدی)
۳ـ ضعیف (۱۲٪ بعدی)
۲ـ خیلی ضعیف (۷٪ بعدی)
۱ـ بیش از حد ضعیف (۴٪ آخر)

هوشبهر انحرافی

در صفحات قبل سن ذهنی و روش محاسبهٔ هوشبهر را یاد گرفتیم. گفتیم که هوشبهر از تقسیم سن ذهنی یا سن عقلی بر سن تقویمی و ضرب نتیجهٔ حاصل در عدد ۱۰۰ به دست می‌آید.

نتایج آزمونهای هوشی قدیم بر حسب نمرات هوشبهر گزارش می‌شدند. اما بعدها آزمونهای هوشی روش فوق را کنار نهاده، به جای آن، از **هوشبهر انحرافی** (DIQ)[1] استفاده کردند. هوشبهر انحرافی یک هنجار مربوط به نمرات معیار است که دارای میانگین ۱۰۰ و انحراف معیار ۱۵ یا ۱۶ است. بنابراین، اگر انحراف معیار ۱۵ (یا ۱۶) باشد فرمول هوشبهر انحرافی به قرار زیر است:

هوشبهر انحرافی = DIQ = ۱۵ Z + ۱۰۰

هوشبهر انحرافی را مانند نمرات Z و T تفسیر می‌کنند. برای مثال، اگر هوشبهر انحرافی دانش‌آموزی ۱۳۰ باشد، این دانش‌آموز دو انحراف معیار بالاتر از میانگین قرار دارد (Z = ۲، T = ۷۰)؛ یا اگر هوشبهر انحرافی دانش‌آموزی ۸۵ باشد، یک انحراف معیار پایین‌تر از میانگین است (Z = ـ۱، T = ۴۰).

نیتکو (۱۹۸۳) می‌گوید از آنجا که هوشبهر انحرافی یک نمرهٔ معیار است، بر این اعتقاد قدیمی که سن ذهنی و سن تقویمی کودکان باید برابر باشند مبتنی نیست. برای مثال، بر اساس مفهوم قدیمی هوشبهر، کودکی که سن تقویمی و سن ذهنی او ۵ سال کامل باشد، هوشبهرش ۱۰۰ خواهد بود، اما هوشبهر واقعی این کودک در آزمون استانفورد بینه بجز این است.

> یک کودک ۵ ساله، که دارای سن ذهنی ۵ است نمرهٔ هوشبهر ۱۰۰ دریافت نمی‌کند، بلکه نمرهٔ ۹۱ به دست می‌آورد. برای اینکه این کودک نمرهٔ هوشبهر ۱۰۰ بگیرد باید نمرهٔ هوشبهر ۶ـ۵ (۵ سال و ۶ ماه) دریافت کند. (ترمن[2] و مریل[3]، ۱۹۷۳، به نقل از نیتکو، ۱۹۸۳، ص ۵۲۵)

نمرهٔ سنی معیار

نمرهٔ معیار دیگری که در آزمونهای میزان‌شدهٔ توانایی کاربرد زیادی دارد **نمرهٔ سنی معیار**[4] (SAS) است. در این مقیاس استاندارد، میانگین ۱۰۰ و انحراف معیار ۱۶ است. بنابراین، فردی که نمرهٔ SAS او

---

1. deviation IQ (DIQ)
2. Terman
3. Merrill
4. standard age score (SAS)

۱۳۲ باشد دو انحراف معیار بالای میانگین قرار دارد و کسی که نمرهٔ SAS او ۸۴ باشد یک انحراف معیار پایین میانگین قرار دارد. نمرات آزمونهای توانایی را می‌توان به همین راحتی به صورت نمرات T یا نمرات NCE نیز گزارش کرد (نگاه کنید به قسمت بعدی). لین و گرانلاند (۲۰۰۰)، در رابطه با استفاده از نمرات SAS، می‌گویند "در گذشته آزمونهای میزان‌شده نمرات را به صورت هوشبهر انحرافی با میانگین ۱۰۰ و انحراف معیار ۱۶ (در بعضی آزمونها ۱۵) گزارش می‌دادند. مقیاس SAS یا مقیاسهای مشابه آن مانند **بهر مهارتهای شناختی**[1] یا **شاخص توانایی آموزشگاهی**[2] به منظور پرهیز از پاره‌ای سوءتفاهمهای وابسته به اصطلاح هوش و هوشبهر جانشین نمرات هوشبهر انحرافی یا نمرات IQ شده‌اند" (ص ۴۹۲).

نمرهٔ معادل منحنی بهنجار

نمرهٔ معیار دیگری به نام **معادل منحنی بهنجار**[3] (یا نمرهٔ NCE) در آزمونهای میزان‌شدهٔ پیشرفت تحصیلی مورد استفاده قرار می‌گیرد. نمرات NCE با همان منظوری که نمرات معادل کلاسی یا هنجارهای کلاسی مورد استفاده قرار می‌گیرند به کار می‌روند، اما معایب آنها را ندارند (لین و گرانلاند، ۲۰۰۰). نمرات NCE نیز، مانند نمرات T، دارای میانگین ۵۰ هستند اما انحراف معیار آنها، به جای ۱۰، رقم ۲۱/۰۶ است. دلیل انتخاب این انحراف معیار آن است که نمرات NCE دامنه‌ای از ۱ تا ۹۹ داشته باشند. نمره‌های ۱، ۵۰، و ۹۹ در مقیاس NCE دارای رتبه‌های درصدی با همان ارزشهای عددی هستند (نگاه کنید به جدول ۳ـ۱۹). در جدول ۳ـ۱۹ رابطهٔ بین تعدادی رتبهٔ درصدی، نمره‌های NCE، و نمره‌های ۹ بخشی نشان داده شده. مقدار NCE با استفاده از فرمول زیر به دست می‌آید:

$$NCE = ۲۱/۰۶Z + ۵۰$$

چنان‌که در جدول ۳ـ۱۹ دیده می‌شود، به استثنای نمره‌های ۱، ۵۰، و ۹۹، در سایر نقاط مقیاس، نمرات NCE دارای رتبه‌های درصدی با ارزشهای عددی متفاوت هستند. در مقایسه با رتبه‌های درصدی، نمرات NCE در دو انتهای مقیاس بازترند. به عنوان مثال، تفاوت بین ۷۷ تا ۹۹ در مقیاس NCE با تفاوت بین ۹۰ تا ۹۹ در مقیاس رتبه‌های درصدی مطابق است. همچنین نمرات NCE در میانهٔ مقیاس جمع‌ترند. برای نمونه، در مقیاس NCE تفاوت بین ۵۰ تا ۵۵ با تفاوت بین ۵۰ تا ۶۰ در مقیاس رتبه‌های درصدی مطابق است.

در جدول ۳ـ۱۹، همچنین به رابطهٔ بین نمرات NCE و نمرات ۹ بخشی توجه کنید. اگر هر نمرهٔ NCE را به صورت اعشاری تصور کنید و آن را تا یک رقم گرد کنید به نمرهٔ ۹ بخشی می‌رسید.

---

1. cognitive skills quotient 2. school ability index
3. normal - curve equivalent (NCE)

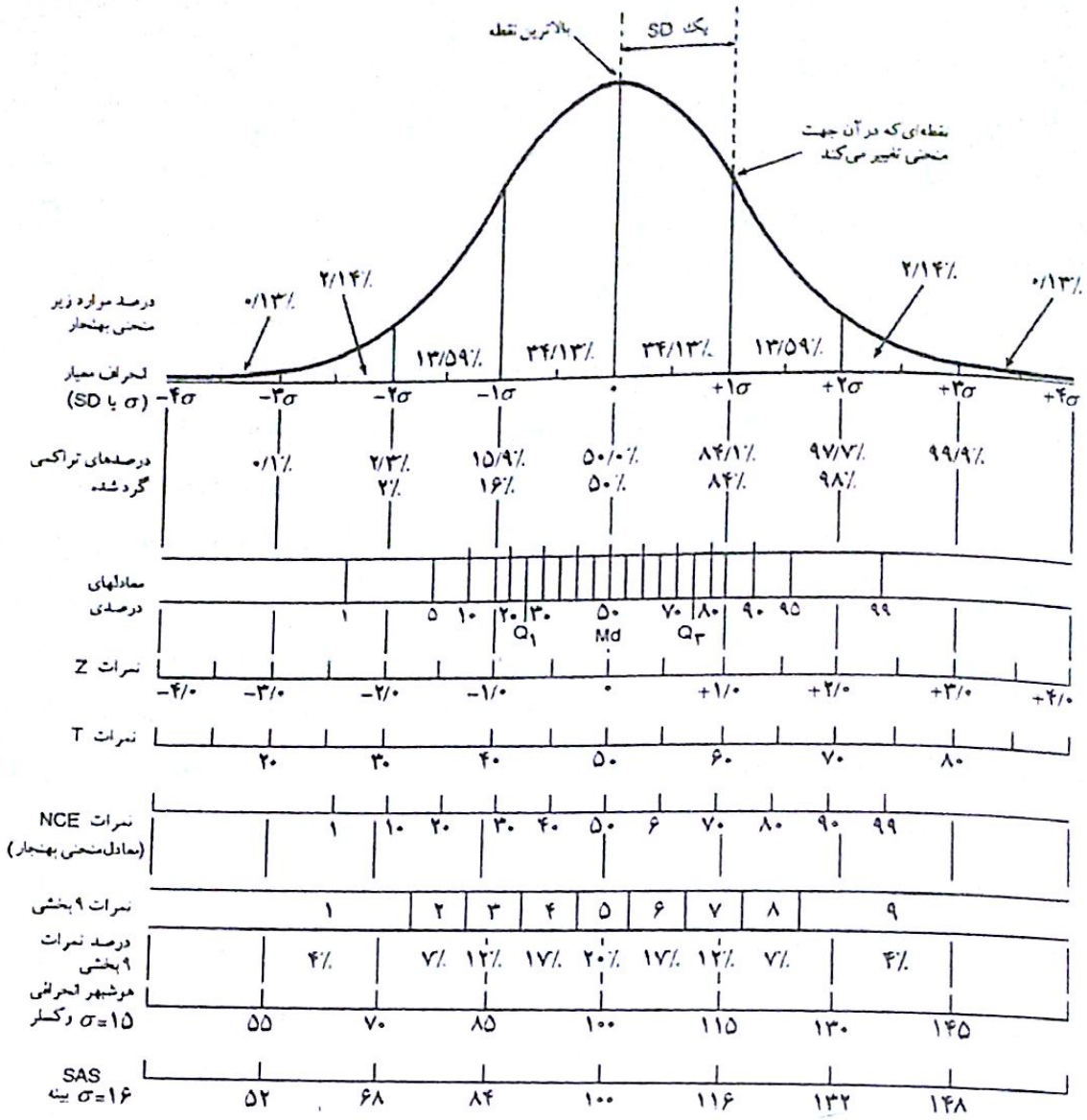
جدول ۳ ـ ۱۹ تعدادی نمرهٔ NCE و معادلهای رتبهٔ درصدی و نمرات ۹ بخشی آنها (نیتکو، ۲۰۰۱، ص ۴۰۸)

نمرهٔ ۹ بخشی	NCE	رتبهٔ درصدی
۱	۱	۱
۲	۱۵	۵
۲	۲۳	۱۰
۳	۳۲	۲۰
۴	۳۶	۲۵
۴	۳۹	۳۰
۴	۴۲	۳۵
۵	۴۵	۴۰
۵	۴۷	۴۵
۵	۵۰	۵۰
۵	۵۳	۵۵
۶	۵۵	۶۰
۶	۵۸	۶۵
۶	۶۱	۷۰
۶	۶۴	۷۵
۷	۶۸	۸۰
۷	۷۲	۸۵
۸	۷۷	۹۰
۸	۸۵	۹۵
۹	۹۹	۹۹

برای مثال، اگر نمرهٔ ۷۲ در مقیاس NCE را گرد کنیم به ۷ می‌رسیم که همان نمرهٔ معادل نمرهٔ ۹ بخشی آن است. اگر نمرهٔ ۵۸ معادل NCE را گرد کنیم به نمرهٔ ۶ معادل نمرهٔ ۹ بخشی می‌رسیم، و غیره. این تطابق نسبی نشان‌دهندهٔ این است که هم نمرات NCE و هم نمرات ۹ بخشی برخاسته از توزیع بهنجارند، و نمرات NCE و رتبه‌های درصدی دامنهٔ یکسانی دارند.

مقایسهٔ نظامهای مختلف نمرات با منحنی بهنجار

در شکل ۱ ـ ۱۹ نظامهای مختلف نمرات استاندارد و رابطهٔ آنها با صدکها و منحنی بهنجار نشان داده شده‌اند. در این شکل ماهیت تبدیل‌پذیری نظامهای مختلف نمرات با یکدیگر به‌خوبی دیده می‌شود. برای نمونه، نمرهٔ خامی که ۱ انحراف معیار پایین میانگین است (-1σ)، برابر است با



شکل ۱-۱۹ مقایسه درصدها و نمرات معیار در توزیع بهنجار (گرانلاند و لین، ۱۹۹۰)

۱- = Z، رتبهٔ درصدی ۱۶، ۴۰ = T، ۲۹ = NCE، ۸۵ = IQ، ۸۴ = SAS، و نمرهٔ ۹ بخشی ۳. بنابراین، نظامهای گوناگون نمره گذاری راههای مختلف بیان مطلبی واحد هستند و ما می‌توانیم به‌سادگی از هر مقیاسی که بخواهیم به مقیاس دیگری برویم. البته این کار در صورتی عملی است که توزیع نمرات مورد نظر ما یک توزیع بهنجار باشد.

توزیع بهنجار و منحنی بهنجار

منظور از **توزیع بهنجار**[1] توزیعی است که بر **منحنی بهنجار**[2] منطبق است. منحنی بهنجار از لحاظ شکل شبیه به منحنی شکل ۱-۱۹ و دارای ویژگیهای زیر است:

1. normal distribution
2. normal curve

۱. دو قسمت منحنی بهنجار نسبت به خط وسط آن قرینه هستند.

۲. در توزیعی که این منحنی مُعرف آن است اکثر نمره‌ها در وسط توزیع انباشته شده‌اند.

۳. هر چه از وسط منحنی به دو طرف آن پیشتر برویم از ارتفاع منحنی (حجم نمرات) کاسته می‌شود، و در دو انتهای منحنی دنباله‌های طویلی به وجود می‌آیند.

۴. دو امتداد منحنی بهنجار در بینهایت محور افقی را قطع می‌کنند، زیرا تعداد مواردی که لازم است تا یک منحنی بهنجار واقعی درست شود از لحاظ نظری (تئوری) بی‌نهایت است.

۵. در منحنی بهنجار، میانگین، میانه، و نما بر روی هم قرار دارند.

۶. در روی این منحنی دو نقطه وجود دارند که در آنها جهت منحنی تغییر می‌کند. این دو نقطه در فواصل ۱± انحراف معیار از میانگین قرار دارند.

۷. در منحنی بهنجار، بین انحراف معیارها و نسبت افراد یک رابطهٔ دقیق ریاضی وجود دارد، و در هر واحد انحراف معیار همیشه یک نسبت ثابت از افراد موجود است. اگر تمام سطح منحنی بهنجار را برابر یک واحد فرض کنیم، بین میانگین و ۱+ یا ۱- انحراف معیار ۳۴/۱۳٪ از کل مساحت منحنی قرار دارد. همچنین ۶۸/۲۶٪ = ۳۴/۱۳ × ۲ از کل سطح منحنی بین ۱+ و ۱- انحراف معیار واقع است. ضمناً در حدود ۹۵/۴۴٪ مساحت زیر منحنی بین ۲+ و ۲- انحراف معیار قرار دارد، و در حدود ۹۹/۷۴٪ این مساحت نیز در فاصلهٔ ۳+ و ۳- انحراف معیار واقع است. بنابراین، تقریباً تمام سطح زیر منحنی بهنجار در فاصلهٔ ۳+ و ۳- انحراف معیار قرار دارد. به همین سبب است که محققان و آمارگران سطح منحنی بهنجار را عملاً بین ۳+ و ۳- انحراف معیار فرض می‌کنند (نگاه کنید به منحنی بهنجار شکل ۱ـ۱۹).

نیمرخ

یکی از فواید تبدیل نمره‌های خام به نمره‌های مشتق این است که با استفاده از نمره‌های مشتق عملکرد یک دانش‌آموز را در آزمونهای مختلف می‌توان مستقیماً با هم مقایسه کرد. چنین کاری معمولاً به توسط **نیمرخ**[1] آزمون امکان‌پذیر است. به کمک نیمرخ که یک بازنمایی نگاره‌ای[2] از نمرات آزمون است، می‌توان نقاط قوت و ضعف یک دانش‌آموز را مشخص کرد.

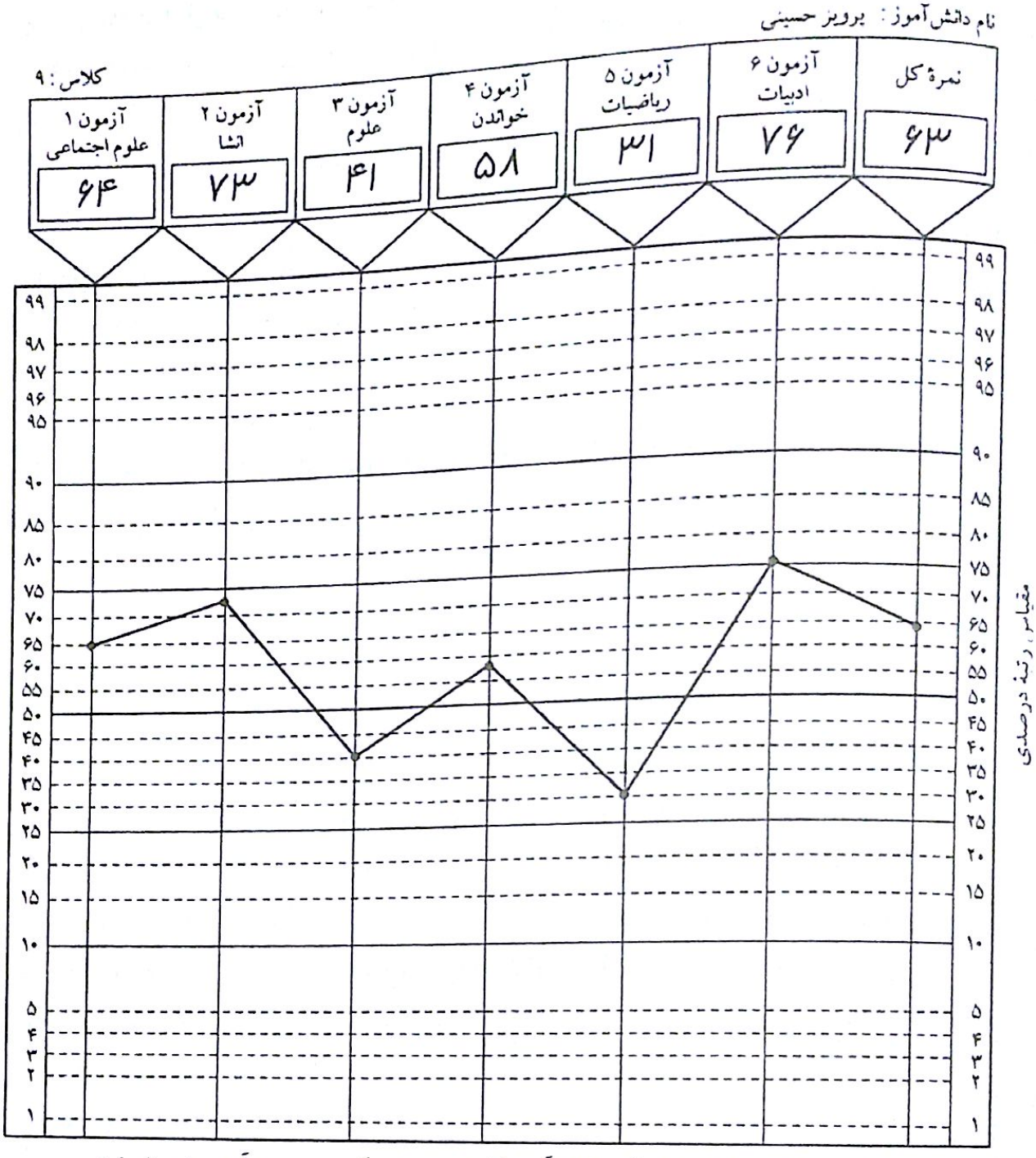
به عنوان مثال، برای مقایسهٔ نمرهٔ یک دانش‌آموز در آزمون درس علوم که تنها به ۱۰ سؤال آن جواب داده است با نمرهٔ همین دانش‌آموز در آزمون درس ریاضی که به ۱۶ سؤال آن پاسخ گفته است، باید این دو نمره را به نوعی هنجار کلاسی، سنی، یا درصدی تبدیل کنیم تا مقایسهٔ آنها امکان‌پذیر شود. بنابراین، هرگاه تعدادی نمرهٔ حاصل از آزمونهای مختلف یا آزمونهای فرعی یک آزمون کلی را برای یک دانش‌آموز به یکی از هنجارها تبدیل کنیم و این نمرات را در یک نمودار

---

1. profile
2. graphic representation

قرار دهیم نیمرخ عملکرد آن دانش‌آموز را درست کرده‌ایم.

شکل ۲ ـ ۱۹ نیمرخ نمره‌های یک دانش‌آموز کلاس نهم را در درسهای علوم‌اجتماعی، انشا، علوم، خواندن، ریاضیات، و ادبیات نشان می‌دهد. این نیمرخ با استفاده از هنجار درصدی یا رتبه‌های درصدی درست شده است. در این نیمرخ رتبه‌های دانش‌آموز برای درسهای مختلف در خانه‌های بالای شکل (زیر نام درسها) نوشته شده‌اند. ضمناً در آخرین خانهٔ سمت راست بالای شکل رتبهٔ کل دانش‌آموز داده شده است. برای ترسیم نیمرخ نمرات دانش‌آموز، رتبه‌های او در درسهای



شکل ۲-۱۹ نیمرخ نمرات یک دانش‌آموز کلاس نهم در یک مجموعه آزمون پیشرفت تحصیلی (اقتباس از اثر نول و اسکانل، ۱۹۷۲)

مختلف، با توجه به درصدهای کنار شکل، در روی شکل تعیین و با خطی به هم وصل شده و به صورت یک نمودار نشان داده شده است. این گونه نمودارها وسیلهٔ مناسبی برای گزارش دادن نمرات درسهای مختلف دانش‌آموزان به اولیای آنان است. همچنین نیمرخها وسایل مناسبی هستند که با یک نگاه کلی به آنها می‌توان نظری اجمالی در بارهٔ نقاط قوت و ضعف دانش‌آموزان به دست آورد.

مقیاس به کار رفته در شکل ۲ ـ ۱۹ مقیاس رتبهٔ درصدی است، اما در تعیین نقاط درصدی عدم تساوی واحدها در نظر گرفته شده است. یعنی نحوهٔ قرار گرفتن صدکها طوری است که با منحنی بهنجار مطابقت دارد، به نحوی که نقاطی که در انتها قرار دارند فاصله‌دارتر و نقاط وسط به یکدیگر نزدیک‌ترند. این مقیاس شبیه به مقیاس رتبه‌های درصدی است که در شکل ۱ ـ ۱۹ نشان داده شده است.

یکی از فواید نیمرخها کسب یک برآورد کلی از وضعیت پیشرفت دانش‌آموز با یک نظر اجمالی است. با توجه به نیمرخ شکل ۲ ـ ۱۹ معلوم می‌شود که درصد نمره‌های این دانش‌آموز در درسهای علوم‌اجتماعی، انشا، خواندن، و ادبیات بالاتر از میانه (۵۰ درصد) گروه هنجار است. اما درصد نمرات او در درسهای علوم و ریاضی پایین‌تر از هنجار است. نمرهٔ کل این دانش‌آموز وضع پیشرفت کلی او را نشان می‌دهد که بالاتر از هنجار است.

در شکل ۲ ـ ۱۹ علاوه بر میانه یا صدک ۵۰، رتبه‌های ۹۰، ۷۵، ۲۵، و ۱۰ نیز با خط ممتد نشان داده شده‌اند. این درصدها برای تفسیر نمرات، نقاط درصدی یا صدکهای مناسبی هستند.

پیش‌تر گفتیم که نیمرخها را، علاوه بر فرد فرد دانش‌آموزان، برای تمام افراد یک کلاس یا یک گروه نیز می‌توان مورد استفاده قرار داد. برای این منظور، باید از نمرات میانگین هر یک از آزمونهای دروس مختلف استفاده کرد. این نوع نیمرخهای گروهی برای تعیین نقاط قوت و ضعف گروه دانش‌آموزان مفید است.

باید توجه داشت که تفاوت نمره‌های یک دانش‌آموز در آزمونهای مختلف ممکن است ناشی از خطای اندازه‌گیری[1] آزمونها باشد نه یک تفاوت واقعی. بنابراین، باید تفاوت میان نمرات آزمونها را با احتیاط تفسیر کنید. این احتیاط به ویژه زمانی خیلی ضروری است که نمرات مورد نظر نمرات خُرده‌آزمونها یا آزمونهای فرعی یک مجموعه آزمون باشند که از تعداد کمی سؤال درست شده‌اند و در نتیجه دارای ضریب پایایی[2] کوچکی هستند. به عنوان یک قانون کلی، نمرات حاصل از خرده‌آزمونهایی راکه دارای ضریب پایایی کمتر از ۰/۹۰ باشند باید با احتیاط تفسیر کرد، و آنها را صرفاً به عنوان شاخصی کلی از پیشرفت دانش‌آموز در یک زمینهٔ بخصوص درسی به حساب آورد، و نیمرخ چنین نمراتی را باید تنها به صورتی خیلی کلی تفسیر کرد.

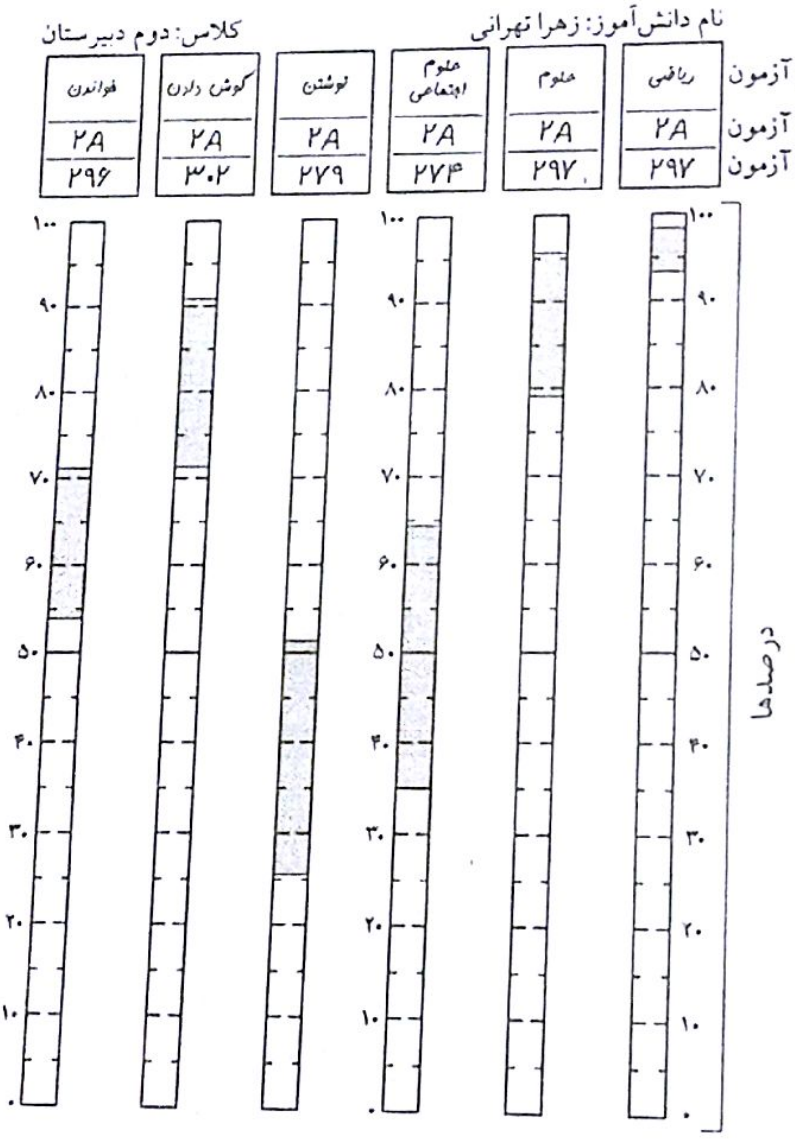
---

۱. مفهوم خطای اندازه‌گیری و خطای معیار اندازه‌گیری در فصل بیست و یکم معرفی خواهد شد.

۲. مفهوم پایایی در فصل بیست و یکم معرفی می‌شود.

روش دیگر ترسیم نیمرخ نمرات که برای رفع مشکل بالا پیشنهاد شده این است که، به عوض نشان دادن نمرات آزمون به صورت نقطه‌هایی در روی نیمرخ نمرات، عملکرد دانش‌آموز به صورت نوارهایی ترسیم شود که هر نوار به اندازهٔ یک خطای معیار اندازه‌گیری بالای نمرهٔ دانش‌آموز و یک خطای معیار اندازه‌گیری پایین آن نمره را شامل شود. در مبحث خطای معیار اندازه‌گیری (نگاه کنید به فصل بیست و یکم) توضیح داده شده است که احتمال دو بر سه وجود دارد که نمرهٔ واقعی دانش‌آموز در حد فاصل یک خطای معیار اندازه‌گیری بالا و پایین نمرهٔ اکتسابی او در آزمون قرار داشته باشد. بنابراین، ترسیم نیمرخ به صورت نوارهای پیش‌گفته، ما را قادر می‌سازد تا هنگام مقایسهٔ نمرات آزمونهای متفاوت، احتیاط لازم را در مورد کم‌دقتی آزمونها رعایت کنیم. نمونه‌ای از این نوع نیمرخها در شکل ۳ـ۱۹ نشان داده شده است.

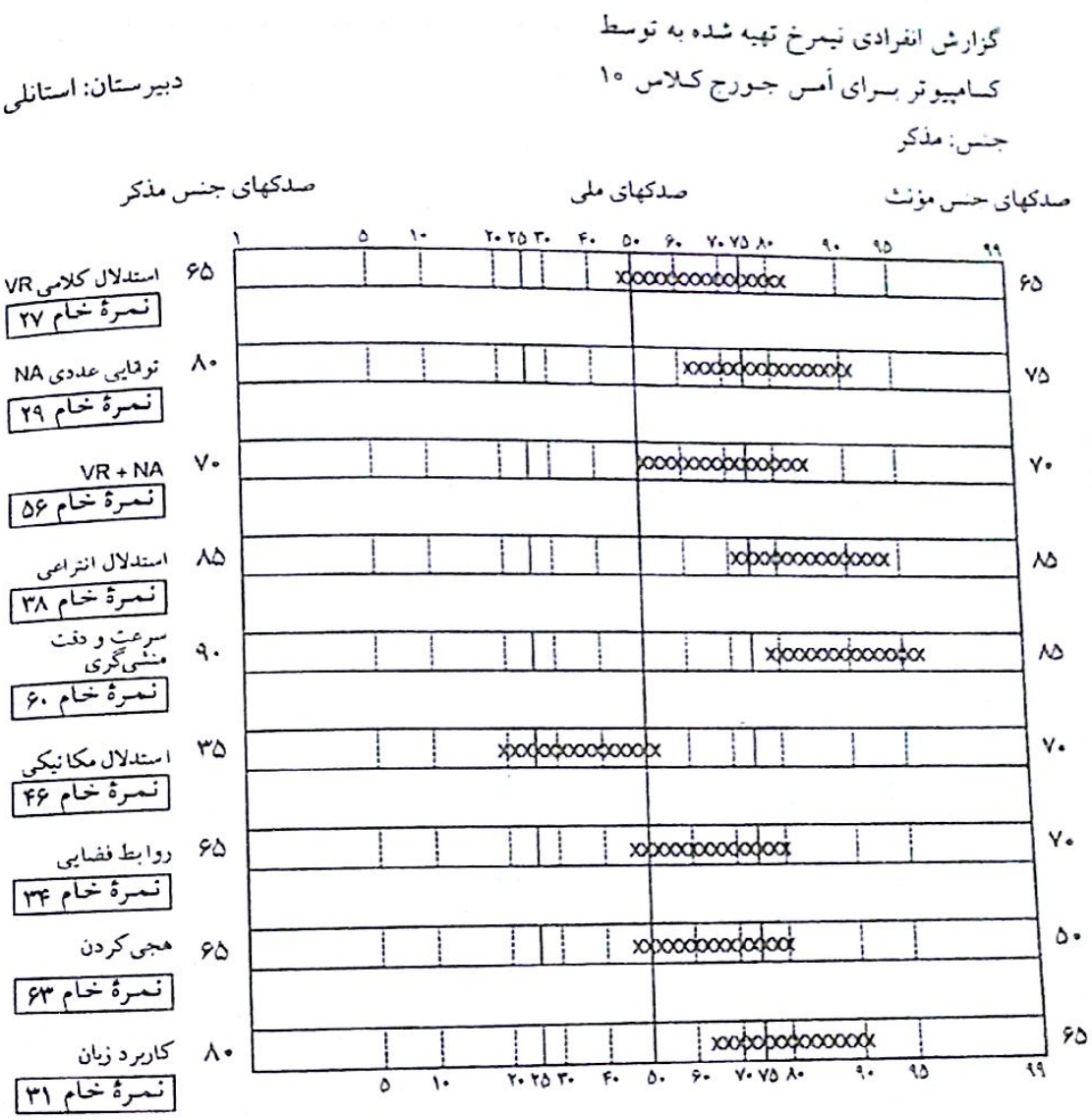
تفسیر نمرات آزمونها با استفاده از این نوارها کار آسانی است. اگر نوارهای مربوط به دو نمرهٔ دانش‌آموز با هم وجه اشتراک یا همپوشی داشته باشند، چنین فرض می‌کنیم که عملکرد دانش‌آموز در این دو آزمون تفاوت چندانی ندارد. اگر نوارها با هم وجه اشتراک نداشته باشند، آن‌گاه می‌توانیم



شکل ۳-۱۹ نیمرخ نمرات یک دانش‌آموز کلاس دوم دبیرستان در آزمونهای متوالی پیشرفت تحصیلی (اقتباس از اثر گرانلاند، ۱۹۷۱)

فرض کنیم که احتمالاً یک تفاوت واقعی بین نمرات آزمونهای مورد نظر موجود است.

غالباً نیمرخ نمرات را هم می‌توان با دست تهیه کرد و هم به توسط کامپیوتر. شکل ۴ـ۱۹ نمونه‌ای از نیمرخ نمرات تهیه‌شده به‌وسیلهٔ کامپیوتر است. این نیمرخ از روی هنجارهای درصدی جدول ۲ـ۱۹ همین فصل (هنجارهای آزمونهای تشخیص استعداد) تهیه شده است. این نیمرخ بر اساس صدکهای یک جنس (جنس مذکر) تهیه شده است و این صدکها در ستون سمت چپ شکل وارد شده‌اند. با این حال، صدکهای جنس مخالف (جنس مؤنث) نیز در ستون کنار سمت راست شکل داده شده‌اند تا مقایسهٔ دو جنس با همدیگر میسر گردد. علت ضروری بودن گزارش دو نوع صدک یا هنجار درصدی برای جنس مذکر و مؤنث تفاوت چشم‌گیر این دو جنس در نمراتشان در بعضی از خرده‌آزمونهاست. برای نمونه عملکرد دو جنس مذکر و مؤنث را در نیمرخ شکل ۴ـ۱۹ در خرده‌آزمونهای استدلال مکانیکی و کاربرد زبان با هم مقایسه نمایید.



شکل ۴ـ۱۹ نیمرخ نمرات یک دانش‌آموز کلاس دهم در آزمونهای تشخیص استعداد
(به نقل از گرانلاند و لین، ۱۹۹۰)

خلاصهٔ فصل

ما در فصل قبل یک رشته از اقداماتی را که برای قابل استفاده کردن نمرات خام حاصل از آزمونها به کار می‌روند با نام روشهای آماری به خوانندگان معرفی کردیم. در فصل حاضر نیز دو اقدام اساسی دیگر را که به‌همین منظور به کار می‌روند موردبحث قراردادیم، یعنی هنجارها و نیمرخها. در قسمت اول فصل زیر عنوان هنجارهای مختلف یا نمرات مشتق موارد زیر را توضیح دادیم:

۱. هنجار سنی
۲. هنجار کلاسی
۳. هنجار سنی‌نمایی
۴. هنجار سن ذهنی
۵. هوشبهر
۶. هنجار درصدی
۷. نمرهٔ Z
۸. نمرهٔ T
۹. نمرات ۹ بخشی
۱۰. هوشبهر انحرافی
۱۱. نمرهٔ سنی معیار
۱۲. نمرهٔ معادل منحنی بهنجار یا نمرهٔ NCE

هنجار وسیله‌ای است که به کمک آن و با استفاده از نمرهٔ خام یک دانش‌آموز عملکرد او را در آزمون با عملکرد یک گروه مرجع مقایسه می‌کنیم یا موقعیت نسبی او را در گروهی که او خود عضوی از آن است نشان می‌دهیم. پس از آنکه نمرات خام را به هنجارها یا نمرات مشتق تبدیل کردیم، می‌توانیم عملکرد یک دانش‌آموز را در آزمونهای مختلف با هم مقایسه نماییم. این کار به توسط نیمرخ آزمون میسر است. ما در قسمت دوم این فصل نیمرخها، انواع آنها، و موارد استفادهٔ نیمرخهای مختلف را نیز توضیح دادیم.

علاوه بر هنجارها و نیمرخها، ما در این فصل منحنی بهنجار و توزیع بهنجار را تعریف و ویژگیهای آنها را با هم مقایسه کردیم. در ضمن به کمک توزیع بهنجار و منحنی بهنجار انواع هنجارها و نمرات معیار را نیز با هم مقایسه نمودیم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. اگر صدک ۸۰ مربوط به عدد ۶۷ باشد، یعنی اینکه:

الف ـ ۶۷ درصد افراد نمرهٔ کمتر از ۸۰ گرفته‌اند.

ب ـ ۲۳ درصد افراد نمرهٔ زیر ۶۷ گرفته‌اند.
ج ـ ۸۰ درصد افراد نمرهٔ پایین ۶۷ گرفته‌اند.
ت ـ ۸۰ درصد افراد نمرهٔ زیر ۸۰ گرفته‌اند.

۲. کاربرد هنجار کلاسی در چه سطحی بهترین نتیجه را به دست می‌دهد؟
الف ـ دبستان
ب ـ راهنمایی
پ ـ دبیرستان
ت ـ دانشگاه

۳. در نمرات کلاسی گروهی از دانش‌آموزان که دارای میانگین ۱۵ و انحراف معیار ۲ است نمرهٔ Z مربوط به ۱۹ چند است؟
الف ـ ۱
ب ـ ۲
پ ـ ۴
ت ـ ۶

۴. در کدام‌یک از نمرات معیار میانگین ۵۰ و انحراف معیار ۱۰ است؟
الف ـ هوشبهر انحرافی
ب ـ نمرات ۹ بخشی
پ ـ نمرات Z
ت ـ نمرات T

۵. با توجه به توزیع بهنجار در یک آزمون هوش که دارای انحراف معیار ۱۵ است، هوشبهر ۱۴۵ برابر با چند نمرهٔ T است؟
الف ـ ۵۰
ب ـ ۶۰
پ ـ ۷۰
ت ـ ۸۰

۶. در منحنی بهنجار حدود چند درصد سطح منحنی بین ۲ + و ۲ ـ انحراف معیار قرار دارد؟
الف ـ ۳۴
ب ـ ۶۸
ج ـ ۹۵
ت ـ ۹۹

۷. هنجار را تعریف کنید و روش تعیین هنجارهای سنی و کلاسی را توضیح دهید.

۸. معایب هنجارهای سنی و کلاسی را توضیح دهید.

۹. هنجارهای سنی نمایی را چگونه تعیین می‌کنند؟

۱۰. سن ذهنی را تعریف کنید و روش تعیین آن را شرح دهید.

۱۱. اگر سن ذهنی دانش‌آموزی ۶ و سن تقویمی او ۵ باشد هوشبهر او چند خواهد بود؟

۱۲. هوشبهر انحرافی را تعریف کنید و تفاوت آن را با هوشبهر شرح دهید.

۱۳. تفاوت بین هنجارهای درصدی با هنجارهای سنی و کلاسی را توضیح دهید.

۱۴. در آزمونی که دارای میانگین ۵۰ و انحراف معیار ۴ است، نمرات Z نمرات خام ۵۵، ۴۰، ۳۰، ۲۰، و ۷۰ را حساب کنید.

۱۵. نمرات ۹ بخشی را تعریف کنید و بهترین مورد استفادهٔ آنها را شرح دهید.

۱۶. با مراجعه به شکل ۱ـ۱۹ تعیین کنید که انحراف معیار ۱+ و ۲- معادل چه مقدار Z، T، صدک و هوشبهر انحرافی است.

۱۷. ویژگیهای منحنی بهنجار را تعریف کنید.

۱۸. تعیین کنید که چه درصدی از سطح کلی منحنی بهنجار بین ۱/۵± انحراف معیار قرار دارد؟

۱۹. نیمرخ را تعریف کنید و نحوهٔ استفاده از آن را توضیح دهید.

۲۰. نمرات مختلف یک دانش‌آموز در چند درس مختلف را به صورت نیمرخ پیشرفت تحصیلی رسم کنید. اگر به نمرات واقعی دسترسی ندارید، تعدادی نمره را به طور دلبخواهی مورد استفاده قرار دهید.

۲۰

روایی آزمون

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. ویژگیهای مطلوب آزمونها را نام ببرد و مهم‌ترین آنها را توضیح دهد.
۲. روایی آزمون را تعریف کند و انواع روایی‌ها را نام ببرد.
۳. روایی محتوایی را تعریف کند.
۴. روایی صوری را تعریف کند و رابطهٔ آن را با روایی محتوایی توضیح دهد.
۵. روش تعیین روایی محتوایی و روایی صوری را توضیح دهد.
۶. روایی پیش‌بینی را تعریف کند.
۷. روایی همزمان را تعریف کند و رابطهٔ آن را با روایی پیش‌بینی توضیح دهد.
۸. روش تعیین روایی پیش‌بینی و روایی همزمان را توضیح دهد.
۹. ویژگیهای مطلوب ملاک را توضیح دهد.
۱۰. روایی سازه را تعریف کند.
۱۱. راههای تعیین روایی سازه را توضیح دهد.
۱۲. ویژگیهای روایی‌های مختلف را با هم مقایسه کند.
۱۳. روشهای تعیین روایی محتوایی را برای آزمونهای وابسته به ملاک توضیح دهد.
۱۴. روشهای تعیین روایی پیش‌بینی را برای آزمونهای وابسته به ملاک توضیح دهد.
۱۵. روشهای تعیین روایی سازه را برای آزمونهای وابسته به ملاک توضیح دهد.
۱۶. عوامل درونی و بیرونی مؤثر بر روایی آزمون را توضیح دهد.
۱۷. با مراجعه به جدول مشخصات یک موضوع درسی که در طرح آزمون پیشرفت تحصیلی برای آن موضوع درسی به کار بسته است، دلایل روا بودن آزمون خود را توضیح دهد.

چنان‌که در فصلهای قبلی این کتاب دیدیم، آزمونها و سایر وسایل اندازه‌گیریِ مورد استفاده در روانشناسی و آموزش و پرورش برای مقاصد مختلفی به کار می‌روند. بعضی از آنها برای اندازه‌گیری مقدار یادگیری دانش‌آموزان در درسهای مختلف مورد استفاده قرار می‌گیرند. بعضی دیگر برای تعیین پیشرفت یادگیرندگان در ضمن یادگیری و اصلاح مشکلات یادگیری آنان مورد استفاده قرار می‌گیرند. دسته‌ای از آنها برای سنجش آمادگیها و تواناییهای داوطلبان یک دورهٔ تحصیلی و انتخاب شایسته‌ترین آنها مورد استفاده قرار می‌گیرند. و بالاخره دسته‌ای دیگر برای مقاصد دیگری به کار می‌روند، مثلاً سنجش هوش و شخصیت و غیره. هر نوع وسیلهٔ اندازه‌گیری به هر منظوری که تهیه می‌شود باید ویژگیهایی را دارا باشد تا آن وسیله برای منظوری که ساخته شده است مفید باشد. متخصصان اندازه‌گیری و ارزشیابی برای آزمونهای مختلف ویژگیهای زیادی را برشمرده‌اند. از جمله می‌توان عینیت، *سهولت اجرا، سهولت نمره‌گذاری، عملی‌بودن، سهولت تعبیر و تفسیر*، **روایی**، و **پایایی** را نام برد. مهم‌ترین این ویژگیها **روایی** و **پایایی**[1] هستند. بنابراین، ما در این فصل و فصل بعدی کتاب به معرفی این دو ویژگی می‌پردازیم.

تعریف و انواع روایی

روایی اصطلاحی است که به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است اشاره می‌کند. لین و گرانلاند (۲۰۰۰) روایی را به صورت "یک ارزشیابی از کفایت و مناسبت تفسیرها و استفاده‌های نتایج سنجش" (ص ۷۳) تعریف کرده‌اند. کاپلان و ساکوزو (۲۰۰۱) گفته‌اند "روایی را می‌توان به صورت توافق بین نمرهٔ آزمون و کیفیتی که قرار است اندازه بگیرد تعریف کرد" (ص ۱۳۲).

بنابر آنچه گفته شد، آزمونی دارای روایی است که برای اندازه‌گیری آنچه مورد نظر است کافی و مناسب باشد. برای مثال، یک آزمون پیشرفت تحصیلی ریاضی کلاس پنجم دبستان در صورتی یک آزمون **روا**[2] است که محتوا و هدفهای آن درس را به خوبی اندازه‌گیری کند و بجز یادگیری دانش‌آموزان از آن درس چیز دیگری را شامل نباشد. به عنوان مثالی دیگر از روایی، یک آزمونِ هوشیِ روا آزمونی است که تنها هوش افراد را اندازه‌گیری کند نه چیز دیگری را. بنابراین، یک آزمون ممکن است از جهتی روا باشد اما از جهاتی دیگر روا نباشد. پس، روایی یک امر نسبی است، نه یک ویژگی که هست یا نیست.

آزمونهای مورد استفاده در آموزش و پرورش و روانشناسی دارای سه هدف عمده هستند. این هدفها به سه پرسش زیر پاسخ می‌دهند:

---

۱. ما دراین کتاب اصطلاح روایی را برای validity و اصطلاح پایایی را برای reliability به کار برده‌ایم. در ترجمهٔ این دو اصطلاح خارجی به زبان فارسی در میان صاحب‌نظران اختلاف نظر وجود دارد. مؤلف از آنجا این دو اصطلاح را ترجیح داده است که اولاً هر دو فارسی و در ثانی از لحاظ مشتقات قابل انعطاف‌اند.

2. valid

۱. آیا سؤالهای آزمون از محتوای درس یا موضوع مورد نظر یک نمونهٔ خوب ارائه می‌دهند؟
۲. آیا نمرات آزمون عملکرد فعلی یا آتی دانش‌آموزان را پیش‌بینی می‌کنند؟
۳. آیا نمرات آزمون به مفاهیم نظری یا سازه‌هایی که آزمون برای سنجش آنها درست شده است مربوط‌اند؟

سه مورد بالا به ترتیب سه نوع روایی یعنی *محتوایی*، *ملاکی*، و *سازه* را نشان می‌دهند. روشهای تعیین روایی محتوایی (و روایی صوری که مشتق روایی محتوایی است) به روشهای **قضاوتی**[1] معروف‌اند و در آنها خودِ ابزار اندازه‌گیری مورد قضاوت و داوری واقع می‌شود. روشهای تعیین روایی ملاکی (روایی پیش‌بینی و روایی همزمان) و روایی سازه روشهای **تجربی**[2] نام دارند، زیرا در آنها ماهیت پاسخهایی که به سؤالهای آزمون داده می‌شوند مورد بررسی قرار می‌گیرد (آندرسون و بورک، ۲۰۰۱).

یعنی اینکه، در مورد اول تعیین می‌شود که ابزار اندازه‌گیری تا چه میزانی می‌تواند ویژگی موردنظر را بسنجد و در مورد دوم تعیین می‌شود که پاسخهای داده شده به ابزار اندازه‌گیری تا چه حدی شواهد لازم را در این باره که ابزار موردنظر واقعاً ویژگی موردنظر را می‌سنجد به دست می‌دهد.

روایی محتوایی

از آنجاکه همهٔ آزمونهای پیشرفت تحصیلی در یک زمان محدود به کار می‌روند، و از آنجاکه هیچ آزمونی نمی‌تواند تمام محتوا و کلیهٔ هدفهای یک درس را شامل شود، پس یک آزمون از مجموعه‌ای سؤال که به عنوان نمونه‌ای از تمامی محتوا و هدفهای آموزشی درس انتخاب شده است تشکیل می‌یابد. (**روایی محتوایی**[3] به این مطلب اشاره می‌کند که نمونه سؤالهای مورد استفاده در یک آزمون تا چه حد معرف کل جامعهٔ سؤالهای ممکن است که می‌توان از محتوا یا موضوع مورد نظر تهیه کرد) هر چه آزمون از این لحاظ بهتر باشد، دارای روایی بیشتری است) بنابراین، اگر معلم درس فیزیک بخواهد برای درس خود یک آزمون پیشرفت تحصیلی روا بسازد، آزمون او باید دربرگیرندهٔ نمونه‌ای درست و دقیق از مطالب فیزیک باشد که به دانش‌آموزان آموزش داده است. در ضمن، این آزمون نباید چیزی خارج از محتوا و هدفهای درس فیزیک این معلم را اندازه‌گیری بکند.

با توجه به توضیحات بالا، برای ساختن آزمونهای پیشرفت تحصیلی روا، باید بکوشیم تا سؤالهای آزمون نمونهٔ کاملی از هدفها و محتوای درس باشند. جدول مشخصات بهترین روش برای انجام دادن چنین کاری است. طرز تهیهٔ جدول مشخصات در فصل ششم کتاب معرفی شده است. چنان‌که در فصل ششم آمده است، جدول مشخصات وسیله‌ای است که در آن محتوای درس و هدفهای آموزشی به‌خوبی ارائه می‌شوند. معلم می‌تواند، با تهیهٔ چنین جدولی برای درسی که می‌خواهد برای آن

1. judgmental
2. empirical
3. content-related validity

ازمون بسازد، در آزمون خود نمونهٔ کاملی از محتوا و هدفهای درس مورد نظر را منظور کند.

روش تعیین روایی محتوایی

هیچ‌گونه روش آماری برای تعیین ضریب روایی محتوایی وجود ندارد. در عوض برای تعیین روایی محتوایی یک آزمون از قضاوت متخصصان در این باره که سؤالهای آزمون تا چه میزانی مُعرف محتوا و هدفهای برنامه یا *حوزهٔ محتوایی*[1] هستند استفاده می‌شود. با این حال، "قضاوتهای مربوط به روایی محتوایی نه قطعی هستند و نه نهایی؛ آزمونها از سطوح مختلف روایی برخوردارند و متخصصان در قضاوتهای خود همیشه با هم توافق نظر ندارند" (مرفی و دیویدشوفر، ۱۹۹۴، ص ۱۱۰).

روایی صوری

روایی صوری[2] یکی از مشتقات روایی محتوایی است. روایی صوری به این مطلب اشاره می‌کند که سؤالهای آزمون تا چه حد در ظاهر شبیه به موضوعی هستند که برای اندازه‌گیری آن تهیه شده‌اند. مورفی و دیوید شوفر (۲۰۰۱) گفته‌اند "اصطلاح روایی صوری به درجه‌ای از منطقی بودن که یک آزمون به نظر کسانی که به آن جواب می‌دهند می‌رسد گفته می‌شود" (ص ۱۵۴).

در واقع روایی صوری نمی‌تواند نوعی روایی باشد، بلکه تنها یک ویژگی آزمون است که در پاره‌ای مواقع وجود آن مفید به نظر می‌رسد. ثرندایک و هیگن (۱۹۷۹) در رد اهمیت روایی صوری گفته‌اند در واقع آنچه که تعیین می‌کند یک آزمون چه چیزی را اندازه می‌گیرد تکالیفی است که به وسیلهٔ سؤالهای آزمون معرفی می‌شوند، و کسی که بخواهد روایی محتوایی یک آزمون را در رابطه با برنامهٔ درسی مورد قضاوت قرار دهد یا قدرت پیش‌بینی یک آزمون را تعیین کند باید سؤالهای آزمون را به‌دقت بررسی نماید، و "برای این منظور، ما به صورت ظاهر آزمون چندان اعتنایی نمی‌کنیم" (ص ۶۰).

با وجود اظهارات فوق، ثرندایک و هیگن (۱۹۷۹) می‌گویند که گاه صورت ظاهر یک آزمون (روایی صوری)، از لحاظ قابلیت پذیرش و معقول بودن آزمون برای آزمون‌شوندگان، دارای اهمیت است. برای مثال، داوطلبانان خلبانی یک آزمون ریاضی را که در آن مسایل مربوط به جریان باد و مصرف بنزین طرح شده است، به آزمون ریاضی دیگری که مسایل را در قالب هزینهٔ غلات یا فهرست مواد خام برای تهیهٔ نوعی شیرینی بیان می‌کند ترجیح می‌دهند. بنابراین، می‌توان گفت که تا آن اندازه که ظاهر آزمون بر انگیزش آزمون شونده اثر می‌گذارد، روایی صوری می‌تواند مهم باشد. در بعضی آزمونها، به ویژه آزمونهای استخدامی، اگر آزمون فاقد روایی صوری باشد، آزمون‌شونده ممکن است علاقه‌ای به جواب‌دادن به سؤالهای آزمون از خود نشان ندهد، زیرا ممکن است چنین تصور شود که آزمون به تصمیمهای مربوط به استخدام او ربطی ندارد. نتیجه

1. content domain
2. face validity

اینکه، می‌توان دید که روایی صوری، هر چند که ضامن اندازه‌گیری دقیق نیست، ممکن است بر انگیزش آزمون‌شونده و در نتیجه بر روایی نمرات حاصل از آزمون تأثیر داشته باشد.

