



سنجش و اندازہ گیری

دکتر حسین زارع دکتر علی مصطفائی حسن امین پور



ASSESSMENT

فهرست مطالب

پیشگفتار	۱
راهنمای مطالعه	۱
هدف‌های کلی	۱
فصل اول. مقدمه‌ای بر سنجش و اندازه‌گیری در علوم تربیتی	۱
هدف کلی	۱
هدف‌های یادگیری	۱
مقدمه	۱
۱-۱ اهمیت و فواید کاربرد روش‌های اندازه‌گیری در علوم تربیتی	۲
۲-۱ تعاریف اصطلاحات و واژه‌ها	۳
۱-۲-۱ اندازه‌گیری	۴
۲-۲-۱ سطوح مختلف اندازه‌گیری	۴
۳-۲-۱ سنجش	۸
۴-۲-۱ ارزیابی	۹
۵-۲-۱ ارزشیابی	۹
۶-۲-۱ ثابت	۹
۷-۲-۱ متغیر	۱۰
۸-۲-۱ آزمون	۱۰
۹-۲-۱ آزمودن	۱۱
۳-۱ اصول کلی سنجش	۱۱
۴-۱ سنجش و فرایند آموزش	۱۳
خلاصه فصل اول	۱۵
خودآزمایی چهارگزینه‌ای فصل اول	۱۵

۱۷	خودآزمایی تشریحی فصل اول
۱۹	فصل دوم. اندازه‌گیری پیامدهای یادگیری
۱۹	هدف کلی
۱۹	هدف‌های یادگیری
۲۰	مقدمه
۲۰	۱-۲ نوشتن هدف‌های آموزشی
۲۳	۲-۲ شناسایی پیامدهای یادگیری
۲۳	۳-۲ شناسایی پیامدهای یادگیری قابل مشاهده و قابل اندازه‌گیری
۲۴	۱-۳-۲ بیان شرایط
۲۵	۲-۳-۲ بیان ملاک‌ها
۲۷	۴-۲ طبقه‌بندی هدف‌های آموزشی
۲۸	۱-۴-۲ طبقه‌بندی تجدیدنظرشده
۳۳	۲-۴-۲ ابعاد دانش اندرسون و کراتول
۳۳	۳-۴-۲ ابعاد فرایند شناختی اندرسون و کراتول
۳۸	۴-۴-۲ طبقه‌بندی مارزانو و کندال
۴۲	۵-۲ نقشه آزمون یا جدول مشخصات
۴۳	۱-۵-۲ بعد محتوا
۴۴	۲-۵-۲ بعد هدف (طبقات)
۴۵	۳-۵-۲ تعداد سؤال‌ها
۴۶	خلاصه فصل دوم
۴۷	خودآزمایی چهارگزینه‌ای فصل دوم
۵۰	خودآزمایی تشریحی فصل دوم
۵۱	فصل سوم. آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم‌ساخته
۵۱	هدف کلی
۵۱	هدف‌های یادگیری
۵۱	مقدمه
۵۱	۱-۳ آزمون‌های استاندارد شده و هدف‌های آن‌ها
۵۲	۲-۳ تعریف آزمون‌های هنجاری و ملاکی
۵۳	۳-۳ استاندارد کردن آزمون‌های هنجاری
۵۴	۴-۳ آزمون‌های ملاکی
۵۵	۵-۳ مقایسه آزمون‌های ملاکی و هنجاری
۵۶	۶-۳ آزمون‌های معلم‌ساخته
۵۸	۷-۳ مقایسه آزمون‌های استاندارد شده و کلاسی
۵۹	۱-۷-۳ مزایای آزمون‌های استاندارد
۶۰	

۶۱	۸-۳ تفاوت‌های آزمون‌های استاندارد با آزمون‌های معلم‌ساخته
۶۲	۹-۳ استفاده از آزمون‌های استاندارد و معلم‌ساخته در تصمیم‌گیری‌های آموزشی
۶۲	خلاصه فصل سوم
۶۳	خودآزمایی چهارگزینه‌ای فصل سوم
۶۵	خودآزمایی تشریحی فصل سوم
۶۷	فصل چهارم. آزمون‌های کوتاه‌پاسخ و عینی
۶۷	هدف کلی
۶۷	هدف‌های یادگیری
۶۷	مقدمه
۶۸	۱-۴ تقسیم‌بندی آزمون‌های عینی
۶۸	۲-۴ چند نکته برای آماده‌سازی سؤال‌ها در طول دوره تحصیلی
۶۹	۳-۴ سؤال‌های صحیح - غلط
۷۱	۱-۳-۴ انواع سؤال‌های صحیح - غلط
۷۴	۲-۳-۴ قواعد تهیه سؤال‌های صحیح - غلط
۷۴	۳-۳-۴ محاسن و معایب آزمون‌های صحیح - غلط
۷۵	۴-۴ سؤال‌های جورکردنی
۷۶	۱-۴-۴ قواعد تهیه سؤال‌های جورکردنی
۷۹	۲-۴-۴ محاسن آزمون‌های جورکردنی
۷۹	۳-۴-۴ معایب آزمون‌های جورکردنی
۷۹	۵-۴ سؤال‌های کامل‌کردنی
۸۰	۶-۴ استفاده از سؤال‌های کوتاه‌پاسخ
۸۱	۱-۶-۴ مزایا و محدودیت‌های سؤال‌های کوتاه‌پاسخ
۸۲	۲-۶-۴ پیشنهادهایی برای نوشتن سؤال‌های کوتاه‌پاسخ
۸۳	۷-۴ قواعد تهیه سؤال‌های عددی
۸۴	۸-۴ آزمون‌های چندگزینه‌ای
۸۶	۱-۸-۴ محاسن آزمون‌های چندگزینه‌ای
۸۷	۲-۸-۴ معایب آزمون‌های چندگزینه‌ای
۸۷	۳-۸-۴ مهم‌ترین قواعد تهیه سؤال‌های چندگزینه‌ای
۹۰	خلاصه فصل چهارم
۹۲	خودآزمایی چهارگزینه‌ای فصل چهارم
۹۵	خودآزمایی تشریحی فصل چهارم
۹۷	فصل پنجم. نوشتن سؤال‌های انشایی
۹۷	هدف کلی
۹۷	هدف‌های یادگیری

۹۷	مقدمه
۹۸	۱-۵ آزمون‌های انشایی
۹۸	۲-۵ انواع آزمون‌های انشایی
۹۹	۳-۵ آزمون‌های انشایی گسترده پاسخ
۱۰۰	۴-۵ آزمون‌های انشایی محدود پاسخ
۱۰۲	۵-۵ مقایسه پیامدهای یادگیری اندازه گیری شده آزمون‌های محدود پاسخ
۱۰۳	۶-۵ قواعد تهیه سؤال‌های انشایی
۱۰۳	۷-۵ قواعد نمره گذاری سؤال‌های انشایی
۱۰۴	۱-۷-۵ روش نمره گذاری تحلیلی
۱۰۴	۲-۷-۵ روش نمره گذاری کلی
۱۰۴	۸-۵ قواعد تصحیح سؤال‌های انشایی
۱۰۸	۹-۵ محاسن و معایب آزمون‌های انشایی
۱۱۰	خلاصه فصل پنجم
۱۱۲	خودآزمایی چهارگزینه‌ای فصل پنجم
۱۱۴	خودآزمایی تشریحی فصل پنجم
۱۱۷	فصل ششم. سنجش کارپوشه
۱۱۷	هدف کلی
۱۱۷	هدف‌های یادگیری
۱۱۷	مقدمه
۱۱۸	۱-۶ ارزشیابی توصیفی
۱۱۸	۲-۶ سنجش کارپوشه
۱۱۹	۳-۶ منطق کارپوشه
۱۲۰	۳-۶ تأمین روایی کارپوشه
۱۲۱	۴-۶ گام‌های ساخت کارپوشه روا
۱۲۶	۵-۶ مزایای کارپوشه
۱۲۷	خلاصه فصل ششم
۱۲۸	خودآزمایی چهارگزینه‌ای فصل ششم
۱۲۹	خودآزمایی تشریحی فصل ششم
۱۳۱	فصل هفتم. هنجار آزمون
۱۳۱	هدف کلی
۱۳۱	هدف‌های یادگیری
۱۳۱	مقدمه
۱۳۱	۱-۷ تفسیر نمره‌ها و هنجارهای آزمون
۱۳۲	۲-۷ روش‌های تفسیر نمره‌های خام آزمون
۱۳۳	

۱۳۳	۳-۷ تفسیرهای ملاکی و هنجاری
۱۳۳	۱-۳-۷ آزمون‌های ملاکی
۱۳۳	۲-۳-۷ نمره‌های هنجاری
۱۳۶	۴-۷ هنجارها
۱۳۶	۱-۴-۷ هنجارهای سنی و کلاسی
۱۴۱	۲-۴-۷ رتبه درصدی
۱۴۱	۳-۴-۷ محاسبه رتبه درصدی
۱۴۳	۴-۴-۷ نمره‌های استاندارد
۱۴۴	۵-۷ توزیع نرمال
۱۴۴	۶-۷ انواع نمره‌های استاندارد
۱۴۵	۱-۶-۷ نمره Z
۱۴۶	۲-۶-۷ نمره T
۱۴۸	۳-۶-۷ نمره‌های ۹ بخشی
۱۴۹	۷-۷ نیم‌رخ
۱۵۱	خلاصه فصل هفتم
۱۵۳	خودآزمایی چهارگزینه‌ای فصل هفتم
۱۵۵	خودآزمایی تشریحی فصل هفتم
۱۵۷	فصل هشتم. اعتبار آزمون
۱۵۷	هدف کلی
۱۵۷	هدف‌های یادگیری
۱۵۷	مقدمه
۱۵۸	۱-۸ تعریف اعتبار
۱۵۹	۲-۸ روش‌های تعیین اعتبار
۱۵۹	۱-۲-۸ روش بازآزمایی
۱۶۱	۲-۲-۸ روش فرم‌های موازی و هم‌تا
۱۶۲	۳-۲-۸ روش‌های برآورد همسانی درونی
۱۶۳	۴-۲-۸ روش دونیمه‌کردن
۱۶۴	۵-۲-۸ روش آلفای کرونباخ
۱۶۵	۶-۲-۸ روش کودر-ریچاردسون
۱۶۷	۳-۸ خطای معیار اندازه‌گیری
۱۶۹	۴-۸ فاصله اطمینان
۱۷۱	۵-۸ تفسیر خطای معیار اندازه‌گیری
۱۷۱	۶-۸ عوامل مؤثر بر افزایش ضریب اعتبار آزمون
۱۷۲	خلاصه فصل هشتم
۱۷۳	خودآزمایی چهارگزینه‌ای فصل هشتم

۱۷۵	خودآزمایی تشریحی فصل هشتم
۱۷۷	فصل نهم- روایی آزمون
۱۷۷	هدف کلی
۱۷۷	هدف‌های یادگیری
۱۷۸	مقدمه
۱۷۹	۱-۹ ماهیت روایی
۱۸۰	۲-۹ تعریف روایی
۱۸۱	۳-۹ انواع روایی
۱۸۲	۱-۳-۹ روایی محتوایی
۱۸۳	۲-۳-۹ روایی صوری
۱۸۴	۳-۳-۹ روایی منطقی
۱۸۴	۴-۳-۹ روایی ملاکی
۱۸۵	۵-۳-۹ روایی پیش‌بینی
۱۸۶	۶-۳-۹ روایی هم‌زمان
۱۸۷	۴-۹ ضرایب روایی ملاکی
۱۸۸	۵-۹ روایی سازه
۱۹۰	۶-۹ روش‌های تعیین روایی سازه
۱۹۰	۱-۶-۹ شواهد همگرا و واگرا
۱۹۱	۲-۶-۹ تفکیک سنی
۱۹۱	۳-۶-۹ تحلیل عاملی
۱۹۱	۳-۶-۹ همسانی درونی
۱۹۱	خلاصه فصل نهم
۱۹۲	خودآزمایی چهارگزینه‌ای فصل نهم
۱۹۳	خودآزمایی تشریحی فصل نهم
۱۹۵	پاسخنامه خودآزمایی‌های چهارگزینه‌ای
۱۹۷	منابع
۱۹۹	

فصل اول

مقدمه‌ای بر سنجش و اندازه‌گیری در علوم تربیتی

هدف کلی

بررسی ساینه روش‌های اندازه‌گیری و تعاریف اصطلاحات مرتبط

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. اهمیت و فواید کاربرد روش‌های اندازه‌گیری در علوم تربیتی را توضیح دهد.
۲. مفاهیم سنجش، اندازه‌گیری، ارزشیابی، متغیر و ثابت را تعریف کند.
۳. آزمون و آزمودن را تعریف کند.
۴. سطوح مختلف اندازه‌گیری را نام‌برد و برای هر کدام مثالی بیاورد.

مقدمه

بدون تردید، سنجش‌های روانی و تربیتی یکی از مهم‌ترین کمک‌های علوم رفتاری به جامعه است که در زمینه‌های دولتی، صنعتی و تربیتی، تغییرات اساسی و معناداری را به وجود آورده است و به جرئت می‌توان گفت که امر استخدام و آموزش را دگرگون ساخته است. کاربرد آزمون‌های تربیتی در چند سال گذشته، هم در مدارس ابتدایی و متوسطه، هم در بخش‌های دولتی و خصوصی و هم در سطح کلی‌تر، به واسطه برگزاری امتحانات همگانی و سراسری، به شکل سریعی افزایش یافته و در زمینه‌های گوناگون تربیتی برای تصمیم‌گیری‌های مختلف از لحاظ امتحانات دبستانی و دبیرستانی، ارتقا از یک کلاس به کلاس دیگر، طبقه‌بندی شاگردان،

ارزشیابی کارآمدی مدارس، طبقه‌بندی و ارزشیابی معلمان و غیره، به‌صورت چشمگیری عمومیت پیدا کرده‌است. افزون‌براین، سنجش‌های روانی در آموزش ویژه، برای کمک به‌روشن‌ساختن انواع یادگیری، مبانی و اندازه مشکلات یادگیری فرد مورد آزمون یا مسائل سازگاری تحصیلی، در برنامه‌ریزی‌های آموزشی و برنامه تحصیلی، جایگزینی در مدرسه یا کلاس‌های ویژه به‌کارمی‌رفته و سرانجام اجرای آزمون‌های گروهی، سیاهه‌های شخصیتی، رغبت‌سنج‌ها، مشاهدات رفتاری، فتون فراقکن، کاربرد آزمون‌ها در همه اقدامات روان‌شناسی و مشاوره، نشان‌دهنده اهمیت فوق‌العاده سنجش‌های روانی است (ثرندایک^۱، ۱۹۸۲؛ ترجمه هومن، ۱۳۷۵).

جنبش اندازه‌گیری‌های روانی، در اواخر قرن نوزدهم به‌وسیله گالتون^۲ پایه‌گذاری شد و سپس به‌وسیله دیگر محققان توسعه یافت. هرچند استفاده از آزمون‌ها، تنها ابزار تشخیص نیست، ولی با توجه به محدودیت‌های سایر روش‌ها و محاسن آزمون‌ها، استفاده از این تکنیک، کاملاً ضرورت دارد (زارع و امین‌پور، ۱۳۹۰).

یادآوری. یکی از کمک‌های روان‌شناسی به‌جامعه، سنجش مفاهیم و متغیرهای روان‌شناختی است.

۱-۱ اهمیت و فواید کاربرد روش‌های اندازه‌گیری در علوم تربیتی

اندازه‌گیری ویژگی‌های روانی و تربیتی، یکی از فعالیت‌های مهم روان‌شناسان و مشاوران است و کمتر فردی را می‌توان یافت که در طول عمر خود، مورداندازه‌گیری روان‌شناختی قرار نگرفته‌باشد؛ بنابراین ساختن ابزارهای سنجش و اندازه‌گیری ویژگی‌های روانی افراد و داشتن هنجارهای متناسب با شرایط فرهنگی و اجتماعی جامعه، امری اجتناب‌ناپذیر است (معنوی‌پور، ۱۳۸۹). برای مطالعه مؤثر در هر حوزه رفتاری بشر، شخص باید اصول اساسی سنجش و اندازه‌گیری را بفهمد. در جامعه پیچیده امروز، ارتباط بین اصول، کاربردها و موضوعات روان‌آزمایی، فراتر از دامنه روان‌شناسی، گسترش یافته‌است. حتی وکلا، پزشکان، مددکاران اجتماعی، مدیران یازرگانی، مربیان و بسیاری از صاحبان حرفه‌های دیگر، اغلب باید گزارش‌های مبتنی بر

مقدمه‌ای بر سنجش و اندازه‌گیری در علوم تربیتی ۳

چنین آزمون‌هایی را مورد بحث قرار دهند (کاپلان^۱ و ساکوزو^۲، ۲۰۰۴؛ ترجمه دلاور، در تاج و فرخی، ۱۳۸۸).