بنا به گفتهٔ کاپلان و ساکوزو (۲۰۰۱)، فرض کنید آزمونی ساخته‌اید تا به وسیلهٔ آن از میان داوطلبان یک دورهٔ آموزش حسابداری مستعدترین را برگزینید. اگر سؤالهای شما دربارهٔ ترازنامه و دفتر حساب طرح شده باشد بیشتر موجب برانگیختگی آزمون‌شوندگان می‌شود تا سؤالهایی که در بارهٔ مصرف سوخت تهیه شده‌اند.

هر چند که روایی صوری یک ویژگی مطلوب آزمون است با این حال در بعضی آزمونها این ویژگی نه‌تنها ضروری نیست بلکه بهتر است که وجود نداشته باشد. برای مثال، "اگر کسی بخواهد بیماری روانی را در افراد تشخیص دهد، بهتر است آزمونی را به کار بندد که روایی صوری اندکی دارد" (دیک و هگرتی، ۱۹۷۱، ص ۹۵).

روش تعیین روایی صوری

از آنجاکه روایی صوری نوع خاصی از روایی محتوایی است، همان روشی که برای تعیین روایی محتوایی به کار می‌رود، یعنی استفاده از نظر متخصصان، در اینجا نیز برای تعیین روایی صوری قابل استفاده است.

روایی ملاکی

منظور از **روایی ملاکی**[1] میزان ارتباط بین نمرات حاصل از یک آزمون با نمرات حاصل از یک آزمون یا وسیلهٔ اندازه‌گیری دیگر است. بنابراین، هر زمان که از نمرات یک آزمون برای پیش‌بینی عملکرد افراد در آزمونی دیگر استفاده می‌کنیم با روایی ملاکی سر و کار داریم. آزمون دوم که عملکرد فرد در آن پیش‌بینی می‌شود **ملاک**[2] نام دارد. گاهی این پیش‌بینی برای آینده است، یعنی نمرات آزمونی که به صورت ملاک به کار خواهد رفت پس از گذشت یک فاصلهٔ زمانی از اجرای آزمون اول گردآوری خواهد شد که به آن **روایی پیش‌بینی**[3] گفته می‌شود؛ و زمانی این پیش‌بینی بدون فاصله و در زمان حال صورت می‌پذیرد که به آن **روایی همزمان**[4] می‌گویند. گرانلاند و لین (۱۹۹۰) در این باره گفته‌اند که استفاده از نمرات آزمون آمادگی خواندن به عنوان وسیله‌ای برای پیش‌بینی پیشرفت آتی دانش‌آموزان در خواندن موردی از رابطهٔ پیش‌بینی است، و کاربرد نمرات آزمون مهارتِ استفاده از فرهنگ لغات که به منظور تعیین توانایی فعلی دانش‌آموزان در استفادهٔ واقعی از فرهنگ لغات (که از طریق مشاهده سنجیده می‌شود) به کار می‌رود نمونه‌ای از رابطهٔ همزمان است.

---

1. criterion-related validity
2. criterion
3. predictive validity
4. concurrent validity

هر قدر رابطهٔ بین نمرات آزمونی که می‌خواهیم روایی ملاکی آن را بسنجیم با نمرات آزمون ملاک بیشتر باشد ضریب روایی ملاکی آن آزمون به همان نسبت بیشتر خواهد بود. زمانی که آزمون مورد نظر ما قرار است برای پیش‌بینی به کار برود رابطهٔ بین نمرات حاصل از آن آزمون و نمرات آزمونی که با فاصله‌ای در آینده اجرا خواهد شد تعیین می‌گردد. اما زمانی که می‌خواهیم از نمرات یک آزمون برای پیش‌بینی نمرات افراد در یک آزمون یا مقیاس اندازه‌گیری دیگر استفاده کنیم رابطهٔ همزمان بین دو دسته نمره را محاسبه می‌کنیم.

در گذشته این دو نوع روایی را تحت نامهای *روایی پیش‌بینی* و *روایی همزمان* مستقلاً مورد بحث و بررسی قرار می‌دادند، اما از سال ۱۹۸۵ به بعد که در کتاب *معیارهای آزمودن پرورشی و روانشناختی* (انجمن روانشناسی آمریکا، ۱۹۸۵) این دو نوع روایی زیر نام *روایی ملاکی* معرفی شدند، کتابهای معتبر روان‌سنجی نیز به همین صورت آنها را معرفی کرده‌اند. سبب عمدهٔ این کار روش مشابهی است که در هر دو نوع تعیین روایی به کار می‌رود. ما در زیر روایی پیش‌بینی و روایی همزمان را به عنوان دو بخش مکمل روایی ملاکی مورد بحث قرار می‌دهیم.

روایی پیش‌بینی

چنان‌که گفته شد، **روایی پیش‌بینی** ویژهٔ آزمونهایی است که برای پیش‌بینی موفقیت افراد در امور تحصیلی یا شغلی به کار می‌روند. آزمونهایی که برای مقاصد پیش‌بینی به کار می‌روند مختلف هستند. گاه از آزمونهای پیشرفت تحصیلی فعلی یا قبلی برای پیش‌بینی پیشرفت تحصیلی آینده استفاده می‌شود (مثلاً استفاده از آزمونهای ورودی دانشگاهها برای پیش‌بینی پیشرفت تحصیلی داوطلبان ورود به دانشگاهها در جریان تحصیل در دانشگاه، یا استفاده از نمرات پیشرفت تحصیلی دورهٔ راهنمایی تحصیلی برای پیش‌بینی پیشرفت تحصیلی در دبیرستان). زمانی هم از نمرات آزمونهای هوش و استعداد برای پیش‌بینی موفقیتهای بعدی تحصیلی و غیرتحصیلی استفاده می‌شود. گاهی هم از آزمونهای روانی و آزمونهای ویژهٔ انتخاب و جایگزینی افراد به پیش‌بینی موفقیت آنها در امور شغلی و مدیریت و سایر امور می‌پردازند. این نوع آزمونها باید از روایی پیش‌بینی برخوردار باشند.

روش تعیین روایی پیش‌بینی

چنان‌که قبلاً گفته شد، از آنجاکه در روایی پیش‌بینی همیشه نوعی ملاک مورد نظر است که نمرات آزمون پیش‌بینی‌کننده یا پیش‌بین را باید با آن مقایسه کرد، به آن روایی ملاکی نیز گفته می‌شود. "روایی ملاکی به فن تجربی مطالعهٔ رابطهٔ بین نمرات آزمون (پیش‌بینی‌کننده) و نوعی اندازهٔ بیرونی مستقل (ملاک) مربوط می‌شود" (مهرنز و لهمان، ۱۹۸۴، ص ۲۹۲). روایی ملاکی را

می‌توان با توجه به درجه‌ای که عملکرد آزمون‌شونده در یک آزمون (آزمون پیش‌بین) با عملکرد او در آزمون دیگر (آزمون ملاک) ارتباط دارد تعریف کرد. پس برای تعیین روایی پیش‌بینی یک آزمون، دو آزمون ضروری هستند. ضریب همبستگی بین نمرات حاصل از این دو آزمون شاخص روایی پیش‌بینی آزمون مورد نظر است. هر چه این ضریب بزرگ‌تر باشد، آزمون دارای روایی بیشتری خواهد بود، و هر چه این ضریب کوچکتر باشد آزمون مورد نظر روایی کمتری خواهد داشت.

بعضی وقتها تعیین ضریب همبستگی بین آزمون پیش‌بین و آزمون ملاک کار آسانی است، اما گاه به دست‌آوردن یک رابطهٔ همبستگی بین این دو آزمون کار ساده‌ای نیست. آنجاکه هر دو آزمون پیش‌بین و ملاک آزمون پیشرفت تحصیلی هستند یا آزمون پیش‌بین آزمون هوش یا استعداد و آزمون ملاک آزمون پیشرفت تحصیلی است محاسبهٔ ضریب همبستگی و در نتیجه تعیین ضریب روایی ساده‌تر است. مثلاً برای اینکه تعیین کنیم که یک آزمون ورودی دانشگاه دارای روایی پیش‌بینی هست یا نه، باید رابطهٔ بین نمرات داوطلبان ورود به دانشگاه را در این آزمون با نمرات آنان در درسهایی که بعداً در دانشگاه می‌گیرند تعیین کنیم و ضریب همبستگی بین آنها را به دست آوریم. در مورد قدرت نمرات دورهٔ راهنمایی تحصیلی برای پیش‌بینی موفقیت دانش‌آموزان در دورهٔ دبیرستان، و قدرت نمرات دورهٔ دبیرستان برای پیش‌بینی توفیق در دانشگاه نیز می‌توان همین کار را انجام داد.

با این حال، گاه تبدیل متغیری که باید پیش‌بینی شود به صورت عددی به سادگی ممکن نیست. برای مثال، اگر بخواهیم یک آزمون یا یک پرسشنامهٔ انتخاب معلمی بسازیم، باید یک وسیلهٔ اندازه‌گیری موفقیت در شغل معلمی نیز درست کنیم تا بتوانیم عملکرد فرد را در این دو نوع وسیلهٔ اندازه‌گیری با هم مقایسه نماییم، و بعد تعیین کنیم که آزمون یا پرسشنامهٔ انتخاب معلم روا هست یا نه. تعیین ملاک موفقیت معلمی مستلزم پژوهشهای وسیع و دقیق در حوزهٔ روانشناسی آموزش و یادگیری است که به رغم سالها تلاش هنوز ملاک دقیق و معتبری برای آن تعیین نشده است. در نتیجه در چنین اوضاع و احوالی باید از همهٔ امکانات موجود بهره گرفت و آنجا که یک ملاک دقیق و مورد پذیرش همگان موجود نباشد، می‌توان از ملاکهای مختلف استفاده کرد. مثلاً برای تعیین میزان موفقیت معلمان در فعالیت آموزشی هم باید از نظر مدیران، همکاران، دانش‌آموزان و دانشجویان استفاده کرد، و هم از میزان پیشرفت تحصیلی دانش‌آموزان و دانشجویان آن معلمان.

در تعیین روایی پیش‌بینی همچنان‌که می‌توان از چندین ملاک استفاده کرد، استفاده از چندین آزمون پیش‌بین نیز این نوع روایی را افزایش می‌دهد و بر قدرت پیش‌بینی می‌افزاید. مثلاً افزودن چند عامل پیش‌بینی دیگر به نمرات آزمون ورودی دانشگاهها می‌تواند قدرت پیش‌بینی آن آزمون را افزایش دهد. یعنی به عوض تکیه بر نمرات آزمون ورودی به عنوان یک عامل واحد پیش‌بینی، می‌توان از نمرات دبیرستانی داوطلبان و نمرات نوعی آزمون استعداد تحصیلی نیز سود جست.

ترکیب این سه نوع عامل به مراتب بیشتر از نمرات یک آزمون ورودی می‌تواند موفقیت آتی داوطلبان ورود به دانشگاه را پیش‌بینی کند، و در نتیجه انتخاب بهتری از داوطلبان به عمل آورد.

روایی همزمان

روایی همزمان نوعی روایی پیش‌بینی است که در آن، به جای تعیین رابطهٔ بین دو آزمون پس از گذشت یک فاصلهٔ زمانی، رابطهٔ بین دو آزمون به طور همزمان تعیین می‌شود. هدف از تعیین روایی همزمان بین دو آزمون این است که معلوم سازیم آیا می‌توان یکی از دو آزمون را به جای آزمون دیگر مورد استفاده قرار داد یا نه.

روش تعیین روایی همزمان

در تعیین روایی همزمان، هنگام اندازه‌گیری متغیر پیش‌بین، علاوه بر این متغیر، متغیر ملاک نیز موجود است، در حالی‌که در تعیین روایی پیش‌بینی، چنان‌که دیدیم، متغیر ملاک در آینده اندازه‌گیری می‌شود. به سخن دیگر، اگر به عوض مقایسهٔ نمرات یک آزمون با نمرات آزمون دیگری که در آینده به دست خواهند آمد، نمرات آزمون اول را با نمرات آزمون دیگری که همزمان با آزمون اول یا در فاصلهٔ کوتاهی بعد از آن اجرا می‌شود مقایسه کنیم، رابطهٔ حاصل بین این دو آزمون همزمان اجرا شده، نشان‌دهندهٔ روایی همزمان است.

فرض کنید برای تشخیص نوعی بیماری روانی پرسشنامه‌ای تهیه کرده‌ایم. حال برای تعیین روایی همزمان این پرسشنامه، نتایج حاصل از اجرای آن را با تعدادی از مراجعان به درمانگاه روانی با نظر متخصصان این بیماری که بلافاصله بعد از اجرای پرسشنامه یا اندک زمانی پس از آن در مورد همان افراد به دست آمده است مقایسه می‌کنیم. در این صورت به تعیین روایی همزمان اقدام کرده‌ایم. اگر رابطهٔ بین این دو نوع تشخیص رابطهٔ بالایی باشد گفته می‌شود که پرسشنامهٔ مورد نظر دارای روایی همزمان است.

مقایسهٔ روایی پیش‌بینی با روایی همزمان

چنان‌که گفتیم، متخصصان روانسنجی دو نوع روایی پیش‌بینی و همزمان را با اصطلاح *روایی ملاکی* معرفی می‌کنند و بین این دو نوع روایی تمایز چندانی قائل نمی‌شوند. اما، با وجود شباهت زیاد بین این دو نوع روایی، می‌توان بین آن دو نوعی تمایز منطقی قائل شد. گفته شد که روایی پیش‌بینی به موقعیتی اشاره می‌کند که در آن داده‌های مربوط به ملاک در آینده به دست می‌آیند. بنابراین، نمرات آزمون واقعاً برای انجام نوعی پیش‌بینی به کار می‌روند. اما چون در روایی همزمان داده‌های پیش‌بینی‌کننده و پیش‌بینی‌شونده همزمان جمع‌آوری می‌شوند، هدف آزمون پیش‌بین

پیش‌بینی ملاک نیست، زیرا داده‌های ملاک در زمان اجرای آزمون پیش‌بین وجود دارند، بلکه هدف آن جانشین‌ساختن آزمون پیش‌بین با آزمون ملاک است. برای مثال، هدف از تعیین روایی همزمان برای پرسشنامهٔ تشخیص بیماری روانی مورد اشاره در بالا ممکن است تعیین جواب برای این پرسش باشد که آیا می‌توان نمرات این پرسشنامه را که به وسیلهٔ دستیار روانپزشک اجرا می‌شود به جای معاینات روانپزشکی برای تشخیص نوع بیماری به کار برد یا نه؟ معلوم است که چنین جانشین‌سازی در صورتی مفید است که: (۱) بین نمرات آزمون و نمرات ملاک رابطهٔ بالایی وجود داشته باشد (در مثال بالا بین تشخیص حاصل از اجرای پرسشنامه و تشخیص روانپزشک) و (۲) استفاده از آزمون از لحاظ اقتصادی و زمانی از جمع‌آوری داده‌های مربوط به ملاک بیشتر به صرفه باشد. بنابراین، روایی همزمان از لحاظ منطقی و عملی با روایی پیش‌بینی تفاوت دارد. با این حال، از آنجا که این دو نوع روایی اساساً با روابط تجربی بین نمرات آزمون پیش‌بین و آزمون ملاک مربوط‌اند، می‌توان آنها را جنبه‌های مختلف روایی ملاکی دانست.

ویژگیهای مطلوب ملاک

ثرندایک و هیگن (۱۹۷۹) برای داده‌های ملاکی، چهار ویژگی مطلوب ذکر کرده‌اند. این ویژگیها به ترتیب اهمیت عبارت‌اند از: **ربط داشتن، بی‌طرفی، پایایی، و در دسترس بودن.**

ربط داشتن داده‌های یک آزمون ملاکی، با توجه به قابلیت انطباق آنها با موقعیت واقعی شخص مورد قضاوت قرار می‌گیرد. برای این کار هیچ‌گونه شواهد تجربی وجود ندارد و باید بر اساس ملاک منطقی تصمیم گرفت. برای مثال، در این مورد که آیا میانگین نمرات دانشجویان سال اول دانشگاه شاخص درستی از دستیابی آنان به هدفهای آموزشی درسهای مختلف هست یا نه هیچ گواه تجربی وجود ندارد. در این باره، بهترین کار آن است که به بهترین قضاوت موجود متخصصان اطمینان کنیم تا مطمئن شویم که محتوای آزمون هدفهای مهم آموزشی را اندازه‌گیری می‌کند یا نمی‌کند. به همین طریق، در رابطه با یک آزمون ملاک نیز لازم است که به قضاوت تخصصی افراد اطمینان کنیم تا معلوم شود که یک آزمون ملاک تا چه حد به ملاک غایی موفقیت در یک برنامهٔ آموزشی یا در یک شغل مربوط است.

مهرنز و لهمان (۱۹۸۴) می‌گویند میزان ربط داشتن اندازهٔ ملاک یک داوری ارزشی است و ممکن است در این باره میان افراد مختلف اختلاف نظر وجود داشته باشد. برای مثال، بعضی مربیان معتقدند که موفقیت در دانشکده عبارت است از میزان دانش کسب‌شده در طول چهار سال تحصیل در دانشگاه که به وسیلهٔ نمرات دانشگاهی نشان داده می‌شود. کسان دیگری می‌گویند که میزان دانش کسب‌شده ملاک درستی از موفقیت دانشگاهی است، اما نمرات دانشگاهی را شاخص خوبی برای آن نمی‌دانند. حتی گروهی دیگر معتقدند که موفقیت در دانشگاه عبارت است از

ازدواج خوب، روابط اجتماعی خوب، شغل خوب، و از این قبیل. لذا برای این افراد نیز نمرات درسی دانشگاه ملاک موفقیت به حساب نمی‌آید.

۳ منظور از بی‌طرفی این است که آزمون باید به گونه‌ای باشد که همهٔ افراد در کسب نمرهٔ خوب فرصت مساوی داشته باشند. مثالهای مربوط به عدم بی‌طرفی، تفاوت بین تجهیزات و شرایط کار برای کارگران یک کارخانه یا تفاوت بین کیفیت آموزش داده شده به دانش‌آموزان مختلف است. تا آنجا که در ارزشیابی از کار یا کیفیت آموزش نمرهٔ حاصل به عوامل مربوط به شرایط کار مربوط می‌شود تا به شایستگی کارگر یا به کیفیت آموزش مربوط می‌شود تا به شایستگی دانش‌آموز، همبستگی بین نتایج آزمون و نمرهٔ ملاک مفهومی ندارد. همچنین، بی‌طرفی زمانی نقض می‌شود که نمرهٔ ملاک تحت تأثیر دانش مربوط به نمرهٔ پیش‌بینی قرار می‌گیرد. فرض کنید در مهر ماه سال تحصیلی معلم ریاضی یک کلاس امتحانی از دانش‌آموزان بگیرد و نمرهٔ آن را برای پیش‌بینی موفقیت دانش‌آموزان در پایان همان سال مورد استفاده قرار دهد. اگر اطلاع او از این نمرات آغاز سال دانش‌آموزان به طور آگاهانه یا ناآگاهانه نمراتی را که او در پایان سال به دانش‌آموزان می‌دهد تحت تأثیر قرار دهد در این صورت ملاک مورد نظر دیگر خالی از بی‌طرفی نیست. در چنین حالتی، بهترین راه جلوگیری از مشکل این است که تهیه کنندهٔ نمرات ملاک از نمرات پیش‌بینی هیچ‌گونه اطلاعی نداشته باشد.

در بارهٔ پایایی در فصل بعدی این کتاب به تفصیل بحث خواهد شد. به طور خلاصه، یک آزمون یا یک وسیلهٔ اندازه‌گیری پایا مربوط به موفقیت شغلی باید باثبات و قابل تکرار باشد. اگر نتایج حاصل از روزی به روز دیگر تغییر کند، به طوری که شخص از روزی به روز دیگر یا از هفته‌ای به هفتهٔ دیگر نمرات متفاوتی بگیرد، در این صورت هیچ آزمونی وجود نخواهد داشت که نمرهٔ ملاک را پیش‌بینی کند. نمره‌ای که از هیچ ثباتی برخوردار نباشد با هیچ وسیله‌ای قابل پیش‌بینی نیست.

۴ در دسترس‌بودن به این معنی است که دستیابی به اندازهٔ ملاک عملی و میسر باشد. دربارهٔ این ویژگی با توجه به زمان، هزینه، و امکانات مورد نیاز قضاوت می‌شود. بنابراین، ملاکی که به سهولت بیشتر در چهارچوب امکانات سهلتر قابل دسترس باشد ملاک بهتری است.

روایی سازه

روایی سازه[1] بیشتر از روایی محتوایی و روایی پیش‌بینی جنبهٔ نظری دارد. بنا به تعریف، یک آزمون در صورتی دارای روایی سازه است که نمرات حاصل از اجرای آن به مفاهیم یا سازه‌های نظریهٔ

1. construct validity

مورد نظر مربوط باشند (کرانباخ[1]، ۱۹۷۰). برای مثال، یک آزمون یا پرسشنامهٔ اضطراب در صورتی دارای روایی سازه است که نمرات حاصل از آن به سازه‌هایی که در نظریه‌های اضطراب آمده‌اند ارتباط داشته باشد.

معنی *سازه* به معنی مفهوم نزدیک است. کرلینجر[2] (۱۹۶۵) سازه را این گونه تعریف کرده است: "سازه یک مفهوم است، اما سازه یک معنی اضافی بر مفهوم دارد و آن این است که سازه برای مقاصد ویژهٔ علمی به طور عمد و از روی آگاهی ابداع می‌شود" (ص ۳۲). سالکیند[3] (۲۰۰۵) گفته است یک سازه به یک دسته متغیر وابسته به هم گفته می‌شود. مثلاً *پرخاشگری* یک سازه است که از متغیرهایی مانند لمس کردن دیگران به طور نادرست، ابراز خشونت، فقدان ارتباط اجتماعی موفقیت‌آمیز، و مانند اینها تشکیل می‌شود. و البته سازه‌ها همواره به باورهای نظری پژوهشگر وابسته‌اند. "مثلاً اینکه مردان پرخاشگر بیشتر از مردان غیرپرخاشگر با مقامات درگیر می‌شوند" (ص ۲۹۶).

همهٔ سازه‌های علمی دارای دو ویژگی مهم هستند: (۱) خلاصه‌ای انتزاعی از پاره‌ای نظام‌مندی‌های طبیعت‌اند، و (۲) به رویدادهای عینی و قابل مشاهده مربوط‌اند. مورفی و دیوید شوفر (۲۰۰۱) نیروی گرانش را به عنوان یک سازهٔ علمی مثال آورده‌اند. وقتی که میوه‌ای از درخت می‌افتد، نیروی جاذبه یا نیروی گرانش برای پیش‌بینی و تبیین رفتار آن شیء به کار می‌رود. نیروی گرانش قابل دیدن نیست. تنها چیزی که دیده می‌شود افتادن میوه از درخت است. "با وجود این، کاملاً منطقی است که نیروی گرانش اندازه‌گیری و دربارهٔ آن نظریه‌پردازی شود" (مورفی و دیوید شوفر، ۲۰۰۱، ص ۱۵۵).

مثالهای سازه در روانشناسی و آموزش و پرورش فراوان‌اند. از جمله می‌توان *هوش*، *انگیزش*، *اضطراب*، *برون‌گرایی*، *درون‌گرایی*، و جز اینها را نام برد.

روش تعیین روایی سازه

برای تعیین روایی سازه نخست باید سازه یا متغیر مورد نظر را تعریف کرد. این متغیر به صورت یک سازه در نظامی از مفاهیم قرار خواهد داشت که در آن نظام این سازه، به طریقی منطقی، با سازه‌های دیگر در رابطه خواهد بود. این روابط در نظریهٔ بخصوصی بیان شده‌اند. پس از تعریف‌کردن سازه، برای اندازه‌گیری آن، آزمونی تهیه می‌شود. سپس، با توجه به پیش‌بینی‌های نظریه، نمرات حاصل از این آزمون باید با متغیرهای دیگر مورد بحث در نظریه روابط پیش‌بینی‌شده را نشان دهند. اگر وجود این روابط نشان داده شد، آزمون مورد نظر دارای روایی سازه است.

برای مثال، سازهٔ هوش گرچه یک مفهوم قابل مشاهده نیست، نظریه‌های روانشناختی و پرورشی مطالب زیادی در بارهٔ آن بیان کرده‌اند. مثلاً اینکه افراد باهوش در یادگیری سریع‌اند، یا

1. Cronbach 2. Kerlinger 3. Salkind

هوش با پیشرفت تحصیلی رابطهٔ مستقیم دارد، و از این قبیل. بنابراین، متخصص هوش‌سنجی، پس از اینکه تعریفی از هوش به دست داد، برای آن یک آزمون می‌سازد و رابطهٔ بین نمرات حاصل از این آزمون را با سرعت یادگیری، میزان پیشرفت تحصیلی، و غیره تعیین می‌کند. چنانچه نتایج حاصل پیش‌بینی‌های نظریهٔ هوشی را تأیید کردند، گفته می‌شود که آزمون تهیه‌شده یک آزمون هوشی رواست. برای تعیین روایی سازه می‌توان راههای مختلف زیر را مورد استفاده قرار داد:

۱. تعیین همبستگی آزمون با سایر آزمونهای روا. یکی از راههای تعیین روایی سازه این است که همبستگی بین نمرات آزمون مورد نظر را با یک آزمون موجود دارای روایی که برای همان سازه ساخته شده است تعیین کنیم. اگر آزمون جدید با آزمون موجود همبستگی زیادی داشته باشد، فرض می‌شود که این دو آزمون سازهٔ واحدی را اندازه می‌گیرند و در نتیجه آزمون جدید نیز برای سازهٔ مورد نظر یک آزمون رواست. به این نوع روایی، روایی **همگرا**[1] می‌گویند. از سوی دیگر، آزمون مورد نظر باید با آزمونهای دیگری که سازه‌های متفاوتی را اندازه‌گیری می‌کنند همبستگی نداشته باشد. این گونه روایی را روایی **واگرا**[2] یا روایی **تمیزی**[3] می‌نامند.

۲. تمایز سنی. راه دیگر تعیین روایی سازه توجه به تمایز سنی است. اگر یکی از ویژگیهای انسان با بالارفتن سن پیشرفت نشان دهد، نمرات آزمونی که برای اندازه‌گیری این ویژگی درست شده است باید منعکس‌کنندهٔ این مطلب باشد. برای مثال، عملکرد کودکان خردسال در آزمونهای هوشی مخصوص کودکان ضعیف است، اما با بالارفتن سن به سؤالهای بیشتری جواب می‌دهند. آزمونهای هوشی وکسلر و بینه این مورد را نشان می‌دهند. البته اگر سن یک عامل مهم ویژگی یا رفتار مورد اندازه‌گیری نباشد، در آن صورت در تعیین روایی سازه تمایزی را نباید در نظر گرفت.

۳. تحلیل عاملی[4]. راه دیگر تعیین روایی سازه استفاده از روش تحلیل عاملی است. تحلیل عاملی یا تحلیل عوامل یک روش آماری است که از طریق آن تعداد و ماهیت متغیرهایی را که یک آزمون اندازه می‌گیرد مشخص می‌کنند. در تحلیل عامل، این کار از طریق ادغام تعداد زیادی از ماده‌ها یا متغیرها با همدیگر و ایجاد تعداد معدودی متغیر قابل بررسی به نام عامل انجام می‌شود. این روش روابط درونی میان داده‌های مورد تحلیل را تعیین می‌کند و برای ایجاد سهولت متغیرها را به خوشه‌ها یا عوامل کاهش می‌دهد. بعد این خوشه‌ها یا عوامل را با توجه به آنچه به وسیلهٔ سؤالهای آزمون اندازه‌گیری می‌شوند نام‌گذاری می‌کنند. بنابراین، یک آزمون را می‌توان بر حسب عواملی که آن را تشکیل می‌دهند نام‌گذاری کرد. مثلاً اگر یک آزمون برای اندازه‌گیری سازهٔ آفرینندگی درست شده است، ترکیب عاملی آن آزمون باید نشان‌دهندهٔ عوامل آفرینندگی در آن آزمون باشد.

---

1. convergent
2. divergent
3. discriminative
4. factor analysis

۴. همسانی درونی[1]. آخرین روش تعیین روایی سازه که در اینجا معرفی می‌شود همسانی درونی است. در این روش، ساخت درونی آزمون بررسی می‌شود. ملاک مورد استفاده برای بررسی همسانی درونی نمرهٔ کل آزمون است. روش معمول برای این کار این است که همبستگی بین نمرات آزمونهای فرعی یا خرده‌آزمونها با نمرهٔ کل آزمون محاسبه می‌شود، و هر یک از این خرده‌آزمونها که با نمرهٔ کل آزمون همبستگی اندکی را نشان داد از آزمون حذف می‌شود. از لحاظ نظری، نمرات خرده‌آزمونها باید با نمرهٔ کل آزمون همبستگی زیادی داشته باشند، زیرا فرض این است که همهٔ آنها بر روی هم حوزهٔ رفتاری واحدی را اندازه‌گیری می‌کنند.

در ضمن، خرده‌آزمونها نباید با یکدیگر همبستگی زیادی داشته باشند، زیرا در چنین حالتی فرض می‌شود که همهٔ آنها یک چیز را اندازه می‌گیرند و لذا تکراری هستند. در این روش، به جای خرده‌آزمونها، از ماده‌ها یا سؤالات آزمون نیز می‌توان استفاده کرد. در این صورت، همبستگی میان ماده‌های آزمون با کل آزمون و با یکدیگر محاسبه می‌شود.

۵. جمعیتهای متفاوت[2]. یک روش نسبتاً سادهٔ تعیین روایی سازه به وسیلهٔ پوفام (۲۰۰۴) به عنوان روش جمعیتهای متفاوت پیشنهاد شده است. در این روش، دو گروه نمونه انتخاب می‌شوند به گونه‌ای که یک گروه دارای ویژگی مورد اندازه‌گیری و گروه دیگر بدون آن ویژگی باشند. آنگاه ابزار ساخته شده بر روی دو گروه اجرا می‌شود. اگر نتایج حاصل نشان دادند که از لحاظ ویژگی یا سازه‌ای که اندازه‌گیری شده است دو گروه با هم فرق دارند، پس ابزار موردنظر دارای روایی سازه است.

برای روشن شدن چگونگی استفاده از روش فوق، فرض کنید پرسشنامه‌ای ساخته‌اید که سازهٔ افسردگی را می‌سنجد. پرسشنامهٔ موردنظر را بر روی دو گروه که یکی از آنها به طور قطع دارای افسردگی است (مثلاً به تشخیص روانپزشکان یک بیمارستان روانی) ولی گروه دیگر مبتلا به افسردگی نیست اجرا کنید. نتایج حاصل از اجرای ابزار شما بر روی دو گروه باید به روشنی تفاوت بین آن دو گروه را نشان دهد. یعنی نمرات پرسشنامهٔ افسردگی گروه افسرده باید به طور معنی‌داری بیشتر از نمرات گروهی باشد که افسرده نیست.

معنای عشق: مثالی از تعیین روایی سازه[3]

مورد جالبی از تعیین روایی سازه کوششهای زیک روبین[4] (۱۹۷۰، ۱۹۷۳) در رابطه با ساختن مقیاسی برای اندازه‌گیری عشق است. عشق از مفاهیمی است که در طول تاریخ همواره مورد بحث انسانها بوده، اما هر کسی از آن برداشت خاصی داشته و هیچ‌کس آن را به طور دقیق معنی نکرده

1. internal consistency 2. differential populations

۳. مطالب این قسمت از کتاب **روان‌آزمایی: اصول، کاربردها، و مسائل**، تألیف رابرت. ام. کاپلان و دنیس پی. ساکوزو (۲۰۰۱) خلاصه شده‌اند.

4. Zick Rubin

است. کار چشم‌گیر روبین این بود که ضمن تهیهٔ یک مقیاس اندازه‌گیری برای هیجان عشق، هم ابعاد آن را شناسایی کرد و هم آن را از سایر احساسهای مشابه متمایز ساخت.

روبین ابتدا فهرستی از همهٔ موردهایی را که افراد مختلف عشق می‌نامند تهیه کرد. برای اینکه معانی و تفسیرهای کلی و مبهم را اندازه‌پذیر کند، لازم بود آنها را در قالب جملاتی درآورد که پاسخ‌دهندگان بتوانند به آنها پاسخ دهند. در نهایت جمله ها و عبارتهای حاصل را به صورت ماده‌های یک پرسشنامه به گونه‌ای تنظیم کرد که پاسخ‌دهندگان توانستند براساس یک مقیاس پنج درجه‌ای (۵ کاملاً موافق تا ۱ کاملاً مخالف) به آنها پاسخ دهند.

روبین، برای دست‌یافتن به بیاناتی که در برگیرندهٔ معانی متنوع موردنظر افراد مختلف از عشق است، به مطالعات گسترده‌ای در این باره اقدام کرد، زیرا لازم بود که تمام تعریفهای گوناگونی را که از عشق به دست آمده بود در مجموعهٔ بیانات خود بگنجاند. گردآوری مجموعه‌ای ماده یا جمله برای تعیین روایی سازه کار آسانی نیست، زیرا معلوم نیست که آخر سر کدام ماده‌ها در وسیلهٔ اندازه‌گیری موردنظر گنجانیده خواهند شد. شواهد مربوط به روایی سازه مستلزم شواهد مربوط به روایی محتوایی است. شواهد مربوط به روایی محتوایی نیز به نوبهٔ خود نیازمند این است که ماده‌های ابزار اندازه‌گیری به طور کامل مُعرف حیطه یا حوزهٔ مورد استنباط (در اینجا عشق) باشند. بنابراین، تمام تعریفهای مربوط به عشق باید به وسیلهٔ روبین جمع‌آوری می‌شدند.

سرانجام روبین مجموعه بیاناتی را که معرف معناهای مختلف عشق بودند فراهم آورد. بعضی از آنها به قصد سنجیدن عشق[1] مورد استفاده قرار گرفتند، در حالی که هدف بعضی دیگر از آنها سنجش **دوست داشتن** یا *خوش آمدن*[2] بود. پس از آماده کردن مجموعه بیانات مختلف، روبین آنها را در اختیار ۱۹۸ دانشجوی دانشگاه میشیگان قرار داد. در مقابل هریک از جمله‌ها یک جای خالی برای نوشتن یک نام منظور شده بود. دانشجویان دوبار به پرسشها پاسخ دادند؛ یک بار نام فردی را که به او عشق می‌ورزیدند نوشتند و بار دوم نام یک دوست را وارد کردند. سپس بیانات یا ماده‌ها با روش تحلیل عاملی تحلیل شدند.

نتایج تحلیل عاملی در مقیاس عشق سه عامل زیر را به دست داد: *وابستگی*[3]، *مراقبت*[4]، و صمیمیت[5]. ماده‌ها یا سؤالهایی که عامل وابستگی را تشکیل می‌دادند بر میلِ بودن با فرد مورد علاقه یا خواستن او به هنگام احساس تنهایی تأکید می‌کردند. عامل مراقبت ماده‌های مربوط به احساس دلواپسی و یگانگی را شامل می‌شدند. و عامل صمیمیت جنبه‌های اختصاصی رابطه با فرد مورد علاقه، مثلاً در میان گذاشتن مسائل و مشکلات شخصی و اعتماد کردن به او را، دربر می‌گرفت. ماده‌های مقیاس *خوش آمدن* یا **دوست داشتن** بر جنبه‌های مثبت شخص موردنظر تأکید می‌کردند و مواردی چون سازگاری، بلوغ، قضاوت درست، و هوش او را شامل می‌شدند.

---

1. love
2. liking
3. attachment
4. caring
5. intimacy

داده‌های حاصل از این مقیاسها با چند روش آماری تحلیل شدند تا پاسخهای *عاشقها*[1] و *دوستدارها*[2] از هم متمایز گردند. نتیجه ایجاد دو مقیاس مستقل عشق و دوستی بود. حال با در دست داشتن این دو ابزار با نامهای مقیاس عشق[3] و مقیاس دوستی[4]، وظیفهٔ روبین این بود که نشان دهد که این دو مقیاس واقعاً آنچه را که قرار بوده است بسنجند (یعنی عشق و دوستی) می‌سنجند.

یکی از تحقیقاتی که با استفاده از مقیاسهای عشق و دوستی روبین انجام گرفت و در آن زوجهایی که با هم رابطه داشتند مورد مطالعه قرار گرفتند نشان داد که دوست داشتن (یا خوش آمدن) و عشق ورزیدن یا دلبسته بودن الزاماً به هم مربوط نیستند. بین نمرات حاصل از این دو مقیاس رابطهٔ اندکی مشاهده شد که برای زنان ضعیف‌تر از مردان بود. نتایج این پژوهش حاکی از آن است که ما انسانها، به ویژه زنان، می‌توانیم عاشق کسی بشویم ولی الزاماً او را دوست نداشته باشیم.

چند چیز نشان دادند که مقیاس عشق واقعاً «عشق» را اندازه می‌گرفت. از جمله اینکه، زنان و مردانی که به سؤالهای پرسشنامه جواب می‌دادند وقتی که نام فرد غیرهمجنسی را که به او عشق می‌روزیدند می‌نوشتند بیشتر از زمانی که نام دوست همجنس را می‌نوشتند نمره می‌آوردند (همهٔ آزمودنیها غیرهمجنس‌خواه بودند). همچنین بین نمرات مقیاس عشق و احتمال ازدواج همبستگی مثبت و قابل توجهی یافت شد.

در اقدامی دیگر، بعضی از زوجهایی که با هم رابطه داشتند به دو گروه "دلبستگان قوی" (افراد دارای نمره‌های بالا در مقیاس عشق) و "دلبستگان ضعیف" (کسانی که نمره‌های آنان در مقیاس عشق پایین بود) تقسیم شدند. پژوهشگران، از پشت یک آئینهٔ یک‌طرفه، تماس چشمی بین زوجهای مختلف را مشاهده کردند. معلوم شد که دلبستگان قوی یا کسانی که در مقیاس عشق نمرهٔ بالایی کسب کرده بودند از دلبستگان ضعیف یا کسانی که نمرهٔ آنان در مقیاس عشق پایین‌تر بود وقت بیشتری صرف خیره شدن به چشمهای یکدیگر می‌کردند. وقتی که فردی از هریک از زوجهای مورد مطالعه با فرد دیگری به غیر از زوج خود (اما غیرهمجنس) روبه‌رو می‌شد، میزان تماس چشمی بین دلبستگان قوی بیشتر از تماس چشمی دلبستگان ضعیف نبود.

کوتاه سخن اینکه، روبین کوششهای خود را زمانی آغاز کرد که نه تعریف روشنی از عشق در اختیار داشت و نه روشی برای اندازه گیری آن موجود بود. او پس از مطالعات گسترده و کوششهای سازمان یافته به تصور بهتری از سازهٔ عشق دست یافت و برای آن یک ابزار اندازه گیری درست کرد. او طی بررسیهای خود کشف کرد که عاشقان یا افراد دلبسته به یکدیگر بعضی از ماده‌های مقیاس اندازه گیری عشق را از کسانی که صرفاً دوست همدیگر هستند متفاوت جواب می‌دهند. بنابراین، بین دوستی و دلبستگی (یا عشق) تمایز قائل شد. او همچنین دریافت که سازهٔ عشق دست کم از سه عامل وابستگی، مراقبت، و صمیمیت تشکیل یافته است.

---

1. lovers
2. friends
3. love scale
4. liking scale

مقایسهٔ ویژگیهای رواییهای مختلف با یکدیگر

حال که با انواع رواییها آشنا شدیم، ویژگیهای مختلف آنها را با هم مقایسه می‌کنیم. جدول ۱ ـ ۲۰ ویژگیهای انواع رواییهایی را که در این فصل مورد بحث قرار دادیم نشان می‌دهد. در این جدول، رواییهای مختلف هم از لحاظ قصد و منظور و هم از جهت روش تعیین روایی توضیح داده شده‌اند.

روایی آزمونهای وابسته به ملاک

اصطلاحات *روایی محتوایی*، *روایی پیش‌بینی*، و *روایی سازه* ابتدا تنها برای آزمونهای وابسته به هنجار به کار می‌رفتند، اما کاربرد آنها در مورد آزمونهای وابسته به ملاک نیز کاملاً امکان‌پذیر است. با وجود این، شواهد مربوط به روایی آزمونهای وابسته به هنجار و آزمونهای وابسته به ملاک قدری متفاوت‌اند، زیرا مقاصدی که این دو نوع آزمون برآورده می‌سازند با هم تفاوت دارند. ما در قسمت باقیماندهٔ این فصل مطالب مربوط به روایی آزمونهای وابسته به ملاک را بحث می‌کنیم.

جدول ۱ ـ ۲۰ خلاصهٔ ویژگیهای رواییهای مختلف

نوع روایی		قصد و منظور	چگونگی تعیین روایی یا روایی‌یابی
محتوایی		تعیین اینکه تا چه حد سؤالهای یک آزمون معرف محتوایی که برای سنجش آن تهیه شده‌اند هستند.	از راه تعیین میزان مطابقت بین سؤالهای آزمون و محتوای مورد نظر.
ملاکی	پیش‌بینی	تعیین این موضوع که نمرات یک آزمون تا چه اندازه نمرات آزمون دیگری را که آزمون ملاک نام دارد پیش‌بینی می‌کنند.	ضریب همبستگی بین نمرات آزمونی که تعیین روایی آن مورد نظر است و نمرات ملاک که بعد از گذشت مدتی از اجرای آزمون پیش‌بین به دست می‌آیند.
	همزمان	تعیین این موضوع که یک آزمون یا وسیلهٔ اندازه‌گیری جانشین مناسبی برای آزمون یا وسیلهٔ اندازه‌گیری دیگری هست یا نه.	ضریب همبستگی بین نمرات آزمونی که تعیین روایی آن مورد نظر است و نمرات ملاک که همزمان به دست می‌آیند.
سازه		تعیین تعداد و ماهیت ویژگیها یا سازه‌هایی که زیربنای مجموعه‌ای نمرهٔ آزمون یا خرده‌آزمون را تشکیل می‌دهند.	(۱) تعیین همبستگی آزمون با سایر آزمونهای روا (۲) تمایز سنی (۳) تحلیل عوامل (۴) همسانی درونی

روایی محتوایی آزمونهای وابسته به ملاک

روایی محتوایی آزمونهای وابسته به هنجار و آزمونهای وابسته به ملاک بسیار شبیه به هم هستند. با وجود این، در تعیین روایی برای آزمونهای وابسته به ملاک وابستگی سؤالهای آزمون به محتوا و هدفهای مشخص یک موضوع بیشتر است. در تعیین روایی آزمونهای وابسته به هنجار، چون هدفهای دقیق آموزشی مورد نظر نیستند، لذا کوشش آزمون‌ساز صرفاً به نمونه‌گیری از محتوای مورد سنجش منحصر می‌شود، در حالی که چون در آزمونهای وابسته به ملاک کاربرد هدفهای دقیق آموزشی ضرورت دارد، این آزمونها در صورتی روا هستند که هدفهای آموزشی درس یا موضوع مورد نظر را به دقت بسنجند.

با توجه به توضیحات بالا، چنان‌که در فصل ششم دیدیم، هر چند که برای ساختن آزمونهای وابسته به هنجار و وابسته به ملاک که باید دارای روایی محتوایی باشند کاربرد جدول مشخصات ضروری است، اما تفاوت عمدهٔ بین این دو نوع روایی آن است که در جدول مشخصاتی که برای تهیهٔ آزمونهای وابسته به هنجار ساخته می‌شود تنها سطوح طبقات یادگیری یعنی دانش، فهمیدن، کاربستن، و غیره منظور می‌گردد، اما در جدول مشخصاتی که برای تهیهٔ آزمونهای وابسته به ملاک تهیه می‌شود هم این سطوح یادگیری و هم خود هدفهای دقیق آموزشی باید در نظر گرفته شوند.

روش تعیین روایی محتوایی آزمونهای وابسته به ملاک

برای تعیین روایی محتوایی آزمونهای وابسته به ملاک بهتر است آزمون‌ساز از گروهی از داوران صاحب‌نظر در موضوع بخواهد تا سؤالهای آزمون را بخوانند و آنها را با توجه به میزان دقتی که هدفهای آموزشی و محتوای مورد نظر را می‌سنجند درجه‌بندی نمایند. لازم نیست که تعداد داوران خیلی زیاد باشد، اما لازم است که آنها هم از موضوع مورد سنجش و هم از کسانی که آزمون برایشان ساخته می‌شود آگاهی کامل داشته باشند. اعضای این گروه داوران می‌توانند معلمان همان موضوع درسیِ همان سطح آموزشی باشند.

ویرسما و جورس (۱۹۹۰) گفته‌اند وقتی که از داوران خواسته می‌شود تا سؤالها را درجه‌بندی کنند، بهتر است از یک فرم مشخص و هماهنگ استفاده شود که بتوان نتایج داوری آنان را با هم مقایسه کرد یا با هم جمع نمود و میانگین‌گیری به عمل آورد. مثالی از یک فرم سادهٔ مقیاس درجه‌بندی برای تعیین روایی محتوایی سؤالهای آزمونهای ملاکی در شکل ۱ ـ ۲۰ داده شده است.

به هنگام استفاده از گروه داوران برای درجه‌بندی سؤالهای آزمون به منظور سنجش روایی محتوایی آزمون، به هر یک از اعضاء نسخه‌ای از آزمون و نسخه‌ای از فرم درجه‌بندی داده می‌شود و از آنان خواسته می‌شود تا میزان مطابقت هر سؤال را با هدفی که از آن استخراج شده است تعیین نمایند. نتایج حاصل از درجه‌بندی افراد مختلف را می‌توان به گونه‌های متفاوت خلاصه کرد تا

هدف آموزشی:

لطفاً سؤالهایی را که برای هدف بالا نوشته شده‌اند با توجه به میزان دقتی که این هدف را می‌سنجند درجه‌بندی کنید.

شمارهٔ سؤال	ضعیف			عالی
۱	۱	۲	۳	۴	۵
۲	۱	۲	۳	۴	۵
۳	۱	۲	۳	۴	۵
۴	۱	۲	۳	۴	۵
۵	۱	۲	۳	۴	۵
۶	۱	۲	۳	۴	۵
۷	۱	۲	۳	۴	۵
۸	۱	۲	۳	۴	۵

در زیر نظر خود را نسبت به سؤالها و هرگونه پیشنهادی که دارید بنویسید:

شکل ۱ ـ ۲۰ فرم درجه‌بندی برای سنجش روایی محتوایی آزمونهای وابسته به ملاک (ویرسما و جورس، ۱۹۹۰)

برای تحلیل و مقایسه قابل استفاده باشند. نمونه‌ای از این کار در جدول ۲ ـ ۲۰ نشان داده شده است. اعداد جدول ۲ ـ ۲۰ درجه‌بندیهایی هستند که از شش داور مختلف به دست آمده‌اند. دقت کنید که سؤال شمارهٔ ۲ از بقیهٔ سؤالها ضعیف‌تر ارزشیابی شده است و احتمالاً نیاز به تجدید نظر دارد. این نتایج همچنین برای تعیین میزان مطابقت بین داوران مختلف قابل استفاده است.

کاملاً آشکار است که هر چه ویژگیهای موضوع مورد آزمون (اجزای محتوایی و هدفهای آموزشی) دقیق‌تر باشند، بر سر اینکه یک سؤال هدف معینی را به خوبی می‌سنجد یا نه میان داوران توافق بیشتری وجود خواهد داشت. اگر داوران در بارهٔ روایی محتوایی آزمون تردید داشته باشند، آن‌گاه لازم است که سؤالهای آزمون مورد تجدید نظر قرار گیرند.

ما در صفحات پیشین این فصل در رابطه با تعیین روایی محتوایی گفتیم که هیچ گونه ضریب آماری برای روایی محتوایی وجود ندارد، اما در سطور بالا (به ویژه در جدول ۲ ـ ۲۰) به نوعی شاخص آماری اشاره کردیم. باید توجه داشت که آنچه در جدول ۲ ـ ۲۰ آمده است شاخص درجهٔ توافق میان داوران است، نه مشخص‌کنندهٔ روایی محتوایی آزمون. بنا به گفتهٔ مرفی و دیویدشوفر (۱۹۹۴) تاکنون هیچ گونه شاخص آماری که نشان‌دهندهٔ روایی محتوایی باشد به دست نیامده است.

جدول ۲ ـ ۲۰ شکل خلاصهٔ درجه‌بندیهای مربوط به روایی محتوایی برای همهٔ داوران (ویرسما و جورس، ۱۹۹۰).

داوران	سؤالات ازمون ۱	۲	۳	۴	۵	۶	۷	۸	میانگین درجه‌بندی داوران
۱	۵	۳	۵	۵	۴	۵	۴	۵	۴/۵۰
۲	۵	۳	۴	۵	۵	۵	۵	۵	۴/۶۳
۳	۴	۵	۳	۵	۵	۴	۳	۵	۴/۲۵
۴	۴	۲	۴	۵	۴	۵	۴	۵	۴/۱۳
۵	۴	۳	۵	۵	۵	۵	۵	۴	۴/۵۰
۶	۵	۳	۵	۵	۵	۵	۵	۵	۴/۵۰
میانگین درجه‌بندی سؤالها	۴/۵	۳/۱۶	۴/۳۳	۵	۴/۶۷	۴/۸۳	۴/۳۳	۴/۸۳	

روش لاشه برای تعیین ضریب روایی محتوایی

یکی از روشهای تبدیل قضاوت کیفی داوران دربارهٔ روایی محتوایی به کمیّت به روش لاشه[1] شهرت دارد و به آن **ضریب روایی محتوایی**[2] (سی وی آر) گفته می‌شود (شولتز و ویتنی، ۲۰۰۵). برای محاسبهٔ این ضریب روایی محتوایی از تعدادی متخصص موضوع خواسته می‌شود تا هریک از سؤالهای آزمون را بررسی کنند و دربارهٔ آن نظر بدهند. در این روش، از داوران خواسته می‌شود تا هر سؤال را در رابطه با محتوایی که سؤال از آن برگرفته شده به سه شکل "ضروری"، "مفید"، یا "غیرضروری" مورد داوری قرار دهند. نظر داوران به کمک فرمول زیر به صورت ضریب روایی محتوایی محاسبه می‌شود.

$$CVR = \frac{n_e - \frac{N}{۲}}{\frac{N}{۲}}$$

n_e = تعداد داورانی که سؤال را ضروری تشخیص می‌دهند

N = تعداد کل داوران

ضریب روایی محاسبه شده با فرمول بالا برای هر سؤال دارای دامنه‌ای از ۱+ تا ۱- است که طبق آن هرچه رقم بزرگ‌تر باشد روایی سؤال بیشتر است. ضریب صفر (۰) نشان‌دهندهٔ این است که نیمی از داوران سؤال را ضروری تشخیص داده‌اند، و ضریب ۱- حاکی از آن است که از نظر همهٔ داوران سؤال غیرضروری است و باید کنار گذاشته شود.

1. Lawshe
2. Content Validity Ratio (CVR)

برای نشان دادن چگونگی محاسبهٔ ضریب روایی محتوایی پیشنهادی لاشه، یک آزمون ۳۰ سؤالی را که برای اندازه‌گیری دانش محتوایی یکی از مشاغل ساخته شده است درنظر می‌گیریم. سازندهٔ آزمون از ۱۲ متخصص آگاه از ویژگیهای شغل موردنظر خواسته است تا هر سؤال را با توجه به مقیاس سه‌درجه‌ای ضروری (۲)، مفید (۱)، و غیرضروری (۰) درجه‌بندی کنند. فرض می‌کنیم یکی از سؤالها را ۹ داور "ضروری"، ۲ داور "مفید" و ۱ داور "غیرضروری" تشخیص داده‌اند. ضریب روایی آن سؤال به قرار زیر محاسبه می‌شود:

$$CVR = \frac{9 - \frac{12}{2}}{\frac{12}{2}} = \frac{9-6}{6} = \frac{3}{6} = ۰/۵$$

ضریب فوق مُعرف روایی یک سؤال است. برای تعیین ضریب روایی کُل آزمون، لازم است میانگین ضریبهای روایی همهٔ سؤالها محاسبه شود که همان **شاخص روایی محتوایی**[1] (سی وی آی) است.

لازم به ذکر است که هنگام داوری دربارهٔ روایی محتوایی یک آزمون تنها اتکاء به شاخص روایی محتوایی کافی نیست. برای روشن شدن این نکته، فرض کنید معلمی برای یکی از درسهای خود آزمونی ساخته است که در آن بیشتر سؤالها تنها از یک فصل کتاب طرح شده‌اند و اکثراً سؤالهای خوبی هستند. پس از محاسبهٔ میانگین ضریبهای روایی سؤالها و به دست آوردن شاخص روایی محتوایی آزمون، معلوم می‌شود که آزمون از روایی خوبی برخوردار است، درحالی که می‌دانیم چنین نتیجه‌گیری درست نیست. بنابراین، لازم است نظر داوران را دربارهٔ اینکه سؤالهای آزمون تا چه میزان محتوای مورد سنجش را پوشش می‌دهند نیز جویا شد. انجام این کار مستلزم داشتن جدول مشخصات آزمون است؛ یعنی همان چیزی که اساس روایی محتوایی را تشکیل می‌دهد.