مشاوران تحصیلی و روانی، آزمون‌ها را در موارد زیر مورد استفاده قرار می‌دهند:

الف) برای به دست آوردن اطلاعات دقیق درباره توانایی‌ها، علایق و ویژگی‌های شخصیتی، به منظور کمک به خودشناسی مراجعان

ب) برای پیش‌بینی موفقیت‌های آینده در امور تحصیلی و شغلی. مشاوران تحصیلی، از آزمون‌های روانی با هدف کمک به تصمیم‌گیری مراجعان خود برای برنامه‌ریزی آینده تحصیلی و شغلی استفاده می‌کنند.

ج) برای تشخیص مشکلات درمان‌جویان

د) برای ارزشیابی فرایند راهنمایی و روش‌های درمانی که در مورد مراجعان خود به کار بسته‌اند (حسن‌زاده و مداح، ۱۳۸۹).

نکته. مشاوران تحصیلی و روان‌شناسان، آزمون‌ها را برای رسیدن به هدف‌های گزینش (انتخاب)^۳، جایگزینی و طبقه‌بندی^۴، تشخیص و ترمیم^۵، بازخورد^۶، راهنمایی و انگیزش یادگیری^۷، ارتقای برنامه آموزشی^۸ و تولید و تدوین نظریه، مورد استفاده قرار می‌دهند (دلاور و زهراکار، ۱۳۸۹).

۱-۲ تعاریف اصطلاحات و واژه‌ها

در این کتاب، با اصطلاحات و واژه‌هایی مواجه خواهید شد که پیوسته تکرار شده‌اند و شایسته است به اختصار معرفی شوند. معمولاً سه اصطلاح اندازه‌گیری، سنجش و آزمون، با هم اشتباه گرفته می‌شوند، زیرا همگی در یک فرایند واحد دخیل‌اند.

1. Kaplan
2. Saccuzzo
3. Selection

۱-۲-۱ اندازه‌گیری

اندازه‌گیری^۱ عبارت است از اختصاص دادن اعداد و ارقام به متغیر مورد مطالعه (برای مثال، محاسبه پاسخ‌های صحیح یا امتیاز دادن به جنبه‌های خاصی از یک انشا) براساس اصول و قواعد معین (میلر^۲، لین^۳ و گراندلند^۴، ۲۰۰۹). هر زمان بتوانیم ارزش متغیر خاصی را در آزمودنی خاصی یا در زمان معینی معلوم کنیم، گوییم آن متغیر را اندازه گرفته‌ایم (هومن، ۱۳۹۰). اندازه‌گیری، دربرگیرنده قواعدی برای اختصاص اعداد به چیزها، افراد یا رویدادها به منظور کمی‌ساختن ویژگی‌های آن‌هاست (هومن، ۱۳۸۶). اندازه‌گیری، به دو صورت وجود دارد که عبارت‌اند از: اندازه‌گیری مستقیم و اندازه‌گیری غیرمستقیم. اندازه‌گیری مستقیم مانند اندازه‌گیری قد با متر (وقتی قد یک دانشجو، با استفاده از متر اندازه‌گیری شود)، نشان دادن زمان سپری شده با ساعت یا ثبت دمای اتاق با استفاده از دماسنج. اندازه‌گیری غیرمستقیم مانند تعیین علاقه آموزشی که با به کار بردن آزمون‌های پیشرفت تحصیلی، پرسش‌نامه‌ها و مانند آن سنجیده می‌شود و اندازه‌گیری میزان اضطراب امتحانی دانشجویان (پوستلتوایت^۵، ۲۰۰۵).

۱-۲-۲ سطوح مختلف اندازه‌گیری

پژوهشگر پیش تصمیم‌گیری درباره چگونگی تحلیل داده‌ها یا به عبارتی قبل از اینکه متغیر وابسته، مورد تجزیه و تحلیل آماری قرار گیرد، باید بتواند نوع داده‌هایش (نوع اندازه‌گیری) را معین کند (هومن، ۱۳۷۷). این شیوه‌های اندازه‌گیری، عبارت‌اند از: اسمی، ترتیبی، فاصله‌ای و نسبی (استیونس^۶، ۱۹۴۶). این طبقه‌بندی، از ساده به پیچیده و به صورت سلسله‌مراتبی تنظیم شده است؛ یعنی هر طبقه، افزون‌بر ویژگی‌های مخصوص خود، دارای همه ویژگی‌های طبقات قبل نیز است. ویژگی‌های مقیاس‌های چهارگانه اندازه‌گیری، به فرض‌هایی که درباره این مقیاس‌ها قائل می‌شوند و همچنین به اعمال ریاضی و آماری که می‌توان درباره آن‌ها انجام داد، مربوط است. در ادامه، ویژگی‌های این مقیاس‌ها مورد بحث قرار می‌گیرند.

مقیاس اسمی. مقیاس اسمی^۱، مطلبی درباره کم یا زیاد بودن بیان نمی‌کند و این ویژگی مهم را دارد که فقط دارای طبقه‌های مشخص و متمایزی است که جنبه کیفی دارند. در این مقیاس، ویژگی‌های مورد اندازه‌گیری، در طبقات مختلف قرار می‌گیرند. اعداد به‌کاررفته، ارزش کمی ندارند و فقط برای مشخص کردن یا نامیدن متغیرها به‌کار می‌روند. بین طبقه‌ها هیچ رابطه تجربی که منطبق با روابط ریاضی اعداد مربوط به آن‌ها باشد، وجود ندارد (امین‌پور، ۱۳۹۰). در مقیاس اسمی، اجازه داده می‌شود که متغیرها به‌طور مشابه یا متفاوت ساخته شوند؛ بنابراین می‌توان گفت که افراد، به‌شیوه مشابه یا متفاوت، از دیگر افراد طبقه‌بندی می‌شوند (پوستلوايت، ۲۰۰۵). برای مثال، گروه خونی افراد. اگر در یک کلاس ۲۵ نفری، ۴ نفر دارای گروه خونی A، ۶ نفر گروه خونی B، ۱۰ نفر گروه خونی O و ۵ نفر دارای گروه خونی AB باشند، می‌توان به‌هریک از گروه‌های خونی، کدی را اختصاص داد. جنسیت نیز دارای دو ارزش زن و مرد است. می‌توان برای زن، کد ۱ و برای مرد، کد ۲ را در نظر گرفت. کدهای به‌کاررفته در پرونده‌های مراجعان مرکز مشاوره نیز از نوع مقیاس اسمی هستند.

نکته ۱. مقیاس اسمی به‌متغیرهایی اشاره دارد که در گروه‌های جداگانه طبقه‌بندی شده‌اند. آزمودنی‌ها بر پایه برخی ویژگی‌های خاص در گروه‌ها گروه‌بندی یا رده‌بندی می‌شوند؛ مانند جنسیت، دین، رنگ مو و منطقه جغرافیایی محل سکونت (یون، ۲۰۰۶).

نکته ۲. طبقاتی که به‌وسیله پژوهشگر در مقیاس اسمی به‌کار می‌روند، باید کاملاً قابل تفکیک از یکدیگر باشند، یعنی نتوان یک فرد را در دو طبقه یا بیشتر قرار داد.

نکته ۳. در مقیاس اسمی، هیچ نوع ترتیب، فاصله، شدت و ضعفی نشان داده نمی‌شود (زارع، طالبی و صیف، ۱۳۸۹).

مقیاس ترتیبی^۲ (رتبه‌ای). متغیرهایی که درجه‌بندی رتبه‌ای داشته باشند، در مقیاس ترتیبی بیان می‌شوند. مشاهده‌ها با یکدیگر مقایسه می‌شوند و به‌ترتیب معینی برای مثال از بهترین تا بدترین یا از بزرگ‌ترین تا کوچک‌ترین، به آن‌ها رتبه‌ای اختصاص داده می‌شود. در اینجا وضعیت نسبی نمره‌ها، به‌وسیله پژوهشگر مشخص می‌شود. مقیاس ترتیبی، مقیاس داده‌ها را

1. Nominal Scale
2. Ordinal Scale

براساس معیار مشخص نشان می‌دهد. در این مقیاس، پژوهشگر با داده‌هایی سروکار دارد که دارای مقیاس با درجه‌بندی رتبه‌ای است و می‌تواند وضعیت نسبی افراد را یا یکدیگر مقایسه کند. اعداد در مقیاس ترتیبی فقط سلسله‌مراتب چیزها، افراد یا رویدادها را از لحاظ ویژگی معینی در یک گروه نشان می‌دهند و بدون بیان کردن مطلبی درباره فاصله بین افراد، وضع نسبی آن‌ها را از لحاظ یک خصیصه، مشخص می‌سازند.

در این مقیاس سعی می‌شود که همه مشاهده‌ها از یکدیگر تفکیک شوند و یک رتبه، بالاتر از دیگری قرارگیرد، اما زمانی که دو متغیر، از ویژگی یکسانی برخوردار باشند، به آن‌ها رتبه یکسانی داده می‌شود. اعداد در مقیاس ترتیبی فقط وضعیت نسبی نمره‌ها را نسبت به همدیگر تعیین می‌کنند. برای مثال رتبه‌ای که در مسابقات یا ورزش‌ها به ورزشکاران یا تیم‌ها داده می‌شود، دارای مقیاس رتبه‌ای است (امین‌پور، ۱۳۹۰). همچنین صورت اسامی ده نفر از برترین‌های نیمسال تحصیلی یا برتری مهارت پرویز نسبت به موسی در درس سنجش و اندازه‌گیری، در این مقیاس قرار می‌گیرند. وقتی که شما قنون مورد استفاده در مشاوره را از لحاظ کارایی، به صورت کارآمدترین تا ضعیف‌ترین دسته‌بندی می‌کنید، از مقیاس رتبه‌ای استفاده کرده‌اید. در این مقیاس، اختلاف رتبه‌ها کمی نیست و بهترین آماره توصیفی برای خلاصه‌سازی داده‌هایی که دارای مقیاس رتبه‌ای هستند، میانه است (زارع، طالبی و صیف، ۱۳۸۹).

نکته ۱. داده‌های ترتیبی به متغیرهایی اشاره دارد که به طور مرتب رتبه‌بندی شده‌اند. داده‌های ترتیبی در بعضی از متغیرها، از بیشترین به کمترین دسته‌بندی می‌شوند. در اندازه‌ی ترتیبی، داده‌ها براساس مکان قرارگرفتن در مجموعه‌ای از داده‌ها به صورت اولین، دومین و سومین مکان، گزارش می‌شوند.

نکته ۲. در اندازه‌ی ترتیبی این سؤال مطرح نیست که آیا فاصله بین اولین و دومین، بیشتر یا کمتر از فاصله بین دومین و سومین است (یون، ۲۰۰۶).

یادآوری. در مقیاس ترتیبی یا رتبه‌ای، افراد از لحاظ خصوصیات یا ویژگی‌ها از بالاترین تا پایین‌ترین یا به صورت بهترین تا بدترین نمره می‌گیرند و از این لحاظ، نظم و ترتیب منطقی وجود دارد.

مقیاس فاصله‌ای^۱. به‌ارزش‌های موجود در این مقیاس، نمره‌های عددی تعلق می‌گیرد. می‌توان نمره‌ها را از هم کم کرد و تفاوت بین آن‌ها را به‌دست آورد. در این مقیاس، واحدهای اندازه‌گیری در سراسر طول مقیاس، یکسان‌اند. نکته مهمی که در این مقیاس وجود دارد این است که متغیرهای آن، صفر واقعی ندارند (پوستلتوایت، ۲۰۰۵). برای مثال مقیاس نگرش‌سنجی است که بیست سؤال دارد. هر سؤال می‌تواند ارزش ۱، ۲، ۳ یا ۴ داشته باشد. بدین معنی که یک آزمودنی می‌تواند نمره‌ای بین ۲۰ تا ۸۰ بگیرد. نمره‌ها در این مقیاس، براساس قاعده در فاصله بین ۲۰ تا ۸۰ قرار می‌گیرند. نقطه صفر در مقیاس فاصله‌ای، قراردادی است و بدین معنی نیست که مشخصه اندازه‌گیری در آن وجود ندارد. وقتی دما صفر درجه فارنهایت خوانده می‌شود، بدین معنی نیست که گرما وجود ندارد (یون^۲، ۲۰۰۶).

نکته ۱. در این مقیاس، فاصله بین اندازه نقاط داده‌ها باید با درجه‌بندی ثابت و برابر باشند. نکته مهم در این مقیاس فاصله‌ای، دقیق بودن آن است. داده‌های موجود در این مقیاس، با درجه‌بندی یا فاصله ثابت‌اند. یکی از مثال‌هایی که اغلب برای مقیاس فاصله‌ای مطرح می‌شود، درجه حرارت (دما) است.

نکته ۲. در مقیاس فاصله‌ای، کمیت‌ها برحسب یک مینا یا صفر قراردادی نسبت به یکدیگر روی یک پیوستار قرار می‌گیرند، نه برحسب صفر مطلق (زارع، طالبی و صیف، ۱۳۸۹).

یادآوری. مقیاس فاصله‌ای، دارای صفر فرضی و قراردادی است.

مقیاس نسبی^۳ (نسبتی). این مقیاس، همه ویژگی‌های مقیاس فاصله‌ای را دارد و دارای صفر مطلق^۴ است و بیشتر در علوم شیمی و فیزیک کاربرد دارد. استفاده از این مقیاس در روان‌شناسی و علوم تربیتی کمتر است. در این مقیاس می‌توان تمام اعمال آماری و ریاضی را انجام داد. برای مثال می‌توان گفت که یک فرد ۶۰ ساله، دو برابر یک فرد ۳۰ ساله سن دارد؛ یا وزن یک فرد ۴۵ کیلوگرمی، نصف وزن کسی است که ۹۰ کیلوگرم

1. Interval Scale

2. Youn, R

3. Ratio Scale

4. Absolute Zero

وزن دارد. در بسیاری از مطالعات برای انتخاب آزمون‌های آماری مناسب، داده‌های مقیاس فاصله‌ای و نسبی همانند هم به کار برده می‌شوند. همه اظهارنظرهایی که دربارهٔ سه مقیاس قبلی وجود داشت، در این مقیاس نیز وجود دارد.

یادآوری ۱. صفر موجود در مقیاس نسبی، بدین معنی است که شیء وزنی ندارد.

یادآوری ۲. مقیاس نسبی، دارای صفر مطلق است و می‌توان تمام فرایندهای آماری و ریاضی را در آن انجام داد.

۱-۲-۳ سنجش

سنجش^۱، شامل فعالیت‌هایی نظیر درجه‌بندی (رسمی^۲ و غیررسمی^۳)، آزمون‌گرفتن، مشاهده، درجه‌بندی عملکرد، طرح‌ها، آزمون‌های مداد-کاغذی، گواهی‌دادن و مانند آن است. [بسیاری از سیستم‌های آموزشی، پیشرفت دانش‌آموز را به صورت تمره^۴، کد حروف^۵ یا به صورت توصیفی نظیر خوب، رضایت‌بخش یا نیاز به پیشرفت ثبت می‌کنند (پوستلتوایت، ۲۰۰۵). به عبارتی سنجش، به فرایند جمع‌آوری، ترکیب و تفسیر اطلاعات برای رسیدن به تصمیم‌گیری دربارهٔ کلاس درس گفته می‌شود و شامل جمع‌آوری اطلاعات دربارهٔ دانش‌آموزان، آموزش و راهنمایی و وضعیت کلاس درس است (آیراسیان^۶، ۲۰۰۲). به طور کلی سنجش، اصطلاحی جامع‌تر و کلی‌تر از اندازه‌گیری یا آزمودن است. اصطلاح اندازه‌گیری، محدود به توصیف‌های کمی دانش‌آموزان است؛ یعنی نتایج اندازه‌گیری، همیشه به صورت عددی بیان می‌شود (برای مثال حمید، ۳۵ مسئله از ۴۰ مسئله را به درستی حل کرد) و شامل توصیف‌های کیفی نمی‌شود (برای مثال کار علی، منظم و تمیز بود) و قضاوت ارزشی دربارهٔ نتیجه به دست آمده ندارد. از طرف دیگر، سنجش ممکن است توصیف‌های کمی (اندازه‌گیری) و کیفی (غیرقابل اندازه‌گیری) دانش‌آموزان را شامل شود. همچنین سنجش همیشه شامل قضاوت‌های ارزشی مربوط به مطلوبیت نتایج می‌شود.