روایی آموزشی

یکی از مفهومهای نسبتاً تازهٔ وابسته به روایی محتوایی **روایی آموزشی**[2] است که به آن **روایی برنامهٔ درسی**[3] نیز گفته می‌شود. هوگان (۲۰۰۷) این نوع روایی را کاربست ویژهٔ روایی محتوایی می‌داند. چنان که دیدیم، روایی محتوایی در ارتباط با این پرسش مطرح می‌شود که "آیا محتوای آزمون یا وسیلهٔ سنجش محتوای موضوع درسی موردنظر را به خوبی پوشش می‌دهد؟" در مقابل، روایی آموزشی یا روایی برنامهٔ درسی در رابطه با پرسش زیر مطرح می‌شود: "آیا محتوای آزمون یا وسیلهٔ سنجش عملاً آموزش داده شده است؟" لذا، برای اینکه آزمون دارای روایی آموزشی باشد باید شواهدی موجود باشند حاکی از اینکه محتوای آزمون در یک برنامهٔ آموزشی به خوبی آموزش

1. Content Validity Index (CVI)
2. instructional validity
3. curricular validity

داده شده است. برای این منظور، این سؤال مطرح می‌شود که "آیا دانش‌آموزان یا دانشجویانی که این آزمون به آنها داده می‌شود عملاً در معرض آموزش مطالبی که در آزمون گنجانیده شده قرار گرفته‌اند؟" اگر جواب سؤال مثبت باشد آزمون دارای روایی آموزشی است.

هرچند به ظاهر روایی آموزشی و روایی محتوایی نزدیک به هم به‌نظر می‌رسند، اما یک چیز نیستند. هوگان (۲۰۰۷) برای تمایزگذاری بین این دو نوع روایی، موضوع ریشهٔ دوم اعداد را مثال زده است. این موضوع بخشی از برنامهٔ درسی و کتابهای درسی مدارس را تشکیل می‌دهد. بنابراین، آزمونهای پیشرفت تحصیلی ریاضیات که از سوی ادارات آموزش و پرورش تهیه می‌شوند سؤالهای مربوط به ریشهٔ دوم یا جذر اعداد را شامل‌اند که این خود حاکی از روایی محتوایی این نوع آزمونهاست. حال فرض کنید معلمان ریاضی به دلایلی جذر اعداد دورقمی را به دانش‌آموزان خود آموزش ندهند. در این صورت آزمونهای نام‌برده دارای روایی آموزشی نیستند.

روایی پیش‌بینی آزمونهای وابسته به ملاک

در قسمت قبلی این فصل دیدیم که روایی پیش‌بینی آزمونهای وابسته به‌هنجار برای این منظور به کار می‌رود که نشان دهد عملکرد آزمون‌شونده در یک آزمون (آزمون پیش‌بینی) تا چه اندازه عملکرد او را در آزمونی دیگر (آزمون ملاک) پیش‌بینی می‌کند. در آزمونهای وابسته به ملاک روایی پیش‌بینی به دو منظور زیر به‌کار می‌رود.

۱. **پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی بر اساس هدف آموزشی دیگر.** در آزمونهای وابسته به ملاک یکی از هدفهای مهم استفاده از روایی پیش‌بینی این است که تعیین می‌کند عملکرد آزمون‌شونده (دانش‌آموز یا دانشجو) در یک هدف آموزشی تا چه میزانی عملکرد او را در هدف آموزشی دیگری که وابسته به هدف اول است پیش‌بینی می‌کند. ممکن است هدف آموزشی اول شامل دانش یا مهارتهایی باشد که برای یادگیری هدف دوم پیش‌نیاز محسوب می‌شود، مانند اینکه دانش‌آموز ابتدا نیاز دارد در منهاکردن اعداد کاملاً مهارت کسب کند (هدف اول) تا بتواند تقسیمهای طولانی را بیاموزد (هدف دوم).

۲. **پیش‌بینی یادگیری در حد تسلط.** علاوه بر پیش‌بینی عملکرد دانش‌آموزان یا دانشجویان در یک هدف آموزشی بر اساس هدف آموزشی دیگر، مورد استفادهٔ دوم روایی پیش‌بینی آزمونهای وابسته به ملاک پیش‌بینی یادگیری در حد تسلط یا تسلط‌یابی یادگیرندگان در هدفهای آموزشی است.

روش تعیین روایی پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی براساس هدف آموزشی دیگر

برای تعیین اینکه عملکرد دانش‌آموزان در یک هدف آموزشی تا چه میزانی عملکرد آنان را در

جدول ۳ ـ ۲۰ همبستگی بین دو آزمون وابسته به ملاک مربوط به دو هدف آموزشی

نام دانش‌آموز	نتایج هدف ۱	نتایج هدف ۲
احمد	۸	۶
جمشید	۱۰	۱۰
حسین	۵	۵
پرویز	۳	۵
داریوش	۸	۹
علی	۱۰	۸
محمود	۹	۱۰
اردشیر	۶	۸
محمد	۷	۸
امیر	۱۰	۱۰

هدف آموزشی دیگر که وابسته به هدف اول است تعیین می‌کند، لازم است ضریب همبستگی عملکرد دانش‌آموزان را در سؤالهای مربوط به هدفهای مورد نظر تعیین نماییم. برای روشن‌شدن موضوع، مثال زیر را مورد توجه قرار دهید. فرض کنید ۱۰ دانش‌آموز به یک آزمون ۱۰ سؤالی مربوط به هدف شمارهٔ ۱ جواب داده‌اند. بعد به این دانش‌آموزان فرصت یادگیری هدف شمارهٔ ۲ داده شده است (دقت کنید که دانش‌آموزان همگی در هدف شمارهٔ ۱ به حد تسلط نرسیده‌اند؛ مقصود از این کار این است که معلوم شود که آیا بعضی از دانش‌آموزان بدون یادگیری کامل هدف ۱ به هدف ۲ می‌رسند یا نه.). درمرحلهٔ بعد، آزمون وابسته به ملاک مربوط به هدف شمارهٔ ۲ به دانش‌آموزان داده می‌شود و نتایج آن تعیین می‌گردد. جدول ۳ ـ ۲۰ نتایج آزمونهای مربوط به دو هدف فوق را نشان می‌دهد. در این جدول تعداد سؤالهایی که دانش‌آموزان در هر یک از دو آزمون درست جواب داده‌اند آمده است. در این جدول همچنین ضریب همبستگی بین نمرات حاصل از دو آزمون محاسبه شده است. دقت کنید که فرمول محاسبهٔ ضریب همبستگی همان فرمول محاسبهٔ ضریب همبستگی گشتاوری پیرسون است که در فصل هجدهم آن را بحث کردیم.

$\Sigma XY = ۶۳۴ \quad \Sigma Y = ۷۹ \quad \Sigma X = ۷۶$

$\Sigma Y^۲ = ۶۵۹ \quad \Sigma X^۲ = ۶۲۸$

$$r = \frac{N\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{\left[N\Sigma X^۲ - (\Sigma X)^۲\right]\left[N\Sigma Y^۲ - (\Sigma Y)^۲\right]}}$$

$$r = \frac{۱۰ \times ۶۳۴ - ۷۶ \times ۷۹}{\sqrt{\left[۱۰ \times ۶۲۸ - (۷۶)^۲\right]\left[۱۰ \times ۶۵۹ - (۷۹)^۲\right]}}$$

$$r = ۰/۸۰$$

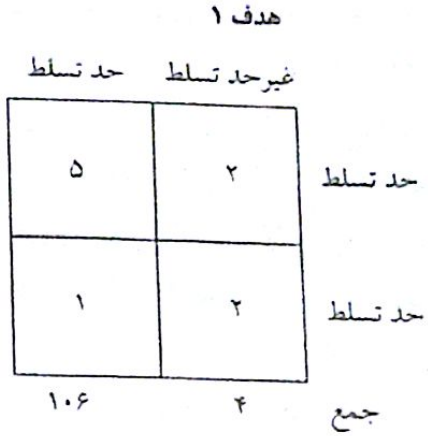
ضریب همبستگی ۰/۸۰+ نشان‌دهندهٔ روایی پیش‌بینی نسبتاً بالایی برای آزمون شمارهٔ ۱ است؛ به این معنی که از نمرات دانش‌آموزان در هدف ۱ به عنوان یک عامل پیش‌بینی توفیق دانش‌آموزان در یادگیری هدف شمارهٔ ۲ می‌توان استفاده کرد. به سخن دیگر، می‌توان این گونه نتیجه گیری کرد که مهارتها و دانش ملحوظ در هدف ۱ پیش‌نیاز یادگیری هدف ۲ به حساب می‌آیند. به تعبیر دیگر، می‌توان گفت دانش‌آموزانی که در آزمون مربوط به هدف ۱ عملکرد ضعیفی داشته‌اند عملکردشان در آزمون مربوط به هدف ۲ نیز ضعیف بوده است، و نیز دانش‌آموزانی که در آزمون مربوط به هدف ۱ نمرات بالایی گرفته‌اند در آزمون مربوط به هدف ۲ نیز موفق‌تر بوده‌اند.

روش تعیین روایی پیش‌بینی یادگیری در حد تسلط

برای آشناشدن با چگونگی تعیین روایی پیش‌بینی یادگیری در حد تسلط به مثال زیر توجه کنید. فرض کنید مرز رسیدن به حد تسلط در دو آزمون وابسته به ملاک حداقل نمرهٔ ۸۰ از صد یا ۸ سؤال از ۱۰ سؤال است. ابتدا باید تعیین کنیم که کدام دانش‌آموزان به حد تسلط رسیده و کدام به این حد نرسیده‌اند. در جدول ۴ ـ ۲۰ این کار برای نمرات جدول ۳ ـ ۲۰ که قبلاً مورد استفاده قرار دادیم انجام شده است. نمرهٔ ۱ معرف رسیدن به حد تسلط و نمرهٔ ۰ نشان‌دهندهٔ نرسیدن به حد تسلط است.

جدول ۴ ـ ۲۰ مقایسهٔ عملکرد دانش آموزان در دو هدف آموزشی از لحاظ رسیدن یا نرسیدن به حد تسلط

دانش‌آموزان	هدف ۱	هدف ۲
احمد	۱	۰
جمشید	۱	۱
حسین	۰	۰
پرویز	۰	۰
داریوش	۱	۱
علی	۱	۱
محمود	۱	۱
اردشیر	۰	۱
محمد	۰	۱
امیر	۱	۱



شکل ۲ ـ ۲۰ درصد عملکرد حدتسلط و غیرحدتسلط دانش‌آموزان در هدفهای ۱ و ۲

ضریب همبستگی حاصل از داده‌های جدول ۴ ـ ۲۰ رقم ۰/۳۶ را نشان می‌دهد که البته بسیار کوچک‌تر از ضریب ۰/۸۰ محاسبه شده در جدول ۳ ـ ۲۰ است. دلیل اینکه ضریب اخیر خیلی کوچک‌تر از ضریب قبلی است این است که در ارقام جدول ۴ ـ ۲۰ تفاوتهای فردی را نادیده گرفته‌ایم و همهٔ حدتسلطی‌ها و غیرحدتسلطی‌ها را یکسان نمره گذاری کرده‌ایم. افزون بر این، ضریبهای همبستگی مبتنی بر متغیرهایی که به صورت ۰ و ۱ یا دوارزشی نمره گذاری می‌شوند به وسیلهٔ درصد اشخاصی که در توزیعهای حاشیه‌ای هستند محدود می‌گردند. در این مثال، درصدهای حاشیه‌ای عبارت‌اند از حدتسلطی‌ها و غیرحدتسلطی‌ها در هدفهای مجزا. وقتی که این درصدها در دو متغیر یکی نیستند، مقدار ضریب همبستگی کاهش می‌یابد. به همین سبب است که ضریب همبستگی ۰/۸۰ که از نمرات خام به دست آمد، وقتی که با استفاده از نمرات دوارزشی ۰ و ۱ محاسبه شد مقدار زیادی کاهش یافت.

بنابر مشکلات فوق، شاید بهتر باشد که روایی پیش‌بینی آزمونهای وابسته به ملاک به عوض ضریب همبستگی به صورت درصد دانش‌آموزانی که عملکرد حدتسلط و عملکرد غیرحدتسلط آنان در هدفهای ۱ و ۲ یکسان است تعیین گردد. از اطلاعات داده‌شده در شکل ۲ ـ ۲۰ می‌بینیم که ۵ دانش‌آموز در هر دو هدف به حدتسلط رسیده‌اند ولی ۲ نفر در هر دو هدف به این حد نرسیده‌اند. بنابراین، ۷ نفر از ۱۰ نفر، یعنی ۷۰٪، در دو هدف عملکرد یکسان داشته‌اند. رقم ۷۰٪ معرف یک ضریب روایی نسبتاً قابل توجه است. پس از این ضریب روایی نتیجه می‌گیریم که تسلط‌یابی در هدف دوم به تسلط‌یابی در هدف اول وابسته است.

سایر استفاده‌ها از روایی پیش‌بینی آزمونهای وابسته به ملاک

علاوه بر تعیین رابطهٔ همبستگی بین عملکرد یادگیرندگان در دو هدف آموزشی، نظیر آنچه در مثال بالا دیدیم، این روش را می‌توان در موقعیتهای دیگری نیز به کار بست. ویرسما و جورس (۱۹۹۰) می‌گویند "از نمرات آزمونهای وابسته به ملاک می‌توان برای پیش‌بینی عملکرد در دیگر

آزمونهای پیشرفت تحصیلی، درجه‌بندیهای معلم از توانایی تحصیلی دانش‌آموزان، نمرات سنجش نگرش، یا هر ملاک مورد نظر دیگر سود جست" (ص ۲۷۸).

چنان‌که قبلاً گفته شد، شواهد مربوط به روایی پیش‌بینی نشان می‌دهند که می‌توان از روی اندازه‌های یک متغیر (یا یک آزمون) عملکرد آزمون‌شوندگان را در متغیری دیگر (یا آزمونی دیگر) پیش‌بینی کرد. اندازهٔ شاخص روایی، یعنی ضریب همبستگی یا درصد تصمیمهای یکجور در بارهٔ حدتسلط، هم به درستی آزمون پیش‌بینی‌کننده وابسته است و هم به درستی آزمون یا وسیلهٔ پیش‌بینی‌شونده یا آزمون ملاک. به سخن دیگر، نه تنها یک مقیاس یا آزمون پیش‌بینی‌کنندهٔ نامطمئن را نمی‌توان برای مقاصد پیش‌بینی به‌کار بست، بلکه یک مقیاس یا ملاک نامطمئن را هرگز نمی‌توان به درستی پیش‌بینی کرد.

روایی سازهٔ آزمونهای وابسته به ملاک

چنان‌که قبلاً دیدیم، سازه به صفت، ویژگی، یا مفهومی که مستقیماً قابل مشاهده نیست ولی از نظریه‌های روانشناختی قابل استنباط است گفته می‌شود. بنابراین، آزمونهایی که برای اندازه‌گیری سازه‌های روانی درست می‌شوند به طور مستقیم سازه‌ها را نمی‌سنجند بلکه عملکرد یا رفتارهایی را که گمان می‌رود منعکس‌کنندهٔ سازه‌ها هستند می‌سنجند. ویرسما و جورس (۱۹۹۰) مثال زیر را ذکر کرده‌اند. ما انتظار داریم که نمرات حاصل از یک مقیاس یا پرسشنامهٔ احساس خوشبختی با نمرات به دست آمده از یک آزمون مفهوم‌خود (خودپنداره)[1] همبسته باشند و نیز میانگین نمرات مفهوم‌خود پسران و دختران کلاس سوم تقریباً برابر باشند. اگر این انتظارهای ما برآورده شوند، شواهدی در دست خواهیم داشت حاکی از روایی آزمون مفهوم‌خود، یعنی اینکه آزمون تهیه‌شده آن چیزی را که به قصد اندازه‌گیری‌اش ساخته شده است می‌سنجد.

روش تعیین روایی سازهٔ آزمونهای وابسته به ملاک

برای تعیین روایی سازهٔ آزمونها، به بیش از یک نوع تحلیل نیاز است. به سخن دیگر، تعیین روایی سازهٔ آزمونها دست‌کم مستلزم دو تحلیل یکی منطقی و دیگری آماری است. بنابراین، روایی سازه و روایی‌های محتوایی و پیش‌بینی دارای وجوه اشتراکی هستند. در زیر به موردی از تعیین روایی سازه با استفاده از هر دو تحلیل منطقی و آماری توجه کنید. این مثال از کتاب ***اندازه‌گیری و آزمودن*** اثر ویرسما و جورس (۱۹۹۰) اقتباس شده است.

فرض کنید ما یک آزمون وابسته به ملاک احیای قلبی-ریوی (CPR)[2] درست کرده‌ایم و برای متقاعد ساختن دیگران به اثربخشی این آزمون می‌خواهیم مطالعاتی در بارهٔ روایی آن انجام دهیم. تحلیلهای زیر اطلاعات مهمی در بارهٔ روایی این آزمون در اختیار ما قرار خواهند داد.

---

1. self-concept
2. Cardio-Pulmonary Resuscitation

تحلیل منطقی. ماده‌های آزمونِ احیای قلبی-ریوی یا CPR باید حوزه‌ای را که مشخص شده است معرفی کنند. بدین منظور، هیأتی مرکب از پزشکان مطلع، مدرسان CPR، و متخصصان آموزش بهداشتْ این آزمون را مورد بازنگری قرار می‌دهند. آنها با روش درجه‌بندیْ آزمون CPR را از چند جهت زیر ارزیابی می‌کنند: (۱) میزان مطابقت ماده‌های آزمون با حوزهٔ مورد نظر؛ (۲) منظورشدن یا نشدن مهارتها یا اندیشه‌های مهم، و (۳) شکل ظاهری و ساختار ماده‌های آزمون. دقت کنید که این تحلیل بسیار شبیه به تحلیل روایی محتوایی است که در صفحات پیش توضیح دادیم. علت آن است که روایی سازه با دیگر روواییها به ویژه روایی محتوایی همپوشی دارد.

تحلیل آماری ـ همبستگی. انتظار ما این است که عملکرد در آزمون CPR با اندازه‌های دیگر مهارتهای شبیه به آن همبسته باشد. برای کسب اطمینان از این مطلب، باید آزمون CPR و مقیاسهای دیگر اندازه‌گیری را که به همان منظور تهیه شده‌اند با گروه واحدی از افراد به اجرا درآوریم و بعد ضرایب همبستگی بین نمرات حاصل را محاسبه نماییم. انتظار ما این است که بین آزمون CPR ما و آزمونهای CPR تهیه‌شده توسط سازمانهای دیگر، مثلاً هلال احمر یا صلیب سرخ، ضرایب همبستگی مثبت و بالایی به دست آیند. همچنین انتظار خواهیم داشت که بین نمرات CPR ما و نمرات مقیاسهای مربوط به فنون دیگر کمکهای اولیه، مانند بندآوردن خونریزش و مانور هیملیچ[1] (بیرون آوردن اجسام خارجی از گلو) ضرایب همبستگی مثبت اما متوسط یافت شوند. همچنین ممکن است بین آزمون CPR ما و یک آزمون دانش کالبدشناسی دستگاه تنفسی همبستگی مثبت یافت شود. حتی نمرات این آزمون ممکن است پاسخهای بعدی به موقعیتهای اضطراری (اورژانس) را پیش‌بینی کنند.

اندازه‌های متغیرهای مشابه باید قویاً با هم همبسته باشند؛ اندازه‌های متغیرهای متفاوت اما به هم مرتبط باید همبستگیهای ضعیف یا متوسطی را به دست بدهند؛ و اندازه‌های متغیرهای نامرتبط، نظیر نمرات CPR و شمارهٔ قفل کمدهای دانشجویان، باید ضرایب همبستگی صفر به دست دهند. بررسی الگوی همبستگیهای نمرات آزمون با متغیرهای دیگر به ما کمک می‌کند تا بفهمیم آزمون مورد نظر برای منظوری که درست شده مناسب است یا نه.

تحلیل آماری ـ تفاوتهای گروهی. در مثال مربوط به آزمون CPR که در بالا توضیح داده شد، ممکن است بتوانیم پیش‌بینی نماییم که گروههای مختلف افراد عملکردهای مختلفی در آزمون خواهند داشت که از لحاظ آماری این تفاوتها معنی‌دار هستند. به سخن دیگر، کسانی که در فنون CPR کارآزموده هستند از کسانی‌که در این زمینه تجربهٔ کافی ندارند عملکردشان بهتر است، و همچنین معلمان CPR در این آزمون نمرات بالاتری از شاگردانشان دریافت خواهند کرد. ما

1. Heimlich

می‌توانیم آزمون خود را با این گروههای مختلف افراد انجام دهیم و نتایج را محاسبه کنیم و ببینیم که پیش‌بینیهای ما درست درمی‌آیند یا نه. اگر معلوم گشت که نتایج حاصل پیش‌بینیهای نظری ما را تأیید می‌کنند آن گاه اطلاعات بیشتری در بارهٔ روایی آزمون خود در دست خواهیم داشت، یعنی اینکه آزمون ما آنچه را که به قصد اندازه‌گیری‌اش درست شده است می‌سنجد.

همچنین مهم است بدانیم که تصمیمات مربوط به حدتسلط مبتنی بر نمرات آزمون وابسته به ملاک که می‌خواهیم روایی آن را تعیین نماییم با تصمیمات مربوط به حدتسلط حاصل از نتایج دیگری که در همان حوزه و یا همان هدف ولی به توسط وسائل یا ابزارهای دیگری که بر روی دانش‌آموزان اجرا شده به دست آمده‌اند توافق دارند یا نه. ابزارهای دیگر می‌توانند درجه‌بندیهای معلم یا نمرات حاصل از یک آزمون دیگر باشند. اگر همهٔ تصمیمات مربوط به وضع حدتسلط با هم همخوانی داشته باشند، آن‌گاه شواهدی در دست خواهیم داشت حاکی از اینکه نمرات آزمون وابسته به ملاک مورد بررسی ما سازهٔ مورد نظر را می‌سنجد.

عوامل مؤثر بر روایی آزمون

عوامل مختلفی وجود دارند که بر روایی آزمونها تأثیر می‌گذارند و بنابراین با کنترل آنها می‌توان سطح روایی یک آزمون را افزایش داد. ما این عوامل را در دو دستهٔ عوامل درونی و عوامل بیرونی معرفی می‌کنیم.

عوامل درونی مؤثر بر روایی آزمون

نخستین عاملی که می‌تواند بر روایی آزمون تأثیر بگذارد *دستورالعمل یا راهنمای آزمون* است. راهنماهایی که دقیقاً به آزمون‌شوندگان نمی‌گویند چگونه به سؤالها پاسخ دهند و چگونه پاسخ خود را ثبت نمایند سطح روایی آزمون را کاهش می‌دهند.

عامل دیگری که بر روایی آزمون تأثیر بسزایی دارد *کیفیت سؤالهای آزمون* است. سؤالهایی که طبق قواعد درست سؤال‌نویسی تهیه شده باشند به روایی آزمون می‌افزایند، و برعکس سؤالهایی که از هر لحاظ مشکل داشته باشند به روایی آزمون لطمه می‌زنند. برای آشنایی با قواعد تهیهٔ سؤالهای آزمونهای مختلف به فصلهای ۷ تا ۱۱ این کتاب مراجعه کنید. علاوه بر این، هر کدام از سؤالهای آزمون که از هدف اصلی آزمون جدا می‌افتد و چیزی جدا از آن را شامل می‌شود نیز به سهم خود از روایی کل آزمون می‌کاهد. همچنین واژگان و جمله‌بندی پیچیده که از حد درک آزمون‌شوندگان فراتر باشد به روایی آزمون لطمه می‌زند.

عامل درونی دیگری که بر روایی آزمون تأثیر دارد *ترتیب قرارگرفتن سؤالهای آزمون به دنبال یکدیگر* است. ما در این باره در فصل ۱۳ سفارشهای لازم را دادیم. از جمله گفتیم که قراردادن

سؤالهای بسیار دشوار در آغاز آزمون ممکن است سبب شود که پاسخ‌دهندگان به آزمون نتوانند وقت خود را برای بخشهای مختلف آزمون به درستی تقسیم کنند. همچنین اگر سؤالهای مختلف آزمون از لحاظ درجهٔ دشواری متنوع نباشند، روایی آزمون صدمه می‌بیند. آزمونهایی که دارای مجموعه سؤالهای بسیار دشوار یا بسیار ساده‌اند از سطح روایی خوبی برخوردار نیستند.

باز هم از عوامل دیگر مؤثر بر روایی *طول یک آزمون یا تعداد سؤالهای آن* است. در فصل بعد در بحث از پایایی آزمون خواهیم گفت که روایی از پایایی تأثیر می‌پذیرد و پایایی نیز با افزایش تعداد سؤالهای یک آزمون افزایش می‌یابد. بنابراین، یکی دیگر از راههای بالابردن سطح روایی آزمون افزایش تعداد سؤالهای آن است. علاوه بر این، در آزمونهایی که با روایی محتوایی معرفی می‌شوند، هر چه تعداد سؤالها بیشتر باشد نمونهٔ بهتری از محتوای مورد اندازه‌گیری را شامل خواهند بود و به همان نسبت به روایی آزمون افزوده می‌شود.

عوامل بیرونی مؤثر بر روایی آزمون

علاوه بر عوامل درونی، چندین عامل بیرونی نیز وجود دارند که بر روایی آزمونها تأثیر می‌گذارند، و برای افزایش دادن سطح روایی یک آزمون می‌توان آنها را کنترل کرد. یکی از این عوامل بیرونی *اجرای نامناسب آزمون* است. کمبود زمان، فضای نامناسب، مکان نامناسب، و مانند اینها از عوامل اجرایی هستند که بر عملکرد آزمون‌شوندگان و در نتیجه بر روایی آزمون تأثیر نامطلوبی به جای می‌گذارند. ما در فصل ۱۷ نکات ضروری مربوط به اجرای درست آزمون را توضیح داده‌ایم.

علاوه بر اجرای نامناسب آزمون، *اشکالات موجود در تصحیح پاسخهای آزمون‌شوندگان و نمره‌گذاری آنها* به ویژه در آزمونهای تشریحی می‌تواند نتایج حاصل را مخدوش سازد و لذا از روایی آزمون بکاهد.

عامل دیگر مؤثر بر روایی آزمون *ترکیب آزمون‌شوندگان از لحاظ توانایی مورد اندازه‌گیری* است. هر چه گروه آزمون‌شوندگان از لحاظ توانایی مورد اندازه‌گیری نامتجانس‌تر باشند روایی آزمون بیشتر خواهد بود. به همین دلیل آزمونهایی که وقتی با گروههای نامتجانس به کار می‌روند دارای روایی قابل ملاحظه‌ای هستند اگر با گروههای دارای تجانس بیشتر به کار گرفته شوند سطح روایی کمتری از خود نشان خواهند داد. برای مثال، آزمونی که برای پیش‌بینی میانگین دانشجویان سال اول دانشگاه تهیه شده است، اگر در مورد دانشجویان همهٔ سالهای تحصیلی دانشگاهی به کار رود روایی پیش‌بینی بیشتری نشان خواهد داد. بنابراین، باید توجه داشت که افزایش و کاهش ضریب روایی در پاره‌ای از مواقع ممکن است به سبب کاهش یا افزایش تجانس آزمون‌شوندگان باشد.

علاوه بر عوامل فوق، *ویژگیهای روانی آزمون‌شوندگان* نیز می‌توانند بر روایی نمرات حاصل از

یک آزمون تأثیر داشته باشند. برای نمونه، بالا و پایین بودن سطح انگیزش آزمون‌شوندگان یا علاقه‌مندبودن و بی‌علاقه‌بودن آنان نسبت به پاسخ‌دادن به سؤالهای آزمون و نیز کمی یا زیادی سطح اضطراب آنان بر نمرات حاصل و لذا بر میزان روایی نمراتِ حاصل تأثیر می‌گذارد. علاقه‌مندی آزمون‌شوندگان به شرکت در آزمون و اندکی از اضطراب بر دقت آنان و روایی نتایج آزمون تأثیر مثبت به جای می‌گذارد، اما بی‌علاقگی یا نفرت آزمون‌شوندگان یا اضطراب زیاد آنان در جلسهٔ آزمون هم بر دقت عمل و هم بر روایی نتایج حاصل تأثیر منفی دارد.

خلاصهٔ فصل

متخصصان اندازه‌گیری و ارزشیابی برای آزمونهای مختلف ویژگیهای زیادی را برشمرده‌اند که از آن جمله می‌توان عینیت، سهولت اجرا، سهولت نمره‌گذاری، عملی‌بودن، سهولت تعبیر و تفسیر، روایی، و پایایی را نام برد. این فصل یکی از این ویژگیها یعنی *روایی* را مورد بحث قرار داده است. روایی به هدفی که آزمون برای تحقق‌بخشیدن به آن درست شده است اشاره می‌کند. مهم‌ترین انواع روایی عبارت‌اند از *روایی محتوایی*، *روایی ملاکی*، و *روایی سازه*. روایی محتوایی به این مطلب اشاره می‌کند که آیا آزمون از مجموعه سؤالهایی درست شده است که معرف یک نمونهٔ واقعی از محتوای درس و هدفهای آموزشی آن باشد یا نه. موردی از روایی محتوایی *روایی صوری* است. روایی صوری نشان می‌دهد که صورت ظاهر سؤالهای یک آزمون به آنچه آزمون می‌خواهد بسنجد شباهت دارد یا نه.

روایی ملاکی دو بخش مهم دارد که عبارت‌اند از *روایی پیش‌بینی* و *روایی همزمان*. روایی پیش‌بینی نشان می‌دهد که آیا بر اساس عملکرد آزمون شوندگان در یک آزمون می‌توان عملکرد آنها را در آزمون دیگری پیش‌بینی کرد یا نه. به آزمون اول که برای پیش‌بینی به کار می‌رود *آزمون پیش‌بین* و به آزمون دوم که نمرات آزمودنی در آن از روی آزمون پیش‌بین پیش‌بینی می‌شود *آزمون ملاک* می‌گویند. اگر نمرات آزمون پیش‌بین عملکرد آزمون‌شونده را در آزمون دیگری که در آینده اجرا خواهد شد پیش‌بینی کند می‌گویند که آزمون دارای روایی پیش‌بینی است؛ اگر نمرات آزمون پیش‌بین عملکرد آزمون‌شونده را در آزمون دیگری که همزمان با آزمون پیش‌بین اجرا می‌شود یا اندک زمانی بعد از آن اجرا می‌شود پیش‌بینی کند، می‌گویند آزمون دارای روایی همزمان است. آزمونی که به عنوان ملاک به کار می‌رود باید دارای ویژگیهای زیر باشد: ربط‌داشتن، بی‌طرفی، پایایی، و در دسترس‌بودن.

روایی دیگری که در این فصل مورد بحث قرار گرفته روایی سازه است. روایی سازه ویژهٔ آزمونهایی است که درستی نظریه‌های علمی را می‌سنجند. یک آزمون در صورتی دارای روایی سازه است که نمرات حاصل از اجرای آن به مفاهیم یا سازه‌های نظریهٔ مورد نظر مربوط باشند.

روشهای تعیین روایی سازه که در این فصل معرفی شده‌اند عبارت‌اند از روش همبستگی، روش تمایز سنی، روش تحلیل عوامل، و روش همسانی درونی.

ما در فصل حاضر روایی آزمون را از دو جهت مورد بحث قرار داده‌ایم: (۱) روایی آزمونهای وابسته به هنجار، و (۲) روایی آزمونهای وابسته به ملاک. در قسمت اول فصل روایی آزمونهای وابسته به هنجار معرفی شدند و در قسمت دوم فصل روایی آزمونهای وابسته به ملاک مورد بحث و بررسی قرار گرفته‌اند. در آخرین قسمت فصل نیز عوامل مؤثر بر روایی آزمون را در دو بخش *عوامل درونی* و *عوامل بیرونی* معرفی کردیم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. می‌توان گفت که روایی محتوایی همان روایی آموزشی است.

الف ـ صحیح

ب ـ غلط

۲. روایی صوری برای کدام یک از وسایل اندازه‌گیری زیر ضروری **نیست**؟

الف ـ پرسشنامهٔ انتخاب معلم

ب ـ آزمون خلاقیت

پ ـ پرسشنامهٔ تشخیص نابهنجاریهای روانی

ت ـ آزمون تاریخ کلاس سوم ابتدایی

ث ـ وسیلهٔ انتخاب شناگر برای شرکت در یک مسابقهٔ ملی

۳. کدام یک از روایی‌های زیر با ضریب آماری مشخص نمی‌شود؟

الف ـ پیش‌بینی

ب ـ همزمان

پ ـ محتوایی

ت ـ سازه

۴. اگر یک آزمون به خوبی مُعرف مهارتها و موضوعهای یک واحد درسی باشد، آن آزمون دارای چه نوع روایی است؟

الف ـ محتوایی

ب ـ سازه

پ ـ پیش‌بینی

ت ـ همزمان

۵. اگر نمره‌های یک آزمون استعداد تحصیلی اجرا شده در سال آخر دبیرستان با نمره‌های سال اول دانشگاه همبستگی مثبت و بالایی داشته باشد، آزمون استعداد دارای کدام نوع روایی است؟

الف ـ محتوایی

ب ـ سازه

پ ـ پیش‌بینی

ت ـ صوری

۶. دو فرم سؤال تشریحی برای درس روانشناسی عمومی دانشگاه، دورهٔ کارشناسی، تهیه و به یک گروه از دانشجویان داده شده است تا به آنها پاسخ دهند. ضریب همبستگی که بین این دو فرم سؤال یا دو آزمون تشریحی محاسبه می‌شود معرف کدام روایی زیر است؟

الف ـ محتوایی

ب ـ پیش‌بینی

پ ـ همزمان

ت ـ صوری

ث ـ سازه

۷. کدام یک از روایی‌های زیر هم مستلزم تحلیل منطقی و هم تحلیل آماری است؟

الف ـ محتوایی

ب ـ سازه

پ ـ پیش‌بینی

ت ـ صوری

۸ یکی از روانشناسان معاصر به نام سلیگمن مفهومی به نام درماندگی آموخته‌شده ارائه داده و ادعا کرده است کسانی که در شرایطی قرار می‌گیرند که هیچ عملی از آنان بر تغییر سرنوشت فلاکت‌بارشان اثری ندارد به یک حالت ناامیدی و درماندگی مبتلا می‌شوند که به آن **درماندگی آموخته شده** می‌گوید. او همچنین اظهار داشته است که این افراد بیشتر از دیگران به حالت افسردگی و سایر بیماریهای وابسته به آن مبتلا می‌شوند. اگر بخواهید یک وسیلهٔ اندازه‌گیری برای تعیین درماندگی آموخته‌شده تهیه کنید این وسیلهٔ شما چه نوع روایی باید داشته باشد؟

الف ـ محتوایی

ب ـ صوری

پ ـ پیش‌بینی

ت ـ همزمان

ث ـ سازه

۹. روایی آزمون به چه چیزی اشاره می‌کند؟
الف ـ قابلیت استفادهٔ آزمون
ب ـ هدفی که آزمون برای آن درست شده است
پ ـ ثبات نتایج حاصل از اجرای آزمون
ت ـ سهولت اجرای آزمون

۱۰. مطمئن‌ترین روش تعیین روایی آزمونهای پیشرفت تحصیلی چیست؟
الف ـ تعیین ضریب همبستگی آزمون با یک آزمون موجود دارای روایی
ب ـ مراجعه به جدول مشخصات آزمون و کسب اطلاع از کیفیت این جدول
پ ـ مقایسهٔ نتایج حاصل از اجرای آزمون با میانگین درسی دانش‌آموزان
ت ـ محاسبهٔ ضریب همبستگی بین دو نیمه‌های آزمون

۱۱. روایی محتوایی را تعریف کنید و بهترین راه تهیهٔ آزمونهای دارای روایی محتوایی را توضیح دهید.

۱۲. ویژگیهای مطلوب آزمونهای روانی و پرورشی را نام ببرید و دو تا از مهم‌ترین آنها را مشخص کنید.

۱۳. برای یکی از کتابهای درسی با استفاده از روش تهیهٔ طرح آزمون که در فصل ششم کتاب آمده است یک آزمون دارای روایی محتوایی بسازید و این آزمون را به عنوان وسیله‌ای برای پیش‌بینی عملکرد دانش‌آموز در یک آزمون دیگر مورد استفاده قرار دهید. توضیح دهید که چگونه این آزمون هم دارای روایی محتوایی و هم دارای روایی پیش‌بینی خواهد بود.

۱۴. روشهای مختلف تعیین روایی سازه را توضیح دهید.

۱۵. روشهای تعیین روایی آزمونهای وابسته به ملاک را توضیح دهید و تفاوت آنها را با روشهای تعیین روایی آزمونهای وابسته به هنجار مشخص کنید.

۱۶. عوامل مختلف مؤثر بر روایی آزمون و راههای بالابردن روواییهای مختلف را توضیح دهید.

۲۱

پایایی آزمون

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. پایایی را تعریف کند و رابطهٔ آن را با روایی توضیح دهد.

۲. انواع پایایی را تعریف کند و محاسبهٔ ضریب پایایی با روشهای مختلف را توضیح دهد.

۳. موارد استفادهٔ پایایی‌های مختلف را تشخیص دهد.

۴. روش آزمون مجدد و روش دو نیمه کردن آزمون را با هم مقایسه کند.

۵. تفاوت بین استفاده از فرمول KR۲۰ و KR۲۱ را شرح دهد.

۶. خطای معیار اندازه‌گیری را تعریف کند و روش محاسبه و نحوهٔ استفاده از آن را شرح دهد.

۷. با استفاده از روشهای مختلف بیان شده در این فصل، ضریب پایایی آزمونی را که خود تهیه کرده است پیدا کند.

۸. خطای معیار اندازه‌گیری آزمون خود را محاسبه کند و آن را تفسیر نماید.

۹. فرمول اصلاح کاهش را توضیح دهد و موارد استفادهٔ آن را بگوید.

۱۰. توضیح دهد که اگر بخواهیم همبستگی بین دو آزمون X و Y را به مقدار معینی افزایش دهیم چه مقدار باید بر ضرایب پایایی آن دو آزمون بیفزاییم.

۱۱. پایایی نمرات اختلاف را توضیح دهد و موارد استفادهٔ آن را بگوید.

۱۲. پایایی روشهای رتبه‌بندی را توضیح دهد.

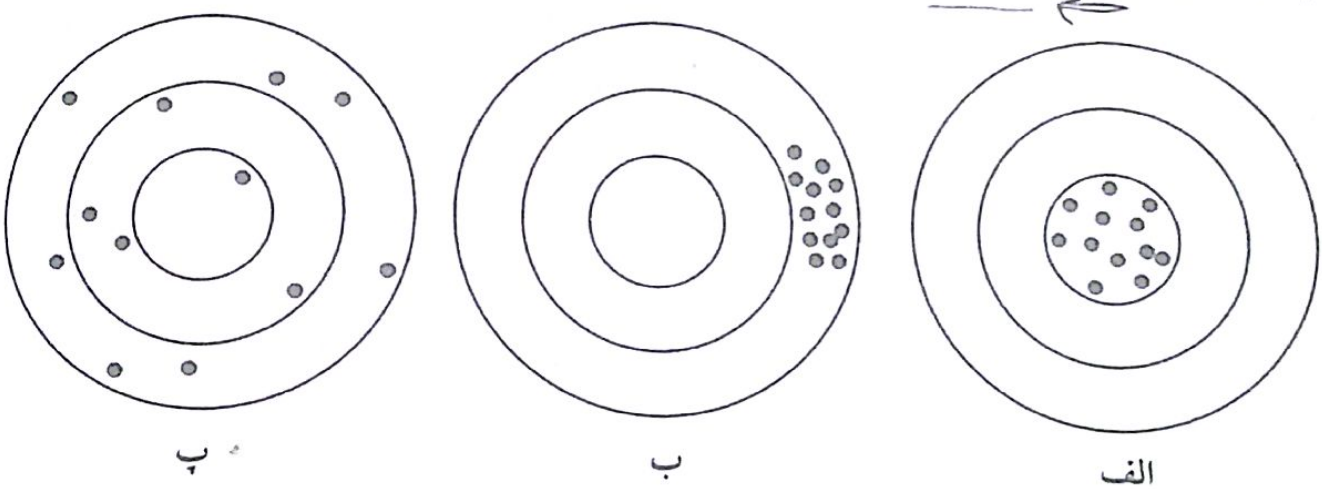
۱۳. فرمول اصلاح کاهش، روش افزایش همبستگی بین دو آزمون از طریق بالا بردن ضرایب پایایی آن آزمونها، روش محاسبهٔ پایایی نمرات اختلاف، و فرمول تعیین پایایی روشهای رتبه‌بندی را با داده‌هایی که در اختیار او گذاشته می‌شود به کاربندد.

در فصل پیش گفتیم که یک آزمون خوب باید از تعدادی ویژگی مطلوب مانند عینیت، سهولت اجرا، سهولت نمره‌گذاری، عملی بودن، سهولت تعبیر و تفسیر، روایی، و پایایی برخوردار باشد. همچنین گفته شد که مهم‌ترین این ویژگیها *روایی* و *پایایی* هستند. در فصل قبل روایی آزمون را معرفی کردیم؛ در این فصل پایایی آزمون را مورد بحث قرار می‌دهیم.

برای مقایسهٔ پایایی و روایی به شکل ۱-۲۱ نگاه کنید. این شکل نتایج تیراندازی با سه تفنگ مختلف را نشان می‌دهد. چنان که می‌بینید، نتایج تیراندازی با تفنگ **الف** هم پایا است و هم روا، زیرا همهٔ تیرها به هدف خورده‌اند. نتایج تیراندازی با تفنگ ب پایا است اما روا نیست، زیرا تیرها همه تقریباً به یک نقطه اصابت کرده‌اند اما آن نقطهٔ موردنظر تیرانداز نیست. نتایج تیراندازی با تفنگ پ نه پایا است و نه روا.

تعریف پایایی

پایایی[1] یک وسیلهٔ اندازه‌گیری عمدتاً به ثبات و دقّت نتایج حاصل از آن اشاره می‌کند. کاپلان و ساکوزا (۲۰۰۱) گفته‌اند "پایایی به دقت، اعتمادپذیری، ثبات، یا تکرارپذیری نتایج آزمون اشاره می‌کند" (ص ۱۱). رابطهٔ بین روایی و پایایی از این قرار است که یک آزمون باید **پایا**[2] باشد تا بتواند روا باشد. اگر آزمونی در هر بار اجرا بر روی تعدادی دانش‌آموز نتایج مختلفی به دست بدهد آن آزمون یک آزمون پایا نخواهد بود و در واقع هیچ چیز را به درستی اندازه نخواهد گرفت؛ و اگر یک آزمون چیزی را به درستی اندازه‌گیری نکند هیچ اطلاع مفیدی به ما نخواهد داد. برای مثال، یک آزمون حساب برای اندازه‌گیری محتوا و هدفهای درس تاریخ روا (مناسب) نیست، اما این آزمون می‌تواند مطالب حساب را که اندازه می‌گیرد با دقت (به طور پایا) اندازه‌گیری کند. پس، ***برای اینکه یک آزمون روا باشد باید نخست پایا باشد. یعنی پایایی شرط روایی است، اما روایی برای پایایی ضروری نیست.***



شکل ۱-۲۱ رابطهٔ بین روایی و پایایی (نتایج تیراندازی با سه تفنگ مختلف)

1. reliability
2. reliable

روشهای تعیین پایایی

روشهای تعیین پایایی متنوع‌اند و ما آنها را با استفاده از پیشنهادهای ایبل (۱۹۷۲)، ویرسما و جورس (۱۹۹۰)، مرفی و دیوید شوفر (۱۹۹۴)، و هوگان (۲۰۰۷) به شرح زیر دسته‌بندی می‌کنیم: (۱) روش پایایی بین مصححان یا مشاهده کنندگان، (۲) روش بازآزمایی، (۳) روش فرمهای موازی، و (۴) روشهای همسانی درونی. به توضیح این روشها توجه کنید.

پایایی بین مصححان

منظور از مصححان در اینجا دو یا چندنفری هستند که برگه‌های امتحانی را تصحیح می‌کنند، رفتارهای افراد یا فرایندها را مورد مشاهده قرار می‌دهند، یا محصولات یا فراورده‌های رفتار را درجه‌بندی می‌کنند. از این‌رو، به این نوع پایایی **پایایی بین مصححان**[1]، **پایایی بین مشاهده‌گران**[2]، و **پایایی بین درجه‌بندی‌کنندگان**[3] گفته می‌شود.

برای تعیین پایایی آزمونهای تشریحی یا به طور کلی آزمونهای غیرعینی که نمرات آنها تحت تأثیر قضاوت مصححان قرار می‌گیرد باید از دو یا چند مصحح که مستقلاً پاسخهای آزمون‌شوندگان را تصحیح می‌کنند استفاده کرد. همبستگی بین نمرات مصححان مختلف شاخص پایایی مصححان به حساب می‌آید. هرچه میزان توافق بین مصححان یا نمره گذاران بیشتر باشد این نوع پایایی مصححان نیز بیشتر است. بنابراین، پایایی مصححان یک آزمون تشریحی را به صورت "میزان توافق مصححان مختلف در نمره گذاری یک مجموعه برگهٔ آزمون تعریف می‌کنند" (ساکس، ۱۹۹۷، ص ۲۸۵). البته این پایایی مربوط به شیوهٔ کار تصحیح کنندگان برگه‌های امتحانی است و نباید آن را با پایایی آزمون که مربوط به سؤالهای خود آزمون است اشتباه گرفت. در این روش، ضریب همبستگی بزرگ بین نمرات مصححان مختلف میزان توافق‌نظر آنان را نشان می‌دهد. با این حال، باید دقت کرد که ضریب پایایی مصححان برای ضریب پایایی نمرات حاصل از اجرای آزمون یک سقف به حساب می‌آید. به عنوان مثال، اگر ضریب توافق بین مصححان مختلف ۰/۷۰ باشد، ضریب پایایی نمرات آزمون احتمالاً کمتر از ۰/۷۰ یا حداکثر ۰/۷۰ خواهد بود. (ما در پایان این فصل در بخش پایایی روشهای سنجش مستقیم در این‌باره توضیحات بیشتری داده‌ایم.)

پایایی آزمون

پایایی آزمون با روشهای مختلفی تعیین می‌شود. به توضیح این روشها در زیر توجه کنید.

---

1. interexaminer (interscorer) reliability
2. interobserver reliability
3. interrater reliability

۲ روش بازآزمایی

یکی از معروف‌ترین روشهای تعیین پایایی یک آزمون روش **بازآزمایی**[1] است. در این روش، آزمون را در دو نوبت (در فاصلهٔ حدود ۲ هفته) با گروه واحدی از آزمون‌شوندگان اجرا می‌کنند و نمره‌های حاصل را با هم مقایسه می‌نمایند. ضریب همبستگی بین نمره‌های به دست آمده از دوبار اجرای آزمون ضریب پایایی آزمون است.

روش بازآزمایی هرچند که یک روش ساده و قابل اجراست، اما متخصصان اندازه‌گیری و ارزشیابی به آن انتقاد کرده‌اند. از جمله گفته‌اند تجربهٔ نوبت اول آزمون سبب می‌شود که آزمون‌شوندگان با سؤالهای آزمون آشنا شوند و این آشنایی سبب خواهد شد که آنان در پاسخ دادن به همان سؤالها در نوبت دوم اجرای آزمون بهتر عمل کنند. همچنین گفته شده است که چون نمونهٔ سؤالهای به کار رفته در هر دو آزمون یک نمونهٔ واحد است، لذا نمرات حاصل از بازآزمایی هیچ‌گونه اطلاعی مبنی بر اینکه اگر نمونهٔ دیگری از سؤالهای مربوط به همان موضوع به کار روند چه تغییراتی حاصل خواهد شد به دست نمی‌دهد. همچنین گفته شده است که اگر فاصلهٔ زمانی بین دوبار اجرای آزمون زیاد باشد، ممکن است خطای اندازه‌گیری با تغییرات واقعی ایجاد شده در توانایی آزمون شوندگان که در نتیجهٔ یادگیری حاصل می‌شود در هم آمیزد. و انتقاد آخر اینکه اجرای دوبارهٔ یک آزمون مورد علاقه دانش‌آموزان و معلمان نیست و آن را کاری بیهوده می‌دانند که سبب اتلاف وقت آموزشی می‌شود؛ در نتیجه، فقدان علاقهٔ دانش‌آموزان به گرفتن آزمون مجدد ممکن است سبب شود که آنها در آزمون دوم نمرات کمتری از بار اول اجرای آزمون بیاورند. نکتهٔ دیگری را که باید خاطرنشان کرد این است که روش بازآزمایی بیشتر با آزمونهای هوش و شخصیت که در طول زمان چندان تغییر نمی‌کنند مفید است (کلاین، ۲۰۰۵).

۳ روش فرمهای موازی

روش دیگری که برای رفع مشکلات روش بازآزمایی پیشنهاد شده است روش **فرمهای موازی** یا **هم‌ارز**[2] است. در این روش دو آزمون معادل یا موازی برای یک موضوع تهیه می‌کنند و آنها را در فاصلهٔ زمانی کوتاهی (حدود دو هفته) به یک گروه واحد از آزمون شوندگان می‌دهند. دو فرم یک آزمون در صورتی موازی هستند که میانگین و واریانس آنها برابر باشند. هرچند عموماً تساوی میانگین و واریانس را برای آزمونهای موازی کافی می‌دانند، اما بعضی صاحب‌نظران، از جمله کلاین (۲۰۰۵)، معتقدند که برای داشتن آزمونهای موازی واقعی "علاوه بر میانگین و واریانس، پایایی آزمونها و رابطهٔ میان نمرات آزمونها و متغیرهای دیگر نیز باید برابر باشند" (ص ۹۴). ضریب همبستگی بین نمرات حاصل از اجرای فرمهای موازی، ضریب پایایی آن آزمونها به حساب می‌آید.

---

1. test-retest
2. equivalent forms

هر چند که در این روش تعیین پایایی برخی از مشکلات روش بازآزمایی رفع شده است، اما این روش دارای مشکلات مخصوص به خود است. مشکل عمدهٔ روش فرمهای هم‌ارز این است که آزمونهای پیشرفت تحصیلی، به ویژه آزمونهایی که معلمان برای ارزشیابی از یادگیری دانش‌آموزان خود می‌سازند، غالباً در یک فرم واحد تهیه می‌شوند، و تهیهٔ فرمهای موازی برای آنها به سادگی میسر نیست. حتی اگر چنین فرمهایی نیز تهیه شوند، باز هم مشکل بی‌علاقگی دانش‌آموزان و ایراد معلمان به اجرای دو فرم از یک آزمون باقی است. ایراد دیگری که به این روش تعیین پایایی می‌گیرند این است که در بعضی موارد آزمون شوندگان در جواب دادن به سؤالهای فرم اول آزمون با اصولی آشنا می‌شوند که عملکرد آنان را در فرم دوم تحت تأثیر قرار می‌دهد. برای رفع این مشکل می‌توان از روشهای همسانی درونی استفاده کرد.

روشهای همسانی درونی

هرچند که روشهای بازآزمایی و فرمهای موازی برای تعیین ضریب پایایی آزمونهایی که ویژگیهای نسبتاً ثابتی (مانند هوش و استعداد یا آموخته‌های افراد) را می‌سنجند روشهای کارآمدی هستند، اما برای تعیین پایایی ابزارهای سنجش ویژگیهای ناپایدار مانند شادمانی که از روزی به روز دیگر می‌تواند فرق کند مناسب نیستند. از آنجاکه ویژگی شادمانی یک ویژگی متغیر آدمهاست، ضریب پایایی حاصل از روش بازآزمایی یا روش فرمهای موازی تغییرپذیری ذاتی آن ویژگی را به صورت خطا نشان خواهد داد (راجرز، ۱۹۹۵). بنابراین، روشهای فوق برای تعیین پایایی ابزارهای اندازه‌گیری ویژگیهای متغیر که به آنها **متغیرهای گذرا**[1] می‌گویند مناسب نیستند. در مقابل، برای این‌گونه ویژگیها از روشهای دیگری استفاده می‌شود که به یک بار اجرای ابزار اندازه‌گیری اکتفا می‌کنند. این روشها **همسانی درونی**[2] نام دارند.

برخلاف روشهای قبلی که برثبات نتایج آزمون (روش بازآزمایی) یا برترازی نتایج آزمون (روش فرمهای هم‌ارز) تأکید دارند، در روش همسانی درونی بر همسانی یا یکنواختی ماده‌ها یا اجزای تشکیل دهندهٔ یک آزمون تأکید می‌شود. برای این منظور، آزمونی را که می‌خواهند ضریب پایایی‌اش را تعیین نمایند یک بار با گروه واحدی از آزمون شوندگان اجرا می‌کنند و بعد با یکی از روشهای زیر به تعیین ضریب پایایی آن می‌پردازند.

روش دو نیمه کردن آزمون

در روش **دو نیمه کردن آزمون**[3]، چنان‌که در بالا گفته شد، آزمون مورد نظر را یک بار با گروه واحدی از آزمون‌شوندگان اجرا می‌کنیم و پس از اجرا آن را به دو نیمه تقسیم می‌نماییم. در این روش،

1. ephemeral variables
2. internal consistency
3. split-halves

بهترین راه دو نیمه کردن آزمون این است که همهٔ سؤالهای فرد را یک آزمون به حساب آوریم، و همهٔ سؤالهای زوج را نیز آزمون دیگری بدانیم. ضریب همبستگی حاصل از نمرات دو نیمهٔ آزمون ضریب پایایی هریک از دو نیمه خواهد بود.