۴-۲-۱ ارزیابی

ارزیابی^۱، یکی از فرایندهای یادگیری است و برای تعیین ارزش یک دوره یا برنامه تحصیلی و مشاوره، مورد استفاده قرار می‌گیرد (دیتل^۲، هرمان^۳ و نوث^۴، ۱۹۹۱). ارزیابی به منظور تشخیص، طبقه‌بندی، جایگزین ساختن، نمره دادن و درجه‌بندی و به وسیله یکی از نظام‌های اندازه‌گیری صورت می‌گیرد (هومن، ۱۳۷۵). ارزیابی می‌تواند به صورت ارزیابی ساده به عنوان قضاوت ذهنی معلم یا مشاور بر اساس مشاهداتش از عملکرد دانش‌آموز و مراجع یا به عنوان ارزیابی پیچیده مانند اجرای یک آزمون استاندارد به مدت زمان ۵ ساعت باشد. ارزیابی می‌تواند در تصمیم‌گیری، پیشرفت دوره، نیازهای آموزشی و برنامه‌های درسی تأثیرگذار دارد.

۵-۲-۱ ارزشیابی

ارزشیابی^۵ در اصطلاح مستلزم وزن کردن ارزش یا اعتبار چیزی و جست‌وجوی آن به طور عینی و مقایسه با سایر برنامه‌ها^۶، برنامه تحصیلی^۷ یا برنامه سازمانی^۸ است (پوستلوايت، ۲۰۰۵). در ارزشیابی، نتایج حاصل از اندازه‌گیری و سنجش، بر اساس معیاری معین یا هدف‌های از پیش تعیین شده مقایسه می‌شود و مورد قضاوت و داوری قرار می‌گیرد.

۶-۲-۱ ثابت

ثابت^۹، ویژگی یا صفتی است که در همه اعضای جامعه (افراد، رویدادها و پدیده‌ها) یکسان است. برای مثال، آب در دمای ۳۲ درجه فارنهایت (از سطح دریا) و صفر درجه سانتی‌گراد یخ می‌بندد. تعداد ثانیه‌های یک دقیقه، به طور ثابت ۶۰ است (یون، ۲۰۰۶). پدیده‌های ثابت، قابل مطالعه نیستند. آنچه یک پدیده را قابل مطالعه می‌کند، امکان و احتمال تغییر است (امین‌پور، ۱۳۹۴).

1. Apprasial
2. Dietel
3. Herman
4. Knuth
5. Evaluation
6. Other Programs

۱-۲-۷ متغیر

متغیر^۱، صفت یا خصیصه‌ای است که از یک مقدار به مقدار دیگر، از یک مشاهده به مشاهده دیگر یا از یک مورد به مورد دیگر، مقادیر و ارزش‌های مختلفی را اختیار می‌کند؛ به عبارتی ویژگی‌های مورد مشاهده و اندازه‌گیری شده به وسیله پژوهشگران، متغیر نامیده می‌شود (رویسون^۲، ۲۰۰۸). جنسیت، یک متغیر است یا دو ارزش زن و مرد. ارزش‌هایی که در تجزیه و تحلیل آماری به آن‌ها داده می‌شود، ممکن است ۱ یا ۲ باشد (پوستلنویت، ۲۰۰۵). در صورت وزن کردن هریک از اعضای یک کلاس، به تعداد اعضا ارزش متفاوت (وزن) از یک آزمودنی به آزمودنی دیگر خواهید داشت. مثال‌های دیگری از متغیر عبارت‌اند از: رنگ چشم، بهره هوشی، سطح تحصیلات، سطح اضطراب، رضایت زناشویی و مانند آن (یون، ۲۰۰۶).

تکته. ویژگی‌های متمایز صفت جامعه آماری، متغیر نامیده می‌شود (شریفی و نجفی زنده، ۱۳۸۴).

۱-۲-۸ آزمون

ارائه تعریفی از آزمون^۳ که مورد توافق همه باشد، مشکل است. در اینجا به چند تعریف از آزمون اشاره می‌شود. آزمون، هرگونه روش یا وسیله عینی بوده که برای اندازه‌گیری نمونه‌ای از رفتار فرد، طرح شده است (ثرندایک، ۱۹۸۲؛ ترجمه هومن، ۱۳۷۵). آزمون، وسیله‌ای است عینی و استاندارد شده که برای اندازه‌گیری نمونه‌ای از حالات و رفتار معین فرد یا افرادی خاص به کار می‌رود (آناستازی^۴، ۱۹۸۸). آزمون، ابزار اندازه‌گیری عینی معیار شده‌ای است که برای ارزیابی یک ویژگی یا مجموعه‌ای از خصوصیات مانند هوش، شخصیت، خلاقیت، استعداد یا علایق به کار می‌رود (برونو، ۱۹۸۶؛ ترجمه یاسایی و طاهری، ۱۳۸۴). آزمون، وسیله اندازه‌گیری یا روشی است که برای کمی کردن رفتار یا کمک به درک و پیش‌بینی رفتار، مورد استفاده قرار می‌گیرد (کاپلان و ساکوزو، ۲۰۰۴؛ ترجمه دلاور و همکاران، ۱۳۸۸، ص ۷). می‌توان گفت که متداول‌ترین وسیله اندازه‌گیری ویژگی‌ها یا صفات روانی، آزمون‌ها هستند (توزنده جانی، ابوالقاسمی و اکبری، ۱۳۸۷).

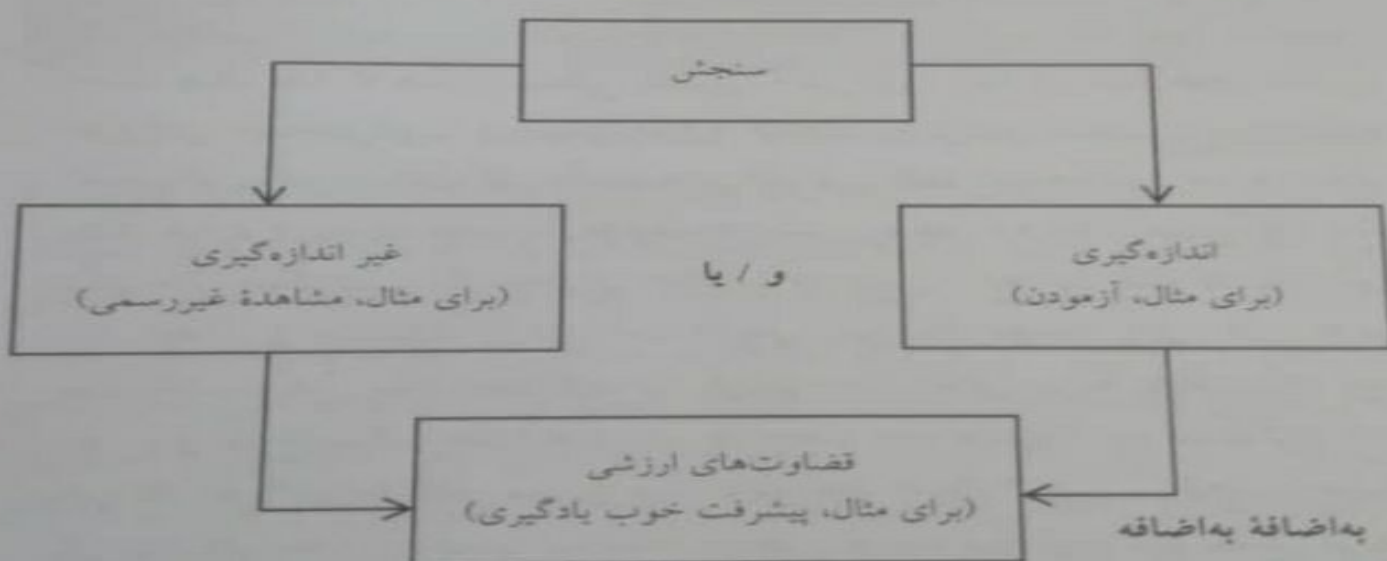
۹-۲-۹ آزمودن

فرایند اجرای آزمون را آزمودن^۱ می‌گویند؛ یا وقتی که برای اندازه‌گیری یکی از ویژگی‌های روانی - تربیتی یک فرد یا گروهی از افراد، از آزمون استفاده می‌شود، به این فعالیت، آزمودن می‌گویند (سیف، ۱۳۹۶).

۱-۳ اصول کلی سنجش

- سنجش، فرایندی یکپارچه برای تعیین ماهیت و گستره یادگیری و رشد دانش‌آموز است. هنگامی که اصول زیر، مد نظر قرارگیرد، این فرایند مؤثرتر خواهد بود:
- آنچه را که سنجیده می‌شود و در فرایند سنجش اولویت دارد، به روشنی مشخص کنید. اثربخشی سنجش، به همان اندازه که به توصیف دقیق آنچه سنجیده می‌شود مربوط است، به کیفیت‌های فنی سنجش روش‌های مورد استفاده نیز بستگی دارد؛ بنابراین تعیین ویژگی‌هایی که باید سنجیده شود، مقدم بر انتخاب یا توسعه روش‌های اندازه‌گیری است. بدان معنا که هنگام سنجش یادگیری دانش‌آموز، ابتدا باید هدف‌های یادگیری به روشنی مشخص شوند و سپس درباره انتخاب روش‌های سنجشی مورد استفاده، تصمیم‌گیری شود. بیانات کلی یا هدف‌های آموزشی، نقطه شروع مناسبی هستند، اما در بیشتر موارد لازم است معلم به آن‌ها وضوح بیشتری بدهد تا فرایند سنجش مؤثر باشد. برای مثال ممکن است یک محتوای استاندارد در درس تاریخ این باشد که دانش‌آموز باید «عقاید و مستندات را در بافت تاریخی خودشان بفهمد». برای چنین هدفی ممکن است چندین سؤال چهارگزینه‌ای، کوتاه پاسخ یا انشایی، مرتبط به نظر برسند. معلم باید برای طرح سؤال، به سؤال‌های زیر پاسخ دهد: چه عقایدی؟ چه مستنداتی؟ کدام بافت تاریخی؟ شواهد کافی مبنی بر درک دانش‌آموز کدام‌اند؟ چنین سؤال‌هایی به وسیله یک بیان کلی جواب داده نمی‌شوند، اما به منظور توسعه سنجش‌ها باید پاسخ سؤال‌ها به صورت آشکار یا تلویحی باشد.
 - انتخاب یک روش سنجش باید براساس ارتباطش با ویژگی‌ها یا عملکرد مورد اندازه‌گیری باشد. هر سؤالی، برای اندازه‌گیری تمام هدف‌های آموزشی مناسب نیست. گاهی سؤال چهارگزینه‌ای مناسب است و گاهی نیز سؤال کوتاه پاسخ، انشایی یا یک آزمون عملکردی. انتخاب این روش‌ها باید متناسب با هدف مورد اندازه‌گیری باشد.

- سنجش جامع، مستلزم بهره‌گیری از روش‌های متنوع است. هیچ نوع ابزار یا روش واحدی برای سنجش مجموعه وسیعی از پیامدهای یادگیری یا رشدی که در برنامه تحصیلی مورد تأکید است، وجود ندارد. آزمون‌های چهارگزینه‌ای و کوتاه‌پاسخ، برای اندازه‌گیری پیامدهای به یاد آوردن، فهمیدن و به کار بستن مفیدند، ولی آزمون‌های انشایی و دیگر طرح‌های مکتوب، برای سنجش توانایی سازمان‌دهی و بیان عقاید و فنون عملکردی نیز بیشتر برای سنجش مهارت‌ها مناسب‌اند.
- استفاده مناسب از روش‌های سنجش، مستلزم آگاهی از محدودیت‌های آن‌هاست. روش‌های سنجش، دامنه‌ای از ابزارهای اندازه‌گیری کاملاً توسعه یافته (برای مثال، آزمون‌های استعداد و پیشرفت هنجار شده) تا ابزارهای سنجش ناپخته‌تر (فنون مشاهده‌ای و خودگزارشی) را شامل می‌شود. حتی بهترین ابزارهای اندازه‌گیری آموزشی و روان‌شناختی نیز محدودیت‌هایی دارند.



شکل ۱-۱. فرایند سنجش

- سنجش وسیله است، به خودی خود هدف نیست. استفاده از روش‌های سنجش به عنوان یک ابزار تشخیصی، سودمند واقع می‌شود، اما جمع‌آوری کورکورانه داده‌ها و سپس بایگانی آن‌ها، ضایع کردن وقت و تلاش است. ضایع کردن وقت و تلاش دانش‌آموزان معلمان و دیگر کاربران اطلاعات، تأثیرات منفی روی سنجش بعدی دارد.

یادآوری ۱. اندازه‌گیری عبارت است از اختصاص دادن اعداد و ارقام به متغیر مورد مطالعه بر اساس اصول و قواعد معین.

یادآوری ۲. سنجش، به فرایند جمع‌آوری، ترکیب و تفسیر اطلاعات برای رسیدن به تصمیم‌گیری درباره کلاس درس گفته می‌شود.

یادآوری ۳. سنجش، فرایندی یکپارچه برای تعیین ماهیت و گستره یادگیری و رشد دانش‌آموز است.

یادآوری ۴. ثابت، ویژگی یا صفتی خاص است که در همه شرایط، یکسان باقی می‌ماند و آنچه یک پدیده را قابل مطالعه می‌کند، امکان و احتمال تغییر است.

یادآوری ۵. متغیر، چیزی است که می‌تواند ارزش‌های گوناگونی داشته باشد.

یادآوری ۶. آزمون، وسیله‌ای است عینی و استاندارد شده که برای اندازه‌گیری نمونه‌ای از حالات و رفتار معین فرد یا افرادی خاص به کار می‌رود.

یادآوری ۷. فرایند اجرای آزمون را آزمودن می‌گویند.

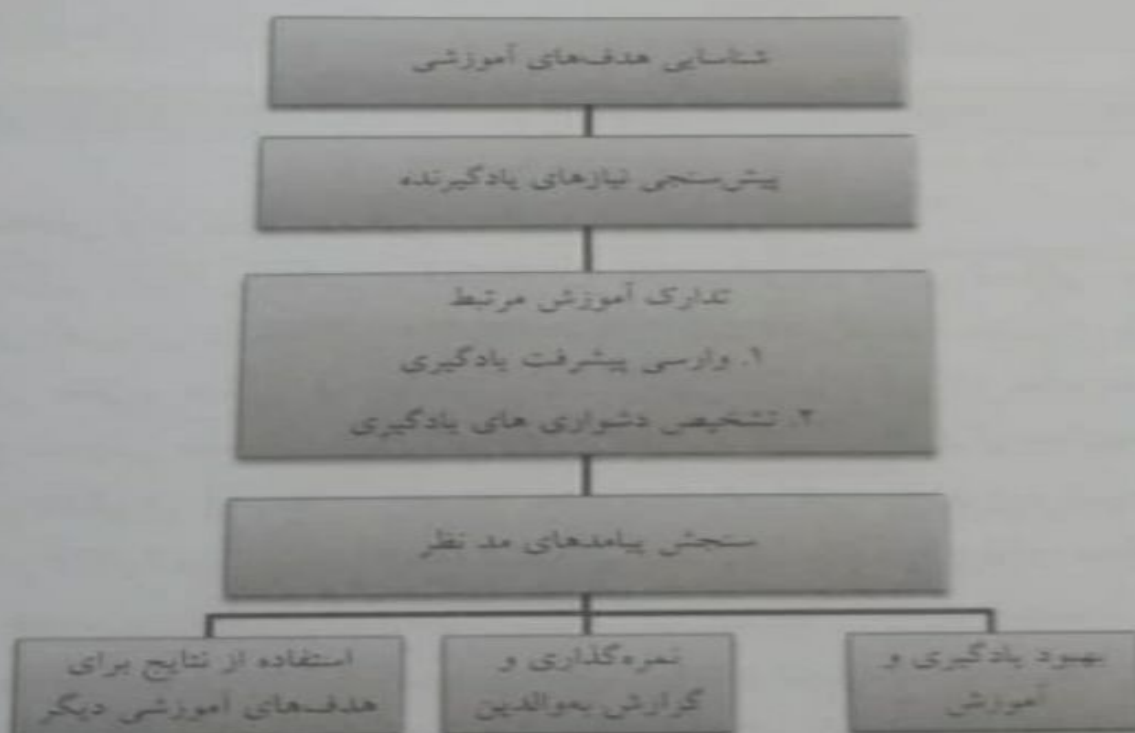
۱-۴ سنجش و فرایند آموزش

هدف اصلی آموزش کلاسی، کمک کردن به دانش‌آموزان برای دستیابی به مجموعه هدف‌های آموزشی است. این هدف‌ها همواره باید دربرگیرنده تغییرات مطلوب در حوزه‌های عقلی، عاطفی و جسمی باشد. وقتی از این دیدگاه به آموزش کلاسی نگاه شود، سنجش به بخش جدایی‌ناپذیری از فرایند تدریس - یادگیری تبدیل می‌شود. پیامدهای یادگیری مد نظر، به وسیله هدف‌های آموزشی پی‌ریزی می‌شود، تغییرات مطلوب در دانش‌آموزان، به وسیله فعالیت‌های یادگیری برنامه‌ریزی شده تحقق می‌یابد و پیشرفت تحصیلی دانش‌آموز به صورت دوره‌ای، به وسیله آزمون‌ها و دیگر ابزارهای سنجش سنجیده می‌شود.