برای محاسبهٔ ضریب پایایی کل آزمون، ضریب همبستگی بین نیمه‌ها را در فرمول زیر که به **فرمول اسپیرمن ـ براون**[1] شهرت دارد قرار می‌دهیم. نتیجهٔ حاصل ضریب پایایی کل آزمون خواهد بود.

$$r_{tt}=\frac{۲r_{\frac{۱}{۲}\frac{۱}{۲}}}{۱+r_{\frac{۱}{۲}\frac{۱}{۲}}}$$

در این فرمول:

r_{tt} = ضریب پایایی کل آزمون

$r_{\frac{۱}{۲}\frac{۱}{۲}}$ = ضریب همبستگی بین نیمه‌های آزمون

به عنوان مثال، اگر ضریب همبستگی بین نیمه‌های یک آزمون برابر با ۰/۷۰ باشد، ضریب پایایی کل آزمون با استفاده از فرمول اسپیرمن ـ براون به نحو زیر محاسبه خواهد شد:

$$r_{tt}=\frac{۲\times ۰/۷}{۱+۰/۷}$$

$$=۰/۸۲$$

فرمول اسپیرمن ـ براون را، با جزیی تغییر، می‌توان برای محاسبهٔ ضریب پایایی یک آزمون، پس از آنکه تعدادی سؤال به آن افزوده شد، نیز مورد استفاده قرار داد. فرمول جدید به قرار زیر است:

$$r'_{tt}=\frac{kr_{tt}}{۱+(k-۱)r_{tt}}$$

در این فرمول:

r'_{tt} = ضریب پایایی آزمون جدید، معمولاً آزمون طولانی‌تر

r_{tt} = ضریب پایایی آزمون اولیه

k = نسبت تعداد سؤالهای آزمون جدید به تعداد سؤالهای آزمون اولیه

دقت کنید که k نسبت طول دو آزمون است نه تعداد سؤالهای جدیدی که به آزمون اولیه افزوده شده است. برای مثال، اگر طول آزمون از ۲۰ سؤال به ۳۰ سؤال افزایش یافته باشد،

$$k=\frac{۳۰}{۲۰}=۱/۵$$

برای نشان دادن چگونگی استفاده از فرمول اسپیرمن ـ براون برای برآورد ضریب پایایی

1. Spearman-Brown formula

آزمونی که تعدادی سؤال به آن افزوده‌ایم، فرض کنید آزمونی با ۳۰ سؤال در دست داریم که ضریب پایایی آن $r_{tt} = ۰/۶۰$ است. اگر تعداد سؤالهای این آزمون را به ۴۰ سؤال افزایش دهیم، ضریب پایایی آزمون جدید چند خواهد شد؟

$$k = \frac{۴۰}{۳۰} = ۱/۳۳$$

$$r'_{tt} = \frac{۱/۳۳(۰/۶۰)}{۱ + (۱/۳۳ - ۱)(۰/۶۰)} = \frac{۰/۸۰}{۱/۲۰} = ۰/۶۷$$

وقتی که روش دو نیمه کردن آزمون را برای محاسبهٔ ضریب پایایی کل آزمون مورد استفاده قرار می‌دهیم، چونکه آزمون را دو برابر می‌سازیم، $k = ۲$ است، در نتیجه، فرمول بالا به فرمولی که در ابتدای این قسمت معرفی کردیم تغییر خواهد کرد.

روش کودر - ریچاردسون

در روش **کودر - ریچاردسون**[1] نیز آزمون تنها یک بار اجرا می‌شود، اما در این روش همهٔ ماده‌های آزمون تحلیل می‌شوند. کودر و ریچاردسون برای بررسی همسانی درونی آزمون و تعیین پایایی آن دو فرمول مورد استفاده قرار داده‌اند که به KR۲۰ و KR۲۱ شهرت دارند. ما در اینجا این دو فرمول و نحوهٔ استفاده از آنها را به ترتیب توضیح می‌دهیم. برای این منظور، از داده‌های جدول ۱-۲۱ استفاده می‌کنیم. در این جدول نمره‌های شش دانش‌آموز در یک آزمون هشت سؤالی نشان داده شده‌اند.

جدول ۱-۲۱ داده‌های لازم برای استفاده از روش کودر - ریچاردسون در محاسبهٔ ضریب پایایی

		سؤالها								
دانش‌آموزان		۱	۲	۳	۴	۵	۶	۷	۸	کل
الف		۱	۱	۰	۱	۱	۱	۱	۱	۷
ب		۱	۱	۱	۱	۰	۰	۱	۰	۵
پ		۰	۱	۰	۱	۱	۰	۰	۱	۴
ت		۱	۱	۱	۰	۱	۰	۰	۰	۴
ث		۰	۱	۱	۱	۱	۱	۰	۰	۵
ج		۰	۰	۱	۱	۰	۰	۰	۰	۲
تعداد کل دانش‌آموزانی که به هر سؤال جواب درست داده‌اند		۳	۵	۴	۵	۴	۲	۲	۲	۲۷
نسبت درست	p	۰/۵۰	۰/۸۳	۰/۶۷	۰/۸۳	۰/۶۷	۰/۳۳	۰/۳۳	۰/۳۳	
نسبت غلط	q	۰/۵۰	۰/۱۷	۰/۳۳	۰/۱۷	۰/۳۳	۰/۶۷	۰/۶۷	۰/۶۷	
	pq	۰/۲۵	۰/۱۴	۰/۲۲	۰/۱۴	۰/۲۲	۰/۲۲	۰/۲۲	۰/۲۲	

$$\Sigma pq = ۱/۶۳ \quad \bar{x} = \frac{۲۷}{۶} = ۴/۵ \quad \sigma = ۱/۵ \quad \sigma^۲ = ۲/۲۵$$

1. Kuder-Richardson

KR۲۰

$$r_{tt} = \left[\frac{n}{n-1}\right]\left[1 - \frac{\Sigma pq}{\sigma^2}\right]$$

در این فرمول:

r_{tt} = ضریب پایایی آزمون

n = تعداد سؤالهای آزمون

σ^2 = واریانس یا مربع انحراف معیار کل سؤالهای آزمون

p = نسبت دانش‌آموزانی که به سؤال پاسخ درست داده‌اند

$q = 1 - p$ = نسبت دانش‌آموزانی که به سؤال پاسخ غلط داده‌اند

Σpq = مجموع pqها برای تمام سؤالها

با توجه به داده‌های جدول ۱ ـ ۲۱، ضریب پایایی (KR۲۰) آزمون به گونهٔ زیر محاسبه می‌شود:

$$r_{tt} = \left(\frac{۸}{۸-۱}\right)\left(۱ - \frac{۱/۶۳}{۲/۲۵}\right)$$

$$= ۱/۱۴ \times \frac{۰/۶۲}{۲/۲۵}$$

$$= ۱/۱۴ \times ۰/۲۷$$

$$= ۰/۳۱$$

فرمول KR۲۰ واریانس هر سؤال را برحسب pq محاسبه می‌کند. اگر بتوان فرض کرد که همهٔ سؤالهای آزمون از لحاظ درجهٔ دشواری متوسط و شبیه به هم هستند، می‌توان به جای Σpq، $n\bar{p}\bar{q}$ را قرار داد، که در آن n همچنان مُعرف تعداد سؤالهای آزمون، $\bar{p}$ میانگین دشواری سؤالها یا خارج‌قسمت میانگین کل آزمون ($\bar{x}$) بخش بر تعداد کل سؤالهای آزمون (n)، و $\bar{q}$ برابر است با $۱-\bar{p}$. دقت کنید که p ضریب دشواری سؤال است. بنابراین، اگر ۹۰ درصد دانش آموزان یک سؤال را درست جواب دهند ضریب دشواری آن سؤال ۰/۹۰ خواهد بود. به فرض اینکه تمام سؤالهای آزمون دارای ضریب دشواری متوسط باشند، می‌توان به جای فرمول KR۲۰ از فرمول KR۲۱ به شرح زیر استفاده کرد:

KR۲۱

$$r_{tt} = \left(\frac{n}{n-۱}\right)\left(۱ - \frac{n\bar{p}\bar{q}}{\sigma^2}\right)$$

در فرمول KR۲۱ همان علائم فرمول KR۲۰ موجودند، با این تفاوت که در فرمول KR۲۱ به جای Σpq، $n\bar{p}\bar{q}$ به کار رفته است که برای محاسبهٔ آن باید میانگین pها و qها را برای تمام سؤالها حساب کنیم و آن را در n، یعنی تعداد سؤالها، ضرب نماییم. با مراجعه به جدول ۱ـ۲۱، می‌توانیم میانگین آزمون (۴/۵) را به تعداد سؤالها (۸) تقسیم کنیم و $\bar{p}$ را محاسبه نماییم:

$$\bar{p} = \frac{\bar{x}}{n} = \frac{۴/۵}{۸} = ۰/۵۶$$

چون $q = ۱-p$ است، مقدار $\bar{q}$ نیز به شرح زیر محاسبه می‌شود:

$$\bar{q} = ۱ - \bar{p} = ۱ - ۰/۵۶ = ۰/۴۴$$

حال مقادیر لازم را در فرمول KR۲۱ قرار می‌دهیم و پایایی آزمون را به شرح زیر حساب می‌کنیم:

$$r_{tt} = \frac{۸}{۷}\left(۱ - \frac{۸(۰/۵۶ \times ۰/۴۴)}{۲/۲۵}\right)$$
$$= ۱/۱۴ \times \frac{۰/۲۸}{۲/۲۵}$$
$$= ۰/۱۴$$

آن‌گونه که می‌بینید، ضریب حاصل از فرمول KR۲۱ کمتر از ضریب حاصل از فرمول KR۲۰ است. دلیلش آن است که شرط متوسط و مشابه بودن درجات دشواری سؤالها برآورده نشده است.

چنان‌که گفتیم، در استفاده از فرمول KR۲۱ فرض براین است که همهٔ سؤالها یا ماده‌های آزمون از لحاظ درجهٔ دشواری شبیه به هم هستند. هر قدر از لحاظ درجهٔ دشواری میان سؤالهای آزمون تفاوت وجود داشته باشد، به همان نسبت از ضریب پایایی حاصل از فرمول KR۲۱ کاسته می‌شود. اگر فرض شباهت درجهٔ دشواری سؤالها صحیح باشد، استفاده از این فرمول بسیار ساده‌تر از فرمول KR۲۰ است، زیرا در استفاده از فرمول KR۲۱ نیازی به محاسبهٔ نسبت پاسخهای درست برای تک تک سؤالهای آزمون نیست و تنها با داشتن سه جزء اطلاع زیر، یعنی تعداد سؤالها (n)، میانگین نمرات آزمون ($\bar{x}$)، و واریانس کل آزمون ($\sigma^۲$)، محاسبهٔ ضریب پایایی میسر است. در چنین حالتی، در فرمول KR۲۱ به جای $\frac{n\bar{p}\bar{q}}{\sigma^۲}$، مقدار $\frac{\bar{x}(n-\bar{x})}{n\sigma^۲}$ را می‌گذاریم و فرمول زیر را به دست می‌آوریم.

KR۲۱

$$r_{tt} = \frac{n}{n-۱}\left[۱ - \frac{\bar{x}(n-\bar{x})}{n\sigma^۲}\right]$$

این فرمول صورت دیگر فرمول KR۲۱ است که در محاسبهٔ آن سه عامل تعداد سؤالها (n)، میانگین آزمون ($\bar{x}$) و انحراف معیار (σ)، کافی است.

به عنوان مثالی از کاربرد این فرمول، فرض کنید می‌خواهیم ضریب پایایی یک آزمون ۵۰ سؤالی ($n = ۵۰$) را که دارای میانگین ۳۰ ($\bar{x} = ۳۰$) و انحراف معیار ۶ ($\sigma = ۶$) است محاسبه نماییم. برای این منظور، طبق فرمول جدید KR۲۱، به گونهٔ زیر عمل می‌کنیم.

$$r_{tt} = \frac{۵۰}{۵۰ - ۱}\left[۱ - \frac{۳۰(۵۰ - ۳۰)}{۵۰ \times ۳۶}\right]$$

$$= \frac{۵۰}{۴۹}\left(۱ - \frac{۶۰۰}{۱۸۰۰}\right)$$

$$= ۰/۶۸$$

فرض اساسی فرمولهای کودر و ریچاردسون این است که سؤالهای آزمون را می‌توان به صورت ۱ برای پاسخ درست و ۰ (صفر) برای پاسخ غلط تصحیح کرد. بنابراین، نمی‌توان این فرمولها را برای محاسبهٔ ضریب پایایی پرسشنامه‌هایی که پاسخهای آنها از ۱ تا ۴ یا ۵ را شامل می‌شوند مورد استفاده قرار داد.

روش ضریب آلفا

سومین روش تعیین پایایی آزمون با تأکید بر همسانی درونی روش **ضریب آلفا**[1] نام دارد که به آن *ضریب آلفا* یا *ضریب آلفای کرانباخ* یا حتی *آلفا* نیز گفته می‌شود. کرانباخ (۱۹۵۱) واضع این روش تعیین پایایی است. در این روش، اجزا یا قسمتهای آزمون برای سنجش ضریب پایایی آزمون به کار می‌روند. اگر قسمتهای آزمون همان سؤالهای آزمون باشند و سؤالها به صورت درست (۱) و غلط (۰) تصحیح بشوند، ضریب آلفا برابر خواهد بود با KR۲۰ که در صفحات قبل معرفی شد. اما اگر، به جای سؤالها یا ماده‌ها، آزمون از بخشها یا قسمتهایی تشکیل شده باشد، مثل زمانی که یک آزمون از تعدادی خرده آزمون تشکیل یافته است، و بخواهیم از آنها در محاسبهٔ ضریب پایایی کل آزمون استفاده کنیم، آنگاه از روش کرانباخ مطابق با فرمول زیر استفاده می‌کنیم.

در این فرمول:

$$ra = \frac{j}{j - ۱}\left(۱ - \frac{\Sigma \sigma_j^{۲}}{\sigma^{۲}}\right)$$

ra = ضریب پایایی کل آزمون

j = تعداد بخشهای آزمون

$\sigma_j^{۲}$ = واریانس نمرات هر بخش آزمون

$\sigma^{۲}$ = واریانس نمرات کل آزمون

1. coefficient alpha

برای توضیح بیشتر چگونگی استفاده از ضریب آلفا، فرض کنید یک آزمون از سه خرده آزمون تشکیل شده است و هریک از این خرده آزمونها نیز شامل تعدادی سؤال است. واریانس این سه خرده آزمون برابرند با:

$\hat{\sigma}_1^2 = ۶$ $\hat{\sigma}_2^2 = ۴$ $\hat{\sigma}_3^2 = ۷$

همچنین واریانس کل آزمون برابر است با: $\sigma^2 = ۳۲$

با توجه به اینکه $j = ۳$ است، ضریب آلفا به صورت زیر محاسبه می‌شود:

$$r_\alpha = \frac{j}{j-۱}\left(۱ - \frac{\Sigma\sigma_j^2}{\sigma^2}\right)$$

$$r_\alpha = \frac{۳}{۳-۱}\left(۱ - \frac{۶+۴+۷}{۳۲}\right)$$

$$= \left(\frac{۳}{۲}\right)\left(۱ - \frac{۱۷}{۳۲}\right)$$

$$= ۰/۷۰$$

روش آلفای کرانباخ در شرایطی که نمرات سؤال نه به صورت دوارزشی صفر و ۱ بلکه به صورت چندارزشی تعیین می‌شود و KR۲۰ برای آن مناسب نیست نیز قابل استفاده است ــ مثلاً در رابطه با نمرات حاصل از پرسشنامه‌های شخصیت یا نگرش‌سنج که پاسخ‌دهنده به هر سؤال در طیفی از مثلاً *کاملاً موافق* تا *کاملاً مخالف* (۵ تا ۱) جواب می‌دهد و هیچ‌یک از آنها درست یا غلط محسوب نمی‌شود. بدین‌منظور، از فرمول زیر استفاده می‌شود:

$$r_\alpha = \frac{k}{k-۱}\left(۱ - \frac{\Sigma\sigma^2 i}{\sigma^2 t}\right)$$

در این فرمول، k برابر است با تعداد سؤالهای آزمون، $\sigma^2 i$ برابر است با مجموع واریانس سؤالها، و $\sigma^2 t$ برابر است با واریانس آزمون.

برای روشن ساختن چگونگی استفاده از آلفای کرانباخ، از اطلاعات جدول ۲-۲۱ استفاده می‌کنیم. این جدول نمرات ۱۰ نفر به ۵ سؤال یک پرسشنامهٔ نظرسنجی را نشان می‌دهد. دامنهٔ نمرات از ۵ (کاملاً موافق) تا ۱ (کاملاً مخالف) است.

$$r_\alpha = \frac{۵}{۵-۱}\left(۱ - \frac{۵/۱۷}{۶/۴}\right)$$

$$r_\alpha = \frac{۵}{۴}(۱ - ۰/۸۰) = ۱/۲۵ \times ۰/۲ = ۰/۲۵$$

ضریب آلفا را به گونهٔ دیگری نیز می‌توان تفسیر کرد. به یاد دارید که در فصل هفدهم در رابطه با تحلیل سؤالهای آزمون گفتیم که همبستگی میان سؤالهای آزمون مستقیماً به پایایی آزمون مربوط است. در واقع روشهای محاسبهٔ پایایی آزمون براساس همسانی درونی ضریب پایایی را با

جدول ۲-۲۱ داده‌های لازم برای استفاده از روش ضریب آلفا در محاسبهٔ ضریب پایایی

افراد	سؤال ۱	سؤال ۲	سؤال ۳	سؤال ۴	سؤال ۵	نمرهٔ کل
۱	۳	۵	۱	۴	۱	۱۴
۲	۴	۴	۳	۵	۳	۱۹
۳	۳	۴	۴	۴	۴	۱۹
۴	۳	۳	۵	۲	۱	۱۴
۵	۳	۴	۵	۴	۳	۱۹
۶	۴	۵	۵	۳	۲	۱۹
۷	۲	۵	۵	۳	۴	۱۹
۸	۳	۴	۴	۲	۱۴	۱۷
۹	۳	۵	۴	۴	۳	۱۹
۱۰	۳	۳	۲	۳	۲	۱۳
واریانس	۰/۳۲	۰/۶۲	۱/۹۶	۰/۹۳	۱/۳۴	۶/۴

مجموع واریانسهای سؤالها = ۵/۱۷

توجه به تعداد سؤالهای آزمون و میانگین همبستگی میان سؤالها برآورد می‌نماید. این دو عامل با هم ترکیب می‌شوند و فرمول زیر را به دست می‌دهند.

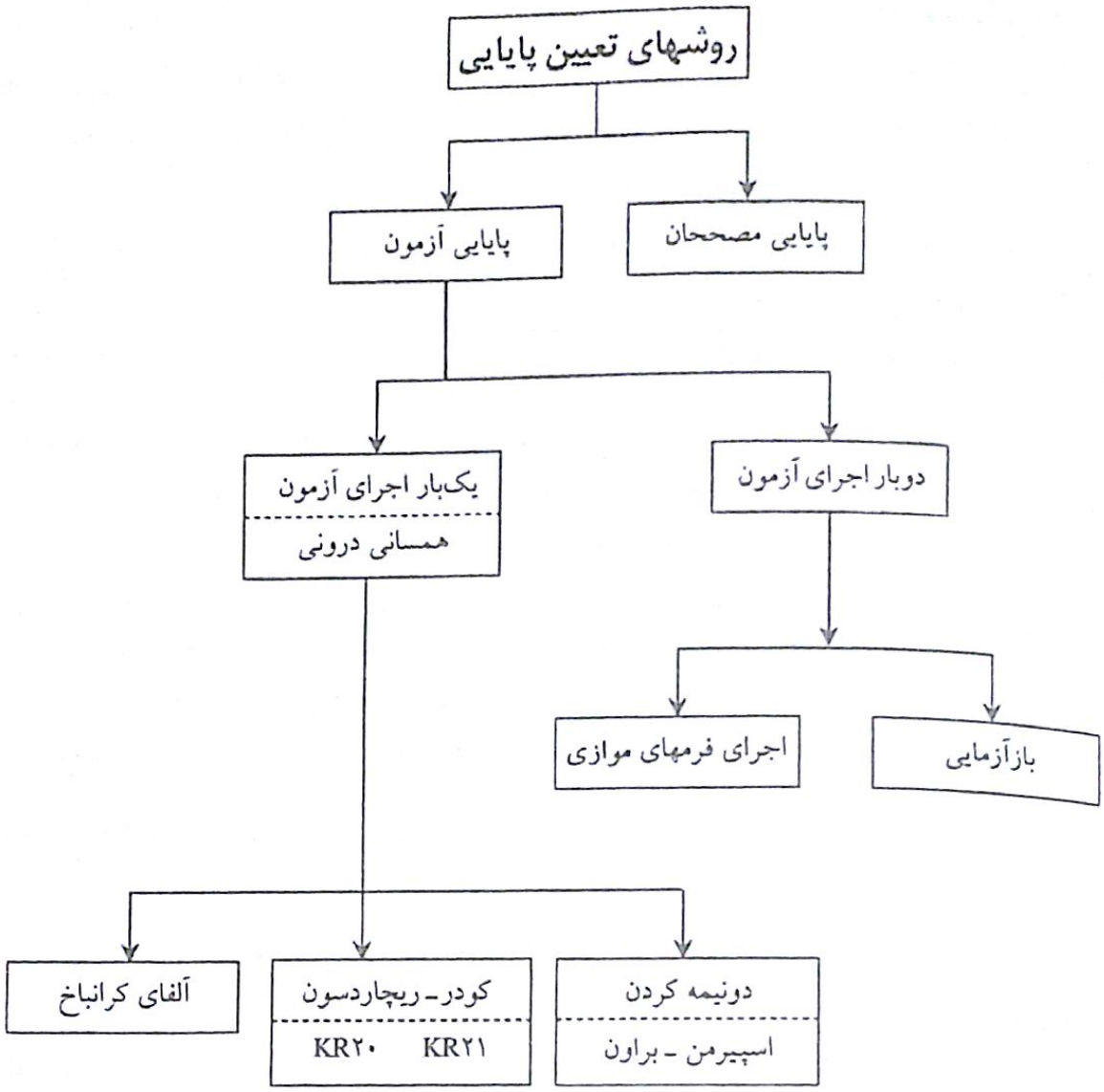
$$r_{xx} = \frac{k\ (\bar{r}_{ij})}{۱ + (k - ۱)\bar{r}_{ij}}$$

در این فرمول، k تعداد سؤالهای آزمون، r_{ij} مُعرف ضریب همبستگی هریک از سؤالهای آزمون با سؤال دیگر همان آزمون، و $\bar{r}_{ij}$ نشان دهندهٔ میانگین همهٔ ضریبهای همبستگی میان سؤالهاست. اگر این فرمول ضریب آلفا را با فرمولِ قبلی ضریب آلفا مقایسه کنیم، خواهیم دید که در فرمول اخیر ضریب همبستگی هر سؤال آزمون با سؤال دیگر واحد محاسبهٔ ضریب پایایی است، در حالی که در فرمول قبلی واریانس سؤالهای مختلف آزمون مورد استفاده قرار می‌گیرد.

فرض‌کنید یک آزمون ۲۰ سؤالی در دست داریم که میانگین ضریبهای همبستگی میان همهٔ سؤالهای آن برابر با ۰/۳۰ است. برای محاسبهٔ ضریب α با فرمول بالا، به گونهٔ زیر عمل می‌کنیم:

$$\alpha = \frac{۲۰(۰/۳۰)}{۱+(۲۰-۱)(۰/۳۰)} = ۰/۸۹۶$$

اگر روش ضریب آلفا و روش دو نیمه کردن آزمون را با هم مقایسه کنیم، خواهیم دید که روش دو نیمه کردن آزمون نوعی روش ضریب آلفاست. به سخن دیگر، ضریب آلفا مُعرف میانگین ضریبهای پایایی حاصل از همهٔ دو نیمه‌های ممکن یک آزمون است.



شکل ۲-۲۱ روشهای مختلف تعیین پایایی

حال که روشهای مختلف تعیین پایایی را توضیح دادیم، می‌توانیم آنها را در رابطه با یکدیگر نشان دهیم. در شکل ۲-۲۱ همین کار انجام گرفته است.

روش تعیین پایایی آزمونهای سرعت

در رابطه با استفاده از ضریب پایایی، آزمونها را می‌توان به دو گروه **آزمون قدرت**[1] و **آزمون سرعت**[2] دسته‌بندی کرد. آزمون قدرت آزمونی است که در اجرای آن به آزمون شوندگان وقت کافی داده می‌شود تا به همهٔ سؤالهای آن جواب دهند. این‌گونه آزمونها شامل سؤالهای چالش‌انگیزند و هدف اصلی آنها این است که سطح عملکرد یا میزان دانش و توانایی آزمون‌شوندگان را بسنجند. "سطح دشواری سؤالهای آزمون قدرت

1. power test 2. speed test

در طول آزمون به تدریج افزایش می‌یابد به گونه‌ای که بعضی سؤالها تا حدی دشوار می‌شوند که هیچ‌یک از آزمون‌شوندگان نمی‌توانند به آنها جواب درست بدهند" (آناستازی[1] و اُربینا[2]، ۲۰۰۷، ص ۱۰۲). در مقابل، آزمونهای سرعت به آزمونهایی گفته می‌شود که از سؤالهای بسیار آسان تشکیل شده و هدف آنها صرفاً سنجش سرعت عمل آزمون‌شوندگان در پاسخ دادن به سؤالهاست. برای جواب دادن به سؤالهای این آزمونها محدودیت زمانی وجود دارد، و اکثر آزمون شوندگان در محدودهٔ زمانی تعیین شده نمی‌توانند آن را به پایان برسانند. از جملهٔ این آزمونها، می‌توان به آزمونی اشاره کرد که شامل ۵۰ سؤال محاسباتی از قبیل ?=۱۷+۴۲، ?=۳۲×۱۵، و مانند اینهاست. آناستازی و اُربینا، در رابطه با آزمونهای قدرت و سرعت گفته‌اند:

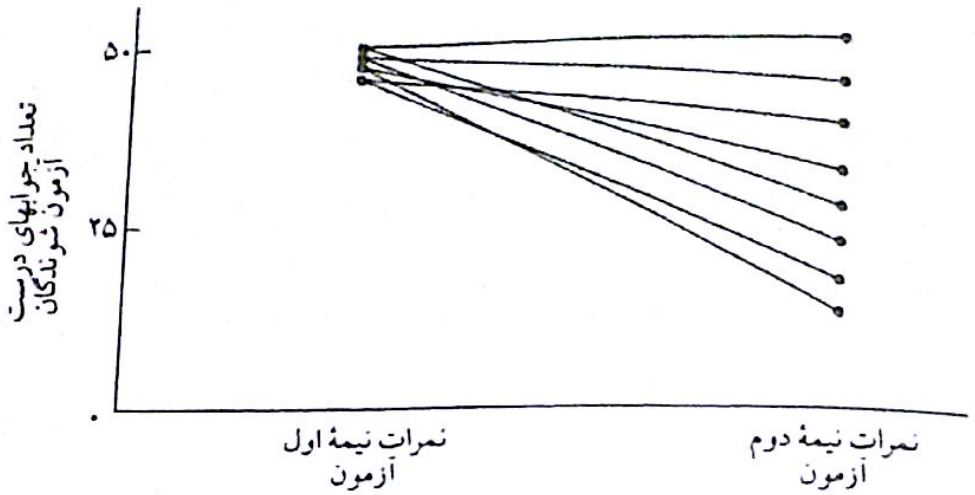
> هم آزمون قدرت و هم آزمون سرعت به گونه‌ای طرح می‌شوند که دستیابی آزمون‌شوندگان را به نمرهٔ کامل محدود می‌کند. دلیل این امر آن است که نمرات کامل نامعلوم‌اند، زیرا غیرممکن است بدانیم که اگر سؤالهای بیشتر یا دشوارتری به آزمون اضافه شود نمرات چقدر افزایش می‌یابند. برای اینکه معلوم نماییم که هریک از آزمون‌شوندگان چه مقدار دانش یا توانایی دارد، آزمون باید از لحاظ تعداد سؤالها یا سطح دشواری یک سقف مناسب داشته باشد. یک استثنا برای این مورد آزمونهای در حد تسلط‌اند. هدف آزمونهای درحد تسلط این نیست که حد توانایی افراد را تعیین کند، بلکه هدف آنها این است که تعیین کند آزمون‌شونده به یک سطح از پیش تعیین شده دست یافته است یا نه. (ص ۱۰۲)

در ارتباط با روشهای سنجش پایایی، هیچ یک از روشهای دو نیمه کردن یا کودر - ریچاردسون برای تعیین پایایی آزمونهای سرعت مناسب نیست. دلیل این امر واضح است. در آزمونهای سرعت، آزمون شوندگان تقریباً همهٔ سؤالهایی را که وقت رسیدگی به آنها را دارند درست جواب می‌دهند، و تنها تفاوت بین آنان در *تعداد* سؤالهایی است که جواب می‌دهند. بنابراین، اگر نمرات نیمهٔ اول آزمون را با نمرات نیمهٔ دوم آزمون همبسته سازیم ضریب حاصل یک ضریب بسیار کوچک خواهد بود، زیرا همهٔ آزمون شوندگان سؤالات نیمهٔ اول را درست جواب می‌دهند، در حالی که در نیمهٔ دوم آزمون تفاوتهای زیادی میان نمرات مشاهده خواهد شد. این وضعیت در شکل ۱ - ۲۱ نشان داده شده است. از سوی دیگر، اگر به جای نیمهٔ اول و نیمهٔ دوم آزمون، سؤالهای فرد و زوج را جدا کنیم و همبستگی بین آنها را به دست آوریم، این ضریب خیلی بزرگ خواهد بود، زیرا برای هریک از آزمون شوندگان نمرات فرد و زوج تقریباً یکسان است.

روش دیگر محاسبهٔ ضریب پایایی آزمونهای سرعت، روش بازآزمایی است. با این حال، از آنجا که یادآوری مطالب و تمرین روی مطالب سبب افزایش تفاوت بین عملکرد آزمون شوندگان در دو بار اجرای یک آزمون واحد می‌شود، بهترین روش برآورد پایایی آزمونهای سرعت این است که دو فرم هم‌ارز یا موازی درست کنیم و این دو فرم را دوبار، اما بلافاصله بعد از یکدیگر، به اجرا در آوریم و ضریب همبستگی حاصل بین نمرات دو فرم را حساب کنیم.

---

1. Anastasi 2. Orbina



شکل ۱ - ۲۱ توزیع نمرات نیمه‌های اول و دوم آزمون سرعت

پایایی نمرات اختلاف

زمانی پیش می‌آید که بخواهیم پایایی **نمرات اختلاف**[1] را به دست آوریم. مهم‌ترین نمرهٔ اختلاف مورد استفاده در آموزش اختلاف بین نمره‌های دانش آموزان در یک آزمون واحد یا در دو فرم موازی یک آزمون است که در دو زمان مختلف اجرا شده‌اند. نام دیگر این نمرهٔ اختلاف **نمرهٔ افزایش**[2] یا **نمرهٔ تغییر**[3] است. علاوه بر این نمرهٔ اختلاف، می‌توان از نمرات اختلاف دیگری نام برد، نظیر نمرهٔ اختلاف بین عملکرد دو گروه دانش‌آموز در یک آزمون یا اختلاف بین نمرات یک دسته دانش‌آموز در آزمونهای مختلف.

برای درک بیشتر مفهوم نمرهٔ اختلاف یا نمرهٔ تفاوت، فرض کنید می‌خواهیم بدانیم که دانش‌آموزان در طول یک سال تحصیلی در یک درس (مثلاً فیزیک) چه مقدار یادگیری داشته‌اند. برای این منظور، در آغاز سال تحصیلی به آنها یک پیش‌آزمون (آزمون ۱) می‌دهیم، و در پایان سال تحصیلی نیز یک آزمون معادل یا پس‌آزمون (آزمون ۲) به آنها می‌دهیم تا جواب بدهند و بعد تفاوت بین نمرات دو آزمون را به دست می‌آوریم (d). به این تفاوتْ نمرهٔ اختلاف می‌گویند که پایایی آن را با فرمول زیر به دست می‌آوریم:

$$r_d = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{12}}$$

در این فرمول:

r_d = پایایی نمرهٔ اختلاف

r_{11} = پایایی آزمون اول

r_{22} = پایایی آزمون دوم

r_{12} = همبستگی بین دو آزمون

1. difference scores
2. gain score
3. change score

باز هم فرض کنید دو آزمون مورد نظر ما دارای همبستگی ۰/۸۰ است. همچنین ضریب پایایی آزمون اول ۰/۹۰ و ضریب پایایی آزمون دوم ۰/۸۵ است. بنا به فرمول، ضریب پایایی نمرات اختلاف برابر با ۰/۴۰ خواهد بود:

$$r_d = \frac{\frac{۰/۹۰ + ۰/۸۵}{۲} - ۰/۸۰}{۱ - ۰/۸۰}$$

$$= \frac{۰/۸۸ - ۰/۸۰}{۰/۲۰}$$

$$= \frac{۰/۰۸}{۰/۲۰}$$

$$= ۰/۴۰$$

در مثال بالا، چنان‌که دیده می‌شود، ضریب پایایی نمرهٔ اختلاف (۰/۴۰) بسیار کوچک‌تر از ضریبهای پایایی هر یک از دو آزمون (۰/۹۰ و ۰/۸۶) است. علت آن است که ضریب همبستگی بین دو آزمون بزرگ است (۰/۸۰). قاعدهٔ کلی این است که، در صورت تساوی بقیهٔ شرایط، *هر چه همبستگی بین دو آزمون بیشتر باشد، به همان نسبت پایایی نمرهٔ اختلاف آن دو آزمون کمتر است.*

بنابر آنچه گفته شد، با کاهش میزان ضریب همبستگی بین دو آزمون ضریب پایایی نمرات اختلاف افزایش می‌یابد. برای نشان دادن این مطلب، فرض کنید ضریب همبستگی بین دو آزمون فوق، به جای ۰/۸۰، ۰/۳۰ باشد. در این صورت ضریب پایایی نمرات اختلاف از ۰/۴۰ به ۰/۸۲ افزایش خواهد یافت، مطابق محاسبات زیر:

$$r_d = \frac{\frac{۰/۹۰ + ۰/۸۵}{۲} - ۰/۳۰}{۱ - ۰/۳۰}$$

$$= \frac{۰/۵۸}{۰/۷۰}$$

$$= ۰/۸۲$$

عوامل مؤثر در افزایش پایایی آزمون (وابسته به هنجار)

عوامل زیر بر پایایی آزمون مؤثرند و می‌توان از آنها برای افزایش ضریب پایایی آزمون استفاده کرد.

۱. با افزودن بر تعداد سؤالات یک آزمون، پایایی آن نیز افزایش می‌یابد.
۲. با متجانس‌تر و همگون‌تر کردن سؤالات یک آزمون، پایایی آن بیشتر می‌شود.
۳. با افزودن تعدادی سؤال با ضریب تمیز زیاد به یک آزمون، پایایی آن بالا می‌رود.
۴. با افزودن تعدادی سؤال دارای ضریب دشواری متوسط به یک آزمون، پایایی آن افزایش می‌یابد.
۵. اگر یک آزمون را در مورد گروهی از افراد که از نظر توانایی مورد سنجش تجانس کمتری با یکدیگر دارند اجرا کنیم، آزمون پایایی بیشتری نشان خواهد داد.

ع. آزمونهای سرعت معمولاً از آزمونهایی که سرعت عملکرد در آنها عامل مهمی به حساب نمی‌آید (آزمونهای قدرت) ضریب پایایی بیشتری را نشان می‌دهند.

استفاده از ضریب پایایی آزمون برای برآورد نمرهٔ حقیقی

مفهوم پایایی آزمون بر این باور استوار است که هیچ آزمونی کاملاً پایا نیست، لذا عموماً با نمره‌های هر آزمون قدری خطا همراه است. این خطا که بخشی از نمرهٔ به دست آمده یا نمرهٔ مشاهده شده (نمرهٔ خام) است، جزء خطای نمره گفته می‌شود. عوامل مختلفی در پدید آمدن این جزء خطا دخیل‌اند و بنابراین واریانس خطا[1] یک واریانس تصادفی یا غیرمنظم است.

هر نمرهٔ مشاهده شده‌ای از دو جزء تشکیل یافته است. چنان‌که گفتیم، یک جزءِ هر نمرهٔ مشاهده شده خطای آن است؛ جزء دیگر آن جزء حقیقی نمره است. پس، علاوه بر واریانس خطا، یک واریانس حقیقی[2] نیز وجود دارد که، برخلاف واریانس خطا، واریانس منظم و غیرتصادفی است. واریانس حقیقی منعکس کنندهٔ توانایی مورد سنجش است، در حالی که واریانس خطا ناشی از یک عامل تصادفی است، مانند اشتباهات نمره‌گذاری (نگاه کنید به جدول ۲-۲۱).

جدول ۲-۲۱ توزیع تعدادی نمرهٔ مشاهده شده، نمرهٔ حقیقی، نمرهٔ خطا و میانگینها، واریانسها، و انحراف معیارهای آنها.

آزمون‌شونده	نمرهٔ مشاهده شده	=	نمرهٔ حقیقی	+	نمرهٔ خطا
الف	۳		۵		-۲
ب	۱۷		۱۵		+۲
پ	۱۶		۲۰		-۴
ت	۲۳		۲۵		-۲
ث	۲۷		۲۵		+۲
ج	۲۵		۲۵		۰
چ	۳۵		۲۵		+۱۰
ح	۲۶		۳۰		-۴
خ	۳۳		۳۵		-۲
د	۴۵		۴۵		۰
جمع	۲۵۰		۲۵۰		۰
میانگین	۲۵		۲۵		۰
واریانس (σ^2)	۱۲۰/۲		۱۰۵		۱۵/۲
انحراف معیار (SD)	۱۰/۹		۱۰/۲		۳/۹

1. error variance
2. true variance

در تفسیر نمره‌های یک آزمون همواره مفاهیم زیر مورد استفاده قرار می‌گیرند: **نمرهٔ مشاهده شده**[1] که همان نمرهٔ خام است، **نمرهٔ حقیقی**[2] و **نمرهٔ خطا**[3]. نمرهٔ مشاهده شده یا نمرهٔ به دست آمده نمره‌ای است که یک آزمون‌شونده در آزمون به دست می‌آورد. نمرهٔ حقیقی به میانگین فرضی تعداد بیشماری آزمون گفته می‌شود که در مورد یک آزمون‌شونده اجرا می‌شوند و فرض بر این است که تمرین بر نمرات پی‌درپی کسب شده تأثیر ندارد. تفاوت بین نمرهٔ مشاهده شده و نمرهٔ حقیقی نمرهٔ خطا یا خطای اندازه‌گیری نام دارد.

اگر X_o معرف نمرهٔ مشاهده شده، X_t نشان دهندهٔ نمرهٔ حقیقی، و X_e علامت نمرهٔ خطا باشند، تساوی زیر را خواهیم داشت:

نمرهٔ خطا + نمرهٔ حقیقی = نمرهٔ مشاهده شده

$$X_o = X_t + X_e$$

به عنوان مثال، با استفاده از داده‌های ردیف اول جدول ۲-۲۱، نمرهٔ مشاهده شده (۳) برابر است با نمرهٔ حقیقی (۵) به اضافهٔ نمرهٔ خطا (۲-) مطابق زیر:

$$۳ = ۵ + (-۲)$$

$$۳ = ۳$$

اگر واریانس نمرات مشاهده شدهٔ آزمونی را با $\sigma_o^۲$، واریانس نمرات حقیقی را با $\sigma_t^۲$، و واریانس نمرات خطای آن را با $\sigma_e^۲$ نشان دهیم، رابطهٔ بین این واریانسها به شرح زیر خواهد بود:

$$\sigma_o^۲ = \sigma_t^۲ + \sigma_e^۲$$

طبق اطلاعات موجود در جدول ۲-۲۱،

$$۱۲۰/۲ = ۱۰۵ + ۱۵/۲$$

حال می‌توانیم پایایی آزمون را به صورت نسبت ***واریانس نمرهٔ حقیقی به واریانس نمرهٔ مشاهده شده*** تعریف کنیم، یعنی به صورت زیر:

$$r_{tt} = \frac{\sigma_t^۲}{\sigma_o^۲} \quad \text{(پایایی آزمون)}$$

همچنین، با توجه به تساوی $\sigma_o^۲ = \sigma_t^۲ + \sigma_e^۲$، پایایی را می‌توان به صورتهای زیر نشان داد:

$$r_{tt} = \frac{\sigma_t^۲}{\sigma_t^۲ + \sigma_e^۲} \qquad r_{tt} = ۱ - \frac{\sigma_e^۲}{\sigma_o^۲}$$

اگر در فرمولهای بالا داده‌های جدول ۲-۲۱ را قرار دهیم خواهیم دید که همواره رقم ثابتی به دست خواهد آمد.

$$r_{tt} = \frac{\sigma_t^۲}{\sigma_o^۲} = \frac{۱۰۵}{۱۲۰/۲} = ۰/۸۷$$

$$r_{tt} = \frac{\sigma_t^۲}{\sigma_t^۲ + \sigma_e^۲} = \frac{۱۰۵}{۱۰۵ + ۱۵/۲} = ۰/۸۷$$

---

1. observed score
2. true score
3. error score

$$r_{tt} = ۱ - \frac{\sigma_e^{۲}}{\sigma_o^{۲}} = ۱ - \frac{۱۵/۲}{۱۲۰/۲} = ۰/۸۷$$

این فرمولها همچنین نشان می‌دهند که ضریب پایایی می‌تواند ارقام بین صفر تا ۱ + را شامل شود. ضریب پایایی صفر نشان دهندهٔ این است که تمامی واریانس مشاهده شده واریانس خطاست. در مقابل، طبعاً ضریب پایایی ۱ معرف این است که صددرصد واریانس مشاهده شده واریانس حقیقی است.

خطای معیار اندازه‌گیری

یکی از فواید ضریب پایایی آزمون این است که با استفاده از آن می‌توان **خطای معیار اندازه‌گیری**[1] آزمون را تعیین کرد. خطای معیار اندازه‌گیری برابر است با *انحراف معیار توزیع نمرات خطا*. برای روشن شدن مفهوم خطای معیار اندازه‌گیری، مثالی را که ثرندایک و هیگن (۱۹۷۹) در این باره ذکر کرده‌اند می‌آوریم. فرض کنید از شخصی بخواهیم تا توپی را ۲۰۰ بار پرتاب کند و فرض کنید که انجام دادن این کار بدون تأثیر پذیرفتن شخص از خستگی یا تمرین ممکن باشد. نتیجهٔ این عمل یک توزیع فراوانی از فواصلی خواهد بود که از این ۲۰۰ بار پرتاب شدن توپ حاصل می‌شود. این توزیع فراوانی مربوط به نمرات مشاهده شده است که دارای یک میانگین است که ما می‌توانیم آن را به‌طور تقریب نمرهٔ حقیقی آن شخص در پرتاب کردن توپ بدانیم. این توزیع همچنین یک انحراف معیار دارد که نحوهٔ پراکندگی نمرات را نشان می‌دهد. از آنجا که این انحراف معیارِ خطاهای پرتاب توپ به وسیلهٔ آن شخص است، به آن خطای معیار اندازه‌گیری می‌گویند. بنابراین، فقدان دقّت نمرات یعنی فقدان دقتی که از ناپایایی آزمون ناشی می‌شود، در خطای معیار اندازه‌گیری انعکاس می‌یابد.

در مورد متغیرهای روانی و پرورشی، معمولاً نمی‌توان آنها را کراراً اندازه گرفت اما با استفاده از ضریب پایایی آزمون می‌توان خطای معیار اندازه‌گیری آزمون را برآورد کرد. فرمول برآورد خطای معیار اندازه‌گیری با استفاده از ضریب پایایی آزمون به شرح زیر محاسبه می‌شود:

پیش‌تر گفتیم که:

$$r_{tt} = ۱ - \frac{\sigma_e^{۲}}{\sigma_o^{۲}}$$

می‌توان این تساوی را طی انجام مراحل زیر به فرمول خطای معیار اندازه‌گیری تبدیل کرد.

$$\frac{\sigma_e^{۲}}{\sigma_o^{۲}} = ۱ - r_{tt} \qquad \sigma_e^{۲} = \sigma_o^{۲}\,(۱ - r_{tt})$$

---

1. standard error of measurment

از آنجاکه خطای معیار اندازه‌گیری برابر است با انحراف معیار نمرات خطا و انحراف معیار نمرات مشاهده شده همان انحراف معیار نمرات آزمون است، لذا با جذر گرفتن از طرفین تساوی بالا به فرمول خطای معیار اندازه‌گیری می‌رسیم:

$$SEM = SDt\sqrt{1 - r_{tt}}$$

در این فرمول:

SEM = خطای معیار اندازه‌گیری برحسب نمرات خام

SDt = انحراف معیار آزمون که همان انحراف معیار نمرات مشاهده شده است

r_{tt} = ضریب پایایی آزمون

فرض کنید یک آزمون دارای ضریب پایایی ۰/۹۱ و انحراف معیار ۱۰ است. طبق فرمول بالا، مقدار خطای معیار اندازه‌گیری این آزمون برابر است با:

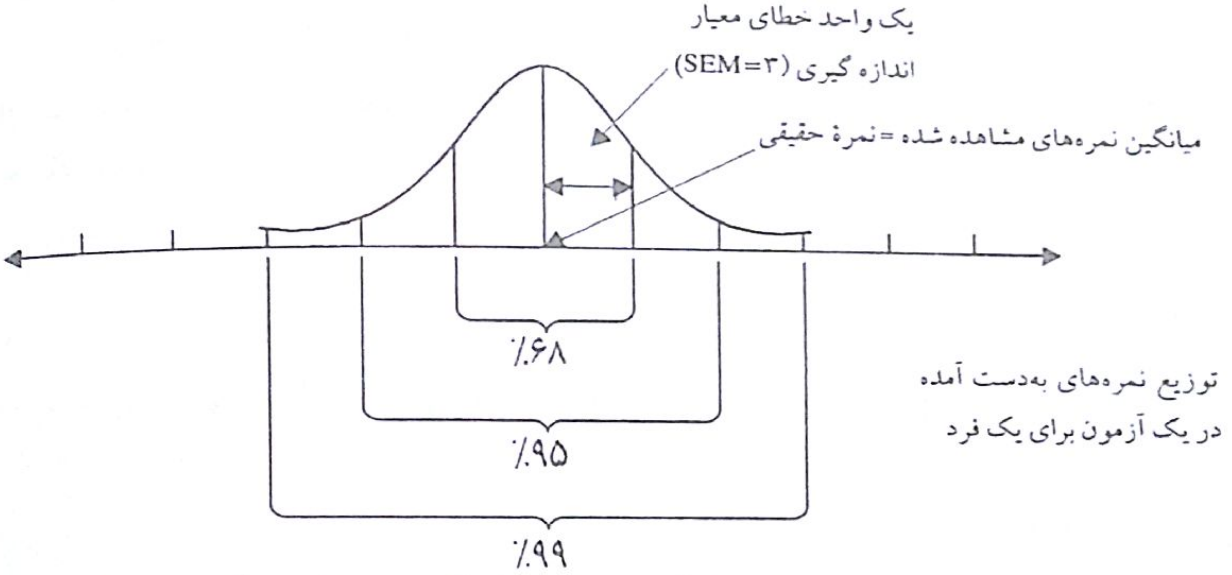
$$SEM = ۱۰\sqrt{۱ - ۰/۹۱}$$

$$= ۱۰\sqrt{۰/۰۹} = ۱۰ \times ۰/۳ = ۳$$

خطای معیار اندازه‌گیری دارای همان مقیاس آزمون است و از همان واحدهای اندازه‌گیری انحراف معیار برخوردار است. بنابراین، می‌توان آن را مستقیماً به نمره‌ها ربط داد و به ما کمک می‌کند تا بتوانیم خطاهای آنها را محاسبه کنیم.

استفاده از خطای معیار اندازه‌گیری برای برآورد نمرهٔ حقیقی

می‌توانیم از خطای معیار اندازه‌گیری برای برآورد نمرهٔ حقیقی استفاده کنیم. برای این منظور، می‌توانیم فرض کنیم که اگر آزمون‌شونده بارها به یک آزمون (یا فرمهای موازی آن) پاسخ دهد، نمره‌های حاصل توزیع طبیعی خواهند داشت که میانگین آن توزیع مُعرف نمرهٔ حقیقی آزمون‌شونده است (نگاه کنید به شکل ۲-۲۱). از آنجاکه نقش خطای معیار اندازه‌گیری شبیه به



شکل ۲-۲۱ استفاده از منحنی بهنجار و خطای معیار اندازه‌گیری برای برآورد نمرهٔ حقیقی از روی نمرهٔ مشاهده شده

نقش انحراف معیار است، می‌توانیم از آن استفاده کنیم تا نشان دهیم که اگر آزمون‌شونده به آزمونهای دیگری جواب دهد چه نتایجی به دست خواهد آمد. با توجه به ویژگیهای منحنی بهنجار که در فصل ۱۹ توضیح دادیم، نتایج حاصل به شرح زیر خواهند بود:

- تقریباً ۶۸٪ (دقیقاً ۶۸/۲۸٪) نمره‌های او در فاصلهٔ ۱SEM± از نمرهٔ حقیقی او قرار خواهند داشت.
- تقریباً ۹۵٪ (دقیقاً ۹۵/۷۴٪) نمره‌های او در فاصله ۲SEM± از نمرهٔ حقیقی او قرار خواهند داشت.
- تقریباً ۹۹٪ (دقیقاً ۹۹/۷۴٪) نمره‌های او در فاصلهٔ ۳SEM± از نمرهٔ حقیقی او قرار خواهند داشت.

البته ما از نمرهٔ حقیقی هیچ آزمون‌شونده‌ای آگاهی نداریم؛ بنابراین، باید آن را برآورد کنیم. برای این منظور، از نمرهٔ مشاهده شده استفاده می‌کنیم. به عنوان مثال، در آزمونی با خطای معیار اندازه‌گیری ۳، احتمال اینکه نمرهٔ حقیقی فردی که نمرهٔ مشاهده شده‌اش ۷۰ است در فاصلهٔ ۳± ۷۰ (یا ۷۳-۶۷) باشد برابر ۶۸٪، برای اینکه در فاصلهٔ ۶± ۷۰ (یا ۷۶-۶۴) باشد برابر ۹۵٪، و برای اینکه در فاصلهٔ ۹± ۷۰ (یا ۷۹-۶۱) باشد برابر ۹۹٪ است (شکل ۲-۲۱). به این فاصله‌ها **فاصلهٔ اطمینان**[1] می‌گویند.

هرچند که اکثر صاحب‌نظران استفاده از نمرهٔ خام (نمرهٔ مشاهده شده) را برای تعیین فاصلهٔ اطمینان پیشنهاد می‌کنند (مانند آنچه در محاسبات بالا انجام شد)، بعضی از متخصصان روانسنجی از جمله کلاین (۲۰۰۵)، نانالی و برنشتاین[2] (۱۹۹۴) معتقدند که باید فاصلهٔ اطمینان را حول **برآوردِ نمرهٔ حقیقی**[3] (T') رسم کرد. برآورد نمرهٔ حقیقی طبق فرمول زیر به دست می‌آید:

$$T' = (R)(X) \qquad R = \text{پایایی} \qquad X = \text{نمرهٔ مشاهده شده}$$

با توجه به این پیشنهاد، در آزمونی با ضریب پایایی ۰/۹۱، باید به جای استفاده از نمرهٔ مشاهده شدهٔ ۷۰، از برآوردِ نمرهٔ حقیقی مربوط به این نمره یعنی ۶۳/۷ استفاده کنیم.

$$T' = ۰/۹۱ \times ۷۰ = ۶۳/۷$$

در فصل نوزدهم، در بحث از گزارش عملکرد کلی دانش‌آموزان به صورت نیمرخ، گفتیم که عملکرد دانش‌آموز در نیمرخ نمره‌های او به صورت نوارهایی در حول نمرهٔ خام یا رتبهٔ درصدی او نشان داده می‌شود و هر نوار به اندازهٔ یک خطای معیار اندازه‌گیری بالا و یک خطای معیار اندازه‌گیری پایین هریک از نمره‌ها رسم می‌شود. این نوار همان فاصلهٔ اطمینان است که چگونگی محاسبهٔ آن را در بالا توضیح دادیم. منظور از فاصلهٔ اطمینان دامنه‌ای از نمرات است که با احتمال معین دربرگیرندهٔ نمرهٔ حقیقی است.

1. confidence interval
2. Bernstein
3. estimated true score

مفهوم پایایی در نظریهٔ سؤال ـ پاسخ

آنچه تاکنون دربارهٔ خطای معیار اندازه‌گیری گفتیم عمدتاً مربوط می‌شود به نظریهٔ کلاسیک اندازه‌گیری. در نظریهٔ سؤال ـ پاسخ یا آی آر تی، به جای خطای معیار اندازه‌گیری، از شاخص دقت اندازه‌گیری[1] استفاده می‌شود (هوگان، ۲۰۰۷). هرچند که این دو شاخص خطای اندازه‌گیری به‌هم شبیه‌اند، اما از جهاتی با یکدیگر تفاوت دارند. یک تفاوت مهم بین آنها این است که شاخص دقت اندازه‌گیری در طول مقیاس نمرات دارای اندازه‌های مختلف است، درحالی که خطای معیار اندازه‌گیری در طول مقیاس نمرات همواره دارای اندازه‌های یکسان است. به عنوان مثال، اگر خطای معیار اندازه‌گیری برای یک آزمون هوش ۵ است، معنی آن این است که نمره‌های هوش ۷۰، ۱۰۰، ۱۰۵، ۱۴۰، و غیره دارای خطای اندازه‌گیری مساوی‌اند، اما شاخص دقت اندازه‌گیری در نظریهٔ سؤال ـ پاسخ ممکن است در یک آزمون برای نمرات پایین مقیاس کوچک و برای نمرات بالای مقیاس بزرگ‌تر باشد؛ در آزمونی دیگر این وضع می‌تواند کاملاً به عکس باشد (هوگان، ۲۰۰۷).

نکته‌ای را که باید مورد توجه قرار دهید این است که شاخص دقت اندازه‌گیری تنها به روشهای پایایی مربوط به همسانی درونی آزمون ارتباط دارد. اگر هدف از تعیین پایایی ثبات نمرات در طول زمان باشد، در آن صورت روش بازآزمایی یا روش فرمهای موازی هم برای نظریهٔ کلاسیک و هم برای نظریهٔ سؤال ـ پاسخ یکسان به کار می‌رود.

نکتهٔ دیگری که لازم است به آن نیز اشاره شود این است که شاخص دقت اندازه‌گیری وابسته به نظریهٔ سؤال ـ پاسخ صرفاً با آزمونهای جامع قابل استفاده با کامپیوتر که در سطح ملی یا منطقه‌ای استاندارد می‌شوند قابل کاربردند. والا، بنا به گفتهٔ هوگان (۲۰۰۷)، "بسیار نامحتمل است که یک معلم یا یک مدرسه از روش آی آر تی برای ساختن ابزارهای سنجش خود استفاده کند، زیرا روشهای وابسته به آی آر تی (نظریهٔ سؤال ـ پاسخ) معمولاً به نمونه‌های بسیار بزرگ آزمون‌شوندگان نیازمندند" (ص ۶۱).

رابطهٔ بین روایی و پایایی (فرمول اصلاح کاهش)

ما در آغاز این فصل در بارهٔ رابطهٔ بین روایی و پایایی گفتیم که پایایی شرط روایی است؛ یعنی اینکه یک آزمون باید پایا باشد تا بتواند روا باشد. بنابراین، هر اندازه از میزان پایایی یک آزمون کاسته شود به همان نسبت از روایی آن کم می‌شود. از سوی دیگر، هر اندازه بر میزان پایایی یک آزمون افزوده شود، به همان نسبت بر روایی آن اضافه می‌شود. سؤال این است که اگر خطاهای اندازه‌گیری کم شوند یا به کلی از بین بروند چه مقدار به روایی آن افزوده می‌شود؟

---

1. precision of measurement index

اسپیرمن (۱۹۰۴) در تدوین نظریهٔ نمرهٔ واقعی مربوط به پایایی آزمون به دنبال جواب دادن به همین سؤال بود. او اظهار داشت که خطاهای اندازه‌گیریْ همبستگی بین دو آزمون X و Y (یا روایی پیش‌بینی) را کاهش می‌دهند. بنابراین، همبستگی بین اندازه‌های ناپایای X و Y از همبستگی واقعی بین آنها (یعنی همبستگی بین اندازه‌های کاملاً پایای X و Y) کمتر خواهد بود. اسپیرمن برای رفع این مشکل فرمول معروف خود را به نام **فرمول اصلاح کاهش**[1] برای برآورد رابطهٔ واقعی بین X و Y بر اساس همبستگی موجود بین دو اندازهٔ ناپایای X و Y به دست داد. این فرمول به قرار زیر است:

$$r'_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} \times r_{yy}}}$$

در این فرمول:

r'_{xy} = همبستگی بین X ، Y (r_{xy}) بعد از اصلاح

r_{xx} = پایایی X

r_{yy} = پایایی Y

فرمول اصلاح کاهش برآوردی از همبستگی بین اندازه‌های کاملاً پایای X و Y به دست می‌دهد. چنان‌که در جدول ۲-۲۱ می‌بینید، این اصلاح می‌تواند بر r_{xy} تأثیر چشم‌گیری داشته باشد، بخصوص وقتی که پایایی X و Y کوچک است.

به عنوان مثال، فرض کنید که ضریب همبستگی بین دو آزمون X و Y رقم ۰/۳۰ محاسبه شده است. باز هم فرض کنید که ضریب پایایی آزمون X برابر است با r_{xx} = ۰/۹۰ و ضریب پایایی آزمون Y مساوی است با r_{yy} = ۰/۸۰، یعنی این دو آزمون کاملاً پایا نیستند. حال، با استفاده از فرمول اصلاح کاهش ضریب همبستگی اصلاح شدهٔ بین X و Y را به نحو زیر به دست می‌آوریم:

$$r'_{xy} = \frac{۰/۳۰}{\sqrt{۰/۹۰ \times ۰/۸۰}} = ۰/۳۵$$

در استفادهٔ عملی از فرمول اصلاح کاهش می‌توان تعیین کرد که اگر پایایی یکی از دو آزمون یا پایایی هر دوی آنها را به مقدار معینی افزایش دهیم در همبستگی بین X و Y یعنی در r_{xy} چه مقدار افزایش خواهیم داشت. برای مثال، فرض‌کنید بین دو آزمون دارای ضرایب پایایی ۰/۷۰ و ۰/۷۵ همبستگی r_{xy} = ۰/۴۰ را به دست آورده‌ایم. اگر بتوانیم پایایی این دو آزمون را به ۰/۹۰ افزایش دهیم چه مقدار بر r_{xy} افزوده خواهد شد؟ فرمول زیر برای جواب دادن به این سؤال به کار می‌رود:

1. correction for attenuation formula

جدول ۲ - ۲۱ تغییرات ایجاد شده در r_{xy} در نتیجهٔ افزایش و کاهش پایایی x و y ارقام اولیه: $r_{xy} = ۰/۴۰$، $r_{xx} = ۰/۷۰$، $r_{yy} = ۰/۷۵$ (مرفی و دیوید شوفر، ۱۹۹۴، ص ۹۵)

اصلاح شده	کاهش پایایی		اصلاح شده	افزایش پایایی	
r_{xy}	r_{yy}	r_{xx}	r_{xy}	r_{yy}	r_{xx}
۰/۳۷	۰/۷۰	۰/۶۵	۰/۴۳	۰/۸۰	۰/۷۵
۰/۳۱	۰/۶۰	۰/۵۵	۰/۴۵	۰/۸۵	۰/۸۰
۰/۲۶	۰/۵۰	۰/۴۵	۰/۴۸	۰/۹۰	۰/۸۵
۰/۲۵	۰/۴۵	۰/۴۵	۰/۵۰	۰/۹۰	۰/۹۰
۰/۲۲	۰/۴۰	۰/۴۰	۰/۵۲	۰/۹۵	۰/۹۵

$$r'_{xy} = \frac{r_{xy}\sqrt{(r'_{xx})(r'_{yy})}}{\sqrt{(r_{xx})(r_{yy})}}$$

در این فرمول:

r'_{xy} = ضریب همبستگی بین x و y پس از اصلاح ضریب پایایی x و (یا) y

r_{xx} = ضریب پایایی آزمون x پیش از تغییر یا ضریب پایایی قدیم x

r_{yy} = ضریب پایایی آزمون y پیش از تغییر یا ضریب پایایی قدیم y

r'_{xx} = ضریب پایایی آزمون x پس از تغییر یا ضریب پایایی جدید x

r'_{yy} = ضریب پایایی آزمون y پس از تغییر یا ضریب پایایی جدید y

در بالا گفتیم که دو آزمون X و Y به ترتیب دارای ضرایب پایایی ۰/۷۰ و ۰/۷۵ هستند و برای آنها ضریب همبستگی ۰/۴۰ را به دست آورده‌ایم. اگر ضرایب پایایی هر دو آزمون را به ۰/۹۰ افزایش دهیم، بر همبستگی بین آندو (r_{xy}) چه مقدار افزوده می‌شود؟ جواب این سؤال با استفاده از فرمول فوق به گونهٔ زیر محاسبه می‌شود.

$$r'_{xy} = \frac{۰/۴۰\sqrt{(۰/۹۰)(۰/۹۰)}}{\sqrt{(۰/۷۰)(۰/۷۵)}}$$

$$= \frac{۰/۴۰ \times ۰/۹}{۰/۷۲}$$

$$= \frac{۰/۳۶}{۰/۷۲}$$

$$= ۰/۵$$

مهم‌ترین فایدهٔ فرمول بالا این است که به ما امکان می‌دهد تا هم تأثیر افزایش و هم تأثیر کاهش پایایی آزمونهای X و Y را بر همبستگی بین آنها تعیین کنیم. در جدول ۲ - ۲۱ نشان داده شده است که با افزایش و کاهش پایایی X و Y همبستگی بین آنها (r_{xy}) به چه صورتی تغییر می‌کند. چنان‌که در این جدول دیده می‌شود، اگر اندازه‌های بزرگ پایایی (یعنی ۰/۹۵) به‌کار بروند، ضریب همبستگی بین دو آزمون (یعنی r_{xy}) تا رقم ۰/۵۲ افزایش می‌یابد. اما چنانچه اندازه‌های کوچک پایایی (یعنی ۰/۴۰) به‌کار روند، ضریب همبستگی بین دو آزمون تا رقم ۰/۲۲ کاهش می‌یابد.

پایایی روشهای درجه‌بندی یا رتبه‌بندی

معروف‌ترین روش تعیین پایاییِ روشهای **درجه‌بندی یا رتبه‌بندی**[1] روش گیلفورد[2] (۱۹۵۴) است. این روش اساساً با داده‌های مربوط به **آرایش رتبه‌های**[2] به دست آمده از داوران مختلف به کار می‌رود. در این روش، تعدادی داور گروهی از افراد را از بالا به پایین یا از زیاد به کم رتبه‌بندی یا درجه‌بندی می‌کنند. بعد نتایج حاصل از این درجه‌بندی در فرمول زیر قرار داده می‌شود و از این طریق پایایی رتبه‌بندی یا درجه‌بندی به دست می‌آید.

فرمول گیلفورد:

$$\bar{r} = ۱ - \frac{M(۴N + ۲)}{(M - ۱)\ (N - ۱)} + \frac{۱۲\Sigma S^{۲}}{M(M - ۱)N(N^{۲} - ۱)}$$

در این فرمول:

$\bar{r}$ = میانگین همبستگیهای بین داوران = پایایی روش درجه‌بندی یک داور

M = تعداد داوران

N = تعداد افراد مورد داوری

S = مجموع رتبه‌ها برای هر نفر

جدول ۳ـ ۲۱ آرایش درجه‌بندی چهار داور را برای ۹ نفر نشان می‌دهد. با استفاده از اطلاعات این جدول ضریب پایایی این روش درجه‌بندی به نحو زیر محاسبه می‌شود.

$$\bar{r} = ۱ - \frac{(۴)\ (۴ \times ۹ + ۲)}{(۴ - ۱)\ (۹ - ۱)} + \frac{۱۲\ (۴۲۳۴)}{۴\ (۴ - ۱)۹(۹^{۲} - ۱)}$$

$$\bar{r} = ۱ - \frac{(۴)\ (۳۸)}{(۳)\ (۸)} + \frac{(۱۲)\ (۴۲۳۴)}{(۴)\ (۳)۹(۸۰)}$$

$$= ۱ - \frac{۱۵۲}{۲۴} + \frac{۵۰۸۰۸}{۸۶۴۰}$$

$$= ۱ - ۶/۳۳۳ + ۵/۸۸۱$$

$$= ۰/۵۴۸$$

---

1. rating 2. Guilford 2. rank ordering

جدول ۳ ـ ۲۱ داده‌های حاصل از درجه‌بندی برای ۹ نفر به‌وسیلهٔ ۴ داور

افراد	داوران ۱	۲	۳	۴	S درجه‌بندی کل	S^2 ^{2}(درجه‌بندی کل)
۱	۸	۶	۹	۷	۳۰	۹۰۰
۲	۲	۱	۱	۳	۷	۴۹
۳	۱	۵	۲	۲	۱۰	۱۰۰
۴	۷	۹	۸	۸	۳۲	۱۰۲۴
۵	۴	۳	۴	۵	۱۶	۲۵۶
۶	۶	۸	۷	۹	۳۰	۹۰۰
۷	۳	۲	۵	۱	۱۱	۱۲۱
۸	۷	۴	۳	۶	۲۰	۴۰۰
۹	۵	۷	۶	۴	۲۲	۴۸۴

$\Sigma S^2 = ۴۲۳۴$

چنان‌که در سطرهای بالا گفتیم، رقم حاصل (۰/۵۴۸) پایایی یک داور است. با استفاده از فرمول اسپیرمن ـ براون می‌توانیم پایایی چهار داور را حساب کنیم. در قسمتهای قبلی این فصل فرمول اسپیرمن ـ براون را به نحو زیر معرفی کردیم:

$$r'_{tt} = \frac{kr_{tt}}{1 + (k - 1)\, r_{tt}}$$

در این فرمول:

r'_{tt} = ضریب پایایی چهار داور

r_{tt} = ضریب پایایی یک داور

$k = (\frac{۴}{۱}$، نسبت تعداد کل داوران به یک داور (در مثال بالا

بنابراین، خواهیم داشت:

$$r'_{tt} = \frac{(۴)\ (۰/۵۴۸)}{۱ + (۴-۱)\ (۰/۵۴۸)}$$

$$= \frac{۲/۱۹۲}{۱ + ۱/۶۴۴}$$

$$= \frac{۲/۱۹۲}{۲/۶۴۴}$$

$$= ۰/۸۸۰$$

پایایی آزمونهای وابسته به ملاک

در رابطه با آزمونهای وابسته به ملاک، ما بیشتر به *ثبات* نتایج آزمون در ارتباط با تصمیمهای مربوط به حد تسلط علاقه‌مندیم؛ به سخن دیگر، این‌گونه پایایی تعیین می‌کند که وضعیت تعدادی

دانش‌آموز از لحاظ دستیابی یا عدم دستیابی به حد تسلط در دو موقعیت مختلف تا چه اندازه از ثبات برخوردار بوده است. اگر وضعیت اکثریت آزمون‌شوندگان از لحاظ دستیابی یا عدم دستیابی به حد تسلط در دو موقعیت آزمون (دوبار اجرای یک آزمون یا اجرای دو فرم موازی از یک آزمون) یکسان باشد در آن صورت گفته می‌شود که آزمون مورد نظر از میزان بالایی از پایایی برخوردار است. بنابراین، در رابطه با آزمونهای وابسته به ملاک توجه ما عمدتاً معطوف است به پایایی (دقّت) تصمیمهایی که می‌گیریم تا پایایی نمره‌های آزمون.

البته باید تأکید کرد که زمان سپری شده بین دو موقعیت آزمون یا دو بار اجرای آزمون باید بسیار کوتاه باشد، به گونه‌ای که در آن فاصله هیچ عامل مشخصی بر بهبود عملکرد آزمون‌شوندگان در بار دوم اجرای آزمون تأثیر نگذارد. همچنین در فاصلهٔ اجرای دو آزمون هیچ‌گونه آموزش یا بازخورد اصلاحی نباید به آزمون‌شوندگان داده شود. کوتاه سخن اینکه، شرایط هر دو بار اجرای آزمون باید کاملاً مشابه باشند.

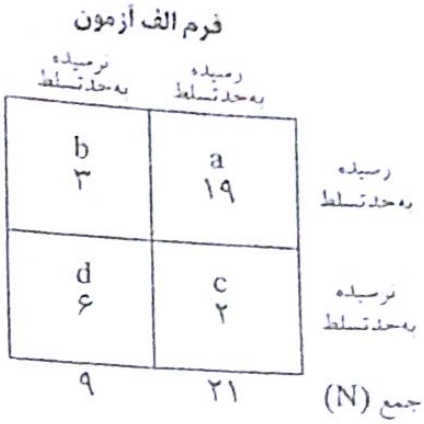
در تعیین پایایی آزمونهای وابسته به ملاک، مانند تعیین آزمونهای وابسته به هنجار، هم روش *بازآزمایی* یک آزمونِ واحد میسر است و هم روش اجرای *فرمهای معادل یا هم‌ارز*. در اینجا نیز استفاده از روش فرمهای هم‌ارز بر روش بازآزمایی ترجیح داده می‌شود.

ضریب پایایی آزمونهای وابسته به ملاک (روش درصد توافق)

چنان‌که گفتیم، مهم‌ترین نوع پایایی آزمونهای وابسته به ملاک به ثبات نتایج آزمون در ارتباط با دسته‌بندی آزمون‌شوندگان به *کسانی که به حد تسلط رسیده* (مسلط‌ها[1]) و *کسانی که به این حد نرسیده‌اند* (غیرمسلط‌ها[2]) مربوط است. بنابراین، برای تعیین این نوع ضریب پایایی لازم است درصد آزمون‌شوندگانی را که در هر دو موقعیت آزمون یک نوع طبقه‌بندی شده‌اند (رسیده به حد تسلط یا نرسیده به حد تسلط) محاسبه نماییم. این روش ساده در کتاب *معیارهایی برای آزمودن در روانشناسی و آموزش و پرورش* (انجمن روانشناسی آمریکا، ۱۹۸۵) پیشنهاد شده است. به این روشِ تعیین پایایی **ثبات تصمیم‌گیری**[3] نیز می‌گویند. ثبات تصمیم‌گیری در مقابل ثبات نمره به کار رفته است. برای مثال، فرض می‌کنیم که دو فرم معادل از یک آزمون مربوط به یادگیری در حد تسلط را اجرا کرده‌ایم. سؤالی که در اینجا مهم است پرسیده شود این است که "آیا هر دو آزمون یک دانش‌آموز را رسیده به حد تسلط یا نرسیده به حد تسلط معرفی خواهند کرد؟" ضریب پایایی که به این منظور به دست می‌آید ضریب *ثبات تصمیم‌گیری* نام دارد (نیتکو، ۲۰۰۱، ص ۷۴).

برای آشنایی بیشتر با چگونگی استفاده از روش پیشنهادی بالا، فرض کنید دو فرم موازی از یک آزمون ۲۰ سؤالی را با یک گروه ۳۰ نفری از دانش‌آموزان اجرا کرده و تصمیم گرفته‌ایم که نمرهٔ رسیدن به

1. masters
2. nonmasters
3. decision consistency



شکل ۴ - ۲۱ طبقه‌بندی دانش‌آموزان از لحاظ رسیدن یا نرسیدن به حد تسلط

حد تسلط برای این افراد جواب دادن به حداقل ۱۶ سؤال از ۲۰ سؤال یعنی کسب دست کم ۸۰ درصد نمرهٔ آزمون باشد. بنابراین، همهٔ دانش‌آموزانی که در هر دو آزمون نمرهٔ ۱۶ یا بیشتر گرفته باشند در گروه *به حد تسلط رسیده‌ها* و کسانی که در هر دو آزمون نمرهٔ ۱۵ یا کمتر گرفته باشند در گروه *به حد تسلط نرسیده‌ها* قرار خواهند گرفت. باقیماندهٔ دانش‌آموزان در یکی از دو آزمون به حد تسلط نرسیده‌اند. می‌توان این اطلاعات را در یک چهارخانه، نظیر شکل ۴ - ۲۱، به طور خلاصه نشان داد. پس از آنکه اطلاعات مورد نیاز برای محاسبهٔ ضریب پایایی مانند آنچه در شکل ۴ - ۲۱ آمده است تکمیل شد، می‌توان با استفاده از فرمول زیر، ضریب پایایی آزمون یا به عبارت دقیق‌تر درصد کسانی که در هر دو آزمون به حد تسلط رسیده یا نرسیده‌اند را مشخص کرد.

$$\text{ضریب پایایی} = \frac{\text{تعداد افرادی که در هر دو آزمون به حد تسلط نرسیده‌اند} + \text{تعداد افرادی که در هر دو آزمون به حد تسلط رسیده‌اند}}{\text{تعداد کل گروه (N)}} \times ۱۰۰$$

با گذاشتن اطلاعات شکل ۴ - ۲۱ در فرمول بالا، ضریب پایایی آزمون مورد نظر به نحو زیر محاسبه می‌شود:

$$\text{ضریب پایایی} = \frac{۱۹ + ۶}{۳۰} \times ۱۰۰$$

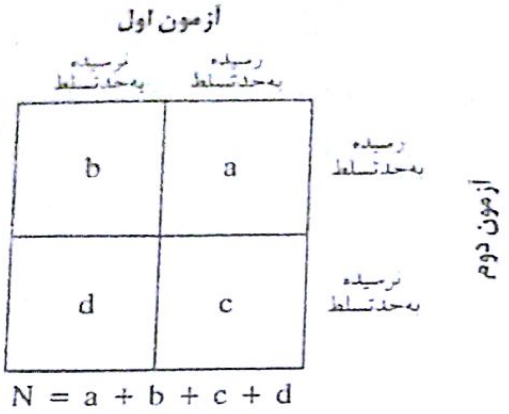
$$= ۰/۸۳ \times ۱۰۰$$

$$= ٪۸۳ \text{ یا } ۸۳$$

ضریب پایایی آزمونهای وابسته به ملاک (روش کاپا)

یکی دیگر از روشهای محاسبهٔ ضریب پایایی آزمونهای وابسته به ملاک براساس توافق بین تصمیمها به روش کاپا[1] معروف است. این روش را کوهن (۲۰۰۰) پیشنهاد داده است. در این روش نیز طرح تعیین پایایی مطابق با شکل ۵-۲۱ است.

1. Kappa



شکل ۵ ـ ۲۱ طرح مورد نیاز برای تعیین پایایی براساس توافق بین تصمیمها در آزمونهای وابسته به ملاک (مرفی و دیوید شوفر، ۲۰۰۱، ص ۱۴۰)

$$\text{توافق تصادفی} = [(\frac{a+b}{N})\ (\frac{a+c}{N})] + [(\frac{c+d}{N})\ (\frac{b+d}{N})]$$

$$\text{Kappa} = \frac{\text{توافق تصادفی} - \text{توافق کلی}}{\text{توافق تصادفی} - ۱}$$

در شکل ۵-۲۱، a نشان‌دهندهٔ تعداد آزمون‌شوندگانی است که در هر دو آزمون به حد تسلط رسیده‌اند و d معرف تعداد آزمون‌شوندگانی است که در هر دو آزمون به حد تسلط نرسیده‌اند. ضریب پایایی براساس روش کاپا به صورت زیر محاسبه می‌شود:

$$\text{توافق کلی} = \frac{a+d}{N}$$

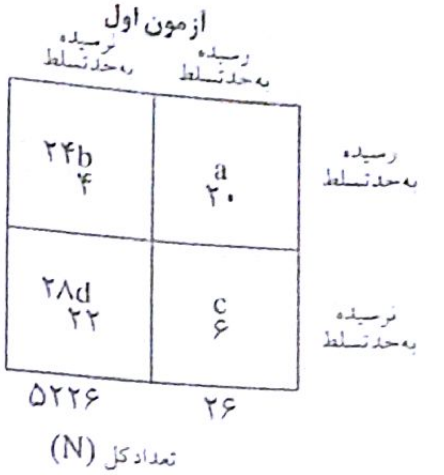
دقت کنید که بخش اول فرمول بالا همان فرمول مورد استفاده در روش قبلی است، اما بخش دوم فرمول توافق تصادفی را حساب می‌کند. علت کم کردن این توافق تصادفی از توافق کلی این است که چنانچه نتایج حاصل از دوبار اجرای یک آزمون یا اجرای دو آزمون موازی به تصادف نتایج مشابهی به دست دهد، ضریب پایایی غیرواقعی برای آن گزارش نشود.

برای نشان دادن چگونگی استفاده از فرمول کاپا، از اطلاعات موجود در شکل ۶-۲۱ استفاده می‌کنیم. ضریب توافق کلی برابر است با:

$$\text{توافق کلی} = \frac{۲۰+۲۲}{۵۲}$$
$$= ۰/۸۰$$

همچنین نسبت توافق تصادفی برابر است با:

$$\text{توافق تصادفی} = [(\frac{۲۰+۴}{۵۲})\ (\frac{۲۰+۶}{۵۲})] + [(\frac{۶+۲۲}{۵۲})\ (\frac{۴+۲۲}{۵۲})]$$
$$= [(\frac{۲۴}{۵۲})\ (\frac{۲۶}{۵۲})] + [(\frac{۲۸}{۵۲})\ (\frac{۲۶}{۵۲})]$$
$$= (۰/۴۶)\ (۰/۵۰) + (۰/۵۲)\ (۰/۵)$$
$$= ۰/۰۶$$



شکل ۶ ـ ۲۱ مثالی برای محاسبهٔ پایایی آزمونهای وابسته به ملاک با روش کاپا

حال با استفاده از ضریب توافق کلی و ضریب توافق تصادفی ضریب کاپا را به نحو زیر محاسبه می‌کنیم:

$$\text{Kappa} = \frac{۰/۸۰ - ۰/۰۶}{۱ - ۰/۰۶}$$

$$= \frac{۰/۷۴}{۰/۹۴}$$

$$= ۰/۷۸$$

ضریب پایایی آزمونهای وابسته به ملاک (روش لیوینگ استون)

روش دیگری که برای محاسبهٔ ضریب پایایی آزمونهای وابسته به ملاک یا آزمونهای درحد تسلط پیشنهاد شده به **روش لیوینگ استون**[1] معروف است. این روش را نخستین بار ساموئل لیوینگ استون (۱۹۷۲a، ۱۹۷۲b) پیشنهاد داده است. او استدلال کرده که می‌توان تفاوت بین نمرهٔ حد تسلط و میانگین آزمون حد تسلط را "واریانس حقیقی" تصور کرد، و حتی زمانی که بین نمره‌های مشاهده شده تفاوت اندکی وجود دارد یا بین آنها هیچگونه تفاوتی دیده نمی‌شود این به اصطلاح نمرهٔ حقیقی می‌تواند چشم‌گیر باشد. لیوینگ استون برای تعیین پایایی آزمونهای وابسته به ملاک فرمول زیر را به دست داده است:

$$r_{ct} = \frac{r_{tt}\, SD^2 + (\text{نمرهٔ حد تسلط} - \text{میانگین آزمون})^2}{SD^2 + (\text{نمرهٔ حد تسلط} - \text{میانگین آزمون})^2}$$

در این فرمول:

r_{ct} = پایایی آزمون وابسته به ملاک

r_{tt} = پایایی آزمون با روش وابسته به هنجار (KR۲۰، دونیمه کردن، و غیره)

SD^2 = واریانس آزمون (براساس تفاوت نمره‌های مشاهده شده و نمرهٔ حد تسلط)

1. Livingstone approach

براساس فرمول بالا، اگر میانگین نمره‌های آزمون و نمرهٔ حد تسلط برابر باشند، ضریب پایایی پیشنهادی لیوینگ استون دقیقاً برابر خواهد بود با ضریب پایایی محاسبه شده با روشهای مورد استفاده با آزمونهای وابسته به هنجار. اما زمانی که بین میانگین آزمون و نمرهٔ حد تسلط تفاوت وجود دارد، ضریب محاسبه شده با روش لیوینگ استون بزرگ‌تر از ضریبهای پایایی محاسبه شده با روشهای وابسته به هنجار خواهد بود.

ضریب پایایی آزمونهای وابسته به ملاک (روش لیبرمن و مرندا)

در روش تعیین پایایی آزمونهای ملاکی پیشنهادی لیبرمن[1] و مرندا[2] (۱۹۷۹) آزمون دوبار با گروه واحدی از آزمون‌شوندگان اجرا می‌شود و نتایج در فرمول زیر قرار داده می‌شوند.

$$C=\frac{nb-sf}{nb-sf+vN}$$

در این فرمول:

n = تعداد آزمون‌شوندگانی که در هر دوبار اجرای آزمون به حد تسلط نرسیده‌اند

b = تعداد آزمون‌شوندگانی که در هر دوبار اجرای آزمون به حد تسلط رسیده‌اند

f = تعداد آزمون‌شوندگانی که فقط در بار اول اجرای آزمون به حد تسلط رسیده‌اند

s = تعداد آزمون‌شوندگانی که فقط در بار دوم اجرای آزمون به حد تسلط رسیده‌اند

v = مقدار کوچک f یا s

N = b + f + n + s = تعداد کل آزمون‌شوندگان

برای نشان دادن چگونگی استفاده از این روش و محاسبهٔ ضریب پایایی C با استفاده از فرمول بالا، فرض کنید به یک آزمون ملاکی ۱۰۰ آزمون‌شونده (N = ۱۰۰) دوبار جواب داده‌اند. از این تعداد ۹۰ نفر در هر دوبار اجرای آزمون به حد تسلط رسیده‌اند (b = ۹۰)، ۵ نفر در هر دوبار اجرای آزمون به حد تسلط نرسیده‌اند (n = ۲)، ۲ نفر فقط در بار اول به حد تسلط رسیده‌اند (f = ۲)، و ۳ نفر فقط در بار دوم به حد تسلط رسیده‌اند (s = ۳). در این مثال فرضی ما، مقدار v یعنی کوچک‌ترین مقدار از f یا s برابر است با ۲، چون s (۲) از f (۳) کوچک‌تر است. با گذاشتن این ارقام در فرمول محاسبهٔ پایایی، ضریب پایایی برابر خواهد شد با:

$$C=\frac{۵\times۹۰-۳\times۲}{۵\times۹۰-۳\times۲+۲\times۱۰۰}$$

$$=\frac{۴۵۰-۶}{۴۵۰-۶+۲۰۰}$$

$$=\frac{۴۴۴}{۶۴۴}$$

$$=۰/۶۹$$

1. Liberman 2. Merenda

عوامل مؤثر در افزایش پایایی آزمون (وابسته به ملاک)

ساکس (۱۹۹۷) برای بالا بردن پایایی آزمونهای وابسته به ملاک یا آزمونهای حدتسلط نکات زیر را پیشنهاد داده است:

۱. در صورت امکان، آزمونی بسازید که بین میانگین نمرات آن و نمرهٔ حد تسلط تفاوت محسوسی وجود داشته باشد.
۲. برای هر هدف آموزشی یا هدف یادگیری تا حد امکان سؤالهای زیادتری بنویسید.
۳. هدفهای خود را تا آنجاکه ممکن است دقیق و مشخص بنویسید.
۴. از روش عینی نمره‌گذاری استفاده کنید.
۵. سؤالهای نمونهٔ تمرینی شبیه به سؤالهای آزمون اصلی در اختیار آزمون‌شوندگان قرار دهید.
۶. سؤالهای آزمون را سازمان یافته و روشن در اختیار آزمون‌شوندگان قرار دهید.
۷. به فرض این پیش‌بینی که بین نمرات اختلاف دیده خواهد شد، از روشهایی که پایایی آزمونهای وابسته به هنجار را بالا می‌برد استفاده کنید (ساکس، ۱۹۹۷، ص ۳۰۰).

پایایی روشهای سنجش مستقیم

در آغاز این فصل، در بحث مربوط به پایایی بین مصححان، گفتیم که منظور ما از مصححان هم کسانی هستند که برگه‌های امتحانات غیر عینی را تصحیح می‌کنند و هم آنانی که به مشاهده و درجه‌بندی رفتار یا محصول رفتار می‌پردازند. در آنجا روش تعیین پایایی بین مصححان را توضیح دادیم. دراینجا روشهای تعیین پایایی بین مشاهده‌گران و پایایی بین درجه‌بندی‌کنندگان را توضیح می‌دهیم.

از آنجاکه روشهای سنجش مستقیم رفتارهای آشکار یا محصولات رفتاری به طور بی‌واسطه موضوعهای موردنظر را می‌سنجند مشکل روایی ندارند، زیرا روایی در پاسخ به این پرسش مطرح می‌شود که "آیا ابزار یا وسیلهٔ سنجش همان چیزی را که موردنظر است می‌سنجد یا نه؟" که البته پاسخ آن در رابطه با سنجش مستقیم مثبت است. اما مسئلهٔ پایایی در ارتباط با این روشها نیز مطرح است، زیرا پایایی به دقت نتایج سنجش اشاره می‌کند و روشهای سنجش مستقیم نیز ممکن است از دقت کافی برخوردار نباشند. لذا برای تعیین اینکه روش موردنظر از دقت یا پایایی لازم برخوردار است یا نه روشهایی پیشنهاد شده که ما در اینجا آنها را توضیح می‌دهیم.

پایایی روش سنجش محصولات ماندنی رفتار یا سنجش فرآوردهٔ رفتار

برای تعیین پایایی سنجش محصولات ماندنی رفتار از فرمول زیر استفاده می‌شود:

$$\text{ضریب پایایی} = \frac{\text{تعداد توافقها}}{\text{تعداد عدم توافقها} + \text{تعداد توافقها}} \times ۱۰۰$$

برای محاسبهٔ ضریب پایایی روش تصحیح تکلیف دانش‌آموزی که از او خواسته شده است

کلماتی را که با کلمهٔ "کتاب" هم قافیه هستند بنویسد، و در آن مصحح اول ۲۰ کلمه را درست ولی مصحح دوم ۱۹ کلمه را درست تشخیص داده است، به نحو زیر عمل می‌شود:

$$\text{ضریب پایایی} = \frac{19}{19+1} \times 100$$

$$= \frac{19}{20} \times 100$$

$$= 95$$

ملاحظه می‌کنید که در محاسبهٔ بالا در مخرج کسر عدد ۲۰ قرار داده شده است. عدد ۲۰ مُعرف تعداد توافقها (۱۹ مورد) به اضافهٔ تعداد عدم توافقها (۱ مورد) است.

پایایی روش ثبت فراوانی رفتار

برای تعیین پایایی روش سنجش فراوانی رفتار از فرمول زیر استفاده می‌شود:

$$\text{ضریب پایایی} = \frac{\text{کوچک‌ترین فراوانی ثبت شده}}{\text{بزرگ‌ترین فراوانی ثبت شده}} \times 100$$

به عنوان مثالی از نحوهٔ محاسبهٔ این نوع ضریب پایایی، فرض کنید یکی از مشاهده‌گرانی که رفتار ریختن غذا توسط یک کودک را مشاهده کرده ۲۷ مورد غذا ریختن را ثبت کرده است و مشاهده‌گر دوم ۲۹ مورد از همین رفتار را گزارش کرده است. بنا به فرمول فوق، ضریب پایایی سنجش مورد نظر ۹۳ خواهد بود، طبق محاسبات زیر:

$$\text{ضریب پایایی} = \frac{27}{29} \times 100$$

$$= 93$$

پایایی روشهای سنجش طول مدت و دورهٔ نهفتگی رفتار

روش محاسبهٔ ضریب پایایی روش سنجش طول مدت و نهفتگی رفتار مانند روش مورد استفاده برای تعیین ضریب پایایی روش سنجش فراوانی رفتار است، منتهیٰ در محاسبهٔ پایایی روش سنجش طول مدت رفتار در صورت کسر، به جای کوچک‌ترین فراوانی ثبت شده، کوچک‌ترین طول مدت ثبت شده برای رفتار و در مخرج کسر، به جای بزرگ‌ترین فراوانی ثبت شده، بزرگ‌ترین طول مدت ثبت شده را قرار می‌دهند.

$$\text{ضریب پایایی} = \frac{\text{کوچک‌ترین طول مدت ثبت شده}}{\text{بزرگ‌ترین طول مدت ثبت شده}} \times 100$$

فرض کنید دو مشاهده‌گر مستقل زمان صرف شده از سوی یک کودک برای حل کردن یک مسئلهٔ ریاضی را مشاهده کرده و یکی از آنها ۳۵ دقیقه و دیگری ۳۳ دقیقه را گزارش کرده‌اند. طبق فرمول بالا پایایی این روش مشاهده برابر است با ۹۴/۲۸:

$$\text{ضریب پایایی} = \frac{33}{35} \times 100$$

$$= 94/28$$

پایایی روشهای ثبت فاصله‌ای و ثبت نمونه‌گیری زمان

چنان که قبلاً گفته شد، در روش ثبت فاصله‌ای و روش نمونه‌گیری زمان، در طول هر یک از زمانی (معمولاً ۵ تا ۱۰ ثانیه)، تنها یک پاسخ ثبت می‌شود. اگر دو مشاهده‌گر مستقل به ثبت رفتار واحدی بپردازند، آنچه باید معین شود این است که فاصله‌های زمانی متوالی ثبت شده از سوی آنها، از لحاظ علامتهای مربوط به وقوع پاسخ و عدم وقوع پاسخ در هر یک از فاصله‌های زمانی، با هم مطابقت دارند یا نه. فرض کنید دو مشاهده‌گر به طور مستقل به ثبت دو نوع رفتار تعاملی یک کودک اقدام کرده‌اند. این رفتارها به صورت "لمس کردن کودک دیگر" و "صحبت کردن با کودک دیگر" تعریف شده‌اند. فرض کنید داده‌های فاصله‌ای حاصل از مشاهدات دو مشاهده‌گر همان است که در شکل ۷-۲۱ نشان داده شده‌اند.

چنان که در شکل ۷-۲۱ دیده می‌شود، هر دو مشاهده‌گر برای کودک ۱۸ مورد لمس کردن را ثبت کرده‌اند، اما تنها در ۱۶ مورد با هم توافق داشته‌اند. هر یک از آنان دو مورد را ثبت کرده که دیگری آن را نادیده گرفته است و در نتیجه در دو مورد با هم اختلاف داشته‌اند.

فرمول محاسبهٔ ضریب پایایی روشهای ثبت فاصله‌ای و ثبت نمونه‌گیری زمان به قرار زیر است:

$$\text{ضریب پایایی} = \frac{\text{تعداد توافقها}}{\text{تعداد عدم توافقها} + \text{تعداد توافقها}} \times ۱۰۰$$

برای محاسبهٔ ضریب پایایی داده‌های شکل ۷-۲۱ به گونهٔ زیر عمل می‌کنیم:

$$\text{ضریب پایایی} = \frac{۱۶}{۱۶+۲} \times ۱۰۰$$

$$= ۸۸$$



شکل ۷-۲۱ برگه‌های ثبت داده پر شده توسط دو مشاهده‌گر مستقل (مارتین و پیبر، ۱۹۹۲، ص ۲۷۰)

خلاصهٔ فصل

در این فصل یکی دیگر از ویژگیهای مهم آزمونهای روانی و پرورشی یعنی پایایی را معرفی کردیم. پایایی به میزان دقت یک آزمون اشاره می‌کند. آزمون پایا آزمونی است که اگر چندین بار به یک گروه از دانش‌آموزان داده شود نتایج یکسانی به دست می‌دهد. برای تعیین ضریب پایایی روشهای مختلفی به نام پایایی مصححان، بازآزمایی، دو نیمه کردن آزمون، روش کودر - ریچاردسون، و روش ضریب آلفا وجود دارند. در روش بازآزمایی، یک آزمون واحد را در فاصلهٔ زمانی کوتاهی دوبار به گروه واحدی از آزمون‌شوندگان می‌دهیم و ضریب همبستگی حاصل را به عنوان ضریب پایایی آزمون حساب می‌کنیم. در روش دو نیمه‌کردن آزمون، آزمونی را به یک گروه یک بار می‌دهیم و بعد آن را به دو نیمهٔ مساوی تقسیم می‌کنیم و ضریب همبستگی بین دو نیمه‌های آزمون را در فرمول اسپیرمن - براون قرار می‌دهیم و نتیجهٔ حاصل ضریب پایایی آزمون است. در روش کودر-ریچاردسون دو فرمول معروف KR۲۰ و KR۲۱ برای محاسبهٔ ضریب پایایی به کار می‌رود. روش ضریب آلفا به همبستگی درونی میان سؤالها یا ماده‌های آزمون مبتنی است.

علاوه بر روشهای بالا، ما در این فصل روشهای تعیین پایایی آزمونهای وابسته به ملاک را نیز توضیح دادیم. در این قسمت فصل روشهای زیر معرفی شدند: (۱) روش درصد توافق، (۲) روش کاپا، (۳) روش لیوینگ استون، و (۴) روش لیندرمن و مرندا.

پس از تعیین ضریب پایایی آزمون، باید آن را در فرمول ویژه‌ای که به فرمول خطای معیار اندازه‌گیری معروف است قرار داد و خطای معیار اندازه‌گیری آزمون را تعیین کرد. خطای معیار اندازه‌گیری را می‌توان برای تعیین میزان دقت نمرات حاصل از اجرای آزمون مورد استفاده قرار داد.

در بارهٔ رابطهٔ پایایی و روایی گفتیم که پایایی شرط روایی است، یعنی هرچه بر میزان پایایی یک آزمون افزوده شود، به همان نسبت بر روایی آن افزوده خواهد شد. در ضمن، فرمول اصلاح کاهش را نیز معرفی کردیم که از طریق آن می‌توان نشان داد که اگر مقدار معینی بر پایایی دو آزمون افزوده شود، ضریب همبستگی بین آنها به چه میزانی افزایش خواهد یافت و اگر بخواهیم ضریب همبستگی بین دو آزمون را به مقدار معینی بالا ببریم چه مقدار به پایایی آن آزمونها باید اضافه کنیم. همچنین در این فصل فرمول محاسبهٔ ضریب پایایی روشهای درجه‌بندی و چگونگی استفاده از آن را توضیح دادیم. در آخرین قسمت فصل نیز روشهای تعیین پایایی سنجش مستقیم رفتار را توضیح دادیم.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. یک آزمون ۴۰ سؤالی دارای ضریب پایایی ۰/۷۰ است. اگر به این آزمون ۸۰ سؤال دیگر اضافه کنیم، ضریب پایایی آن چند خواهد شد؟

الف - ۰/۹۹

ب - ۰/۹۰

ب - ۰/۸۷

ت - ۰/۵۴

۲. اگر انحراف معیار آزمونی ۲۰ و پایایی آن ۰/۸۴ باشد، خطای معیار اندازه‌گیری آن چند است؟

الف - ۳/۲

ب - ۸

پ - ۱۶/۸۰

ت - ۱۸/۳۳

۳. اگر پایایی یک آزمون ۱ باشد خطای معیار اندازه‌گیری آن چند است؟

الف - صفر

ب - ۱

پ - بزرگ‌تر از ۱

ت - غیرقابل محاسبه

۴. یک آزمون حساب را به دو گروه دانش‌آموز دورهٔ راهنمایی داده‌ایم. یک گروه شامل دانش‌آموزان قوی (هوشبهر بالای ۱۲۰) و گروه ب شامل ترکیبی از دانش‌آموزان دارای تواناییهای مختلف (از هوشبهر ۹۰ تا ۱۵۰) هستند. کدام یک از بیانات زیر در بارهٔ پایایی نتایج حاصل از اجرای این آزمون در دو موقعیت مختلف درست است؟

الف - پایایی گروه اول از دوم بیشتر است.

ب - پایایی گروه دوم از اول بیشتر است.

پ - پایایی هر دو گروه یکسان است.

ت - پایایی هر دو گروه کوچک است.

۵. یک آزمون ۱۰۰ سؤالی پیشرفت تحصیلی از ۵ خرده آزمون ۲۰ سؤالی مربوط به موضوعهای درسی مختلف تشکیل شده است. اگر بخواهیم ضریب پایایی این آزمون را حساب کنیم بهترین روش کدام است؟

الف ـ اسپیرمن ـ براون

ب - آلفای کرانباخ

پ - KR۲۰

ت - KR۲۱

۶. کدام یک از عوامل زیر پایایی آزمون را افزایش می‌دهد؟

الف - افزودن تعدادی سؤال با ضریب تمیز متوسط و ضریب دشواری زیاد به آزمون

ب - اضافه کردن تعدادی سؤال با ضریب دشواری متوسط و ضریب تمیز بالا به آزمون

پ - اجرای آزمون در مورد گروهی از افراد که از لحاظ توانایی مورد سنجش تجانس بیشتری دارند

ت - اجرای آزمون در مورد گروهی از افراد که از لحاظ توانایی مورد سنجش تجانس کمتری دارند

ث - ب و ت

۷. در چه صورتی پایایی نمرهٔ اختلاف بین یک پیش آزمون و یک پس آزمون زیاد است؟

الف - همبستگی بین پیش‌آزمون و پس‌آزمون کم باشد.

ب - پایایی پیش آزمون و پس‌آزمون زیاد باشد.

پ - شرایط الف و ب هر دو برقرار باشد.

ت - شرایط الف و ب هیچ‌کدام برقرار نباشد.

۸ ضریب پایایی شامل چه مقادیری است؟

الف - از صفر تا ۱+

ب - از ۱- تا ۱+

پ - از صفر تا ۱-

ت - محدودیتی وجود ندارد

۹. در رابطه با روایی و پایایی می‌توان گفت که،

الف - آزمون روا پایا نیز هست.

ب - پایایی شرط لازم برای روایی است.

پ - آزمون پایا حتماً رواست.

ت - این دو از هم مستقل‌اند.

۱۰. برای تعیین پایایی روشهای سنجش مستقیم رفتار از کدام روش استفاده می‌شود؟

الف - همسانی درونی

ب - فرمهای موازی

پ - میان مشاهده‌گران

ت - تکرار آزمون

۱۱. فرض کنید یک آزمون هوشی دارای میانگین ۱۰۰ و انحراف معیار ۱۴ و پایایی ۰/۸۹ است. نمرهٔ حقیقی هوش‌فردی که در این آزمون ۱۰۶ گرفته با احتمال ۹۵ درصد در چه فاصله‌ای قرار می‌گیرد؟

الف - ۱۰۸/۳۲ - ۱۰۳/۶۸

ب ـ ۱۱۰/۶۴ ـ ۱۰۱/۳۶

پ ـ ۱۱۵/۲۸ ـ ۹۶/۷۲

ت ـ ۱۰۴/۶۴ ـ ۹۵/۳۶

ث ـ ۱۱۹/۹۲ ـ ۹۲/۵۸

۱۳. پایایی را تعریف کنید و رابطهٔ آن را با روایی توضیح دهید.

۱۳. روشهای مختلف تعیین پایایی را توضیح دهید.

۱۴. روش بازآزمایی را شرح دهید و اشکالهای آن را نام ببرید.

۱۵. ضریب همبستگی بین دو نیمهٔ یک آزمون را با چه محاسباتی به ضریب پایایی کل آن آزمون تبدیل می‌کنیم؟

۱۶. پایایی یک آزمون ۰/۷۵ و انحراف معیار آن ۷ است، خطای معیار اندازه‌گیری این آزمون را حساب کنید.

۱۷. اگر آزمونی دارای ضریب پایایی ۰/۷۵ و انحراف معیار ۱۴ باشد، خطای معیار اندازه‌گیری آن چقدر خواهد بود؟ نتیجهٔ حاصل را با جواب سؤال ۱۳ مقایسه کنید و دلیل اختلاف حاصل را توضیح دهید.

۱۸. فرمول محاسبهٔ پایایی روشهای درجه‌بندی را توضیح دهید.

۱۹. ارقام زیر نمرات ۱۰ دانش‌آموز در یک آزمون ۶ سؤالی است. با استفاده از فرمول KR۲۰، ضریب پایایی این آزمون را حساب کنید.

سؤالها

دانش آموزان	۱	۲	۳	۴	۵	۶	نمرهٔ کل
۱	۱	۱	۱	۰	۱	۱	۵
۲	۰	۱	۱	۰	۰	۱	۳
۳	۱	۰	۱	۱	۰	۰	۳
۴	۱	۱	۰	۱	۱	۱	۵
۵	۱	۱	۱	۱	۰	۰	۴
۶	۰	۱	۰	۰	۱	۰	۲
۷	۱	۱	۰	۱	۱	۰	۴
۸	۱	۱	۰	۱	۰	۱	۴
۹	۱	۱	۱	۱	۱	۱	۶
۱۰	۱	۱	۱	۱	۱	۱	۶

۲۰. دو مشاهده‌گر مستقل رفتار تیک یک فرد را به مدت یک ساعت مشاهده کرده و ارقام ۱۱ و ۱۲ را گزارش کرده‌اند. ضریب پایایی این روش سنجش را تعیین کنید.

۲۲

ارزشیابی معلم

هدفهای یادگیری

از خواننده انتظار می‌رود پس از مطالعه و یادگیری مطالب این فصل بتواند:

۱. هدفهای ارزشیابی معلم را توضیح دهد.
۲. ارزشیابی معلم را تعریف کند.
۳. روشهای ارزشیابی معلم را تعریف کند.
۴. هر یک از روشهای ارزشیابی معرفی‌شده در این فصل را توضیح دهد و چگونگی استفاده از آن را مشخص نماید.
۵. چگونگی استفاده از روش مصاحبه را برای سنجش تواناییهای داوطلبان ورود به شغل معلمی و برای ارزشیابی از کیفیت کار و تواناییهای آموزشی معلمان شاغل توضیح دهد.
۶. محاسن و معایب روش مصاحبه را توضیح دهد.
۷. شاخصهای اثربخشی معلم را که در روش مشاهدهٔ کلاسی مورد ارزشیابی قرار می‌گیرند نام ببرد.
۸. روشهای مختلف مشاهدهٔ کلاسی را توضیح دهد.
۹. روش نمونه گیری از آموزش معلم به کمک مشاهده را توضیح دهد.
۱۰. محاسن و معایب روش مشاهدهٔ کلاسی را توضیح دهد.
۱۱. انواع فرمهای نظرسنجی دانشجویی و دانش‌آموزی را توضیح دهد و موارد استفادهٔ آنها را مشخص نماید.
۱۲. انواع سؤالهای پرسشنامه‌های نظرخواهی از دانشجویان و دانش‌آموزان را توضیح دهد.
۱۳. شیوهٔ اجرا، نمره گذاری، و گزارش نتایج پرسشنامه‌های نظرخواهی را توضیح دهد.
۱۴. محاسن و معایب روش استفاده از نظرخواهی از دانشجویان و دانش‌آموزان را توضیح دهد.
۱۵. چگونگی استفاده از روش بهره گیری یا استفاده از یادگیری دانش‌آموزان و دانشجویان را شرح دهد.
۱۶. روش استفاده از آزمونهای عملکرد آموزشی و درس‌خُرد را توضیح دهد.
۱۷. محاسن و معایب روش استفاده از یادگیری شاگردان را توضیح دهد.
۱۸. وسایل مورد استفاده در روش خودسنجی را توضیح دهد.
۱۹. محاسن و معایب روش خودسنجی را توضیح دهد.
۲۰. محاسن و معایب روش ارزشیابی به وسیلهٔ همکاران را توضیح دهد.
۲۱. محاسن و معایب روش بررسی آثار علمی اعضای هیأت علمی را به عنوان وسیله‌ای برای ارزشیابی استادان توضیح دهد.
۲۲. برای ارزشیابی از میزان اثربخشی فعالیتهای آموزشی و علمی یک مؤسسهٔ آموزشی یک طرح ارزشیابی معلم تهیه کند، آن را به اجرا درآورد، و نتایج را استخراج نماید.

ما در فصل چهارم کتاب، در رابطه با ارزشیابی آموزشی، گفتیم که یکی از ارزشیابیهای آموزشی مهم **ارزشیابی معلم**[1] است. به خاطر اهمیت ارزشیابی معلم و نقش آن در بهبود کیفیت آموزشی، این فصل را به ارزشیابی معلم اختصاص داده‌ایم.

اهمیت ارزشیابی معلم

ارزشیابی از کار معلم همواره مورد توجه متخصصان آموزشی، به‌ویژه در کشورهای مغرب زمین، بوده است. با این حال، این نوع ارزشیابی به تازگی مورد تأکید قرار گرفته است. به نظر استرانگ[2] (۲۰۰۶) در سالهای اخیر ارزشیابی معلم، به عنوان یکی از جنبه‌های اساسی توسعهٔ مدارس، مورد توجه واقع شده است. به گفتهٔ او "این توجه به ارزشیابی معلم از این شناخت ناشی می‌شود که معلم در قلب هر گونه پیشرفت در نظام آموزشگاهی قرار دارد" (ص XI). و البته همان‌گونه که بین پیشرفت مدارس و عملکرد معلمان یک پیوند منطقی وجود دارد، بین پیشرفت معلمان در کار خود و ارزشیابی از کار آنان نیز یک پیوند منطقی و ضروری موجود است.

ارزشیابی آموزشی، از جمله ارزشیابی معلم، همراه با نهضت **پاسخگویی (مسئولیت‌پذیری)**[3] در آموزش و پرورش ایالات متحد آمریکا و بعضی کشورهای دیگر دنیا اهمیت یافته است. پاسخگویی به این واقعیت اشاره می‌کند که مدارس پاسخگوی محصول خود یعنی یادگیری دانش‌آموزان هستند (هوگان، ۲۰۰۷). کروکشانک[4]، جنکینس[5]، و متکالف[6] (۲۰۰۶) می‌گویند در پاسخ به تقاضای گستردهٔ عمومی در کشور آمریکا ایالتهای مختلف به اجرای نظامهای پاسخگویی اقدام کرده‌اند. این نظامهای جدیدْ مدارس و معلمان را مسئول حفظ سطح مورد قبول پیشرفت تحصیلی دانش‌آموزان می‌دانند. بنا به گفتهٔ استرانگ (۲۰۰۶) "در سالهای اخیر، همزمان با تأکید بیشتر بر پاسخگویی و تحلیل دقیق‌تری از متغیرهای تأثیرگذار بر بازده‌های آموزشی، کاملاً ثابت شده است که معلم مهم‌ترین نیروی مدرسه‌ای مؤثر بر پیشرفت تحصیلی دانش‌آموزان است" (ص ۱).

هدف ارزشیابی معلم

ارزشیابی معلم دارای هدفهای گوناگونی است، اما مهم‌ترین آنها عبارت‌اند از: (۱) کمک به معلمان برای بهبود بخشیدن به عملکرد خودشان، و (۲) تعیین میزان موفقیت آنان در ارتباط با وظیفهٔ معلمی، یعنی پاسخگویی آموزشی.

هدف پاسخگویی آموزشی دربرگیرندهٔ نیاز به تعیین شایستگی معلمان به این منظور است که نشان دهد خدمتهایی که آنان انجام می‌دهند سالم و اثربخش‌اند، و این نوع ارزشیابی نوعاً یک ارزشیابی تراکمی یا مجموعی است. هدف بهبود بخشیدن به عملکرد منعکس‌کنندهٔ نیاز به

1. teacher evaluation
2. Strong
3. accountability
4. Cruckshank
5. Jenkins
6. Metcalf

رشد و توسعهٔ حرفه‌ای فرد فرد معلمان است و این نوعاً یک ارزشیابی تکوینی است (استرانگ، ۲۰۰۶، ص ۴).

تردیدی نیست که ارزشیابی از کار معلم اقدامی مهم و دارای هدفهای والایی است. به همین قیاس، ارزشیابی از کار سایر کارکنان مؤسسه‌های آموزشی، به‌ویژه مدیران، نیز اقدامی ضروری و سازنده است. بنابراین، درست‌تر آن است که ارزشیابی به صورت یک فعالیت نظام‌دار با هدفِ تعیین و تشخیص نقاط قوت و ضعف فعالیتهای همهٔ کارکنان مؤسسهٔ آموزشی طرح و اجرا گردد، زیرا آموزش معلم همواره از عملکرد سایر بخشهای مؤسسهٔ آموزشی اثر می‌پذیرد. علاوه بر این، اگر در مؤسسات آموزشی تنها معلمان مورد ارزشیابی قرار گیرند و فعالیتهای سایر کارکنان به‌ویژه مدیران مورد داوری واقع نشود، آن گاه ارزشیابی معلم قابل دفاع نخواهد بود و اعتراض اعضای هیأتهای علمی و آموزشی را در پی خواهد داشت. بنا به گفتهٔ اسکریون (۱۹۸۱)، "برای جلوگیری از مقاومت کاملاً موّجه 'رُعایا' [معلمان] در برابر ارزشیابیهای به عمل‌آمده از آنان به وسیلهٔ 'اربابان' [مدیران] که خود را ورای این گونه چیزها می‌دانند، ارزشیابی از مدیران مؤسسات آموزشی اقدامی بجاست" (ص ۲۴۵).