در گام اول، تدریس و سنجش، تعیین پیامدهای یادگیری مورد انتظار از آموزش کلاسی است.

در گام دوم، هنگامی که هدف‌های آموزشی به‌روشنی مشخص شدند، معمولاً مطلوب است سنجشی از نیازهای یادگیرنده در ارتباط با پیامدهای یادگیری که قرار است به آن‌ها برسد، به‌عمل‌آید. به‌این‌توجه سنجش، سنجش آغازین گفته‌می‌شود تا مشخص شود آیا یادگیرنده، بر پیش‌نیازهای درس جدید مسلط است یا به‌آموزش کمکی نیاز دارد.

در گام سوم، آموزش رسمی مرتبط با هدف‌های آموزشی شروع می‌شود و در طول آموزش، ارزیابی‌هایی به‌عمل می‌آید تا مشخص شود آیا یادگیرنده پیشرفت داشته‌است یا خیر؟ در این مرحله، موانع پیشرفت یادگیرنده واری می‌شود و اگر مشکل خاصی وجود داشته‌باشد، از آموزش کمکی بهره گرفته می‌شود. گام نهایی فرایند آموزشی، تعیین این نکته است که آیا یادگیرنده به‌هدف‌های آموزشی موردنظر رسیده است یا خیر. در این مرحله، از آزمون‌ها و دیگر ابزارهای سنجش برای تعیین این نکته استفاده می‌شود.



شکل ۱-۲. الگوی آموزشی ساده‌شده (اقتباس از میلر، لین و گراندلند، ۲۰۰۹)

استفاده از نتایج چهار گام بالا، هم کمک به معلم است برای اصلاح ایرادهای احتمالی روش آموزشی و هم کمک به دانش‌آموز که به واسطه سنجش، به‌ویژه بازخورد لازم داده می‌شود. انعکاس نتایج به والدین نیز سودمند است و در صورت لزوم، کمک‌های لازم را به منظور تحقق هدف‌های آموزشی به عمل می‌آورند. تدریس، یادگیری و سنجش، فرایندهای مرتبط و درهم‌تنیده‌ای هستند و ماهیت پیچیده آن‌ها، به وسیله میلر، لین و گراندلند (۲۰۰۹)، به صورت زیر خلاصه شده است:

یادآوری ۱. هدف اصلی آموزش کلاسی، کمک به دانش‌آموزان برای دستیابی به مجموعه هدف‌های آموزشی است.

خلاصه فصل اول

اندازه‌گیری ویژگی‌های روانی و تربیتی، یکی از فعالیت‌های مهم روان‌شناسان و مشاوران است و برای مطالعه مؤثر در هر حوزه رفتاری بشر، شخص باید اصول اساسی سنجش و اندازه‌گیری را بفهمد. مشاوران تحصیلی و روانی، آزمون‌ها را در موارد مختلفی از جمله پیش‌بینی موفقیت‌های آینده در امور تحصیلی، شغلی و برنامه‌ریزی تحصیلی، مورد استفاده قرار می‌دهند. سپس تعریف اندازه‌گیری اشاره و مقیاس‌های اندازه‌گیری اسمی، ترتیبی، فاصله‌ای و نسبی مطرح شد. همچنین اصطلاحات سنجش، ارزیابی، ارزشیابی، ثابت، متغیر، آزمون و آزمودن تعریف شدند.

در ادامه، به اصول کلی سنجش اشاره شد که عبارت‌اند از: آنچه اولویت دارد و سنجیده می‌شود، به روشنی مشخص سازید. انتخاب یک روش سنجش باید بر مبنای ارتباطش با ویژگی‌ها یا عملکرد مورد نظر باشد. سنجش جامع، مستلزم بهره‌گیری از روش‌های متنوع است. استفاده مناسب از روش‌های سنجش، مستلزم آگاهی از محدودیت‌های آن‌هاست و سنجش، وسیله است، به خودی خود هدف نیست.

خودآزمایی چهارگزینه‌ای فصل اول

۱. جنبش اندازه‌گیری‌های روانی، در کدام قرن و به وسیله چه کسی پایه‌گذاری شد؟

ب) قرن بیستم، نرندایک

د) قرن بیستم، کاپلان

الف) قرن نوزدهم، گالتون

ج) قرن هجدهم، گالتون

۲. کدام گزینه، جزء موارد کاربردی استفاده از آزمون‌های روان‌شناختی نیست؟
 الف) کمک به مراجعان در کسب سازگاری
 ب) تشخیص مشکلات درمان‌جویان
 ج) پیش‌بینی موفقیت‌های آینده
 د) اندازه‌گیری ویژگی‌های فیزیکی
۳. کدام مقیاس، دارای طبقه‌های مشخص و متمایزی است که جنبه کیفی دارند؟
 الف) نسبی
 ب) فاصله‌ای
 ج) ترتیبی
 د) اسمی ✓
۴. بعد از اجرای آزمون، آزمونگر بایستی وارد کدام مرحله شود؟
 الف) آشنایی بیشتر با آزمودنی
 ب) تفسیر
 ج) نمره‌گذاری
 د) وارد کردن نمره‌ها به نیم‌رخ
۵. به جمع‌آوری اطلاعات درباره دانش‌آموزان، آموزش و راهنمایی گفته می‌شود.
 الف) سنجش
 ب) اندازه‌گیری
 ج) ارزشیابی
 د) آزمون
۶. کدام یک از گزینه‌های زیر، اندازه‌گیری غیرمستقیم است؟
 الف) اندازه‌گیری قد با متر
 ب) اضطراب امتحانی
 ج) نمره درس سنجش و اندازه‌گیری
 د) مدت زمان صرف‌شده برای مطالعه
۷. کدام گزینه، مستلزم وزن کردن ارزش یا اعتبار چیزی و مقایسه با سایر برنامه‌هاست؟
 الف) سنجش
 ب) ارزشیابی
 ج) ارزیابی
 د) آموزش
۸. روشی که برای کمی کردن رفتار با کمک به درک و پیش‌بینی رفتار، مورد استفاده قرار می‌گیرد نام دارد.
 الف) آزمون
 ب) آزمودنی
 ج) ارزشیابی
 د) اندازه‌گیری
۹. شیوه رسمی و نظام‌دار برای گردآوری رفتار نمونه‌ای از دانش‌آموز، چه نام دارد؟
 الف) آزمون
 ب) آزمودن
 ج) ارزیابی
 د) سنجش

۱۰. اولین گام در سنجش و فرایند آموزش چیست؟
(الف) سنجش نیازهای یادگیرنده
(ب) آموزشی رسمی مرتبط با هدف‌های آموزشی
(ج) دستیابی یادگیرنده به هدف‌های آموزشی
(د) تعیین پیامدهای یادگیری
۱۱. نخستین گام در تجزیه و تحلیل داده‌ها چیست؟
(الف) تعیین نوع مقیاس داده‌ها
(ب) استفاده از روش‌های آماری
(ج) آگاهی از عملکرد مراجعان
(د) گردآوری داده‌ها
۱۲. کدام مقیاس اندازه‌گیری، دارای صفر فرضی است؟
(الف) اسمی
(ب) ترتیبی
(ج) فاصله‌ای
(د) نسبی

خودآزمایی تشریحی فصل اول

۱. جامعه و نمونه آماری را با ارائه مثال تعریف کنید؟
۲. اهمیت و فواید کاربرد روش‌های اندازه‌گیری در علوم تربیتی را بنویسید؟
۳. سنجش را تعریف و برای آن مثالی بیان کنید؟
۴. انواع مقیاس‌های اندازه‌گیری را نام ببرید و برای هر کدام مثالی ارائه کنید؟
۵. ارزشیابی چیست؟
۶. ثابت و متغیر را تعریف و برای هر کدام مثالی بیان کنید؟
۷. آزمون و آزمودن را تعریف کنید؟

فصل دوم

اندازه‌گیری پیامدهای یادگیری

هدف کلی

آشنایی با طبقه‌بندی و اندازه‌گیری پیامدهای یادگیری

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. غایت‌ها و هدف‌های آموزشی را توضیح دهد.
۲. پیامدهای یادگیری را شرح دهد.
۳. طبقه‌بندی هدف‌های آموزشی بلوم و زیرگروه‌های آن را تعریف کند.
۴. طبقه‌بندی آموزشی بلوم و اندرسون و کراتول را با هم مقایسه کند.
۵. ابعاد دانش اندرسون و کراتول را نام‌برد و مثال بیاورد.
۶. برای یک درس، نقشه‌آزمون با جدول مشخصات را تهیه کند.

مقدمه

از تدریس خود، چه نوع پیامد یادگیری را انتظار دارید؟ دانش؟ فهمیدن؟ به‌کار بستن؟ مهارت‌های تفکر؟ حل مسئله؟ مهارت‌های عملکردی؟ نگرش‌ها؟ بدیهی است تعریف پیامدهای یادگیری مطلوب، اولین گام در تدریس خوب و برای سنجش یادگیری دانش‌آموز، ضروری است. سنجش صحیح ایجاب می‌کند تا حد امکان شیوه‌های سنجش، با پیامدهای یادگیری مد نظر، ارتباط مستقیم داشته باشد. غایت‌های آموزشی^۱ و هدف‌ها^۱ در

فرایند آموزشی^۱ و سنجش، نقش کلیدی دارند و راهنمای تدریس و یادگیری‌اند. هدف‌ها آموزش را به‌دیگران انتقال می‌دهند و راهنمایی‌هایی برای سنجش یادگیری دانش‌آموز تدارک می‌بینند. این هدف‌های عمده، در شکل ۱-۲ توضیح داده شده‌اند.

۱-۲ نوشتن هدف‌های آموزشی

یک هدف آموزشی^۲ باید جمله‌ای روشن و دقیق از یک مهارت یا مهارت‌هایی باشد که انجام آن پس از آموزش واحد درسی، از دانش‌آموز، مورد انتظار است (کوبیزین^۳ و بوریش^۴، ۲۰۱۳). این دقت باید قادر به توضیح سطح مهارت ارائه‌شده و شرایط نشان‌دادن آن مهارت باشد. افزون‌براین، هدف آموزشی باید در قالب عبارت‌های رفتاری قابل مشاهده و قابل اندازه‌گیری بیان شود؛ به‌گونه‌ای که برداشت دو دانش‌آموز از آن جمله یا هدف آموزشی یکسان باشد. به‌طور خلاصه یک هدف آموزشی، شامل موارد زیر است:

- رفتار قابل مشاهده^۵ (فعل هدف آموزشی به‌گونه‌ای نوشته‌شود که پیامد یادگیری را نشان دهد).

- شرایط^۶ (شرایط ویژه‌ای که تحت آن شرایط، رفتار باید نشان داده‌شود. برای مثال در امتحان آمار، آیا آوردن ماشین حساب در جلسه امتحان مجاز است یا خیر؟).

- سطح عملکردی^۷ که برای رسیدن به حد تسلط، مد نظر است، توضیح داده‌شود.

مثال‌های زیر کمک می‌کند تا با هریک از این مؤلفه‌ها آشنا شوید. با تمرین، باید تسلط شما در نوشتن هدف‌های آموزشی افزایش یابد.

۲-۲ شناسایی پیامدهای یادگیری

یک هدف آموزشی باید شامل یک فعل کنشی^۸ باشد (فعلی) که هدف آموزشی را به‌صراحت مشخص سازد. با وجود این، همه افعال کنشی، پیامدهای یادگیری^۱ را

1. Objectives
2. Instructional Process
3. Instructional Objective
4. Kubiszy
5. Borich
6. Observable Behavior
7. Circumstances
8. Criterion Level

به‌صراحت مشخص نمی‌کنند. پیامدهای یادگیری، اغلب با فعالیت‌های یادگیری^۱ اشتباه گرفته می‌شوند. در مثال‌های زیر نسبت به‌تعیین اینکه کدام مورد، معرف پیامدهای یادگیری است و کدام یک، معرف فعالیت‌های یادگیری، تلاش کنید.

۱. کودک تا آخر نیمسال تحصیلی، تصاویر کلماتی را که صدای مشابه دارند، شناسایی خواهد کرد.

۲. کودک، نقد شعر را توضیح خواهد داد.

۳. دانش‌آموز، تفریق اعداد یک‌رقمی را تا ۸۰٪ به‌درستی انجام می‌دهد.

۴. دانش‌آموز، دانش نقطه‌گذاری صحیح را نشان خواهد داد.

۵. دانش‌آموز، جدول ضرب را تمرین خواهد کرد.

۶. دانش‌آموز، «سرود ملی» را خواهد سرود.

در چهار هدف اول، افعال کنشی شناسایی کردن، توضیح دادن، تفریق کردن و نشان دادن، همگی به‌پیامدها یا بازده نهایی واحد درسی اشاره دارند. با وجود این، فعل کنشی تمرین کردن در هدف آموزشی شماره ۵، تنها بر عملی دلالت دارد که به‌پیامد یادگیری منجر خواهد شد؛ بنابراین هدف شماره ۵، هیچ‌گونه پیامد یادگیری ندارد و یک فعالیت یادگیری است. این درواقع یک هدف میانی^۲ است تا یک هدف پایانی^۳.

هدف شماره ۶ نیز کمی پیچیده است. آیا سرودن، یک پیامد است یا یک فعالیت؟ بدون اطلاعات بیشتر، تشخیص آن دشوار است. اگر هدف شما این است که ترس شاگرد از سرودخواندن در مکان عمومی از بین برود، ممکن است یک پیامد یادگیری باشد. با وجود این اگر سرودخواندن، تنها تمرینی برای عملکرد بعدی باشد، این یک فعالیت یادگیری است. تمرین زیر به‌شما کمک می‌کند پیامد یادگیری و هدف یادگیری را از همدیگر تمیز دهید.

پیامدهای یادگیری (هدف پایانی)	فعالیت‌های یادگیری (هدف میانی)
شناسایی کردن	مطالعه کردن
به‌یاد آوردن	تماشا کردن
فهرست کردن	گوش دادن
نوشتن	خواندن

تمرین. پیامدهای یادگیری و فعالیت یادگیری را با گذاشتن «پ» بعد از پیامدها و گذاشتن «ف» بعد از فعالیت‌ها از همدیگر تمیز دهید.

۱. نصب رادیوی ماشین:
۲. از حفظ گفتن چهار مؤلفه انشای خوب:
۳. جمع‌زدن اعداد علامت‌گذاری‌شده به‌طور صحیح:
۴. تمرین موسیقی:
۵. از حفظ گفتن بخش‌هایی از یک سخنرانی:
۶. عنوان‌کردن موضوع اصلی «داستان موش و گربه»:
۷. حفظ‌کردن حروف الفبا:
۸. نقطه‌گذاری یک انشا به‌طور صحیح:

پاسخ‌ها. ۱) پ؛ ۲) پ؛ ۳) پ؛ ۴) ف؛ ۵) پ؛ ۶) پ؛ ۷) ف؛ ۸) پ

اگر فعالیتی بر یک بازده^۱ یا نتیجه خاص دلالت‌کند، یک پیامد تلقی می‌شود. چیزی که می‌خواهیم دربردارنده هدف آموزشی باشد، بازده نهایی شیوه آموزشی است. سؤال‌های آزمون، بر مبنای این بازده نهایی تدوین می‌شوند. اگر دریافتید که هدف شما شامل یک فعالیت یادگیری است (هدف میانی) و نه یک پیامد (هدف پایانی)، سؤال را به‌گونه‌ای بازنویسی کنید که بازده فعالیت موردنظر، بیان شده باشد. حالا دو پیامد یادگیری را در نظر بگیرید: یکی قابل اندازه‌گیری^۲ و قابل مشاهده^۳ و دیگری غیرقابل مشاهده.

یادآوری ۱. غایت‌های آموزشی و هدف‌ها راهنمای تدریس و یادگیری‌اند.

یادآوری ۲. یک هدف آموزشی باید شامل رفتار قابل مشاهده، شرایط و سطح عملکرد باشد.

یادآوری ۳. یک هدف آموزشی باید شامل یک فعل کنشی باشد که هدف آموزشی را به‌صراحت مشخص سازد.

یادآوری ۴. اگر فعالیتی، بر یک بازده یا نتیجه خاص دلالت‌کند، یک پیامد تلقی می‌شود.

۳-۲ شناسایی پیامدهای یادگیری قابل مشاهده و قابل اندازه‌گیری

در این مرحله، کار شما تعیین این موضوع است که آیا پیامد، به صورت رفتار قابل مشاهده و قابل اندازه‌گیری یا به شکل رفتار غیرقابل اندازه‌گیری و غیرقابل مشاهده بیان شده است؛ یعنی آیا با مشاهده دو یا چند دانش‌آموز، اتفاق نظر دارند که دانش‌آموز، پیامد یادگیری را نشان داده است؟ گاهی لازم است رفتار غیرقابل مشاهده، با یک شاخص قابل مشاهده پیامد یادگیری جایگزین شود. برای مثال اگر هدف ما این باشد: «دانش‌آموز، دانش نقطه‌گذاری را نشان خواهد داد»، پیامد یادگیری «نشان دادن دانش نقطه‌گذاری»، غیرقابل اندازه‌گیری است. چطور می‌دانیم دانش، نشان داده شده است؟ از دانش‌آموز می‌پرسیم؟ آیا فرض می‌کنیم که اگر دانش‌آموز برای سخنرانی آماده باشد یا بخش مناسبی از متن را بخواند، دانش به دنبال آن خواهد آمد؟ احتمالاً خیر. در عوض، ما به شاخصی نیاز داریم که نشان دهد دانش‌آموز، قواعد نقطه‌گذاری را می‌داند. به طور مثال برای نشان دادن دانش نقطه‌گذاری، باید در جمله «هرجا مناسب است، ویرگول بگذارد»، «قواعد حاکم بر دو نقطه روی هم یا نقطه ویرگول را فهرست کند» و مانند آن.

مثال‌های زیر را بخوانید و پیامدهای قابل مشاهده و غیرقابل مشاهده را مشخص‌سازیید.

غیرقابل مشاهده

ارزش داشتن

دانستن

درک کردن

فهمیدن

قابل مشاهده

فهرست کردن

ازبرخواندن

ساختن

کشیدن

تمرین. با گذاشتن «ق» بعد از پیامدهای قابل مشاهده و گذاشتن «غ» بعد از پیامدهای غیرقابل مشاهده، پیامدهای یادگیری قابل مشاهده را از پیامدهای یادگیری غیرقابل مشاهده تمیز دهید.

۱. دور حرف اول کلمات، دایره بکشد.

۲. با قانون آشنا باشد.

۳. اعداد دورقمی را روی کاغذ جمع بزنند.

۴. فرایند آسمزی را درک کند.
۵. از صحبت کردن لذت ببرد.
۶. پیچی را با آچار باز کند.
۷. نام ده شخصیت سینمایی ایرانی را از حفظ بگوید.
۸. نظریه نمره واقعی را درک کند.
۹. از هنر کننده کاری لذت ببرد.
۱۰. شعر کوتاهی را از حفظ بخواند.

پاسخ‌ها: (۱) ق؛ (۲) غ؛ (۳) ق؛ (۴) غ؛ (۵) غ؛ (۶) ق؛ (۷) ق؛ (۸) غ؛ (۹) غ؛ (۱۰) ق

۲-۳-۱ بیان شرایط

یک هدف آموزشی^۱، هر نوع شرایط خاصی را که یادگیری در آن به وقوع خواهد پیوست، توصیف می‌کند. اگر پیامد یادگیری قابل مشاهده، در یک زمان خاص، مکان خاص، با مواد، دستگاه، ابزار یا منابع دیگری رخ می‌دهد، پس شرایط، به وضوح در قالب کلمات عینی بیان شود.

مثال.

- با دراختیار داشتن ماشین حساب، اعداد دورقمی را در هم ضرب کند و هیچ خطایی نداشته باشد.
- با دراختیار داشتن یک لیست، همه غلط‌های چاپی آن را اصلاح کند.
- با دراختیار داشتن فهرستی از شش کلمه بهم ریخته و با مرتب کردن کلمات، یک جمله بسازد.

تمرین. شرایط را برای پیامدهای یادگیری زیر بنویسید.

۱. با دراختیار داشتن، روغن و روغن فیلتر را عوض کنید.
۲. با دراختیار داشتن، دما را مشخص کنید.
۳. با دراختیار داشتن، اعداد سه رقمی را جمع بزنید.

پاسخ‌های احتمالی. (۱) یک ماشین داخلی؛ (۲) دماسنج؛ (۳) ماشین حساب

۲-۳-۲ بیان ملاک‌ها

- یک هدف آموزشی نشان می‌دهد شیوه انجام یک رفتار چگونه باشد که مورد پذیرش واقع شود. برای هر هدف ارائه شده، تعدادی سؤال نوشته خواهد شد. سطح ملاک عملکرد قابل قبول، مشخص می‌سازد که برای تعیین دستیابی یا عدم دستیابی دانش‌آموز به هدف، باید به چند سؤال از این سؤال‌ها پاسخ صحیح دهد.
- با دراختیارداشتن ۲۰ جمع دورقمی، دانش‌آموز همه عددها را به‌طور صحیح محاسبه خواهد کرد.
 - با دراختیارداشتن ۲۰ جمع دورقمی، دانش‌آموز ۹۰٪ جمع‌ها را به‌درستی محاسبه خواهد کرد.

تمرین. ملاک عملکرد قابل قبول را برای هدف‌های زیر بنویسید.

۱. با دراختیارداشتن ۱۰ کلمه، با صحت، دور کلماتی که «وا» دارند و «ا» تلفظ می‌شوند، دایره بکشید.
 ۲. دانش‌آموز، ۱۰۰ متر شنا را در کمتر از دقیقه، شنا خواهد کرد.
 ۳. دانش‌آموز، قادر باشد یک صفحه مشق را در کمتر از دقیقه بنویسد.
- پاسخ‌های احتمالی. ۱) ۸۰٪؛ ۲) ۶۰ ثانیه؛ ۳) پنج
- ملاک را می‌توان برحسب موارد ذیل در نظر گرفت.
۱. تعداد سؤال‌های صحیح (یا درصد سؤال‌های صحیح).
 ۲. تعداد سؤال‌های پیاپی صحیح (برای مثال ۵ سؤال اول، بدون غلط باشد).
 ۳. ویژگی ضروری (برای مثال اندازه‌گیری تا میلی‌گرم در آزمایش شیمی).
 ۴. کامل کردن در یک محدوده زمانی (هنگامی که سرعت عمل، مهم است).
- تا اینجا یادگرفتید که:

- پیامدهای یادگیری را از فعالیت‌های یادگیری تمیز دهید.
- پیامدهای یادگیری قابل مشاهده / قابل اندازه‌گیری را از پیامدهای غیرقابل مشاهده / غیرقابل اندازه‌گیری یادگیری تمیز دهید.
- شرایط را بیان کنید.
- ملاک عملکرد را بیان کنید.

قبل از پرداختن به موضوع بعدی، به موارد دیگری می‌پردازیم که در ساخت هدف‌های آموزشی باید مد نظر قرار گیرند.

اغلب برای اندازه‌گیری پیامدهای یادگیری، به اشتباه آن‌ها را خیلی پیچیده می‌کنیم. در نتیجه سؤال برای دانش‌آموز پیچیده خواهد شد و آن را درک نخواهد کرد. اگر می‌خواهید بدانید دانش‌آموز می‌تواند نام خود را بنویسد، از او بخواهید نام خودش را بنویسد، اما نه چشم‌پسته. در مقابل وسوسه فریب‌دادن دانش‌آموز مقاومت کنید. مثال‌های زیر را در نظر بگیرید:

- دانش‌آموز، توانایی خود را در یادآوری نام شخصیت‌های کتاب بینوایان، با نقاشی تصویر هریک نشان دهد.

- تلویزیون و تلفن را با رسم نمودار الکتریکی هر کدام، از یکدیگر تشخیص دهد.

در مثال اول، نقاشی تصاویر شخصیت‌ها احتمالاً به‌ما اجازه می‌دهد تعیین کنیم آیا دانش‌آموز می‌تواند شخصیت‌های کتاب بینوایان را به یاد بیاورد، اما آیا راه بهتری برای اندازه‌گیری یادآوری (که زمان کمتری هم می‌طلبد) وجود ندارد؟ اگر از دانش‌آموز بخواهید فقط نام شخصیت‌ها را بنویسد، راحت‌تر نیست؟ اگر هدف آموزشی شما فقط یادآوری است، نوشتن اسامی کفایت می‌کند.

در مثال دوم، سؤال، بدون ضرورت پیچیده شده است. به جای آن می‌توانستید دو توضیح ارائه دهید، یکی برای تلویزیون و دیگری برای تلفن و فقط از آن‌ها بخواهید (شفاهی یا کتبی) مشخص کنند کدام تعریف، مربوط به کدام یک است.

یادآوری ۱. یک هدف آموزشی، هر نوع شرایط خاصی را که یادگیری در آن به وقوع خواهد پیوست، توصیف می‌کند.

یادآوری ۲. یک هدف آموزشی نشان می‌دهد شیوه انجام یک رفتار چگونه باشد که مورد پذیرش واقع شود.

یادآوری ۳. هر هدف آموزشی، سه جزء دارد: پیامد یادگیری قابل مشاهده، شرایط و ملاک عملکرد.

تکته. هنگامی که هدف‌های آموزشی را می‌نویسید، بررسی همه اجزاء هدف‌های آموزشی برای آگاهی از نوشتن موارد ضروری، ایده خوبی است.

۲-۴ طبقه‌بندی هدف‌های آموزشی

هدف‌های آموزشی، بسیار متنوع‌اند و برای سهولت مطالعه، به گروه‌های مختلفی طبقه‌بندی شده‌اند. قدیمی‌ترین طبقه‌بندی، مربوط به بلوم^۱ و همکاران (۱۹۵۶؛ به نقل از بلوم، ۱۹۸۴) است. سپس اندرسون و کراتول^۲ (۲۰۰۱) در آن تجدیدنظر به عمل آوردند. مارزانو^۳ و کندال^۴ (۲۰۰۷؛ به نقل از مارزانو و همکاران، ۲۰۱۰) نیز طبقه‌بندی خاص خود را معرفی کردند. در طبقه‌بندی بلوم و همکاران، هدف‌های آموزشی به سه حوزه شناختی^۵، عاطفی^۶ و روانی-حرکتی^۷ تقسیم شده‌اند.

هدف‌های حوزه شناختی، به جریان‌هایی که با فعالیت‌های ذهنی و فکری آدمی سروکار دارند، مربوط می‌شوند. سطح دشواری هدف، به پیچیدگی سطح شناختی، ذهنی یا فکری‌ای بستگی دارد که مد نظر هدف است. برای مثال، هدف «دانش‌آموز حداقل نام سه بازیکن از چهار بازیکن مشهور تاریخ فوتبال ایران را بگوید»، یک تکلیف شناختی نسبتاً آشکار است. این هدف، صرفاً نیازمند بازیابی اطلاعات است. چنین هدفی، هدف سطح پایین‌تر شناختی تلقی می‌شود. از طرف دیگر، هدفی مانند «با توجه به گفته‌های شاهد عینی درباره یک رویداد، با ۷۵٪ صحت، واقعیت را از دیدگاه جدا کنند»، یک تکلیف شناختی نسبتاً پیچیده است. این هدف، توانایی تحلیل توجیه شاهد را دارد و می‌تواند معیارهایی را به کارگیرد که به کمک آن بتواند تعیین کند آیا جمله‌ها، قابل مشاهده و عینی هستند یا ذهنی و مبتنی بر برداشت شخصی. چنین هدفی، یک هدف شناختی سطح بالاتر تلقی می‌شود.

طبقه‌بندی بلوم از هدف‌های حوزه شناختی، شش سطح دارد که عبارت‌اند از: دانش^۸، درک^۹، به کار بستن^{۱۰}، تحلیل^{۱۱}، ترکیب^{۱۲} و ارزشیابی. همان‌طور که در بالا اشاره شد، این طبقه‌بندی به وسیله اندرسون و کراتول (۲۰۰۱) مورد تجدیدنظر

1. Bloom
2. Anderson & Krathwohl
3. Marzano
4. Kendall

قرار گرفت. در طبقه‌بندی اندرسون و همکاران (۲۰۰۱)، یک بعد جدید به طبقه‌بندی بلوم اضافه شده است. این بعد، شامل انواع دانش است که عبارت‌اند از: دانش امور واقعی^۱، دانش مفهومی^۲، دانش روندی^۳ و دانش فراشناختی^۴. پس از مشخص شدن نوع دانش، یک هدف در یکی از سطوح فرایند شناختی طبقه‌بندی می‌شود.

یادآوری ۱. قدیمی‌ترین طبقه‌بندی هدف‌های آموزشی، مربوط به بلوم و همکاران (۱۹۵۶) است.

یادآوری ۲. در طبقه‌بندی بلوم و همکاران (۱۹۵۶)، هدف‌های آموزشی به سه حوزه شناختی، عاطفی و روانی-حرکتی تقسیم شده‌اند.

یادآوری ۳. هدف‌های حوزه شناختی بلوم عبارت‌اند از: دانش، درک، به‌کار بستن، تحلیل، ترکیب و ارزشیابی.

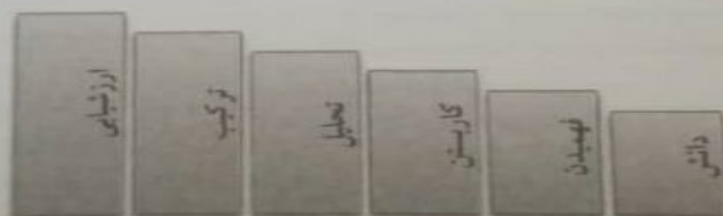
۲-۴-۱ طبقه‌بندی تجدیدنظرشده

همان‌طور که اشاره شد، در طبقه‌بندی تجدیدنظرشده، حوزه شناختی، یک بعد دانش^۵ و یک بعد فرایند شناختی^۶ دارد. بعد فرایند شناختی، دربرگیرنده به‌یادآوردن^۷، فهمیدن^۸، به‌کار بستن، تحلیل کردن، ارزشیابی کردن^۹ و آفریدن^{۱۰} است. این طبقه‌بندی، به صورت سلسله‌مراتبی و از عینی به انتزاعی و ساده به پیچیده تنظیم شده است. در این طبقه‌بندی، شش مقوله اصلی، از شکل اسم^{۱۱} به فعل تغییر داده شده است. دلیل این تغییر این است که طبقه‌بندی، منعکس‌کننده اشکال متفاوت تفکر و تفکر نیز یک فرایند فعال است. افعال^{۱۲}، وقایع و فعالیت‌ها را

1. Factual Knowledge
2. Conceptual knowledge
3. Procedural Knowledge
4. Metacognitive Knowledge

توصیف می‌کنند، درحالی‌که اسم، قادر به توصیف این ویژگی‌ها نیست. دانش در طبقه‌بندی اندرسون، به فعل به خاطر سپردن یا به یاد آوردن، تغییر شکل داده است، زیرا دانش، نتیجه یا محصول تفکر است، نه یک شکل خاص از تفکر. در این تجدیدنظر، در سطح دوم به جای کلمه مبهم درک، کلمه دقیق‌تر فهمیدن و به جای کلمه ترکیب، کلمه آفریدن، جایگزین شده است. همچنین در این تجدیدنظر، جای سطوح ارزشیابی و آفریدن، با هم عوض شده است. این طبقه‌بندی نه تنها در شناسایی پیامدهای مطلوب، بلکه برای شناسایی و سنجش فرایندهای تفکری^۱ که دانش‌آموزان باید به کارگیرند تا به پیامدها دست یابند، به معلمان کمک می‌کند. در ادامه این فصل، این طبقه‌بندی با تفصیل بیشتری مورد بحث قرار خواهد گرفت.

همان‌طور که در شکل ۱-۲ می‌بینید، در طبقه‌بندی بلوم فرض می‌شود که این سطوح، حالت سلسله‌مراتبی دارند؛ یعنی هدف‌های سطح بالاتر، هم شامل هدف‌های سطح پایین‌تر می‌شوند و هم به مهارت‌های شناختی سطح پایین‌تر وابسته‌اند. هر سطح طبقه‌بندی، ویژگی‌های مختلفی دارد. هر کدام از این سطوح به همراه ویژگی‌های آن، در زیر توصیف می‌شود.