تعریف ارزشیابی معلم

ما در فصل اول این کتاب ارزشیابی را به عنوان یک فرایند نظامدار برای جمع‌آوری، تحلیل، و تفسیر اطلاعات به منظور تعیین این مطلب که آیا هدفهای مورد نظر تحقق یافته‌اند یا در حال تحقق‌یافتن هستند و به چه میزانی تعریف کردیم. اگر بخواهیم این تعریف را به ارزشیابی معلم تعمیم دهیم می‌توانیم بگوییم ارزشیابی معلم عبارت است از *تعیین میزان موفقیت معلمان در رسیدن به هدفهای آموزشی خود*| انجام چنین کاری مستلزم جمع‌آوری اطلاعات لازم در بارهٔ فعالیتهای آموزشی معلم و انتخاب معیارهایی برای مقایسهٔ اطلاعات به دست آمده با آن معیارها و سپس قضاوت در این باره است که معلم تا چه میزانی به هدفهای از پیش تعیین‌شده دست یافته است. بنابراین، در ارزشیابی از معلم دو دسته اطلاع ضروری‌اند: (۱) ملاکها یا معیارهای موفقیت یا شایستگی آموزشی و (۲) شاخصی از میزان توفیق معلمان در رسیدن به این ملاکها.

روشهای ارزشیابی معلم

ملاکهای شایستگی آموزشی معلم انواع مختلفی دارند[1] و بسته به اینکه هدف آموزشی معلم را چه چیزی بدانیم، ملاک شایستگی آموزشی او نیز باید متناسب با آن هدف انتخاب شود. از آنجا که برای فعالیتهای آموزشی معلمان هدفهای مختلفی وجود دارند، به همان نسبت ملاکهای شایستگی

۱. در این باره نگاه کنید به مقالهٔ سیف (۱۳۷۲) با عنوان "ملاکهای شایستگی معلم".

آموزشی نیز متنوع‌اند. بنابراین، چون ملاکهای شایستگی آموزشی معلمان متعددند، روشهای ارزشیابی از کار معلمان نیز به همان نسبت متعددند. با وجود تنوع روشهای ارزشیابی موجود، هیچ‌یک از آنها برای ارزشیابی از کار همهٔ معلمان و حتی برای ارزشیابی از کار یک معلم در شرایط متفاوت مفید نیست. بسته به اینکه هدف آموزشی معلم چه باشد و در شرایط خاصی از او چه انتظاری داشته باشیم، روش ارزشیابی متناسب با آن هدف و شرایط و انتظار را باید به کاربندیم. بنابراین، اگر بخواهیم ارزشیابی دقیق و قابل اعتمادی از معلمان به عمل آوریم، لازم است روشهای مختلف ارزشیابی معلم را مورد استفاده قرار دهیم تا تصویر کامل‌تری از نحوهٔ عملکرد معلمان به دست آوریم و به قضاوت معقول‌تری در بارهٔ آنان برسیم. به همین منظور، ما در ادامهٔ این فصل به معرفی روشهای متداول ارزشیابی معلم خواهیم پرداخت. پس از معرفی هر یک از روشهای ارزشیابی معلم، محاسن و معایب آن را نیز مورد بحث قرار خواهیم داد.

چنان‌که گفتیم، روشهای ارزشیابی معلم متنوع‌اند. استرانگ (۲۰۰۶) به عوض **روشهای**[1] ارزشیابی از **منبعهای**[2] ارزشیابی نام می‌برد. مهم‌ترین روشها یا منبعهای ارزشیابی معلم را می‌توان در ۸ دستهٔ زیر قرار داد:

- مصاحبه با معلم
- مشاهده از کلاس درس معلم
- نظر دانش‌آموزان و دانشجویان
- نظر والدین دانش‌آموزان و دانشجویان
- میزان یادگیری یا پیشرفت دانش‌آموزان و دانشجویان
- خودسنجی یا ارزشیابی از خود
- نظر همکاران معلم
- کارپوشه
- آثار علمی معلم

دنبالهٔ این فصل را به توضیح روشهای فوق اختصاص داده‌ایم.

مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلمان

نخستین روشی که به معرفی آن می‌پردازیم روش **مصاحبه**[3] است. این روش می‌تواند به دو منظور مورد استفاده قرار گیرد: یکی برای گزینش معلمان از میان داوطلبان شغل معلمی، دوم برای نگهداری معلمان و دادن امتیازهایی به آنان. ابتدا کاربرد مصاحبه را برای گزینش معرفی می‌کنیم و

1. methods
2. sources
3. interview

آموزشی نیز متنوع‌اند. بنابراین، چون ملاکهای شایستگی آموزشی معلمان متعددند، روشهای ارزشیابی از کار معلمان نیز به همان نسبت متعددند. با وجود تنوع روشهای ارزشیابی موجود، هیچ‌یک از آنها برای ارزشیابی از کار همهٔ معلمان و حتی برای ارزشیابی از کار یک معلم در شرایط متفاوت مفید نیست. بسته به اینکه هدف آموزشی معلم چه باشد و در شرایط خاصی از او چه انتظاری داشته باشیم، روش ارزشیابی متناسب با آن هدف و شرایط و انتظار را باید به کاربندیم. بنابراین، اگر بخواهیم ارزشیابی دقیق و قابل اعتمادی از معلمان به عمل آوریم، لازم است روشهای مختلف ارزشیابی معلم را مورد استفاده قرار دهیم تا تصویر کامل‌تری از نحوهٔ عملکرد معلمان به دست آوریم و به قضاوت معقول‌تری در بارهٔ آنان برسیم. به همین منظور، ما در ادامهٔ این فصل به معرفی روشهای متداول ارزشیابی معلم خواهیم پرداخت. پس از معرفی هر یک از روشهای ارزشیابی معلم، محاسن و معایب آن را نیز مورد بحث قرار خواهیم داد.

چنان‌که گفتیم، روشهای ارزشیابی معلم متنوع‌اند. استرانگ (۲۰۰۶) به عوض **روشهای**[1] ارزشیابی از **منبعهای**[2] ارزشیابی نام می‌برد. مهم‌ترین روشها یا منبعهای ارزشیابی معلم را می‌توان در ۸ دستهٔ زیر قرار داد:

- مصاحبه با معلم
- مشاهده از کلاس درس معلم
- نظر دانش‌آموزان و دانشجویان
- نظر والدین دانش‌آموزان و دانشجویان
- میزان یادگیری یا پیشرفت دانش‌آموزان و دانشجویان
- خودسنجی یا ارزشیابی از خود
- نظر همکاران معلم
- کارپوشه
- آثار علمی معلم

دنبالهٔ این فصل را به توضیح روشهای فوق اختصاص داده‌ایم.

مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلمان

نخستین روشی که به معرفی آن می‌پردازیم روش **مصاحبه**[3] است. این روش می‌تواند به دو منظور مورد استفاده قرار گیرد: یکی برای گزینش معلمان از میان داوطلبان شغل معلمی، دوم برای نگهداری معلمان و دادن امتیازهایی به آنان. ابتدا کاربرد مصاحبه را برای گزینش معرفی می‌کنیم و

1. methods
2. sources
3. interview

بعد استفاده از آن را برای نگهداری مورد بحث قرار می‌دهیم.

در گزینش معلمان آتی از میان داوطلبان در نظرگرفتن سه عامل اساسی زیر ضروری است: (۱) علاقه‌مندی به شغل معلمی، (۲) داشتن ویژگیهای شخصیتی مناسب شغل معلمی، و (۳) داشتن اطلاعات کافی از دانش موضوعی که فرد داوطلب آموزش‌دادن آن است. این عوامل از ضرورتهای شغل معلمی‌اند که داوطلب پیش از ورود به شغل معلمی باید از آنها بهره‌مند باشد، زیرا هیچ مؤسسهٔ آموزشی توانایی آن را ندارد که آنها را به طور تمام و کمال در معلمان داوطلب یا معلمان شاغل ایجاد نماید، ولو که به تأسیس و ادارهٔ دوره‌های آموزشی پیش از خدمت و ضمن خدمت اقدام کند. اما ایجاد مهارتهای آموزشی در معلمان، مانند هر مهارت حرفه‌ای دیگر، هم در دوره‌های آموزشی پیش از خدمت و هم در ضمن خدمت امکان‌پذیر است.

بنابراین، روشهای گزینش معلمان باید بر سه عامل فوق متمرکز باشند. عامل سوم، یعنی داشتن اطلاع کافی از دانش موضوع مورد تدریس، با استفاده از آزمونهای متداول قابل سنجش است و نیاز حتمی به استفاده از روش مصاحبه ندارد. اما دو ویژگی اول، یعنی علاقه‌مندی به شغل معلمی و داشتن ویژگیهای شخصیتی مناسب شغل معلمی با هیچ روش دیگری به خوبی مصاحبه قابل سنجش نیستند. بنابراین، مصاحبه یکی از روشهای ضروری ارزشیابی داوطلبان ورود به شغل معلمی است.

مصاحبه یکی از روشهای شناخته‌شدهٔ پرکاربرد در زمینه‌های بالینی، اجتماعی، و تربیتی است. مصاحبه دارای فنون و قواعد مختلفی است. بحث در بارهٔ فنون و قواعد مختلف مصاحبه در اینجا میسر نیست. علاقه‌مندان می‌توانند به منابع مربوط به آن از جمله هافل[1] (۱۹۸۱)، گرت[2] (۱۹۷۲) و سیف (۱۳۸۷) مراجعه نمایند. با این حال در اینجا ذکر چند نکته در رابطه با کاربرد روش مصاحبه مفید است.

۱. پیش از مصاحبه، با بررسی سوابق و مدارک داوطلبان، به سرند کردن آنان بپردازید و بهترینشان را انتخاب کنید تا در جریان مصاحبه با گروه کوچک‌تری سر و کار داشته باشید و بتوانید برای هر یک از آنها وقت کافی صرف مصاحبه بکنید.

۲. گروه یا کمیتهٔ مصاحبه کننده باید از میان صاحب‌نظران آموزشی و معلمان موفق و باتجربه برگزیده شوند. هافل (۱۹۸۱) پیشنهاد می‌کند که کمیتهٔ مصاحبه کننده برای گزینش استاد در سطح آموزش عالی از مدیر گروه آموزشی و چند تن از اعضای هیأت علمی تشکیل یابد.

۳. پیش از تشکیل جلسات مصاحبه، لازم است مصاحبه کنندگان خط مشی معین و ترتیب خاصی را پیش‌بینی کنند و در بارهٔ آن به توافق برسند و در مورد همهٔ داوطلبان یکنواخت عمل کنند.

۴. مصاحبه کنندگان، در ضمن مصاحبه با داوطلبان، باید بیشتر به دنبال شاخصهای

1. Haefele
2. Garrett

علاقه‌مندی به کار معلمی و تناسب شخصیتی داوطلبان با ویژگیهای شغل معلمی باشند. معلمی مانند هر شغل دیگر ویژگیها و ملزومات خاص خودش را می‌طلبد. این کار از هر کسی ساخته نیست و تنها افراد خاصی هستند که به این کار علاقه‌مندند و شخصیت متناسب با آن را دارند. مصاحبه کنندگان باید در میان داوطلبان به‌دنبال این‌گونه افراد بگردند.

۵. برخی از صاحب‌نظران توصیه کرده‌اند که از داوطلبان معلمی خواسته شود تا در حضور مصاحبه کنندگان (ارزشیابان) به آموزش عملی موضوعی از درسی که داوطلب تدریس آن هستند بپردازند. رعایت این نکته برای استخدام داوطلبان شغل معلمی در سطوح آموزش عالی ضروری‌تر به نظر می‌رسد. هافل (۱۹۸۱) در این باره گفته است در جلسات مصاحبهٔ استخدامی اعضای هیأت علمی دانشگاهها، بهتر است از داوطلب خواسته شود تا در حضور جمعی از استادان و دانشجویان دوره‌های دکتری به تدریس یک موضوع علمی اقدام کند.

نکاتی که در بالا گفته شد ویژهٔ مصاحبه‌های استخدامی برای سنجش تواناییهای داوطلبان ورود به شغل معلمی است. چنان‌که پیشتر گفته شد، از روش مصاحبه برای ارزشیابی از کیفیت کار و تواناییهای آموزشی معلمان شاغل نیز می‌توان سود جست. هدفهای این روش ارزشیابی، مانند سایر روشهای ارزشیابی معلم، فراوان‌اند و دست‌کم موارد زیر را شامل می‌شوند: تعیین و تشخیص نقاط قوت و ضعف کار معلمان و کمک به رفع نقاط ضعف و تشویق و پاداش نقاط قوت.

در استفاده از روش مصاحبه برای ارزشیابی از کیفیت کار معلمان شاغل، باز هم گروهی همکار متخصص که مورد قبول خود معلم و صاحب‌نظر در مسائل ارزشیابی باشند باید جلسات مصاحبه را اداره نمایند. این نوع مصاحبه معمولاً بر اساس اطلاعاتی که از سایر روشهای ارزشیابی معلم گردآوری شده‌اند هدایت می‌شود و مصاحبه کنندگان اطلاعات به دست‌آمده از سایر روشهای ارزشیابی را با روش مصاحبه تکمیل می‌کنند. در استفاده از روش مصاحبه برای ارزشیابی از کار معلمان شاغل رعایت دو نکتهٔ زیر مورد تأکید صاحب‌نظران بوده است:

۱. لازم است بین مصاحبه کنندگان (ارزشیابان) و مصاحبه‌شونده (معلم) بر سر هدفهای ارزشیابی و اینکه مصاحبه چگونه به این هدفها ربط پیدا می‌کند توافق حاصل شود. لازم است هدف مصاحبه برای معلمان شاغل از پیش به دقت توضیح داده شود و فواید و پیامدهای آن به اطلاع معلم رسانیده شود و موافقت و همکاری او برای شرکت در مصاحبه جلب گردد.

۲. در اجرای مصاحبه برای ارزشیابی از آموزش معلمان شاغل بهتر است از نتایج روشهای دیگر ارزشیابی معلم مانند نظرخواهی از دانش‌آموزان و دانشجویان و مشاهدهٔ کلاسی نیز سود جست و در جریان مصاحبه نتایج این ارزشیابیها را به اطلاع معلم رسانید و در بارهٔ آنها با او به بحث و گفتگو پرداخت. نکته‌ای را که در این باره باید با دقت زیاد در نظر گرفت این است که مصاحبه کنندگان یا ارزشیابان نتایج سایر ارزشیابیهای انجام‌شده در بارهٔ معلم مانند نتایج

مشاهدات کلاسی و نظر دانشجویان او را، بدون تهدید و به نحوی کاملاً دوستانه، با او در میان بگذارند. معلمان مصاحبه‌هایی راکه بر اقدامات بهبودبخش تأکید می‌کنند و کمتر جنبهٔ عیب‌جویی دارند کاملاً پذیرا می‌شوند. برای پیشگیری از هر گونه اتفاق ناگوار در جریان مصاحبه، بهتر است نسخه‌ای از گزارش نتایج ارزشیابیهای معلم از قبل در اختیار او گذاشته شود. هافل (۱۹۸۱) می‌گوید "معلمان غالباً از این کار استقبال می‌کنند زیرا طرح کردن جنبه‌های ضعیف آموزش معلم در جلسهٔ مصاحبه، به طور غیرمترقبه، ممکن است او را در شرایط ناگواری قرار دهد" (ص ۵۲).

محاسن و معایب روش مصاحبه

روش مصاحبه را نمی‌توان یک روش کامل ارزشیابی معلم دانست و به هیچ‌وجه نباید از آن به تنهایی استفاده کرد، زیرا در جلسات مصاحبه ویژگیهایی از معلم مورد ارزیابی قرار می‌گیرند که ممکن است با ملاکهای واقعی اثربخشی معلم، از جمله یادگیری شاگردان، ارتباط چندانی نداشته باشند. پژوهشهای زیادی رابطهٔ بین ویژگیهای کلی معلمان و یادگیری شاگردان را بررسی کرده و به نتایج خیلی مفیدی دست نیافته‌اند. شالوک[1] (۱۹۷۹) این گونه پژوهشها را مورد بازنگری قرار داده و نشان داده است که ویژگیهای شخصیتی معلمان رابطهٔ چندانی با یادگیری شاگردان ندارند.

با وجود کم و کسرهای روش ارزشیابی معلم از طریق مصاحبه، این روش اگر به درستی انجام شود می‌تواند فواید زیادی در بر داشته باشد. مخصوصاً در ارتباط با انتخاب معلمان واجد شرایط از میان داوطلبان، مصاحبه یک امر اجتناب‌ناپذیر است. روش مصاحبه نه‌تنها به مسئولان مؤسسهٔ آموزشی کمک می‌کند تا از میان داوطلبان شغل معلمی آنانی را که شرایط مناسب‌تری دارند انتخاب کنند، بلکه به داوطلبان نیز کمک می‌کند تا آگاه شوند شغلی که داوطلب آن هستند انتظاراتشان را برآورده می‌سازد یا نه.

افزون بر امتیاز فوق، مصاحبه روش مناسبی برای استفاده از نتایج سایر روشهای ارزشیابی معلم است. بهترین استفاده‌ای که از مصاحبه می‌توان به عمل آورد این است که، پس از جمع‌آوری نتایج ارزشیابیهایی که با روشهای مختلف از کار معلم به دست آمده است، در یک مصاحبهٔ دوستانهٔ رو در رو، نقاط قوت و ضعف کارش با او در میان گذاشته شود و در رفع مشکلاتش هدایت گردد.

مشاهده از کلاس درس

یکی از روشهای بسیار متداول ارزشیابی معلم روش مشاهده[2] است. در این روش، اطلاعات مورد نیاز، به طور مستقیم، از اعمال و رفتار معلم در کلاس درس گردآوری می‌شوند. عمده‌ترین شاخصهای اثربخشی معلم که مشاهده‌کننده (ارزشیاب) به مشاهده و ثبت و ارزشیابی آنها می‌پردازد به قرار زیرند:

1. Schalock
2. observation

۱. انتخاب روشهای آموزشی متناسب با هدفهای آموزشی درس
۲. استفاده از روشهای متنوع آموزشی، از جمله روش سخنرانی، بحث و گفتگو، پرسش و پاسخ، و غیره
۳. رعایت ویژگیهای فردی یادگیرندگان
۴. استفاده از تهیهٔ طرح درس و سازمان‌بندی درست مطالب
۵. درست اداره‌کردن کلاس و استفادهٔ حداکثر از وقت کلاس

معمولاً مشاهده از کلاس درس معلم به دو صورت **غیررسمی**[1] و **رسمی**[2] صورت می‌پذیرد. مشاهدهٔ غیررسمی "سری به کلاس زدن"، "خبری از کار معلم گرفتن"، و "دیداری از کلاس به عمل آوردن" نیز نامیده می‌شود (زپدا[3]، ۲۰۰۶). در این روش، مدیر، راهنمای آموزش، یا متخصص ارزشیابی روزانه دو سه نوبت به کلاس معلم سرکشی می‌کند و هر نوبت چهار پنج دقیقه به مشاهدهٔ کار او می‌پردازد. "این روش امکان گردآوری داده‌های متنوعی را دربارهٔ رویدادهای کلاسی فراهم می‌آورد، و به عنوان یک روش ارزشیابی فرصت کسب نمونه‌های مختلفی از آموزش معلم را به مسئولان آموزشگاه می‌دهد" (زپدا، ۲۰۰۶، ص ۱۱۲).

مشاهدهٔ رسمی از کلاس درس به زمان بیشتری نیازمند است و در آن روشها و فنون مختلفی مورد استفاده قرار می‌گیرند و نمونه‌های بزرگ‌تری از آموزش معلم ثبت و ضبط می‌شوند. مهم‌ترین روشها و فنون جمع‌آوری اطلاعات در روش مشاهدهٔ رسمی به قرار زیرند:

۱. روش تعیین فراوانی یا دفعات رفتار. روزنشاین[4] و فرست[5] (۱۹۷۳) چندین روش تعیین فراوانی یا تعیین تعداد رفتارها یا رویدادهای کلاسی را معرفی کرده‌اند. روشی که وقوع رفتار، رویداد، یا تعامل خاصی را تعیین می‌کند روش طبقه‌بندی نام دارد. در این روش، طبقات مختلفی از رویدادهای کلاسی تعریف می‌شوند و مشاهده‌گر فراوانی یا دفعات وقوع آنها را در کلاس درس معلم مورد مشاهده قرار داده آن را ثبت می‌نماید.

۲. روش درجه‌بندی. در این روش مشاهده‌گر، در طول مدت مشاهده از کلاس معلم، میزان وقوع یک رفتار خاص را به صورت درجه‌بندی (مثلاً کم، زیاد، متوسط) ثبت می‌کند. این روش عموماً برای ثبت عوامل آموزشی که جنبهٔ استنباطی و تفسیری بیشتری از سوی مشاهده‌گر دارند به کار می‌رود، مثلاً میزان گرمی کلام یا اشتیاق معلم به کار آموزشی خود.

۳. روش نقلی. این روش به روش واقعه‌نگاری معرفی شده در فصل ۱۴ شبیه است. در این روش مشاهده‌گر یک گزارش نسبتاً کامل و طبیعی از آنچه در کلاس درس معلم رخ می‌دهد تهیه

---

1. informal
2. formal
3. Zepeda
4. Rosenshine
5. Furst

می‌نماید، اما برخلاف روشهای بالا هیچ‌گونه طبقه‌بندی یا تعبیر و تفسیری به گزارش خود نمی‌افزاید. هر چند که در روش نقلی هدف ثبت کامل وقایع مشاهده‌شده است، با این حال مشاهده‌گران خبره می‌توانند در نقل وقایع بسیار به اختصار و در عین حال دقیق عمل کنند و از حشو و زواید بکاهند و تنها بر وقایع مهمی که در کلاس درس رخ می‌دهند تأکید بورزند. هولی[1] و اورستون[2] (۱۹۸۱) روش نقلی را از دو جهت بر روشهای دیگر ترجیح می‌دهد: (۱) از جهت رویکرد طبیعی آن به آموزش و سهولت گزارش آن به دیگران، و (۲) از جهت گستردگی آن که موجب می‌شود هر شخصی که گزارش مشاهده‌گر را مطالعه می‌نماید وضعیت آموزشی او را به صورت یک واحد کامل مورد ارزیابی قرار دهد.

مشکل نمونه گیری از آموزش معلم به کمک مشاهده

یکی از مشکلات روش مشاهدهٔ کلاسی این است که نمونهٔ کاملی از آموزش معلم به دست نمی‌دهد. برای رفع این مشکل، لازم است آموزش معلم بیش از یک بار مورد مشاهده قرار گیرد. هر چه تعداد دفعات مشاهده از کلاس معلم بیشتر باشد، به همان نسبت ثبات و پایایی نتایج حاصل نیز افزایش می‌یابند. راولی[3] (۱۹۷۸) نشان داده است که استفاده از تعداد زیادی مشاهدهٔ کوتاه‌مدت از تعداد کمی مشاهدهٔ درازمدت پایایی بیشتری را به دست می‌دهد. استودولسکی[4] (۱۹۹۱) گفته است "گر چه به ظاهر چنین به نظر می‌رسد که مشاهدهٔ یک درس کامل معلم از مشاهدهٔ قسمتهای کوتاه درسها بهتر است، اما از لحاظ پایایی نتایج مشاهدات کوتاه‌مدت شواهد بهتری به دست می‌دهند" (ص ۱۸۳).

مسئلهٔ پایایی و ثباتِ نتایج حاصل از مشاهدات کلاسی در ارزشیابیهایی که به منظور تصمیم‌گیری در بارهٔ معلمان به کار برده می‌شوند باید قویاً مورد توجه قرار گیرد، زیرا نتایج حاصل از جلسات مشاهده باید قابل تعمیم به همهٔ انواع آموزشهای معلم باشند. اما اگر قرار است که نتایج مشاهده برای بهبود آموزش معلم به کار روند، در آن صورت ناپایایی و تعمیم‌ناپذیری نتایج چندان نگران‌کننده نخواهد بود. یک مشاهدهٔ واحد یا تعداد کمی مشاهده در رابطه با هر درس معلم می‌تواند منجر به پیشنهادهای مفیدی به وی گردد، به شرط آنکه فراموش نشود که این مشاهدات تصویر کاملی از آموزش معلم به دست نمی‌دهند.

محاسن و معایب روش مشاهدهٔ کلاسی

روش ارزشیابی معلم با استفاده از مشاهدهٔ کلاسی روش کاملی نیست و انتقادهای زیادی بر آن وارد شده است. اسکریون (۱۹۸۱) یکی از صاحب‌نظران حوزهٔ ارزشیابی معلم در این باره گفته است "استفاده از دیدارهای کلاسی توسط همکاران معلم (یا مدیران و متخصصان) برای ارزشیابی

---

1. Holley
2. Everston
3. Rowley
4. Stodolsky

از آموزش نه‌تنها نادرست بلکه توهین‌آمیز است" (ص ۲۵۱). از جمله ایرادهای اسکریون از مشاهدهٔ کلاسی یکی این است که رفتن به کلاس درس معلم جریان طبیعی آموزش او را به هم می‌زند و مانع می‌شود که مشاهده‌گر نمونهٔ درستی از آموزش معلم به دست آورد. دوم اینکه تعداد دیدارها از کلاس درس حتی اگر تصادفی هم باشند نمی‌توانند آن قدر زیاد انجام پذیرند که قابل تعمیم به کل آموزش معلم باشند. سوم اینکه مشاهده‌گران معمولاً در مشاهده از آموزش و درس معلم خالی از تعصب عمل نمی‌کنند. و بالاخره اینکه هیچ‌چیز را در کلاس درس نمی‌توان مشاهده کرد که به عنوان پایه‌ای برای یک نتیجه‌گیری درست در بارهٔ شایستگی معلم بتوان از آن سود جست. پیترسون[1] (۲۰۰۴) انتقادهای وارد شده به روش مشاهدهٔ کلاسی را به صورت زیر خلاصه کرده است:

> پژوهشهای انجام شده دربارهٔ روش ارزشیابی معلم از راه مشاهده و گزارش توسط مدیران نشان داده‌اند که این روش نمونهٔ غیرمُعرفی از تدریس و کلاسداری معلم به دست می‌دهد؛ گزارشهای حاصل از این مشاهده‌ها سوگیری دارند؛ حضور مشاهده‌گر در کلاس درس در کار آموزشی معلم اخلال ایجاد می‌کند؛ و ویژگیهای گمراه‌کنندگی یا بی‌سروته بودن نظامهای گزارش نتایج مشاهده، مانند فهرست‌وارسی و واقعه‌نگاری‌های مختصر، نتایج مشاهدات کلاسی را محدود می‌سازند. (ص ۱۱۱)

به رغم انتقادهای بالا، گروهی دیگر از صاحب‌نظران از روش مشاهده به عنوان وسیله‌ای مفید برای ارزشیابی از کار معلم دفاع کرده‌اند. برای مثال، هولی (۱۹۸۱) گفته است روش مشاهده اطلاعاتی در اختیار ما می‌گذارد که از هیچ روش دیگری به دست نمی‌آید. بنا به گفتهٔ او، "تنها مشاهدهٔ کلاسی است که به ما دید درستی از جو کلاس، ارتباط و تعامل معلم و شاگرد، و کارکردهای دیگر کلاس به دست می‌دهد. بنابراین، درک بهتر مشکلات روش کار معلم و یافتن راههای درست آن تنها از راه مشاهدهٔ مستقیم کلاس درس او میسر است" (ص ۹۰).

گذشته از آنچه گفته شد، از آنجاکه نقش عمدهٔ مدیر و راهنمای آموزشی هدایت فعالیتهای آموزشی معلم است و این کار بدون مشاهدهٔ معلم هنگام آموزش امکان‌پذیر نیست، پس مشاهدهٔ کلاس درس به عنوان یک روش ارزشیابی، به‌ویژه با هدفِ هدایتِ معلم و کمک به رفع مشکلات آموزشی او، روشی مهم و اجتناب‌ناپذیر است. افزون بر این، با تأکید بر پاسخگویی یا مسئولیت‌پذیری در سالهای اخیر (به‌ویژه در کشور ایالات متحد آمریکا)،

> مدیران و دیگران این روزها بیشتر راهی کلاسهای درس می‌شوند تا به معلمان در کسب و بهبود مهارتهای آموزشی کمک کنند با این هدف که به پیشرفت تحصیلی دانش‌آموزان افزوده شود. فعالیتهای مربوط به یاری دادن معلمان باید از جایی آغاز شوند که مؤثرترین‌اند ــ و آن کلاس درس

---

1. Peterson

است. به منظور استفاده از این گونه حمایتها در راستای راهنمایی آموزشی به عنوان وسیله‌ای برای کمک به معلمان در بهبود فعالیتشان هیچ جایگزینی برای مشاهدهٔ کلاس درس قابل تصور نیست. (زپدا، ۲۰۰۶، ص ۱۱۲)

صرف‌نظر از موضع‌گیریهای موافق و مخالف صاحب‌نظران نسبت به روش مشاهدهٔ کلاسی، استفاده از این روش برای ارزشیابی از کار معلم، چه به منظور بهبود کیفیت کار او و چه به منظور اتخاذ تصمیمهای شغلی در بارهٔ معلم صورت پذیرد، روش مفیدی است. با این حال، هرگز نباید از آن به تنهایی استفاده کرد، خصوصاً اگر قرار است در بارهٔ وضعیت شغلی معلم تصمیمی اتخاذ شود. کار درست آن است که نتایج حاصل از روش مشاهدهٔ کلاسی با نتایج حاصل از روشهای دیگر ارزشیابی معلم کامل شوند. استودولسکی (۱۹۹۱) در رابطه با نقش مشاهدهٔ کلاسی در ارزشیابی معلم گفته است:

حذف مشاهدهٔ مستقیم از ارزشیابی معلمان غیرقابل تصور است. چگونه کسی می‌تواند از معلمان ارزشیابی به عمل آورد، بدون اینکه شاهد فعالیت آنها در کلاس درس باشد؟ با این حال، نباید به ارزشیابی تا آن اندازه اهمیت داده شود که بسیاری جنبه‌های غیرقابل مشاهده و مهم آموزش نادیده گرفته شوند. شاید یک ارزشیابی آرمانی از معلم آن باشد که منابع چندگانهٔ اطلاعاتی در بارهٔ آموزش معلمان و تأثیراتی که بر شاگردانشان به جای می‌گذارند را با هم به کار گیرد. (ص ۱۸۸)

استفاده از نظر دانش‌آموزان و دانشجویان

در روش *استفاده از نظر دانش‌آموزان و دانشجویان*، مدیران و مسئولان مؤسسه‌های آموزشی، به طور رسمی یا غیررسمی، نظر دانش‌آموزان یا دانشجویان را در بارهٔ ویژگیهای مختلف شخصی و حرفه‌ای معلمان مؤسسهٔ خود جویا می‌شوند و بر اساس نظر آنان به قضاوت در بارهٔ میزان شایستگی معلمان می‌پردازند. به رغم اختلاف نظرهای زیادی که در بارهٔ استفاده از نظرهای شاگردان برای ارزشیابی از معلمان وجود دارد، این روش کاربرد زیادی یافته است و برای ارزشیابی معلم، به‌ویژه در سطوح آموزش عالی، مورد استفاده قرار می‌گیرد. بنا به نظر تورن[1]، اسکات[2]، و بیرد[3] (۱۹۷۶) و اُکانل[4] و اسمارت[5] (۱۹۷۹)، ظاهراً استفاده از نظر دانشجویان برای ارزشیابی از کار استادان دانشگاهها تنها منبع اطلاعاتی ملموس در دانشکده‌ها و دانشگاههای سراسر جهان است. آلیمونی[6] (۱۹۸۱) می‌گوید "با توجه به نارضایتی اکثریت استادان از روش نظرخواهی از دانشجویان، این یک وضع نگران‌کننده را ایجاد کرده است" (ص ۱۱۰). با این حال، اگر در روش استفاده از نظر دانشجویان احتیاطهای لازم به کار بسته شود و نتایج حاصل عمدتاً به صورت یک

1. Thorne
2. Scott
3. Beaird
4. O'Connel
5. Smartt
6. Aleamonie

منبع بازخوردی برای اصلاح کار معلمان مورد استفاده قرار گیرد (یعنی به صورت ارزشیابی تکوینی به کار رود) و برای مقاصد تصمیم‌گیری شغلی (یا ارزشیابی تراکمی) نتایج این روش ارزشیابی با نتایج حاصل از روشهای دیگر ارزشیابی معلم تکمیل گردد جای نگرانی وجود نخواهد داشت. در یکی از پژوهشهایی که تأثیر بازخوردی نتایج حاصل از نظرسنجی دانشجویان را به اثبات رسانیده، مارش[1] و روچه[2] (۱۹۹۳) نشان داده‌اند که بازخورد حاصل از ارزشیابی دانشجویان از اثربخشی آموزش اگر با مشورت و راهنمایی همراه گردد در بالابردن سطح کیفی آموزش مؤثر است.

هر چند به راههای مختلفی می‌توان نظر شاگردان را نسبت به معلمانشان جویا شد، بهترین و متداول‌ترین راه نظرخواهی از شاگردان برای ارزشیابی از کار معلمان، تهیه و اجرای پرسشنامه‌های نظرخواهی است. پرسشنامه‌های نظرخواهی شاگردی دارای شکلهای مختلفی است که در زیر به مهم‌ترین آنها اشاره می‌کنیم.

پرسشنامه‌های نظرخواهی از دانش‌آموزان

هر چند که روش استفاده از نظر شاگردان برای ارزشیابی از کار معلمان عمدتاً در سطح آموزش عالی مورد استفاده قرار می‌گیرد، اما کاربرد آن در دبستان و دبیرستان نیز متداول شده است و اینک در بسیاری از کشورهای جهان، به‌ویژه در ایالات متحد آمریکا، به عنوان یک منبع اطلاعاتی برای ارزشیابی تکوینی مورد استفاده قرار می‌گیرد. شکلهای ۱ـ۲۲ و ۲ـ۲۲ دو نمونه از پرسشنامه‌های نظرخواهی دانش‌آموزی را نشان می‌دهند. دقّت کنید که فرم شکل ۱ـ۲۲ ویژهٔ دانش‌آموزان پایه‌های

راهنمایی: پس از آنکه معلم هریک از سؤالها را می‌خواند، صورتی را که نشان می‌دهد دربارهٔ آن سؤال چگونه فکر می‌کنید با مداد رنگی رنگ کنید.

☹	😐	☺
☹	😐	☺
☹	😐	☺
☹	😐	☺
☹	😐	☺
☹	😐	☺
☹	😐	☺

سؤال نمونه: من با اتوبوس مدرسه به مدرسه می‌روم.

۱. معلم من به حرفهای من گوش می‌دهد.
۲. وقتی که به کمک نیاز دارم معلمم کمکم می‌کند.
۳. معلم من به من نشان می‌دهد که چگونه کارهای تازه انجام دهم.
۴. می‌دانم که در کلاس چه کار باید بکنم.
۵. می‌توانم کارهای کلاس را انجام دهم.
۶. من در کلاسم چیزهای تازه یاد می‌گیرم.

شکل ۱ـ۲۲ یک پرسشنامهٔ نظرخواهی از دانش‌آموزان پایه‌های ۲ـ۱ (ویژهٔ دانش‌آموزان مدارس دولتی یکی از ایالتهای آمریکا، به نقلِ استرانگ و اُستراندر، ۲۰۰۶، ص ۱۳۸)

1. Marsh 2. Roche

راهنمایی: با انتخاب اعداد ۱ تا ۵ برای جمله‌های زیر، نظر خود را نسبت به این کلاس و معلم آن مشخص کنید. برای کاملاً مخالف عدد ۱، مخالف عدد ۲، بی‌نظر عدد ۳، موافق عدد ۴، و کاملاً موافق عدد ۵ را انتخاب کنید. جای خالی پایین صفحه برای سایر اظهارنظرهای شما پیش‌بینی شده است.
نام خود را در این پرسشنامه ننویسید.

سؤال نمونه: دوست دارم موسیقی گوش کنم.

	کاملاً مخالف	مخالف	بی‌نظر	موافق	کاملاً موافق
	۵	۴	۳	۲	۱
معلم من در این درس:					
۱. راهنماییهای روشنی می‌دهد.	☐	☐	☐	☐	☐
۲. با همه عادلانه رفتار می‌کند.	☐	☐	☐	☐	☐
۳. در خارج از کلاس برای کمک گرفتن به او دسترسی داریم.	☐	☐	☐	☐	☐
۴. هدفهای درس را به روشنی بیان می‌کند.	☐	☐	☐	☐	☐
۵. تکالیف ما را به موقع تصحیح می‌کند و به ما بازمی‌گرداند.	☐	☐	☐	☐	☐
۶. مطالب این درس را به درسهای دیگر و زندگی واقعی ربط می‌دهد.	☐	☐	☐	☐	☐
۷. اجازهٔ ابراز عقاید مختلف می‌دهد.	☐	☐	☐	☐	☐
۸. همهٔ دانش‌آموزان را به یادگیری تشویق می‌کند.	☐	☐	☐	☐	☐
۹. در کلاس از فعالیتهای مختلفی استفاده می‌کند.	☐	☐	☐	☐	☐
۱۰. توضیحاتش را به خوبی می‌فهمم.	☐	☐	☐	☐	☐
۱۱. کلاس را به خوبی اداره می‌کند.	☐	☐	☐	☐	☐
۱۲. برای همهٔ دانش‌آموزان احترام قائل است.	☐	☐	☐	☐	☐
۱۳. مقررات انضباطی را عادلانه اعمال می‌کند.	☐	☐	☐	☐	☐
۱۴. می‌کوشد تا از وقت کلاس برای یادگیری استفاده شود.	☐	☐	☐	☐	☐
۱۵. در زمینهٔ موضوعی که تدریس می‌کند اطلاعات کافی دارد.	☐	☐	☐	☐	☐
۱۶. چگونگی انجام پروژه‌های تحقیقی را به خوبی توضیح می‌دهد.	☐	☐	☐	☐	☐
۱۷. سطح انتظارات کلاس را بالا می‌برد.	☐	☐	☐	☐	☐
۱۸. کمک می‌کند تا به انتظاراتی که برای ما تعیین کرده است برسیم.	☐	☐	☐	☐	☐
۱۹. به ما تکالیف متناسب با درس می‌دهد.	☐	☐	☐	☐	☐
۲۰. با ما صادقانه ارتباط برقرار می‌کند.	☐	☐	☐	☐	☐

اظهارنظر:

شکل ۲ـ۲۲ یک پرسشنامهٔ نظرخواهی از دانش‌آموزان پایه‌های ۱۲ـ۹ (دانش‌آموزان مدارس دولتی یکی از ایالتهای آمریکا، به نقلِ استرانگ و استراندر، ۲۰۰۶، ص ۱۳۹)

اول و دوم دبستان است. بنابراین، هنگام اجرا، معلم صورت سؤالها را برای شاگردان می‌خواند و آنها برای هر سؤال یکی از شکلهای سه گانه را که نشان دهندهٔ *موافق*، *مخالف*، یا *بی‌نظر* است به عنوان جواب سؤال برمی‌گزینند. فرم شکل ۲ـ۲۲ برای استفادهٔ دانش‌آموزان پایه‌های ۱۲ـ۹ تهیه شده است، و چنان‌که خواهید دید به فرمهای مورد استفاده در سطح دانشگاه شبیه‌ترند (استرانگ و اُستراندر[1]، ۲۰۰۶).

1. Ostrander

پرسشنامه‌های نظرخواهی از دانشجویان

پرسشنامه‌های نظرخواهی از دانشجویان به عنوان ابزاری برای جمع‌آوری اطلاعات دربارهٔ کم و کیف فعالیتهای آموزشی اعضای هیأت علمی در آموزش عالی فراوان مورد استفاده قرار می‌گیرند. به همین دلیل، شکلهای مختلف این گونه پرسشنامه‌ها برای استفاده‌های مختلف تهیه شده‌اند. ما در اینجا شناخته‌شده‌ترین نمونه‌های آنها را معرفی می‌کنیم.

ساده‌ترین شکل پرسشنامهٔ نظرخواهی دانشجویی پرسشنامه‌ای است که استاد برای سنجش نظرهای دانشجویان خود در یک درس خاص به کارمی‌گیرد. این‌گونه پرسشنامه‌ها برای دروس معین بسیار مفید، آموزنده، و قابل استفاده است. معلم با اجرای چنین پرسشنامه‌ای در یکی از کلاسهای درس خود می‌تواند نظر دانشجویانش را نسبت به ویژگیهای آموزش خود در یک درس خاص به دست آورد. بنابراین، نتایج حاصل از این پرسشنامه‌ها برای مقاصد بازخوردی و اصلاحی می‌توانند بسیار زیاد مورد استفادهٔ معلم واقع شوند. اما این‌گونه پرسشنامه‌ها نتایج قابل تعمیم برای معلمان دیگر و شاید برای درسهای دیگر همان معلم به دست نمی‌دهند. لذا، برای به‌دست آوردن نتایج قابل تعمیم‌تر باید پرسشنامه‌های دیگری را تهیه و اجراکرد. شکل ۳ـ۲۲ نمونهٔ پرسشنامهٔ قابل استفاده برای دروس خاص است.

پرسشنامه‌های نظرخواهی ویژهٔ دروس خاص، نظیر نمونهٔ شکل ۳ـ۲۲، صرفاً به عنوان ارزشیابی تکوینی برای مقاصد بازخورد و اصلاح و بهسازی آموزش مفیدند، اما نمی‌توان آنها را برای تصمیم‌گیریهای شغلی مورد استفاده قرار داد. اگر قرار باشد از نظر شاگردان برای ارزشیابی از کار معلم به عنوان ارزشیابی تراکمی به منظور تصمیم‌گیری در بارهٔ استخدام قطعی، ارتقا و غیره استفاده شود، لازم است پرسشنامه‌های جامع‌تری را تهیه و اجراکرد که سؤالهای آن کلی‌تر و تعمیم‌پذیرتر باشند. این نوع پرسشنامه‌ها باید طوری تهیه شوند که برای همهٔ دروس و همهٔ معلمان قابل استفاده باشند. شکل ۴ـ۲۲ نمونه‌ای از این گونه پرسشنامه‌های کلی را نشان می‌دهد.

یکی از پرسشنامه‌های جامع و چند بعدی که برای اکثر دروس غالب مدرسان دانشگاهی قابل استفاده است، پرسشنامهٔ *ارزشیابی دانشجویان از کیفیت آموزشی*[1]، معروف به (اس ای ای کیو)، است. این پرسشنامه که در دانشگاه سیدنی غربی کشور استرالیا تهیه شده و اثربخشی آن به عنوان یک وسیلهٔ مفید بهسازی آموزشی مدرسان دانشگاهی در پژوهشهای مختلف به تأیید رسیده (از جمله مارش و روچه، ۱۹۹۳) دارای ۹ بخش به شرح زیر است:

۱. ارزش علمی ـ یادگیری درس، شامل ۴ سؤال در بارهٔ ارزش و اعتبار محتوای درس و اهمیت علمی آن، نظیر این سؤال: "شما از این درس چیز با ارزشی یاد گرفتید."

---

1. Student's Evaluation of Educational Quality (SEEQ)

پرسشنامهٔ بهسازی آموزش

ویژهٔ درس امار

لطفاً از ذکر نام و مشخصات فردی خودداری کنید.

۱. در حال حاضر، ۱۲/۵٪ وقت کلاس صرف مرور درس می‌شود، ۱۲/۵٪ صرف مرور آزمون می‌شود، و ۷۵٪ صرف آموزش می‌شود. به نظر شما این زمان‌بندی مناسب است؟ _____ اگر جوابتان منفی است، لطفاً تقسیم‌بندی زمانی مناسبی را که در نظر دارید بنویسید. _____

۲. آیا معمولاً یک تمرین یادگیری را، پیش از نگاه کردن به جواب داده‌شده برای آن، جواب می‌دهید؟ ___ (بله یا نه)

۳. تا چه اندازه نمی‌توانید بفهمید که چرا جوابهای داده‌شده برای تمرینهای یادگیری درست‌اند؟ ___ (توجه: منظور ما این نیست که شما به تمرینها جواب درست می‌دهید، بلکه تا چه اندازه نمی‌توانید بفهمید که چرا جوابها درست‌اند؟)

۴. اگر در مقابل جواب داده‌شده برای هر تمرین، منبعی که مطالعهٔ آن درک شما را از مطلب یادگیری افزایش می‌دهد داده شود، به شما کمک می‌کند؟

الف) خیلی کمک می‌کند
ب) کمک زیادی نخواهد کرد

۵. با فرض اینکه مقدار مطلب یادگیری تغییری نکند، کدام‌یک از موارد زیر را ترجیح می‌دهید:

الف) هدفها کمتر ولی جامع‌تر
ب) هدفهای بیشتر و دقیق‌تر
ج) همان تعداد هدفهای فعلی

۶. تعداد تمرینهای یادگیری مربوط به هر هدف:

الف) خیلی کم است (کاش بیشتر بودند)
ب) خیلی زیاد است
ج) کافی است

۷. نظرتان نسبت به انواع مختلف امتحانهای مختصر هفتگی چیست؟ _____

۸. بخشی از وقت آموزش صرف توضیح مفاهیم کتاب یا جزوه و بخش دیگر صرف کاربردهای این مطالب می‌شود. آیا با این تقسیم‌بندی موافق‌اید؟ _____ اگر جوابتان منفی است، توضیح دهید بهتر است وقت بیشتری صرف کدام قسمت بشود. _____

۹. برای کمک بیشتر، به معلم این درس در خارج از کلاس مراجعه کرده‌اید؟ _____ اگر مراجعه کرده‌اید، فکر می‌کنید با آمادگی و علاقه با شما برخورد کرده است؟

۱۰. آیا چیزی در نظر داشته‌اید که فکر می‌کردید از این درس (درس آمار) یاد خواهید گرفت، اما یاد نگرفتید؟ _____ اگر جوابتان مثبت است، لطفاً بنویسید آن مطلب چیست.

۱۱. لطفاً هر گونه نظر دیگری دارید که فکر می‌کنید به بهسازی این درس کمک می‌کند را بنویسید.

شکل ۳-۲۲ یک پرسشنامهٔ نظرخواهی از دانشجو ویژهٔ یک درس خاص (اقتباس از اثر آلیمونی، ۱۹۸۱)

شکل ۴ـ۲۲ یک پرسشنامهٔ نظرخواهی از دانشجو ویژهٔ درسهای مختلف معلمان (سیف، ۱۳۵۶)

فرم ارزشیابی فعالیتهای آموزشی معلمان و محتوای دروس

دانشجوی عزیز:

این دفترچه برای سنجش عقاید شما نسبت به روش تدریس و سایر فعالیتهای آموزشی معلمان و موضوعهای درسی شما تهیه شده است. نتایج این فرم ارزشیابی برای اصلاح روش تدریس معلمان و ارزشیابی از کار آنان و همچنین برای تجدید نظر و اصلاح محتوای دروس مؤثر واقع خواهد شد. دقت و قضاوت صحیح شما در تکمیل این فرم بسیار ضروری و مایهٔ سپاسگزاری است.

لطفاً از نوشتن نام خود در این فرم ارزشیابی و پاسخنامهٔ مربوط خودداری کنید. معلمان شما پس از اینکه نمرات شما را به ادارهٔ آموزش تحویل دادند از نتیجهٔ این ارزشیابی مطلع خواهند شد. بنابراین، نتایج این ارزشیابی در نمرات شما هیچ‌گونه تأثیری نخواهد داشت.

بخش اول: ارزشیابی فعالیتهای آموزشی معلم

عبارتهای این بخش به منظور ارزشیابی فعالیتهای آموزشی معلمان شما تهیه شده‌اند. با گذاشتن علامت (x) در محل خالی مربوط به حروف الف، ب، ج، یا د در پاسخنامه فعالیتهای آموزشی معلمان خود را ارزشیابی کنید.

اگر **الف** را برای عبارتی انتخاب کنید، منظور شما این است که با آن عبارت **کاملاً موافق** هستید.

اگر **ب** را برای عبارتی انتخاب کنید، منظور شما این است که با آن عبارت **موافق** هستید.

اگر **ج** را برای عبارتی انتخاب کنید، منظور شما این است که با آن عبارت **مخالف** هستید.

اگر **د** را برای عبارتی انتخاب کنید، منظور شما این است که با آن عبارت **کاملاً مخالف** هستید.

فقط یکی از حروف فوق را برای هر عبارت انتخاب کنید. عبارتها را به دقت بخوانید و توجه داشته باشید که عبارتهای این بخش صرفاً برای ارزشیابی **روش تدریس و فعالیتهای آموزشی معلم** شما تهیه شده‌اند، نه ارزشیابی شخصیت او. دقت کنید که قضاوت شما در بارهٔ روش تدریس معلم تحت تأثیر محتوا یا موضوع درس و مواد آموزشی (چون کتاب درسی و جزوهٔ درسی) که به‌طور جداگانه در بخش دوم به وسیلهٔ شما ارزشیابی خواهند شد قرار نگیرد. چنانچه به عبارتی برخورد کردید که با یک درس یا معلم بخصوصی مناسبت نداشت، آن عبارت را نادیده بگیرید. فقط برای آن عبارتهایی که به نحوی با درس و معلم شما مربوط هستند حروف الف، ب، ج، و یا د را انتخاب کنید و در پاسخنامه علامت بزنید.

۱. هدفها و ریز مواد درسی را برای تمام ترم مشخص کرد.

۲. هدفها و ریز مواد درسی هر جلسهٔ خود را در اول ساعت مشخص کرد.

۳. هدفهای درس هر جلسه با هدفهای کلی درس مناسبت داشتند.

۴. نحوهٔ ارزشیابی از کار دانشجویان و ملاکهای نمره گذاری خود را در آغاز ترم به دانشجویان اعلام کرد.

۵. سر وقت وارد کلاس می‌شد و تا آخر وقت در کلاس باقی می‌ماند.

۶. همهٔ وقت کلاس را صرف فعالیتهای آموزشی می‌کرد و اتلاف وقت نمی‌کرد.

۷. جزوهٔ درسی و سایر مواد آموزشی مربوط به یک موضوع درسی را، پیش از تدریس آن موضوع در اختیار دانشجویان قرار می‌داد.

۸. پیش از شروع درس تازه، با ذکر خلاصهٔ مطالب درس گذشته و ربط‌دادن آنها با مطالب جدید، دانشجویان را برای یادگیری درس تازه آماده می‌ساخت.

۹. در آغاز هر درس تازه، ابتدا کلیات آن درس را مطرح می‌کرد، سپس به بحث و معرفی جزئیات آن درس اقدام می‌کرد.

۱۰. بعد از خاتمهٔ درس به خلاصه کردن اصول کلی مطالب بحث‌شده و نتیجه گیری از آنها می‌پرداخت.

۱۱. تدریس او تنها به ارائهٔ دانش و اطلاعات ختم نمی‌شد و به مواردی چون بسط معنا، روابط، علل، عوامل و دلایل کشانده می‌شد.

۱۲. تدریس خود را با استفاده از روشهای مختلف آموزشی متنوع می‌ساخت.

۱۳. نکات مهم و اصول کلی مطالب درسی را بر روی تخته سیاه می‌نوشت.

۱۴. دانشجویان را به طرح سؤالاتی که به روشن‌شدن مطالب درسی کمک می‌کردند تشویق می‌کرد.

۱۵. با استفاده از مثالهای عینی و عملی گوناگون، مطالب انتزاعی و نظری را قابل فهم می‌ساخت.

۱۶. در سخنرانیها و توضیح و تشریح مطالب از کلمات و عبارات روشن و قابل فهم دانشجویان استفاده می‌کرد.

۱۷. در لفافه سخن نمی‌گفت و در تدریس او ابهامی وجود نداشت.

۱۸. هر گاه لازم می‌آمد اصطلاحات علمی و تخصصی به کار می‌برد آنها را دقیقاً تعریف می‌کرد.

۱۹. به سؤالات و اظهارات دانشجویان با دقت گوش می‌داد.

۲۰. به دانشجویان اجازه می‌داد تا مخالفت خود را با گفته‌های او ابراز دارند، و انتقادها و پیشنهادهای مفید آنان را می‌پذیرفت.

۲۱. درس خود را منظم و با روال منطقی ارائه می‌داد.

۲۲. من در هر جای کلاس که می‌نشستم می‌توانستم صدای او را به‌خوبی بشنوم.

۲۳. لحن صدای خود را مناسب با اوضاع و احوال کلاس تغییر می‌داد.

۲۴. در بارهٔ تکلیف و یا تمرینهایی که به دانشجویان می‌داد راهنماییهای لازم را می‌کرد.

۲۵. به نظر من از اینکه دانشجویان درس او را می‌فهمیدند یا نمی‌فهمیدند بااطلاع بود.

۲۶. سؤالاتی راکه دانشجویان در مورد مطالب درسی او مطرح می‌کردند به‌خوبی جواب می‌داد.

۲۷. به نظر من نسبت به این درس علاقه‌مند بود.

۲۸. تکالیف و اوراق امتحانی را با دقت تصحیح می‌کرد و به‌سرعت به دانشجویان باز می‌گرداند.

۲۹. تکالیف او به فهم بیشتر من از مطالب درسی کمک می‌کردند.

۳۰. برای رفع مشکلات درسی در خارج از کلاس به قدر لازم به او دسترسی داشتم.

۳۱. روش کار او مرا به موضوع درس علاقه‌مند کرد.

۳۲. تدریس او مرا وامی‌داشت تا قدمی فراتر از معلوماتی که او عرضه می‌کرد بردارم و خود به استنباط مطالب و استنتاج مطالب بپردازم.

۳۳. وقتی که جواب سؤالی را نمی‌دانست و یا اشتباهی می‌کرد آن را می‌پذیرفت.

۳۴. به نظر من در مورد مطالبی که درس می‌داد اطلاعات کافی داشت.

۳۵. علاوه بر کتاب درسی و (یا) جزوهٔ درسی، دانشجویان را از مراجع دیگر از جمله کتابها و مقالات و نشریات مطلع می‌ساخت و آنها را در مطالعهٔ این مراجع تشویق می‌کرد.

۳۶. توانست هدفهایی راکه در آغاز ترم تعیین کرده بود با موفقیت پیاده کند.

۳۷. ملاکهای مشخصی برای نمره گذاری تکالیف و ورقه‌های امتحانی به کار برد.

۳۸. امتحانات (امتحان) این درس کلیهٔ مطالب تدریس‌شده را به‌خوبی در برداشتند.

۳۹. امتحانات این درس بیشتر فهم و درک دانشجو و کمتر محفوظات را اندازه گیری می‌کردند.

۴۰. به تقاضای من برای روشن‌شدن اشتباهاتم در ورقهٔ امتحان توجه می‌کرد.

بخش دوم: ارزشیابی محتوای درس و مواد آموزشی

عبارتهای این بخش به منظور ارزشیابی موضوع درس و مواد آموزشی (کتاب و جزوهٔ درسی) تهیه شده‌اند. باز هم، با گذاشتن علامت (×) در جای خالی مربوط به حروف الف، ب، ج، یا د، در پاسخنامه، محتوا و مواد آموزشی این درس را ارزشیابی کنید.

عبارتها را با دقت بخوانید و توجه داشته باشید که عبارتهای این بخش صرفاً برای ارزشیابی محتوای دروس و مواد آموزشی تهیه شده‌اند. سعی کنید که نظر شما در بارهٔ معلم و روش کار او قضاوت شما را در بارهٔ محتوای درس و مواد آموزشی تحت تأثیر قرار ندهد. باز هم چنانچه به عبارتی برخوردید که با این درس مناسبت نداشت، آن عبارت را نادیده بگیرید.

۴۱. یادگیری مطالب این درس تجربهٔ آموزشی مفیدی برای من بود.

۴۲. مطالب این درس به نظر من جالب توجه بودند.

۴۳. از نظر من مطالب این درس کمتر تکراری و بیشتر تازه بودند.
۴۴. این درس در ردیف بهترین دروسی است که تاکنون گرفته‌ام.
۴۵. فکر می‌کنم یادگیری مطالب این درس در کار آتی من مفید واقع شوند.
۴۶. موضوع این درس طرز فکر مرا نسبت به مسائل شغلی آتی‌ام به نحو مطلوبی تغییر داد.
۴۷. موضوع این درس در من علاقهٔ بیشتری نسبت به حرفهٔ آتی‌ام ایجاد کرد.
۴۸. مایلم درس دیگری در این زمینه داشته باشم.
۴۹. مطالب این درس مفیدتر از مطالب دروس دیگری که من در این ترم دارم هستند.
۵۰. در مقایسه با درسهای دیگر، فکر می‌کنم مطالب کمتری از این درس را فراموش خواهم کرد.
۵۱. مطالب این درس کمتر به حافظه و بیشتر به تفکر و استدلال مربوط می‌شدند.
۵۲. فکر می‌کنم این درس مرا در یادگیری مطالب درسهای دیگر یاری خواهد کرد.
۵۳. مطالب این درس منطبق بر احتیاجات آیندهٔ شغلی من است.
۵۴. مقدار کار این درس درخور توانایی من بود و در این درس فشار زیادی بر من وارد نیامد.
۵۵. کتاب (کتابهای) درسی تعیین شده برای من قابل فهم بود.

۲. **علاقه و اشتیاق مدرس به موضوع درس**، شامل ۴ سؤال در بارهٔ میزان علاقه‌مندی معلم به موضوعی که تدریس می‌کند. نمونه‌ای از سؤالهای این بخش عبارت است از: "معلم شما نسبت به تدریس این درس از خود علاقه نشان می‌داد."

۳. **روشنی و سازمان‌بندی مطالب**، شامل ۴ سؤال در بارهٔ روشنی و سازمان‌بندی مطالبی که معلم تدریس می‌کند، مانند سؤال زیر: "توضیحات معلم روشن بودند."

۴. **تعامل گروهی**، شامل ۴ سؤال در بارهٔ مشارکت شاگردان در فعالیت کلاس، مانند این سؤال: "شاگردان به شرکت در بحث کلاسی تشویق می‌شدند."

۵. **در دسترس‌بودن معلم**، شامل ۴ سؤال در بارهٔ قابلیت دسترسی شاگردان به معلم از جمله سؤالات این بخش سؤال زیر است: "شاگردان برای رفع مشکلات خود به اندازهٔ کافی به معلم دسترسی داشتند."

۶. **جامعیت مطالب آموزش داده شده**، شامل ۴ سؤال در بارهٔ گستردگی و جامعیت مطالب تدریس شده، نظیر این سؤال: "در این درس معلم نظریه‌های متفاوت را با هم مقایسه می‌کرد."

۷. **امتحان و نمره گذاری**، شامل ۳ سؤال در بارهٔ کیفیت امتحانات و نحوهٔ نمره گذاری معلم، مانند سؤال زیر: "تأکید مواد و سؤالهای امتحانی با تأکید معلم بر موضوعهای درس همخوانی داشت."

۸. **مطالب خواندنی و تکالیف**، شامل تعدادی سؤال در بارهٔ کیفیت کتاب، جزوه، و تکالیف درسی، مانند این سؤال: "کتاب درسی شامل مطالب با ارزش بود."

۹. **نظر کلی**، شامل ۲ سؤال در بارهٔ کل درس و کیفیت آموزش معلم، مانند نمونهٔ زیر: "به طور کلی این معلم را در مقایسه با سایر معلمان این مؤسسه چگونه ارزیابی می‌کنید؟"

علاوه بر سؤالات بالا، مدرس می‌تواند از میان مجموعهٔ بانک سؤال که دانشگاه در اختیار او می‌گذارد تعداد ۲۵ سؤال به دلخواه خود برگزیند و به پرسشنامهٔ *ارزشیابی دانشجویان از کیفیت آموزشی* اضافه کند.

همچنین در بخش پایانی پرسشنامهٔ *ارزشیابی دانشجویان از کیفیت آموزشی* سؤالهایی در بارهٔ دانشجویان آمده است. این بخش میزان سهولت و دشواری و حجم مطالب و تکالیف درسی را از نظر دانشجو و نیز جنس (مرد یا زن بودن) دانشجو، نمرهٔ مورد انتظار دانشجو، میانگین نمرات درسی دانشجو، و دلایل انتخاب درس توسط او را شامل می‌شود.

در جلسهٔ اجرای پرسشنامه، همراه با فرم *ارزشیابی دانشجویان از کیفیت آموزشی*، به دانشجویان یک برگهٔ سفید داده می‌شود تا نظر خود را در بارهٔ کم و کیف تدریس استاد به صورت تشریحی نیز بیان نمایند. در بالای این برگه عنوان *اظهارنظرهای آزاد* ذکر شده و در زیر آن از دانشجو خواسته شده است تا نظرات خود را دربارهٔ ویژگیهای مهم درس و مدرس درس که در بالا بردن سطح یادگیری او مهم بوده، نقاط ضعف درس و مدرس که نیاز به اصلاح دارند، و هرگونه مطلب دیگری که دانشجو احساس می‌کند لازم است در بارهٔ آن توضیح دهد، به‌ویژه مطالبی که در پرسشنامهٔ *ارزشیابی دانشجویان از کیفیت آموزشی* ذکر نشده‌اند به تفصیل بنویسد.

یک ابزار جامع دیگر ارزشیابی معلم که آن هم بر نظرخواهی از دانشجویان استوار است سیستم *ارزشیابی معلم و درس*[1] نام دارد که در دانشگاه ایلینویز[2] درست شده و یک سیستم کامپیوتری است. این سیستم ارزشیابی شامل ۴۰۰ سؤال است که بر حسب محتوا دسته‌بندی شده‌اند. معلمان مختلف می‌توانند، از میان این سؤالها، هر تعدادی را که مناسب کار خود می‌دانند برگزینند و مورد استفاده قرار دهند (سنترا[3]، ۱۹۸۹).

1. Instructor and Course Evaluation System (ICES)
2. Illinois University
3. Centra

انواع سؤالهای پرسشنامه‌های نظرخواهی

پرسشنامه‌های نظرخواهی از شاگردان، برای اینکه اطلاعات جامعی از همهٔ ویژگیهای آموزشی معلم به دست بدهند، باید سؤالهای مختلفی را شامل گردند. سؤالهای پرسشنامه باید از چند دستهٔ مختلف تشکیل شوند و هر دسته سؤال جنبهٔ خاصی از ویژگیهای مورد نظر را بسنجند. یک دسته از سؤالها باید مختص درس باشند که جنبه‌های مختلف آن یعنی هدفها، سازمان‌بندی، مناسب‌بودن، فایده‌بخشی، دشواری، سهولت، و غیره را معرفی کنند. دستهٔ دیگر سؤالها باید ویژهٔ آموزش معلم باشند که مهارتهای آموزشی معلم، روشنی توضیحات او، تعامل با دانشجویان، و غیره را شامل شوند. سؤالهای مخصوص یادگیری باید نظر شاگرد را نسبت به قابل استفاده و قابل فهم‌بودن مطالب برای او بپوشانند. و سؤالهای مربوط به کتاب و سایر منابع یادگیری باید شامل سازمان درست مطالب، قابل‌فهم بودن، خاصیت کاربردی داشتن، و از این قبیل را شامل شوند. دقت کنید که در پرسشنامهٔ نمونهٔ شکل ۴-۲۲ و در پرسشنامهٔ ***ارزشیابی دانشجویان از کیفیت آموزشی*** (اس ای ای کیو) برای هریک از جنبه‌های بالا چندین سؤال در نظر گرفته شده است. علاوه بر این، در این پرسشنامه‌ها سؤالهای مربوط به آموزش معلم از سؤالهای مربوط به محتوای درس و مواد آموزشی مجزا شده‌اند. حُسن این کار آن است که پاسخ‌دهندگان را وامی‌دارد تا حتی‌المقدور ویژگیهای آموزشی معلم را جدا از ویژگیهای درس و مواد آموزشی (مانند کتاب و جزوهٔ درسی) مورد ارزشیابی قرار دهند. بسیاری از اوقات دیده می‌شود که معلمی در آموزش خود موفق است، اما به دلیل اینکه درس او مورد علاقهٔ شاگردانش نیست یا کتاب درسی کتاب خیلی خوبی نیست آموزش او نیز تحت‌الشعاع درس و کتاب درسی قرار می‌گیرد یا بالعکس.

شیوهٔ اجرا، نمره گذاری، و گزارش نتایج پرسشنامه‌های نظرخواهی

پرسشنامه‌های نظرخواهی که به وسیلهٔ معلمان تهیه می‌شوند و برای دروس خاص خودشان مورد استفاده قرار می‌گیرند، باید به وسیلهٔ خود معلمان در کلاس درس به اجرا درآیند و مستقیماً نتایجش را خود در اختیار بگیرند، و ضرورتی ندارد که کس دیگری از نتایج آنها با اطلاع شود، زیرا هدف اصلی تهیه و اجرای این پرسشنامه‌ها بهسازی کیفیت آموزش و دروس خاص معلمان است. اما پرسشنامه‌هایی که به منظور ارزشیابی از آموزش همهٔ معلمان یک مؤسسهٔ آموزشی مورد استفاده قرار می‌گیرند باید به توسط متخصصان ارزشیابی و با رعایت کلیهٔ اصول ضروری تهیه شوند و زیر نظر بالاترین مقام آموزشی مؤسسه و با نظارت ایشان اجرا شوند. این پرسشنامه‌ها باید به گونه‌ای یکنواخت و در زمانی معین به اجرا درآیند، ترجیحاً در یکی از جلسات رسمی کلاس (جلسات آخر ترم تحصیلی یا سال تحصیلی و پیش از برگزاری امتحانات پایان ترم). هرگز نباید پرسشنامه‌ها را به دست شاگردان داد تا با خود از کلاس بیرون ببرند و در خانه به آن پاسخ دهند.

گذشته از اینکه، با چنین کاری، شاگردان ممکن است با هم بر سر نمره‌دادن به معلم تبانی کنند و تحت تأثیر نظرات یکدیگر قرار گیرند، تعداد زیادی از شاگردان پرسشنامه‌ها را پر نمی‌کنند و بازنمی‌گردانند. نتایج نظرخواهی‌ها را نباید پیش از امتحان، نمره‌گذاری و تحویل نمرات توسط معلمان در اختیار آنان قرار داد. همچنین باید به شاگردان اطمینان خاطر داد که معلمانشان پیش از تحویل نمرات از نتایج ارزشیابیها مطلع نخواهند شد. با این حال، برای اینکه تأثیر بازخوردی ارزشیابیها به حداکثر برسد، لازم است هر چه زودتر معلمان را از نتایج ارزشیابی بااطلاع ساخت.

بهتر است نتایج ارزشیابی حاصل از پرسشنامه‌های نظرخواهی دانشجویی و دانش‌آموزی به گونه‌ای در اختیار معلمان قرار گیرند که حداکثر استفاده را به آنها برسانند. در این باره رعایت نکات زیر مفید است:

۱. نمره‌ای که معلم در هر یک از سؤالها گرفته است، به تفکیک درجات از جمله عالی، خوب، متوسط، و غیره محاسبه و در اختیار او گذاشته شود.

۲. نمره‌ای که معلم در هر یک از بخشهای پرسشنامه گرفته است از جمله در آموزش، منابع درسی، و غیره باز هم به تفکیک درجات تعیین شود و در اختیار معلم گذاشته شود.

۳. میانگین نمرات معلم برای کل سؤالها و مجموعهٔ سؤالهای هربخش تعیین و به وی گزارش شود.

۴. همراه با نمرات معلم، بالاترین نمره و کمترین نمرهٔ گروه معلمان شرکت‌کننده در ارزشیابی (بدون ذکر شخصی که بالاترین یا کمترین نمره را گرفته است) نیز مشخص و در اختیار معلم گذاشته شود.

۵. نمرات هم به صورت خام، با ذکر تعداد پاسخ‌دهندگان به هر سؤال، و هم به صورت درصدی تعیین و گزارش شوند.

۶. برای هر یک از کلاسهایی که معلم را مورد ارزشیابی قرار داده‌اند نتایج به شرح فوق گزارش شوند، و علاوه بر آن میانگین نمرات همهٔ کلاسهای یک معلم نیز مشخص شوند و در اختیار وی قرار گیرند.

محاسن و معایب روش استفاده از نظر دانش‌آموزان و دانشجویان

بحث‌انگیزترین روش ارزشیابی معلم روشی است که در آن از شاگردان معلم خواسته می‌شود تا او را مورد ارزشیابی قرار دهند. گروهی شدیداً از آن انتقاد کرده و گروهی دیگر قویاً از آن حمایت به عمل آورده‌اند. انتقادهای وارده به این روش ارزشیابی معلم را می‌توان در چند مورد زیر خلاصه کرد:

۱. نظر شاگردان نسبت به کار معلم به درستی معرف شایستگی معلم نیست، بلکه صرفاً منعکس‌کنندهٔ رضایت خاطر یا عدم رضایت خاطر شاگردان معلم از اوست؛ شاگردانی که نمرهٔ بهتری از یک معلم می‌گیرند از شاگردانی که نمرهٔ خوبی از او دریافت نمی‌کنند، در ارزشیابیهای

خود از او، نمرهٔ بیشتری به وی می‌دهند.

۲. جنسیت معلم و شاگرد در نحوهٔ ارزشیابی شاگردان از معلمان مؤثر است. مثلاً بندینگ[1] (۱۹۵۲) دریافت که دانشجویان مؤنث از دانشجویان مذکر به معلمان مرد نمرهٔ کمتری می‌دهند. همچنین واکر[2] (۱۹۶۹) به این نتیجه دست‌یافت که دانشجویان دختر معلمان زن را از معلمان مرد مثبت‌تر ارزشیابی می‌کنند.

۳. تعداد شاگردان کلاس، سطح دشواری درس، نوع درس، و سطح تحصیلات بر نحوهٔ ارزشیابی شاگردان از معلمان تأثیر می‌گذارد. شاگردان کلاسهای شلوغ از شاگردان کلاسهای کم‌جمعیت معمولاً به معلمان خود نمرهٔ کمتری می‌دهند. شاگردان دروس انتخابی از شاگردان دروس اجباری به معلمان خود نمرهٔ بهتری می‌دهند. شاگردان سطوح بالاتر، مثلاً دانشجویان دوره‌های فوق‌لیسانس (کارشناسی ارشد) و دکتری، از شاگردان سطوح پایین‌تر، مثلاً دانشجویان دورهٔ لیسانس، به معلمان خود نمرهٔ بهتری می‌دهند.

۴. نظر دانشجویان و دانش‌آموزان نسبت به معلمان بیشتر تحت تأثیر رفتار دوستانهٔ معلم، شیوه‌های سرگرم‌کننده و به اصطلاح اغواگرانهٔ او، و ویژگیهای کلی و شهرت و اعتبار اوست [برای نمونه پژوهشهایی در این باره نگاه کنید به سیف (۱۳۷۰) و گرجی (۱۳۷۳).]. گرجی، در پژوهش خود که در این باره انجام داد، بین نمراتی که دانشجویان دانشگاه در یک فرم محبوبیت‌سنج استادی به استادان خود دادند و نمراتی که به عنوان ارزشیابی از کیفیت آموزش در فرمهای نظرخواهی برای آنان در نظر گرفتند ضریب همبستگی ۰/۷۵+ را به دست آورد. این اثرهای هاله‌ای[3] استفاده از فرمهای نظرسنجی دانشجویی برای ارزشیابی از آموزش معلمان را با کمبود روایی مواجه می‌سازد. اما اگر هدف استفاده کننده از آنها سنجش احساسات دانشجویان و دانش‌آموزان نسبت به معلمان باشد، بی‌تردید این فرمها از روایی کافی برخوردارند. هریس[4] (۱۹۸۶) در این باره می‌گوید دخالت شاگردان در سنجش معلم باید به توصیف آنان از زندگی در کلاس درس محدود باشد و به داوری دربارهٔ ارزش کار معلم نسبت داده نشود.

مورش[5] و وایلدر[6] (۱۹۵۴) در یک بررسی از پژوهشهای اولیه در بارهٔ روایی این‌گونه فرمهای نظرسنجی شواهدی یافتند حاکی از اینکه اگر معلم کوششهای آموزشی خود را متوجه شاگردان قوی کلاس بکند مورد تأیید این شاگردان قرار خواهد گرفت و در نتیجه ضریب همبستگی بین نمرات درسی شاگردان و نمراتی که این دانش‌آموزان در ارزشیابی معلم به او می‌دهند مثبت خواهد بود؛ اما اگر معلم آموزش خود را بر شاگردان ضعیف کلاس متمرکز کند، در آن صورت دیگر مورد تأیید شاگردان قوی کلاس نخواهد بود؛ در نتیجه ضریب همبستگی بین نمرات درسی شاگردان و نمراتی که شاگردان در فرم نظرسنجی به معلم می‌دهند منفی خواهد بود.

1. Bending
2. Walker
3. halo effects
4. Harris
5. Morsh
6. Wilder

از سوی دیگر، طرفداران روش ارزشیابی معلم توسط شاگردان نیز مزایایی برای آن ذکر کرده‌اند. این امتیازها را می‌توان به نحو زیر خلاصه کرد:

۱. شاگردان منبع اصلی اطلاعات در بارهٔ توفیق معلم در رسیدن به هدفهای آموزشی درس و نحوهٔ برخورد او با شاگردان و روشهای ارائهٔ درس و کلاسداری هستند. بنابراین، هیچ‌کسی به‌خوبی شاگردان یک معلم از کمّ و کیف کار او بااطلاع نیست. لذا بهترین داوران اثربخشی کار معلم همانا شاگردان او هستند. بنا به گفتهٔ استرانگ (۲۰۰۶)، دانش‌آموزان و دانشجویان اصلی‌ترین مصرف‌کنندگان خدمتهای معلم‌اند، پس باید به نظر آنان در این باره اهمیت داده شود.

۲. استفاده از نظرهای شاگردان یک معلم برای داوری در بارهٔ کار او از هر روش دیگر ارزشیابی امکان‌پذیرتر است، زیرا برخلاف روشهایی چون مشاهدهٔ کلاسی توسط مدیران یا سایر افراد خارج از کلاس، در روش ارزشیابی معلم توسط شاگردان جریان طبیعی درس و بحث معلم مختل نمی‌شود.

۳. نظر شاگردان نسبت به معلمانشان وسیلهٔ مناسبی برای ایجاد ارتباط بین شاگردان و معلم فراهم می‌آورد و بهترین راه تأثیرگذاری متقابل بر هر دو طرف ذینفع است. در نتیجه، با استفاده از نظر شاگردان، هم می‌توان روند آموزشی معلمان را اصلاح کرد و آنان را از سطح توقع و خواسته‌های شاگردانشان باخبر گردانید و هم شاگردان را نسبت به کار و کوشش معلمان حساس و علاقه‌مند نمود.

به‌رغم اختلاف نظرهای فراوان در بارهٔ استفاده از نظر دانشجویان و دانش‌آموزان برای ارزشیابی از کار معلمان، اگر پرسشنامه‌های نظرسنجی مطابق اصول صحیح تهیه و اجرا گردند و نتایج آنها برای اصلاح و بهسازی آموزش معلمان به کار بروند و آنجا که قرار است تصمیمهایی دربارهٔ معلم گرفته شود، این نتایج با نتایج حاصل از روشهای دیگر ارزشیابی تکمیل شوند، استفاده از روش ارزشیابی معلم توسط دانشجویان و دانش‌آموزان روشی مفید و قابل استفاده‌ای است. بنا به گفتهٔ استرانگ و استراندر (۲۰۰۶) "نتایج ارزشیابی شاگردان از معلمان پنجرهٔ دیگری به روی متخصصان ارزشیابی برای تعیین اثربخشی معلم فراهم می‌آورند" (ص ۱۴۱). نکتهٔ مهم دیگری که در رابطه با ارزشیابی معلم توسط دانش‌آموزان و دانشجویان مورد تأکید قرار گرفته این است که معلمان خود باید تصمیم بگیرند که نتایج ارزشیابی شاگردان از کار آنان در اختیار مدیران و سرپرستان قرار بگیرد یا نه (لارسون[1]، ۱۹۸۴).

استفاده از نظر والدین دانش‌آموزان و دانشجویان

روش استفاده از نظر والدین دانش‌آموزان و دانشجویان می‌تواند مکمل روش استفاده از نظر دانش‌آموزان و دانشجویان باشد. استفاده از نظر والدین دانش‌آموزان به عنوان منبعی برای

1. Larson

ارزشیابی از کار معلمان به تازگی در کشورهای مغرب زمین مورد استفاده قرار گرفته است. هوگان (۲۰۰۷) دلیل آن را این‌گونه بیان می‌کند که، بعد از دانش‌آموزان و دانشجویان، پدران و مادران آنان ذینفع‌ترین کسان در آموزش و پرورش فرزندانشان هستند. استرانگ و اُستراندر (۲۰۰۶) نیز گفته‌اند:

> کوششهای متعدد وابسته به اصلاحات آموزشی در دو دههٔ اخیر خواستار افزایش درگیری والدینِ دانش‌آموزان در آموزش و پرورش آنان و دادن اختیار بیشتر به والدین برای تصمیم‌گیری دربارهٔ مسائل مربوط به آموزش و پرورش بوده است. شواهد کافی وجود دارند که نشان می‌دهند رابطهٔ خوب بین معلم و والدین بر موفقیت مدارس تأثیر مثبت دارد. یکی از راههای درگیر کردن والدین در مسائل آموزش و پرورش استفاده از نظرات آنان برای ارزشیابی از کار معلمان است. در این باره، دیدگاه و بینش والدین منشأ اطلاعات ارزشمندی است که از هیچ راه دیگری قابل اکتساب نیست. (ص ۱۳۰)

پرسشنامه‌های نظرخواهی از والدین

پرسشنامه‌های نظرخواهی از والدین بسیار شبیه به پرسشنامه‌های نظرخواهی از دانش‌آموزان و دانشجویان‌اند. با این حال، این پرسشنامه‌های نظرخواهی از جهاتی با پرسشنامه‌های دانش‌آموزی و دانشجویی فرق دارند. مهم‌ترین تفاوت در نوع سؤالهای این دو نوع پرسشنامه است. در پرسشنامه‌های دانش‌آموزی و دانشجویی، چنان که دیدیم، سؤالها عمدتاً به نحوهٔ تدریس و کلاسداری معلمان اختصاص دارند، اما سؤالهای پرسشنامه‌های نظرخواهی از والدین غالباً موضوعاتی را در رابطه با کم و کیف تکالیفی که معلمان به دانش‌آموزان می‌دهند تا در خانه انجام دهند، چگونگی بازخوردی که والدین در رابطه با فرزندانشان از معلم و مدرسه دریافت می‌کنند، و مواردی از این قبیل را شامل می‌شوند. به عنوان نمونه‌ای از پرسشنامه‌های نظرخواهی از والدین نگاه کنید به شکل ۵ـ۲۲.

اجرا و نمره‌گذاری پرسشنامه‌های نظرخواهی از والدین

به خلاف پرسشنامه‌های نظرخواهی دانش‌آموزی و دانشجویی که معمولاً در کلاس درس اجرا می‌شوند، پرسشنامه‌های نظرخواهی از والدین باید برای پدران و مادران دانش‌آموزان فرستاده شوند و پاسخ آنها نیز از همان طریق دریافت می‌گردند. از این رو، افت اطلاعات مربوط به روش نظرخواهی از والدین بسیار زیاد است.

در ارتباط با چگونگی تصحیح و نمره‌گذاری پرسشنامه‌های نظرخواهی از والدین، همان روش مورد استفاده با پرسشنامه‌های نظرخواهی از دانش‌آموزان و دانشجویان به‌کار بسته می‌شود.

شکل ۵ـ۲۲ یک پرسشنامهٔ نظرخواهی از والدین، ویژهٔ مدارس دولتی شهر ویرجینیابیچ کشور آمریکا (نقل شده به وسیلهٔ استرانگ و اُستراندر، ۲۰۰۶، صص ۱۳۴ـ۱۳۲)

نام معلم: پایهٔ تحصیلی: نام مدرسه:

معلمی که نام آن در بالای این فرم نوشته شده علاقه‌مند به دریافت نظرات شماست. لطفاً به سؤالهای زیر صرفاً در رابطه با همین معلم پاسخ دهید. ابتدا سؤالهای زیر را جواب دهید:

از آغاز سال تحصیلی تاکنون، شما...	بله	نه
• در یکی از جلسات عمومی مدرسه، از جمله جلسات والدین و معلمان شرکت کرده‌اید؟	☐	☐
• در یکی از رویدادهای کلاس یا مدرسه، از جمله مسابقهٔ ورزشی یا نمایشگاه علمی شرکت کرده‌اید؟	☐	☐
• داوطلب کمک به فعالیت‌های مدرسه یا عضو کمیته‌های مدرسه شده‌اید؟	☐	☐

راهنمایی

لطفاً ابتدا به سه سؤال بالا جواب دهید. بعد به دقت هریک از جمله‌های زیر را که نشان‌دهندهٔ تصورات شما دربارهٔ معلم فرزندتان است مطالعه کنید و نظر خود را به صورت **کاملاً موافق، موافق، مخالف، یا کاملاً مخالف** نسبت به آن جمله در جای تعیین شده در مقابل آن علامت بزنید. اگر در رابطه با هریک از سؤالها نظری ندارید جای مخصوص بی‌نظر را علامت بزنید. لطفاً به همهٔ سؤالها جواب دهید. اگر نظر دیگری دارید از قسمت سفید پشت این فرم استفاده کنید. در این فرم نام خود را ننویسید. پس از تکمیل، فرم را در پاکت مخصوص آن قرار داده و به دفتر پژوهش آموزش و پرورش ارسال کنید. لطفاً در طول یک هفته به ما جواب دهید.

از وقتی که برای این کار صرف می‌کنید سپاسگزاریم

معلم فرزند من...	کاملاً مخالف	مخالف	بی‌نظر	موافق	کاملاً موافق
۱. با من ارتباط برقرار می‌کند.	☐	☐	☐	☐	☐
۲. دوست دارد با او ارتباط برقرار کنم.	☐	☐	☐	☐	☐
۳. هنگام ملاقات با او اطلاعات مفیدی در اختیار من می‌گذارد.	☐	☐	☐	☐	☐
۴. از پیشرفت فرزندم در مدرسه مرا آگاه می‌سازد.	☐	☐	☐	☐	☐
۵. آدم خوش برخوردی است.	☐	☐	☐	☐	☐
۶. به حرفهای من گوش می‌دهد.	☐	☐	☐	☐	☐
۷. احترام مرا به خود جلب می‌کند.	☐	☐	☐	☐	☐
۸. از آنچه در کلاس اتفاق می‌افتد مرا آگاه می‌سازد.	☐	☐	☐	☐	☐
۹. روشهای نمره گذاری‌اش را به من اطلاع می‌دهد.	☐	☐	☐	☐	☐
۱۰. دانش‌آموزان را با مقررات کلاس آشنا می‌کند.	☐	☐	☐	☐	☐
۱۱. از روشهای انضباطی معتدل استفاده می‌کند.	☐	☐	☐	☐	☐

۱۲. فرزندم را به فعالیت وامی‌دارد.	□	□	□	□	□
۱۳. به افزایش توانایی فرزندم کمک می‌کند.	□	□	□	□	□
۱۴. پیشرفت دانش‌آموزان را می‌فهمد.	□	□	□	□	□
۱۵. رفتارهای مثبت را تشخیص می‌دهد.	□	□	□	□	□
۱۶. مورد احترام فرزندم است.	□	□	□	□	□
۱۷. به رشد زمینه‌های ضعیف فرزندم کمک می‌کند.	□	□	□	□	□
۱۸. مشارکت کلاسی را تشویق می‌کند.	□	□	□	□	□
۱۹. از روشهای متنوع آموزشی استفاده می‌کند.	□	□	□	□	□
۲۰. به رفتارهای کلی فرزندم در کلاس نمره می‌دهد.	□	□	□	□	□
۲۱. تدریس خود را به فعالیتهای زندگی ربط می‌دهد.	□	□	□	□	□
۲۲. تکالیفی می‌دهد که به یادگیری کلاسی کمک می‌کنند.	□	□	□	□	□
۲۳. نحوهٔ انجام تکالیف خانه را برای دانش‌آموزان به خوبی توضیح می‌دهد.	□	□	□	□	□
۲۴. وقتی لازم باشد کمکهای اضافی در اختیار دانش‌آموزان می‌گذارد.	□	□	□	□	□
۲۵. محیط یادگیری تشویق‌کننده‌ای ایجاد می‌کند.	□	□	□	□	□
۲۶. محیط یادگیری سالمی ایجاد می‌کند.	□	□	□	□	□
۲۷. محیط یادگیری مشوق یادگیری را ایجاد می‌کند.	□	□	□	□	□

در ضمن، نمرات این پرسشنامه‌ها نیز باید مستقیماً در اختیار معلمان قرار گیرند. از نتایج این نوع ارزشیابی باید به صورت یک منبع بازخوردی و به عنوان نوعی ارزشیابی تکوینی برای رشد شخصی و حرفه‌ای معلمان استفاده شود.

پژوهشهای انجام شده دربارهٔ تأثیر استفاده از نظر والدین برای بهبود کار معلمان نتایج مثبتی به دست داده‌اند. از جمله در پژوهشی که در آن پرسشنامهٔ شکل ۵ـ۲۲ با معلمان مدارس عمومی یکی از نواحی آموزش و پرورش کشور آمریکا به مدت سه سال اجرا شد نتایج زیر گزارش شدند:

> هم برای فرد فرد معلمان شرکت‌کننده در پژوهش و هم برای ناحیهٔ آموزش و پرورش نتایج مثبت بودند. بیشتر از ۷۰ درصد همهٔ معلمان ناحیه دست‌کم یک بار خواستار شرکت در پژوهش و دریافت نظرات والدین دانش‌آموزانشان شدند، و خیلی‌ها هم هر سه سال داوطلب بودند. هر سال میانگین نمره‌ها افزایش یافت که نشان می‌دهد معلمان اطلاعات دریافتی از والدین را برای بهبود کارشان مورد استفاده قرار دادند. بیشترین تأثیر مثبت در زمینهٔ بهبود ارتباط بین معلمان و والدین گزارش شد. (استرانگ و اُستراندر، ۲۰۰۶، ص ۱۳۴)

محاسن و معایب روش استفاده از نظر والدین

روش استفاده از نظر والدین دانش‌آموزان و دانشجویان به عنوان یک منبع بازخورد اطلاعاتی برای بهبود کیفیت کار معلمان، به‌ویژه بالا بردن سطح ارتباط بین معلم و والدین، روش بسیار مفیدی است. هوگان (۲۰۰۷)، در دفاع از این روش، گفته است "پس از دانش‌آموزان، والدین آنها ذینفع‌ترین افراد در

ارتباط با نهاد آموزش و پرورش‌اند" (ص ۳۷۹). افزون بر این، چنان که قبلاً اشاره کردیم، پژوهشها نشان داده‌اند که استفاده از بازخورد حاصل از نظر والدین می‌تواند در بهبود کار معلم مؤثر باشد.

به رغم امتیازهای مختلف، استفاده از نظر والدین به عنوان یک روش ارزشیابی معلم دارای همان معایبی است که در رابطه با روش استفاده از نظر دانش‌آموزان و دانشجویان گفتیم. عیب مهم دیگر این روش اُفت اطلاعاتی آن است. برای اجرای این روش، باید پرسشنامه‌ها و فرمهای ارزشیابی یا توسط فرزندان یا پست برای والدین ارسال شوند و آنها نیز به همان طریق فرمهای پر شده را بازگردانند. هیچ تضمینی وجود ندارد که والدین همهٔ فرمها یا حتی اکثریت آنها را جواب دهند و بازگردانند. هوگان (۲۰۰۷) در این باره می‌گوید:

> ممکن است به اندازهٔ کافی جواب دریافت نکنید. اگر درصد کوچکی از والدین فرمهای ارزشیابی را پر کنند و برای شما بازگردانند با یک مشکل جدّی روبه‌رو هستید. شما هرگز مطمئن نخواهید بود کسانی که فرمها را بازمی‌فرستند نمایندهٔ واقعی همهٔ والدین هستند. مشکل اینجاست که کسانی که جواب نداده‌اند ممکن است نظری کاملاً متفاوت با نظر کسانی که فرمها را فرستاده‌اند داشته باشند. بنابراین، لازم است نسبت به نتایج محتاط باشید. (ص ۳۷۹)

به رغم اختلاف‌نظرهای موجود، روش استفاده از نظرات والدین به عنوان یک روش ارزشیابی معلم روش مفیدی است. اما شرط مفید بودن این روش آن است که پرسشنامه‌های مورد استفاده مطابق اصول صحیح تهیه شوند و نتایج آنها عمدتاً برای اصلاح و بهسازی آموزش معلمان به کار روند، و آنجا که قرار است تصمیمهایی دربارهٔ معلم گرفته شود، لازم است این نتایج با نتایج حاصل از روشهای دیگر ارزشیابی تکمیل شوند.

استفاده از یادگیری دانش‌آموزان و دانشجویان

میزان یادگیری یا پیشرفت تحصیلی دانش‌آموزان و دانشجویان یکی از شاخصهای مهم شایستگی معلم یا کیفیت کار اوست. درواقع، بسیاری متخصصانِ آموزشی و ارزشیابی آن را معتبرترین ملاک می‌دانند، زیرا بر این باورند که وظیفهٔ اصلی معلم کمک به دانش‌آموزان و دانشجویان در کسب یادگیری و پیشرفت تحصیلی است. از جمله، سنترا (۱۹۸۹) بر این باور است که "استفاده از یادگیری شاگردان عادلانه‌ترین و عینی‌ترین روش سنجش اثربخشی معلم است" (ص ۱۰۴). همچنین تاکر[1] و استرانگ (۲۰۰۶) می‌گویند "اطلاعات مربوط به یادگیری دانش‌آموزان به هر شکلی که باشد برای ارزشیابی درست از اثربخشی کار معلم ضروری است" (ص ۱۵۶).

معمولاً هدف عمدهٔ اندازه‌گیری پیشرفت تحصیلی یادگیرندگان نمره‌دادن به آنان است. برای این منظور، اصول ضروری حاکم بر موقعیتهای امتحانی باید رعایت شوند. از جمله اینکه

1. Tucker

دانش‌آموزان و دانشجویان باید بدانند که از چه چیزی امتحان خواهند داد، سؤالهای امتحانی باید هدفهای از پیش تعیین شدهٔ درس را بسنجند، و همهٔ دانش‌آموزان یک کلاس باید به سؤالهای واحدی جواب دهند. مورد آخر از این بابت ضروری است که بتوان نمرات همهٔ شاگردان را با هم مقایسه کرد. مثلاً اگر به تعدادی دانش‌آموز امتحان آسانی داده شود و از گروهی دیگر خواسته شود تا امتحان مشکل‌تری را جواب دهند آن‌گاه معلم نخواهد توانست مطمئن شود که نمرهٔ بهتر دانش‌آموزان ناشی از پیشرفت بیشتر آنان است یا سهولت امتحان. در این صورت دانش‌آموزانی که امتحان مشکل را داده‌اند در شرایط ناعادلانه‌ای قرار داده شده‌اند.

از سوی دیگر، در امتحاناتی که برای ارزشیابی از آموزش و اندازه‌گیری اثربخشی معلم اجرا می‌شوند، رعایت این شرایط خیلی ضروری نیست. در اینجا رعایت حال شاگردان را نباید زیاد مورد توجه قرار داد، زیرا قضاوتها و تصمیمات مبتنی بر نتایج امتحانات بر فرد فرد دانش‌آموزان تأثیر نخواهد داشت و صرفاً برای ارزشیابی از کار معلم به‌کار خواهند رفت. بنابراین، در استفاده از مقدار یادگیری شاگردان برای ارزشیابی از کار معلم، علاوه بر روشهای معمول سنجش پیشرفت تحصیلی شاگردان، از روشهای دیگری نیز می‌توان سود برد. مثلاً شواهد مربوط به یادگیری دانش‌آموزان را، علاوه بر عملکرد آنان در آزمونها، هم می‌توان از تکالیفی که در خانه انجام می‌دهند به دست آورد و هم از تمرینهایی که در کلاس درس انجام می‌دهند. در ضمن، انواع و اقسام فعالیتهای یادگیری شاگردان که نشان‌دهندهٔ یادگیری آنان است قابل استفاده هستند، مانند پروژه‌های تحقیقی، کارهای آزمایشگاهی و غیره.

علاوه بر موارد فوق، راههای دیگری را می‌توان به کار بست که در آن میزان پیشرفت تحصیلی شاگردان را برای سنجش اثربخشی معلمان بتوان مورد استفاده قرار داد. مثلاً می‌توان از معلم خواست تا به عنوان نمونهٔ کار خود موضوع مشخصی را با هدفهای آموزشی محدودی به گروهی دانش‌آموز تدریس کند و سپس میزان یادگیری شاگردان را با وسایل مختلف اندازه بگیرد. در این صورت اگر دانش‌آموزان به هدفهای از پیش تعیین شده رسیده باشند روش معلم موفقیت‌آمیز تلقی خواهد شد، در غیر این صورت روش او ضعیف است و به کمک و اصلاح نیازمند خواهد بود.

یکی از نمونه‌های مهم استفاده از یادگیری دانش‌آموزان برای ارزشیابی از میزان موفقیت معلمان و سایر مسئولان مدرسه در به ثمر رسانیدن هدفهای آموزشی یا برآورده ساختن استانداردهای تحصیلی برنامه‌های پاسخگویی یا مسئولیت‌پذیری آموزشی[1] آمریکاست. "پاسخگویی به یک نهضت آموزشی گفته می‌شود که از مدارس و مسئولان می‌خواهد تا موفقیت برنامه‌های آموزشی را، عمدتاً به کمک آزمونها، نشان بدهند" (هوگان ۲۰۰۷، ص ۴۰۴).

1. educational accountability

نهضت اصلاحات آموزشی اواخر قرن بیستم میلادی در کشور ایالات متحد امریکا که برنامه‌های پاسخگویی را به دنبال آورد بر بالا بردن سطح استانداردهای تحصیلی تأکید ورزید. در این زمان قانون‌گذاران در غالب ایالتهای آمریکا مجموعه قانونهایی را به تصویب رسانیدند که استفاده از آزمونهای به اصطلاح خطیر[1] را اجباری می‌کرد. "از آنجا که دولت و مردم خواستار پاسخگویی هرچه بیشتر مدارس در قبال تربیت کودکان ملت شدند، آزمونهای اجباری نقش بیشتری را به عهده گرفتند" (سانتروک، ۲۰۰۴، ص ۴۹۴).

آزمونهای خطیر یا آزمودن خطیر به عنوان یک وسیلهٔ ارزشیابی آموزشی که در آن ملاک داوری دربارهٔ شایستگی معلمان و کُل فعالیتهای آموزشی مدرسه میزان یادگیری دانش‌آموزان است پیامدهای مهمی به دنبال دارد. براساس نتایج این آزمونها کارکنان یک آموزشگاه مورد تشویق و تنبیه قرار می‌گیرند. به عنوان نمونه،

> مدارسی که دانش‌آموزان آنها در سنجشهای اجباری ایالتی عملکرد چشم‌گیری به دست می‌آورند به گونه‌های مختلف مورد تشویق قرار می‌گیرند و حتی پاداشهای پولی دریافت می‌کنند. در بعضی مواقع سهمی از پولها به معلمان تعلق می‌گیرد. مدارسی که عملکرد دانش‌آموزانشان در آزمونهای اجباری در سطح پایینی قرار دارند تنبیه و توبیخ می‌شوند. از جمله تنبیه‌هایی که براساس نظام پاسخگویی مبتنی بر سنجش یادگیری بر مدارس اعمال می‌شود این است که گروهی از افراد بیرون از مدرسه را برای نظارت بر مدرسه می‌گمارند، یا معلمان مدرسه را به مدرسه‌های دیگر منتقل می‌کنند، و یا اینکه مدیر تازه‌ای به جای مدیر قبلی برمی‌گزینند. (لین و گرانلاند، ۲۰۰۰، ص ۷)[2]

استفاده از آزمونهای عملکرد آموزشی

یکی از فنون معروف روش استفاده از مقدار یادگیری دانش‌آموزان برای ارزشیابی معلمان **آزمونهای عملکرد آموزشی**[3] نام دارد (یوفام، ۱۹۷۱). در این روش، ابتدا معلمان در معرض یک **درسِ خُرد** قرار می‌گیرند. درسِ خُرد یک مطلب از پیش تهیه شده است در بارهٔ موضوع خاصی که از معلمان خواسته می‌شود تا آن را آموزش بدهند. موضوع انتخابی برای درس خُرد باید چیزی باشد که معلمان از آن بی‌اطلاع‌اند، چنان‌که تسلط‌یابی معلمان بر آن موضوع خود بخشی از توفیق آموزشی آنان به حساب می‌آید. مطالب درس خُرد شامل بیان هدفهای آموزشی، مواردی از نوع سؤالهایی که شاگردان قرار است به آن جواب دهند، و اطلاعات کافی در بارهٔ موضوع تدریس است به گونه‌ای که خود معلمان قادر به جواب دادن به سؤالها باشند. برای جلوگیری از اینکه تدریس

---

1. high-states

۲. برای کسب اطلاعات بیشتر در این باره نگاه کنید به اثر دیگر مؤلف، *سنجش فرآیند و فرآوردهٔ یادگیری: روشهای قدیم و جدید* (سیف، ۱۳۸۷).

3. teaching performance tests

معلمان بر سؤالها متمرکز نشود، بهتر است اصل سؤالها در اختیار معلمان قرار نگیرد، بلکه سؤالهای مشابه با آنها در اختیارشان گذاشته شود.

پس از این مقدمات، به معلمان وقت کافی داده می‌شود تا شیوه‌های آموزشی خود را تدارک ببینند. به معلمان گفته نمی‌شود چگونه آموزش بدهند، بلکه کلیهٔ فنون آموزش، از قبیل تهیهٔ طرح درس، انتخاب روش آموزشی مناسب، و غیره بر عهدهٔ خود معلمان است. به آنها گفته می‌شود هر کاری که خود می‌خواهند انجام دهند، اما در هر حال هدفشان باید بالابردن سطح یادگیری دانش‌آموزان باشد.

بعد معلمان به گروهی از دانش‌آموزان یا دانشجویان برای مدت معینی آموزش می‌دهند. و سرانجام در پایان درس دانش‌آموزان با سؤالهایی شبیه به سؤالهایی که در ابتدا به معلمان نشان داده شده بود امتحان می‌شوند. میزان پیشرفت یادگیرندگان در رسیدن به هدفهای درس به عنوان شاخصی برای اثربخشی کار معلمان مورد استفاده قرار می‌گیرد.

محاسن و معایب روش استفاده از یادگیری دانش‌آموزان و دانشجویان

مانند روشهای دیگر ارزشیابی معلم، روش استفاده از یادگیری دانش‌آموزان و دانشجویان نیز بی‌چون و چرا مورد پذیرش قرار نگرفته است، بلکه گروهی از آن انتقادکرده و گروهی دیگر آن را قویاً مورد تأیید قرار داده‌اند. ما در اینجا مختصراً به این بحثهای موافق و مخالف اشاره می‌کنیم. از جمله ایرادهایی که به این روش وارد شده به قرار زیرند:

۱. بعضی‌ها معتقدند که بازده‌های یادگیری شاگردان ملاک درستی برای اثربخشی آموزش معلم نیستند، زیرا معلمان، به‌ویژه معلمان دانشگاه را نباید مسئول اصلی یادگیری یا عدم یادگیری دانشجویان دانست. بلکه عوامل دیگری در این کار دخیل‌اند که خود شاگردان عمده‌ترین آنها هستند. این منتقدان می‌گویند، صرف‌نظر از مهارت و جدیت معلم، اگر سایر عوامل به یادگیری شاگردان کمک نکنند یا مخالف آن عمل نمایند، به‌ویژه اگر خود دانش‌آموزان و دانشجویان حاضر به یادگیری مطالب درسی نباشند، از معلمان کاری ساخته نیست و گناه یادنگرفتن شاگردان را نباید به گردن معلمان انداخت.

۲. منتقدان دیگر می‌گویند یادگیری مطالب درسی که با آزمونها سنجیده می‌شوند تنها دستاوردهای آموزشی معلمان نیستند، لذا نباید خیلی به آنها اعتنا کرد. کار مهم دیگری که معلمان انجام می‌دهند کمک به رشد شخصیت یادگیرندگان است که به هیچ‌وجه با وسایل متداول سنجش قابل ارزیابی نیست.

۳. انتقاد دیگر یک انتقاد فنی است. میلمن[1] (۱۹۸۱) می‌گوید "اطلاعات جمع‌آوری شده

1. Millman

نشان می‌دهند که آزمونهای عملکرد آموزشی ناپایا[1] هستند.... زمانی که قرار است از نتایج این آزمونها برای مقاصد تصمیم‌گیری در بارهٔ معلمان استفاده شود، مسئلهٔ ناپایایی آزمونها امری جدی است، اما اگر از این نتایج برای مقاصد اصلاحی استفاده شود نگرانی چندانی وجود ندارد" (ص ۱۵۳).

۴. انتقاد دیگر بر این نکته تأکید می‌کند که استفاده از مقدار یادگیری دانش‌آموزان به عنوان تنها ملاک توفیق معلمان برای معلمانی که دارای شاگردان غیرکوشا، کندآموز، و دارای مشکلات دیگر هستند روشی ناعادلانه است.

در مقابل این انتقادها، مدافعان این روش می‌گویند از آنجاکه هدف اصلی آموزش معلمان و سایر کوششهای مدارس و مسئولان ایجاد یادگیری در دانش‌آموزان است، مراجعه به یادگیری شاگردان بهترین ملاک ارزشیابی معلمان است. بنابراین، به جای ایرادگیری از آن، باید سعی کرد که ملاکهای بهتری برای سنجش یادگیری دانش‌آموزان و دانشجویان فراهم آورد و این روش را کامل‌تر کرد. تاکر و استرانگ (۲۰۰۶) می‌گویند، به رغم بحث‌انگیز بودن روش استفاده از یادگیری دانش‌آموزان و دانشجویان برای ارزشیابی از کار معلم، بسیاری از نظامهای آموزشی مدارس (در کشور آمریکا) سالهاست که از این روش سود می‌برند. "در سال ۱۹۸۸ پژوهشی که در این زمینه انجام گرفت نشان داد که ۶۷ درصد مدارس بر نوعی شاخص یادگیری شاگردان تکیه داشتند.... اندیشهٔ تصمیم‌گیری دربارهٔ معلمان براساس میزان یادگیری دانش‌آموزان مورد حمایت مردم و نمایندگان منتخب آنان است" (ص ۱۵۶). بنابراین، استفاده از میزان یادگیری و عملکرد تحصیلی دانش‌آموزان برای داوری دربارهٔ اثربخشی کار معلمی امری اجتناب‌ناپذیر است، اما آنچه که در این باره مورد تأکید قرار دارد این است که اطلاعات موردنیاز باید با روشهای دقیق، منصفانه، و روا گردآوری شود.

خودسنجی یا ارزشیابی از خود

روش **خودسنجی**[2] یا **ارزشیابی از خود**[3] مناسب‌ترین روش برای مطالعه و بهسازی آموزش معلمان است. آنجاکه هدف اصلی ارزشیابی فعالیتهای آموزشی معلمان کشف نواقص و اشکالات روش آموزش و کلاسداری معلم است، استفاده از نتایج خودسنجی‌های معلمی اجتناب‌ناپذیر است. حتی نتایج روشهای دیگر ارزشیابی معلمان، نظیر روش مشاهدهٔ کلاسی یا روش نظرخواهی از شاگردان، وقتی بهترین فایده را برای اصلاح کار معلمان فراهم می‌آورند که نتایج آنها در اختیار معلمان گذاشته شوند و آنان با اطلاع از این نتایج به سنجش کار خود بپردازند.

---

1. unreliable
2. self-assessment
3. self-evaluation

کارول[1] (۱۹۸۱، ص ۱۸۰) روش خودسنجی معلم را به عنوان "قضاوت دربارهٔ آموزش خود"، تعریف کرده است. آیراسیان[2] و گولیکسون (۲۰۰۶) نیز تعریف زیر را از این روش به دست داده‌اند: "روش ارزشیابی از خود فرآیندی است که در آن معلمان دربارهٔ شایستگی و اثربخشی دانش، عملکرد، باورها، و تأثیرشان در رابطه با هدف بهسازی شخصی به داوری می‌پردازند" (ص ۱۸۶). کارول برای خودسنجی معلم ۵ وسیلهٔ عمده را معرفی کرده است: (۱) فرمهای درجه‌بندی شخصی، (۲) گزارشهای شخصی، (۳) مواد مطالعهٔ شخصی، (۴) مشاهدهٔ آموزش همکاران، و (۵) ضبط صوت و ضبط ویدیویی از تدریس، و (۶) خوداندیشی. باربر[3] (۱۹۹۱) نیز بر استفاده از این وسایل در روش خودسنجی معلمان تأکید کرده است. توضیح مختصر این روشها در زیر می‌آید.

فرم درجه‌بندی شخصی

منظور از **فرم درجه‌بندی شخصی**[4] گونه‌ای پرسشنامه است که معلم در آن خودش را با توجه به مهارتهای مختلف آموزشی که در آن فرم آمده است درجه‌بندی یا ارزشیابی می‌کند. فرمهای متداول خودسنجی معلم غالباً از پرسشنامه‌های نظرسنجی دانش‌آموزان و دانشجویان اقتباس می‌شوند. به نمونه‌ای از این فرمهای خودسنجی که در شکل ۵ـ۲۲ آمده است توجه کنید.

فرمهای درجه‌بندی شخصی بهترین وسیله برای مقایسهٔ بین نظر معلم نسبت به خود و سایر منابع اطلاعاتی، مانند درجه‌بندی شاگردان از معلمان، مشاهدات کلاسی، درجه‌بندی همکاران، و تعبیر و تفسیر آنهاست. پژوهشهایی که در این باره انجام گرفته نشان داده‌اند که وقتی تفاوتهای بین درجه‌بندی معلمان از خود و درجه‌بندی شاگردانشان از آنها بر معلمان معلوم می‌گردند، نمرات بعدی درجه‌بندی شاگردان از معلمان افزایش می‌یابند (پامبوکیان[5]، ۱۹۷۴).

گزارش شخصی

در **گزارش شخصی**[6]، سؤالهای درجه‌بندی در اختیار معلم گذاشته نمی‌شود بلکه از او خواسته می‌شود تا به طور تشریحی نظرش را در بارهٔ ویژگیها و فعالیتهای آموزشی خودش بنویسد. به سخن دیگر، سؤالهای فرمهای گزارش شخصی سؤالهای تشریحی هستند نه سؤالهای علامت‌زدنی. در زیر به نمونه سؤالهایی که از فرم گزارش شخصی هیأت علمی کالج گرین فیلد گرفته شده است توجه کنید (ژنوا[7]، مادوف[8]، چین[9]، و توماس[10]، ۱۹۷۶):

- فکر می‌کنید در چه زمینه‌هایی از موضوعی که درس می‌دهید مسلط‌تر هستید؟

---

1. Carroll
2. Airasian
3. Barber
4. self-rating form
5. Pambookian
6. self-report
7. Genova
8. Madoff
9. Chin
10. Thomas

خودسنجی آموزشی

نام معلم ـــــ نام درس ـــــ ترم تحصیلی ـــــ سال تحصیلی ـــــ

ارزشیابیهای شخصی به‌دقت انجام‌شده از سوی معلمان می‌توانند اثربخشی آموزش معلمان را افزایش دهند. در این پرسشنامه از شما خواسته شده است تا به عملکرد خود در کلاس درس توجه کنید.