شکل ۱-۲. طبقه‌بندی هدف‌های آموزشی: حیطه شناختی بلوم

طبقه‌بندی تجدیدنظرشده، یک ابزار قابل اعتماد برای برنامه‌ریزی آموزشی، انتقال یادگیری و ارزیابی است. همچنین این طبقه‌بندی، طیف گسترده‌تری از مخاطبان را هدف قرار می‌دهد، ولی طبقه‌بندی بلوم، به طور سنتی مانند یک ابزار به نظر می‌آید که در سال‌های نخست مدرسه (دوره ابتدایی)، بهترین عملکرد را دارد. درحالی‌که اندرسون، با تغییراتی که در آن ایجاد کرد، آن را جهانی، گسترده‌تر و قابل بهره‌برداری برای سطوح دبستان، راهنمایی، دبیرستان و حتی بالاتر کرده است.

۳۰ سنجش و اندازه‌گیری

در نهایت، طبقه‌بندی تجدیدنظرشده اندرسون، با مبنای قرار دادن طبقه‌بندی بلوم، تغییراتی یا هدف‌های زیر به وجود آورده است:

۱. به‌روزر کردن طبقه‌بندی بلوم برای استفاده در قرن ۲۱
۲. رفع اشکالات طبقه‌بندی قبلی
۳. گسترده کردن طیف مخاطبان این طبقه‌بندی از مقطع دبستان به مقاطع بالاتر
۴. توصیف و تشریح مقوله‌های اصلی با ارائه زیرمقوله‌هایی برای هر طبقه

یادآوری ۱. دانش در طبقه‌بندی اندرسون، به فعل به‌خاطر سپردن یا به‌یاد آوردن، تغییر شکل داده است.

یادآوری ۲. طبقه‌بندی بلوم، به‌طور سنتی به‌شکل ابزاری است که در دوره ابتدایی، بهترین عملکرد را دارد.

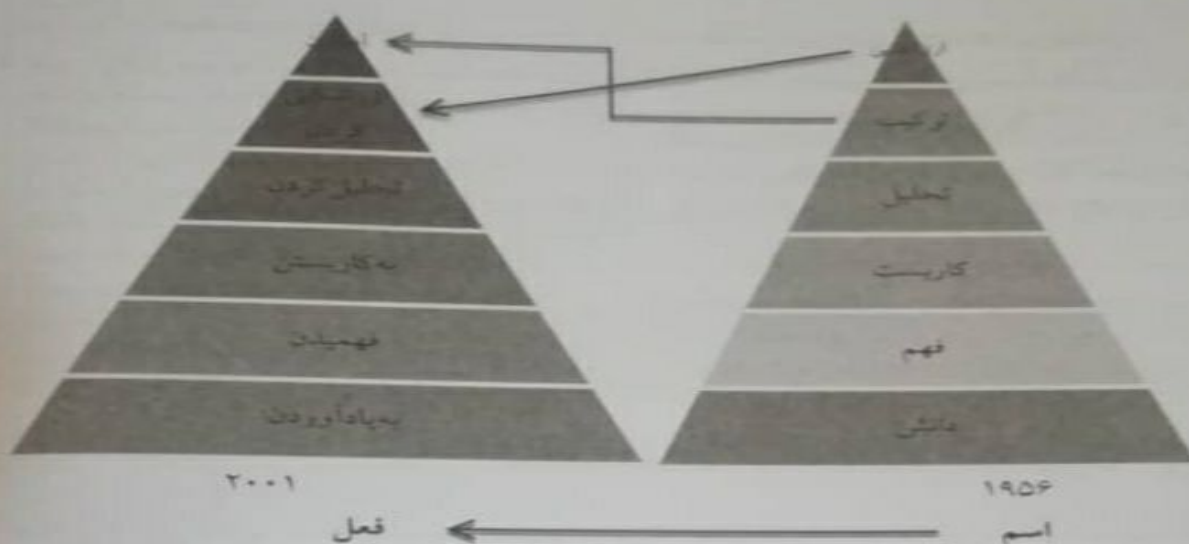
جدول ۱-۲. مقوله‌ها و زیرمقوله‌های طبقه‌بندی بلوم

مقوله اصلی	زیرمقوله‌ها
به‌یاد آوردن	تشخیص دادن، فهرست کردن، توصیف کردن، نام بردن
فهمیدن	تفسیر کردن، الگوسازی، خلاصه کردن، برداشت کردن
به کار بستن	اجرا کردن، مورد استفاده قرار دادن
تحلیل کردن	متمایز کردن، سازمان دادن، نسبت دادن
ارزشیابی کردن	وارسی کردن، نقد کردن، قضاوت و داوری کردن
آفریدن	تولید کردن، برنامه‌ریزی کردن، ساختن

جدول ۲-۲. مقایسه طبقه‌بندی حیطه شناختی بلوم با طبقه‌بندی اندرسون و کراتول

طبقه‌بندی بلوم ۱۹۵۶	طبقه‌بندی اندرسون و کراتول ۲۰۰۱
۱. دانش. دانش یا بازایی مواد از قبل یادگرفته شده. بعضی از افعال کنشی که در سطح دانش به کار گرفته می‌شوند، عبارت‌اند از: تعریف کردن، لیست کردن، ازیرخواندن، توصیف کردن، جور کردن، انتخاب کردن، تشخیص دادن، نام بردن، بیان کردن، نام گذاری کردن، به‌یاد آوردن و بیان کردن.	۱. به‌یاد آوردن (دانش). در سطح به‌یاد آوردن، دانش آموز باید اطلاعات آموخته شده را به‌یاد آورد. سؤال‌های آزمون، از دانش آموز می‌خواهند واقعیت‌ها، معانی و اژه‌ها، راهبردهای حل مسئله یا قواعد را به‌یاد آورد یا تشخیص دهد.

طبقه‌بندی بلوم ۱۹۵۶	طبقه‌بندی اندرسون و کراتول ۲۰۰۱
۱. فهم. توانایی فهم یا کسب معنا از مواد. بعضی از افعال کنشی که در سطح فهمیدن به‌کار گرفته می‌شوند، عبارت‌اند از: تبدیل کردن، تبیین کردن، استنباط کردن، حمایت کردن، گسترش دادن، پیش‌بینی کردن، تفکیک کردن، برآورد کردن، خلاصه کردن، تمیز دادن، تعمیم دادن و تفسیر کردن.	۲. فهمیدن. وقتی گفته می‌شود «دانش آموزان می‌فهمند که بتوانند از مواد شفاهی، کتبی یا تصویری، اطلاعاتی دریافت کنند». فهمیدن، زمانی صورت می‌پذیرد که دانش آموز بتواند بین دانش تازه و دانش‌های قبلی خود، ارتباط برقرار کند.
۳. کاربرد. توانایی آنچه یادگیرنده به‌طور انفرادی یاد گرفته است، در موقعیت‌های عملی به‌کار می‌گیرد. بعضی از افعال کنشی که در سطح کاربرد به‌کار گرفته می‌شوند، عبارت‌اند از: تغییر دادن، اصلاح کردن، حل کردن، محاسبه کردن، عمل کردن، انتقال دادن، توسعه دادن، سازمان دادن، تولید کردن، به‌کار گرفتن، آماده کردن و استفاده کردن.	۳. به‌کار بستن. به‌کار بستن، شامل استفاده از روش‌ها و روندها برای انجام تمرین‌ها یا حل کردن مسئله‌هاست. به‌کار بستن، به موقعیت‌هایی اشاره دارد که مواد یاد گرفته شده در آن، مورد استفاده قرار می‌گیرد. موقعیت‌هایی مانند سخنرانی‌ها، مصاحبه‌ها یا شبیه‌سازی‌ها.
۴. تحلیل. توانایی تجزیه یا تفکیک بخش‌هایی از مواد به مؤلفه‌های آن، به گونه‌ای که باعث درک بهتر آن شود. بعضی از افعال کنشی که در سطح تحلیل به‌کار گرفته می‌شوند، عبارت‌اند از: تجزیه کردن، تمیز دادن، دریافتن، قیاس کردن، شرح دادن، مرتبط ساختن، جدا کردن، استنباط کردن، دسته‌بندی کردن، طبقه‌بندی کردن، طرح کردن و تحلیل کردن.	۴. تحلیل کردن. مواد یا مفاهیم را به بخش‌هایی تبدیل کردن، تعیین اینکه مواد چگونه با یکدیگر و با ساختار کلی مرتبط هستند. کنش‌های ذهنی در این سطح، شامل سازمان‌دهی، تفکیک، نسبت دادن و توانایی تمیز دادن مؤلفه‌ها یا بخش‌ها از یکدیگر است.
۵. ترکیب. توانایی کنار هم قرار دادن بخش‌ها به گونه‌ای که یک کل یکپارچه و منحصر به فرد جدید ایجاد شود. بعضی از افعال کنشی که در سطح ترکیب به‌کار گرفته می‌شوند، عبارت‌اند از: طبقه‌بندی کردن، خلق کردن، بازتولید کردن، تدوین کردن، طرح کردن، خلاصه کردن، ترکیب کردن، ابداع کردن و تنظیم کردن.	۵. ارزشیابی کردن. قضاوت کردن بر اساس معیارها و استانداردها به صورت انتقادی. نقدها، توصیه‌ها و گزارش‌ها، بعضی از بازده‌هایی هستند که می‌توان به کمک آن‌ها فرایند ارزشیابی را ایجاد کرد. در طبقه‌بندی جدید، این بخش قبل از ترکیب قرار گرفته است.
۶. ارزشیابی. توانایی قضاوت، بررسی و حتی نقد ارزش مواد به منظور دستیابی به یک هدف معین. بعضی از افعال کنشی که در سطح ارزشیابی به‌کار گرفته می‌شوند، عبارت‌اند از: ارزشیابی کردن، نقد کردن، اعتبار بخشیدن، مقایسه کردن، حمایت کردن، انتقاد کردن، مقابله کردن، قضاوت کردن، تصمیم گرفتن، نتیجه‌گیری کردن، تفسیر کردن و توجیه کردن.	۶. آفریدن. کنار هم قرار دادن عناصر به منظور ایجاد یک کل یکپارچه؛ چیدن عناصر کنار هم به گونه‌ای جدید، به منظور خلق یک اثر جدید. این بخش، سخت‌ترین قسمت طبقه‌بندی جدید به‌شمار می‌آید.



شکل ۲-۲. مقایسه طبقه‌بندی بلوم با طبقه‌بندی اندرسون و کراتول

۲-۴-۲ ابعاد دانش اندرسون و کراتول

۱. دانش واقعی. دانش عناصر موردنیاز یادگیرنده برای آشناسدن با یک رشته علمی یا حل مسائل مربوط به آن.
۲. دانش مفهومی. دانش مقوله‌ها و طبقه‌ها (دانش انواع گوناگون حشرات)، دانش اصل‌ها و تعمیم‌ها (دانش اصول تغییر رفتار)، دانش نظریه‌ها، الگوها و روابط میان آن‌ها (دانش مربوط به نظریه‌های قدیم و جدید یادگیری).
۳. دانش روندی. اشاره به اطلاعات یا دانشی دارد که به دانش‌آموز کمک می‌کند در یک رشته، موضوع، یا زمینه مطالعاتی، با چگونگی انجام کار آشنا شود.
- دانش مهارت‌ها و الگوریتم‌های خاص یک رشته (دانش مهارت‌های موردنیاز در نقاشی آبرنگ)
- دانش فن‌ها و روش‌های خاص یک رشته (دانش روش‌های پژوهش متناسب با علوم اجتماعی)
- دانش معیارهای تعیین زمان استفاده فرد درباره خودش

جدول ۲-۳. طبقه‌بندی دوبعدی ابعاد دانش و فرایندهای شناختی

ابعاد دانش	فرایندهای شناختی					
	۱	۲	۳	۴	۵	۶
واقعی	به یاد آوردن	فهمیدن	به کار بستن	تحلیل کردن	ارزشیابی کردن	آفریدن
واژه‌شناسی، عناصر و مؤلفه‌ها	برجسب نقشه فهرست نام‌ها	تفسیر پاراگراف، خلاصه کردن کتاب	به کارگیری الگوریتم‌های ریاضی	دسته‌بندی کلمات	نقد مقاله	خلق داستان کوتاه
مفهومی	طبقه‌بندی‌ها اصول نظریه‌ها	تعریف سطوح طبقه‌بندی	توصیف طبقه‌بندی به‌زیان خود	نوشتن هدف‌ها با استفاده از طبقه‌بندی	نقد هدف‌های مکتوب	ایجاد نظام طبقه‌بندی جدید
روندی	مهارت‌های خاص، فنون معیارهای کاربرد	فهرست کردن گام‌های حل مسئله	بازگویی فرایند حل مسئله به‌زیان خود	استفاده از فرایند حل مسئله برای تکالیف تعیین شده	نقد تناسب فنون به کار گرفته شده در تحلیل کیس	گسترش رویکرد اصلی حل مسئله
فراشناختی	دانش عمومی، دانش خود	فهرست عناصر سبک یادگیری فردی	توصیف تلویحات سبک یادگیری	توسعه مهارت‌های مطالعه متناسب با سبک یادگیری	نقد تناسب سبک یادگیری خاص یا یادگیری خود	ایجاد نظریه سبک یادگیری خاص

۴. دانش فراشناختی. به آگاهی فرد از شناخت و فرایندهای شناختی خاص خودش اشاره دارد.

- دانش راهبردی. علم به اینکه تکرار اطلاعات، یکی از روش‌های حفظ اطلاعات است.
- دانش مربوط به تکلیف‌شناختی. شامل دانش موقتی و مشروط. علم به اینکه درک کتاب تخصصی از کتاب عمومی سخت‌تر است.
- دانش درباره خود. دانش اینکه فرد در بعضی از زمینه‌ها بی‌اطلاع است.

۲-۴-۳ ابعاد فرایند شناختی اندرسون و کراتول

۱. به یاد آوردن. وقتی هدف آموزش، حفظ مطالب آموزش داده شده به همان صورت اولیه از سوی دانش‌آموز است، این فرایند، به یاد آوردن نام دارد. این بخش، فرصتی برای دانش‌آموزان فراهم می‌کند تا بازخوانی و بازشناسی ساده‌ای از عناوین به دست آورند.

مثال.

- دانش‌آموز، تا روز شنبه و بدون خطا، چهار گروه غذایی را به‌خاطر خواهد آورد.
- دانش‌آموز، شعر رویه و زاغ را بدون اشتباه، از حفظ می‌خواند.
- قطب‌های آهن‌ریا معمولاً به‌این صورت نام‌گذاری می‌شوند:

الف) منفی و مثبت

ب) قرمز و آبی

پ) غربی و شرقی

ت) شمالی و جنوبی

۲. فهمیدن. طبقه فهمیدن، شامل هفت خرده طبقه است: (۱) تفسیر کردن؛ (۲) مثال آوردن؛ (۳) طبقه‌بندی کردن؛ (۴) مقایسه کردن؛ (۵) خلاصه کردن؛ (۶) استنباط کردن؛ (۷) تبیین کردن.

- تفسیر کردن. یعنی تبدیل اطلاعات از شکلی به شکل دیگر، تفسیر شامل تغییر مطلب از کلامی به کلام دیگر، از شکل به کلام، از کلام به شکل، از اعداد به کلام، از کلام به اعداد، از نت به موسیقی یا به آهنگ و مانند این‌هاست. اصطلاح دیگر برای تفسیر کردن، برگرداندن (ترجمه کردن) است.

مثال. دانش‌آموز تا آخر نیمسال تحصیلی، رویدادهای اصلی یک داستان را با دستور زبان صحیح، خلاصه خواهد کرد.

دانش‌آموزان با ۸۰٪ صحت، یک متن انگلیسی را به فارسی برمی‌گردانند.

- مثال آوردن (توضیح دادن). بیان مورد یا مثالی خاص از یک مفهوم یا اصل کلی.

مثال. دانش‌آموز بتواند برای رفتار آشکار، چهار مثال عینی بیاورد.

- طبقه‌بندی کردن. طبقه‌بندی کردن، فرایند مکمل مثال آوردن است و مشخص می‌سازد چیزی به یک طبقه تعلق دارد یا خیر.

مثال. طبق معیارهای تشخیصی DSM-5، اختلال استرس پس از سانحه، جزء کدام طبقه از اختلالات است؟

- خلاصه کردن. استفاده از یک بیان مختصر که معرف یک توضیح مفصل‌تر است. دانش‌آموز برای خلاصه کردن مطالب مفصل، باید آن‌ها را در قالب مضمون اصلی یا نکات مهم، مختصر کند.

مثال. دانش‌آموز بتواند وقایعی را که به صورت تصویر نشان داده شده‌اند، به‌زبانی ساده خلاصه کند.

- استنباط. استنباط، زمانی رخ می‌دهد که دانش‌آموز بتواند یک مفهوم یا اصل را از تعدادی مثال یا مورد (با توجه به روابط میان آن‌ها)، انتزاع کند.

مثال. جمعیت ایران در حال حاضر، هشتاد میلیون نفر است. با این قرض که رشد سالیانه جمعیت، $\frac{1}{5}$ درصد باشد، در سال ۱۴۰۰، جمعیت ایران چند نفر خواهد بود؟
- مقایسه کردن. تشخیص شباهت‌ها و تفاوت‌های بین دو یا چند شیء، رویداد، اندیشه، مسئله یا موقعیت.

مثال. دانش‌آموز بتواند به کمک توضیحات مربوط به کار مغز و کامپیوتر در زمینه دریافت اطلاعات، این دو را با هم مقایسه کند.

- تبیین. توانایی ساختن یک الگوی علت و معلولی از یک نظام و استفاده از آن. الگو ممکن است از یک نظریه علمی (مانند نظریه‌های علوم طبیعی) استخراج شود یا مبنای آن، پژوهش یا تجربه باشد (مانند آنچه که اغلب در علوم اجتماعی و انسانی رخ می‌دهد).

مثال. دانش‌آموز بتواند با استناد به منابع تاریخی، مفاد عهدنامه ترکمن‌چای را بیان کند.

۳. به کار بستن. وقتی دانش‌آموز بتواند از مطالبی که یاد گرفته است، به خوبی استفاده کند، در واقع به مرحله به کار بستن که بخشی از یادگیری محسوب می‌شود، دست یافته است. در این مرحله، دانش‌آموزان با به کارگیری افکار انتزاعی، مشکلات خود را حل می‌کنند. آن‌ها می‌توانند به اندیشه‌های کلی براساس اطلاعاتی که حفظ کرده و به آن معنا بخشیده‌اند، دست یابند. به کار بستن، شامل استفاده از روش‌ها و روندها برای انجام تمرین‌ها یا حل کردن مسئله‌هاست؛ بنابراین به کار بستن بسیار وابسته به دانش روندی است و شامل دو خرده طبقه زیر می‌شود:

- اجرا کردن. استفاده از روشی برای انجام یک تکلیف آشنا. این همان انجام تمرین است؛ بنابراین ویژگی مهم این خرده طبقه، آشنایی دانش‌آموز با تمرین‌هایی است که انجام می‌دهد. اندرسون، کراتول و همکاران در توضیح این خرده طبقه گفته‌اند: «اجرا کردن، بیشتر به استفاده از مهارت‌ها و الگوریتم‌ها وابسته است تا تکنیک‌ها و

روش‌ها، مهارت‌ها و الگوریتم‌ها دارای دو کیفیت‌اند که آن‌ها را به‌اجرا وابسته می‌سازد. نخست اینکه آن‌ها از یک رشته مرحله تشکیل می‌شوند که ترتیب ثابتی دارند. دوم اینکه وقتی مراحل به‌درستی طی شوند، نتیجه نهایی، قابل پیش‌بینی است. اصطلاح دیگر برای اجراکردن، از مرحله طرح به‌عمل درآوردن است.

مثال. دانش‌آموز بتواند یک عدد صحیح را بر عدد صحیح دیگر تقسیم کند.

- انجام‌دادن (مورد استفاده قرار دادن). انتخاب یک روش یا روند و استفاده از آن برای انجام یک تکلیف ناآشنا. برخلاف اجراکردن که تمرین کردن است، خرده‌طبقه حاضر (انجام‌دادن)، حل مسئله است.

مثال. دانش‌آموز بتواند بدون استفاده از ماشین‌حساب، مسائل ریاضی خود را حل کند.

یادگیرنده باید بتواند اصول تقویت را در آموزش، به‌کاربندد.

در پایان درس امروز، دانش‌آموزان می‌توانند وضعیت زاویه‌های مساوی و روبه‌رو در متوازی‌الاضلاع را با اندازه‌گیری اضلاع سقف خانه که به‌صورت متوازی‌الاضلاع است، حل کنند.

۴. تحلیل کردن. هدف‌های نوشته‌شده در سطح تحلیل، نیازمند تشخیص خطاهای منطقی (برای مثال، تضاد یا استنباط اشتباه را دریابد) یا واقعیت‌ها، دیدگاه‌ها، فرض‌ها، فرضیه‌ها، نتیجه‌گیری‌ها به‌وسیله دانش‌آموز است. هدف این بخش، تجزیه و تحلیل همه اطلاعات به‌دست‌آمده از انواع مقوله‌های کاربردی یا بخش‌های متفاوت به‌وسیله یادگیرنده است. بدین معنا که مطالب گردآوری‌شده، روان‌تر و ساده‌تر بیان شود.

- متمایز کردن. جداکردن اجزاء از یک ساخت کلی برحسب ارتباط با اهمیت آن‌ها. متمایز کردن یا ویژه‌سازی کردن وقتی رخ می‌دهد که دانش‌آموز، اطلاعات مربوط از نامربوط یا مهم از غیرمهم را تشخیص می‌دهد و سپس به اطلاعات مربوط یا مهم توجه می‌کند. برای مثال در متمایز کردن سیب از پرتقال به‌عنوان میوه، دانه‌های درون آن‌ها مربوط، اما شکل و رنگ نامربوط‌اند. در مقایسه کردن که یکی از خرده‌طبقه‌های فهمیدن است، همه این جنبه‌ها (یعنی دانه، رنگ و شکل) مربوط‌اند، بنابراین متمایز کردن و مقایسه کردن، دو فرایند شناختی متفاوت‌اند.

مثال. دانش‌آموز بتواند نکات عمده گزارش‌های تحقیقی را مشخص سازد.

- سازمان‌دادن. تشخیص عناصر یک ارتباط و شناختن اینکه چگونه آن عناصر، در ایجاد یک کل یکپارچه، به هم پیوند می‌یابند. سازمان‌دادن معمولاً همراه با متمایزکردن اتفاق می‌افتد. ابتدا دانش‌آموز عناصر مربوط یا مهم را تشخیص می‌دهد و سپس ساختار کل را که در آن، عناصر به هم پیوند می‌خورند، تعیین می‌کند.

مثال. دانش‌آموز بتواند با توجه به شواهد موجود در یک توصیف تاریخی، به شواهد لازم له و علیه یک تبیین تاریخی خاص سازمان‌دهد.

- نسبت‌دادن. مشخص کردن دیدگاه، سوگیری‌ها، ارزش‌ها یا قصدهای زیربنایی ارتباط‌ها یا اثرها. نسبت‌دادن یا واسازی، مستلزم فرایند ساختارشکنی است که در آن، دانش‌آموز هدف‌های نویسنده را از ارائه مطالب تعیین می‌کند. اندرسون، کراتول و همکاران، در توضیح این خرده‌طبقه می‌گویند: «در مقایسه با تفسیرکردن که در آن دانش‌آموز می‌کوشد معنی مطالب ارائه‌شده را بفهمد، نسبت‌دادن فراتر از فهمیدن است، زیرا در آن دانش‌آموز می‌کوشد قصد یا دیدگاه نویسنده را که زیربنای مطلب ارائه‌شده است استنباط کند».

مثال. یادگیرنده بتواند در یک مقاله، ایده‌ها را از واقعیت‌ها تمیز دهد.

یادگیرنده بتواند هدف اصلی یک مطلب داستانی را تشخیص دهد و به‌زیان بیاورد.

۵. ارزشیابی کردن. توانایی داوری یا قضاوت کمی و کیفی دربارهٔ امور، با توجه به ملاک‌های معین. در این بخش، از دانش‌آموزان خواسته می‌شود تمام اطلاعات فراگرفته را یکجا به‌ذهن بسپارند و به‌بررسی و قضاوت دربارهٔ اطلاعات گردآوری‌شده بپردازند.

- واریسی کردن. آزمون همخوانی‌ها و ناهمخوانی‌های یک اثر یا محصول، برای مثال وقتی دانش‌آموز برای تعیین این موضوع می‌کوشد که آیا شواهد از یک فرضیه دفاع می‌کنند یا یک مطلب، از بخش‌هایی تشکیل شده‌است که همدیگر را نقض می‌کنند، فعالیت او، در این خرده‌طبقه جای می‌گیرد.

مثال. دانش‌آموز بتواند در پیام‌های ترغیب‌کننده، ناهمخوانی‌ها را تبیین کند.

دانش‌آموز با ارائه توصیفی از نظام اقتصادی کشور براساس اصولی که درباره اقتصاد پویا خوانده‌است، نقاط قوت و ضعف آن را ارزیابی کند.

- نقد کردن. داوری کردن درباره یک عمل یا بازده براساس معیارها یا استانداردهای بیرونی. در نقد کردن، دانش‌آموز متوجه جنبه‌های مثبت و منفی یک بازده می‌شود و براساس آن‌ها درباره آن بازده داوری می‌کند.

۶. آفریدن. آفریدن، همان خلاقیت است. هدف‌های نوشته‌شده در سطح آفریدن، نیازمند تولید چیزی منحصر به فرد یا اصیل به وسیله دانش‌آموز است. سؤال‌های سطح آفریدن، نیازمند آن است که دانش‌آموز مسائل ناآشنا را به شیوه‌ای منحصر به فرد، حل یا بخش‌ها را به گونه‌ای ترکیب کند که چیز جدیدی حاصل شود.

مثال. دانش‌آموز با در دست داشتن یک داستان، در مدت زمان ۱۵ دقیقه، یک پایان متفاوت، اما قابل قبول برای داستان خواهد نوشت.

بدون کمک گرفتن از کسی، یادگیرنده بتواند برای تدریس یک موضوع، یک طرح آموزشی ابتکاری ارائه دهد.

یادآوری ۱. ابعاد دانش اندرسون و کراتول عبارت‌اند از: دانش واقعی، دانش مفهومی، دانش روندی و دانش فراشناختی.

یادآوری ۲. ابعاد فرایند شناختی اندرسون و کراتول عبارت‌اند از: به یاد آوردن، فهمیدن، به کار بستن، تحلیل کردن، ارزشیابی و آفریدن.

۲-۴-۴ طبقه‌بندی مارزانو و کندال

این طبقه‌بندی، در سال ۲۰۰۷ به وسیله مارزانو و کندال ارائه شد. در این الگو، سه نظام ذهنی^۱ وجود دارد: نظام شخصی^۲، نظام فراشناختی^۳ و نظام شناختی^۴. مؤلفه چهارم در این الگو، دانش است. نظام شخصی، شامل چهارچوبی از اعتقادات و هدف‌های مرتبط به هم است که در مورد مداخله در یک تکلیف جدید،

1. Mental System
2. Personal System
3. Metacognitive System
4. Cognitive System

مصلحت‌اندیشی می‌کند. نظام شخصی، تعیین‌کننده اصلی میزان انگیزش فرد برای انجام یک تکلیف است. اگر تکلیفی مهم ارزیابی شود، احتمال موفقیت در آن نیز بالا خواهد بود و به لحاظ هیجانی، اثر مثبتی تولید خواهد کرد و فرد برای انجام آن، برانگیخته خواهد شد.

به محض اینکه ورود به یک تکلیف تصویب شد، نظام فراشناختی شروع به فعالیت می‌کند. یکی از اولین اقداماتی که نظام فراشناختی انجام می‌دهد، تعیین هدف‌های مرتبط با تکلیف جدید است. این نظام، راهبردهایی مورد نیاز برای دستیابی به هدف‌های تعیین شده را نیز طراحی می‌کند. نظام فراشناختی، به‌طور مداوم با نظام شناختی فرد در تعامل است.

نظام شناختی، مسئول پردازش اثربخش اطلاعاتی است که برای به‌انجام‌رساندن تکلیف، ضروری است. این نظام، عملیات تحلیلی نظیر استنباط، مقایسه، طبقه‌بندی و ترجیح‌دادن اطلاعات موجود را هدایت می‌کند. در نهایت، موفقیت در انجام یک تکلیف، تا حد زیادی به میزان دانش فرد در مورد آن تکلیف وابسته است. هرچه دانش فرد از یک موضوع کامل‌تر باشد، نظام‌های شخصی، فراشناختی و شناختی، وظایف خود را در مورد تکلیف مربوط به آن، بهتر و کارآمدتر انجام می‌دهند.

- **حوزه عاطفی.** موضوعات مربوط به این حوزه معمولاً درباره نگرش‌ها و عواطف دانش‌آموزان است. حیطه عاطفی، شامل رفتارهایی است که به علائق، احساسات، ارزش‌ها، اخلاقیات و عواطف مربوط می‌شوند. وسعت و دامنه رفتارها در حیطه عاطفی زیاد است و ارزشیابی درجه تحقق‌بخشیدن به هدف‌های رفتاری در آن مشکل‌تر است، زیرا سنجش میزان توجه، نگرش و عمق ارزش کسب‌شده، کار لحظه‌ای نیست و در بسیاری از موارد، نیازمند مراقبت مستمر و تجربه طولانی از شخص است. برای یک معلم، سنجش میزان نگرش و احساسات دانش‌آموزانش بسیار دشوار است. برای مفید واقع شدن این حوزه در مدرسه باید برای دانش‌آموزان، فرصت‌های قابل دستیابی فراهم شود.

این حوزه، دارای طبقاتی به شرح زیر است:

۱. دریافت کردن^۱ (توجه کردن^۲). پاسخ‌ها در سطح دریافت کردن، نیازمند داشتن حداقل آگاهی دانش‌آموز درباره محرک‌هاست؛ بدین معنی که آگاهانه به سمت

محركی گرایش پیدا کند و به صورت انتخابی، به آن توجه نشان دهد. این دسته از هدف‌ها شامل رفتارهایی است که از وجود پدیده خبر می‌دهد. توجه به یک پدیده، نخستین گام برای طی کردن مرحله‌های بعدی است. دریافت و توجه در تدریس، عبارت است از جلب، حفظ و هدایت توجه شاگرد نسبت به فعالیت آموزشی.

۲. پاسخ‌دادن^۲: دانش‌آموز در سطح پاسخ‌دهی، فعالانه به محرك‌ها توجه می‌کند و در فرایند یادگیری شرکت فعال دارد. در این سطح، از شاگرد انتظار می‌رود احساس خود را فعالانه نمایان کند. در حقیقت شخص باید واکنشی درباره مسئله مورد نظرش نشان دهد. دامنه سطح پاسخ‌دادن، از پیروی و متقاعد شدن شروع می‌شود و تا تمایل به پاسخ‌دادن و اقدام داوطلبانه و سرانجام پاسخ همراه با رضایت و خشنودی پیش می‌رود. بالاترین سطح این مقوله، احساس رضایت آزمودنی بعد از مشارکت در فعالیت‌هاست (شعبانی، ۱۳۹۵).

۳. ارزش‌گذاردن^۳: این سطح، مبتنی بر درونی‌کردن دسته معینی از ارزش‌هاست. بنیادی‌ترین زیرسطح، مستلزم پذیرش^۴ یک باور، عقیده، نگرش و مانند آن است. هنگامی که فرد، ایده‌ای را دنبال می‌کند، آن را بر بقیه ترجیح^۵ می‌دهد.

۴. سازمان‌دادن^۶: با درونی‌شدن ارزش‌ها آن‌ها به صورت یک نظام ارزشی، سازمان داده می‌شوند. در واقع فرد پس از پذیرش ارزش‌ها، ناگزیر از تطبیق یا سازمان‌دهی ارزش‌های مختلف است. در این سطح، ارزش‌های مهم‌تر در سطح بالاتری قرار می‌گیرند (شعبانی، ۱۳۹۵).

الف) تلفیق ارزش‌های مختلف به وسیله درک و سازمان‌دهی آن‌ها

ب) حل و رفع تناقضات موجود بین ارزش‌های مختلف

ج) ایجاد یک نظام ارزشی پایدار و منسجم (رومی‌سوزوسکی، ترجمه فردانش هاشمی، ۱۳۹۳).

۵. تشخیص به وسیله یک ارزش یا مجموعه‌ای از ارزش‌ها^۱. در این سطح، دانش‌آموز مطابق ارزش‌هایش عمل می‌کند. وی در درون خود، سلسله‌مراتبی از ارزش‌ها به وجود آورده است که رفتار فردی‌اش را کنترل می‌کند.

حالا که با طبقه‌بندی عاطفی آشنا شدید، ممکن است بپرسید: «من در کلاس، چه کاری می‌توانم با این طبقه‌بندی‌ها انجام دهم؟» در ابتدا، محدود کردن همه تجارب دانش‌آموز از سوی معلم به این پنج سطح در طول سال تحصیلی، منطقی نیست. معلم، قادر به انجام اقدامات سودمند زیادی است. تشخیص درست معلم، راهگشای حل مشکلات انگیزشی دانش‌آموز خواهد بود.