سؤالهای ۱۲ و ۱۳ را خودتان تهیه کنید. از پشت این فرم برای نوشتن هرگونه نظری که دارید استفاده نمایید. در اظهار نظر خود هر رویداد غیرمعمول را که در کلاس شما و در رابطه با آموزش شما رخ داده است بنویسید.

راهنمایی:

در رابطه با هریک از سؤالهای زیر خود را درجه‌بندی کنید و بالاترین نمره را به عملکردهای فوق‌العادهٔ خود بدهید. در جای خالی مقابل هر سؤال نمره‌ای را که معرف نظرتان است قرار بدهید.

بالاترین نمره			نمرهٔ متوسط			نمی‌دانم
۷	۶	۵	۴	۳	۲	۱	×

ـــ ۱. هدفهای اصلی درس خود را کاملاً روشن بیان کرده‌اید؟

ـــ ۲. میزان همخوانی بین هدفهای درس و تکالیف درسی چقدر است؟

ـــ ۳. مطالبی را که آموزش می‌دهید به خوبی طراحی و سازمان‌بندی می‌کنید؟

ـــ ۴. نکات مهم را به خوبی توضیح می‌دهید؟

ـــ ۵. میزان تسلط خود را بر محتوای درس چگونه ارزیابی می‌کنید؟

ـــ ۶. از وقت کلاس به خوبی استفاده می‌کنید؟

ـــ ۷. در شاگردان قدرت تفکر انتقادی و تحلیلی را ایجاد می‌کنید؟

ـــ ۸. دانش‌آموزان را تشویق می‌کنید تا موقع لزوم از شما کمک بطلبند؟

ـــ ۹. دانش‌آموزان را به شرکت فعال در کلاس تشویق می‌کنید؟

ـــ ۱۰. تا چه اندازه نظرات شاگردانی را که با شما مخالفت می‌کنند تحمل می‌کنید؟

ـــ ۱۱. با توجه به ۱۰ نکتهٔ بالا، عملکرد خود را در این درس با عملکرد سایر همکارانتان که همین درس را آموزش می‌دهند چگونه ارزیابی می‌کنید؟

ـــ ۱۲.

ـــ ۱۳.

ـــ. نمرهٔ کل

شکل ۵ـ۲۲ یک نمونه فرم خودسنجی معلم (اقتباس از اثر میلر[1]، ۱۹۷۲)

1. Miller

- به نظر خودتان در چه زمینه‌هایی ضعیف‌تر هستید؟
- موفق‌ترین جنبهٔ آموزشی شما چیست؟
- بزرگ‌ترین نقطه ضعف آموزشی شما چیست؟

گزارشهای شخصی نیز، مانند خودسنجی‌ها، برای تفسیر داده‌های حاصل از منابع دیگر ارزشیابی معلم مفیدند. برای مثال، داده‌های حاصل از روش مشاهدهٔ کلاسی یا روش درجه‌بندی شاگردان از معلمان که در بارهٔ دو کلاس متفاوت یک معلم به دست آمده‌اند ممکن است با هم متفاوت باشند و دلیل آن روشن نباشد. اما زمانی که گزارش شخصی معلم فراهم آمد ممکن است دلیل آن به خوبی روشن گردد. کارول (۱۹۸۱) می‌گوید گزارشهای شخصی معلمان معمولاً از فرمهای درجه‌بندی اطلاعات کامل‌تری در بارهٔ آموزش معلم به دست می‌دهند. باربر (۱۹۹۱)، در رابطه با نحوهٔ تهیه و استفاده از گزارش شخصی معلمان، می‌گوید "برای اینکه بهترین نتایج را به دست آورید، فرمهای گزارش شخصی را روشن، بدون ابهام و خلاصه تهیه کنید، و مرتباً آنها را مورد استفاده قرار دهید" (ص ۲۲۱).

مواد آموزش شخصی

مواد آموزش شخصی[1] مطالبی هستند در بارهٔ روشهای آموزش و کلاسداری که به صورت نظامدار و خودآموز تهیه شده و در اختیار معلمان قرار داده می‌شوند. معلمان با مطالعهٔ این مواد و آشناشدن با روشهای درست آموزشی هم به ارزشیابی از کار خود می‌پردازند و یا مشکلات خود آشنا می‌شوند و هم روشهای درست کلاسداری و اموزش را می‌آموزند. امتیاز این روش آن است که معلمان را از شرکت در کلاسها و دوره‌های آموزشی بی‌نیاز می‌کند و موجب می‌شود آنها با صرف اوقات فراغت خود بتوانند کم و کسرهای آموزش خود را برطرف سازند.

روش استفاده از مواد آموزش شخصی عمدتاً برای معلمان سطوح آموزش عالی مفید است که غالباً بدون گذراندن دوره‌های تخصصی معلمی وارد این حرفه می‌شوند. این مشکل نه‌تنها در ایران بلکه در سایر کشورهای جهان نیز به چشم می‌خورد. به قول کارول (۱۹۸۱)، "تعداد کمی از معلمان سطوح آموزش عالی، در روانشناسی پرورشی، علم آموزش، یا برنامه‌ریزی درسی آموزش رسمی می‌بینند. برخلاف آماده‌سازی معلمان ابتدایی و متوسطه، آموزشهای دوره‌های دکتری استادان دانشگاه معمولاً به‌علم و پژوهش اختصاص دارد و در آن کمترین توجهی به روشها و فرایندهای آموزشی نمی‌شود" (ص ۱۹۱). با این توضیحات معلوم می‌شود که مواد آموزشی هم برای سنجش آموزش و هم برای رفع نواقص آموزش مورد استفادهٔ معلمان قرار می‌گیرند.

1. self-study materials

مشاهدهٔ آموزش همکاران

یکی دیگر از راههای خودسنجی معلم *مشاهدهٔ آموزش همکاران* و قضاوت در بارهٔ روش خودش است. این‌گونه مشاهدات شکلهای مختلفی دارد. یک شکل آن آموزش گروهی یک درس است که در آن چند معلم مشترکاً آموزش یک درس را به عهده می‌گیرند و هر یک به نوبت بخشی از آن را در حضور همکاران خود تدریس می‌کند. این کار شرایط مناسبی برای مقایسهٔ روش معلم با دیگران و قضاوت در بارهٔ کار خودش فراهم می‌آورد.

همچنین تعدادی از معلمان می‌توانند با هم قرار بگذارند که به نوبت از کلاسهای همدیگر بازدید به عمل آورند و شاهد روشهای تدریس یکدیگر باشند. در این روش، معلمان مشاهده‌گر و مورد مشاهده می‌توانند نظرهای خود را مبادله کنند و به بهبود کار همدیگر کمک نمایند.

بهترین نتیجه زمانی از روش مشاهدهٔ آموزش دیگران می‌تواند عاید معلمان بشود که ابتدا مواد آموزشی شخصی را مطالعه کنند و سپس به مشاهده از کار یکدیگر بپردازند. در ضمن اگر روشهای خودسنجی مورد بحث در این قسمت به طور مکمل یکدیگر به کار روند بهترین نتیجه را به دست می‌دهند.

بازخورد حاصل از ضبط صدا و ضبط تصویر

بهترین فرصت برای خودسنجی *مشاهدهٔ فعالیتهای خود در نوارهای ویدیویی* یا *شنیدن صدای خود از نوارهای ضبط صوت* و قضاوت کردن در بارهٔ آن است. کارول (۱۹۸۱) می‌گوید استفاده از تکنولوژی ثبت الکترونیکی در بهبود روش آموزش معلمان تأثیر فراوانی داشته است.

یکی از مشکلات استفاده از نوار ویدئویی این است که معلمان نمی‌دانند که دقیقاً در مشاهدهٔ رفتار خود در نوار چه چیزی را مورد توجه قرار دهند. به همین منظور، روشهای مختلفی ابداع شده‌اند تا به معلم کمک کنند دقیقاً چه چیزی را جستجو نماید. بهترین آنها روش **آموزش خُرد**[1] است که اولین بار به وسیلهٔ آلن[2] و ریان[3] (۱۹۶۹) معرفی شد.

آموزش خُرد روشی است برای تمرین و انتقاد از مهارتهای آموزش به صورت مشاهدهٔ مستقیم. پیام عمدهٔ آن این است که با تأکید بر اجزای آموزش مانند مقدمه‌چینی، سؤال کردن، استفاده از مثال، و نتیجه‌گیری می‌توان فرایند آموزش را تحلیل کرد و آن را بهبود بخشید. فرایند آموزش خُرد شامل بخش کوتاهی است از آموزش تمرینی در حضور شاگردان و همکاران که حدود ۵ تا ۱۰ دقیقه به طول می‌انجامد. جریان درس بر روی نوار ویدیو ضبط می‌شود و بلافاصله به نمایش درمی‌آید. پس از آنکه نمایش فیلم پایان یافت، مدرس و سایر شرکت‌کنندگان در بحث مهارتهای ارائه‌شده در درس را طبق ملاکهای آموزش اثربخش مورد نقد و بررسی قرار می‌دهند و رهنمودهای لازم را به مدرس ارائه می‌نمایند.

---

1. microteaching 2. Allen 3. Ryan

خوداندیشی

آیراسیان و گولیکسون (۲۰۰۶) مهم‌ترین جنبهٔ روش ارزشیابی از خود را **خوداندیشی**[1] می‌دانند. آنها دربارهٔ خوداندیشی یا تفکر دربارهٔ خود گفته‌اند:

> عنصر تکمیل کنندهٔ ارزشیابی از خود در کار معلمی خوداندیشی است؛ یعنی عملِ فکر کردن دربارهٔ یک موضوع یا مشکل.... معلم باید دربارهٔ عملکرد خود بیندیشد تا آن را درک کند، از آن انتقاد نماید، و تغییرش دهد.... **پاسخگویی به خود**[2] یک جنبهٔ مهم حرفهٔ معلمی است زیرا گویای این تعهد اخلاقی است که باید به دانش‌آموزان خدمت کرد و ماهیت فعالیت خود را، به قصد بهبود بخشیدن به آن، مورد بازبینی قرار داد. (ص ۱۹۰)

شکل ۶ـ۲۲ یک پرسشنامهٔ ویژهٔ خوداندیشی برای معلمان را نشان می‌دهد.

راهنمایی: هر سؤال را با کشیدن دایره‌ای به دور جوابی که معرف نظر شماست جواب دهید.

۱. من به طور فعال مشغول یادگیری و گسترش دانشم دربارهٔ موضوعهایی که درس می‌دهم هستم.

کاملاً موافق　موافق　مخالف　کاملاً مخالف

موضوعهایی که فعلاً مشغول یادگیری آنها هستم.　موضوعهایی که بعداً به یادگیری آنها خواهم پرداخت.

.....　.....

.....　.....

۲. من به طور فعال مشغول فکر کردن به موضوعهای مربوط به خودم به عنوان معلم و موضوعهای مربوط به حرفه‌ام هستم.

کاملاً موافق　موافق　مخالف　کاملاً مخالف

موضوعهایی که فعلاً دربارهٔ آنها فکر می‌کنم.　موضوعهایی که تصمیم دارم بعداً درباره‌شان فکر کنم.

.....　.....

.....　.....

۳. من به طور فعال مشغول یادگیری و گسترش مهارتهای خاص آموزشی و سنجشی‌ام هستم.

کاملاً موافق　موافق　مخالف　کاملاً مخالف

چیزهایی که فعلاً مشغول یادگیری‌شان هستم.　چیزهایی که تصمیم دارم بعداً یاد بگیرم.

.....　.....

.....　.....

۴. یکی از زمینه‌های تدریس خود را که بیشتر نیاز به یادگیری چیزهای تازه دربارهٔ آن دارید کدام است؟ براساس چه شواهدی موجب شده که شما این زمینه را انتخاب کنید؟

.....

.....

شکل ۶ـ۲۲ یک نمونه پرسشنامهٔ خوداندیشی (آیراسیان و گولیکسون، ۲۰۰۶، ص ۲۰۱)

1. reflection　2. accountability to oneself

محاسن و معایب روش خودسنجی

چنان‌که گفتیم، روش خودسنجی برای کشف نواقص آموزشی معلم و کمک به رفع آنها از بهترین روشهای ارزشیابی معلم است. در این روش خود معلم مسئول ارزشیابی از کار خود و مهم‌ترین عامل در رفع مشکلات کار خویش است.

با وجود این، کسانی به استفاده از روش خودسنجی معلم، به‌ویژه استفاده از آن برای اخذ تصمیمهای شغلی، ایراد گرفته‌اند. از جمله گفته‌اند اصطلاح *خودسنجی یا ارزشیابی از خود* نقض غرض است، زیرا هدف اصلی ارزشیابی از معلم قضاوت کردن در بارهٔ اثربخشی کار او به منظور استخدام، ارتقای مرتبه، افزایش حقوق، و مانند اینهاست، و از این لحاظ معلمان نمی‌توانند خود به قضاوت در بارهٔ کار خویش بپردازند. کارول (۱۹۸۱) می‌گوید "ارزشیابیها معمولاً اندازه‌هایی هستند که از بیرون بر اعضای هیأت علمی تحمیل می‌شوند، مانند نمراتی که راهنمایان تعلیماتی، همکاران، یا دانشجویان به معلمان می‌دهند. اگر هدف ارزشیابی حسابرسی بین هزینه و درآمد باشد آنگاه اصطلاح ارزشیابی از خود چیز بی‌موردی خواهد بود" (ص ۱۸۰).

به‌رغم اختلاف نظرهای موجود دربارهٔ روش خودسنجی، اگر قرار باشد معلمان از جنبه‌های مختلف مورد ارزشیابی قرار گیرند و اطلاعات کاملی در بارهٔ کمّ و کیف کار آنان فراهم آید، استفاده از این روش اجتناب‌ناپذیر است. به‌ویژه زمانی که هدف عمدهٔ ارزشیابی معلم کشف نواقص کار او و کمک به بهبود روش آموزش اوست، خودسنجی یا ارزشیابی شخصی روش بسیار مفیدی است.

ارزشیابی به وسیلهٔ همکاران

اگر چه روش **ارزشیابی به وسیلهٔ همکاران**[1] به منظور داوری در بارهٔ فعالیتهای علمی و پژوهشی معلمان، به‌ویژه اعضای هیأت علمی دانشگاهها، فراوان به کار می‌رود، با این حال، از این روش برای ارزشیابی اثربخشی کیفیت آموزشی معلمان خیلی کم استفاده شده است (فرنچ ـ لازویک[2]، ۱۹۸۱). در بارهٔ چگونگی استفاده از نظر همکاران، لازم است به همان طریقی که سوابق و مدارک مربوط به فعالیتهای پژوهشی عضو هیأت علمی به منظور داوری در بارهٔ بارآوری علمی و پژوهشی او جمع‌آوری می‌شوند، اطلاعات مربوط به اثربخشی آموزشی او نیز گردآوری شوند. این گونه اطلاعات باید تا حد امکان عینی و ملموس باشند. فرنچ ـ لازویک مواردی راکه می‌توانند در پروندهٔ سوابق آموزشی معلم به منظور داوری همکاران او گنجانیده شوند به شرح زیر بیان داشته است: توصیف مختصری از هریک از دروسی که معلم تدریس می‌کند، هدفهای آن درس، تعداد دانشجویان کلاس، تعداد واحدهای درسی، نوع درس (اصلی، تخصصی، اختیاری)، سطح درس (مقدماتی، پیشرفته، کارشناسی، کارشناسی ارشد، دکتری)، تدریس داخل دانشکده‌ای که

1. colleagues evaluation 2. French-Lazovik

فرد عضو آن است یا خارج از دانشکدهٔ او؛ توصیف مختصری از نوع مشکلات وابسته به تدریس آن درس؛ و نحوهٔ نمره گذاری معلم در آن درس.

با در دسترس‌بودن اطلاعات فوق در بارهٔ هریک از دروس معلم، همکاران متخصص و هم‌رشتهٔ او می‌توانند به ارزشیابی از کیفیت آموزشی او بپردازند. فرنچ ـ لازویک (۱۹۸۱) در رابطه با ارزشیابی از کیفیت آموزش دروس طرح سؤالهای زیر و تهیهٔ جواب برای آنها را پیشنهاد داده است:

۱. کیفیت مواد مورد استفادهٔ معلم در آموزش چگونه است؟ منابعی که با مراجعه به آنها می‌توان این سؤال را جواب داد عبارت‌اند از: فهرست موضوعهای درس، فهرست تکالیفی که به دانشجویان داده می‌شود، کتاب یا جزوهٔ درسی و سایر منابع خواندنی. در رابطه با این مطالب سؤالهای زیر را می‌توان پرسید: مطالب تازه هستند؟ معرف بهترین کارهای رشتهٔ علمی مورد نظر هستند؟ مطالب و موضوعها می‌توانند دانشجویان را به هدفهای پیش‌بینی شده برای درس برسانند؟

۲. چه نوع تکالیف ذهنی برای دانشجویان در نظر گرفته شده است و دانشجویان چگونه به آنها عمل کردند؟ برای جواب‌دادن به این سؤال باید به امتحانهای معلم، نمونه مقالات تحقیقی انجام‌شده به توسط دانشجویان، و نمونه تکالیف انجام‌شدهٔ دانشجویان مراجعه کرد. در دسترس‌بودن این مطالب به اعضای کمیتهٔ ارزشیابی امکان می‌دهد تا سطح کیفی عملکرد ذهنی یادگیرندگان را مورد داوری قرار دهند.

۳. میزان اطلاعات معلم در موضوعی که تدریس می‌کند چه اندازه است؟ برای جواب‌دادن به این سؤال لازم است که "داوری همکاران معلم بر مشاهداتی که به طور روزانه انجام می‌گیرند استوار باشد: یعنی اظهارنظرهای معلم در بارهٔ اندیشه‌ها و پیشرفتهای تازهٔ علمی، انتخاب مواد و کتابهای درسی، و شرکت در جلسات بحث و سخنرانی" (فرنچ ـ لازویک، ۱۹۸۱، ص ۷۷).

۴. تا چه اندازه برای بهبود کیفیت کار خود کوشش می‌کند؟ نمونه کارهایی که معلم می‌تواند انجام دهد تا سطح کیفیت آموزشی خود را بالا ببرد و همکاران معلم می‌توانند با بررسی آنها به قضاوت در بارهٔ معلم مورد ارزشیابی بپردازند از این قرار است: بازخوردهایی که معلم در رابطه با درس خود از طریق پرسشنامه‌هایی که به دانشجویان می‌دهد کسب می‌نماید؛ تغییراتی که بر اساس این بازخوردها در کار خود به وجود می‌آورد؛ ایجاد و گسترش اندیشه‌های نو در زمینهٔ آموزش دروس خود؛ و شرکت در فعالیتهایی که هدف آنها بالابردن سطح کیفی آموزش است.

پس از جمع‌آوری اطلاعات و شواهد لازم برای ارزشیابی از کیفیت کار معلم، سه تا شش نفر از اعضای هیأت علمی همکار معلم که بیشترین آشنایی را با رشتهٔ علمی و تخصصی او دارند توسط مدیر گروه آموزشی یا سرپرست مؤسسهٔ آموزشی برگزیده می‌شوند و به آنان مأموریت داده

می‌شود تا بر اساس مدارک و شواهد موجود به داوری در بارهٔ عضو هیأت علمی بپردازند. لازم است برای این منظور یک فرم کتبی درجه‌بندی از پیش آماده شود تا هریک از اعضای هیأت داوران با توجه به سؤالات بالا به طور مستقل در بارهٔ عضو هیأت علمی داوری کند. بر اساس میانگین نظرات اعضای هیأت داوری قضاوت نهایی دربارهٔ عضو هیأت علمی مورد ارزشیابی صورت می‌پذیرد.

علاوه بر بررسی اطلاعات و شواهد فوق، دیدار معلمان از کلاسهای درس یکدیگر و صرف مقداری وقت برای مشاهدهٔ فعالیتهای آموزشی همدیگر، کیفیت ارزشیابی به وسیلهٔ همکاران را بهبود می‌بخشد. سنترا (۱۹۸۹) گفته است آموزش اصول مشاهدهٔ دقیق به اعضای هیأت علمی پایایی روش مشاهدهٔ آنان را افزایش می‌دهد. با این حال، مشاهده به تنهایی نمی‌تواند برای داوری همکاران معلم در بارهٔ کیفیت کار او کافی باشد، بلکه لازم است در رابطه با فعالیتهای مختلف آموزشی معلم هر گونه مدرک و گواه مستندی که ممکن باشد فراهم آورد.

محاسن و معایب روش ارزشیابی به وسیلهٔ همکاران معلم

روش ارزشیابی معلم به وسیلهٔ همکاران او نیز مانند سایر روشهای ارزشیابی معلم خالی از اشکال نیست. یکی از مشکلات درگیر با این ارزشیابی پایین بودن سطح پایایی آن است، بدین معنی که بین نمراتی که معلمان مختلف به یک معلم می‌دهند همخوانی زیادی دیده نمی‌شود. علاوه بر این، اعضای هیأت علمی در ارزشیابی از کیفیتِ آموزشی یکدیگر نمرات سطح بالایی به هم می‌دهند و در مشاهدات کلاسی خود از کار یکدیگر میان آنها توافق نظر زیادی به چشم نمی‌خورد (سنترا، ۱۹۷۵).

به‌رغم مشکلات فوق، استفاده از نظر همکاران معلم برای قضاوت در بارهٔ کیفیت آموزش او اجتناب ناپذیر است. این روش به عنوان مکمل سایر روشهای ارزشیابی معلم، به‌ویژه در سطح آموزش عالی، بسیار مفید است. سنترا (۱۹۸۹)، در دفاع از این روش گفته است، "اگر بخواهیم ارزشیابی معلم را از انحصار فرمهای نظرخواهی از دانشجویان درآوریم، لازم است طراحان نظام ارزشیابی راههای معتبر سنجش توسط همکاران معلم را مورد رسیدگی قرار دهند" (ص ۱۰۹).

ارزشیابی به کمک کارپوشه

ما در فصل ۱۳ کتاب **کارپوشه**[1] را به عنوان یکی از روشهای عملکردی سنجش یادگیری دانش‌آموزان و دانشجویان توضیح دادیم. در اینجا نیز آن را به عنوان یک روش ارزشیابی معلم به نام **کارپوشهٔ آموزشی**[2] مورد بحث قرار می‌دهیم. به معنی کلی، کارپوشه "به مجموعه کارهایی که یک فرد در طول یک مدت جمع‌آوری می‌کند گفته می‌شود" (هوگان، ۲۰۰۷، ص ۴۰۹). برای

1. portfolio
2. teaching portfolio

ارزشیابی اثربخشی کار معلمان، کارپوشهٔ آموزشی "به مجموعه اطلاعاتی دربارهٔ کارهای معلم گفته می‌شود" (ولف، ۲۰۰۶، ص ۱۶۸). از کارپوشهٔ آموزشی، مانند کارپوشهٔ دانش‌آموزی، هم برای ارزشیابی از کار معلمان و هم آموزش و توسعهٔ دانش و مهارتهای آنان استفاده می‌شود (سیف، ۱۳۸۷). ولف (۲۰۰۶) در این باره گفته است "اکنون کارپوشه‌ها در موقعیتهای مختلفی به کار می‌روند ــ در برنامه‌های دانشگاهی، برای تربیت معلمان در آموزشهای پیش از خدمت و در آموزش و پرورش، برای تأیید شایستگی و تشویق معلمان و نیز برای استخدام معلمان و ارزشیابی از کار آنان" (ص ۱۶۸). (ما در فصل ۱۳ کتاب حاضر مثالی از کاربرد کارپوشه برای تربیت معلمان و ارزشیابی از کار آنان توضیح دادیم.)

ویژگیهای کارپوشهٔ آموزشی

چنان که دیدیم، کارپوشهٔ آموزشی به مجموعه‌ای از دستاوردهای حرفه‌ای معلم گفته می‌شود. ولف (۲۰۰۶) ویژگی عمدهٔ کارپوشهٔ آموزشی را به گونهٔ زیر بیان کرده است: "یک تاریخچهٔ مستند از دستاوردهای معلم که به دقّت انتخاب شده و با نمونه کارهای دانش‌آموزان تکمیل گردیده است" (ص ۱۶۹). ویژگیهای دیگر کارپوشهٔ آموزشی به قرار زیرند:

- باید دربرگیرندهٔ مثالها و موارد به دقت انتخاب شده از کارهای معلم و دانش‌آموزان باشد، به گونه‌ای که جنبه‌های مهم فعالیت معلم را نشان دهد.
- محتوای کارپوشه باید با توضیحات و شرح و تفصیل همراه باشد.

هدفهای کارپوشهٔ آموزشی

کارپوشه‌های آموزشی هدفها یا مقاصد زیادی را برآورده می‌سازند. از جمله می‌توان به موارد زیر اشاره کرد:

- مقاصد ارزشیابی
- کمک به رشد حرفه‌ای معلمان
- کمک به کارفرمایان در استخدام معلمان

هرچند که کارپوشه‌های به دقّت تنظیم یافته را می‌توان برای هر سه هدف فوق مورد استفاده قرار داد، با این حال، این کارپوشه‌ها از جهاتی با هم فرق دارند. از جمله اینکه، کارپوشه‌هایی که هدف آنها ارزشیابی است باید از کارپوشه‌هایی که به قصد رشد حرفه‌ای ساخته می‌شوند از ساختار دقیق‌تر و منسجم‌تری برخوردار باشند. بدین منظور، لازم است که ساختار و محتوای کارپوشه‌های ارزشیابی از پیش مشخص گردند تا اینکه شرایط تکمیل کارپوشه و فرایند ارزشیابی

برای افراد مختلف همسان باشد. در مقابل، در کارپوشه‌هایی که به منظور رشد حرفه‌ای معلمان درست می‌شوند، تنوع انتخاب موضوعها باید منظور شود، زیرا هر معلمی کارپوشه‌اش را مطابق با نیازها و هدفهای ویژهٔ خودش تهیه می‌کند.

از سوی دیگر، کارپوشه‌های استخدامی که برای ملاحظهٔ کارفرمایان درست می‌شوند باید از لحاظ موضوعها محدودتر و از جهت محتوا مختصرتر باشند، زیرا کسانی که مسئول استخدام معلمان‌اند فرصت زیادی برای مطالعهٔ محتوای کارپوشه ندارند و عموماً به دنبال یک برداشت کلی از کارپوشه هستند. از این رو، تهیه کنندگان این نوع کارپوشه‌ها بهتر است به نحوهٔ ارائهٔ مطالب کارپوشه بیشتر دقت کنند ــ مانند جذابیت کارپوشه و سهولت مراجعه به اطلاعات موردنیاز. به رغم اختلافات فوق، "یک کارپوشهٔ واحد می‌تواند هر سه هدف بالا را برآورده سازد، به شرط آنکه شخص تهیه کنندهٔ آن نسبت به هدفها با آگاهی کامل عمل کند و در طراحی و تهیهٔ آن دقت نظر به کار بندد" (ولف، ۲۰۰۶، ص ۱۷۰).

محتوای کارپوشهٔ آموزشی

مطالبی که در درون کارپوشه‌های آموزشی قرار می‌گیرند بسیار متنوع‌اند. از جمله می‌توان به نمونه کارهای معلم و دانش‌آموزان، مانند تصویر پروژه‌های کلاسی، طرح درسها، تکالیف دانش‌آموزان، و شواهد مربوط به فعالیتهای حرفه‌ای معلم اشاره کرد. همچنین موضوعهای متنوع دیگر از جمله نامه‌های رضایت والدین دانش‌آموزان، ارزشیابیهای راهنمایان آموزشی، و مدارک و سوابق تحصیلی معلم را می‌توان در کارپوشه جای داد. ولف (۲۰۰۶) موارد زیر را برای گنجاندن در کارپوشهٔ معلم سفارش داده است:

- بیان فلسفه یا هدفهای آموزشی
- نمونه‌هایی از کارهای معلم، مانند طرح درسها و تکالیف دانش‌آموزان
- نمونه‌هایی از کارهای دانش‌آموزان، تکالیف خواندنی، و پروژه‌های تحقیقی
- توضیحات معلم برای معرفی نمونه‌کارها
- اظهارنظرهای معلم دربارهٔ مدارک مربوط به آموزش و یادگیری موجود در کارپوشه

ارزشیابی از کارپوشهٔ آموزشی

ارزشیابی از کارپوشهٔ معلم باید براساس دو نوع استاندارد *محتوایی* و *عملکردی* صورت پذیرد. استاندارد محتوایی به اینکه معلم چه چیزی می‌داند اشاره می‌کند. استاندارد عملکردی به اینکه معلم به چه خوبی آنچه را که می‌داند می‌تواند انجام دهد گفته می‌شود. این استانداردها باید از پیش به معلم گفته شوند، تا اینکه بداند از او چه انتظاراتی می‌رود. در ضمن، بیان روشن این

استانداردها هم معلم را در تهیهٔ کارپوشه‌اش یاری می‌دهد و هم به ارزشیاب در ارزشیابی از محتوای کارپوشه کمک می‌کند.

در رابطه با نمره گذاری محتوای کارپوشه، ما در فصل ۱۳ دربارهٔ چگونگی نمره گذاری کارپوشه‌های دانش‌آموزان و دانشجویان توضیحاتی داده‌ایم. آن روشها برای نمره گذاری کارپوشه‌های معلمان نیز قابل استفاده‌اند.

محاسن و معایب روش ارزشیابی به کمک کارپوشه

برای استفاده از روش کارپوشه، چه در ارتباط با معلمان و چه دانش‌آموزان و دانشجویان، محاسن زیادی ذکر شده است. بزرگ‌ترین آنها این است که به کمک کارپوشه هم می‌توان آموزش داد و هم ارزشیابی کرد. امتیاز دیگر این است که، به گفتهٔ ولف (۲۰۰۶)، کارپوشه‌ها تصویر روشن‌تری از یادگیری و آموزش ارائه می‌دهند و چشم‌انداز مطمئن‌تری از آنچه معلمان می‌دانند و می‌توانند انجام دهند فراهم می‌آورند.

به رغم امتیازهای زیاد روش کارپوشه، این روش نیز، مانند سایر روشهای ارزشیابی، خالی از اشکال نیست. بزرگ‌ترین اشکال این روش ضعف روایی و پایایی آن است. از مشکلات دیگر روش کارپوشه وقت‌گیر بودن تهیه، جاگیر بودن نگهداری، و دشواری ارزشیابی آن است. با وجود این مشکلات، "اهمیت کارپوشه برای مقاصد ارزشیابی و توسعهٔ یادگیری حرفه‌ای معلمان حکم می‌کند که مدیران برای کارپوشه‌های آموزشی در نظام ارزشیابی و توسعهٔ هیأت آموزشی خود نقش ارزنده‌ای درنظر بگیرند" (ولف، ۲۰۰۶، ص ۱۶۸).

بررسی آثار علمی معلم

روش بررسی آثار علمی یک روش ارزشیابی ویژهٔ اعضای هیأتهای علمی دانشگاهها و سایر مؤسسات آموزش عالی است. در این روش ارزشیابی، کمیته‌ای متشکل از اعضای هیأت علمی، متخصص در رشتهٔ مورد ارزشیابی یا متخصص در رشته‌های نزدیک به آن، از درون یا بیرون مؤسسهٔ مربوط به او، برگزیده می‌شوند و آثار علمی عضو مورد نظر را ارزشیابی می‌کنند. ملاک ارزشیابی در این کمیته‌ها هم کمّیت و هم کیفیت آثار علمی است. آثار علمی مورد ارزیابی عبارت‌اند از:

- کتابهای تألیف شده
- کتابهای ترجمه یا تصحیح شده
- مقاله‌های چاپ شده در مجله‌های تخصصی
- فصلهای چاپ شده در کتابهایی که به وسیلهٔ مؤلفان مختلف تألیف شده‌اند
- خلاصه مقاله‌های چاپ‌شده در مجموعهٔ خلاصه مقالات ارائه‌شده در سمینارها

علاوه بر موارد بالا، درجات و افتخارات کسب‌شده از سوی سازمانهای علمی، بورسهای علمی ـ تحقیقاتی، وعضویت در شورای دبیران و داوران مجله‌های علمی نیز امتیاز می‌گیرند.

از لحاظ سنجش کمّی، آثار مختلف عضو هیأت علمی شمارش می‌شوند و با توجه به امتیازهایی که کمیتهٔ ارزشیابی برای هر مورد از پیش تعیین کرده است مجموع امتیازهای عضو هیأت علمی تعیین می‌گردد. از لحاظ سنجش کیفی، محتوای آثار توسط اعضای متخصص کمیتهٔ بررسی‌کننده ارزیابی و امتیازبندی می‌شوند.

یکی از روشهای نسبتاً جدید بررسی کیفیت آثار علمی، تعیین دفعاتی است که آن آثار به وسیلهٔ کسان دیگر مورد استناد قرار گرفته‌اند (کروک[1]، ۱۹۸۴، سنترا، ۱۹۷۵). در این روش، با مراجعه به آثار مکتوب سایر متخصصان هم‌رشتهٔ عضو هیأت علمی یا با مراجعه به ادبیات پژوهشی تخصصی مربوط به زمینهٔ کاری عضو هیأت علمی مورد ارزشیابی، تعیین می‌کنند که آثار عضو هیأت علمی مورد ارزشیابی چند بار مورد استناد قرار گرفته‌اند، یعنی چند بار به افکار، اندیشه‌ها، یافته‌ها، و پژوهشهای او استناد شده است. نخستین بار در سال ۱۹۶۶ *نمایهٔ استناد در علوم اجتماعی*[2] پایه گذاری شد. در این نمایه، تعداد دفعاتی که یک مقالهٔ علمی یا یک کتاب در طول یک سال مورد استناد قرار گرفته است مشخص می‌شود. همچنین در این نمایه منبعی که به مقاله یا کتاب استناد کرده است نیز مشخص می‌شود. این اطلاعات زیر نام مؤلف آثار علمی و عنوان اثر فهرست شده‌اند. این فهرست بسیار جامع است و تعداد زیادی مجلهٔ علمی را شامل می‌شود (کروک، ۱۹۸۴).

روش بررسی آثار علمی اعضای هیأت علمی نه‌تنها به عنوان یک روش سودمند برای ارزشیابی از فعالیتهای علمی خود اعضای هیأت علمی به کار می‌رود، بلکه به منظور ارزشیابی از گروهها، دانشکده‌ها، و دانشگاهها نیز مورد استفاده قرار می‌گیرد. بدین‌منظور، شمار دفعاتی که آثار اعضای هیأت علمی یک واحد یا مؤسسهٔ آموزشی مورد استناد قرار گرفته است ملاک داوری در بارهٔ اثربخشی یا بهره‌دهی آن واحد یا مؤسسه واقع می‌شود.

کروک (۱۹۸۴) روش بالا را برای درجه‌بندی ۵۱ دانشکدهٔ علوم تربیتی معروف کشور آمریکا به کار بست. نام اعضای هیأت علمی این دانشکده‌ها از کتابچهٔ راهنمای دانشگاه مربوط یا از طریق مکاتبه با دانشکده مشخص شد. در زمان انجام مطالعهٔ کروک، این ۵۱ دانشکده در کل ۴۶۰۰ عضو هیأت علمی داشتند که از ۱۸ تا ۱۹۹ بین دانشکده‌های مختلف پراکنده بودند. برای هریک از اعضای هیأت علمی دانشکده‌های مورد مطالعه تعداد استنادهای مندرج در *نمایهٔ استناد در علوم اجتماعی* برای سال ۱۹۸۱ شمارش شد و مجموع آنها برای هریک از دانشکده‌ها مشخص گردید. نتایج به‌دست‌آمده حاکی از وجود ۲۸۰۰۰ بار استناد برای ۴۶۰۰ عضو هیأت علمیِ

1. Kroc
2. Social Science Citation Index (SSCI)

۵۱ دانشکدهٔ علوم تربیتی با میانگین ۶/۰۲ بار استناد بود. بالاترین رقم استناد ۶۱۸ مورد و کمترین آن صفر مورد بود. جدول ۱ـ۲۲ نتایج پژوهش کروک (۱۹۸۴) را نشان می‌دهد. در این جدول، برای هریک از دانشگاههای مورد مطالعه، میانگین دفعات استناد، درصد اعضای هیأت علمی دارای ۱۰۰ـ۱۰ مورد استناد، درصد اعضای هیأت علمی بدون استناد، و رتبهٔ دانشگاه نیز داده شده است. چنان‌که دیده می‌شود، دانشگاه هاروارد بالاترین و دانشگاه مونتانا پایین‌ترین رتبه را کسب کرده‌اند.

محاسن و معایب روش بررسی آثار علمی معلم

چنان‌که در رابطه با سایر روشهای ارزشیابی معلم دیدیم، روش بررسی آثار علمی اعضای هیأت علمی برای قضاوت در بارهٔ اثربخشی فعالیتهای آنان نیز بی‌ایراد نبوده است. بزرگ‌ترین اشکالی که می‌توان به این روش گرفت این است که گرچه شمارش تعداد آثار علمی عضو هیأت علمی (یعنی ارزشیابی کمّی) آسان است، اما بررسی سطح علمی آثار (ارزشیابی کیفی) کار ساده‌ای نیست.

روش استفاده از تعداد استنادهای عضو هیأت علمی برای ارزشیابی از میزان بهره‌دهی فعالیتهای او و ارزشیابی از مؤسسه‌ای که عضو هیأت علمی به آن وابسته است نیز دارای اشکالاتی است. مهم‌ترین مشکل آن یافتن راهی برای شمارش دفعاتی است که آثار یک فرد مورد استناد دیگران قرار می‌گیرد. حتی وجود تدابیری نظیر *نمایهٔ استناد در علوم اجتماعی* که در کشور آمریکا متداول است نیز همراه با مشکلاتی است. از جمله اینکه گفته شده است که گاه شباهت اسمی بین مؤلفان موجب دشواری می‌شود. علاوه بر این، در *نمایهٔ استناد در علوم اجتماعی* تنها نام اولین مؤلف آثار داده می‌شود و لذا امتیاز استناد برای مؤلفان دوم به بعد منظور نمی‌شود. علاوه بر این، بعضی پژوهشها (از جمله گارفیلد[1]، ۱۹۷۹، لیندسی[2]، ۱۹۷۸) نشان داده‌اند که پژوهشهای **نظری**[3] و **روش‌شناختی**[4] بیشتر از پژوهشهای **کاربردی**[5] مورد استناد قرار می‌گیرند و این به ضرر کسانی تمام می‌شود که بیشتر پژوهشهای کاربردی انجام می‌دهند.

به‌رغم مشکلات درگیر با روش بررسی کمیت و کیفیت آثار اعضای هیأت علمی و شمارش تعداد استنادها به آثار آنان به منظور ارزشیابی اثربخشی فعالیتهای علمی اعضای هیأت علمی و مؤسسات آموزشی، این روش بهترین روش ممکن برای سنجش بارآوری علمی اعضای هیأت علمی دانشگاهها و سایر مؤسسات آموزش عالی است. لذا صاحب‌نظران استفاده از آن را توصیه می‌کنند.

1. Garfield
2. Lindsey
3. theoretical
4. methodological
5. applied

جدول ۱ ـ ۲۲ رتبه‌بندی ۵۱ دانشکدهٔ علوم تربیتی بر اساس تعداد استنادهای اعضای هیأت علمی آنها (کروک، ۱۹۸۴)

	میانگین استنادها	درصد اعضای هیأت علمی دارای ۱۰۰ ـ ۱۰ استناد	درصد اعضای هیأت علمی بدون استناد	رتبهٔ دانشگاه
دانشگاه شیکاگو	۴۶/۰۰	۵۰	۱۳	۲
دانشگاه هاروارد	۳۸/۰۲	۳۵	۱۸	۱
دانشگاه استانفورد	۲۹/۲۹	۵۷	۱۲	۷
دانشگاه کالیفرنیا در برکلی	۲۷/۳۸	۳۹	۸	۹
دانشگاه پورت وسترن	۲۱/۱۱	۱۸	۴۲	۱۰
دانشگاه کالیفرنیا در لوس آنجلس	۱۶/۰۰	۳۷	۱۹	۱۱
دانشگاه واشنگتن (سنت لوئیس)	۱۳/۵۶	۲۸	۱۷	۴۱
دانشگاه پنسیلوانیا	۱۱/۹۴	۲۹	۲۳	۲۵
دانشگاه آرگون	۱۱/۳۶	۱۲	۴۴	۸
دانشگاه مینه‌سوتا	۱۰/۷۸	۲۰	۴۹	۳
دانشگاه کالیفرنیا در سانتا باربارا	۱۰/۲۰	۲۸	۲۰	۳۷
دانشگاه ایلینویز در اوربانا ـ شامپین	۹/۸۲	۱۹	۴۲	۴
دانشگاه کلمبیا	۹/۸۰	۱۸	۳۳	۵
دانشگاه روچستر	۹/۱۶	۱۶	۳۲	۳۲
دانشگاه کلرادو	۸/۶۴	۸	۶۴	۲۸
دانشگاه ایلینویز در حلقهٔ شیکاگو	۸/۱۲	۱۴	۵۲	۲۴
دانشگاه آریزونا	۶/۳۷	۸	۶۵	۱۳
دانشگاه ایالتی اهایو	۶/۱۷	۱۶	۶۲	۶
دانشگاه ویسکانسین	۵/۶۶	۱۵	۴۰	۱۲
دانشگاه ایالتی میشیگان	۵/۵۵	۱۴	۵۹	۱۴
دانشگاه کرنل	۵/۰۰	۱۴	۵۸	۴۴
دانشگاه کارولینای شمالی	۴/۶۸	۱۰	۵۱	۳۵
دانشگاه بوستون	۴/۶۰	۸	۵۱	۲۹
دانشگاه سیراکیوز	۴/۴۶	۲۰	۵۱	۱۹
دانشگاه ایالتی پنسیلوانیا	۴/۳۲	۱۳	۴۹	۲۲
دانشگاه نیویورک	۴/۲۷	۹	۶۵	۱۸
دانشگاه واشنگتن	۴/۱۹	۱۲	۴۱	۴۰
دانشگاه آیوا	۴/۰۶	۱۲	۶۵	۲۱
دانشگاه تگزاس	۴/۰۳	۱۶	۵۴	۱۶
دانشگاه وندربیلت	۳/۷۲	۹	۵۱	۳۹
دانشگاه پیتسبرگ	۳/۴۷	۸	۵۸	۱۵
دانشگاه روتجرز	۳/۴۳	۸	۴۹	۴۳
دانشگاه میشیگان	۳/۳۵	۷	۵۴	۲۳
دانشگاه ایالتی جرجیا	۳/۳۴	۶	۶۳	۱۷
دانشگاه پوردو	۳/۲۱	۱۰	۵۲	۲۷
دانشگاه تمپل	۲/۸۷	۸	۶۱	۳۴
دانشگاه ویرجینیا	۲/۷۸	۷	۶۲	۳۰
دانشگاه هیوستون	۲/۶۶	۷	۶۰	۳۸
دانشگاه ایالتی آریزونا	۲/۴۷	۷	۶۴	۲۶
دانشگاه کانزاس	۲/۳۵	۶	۶۱	۳۶
دانشگاه کانکتیکت	۲/۳۳	۴	۶۸	۴۲
دانشگاه کاتولیک	۲/۲۸	۶	۶۱	۵۰
دانشگاه ایندیانا	۲/۲۳	۵	۵۳	۲۰
دانشگاه کارولینای جنوبی	۲/۱۰	۵	۶۰	۴۵
دانشگاه میسوری	۱/۷۵	۲	۷۱	۳۳
دانشگاه نبراسکا	۱/۷۴	۴	۶۱	۳۱
انستیتوی پلی‌تکنیک ویرجینیا	۱/۷۳	۲	۵۳	۴۷
دانشگاه شهر نیویورک	۱/۷۰	۴	۷۲	۴۶
دانشگاه کالیفرنیای جنوبی	۱/۶۷	۵	۶۰	۴۸
دانشگاه فلوریدا	۱/۳۷	۰	۴۷	۴۹
دانشگاه مونتانا	۰/۴۳	۰	۹۰	۵۱

توضیح: افزودن تعداد استنادهای یکی از استادان بازنشستهٔ دانشگاه استانفورد این دانشگاه را در ردیف دوم قرار خواهد داد.

خلاصهٔ فصل

ارزشیابی از کار معلمان دست‌کم برای دو منظور به کار می‌رود: (۱) کمک به معلمان برای بهبود بخشیدن به عملکرد خودشان و (۲) تعیین میزان موفقیت آنان در ارتباط با وظیفهٔ معلمی، یعنی پاسخگویی آموزشی. اسکریون (۱۹۸۱) این دو نوع هدف ارزشیابی را با اصطلاحات *ارزشیابی تکوینی* و *ارزشیابی تراکمی* نام‌گذاری کرده است. ارزشیابی تکوینی، از راه فراهم آوردن اطلاعات و داوریها و پیشنهادهایی دربارهٔ نحوهٔ آموزش معلمان، به آنها کمک می‌کند تا عملکردهای آموزشی خود را بهبود بخشند. از سوی دیگر، ارزشیابی تراکمی به مدیران و مسئولان مؤسسات آموزشی در تصمیم‌گیری‌های مربوط به استخدام، ارتقاء، افزایش حقوق و سایر امتیازهای شغلی مربوط به معلمان مدد می‌رساند.

متخصصان ارزشیابی معلم، برای مقاصد مختلف ارزشیابی، روشهای گوناگونی را ابداع کرده‌اند. عمده‌ترین آنها با عناوین زیر در این فصل معرفی شدند:

۱. مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلمان
۲. مشاهده از کلاس درس
۳. استفاده از نظر دانش‌آموزان و دانشجویان
۴. استفاده از نظر والدین دانش‌آموزان و دانشجویان
۵. استفاده از یادگیری دانش‌آموزان و دانشجویان
۶. خودسنجی یا ارزشیابی از خود
۷. ارزشیابی به وسیلهٔ همکاران
۸. ارزشیابی به کمک کارپوشه
۹. بررسی آثار علمی معلم

هریک از روشهای ارزشیابی معلم تنها بخشی از فعالیتهای علمی ـ آموزشی معلم را می‌سنجد و بنابراین برای به دست‌آوردن تصویر کاملتری از آموزش و فعالیتهای علمی معلم لازم است روشهای مختلف ارزشیابی مورد استفاده قرار گیرند. این تأکید بر استفاده از روشهای مختلف ارزشیابی به‌ویژه زمانی که قرار است ارزشیابی تراکمی انجام گیرد و نتایج آن برای تصمیم‌گیریهای مؤثر بر سرنوشت شغلی معلمان به کار گرفته شود الزامی است.

سؤالها و تمرینهایی برای بهتر یادگرفتن مطالب فصل

۱. بهترین روش ارزشیابی معلم که بر ملاک واقعی شایستگی معلمی استوار است چیست؟
الف ـ استفاده از نظر یادگیرندگان

ب ـ استفاده از نظر همکاران معلم
پ ـ مشاهده از کلاس درس معلم
ت ـ سنجش آموخته‌های یادگیرندگان

۲. عمده‌ترین دلیل برای اینکه چرا روشهای ارزشیابی معلم گوناگون‌اند چیست؟
الف ـ معلمی کار بسیار پیچیده‌ای است.
ب ـ ملاکهای شایستگی معلمی متنوع‌اند.
پ ـ فعالیتهای معلمان مختلف‌اند.
ت ـ دروسی که معلمان تدریس می‌کنند با هم فرق دارند.

۳. در گزینش معلمان آینده از میان داوطلبان شغل معلمی به هنگام مصاحبه با آنان کدام مورد زیر از اهمیت **کمتری** برخوردار است؟
الف ـ داشتن ویژگیهای شخصیتی مناسب شغل معلمی
ب ـ داشتن اطلاعات کافی از دانش موضوعی که فرد داوطلب آموزش‌دادن آن است
پ ـ علاقه‌مندی به شغل معلمی
ت ـ داشتن شرایط جسمی مناسب شغل معلمی

۴. با کدام‌یک از روشهای جمع‌آوری اطلاعات می‌توان یک گزارش نسبتاً کامل و طبیعی از آنچه در کلاس درس می‌گذرد تهیه نمود؟
الف ـ روش نقلی
ب ـ روش درجه‌بندی
پ ـ روش تعیین فراوانی

۵. کدام‌یک از موارد زیر به عنوان دلیل اصلی مخالفت با کاربرد روش استفاده از یادگیری دانش‌آموزان و دانشجویان در ارزشیابی معلم است؟
الف ـ هیچ‌گونه وسیلهٔ مناسبی برای سنجش میزان یادگیری دانش‌آموزان و دانشجویان وجود ندارد.
ب ـ نظر یادگیرندگان بیشتر از مقدار یادگیری آنان در ارزشیابی از معلم مؤثر است.
پ ـ عوامل دیگری بجز معلم وجود دارند که جریان یادگیری شاگردان را تحت تأثیر قرار می‌دهند.
ت ـ معلمان را به هیچ‌وجه نباید مسئول یادگیری یا عدم‌یادگیری شاگردان به حساب آورد.

۶. هدف ارزشیابی معلم را توضیح دهید.

۷. ارزشیابی معلم را تعریف کنید و آن را با تعریف کلی ارزشیابی آموزشی مقایسه نمایید.

۸. روشهای مختلف ارزشیابی مورد بحث در این فصل را نام ببرید و ویژگیهای مهم هریک را توضیح دهید.

۹. موارد استفادهٔ روش مصاحبه را در رابطه با ارزشیابی معلم توضیح دهید.

۱۰. محاسن و معایب روش مصاحبه را توضیح دهید.

۱۱. عمده‌ترین شاخصهای اثربخشی معلم راکه مشاهده کننده (ارزشیاب) باید به مشاهده و ثبت و ارزشیابی آنها بپردازد توضیح دهید.

۱۲. روشهای مختلف جمع‌آوری اطلاعات را با روش مشاهده توضیح دهید.

۱۳. نمونه گیری از آموزش معلم به کمک مشاهده را توضیح دهید.

۱۴. محاسن و معایب روش مشاهدهٔ کلاسی را توضیح دهید.

۱۵. روش استفاده از نظر دانش‌آموزان و دانشجویان را توضیح دهید.

۱۶. انواع فرمهای نظرسنجی را توضیح دهید.

۱۷. انواع سؤالهای پرسشنامه‌های نظرسنجی را توضیح دهید.

۱۸. محاسن و معایب روش استفاده از نظر دانش‌آموزان و دانشجویان را توضیح دهید.

۱۹. روش استفاده از آزمونهای عملکردی آموزشی برای ارزشیابی از کار معلمان را توضیح دهید.

۲۰. محاسن و معایب روش استفاده از یادگیری دانش‌آموزان و دانشجویان را توضیح دهید.

۲۱. فرم درجه‌بندی شخصی مورد استفاده در روش خودسنجی را توضیح دهید.

۲۲. محاسن و معایب روش خودسنجی را توضیح دهید.

۲۳. محاسن و معایب روش بررسی آثار علمی عضو هیأت علمی را توضیح دهید.

۲۴. برای ارزشیابی از میزان اثربخشی فعالیتهای آموزشی و علمی یک مؤسسهٔ آموزشی یک طرح ارزشیابی معلم تهیه کنید و مراحل اجرا و استفاده از نتایج آن را شرح دهید.

جواب سؤالهای عینی پایان فصلها

	سؤال	جواب درست
فصل ۱	۱	ب
	۲	الف
	۳	الف
	۴	ب
	۵	پ
	۶	ت
	۷	ج
فصل ۲	۱	ث
	۲	ث
	۳	الف
	۴	الف
فصل ۳	۱	الف
	۲	ب
	۳	پ
فصل ۴	۱	ب
	۲	الف
	۳	الف
	۴	ب
	۵	ث
	۶	الف
	۷	ب
	۸	ث
	۹	الف
فصل ۵	۱	الف
	۲	ب
فصل ۶	۱	الف
	۲	ب
	۳	الف
	۴	پ
	۱۰	شناختی
	۱۱	روانی ـ حرکتی
	۱۲	عاطفی
	۱۳	شناختی
	۱۴	شناختی
	۱۵	روانی ـ حرکتی
فصل ۷	۱	الف
	۲	الف
	۳	پ
فصل ۸	۱	ب
	۲	پ
فصل ۹	۱	ب
	۲	ت
	۳	ب
	۴	الف

	سؤال	جواب درست
فصل ۱۰	۱	ب
	۲	ب
	۳	ب
	۴	ث
	۵	الف
فصل ۱۱	۱	ب
	۲	الف
فصل ۱۲	۱	ب
	۲	الف
	۳	ب
	۴	ت
فصل ۱۳	۱	الف
	۲	ت
فصل ۱۴	۱	الف
	۲	ث
	۳	ب
فصل ۱۵	۱	الف
	۲	الف
	۳	ب
فصل ۱۶	۱	ب
	۲	الف
فصل ۱۷	۱	الف
	۲	ب
	۳	ب
فصل ۱۷	۴	پ
	۵	ب
	۶	ب
	۷	الف
	۸	ت
	۹	الف
	۱۰	ب
	۱۱	ت
	۱۲	ب
	۱۳	ت
فصل ۱۸	۱	الف
	۲	ت
	۳	ب
	۴	ب
	۵	ب
	۶	ث
	۷	الف
	۸	الف
	۹	ب
	۱۰	ب
فصل ۱۹	۱	پ
	۲	الف
	۳	ب
	۴	ت
	۵	ت
	۶	پ
فصل ۲۰	۱	ب
	۲	پ
	۳	پ
	۴	الف

	سؤال	جواب درست
فصل ۲۰	۵	پ
	۶	پ
	۷	ب
	۸	ث
	۹	ب
	۱۰	ب
فصل ۲۱	۱	پ
	۲	ب
	۳	الف
	۴	ب

فصل ۲۱	۵	ب
	۶	ث
	۷	پ
	۸	الف
	۹	ب
	۱۰	پ
	۱۱	پ
فصل ۲۲	۱	ت
	۲	ب
	۳	ت
	۴	الف
	۵	پ



اداره کل، سنجش و
10BF0134397
کتابخانه مرکزی

نشر دوران