مثال. فرض کنید بیشتر کودکان کلاس شما علاقه‌مند به تاریخ نیستند و از تمام محرک‌های مربوط به آن اجتناب می‌کنند. اگر از آن‌ها بخواهید مقاله‌ای درباره اهمیت درس تاریخ بنویسند، هم برای آن‌ها و هم شما تجربه ناموفقی خواهد بود، زیرا آن‌ها حتی به سطح دریافت یا توجه کردن نرسیده‌اند. دادن مقاله به آن‌ها کمکی نمی‌کند و نتیجه‌ای ندارد. به جای آن، در ابتدا سعی در جلب توجه آن‌ها به این درس با استفاده از مثال‌های جذاب داشته باشید. در صورت موفقیت در این روش، آنگاه دانش‌آموزان به توضیحات شما توجه می‌کنند و ممکن است به آن علاقه‌مند شوند و مطالعات غیردرسی خود را روی تاریخ متمرکز کنند.

- **حوزه روانی - حرکتی.** آخرین حوزه مطرح شده بلوم، حوزه روانی - حرکتی^۲ است. این حوزه، در عمل همه رفتارها از جمله: صحبت کردن، نوشتن، خوردن، پریدن، پرتاب کردن، گرفتن، قدم زدن، رانندگی کردن، باز کردن در، پرواز با هواپیما و همانند آن را دربرمی‌گیرد. این حوزه، به پرورش حرکات کلی بدن، آموزش موسیقی، هنر و انواع آموزش‌های حرفه‌ای از قبیل کشاورزی و تجارت مربوط می‌شود. چهار بخش از یادگیری برای این حوزه تعیین شده است که عبارت‌اند از: حرکات بنیادی (کنترل کم از سوی دیگران)، حرکات کلی (حرکات ضروری)، حرکات معمولی (انواع کنترل و بداهه‌گویی‌ها)، حرکات خلاق (حرکات انحصاری یک دانش‌آموز). برای این حوزه نیز سلسله‌مراتبی

- وجود دارد و دانش‌آموزان پیش از گذراندن سطوح اول (ادراک) و دوم (آمادگی)، قادر به گذراندن سطوح دیگر نیستند.
- براساس نظر سیمپسون، حوزه روانی-حرکتی، شامل هفت مرحله زیر است:
- ادراک. استفاده از حواس برای هدایت کنش‌های حرکتی.
 - آمادگی. آماده‌بودن برای انجام یک عمل.
 - پاسخ هدایت‌شده. عمل کردن به کمک هدایت یک الگو.
 - مکانیسم. انجام اعمالی که نسبتاً مشخص و عاداتی هستند.
 - پاسخ پیچیده آشکار. ترکیب تعدادی از عادت‌ها.
 - انطباق. استفاده از مهارت‌های از پیش آموخته‌شده برای انجام تکالیف تازه، اما مربوط به گذشته.
 - ابتکار. خلق الگوهای تازه حرکتی برای حل مسائل غیرمعمول.

یادآوری ۱. در طبقه‌بندی مارزاتو و کندال (۲۰۰۷)، سه نظام ذهنی وجود دارد: نظام شخصی، نظام فراشناختی و نظام شناختی.

یادآوری ۲. حوزه عاطفی مارزاتو و کندال (۲۰۰۷)، شامل دریافت کردن (توجه کردن)، پاسخ دادن، ارزش‌گذاردن، سازمان‌دادن و تشخیص به‌وسیله یک ارزش یا مجموعه‌ای از ارزش‌هاست.

یادآوری ۳. حوزه روانی-حرکتی بلوم، شامل همه رفتارها از جمله: صحبت کردن، نوشتن، خوردن، پریدن، پرتاب کردن، گرفتن، قدم زدن، رانندگی کردن، بازکردن در، پرواز با هواپیما و مانند آن است.

۲-۵ نقشه آزمون یا جدول مشخصات

نوشتن سؤال‌های آزمون، نیازمند صرف وقت و دقت است. سؤال‌های آزمون باید فرایندهای شناختی سطوح پایین و بالای هدف‌های آموزشی را پوشش دهند. یکی از ابزارهای راهگشا در این امر، نقشه آزمون^۱ است. این نقشه، بسیار شبیه نقشه‌ای است

که بناها از آن استفاده می‌کنند. همان‌طور که نقشه ساختمان، راهنمای بنا در ساخت آن است، نقشه آزمون نیز در ساخت آزمون به معلم کمک می‌کند. نقشه آزمون، جدول مشخصات^۱ نیز نامیده می‌شود.

نقشه ساختمان، مانع چشم‌پوشی بنا از جزئیات ضروری نقشه می‌شود. بدین ترتیب نقشه آزمون هم تضمین می‌کند که مانع غفلت معلم از جزئیات ضروری برای ساخت یک آزمون خوب شود. جدول مشخصات کمک می‌کند:

۱. سؤالات مورد استفاده معلم در آزمون، در راستا و سطح مطالب تدریس‌شده در کلاس باشد.
۲. فرایندهای شناختی مد نظر مهم‌اند.
۳. از کل محتوای کتاب، سؤال طرح شده‌است.

۴. سؤال‌هایی را شامل می‌شود که از نظر پیچیدگی شناختی، متفاوت هستند. به عبارت دیگر، آزمون هم سؤال ساده دارد، هم متوسط و هم دشوار. به کارگیری سؤال‌هایی از سطح به یاد آوردن تا پیچیده‌ترین سطح، یعنی سطح آفریدن.

نقشه آزمون یا جدول مشخصات، یک جدول دوبعدی است با تعدادی ردیف و ستون، متناسب با محتوا و هدف‌های آموزشی درس. در بعد هدف که در ردیف بالای جدول نوشته می‌شود، سطوح مختلف هدف‌های آموزشی و در ردیف محتوا، همه بخش‌ها و فصل‌های کتاب یا درسی که قرار است سنجش‌شود، نوشته می‌شوند.

یک آزمون خوب، به‌بهترین شکل، منعکس‌کننده تمام هدف‌های آموزشی و همه محتوای برنامه درسی است. در صورتی که آزمون، دارای چنین مشخصه‌ای باشد، می‌توان آن را یک آزمون دارای روایی مناسب دانست. جدول مشخصات زیر، به روشن شدن بیشتر موضوع، کمک خواهد کرد.

۲-۵-۱ بعد محتوا

بعد محتوا شامل عنوان و هدف‌های آموزشی مهمی است که زیر آن عنوان قرار می‌گیرند. سؤال‌های آزمون، براساس این هدف‌های آموزشی نوشته خواهند شد.

جدول ۲-۴. جدول مشخصات آزمون برای بخش هدف‌های آموزشی

ردیف	نوع هدف	بعد هدف					بعد محتوا
		یادآوردن	فهمیدن	به کار بستن	تعمیل کردن	ارزیابی	
۱	تقش هدف‌های آموزشی						
۱۱/۲۲	الف) دانش آموز بتواند کاربردهای هدف‌های آموزشی را توضیح دهد.	۴				۴	
۲۲/۸۵	ب) دانش آموز بتواند الگوی نظام کلاسی و تقش هدف‌های آموزشی در آن را توضیح دهد.		۱			۱	
۲	نوشتن هدف‌های آموزشی						
۱۳/۲۹	الف) با توجه به یک هدف آموزشی کلی، دانش آموز بتواند یک هدف آموزشی بنویسد که آن هدف کلی، در آن گنجانده شده باشد.			۵		۵	
۱۳/۲۹	ب) دانش آموز بتواند هدف‌های آموزشی را با سطح مناسب آن‌ها در حوزه شناختی هماهنگ سازد.		۵			۵	
۱۳/۲۹	ج) دانش آموز بتواند سه بخش یک هدف آموزشی (فعل کنشی، شرایط، ملاک‌ها) را مشخص سازد.		۵			۵	
۲۸/۵۷	د) دانش آموز بتواند هنگام ارائه مثال‌هایی از هدف‌های آموزشی و فعالیت آموزشی، آن‌ها را از همدیگر تشخیص دهد.		۱۰			۱۰	
۳	رمزگشایی از هدف‌های آماده						
۱۳/۲۹	یا در اختیار داشتن هدف‌های آموزشی که نیاز به اصلاح دارند، دانش آموز خواهد توانست هدف‌ها را به گونه‌ای بازنویسی کند که به یک هدف آموزشی مناسب تبدیل شود.				۵	۵	
جمع		۴	۲۱	۵	۵	۳۵	
درصد		۱۱/۲۲	۶۰/۲۱	۱۳/۲۹	۱۳/۲۹	۱۰۰/۳۵	

۲-۵-۲ بعد هدف (طبقات)

طبقات، به عنوان یادآور یا معیار، در برابر «پیچیدگی شناختی» آزمون عمل می‌

بدریهی است بسیاری از واحدهای مورد آزمون، شامل آن نوع هدف‌های آموزشی هستند که از سطح فهمیدن تجاوز نمی‌کنند. با وجود این، بعد هدف عنوان کمک‌می‌کند که بتوانید سطوح یادگیری بالاتر را به آموزش و ارزشیابی‌تان اضافه کنید. در خانه‌های طبقات جدول، سؤال‌های آزمون شما برای یک هدف آموزشی خاص که در آن سطح قرار دارند، نشان داده می‌شوند. برای مثال، در جدول ۲-۵، برای اندازه‌گیری سطح فهمیدن، پنج سؤال باید نوشته شود.

۲-۵-۳ تعداد سؤال‌ها

خانه‌های جدول ۲-۴ را به کمک روش‌های زیر پر کنید:

۱. طبقه‌بندی هر هدف آموزشی را تعیین کنید.
۲. تعداد سؤال‌هایی که بایستی برای هر هدف در هر خانه تهیه شود، باید مطابق با آن طبقه‌ای باشد که به آن تعلق دارد. برای مثال در بعد محتوا (جدول ۲-۵)، بند ۱ قسمت الف که قرار است چهار سؤال در سطح به یاد آوردن نوشته شود، باید هر چهار سؤال در خانه‌ای نوشته شوند که زیر طبقه به یاد آوردن قرار دارد.
۳. گام ۱ و ۲ را برای همه هدف‌های آموزشی در بعد محتوا تکرار کنید.
۴. تعداد سؤال‌های هر هدف آموزشی را جمع بزنید و زیر ستون «جمع» بنویسید.
۵. گام ۱ تا ۴ را برای هر موضوع تکرار کنید.
۶. تعداد سؤال‌های مربوط به هر سطح را جمع بزنید و در ردیف «جمع» در پایین جدول بنویسید.
۷. درصد ستون‌ها و سطرها را جمع بزنید تا درصد کل به دست آید.

یادآوری ۱. سؤال‌های یک آزمون باید فرایندهای شناختی سطوح پایین و بالای هدف‌های آموزشی را پوشش دهند.

یادآوری ۲. نقشه آزمون، جدول مشخصات نیز نامیده می‌شود.

یادآوری ۳. در بعد هدف نقشه آزمون، سطوح مختلف هدف‌های آموزشی و در ردیف محتوا، همه بخش‌ها و فصل‌های کتاب یا درسی که قرار است سنجش شود، نوشته می‌شوند.

خلاصه فصل دوم

غایت‌های آموزشی و هدف‌های آموزشی، راهنمای تدریس و یادگیری‌اند. یک هدف آموزشی موارد زیر را شامل می‌شود: رفتار قابل مشاهده، شرایط و سطح عملکردی که برای رسیدن به حد تسلط، مد نظر است. یک هدف آموزشی باید شامل یک فعل کنشی باشد که هدف آموزشی را به‌صراحت مشخص‌سازد و توصیف‌کننده هر نوع شرایط خاصی باشد که یادگیری در آن به‌وقوع خواهد پیوست و چگونگی انجام رفتار مورد پذیرش را نشان دهد.

قدیمی‌ترین طبقه‌بندی هدف‌های آموزشی، مربوط به بلوم و همکاران (۱۹۵۶) است. سپس اندرسون و کراتول (۲۰۰۱)، در آن تجدیدنظر به‌عمل آوردند. مارزانو و کندال (۲۰۰۷) نیز طبقه‌بندی خاص خودشان را معرفی کردند. در طبقه‌بندی بلوم و همکاران، هدف‌های آموزشی، به سه حوزه شناختی، عاطفی و روانی - حرکتی تقسیم شده‌اند.

حوزه شناختی. طبقه‌بندی بلوم از هدف‌های حوزه شناختی، شش سطح دارد که عبارت‌اند از: دانش، درک، به‌کار بستن، تحلیل، ترکیب و ارزشیابی. در طبقه‌بندی اندرسون و همکاران (۲۰۰۱)، یک بعد جدید به طبقه‌بندی بلوم اضافه شده است. این بعد، شامل انواع دانش است که عبارت‌اند از: دانش امور واقعی، دانش مفهومی، دانش روندی و دانش فراشناختی. طبقه‌بندی تجدیدنظرشده، یک ابزار قابل اعتماد برای برنامه‌ریزی آموزشی، انتقال یادگیری و ارزیابی است و طیف گسترده‌تری از مخاطبان را هدف قرار می‌دهد.

ابعاد دانش اندرسون و کراتول عبارت‌اند از: دانش واقعی، دانش مفهومی، دانش روندی و دانش فراشناختی. ابعاد فرایند شناختی اندرسون و کراتول نیز عبارت‌اند از: به‌یاد آوردن، فهمیدن، به‌کار بستن، تحلیل کردن، ارزشیابی کردن و آفریدن. در طبقه‌بندی مارزانو و کندال (۲۰۰۷)، سه نظام ذهنی وجود دارد که عبارت‌اند از: نظام شخصی، نظام فراشناختی و نظام شناختی.

حوزه عاطفی. موضوعات مربوط به حوزه عاطفی، شامل رفتارهایی است که به‌علاوه احساسات، ارزش‌ها، اخلاقیات و عواطف مربوط می‌شوند. این حوزه، دارای طبقات به‌شرح زیر است: دریافت کردن (توجه کردن)، پاسخ دادن، ارزش‌گذاران، سازمان دادن، تشخیص به‌وسیله یک ارزش یا مجموعه‌ای از ارزش‌ها.

حوزه روانی - حرکتی. حوزه روانی - حرکتی در عمل، همه رفتارها از جمله صحبت کردن، نوشتن، خوردن، پریدن، پرتاب کردن، گرفتن، قدم زدن، رانندگی، باز کردن در، پرواز با هواپیما و مانند آن را دربرمی‌گیرد. این حوزه، به پرورش حرکات کلی بدن، آموزش موسیقی، هنر و انواع آموزش‌های حرفه‌ای از قبیل کشاورزی و تجارت مربوط می‌شود. چهار بخش از یادگیری، برای این حوزه تعیین شده است که عبارت‌اند از: حرکات بنیادی (کنترل کم از سوی دیگران)، حرکات کلی (حرکات ضروری)، حرکات معمولی (انواع کنترل و بداهه‌گویی‌ها)، حرکات خلاق (حرکات انحصاری یک دانش‌آموز).

نقشه آزمون یا جدول مشخصات. نوشتن سؤال‌های آزمون، نیازمند صرف وقت و دقت است. سؤال‌های آزمون باید فزاینده‌ای شناختی سطوح پایین و بالای هدف‌های آموزشی را پوشش‌دهند. یکی از ابزارهایی که به این امر کمک می‌کند، نقشه آزمون است. نقشه آزمون نیز در ساخت آزمون به معلم کمک می‌کند و جدول مشخصات نیز نامیده می‌شود.

خودآزمایی چهارگزینه‌ای فصل دوم

۱. در نوشتن هدف‌های آموزشی، کدام گزینه صحیح است؟
 - الف) هدف آموزشی باید در قالب عبارات رفتار غیرقابل اندازه‌گیری بیان شوند.
 - ب) هدف آموزشی باید به گونه‌ای نوشته شود که دو دانش‌آموز، برداشت متفاوتی از آن داشته باشند.
 - ج) هدف آموزشی باید به گونه‌ای نوشته شود که پیامد یادگیری را نشان دهد.
 - د) شرایط ویژه برای رسیدن به حد تسلط، مورد نیاز است.
۲. در مورد شناسایی پیامدهای یادگیری، کدام گزینه صحیح است؟
 - الف) همه افعال کنشی، پیامدهای یادگیری را به صراحت مشخص سازند.
 - ب) یک هدف آموزشی، شامل یک فعل کنشی است.
 - ج) پیامدهای یادگیری یا فعالیت‌های یادگیری، یکسان تلقی می‌شوند.
 - د) مثال «کودک نقد شعر را توضیح دهد»، به فعالیت یادگیری اشاره دارد.

۳. کدام یک از گزینه‌های زیر، یک فعالیت یادگیری است؟
 الف) تمرین موسیقی
 ب) نصب رادیوی ماشین
 ج) دانش آموز، تقریب اعداد یک‌رقمی را با صحت ۸۰٪ انجام می‌دهد
 د) از حفظ گفتن ۴ مؤلفه انشای خوب
۴. سؤال‌های آزمون، بر مبنای کدام گزینه تدوین می‌شود؟
 الف) فعل کنشی
 ب) پیامد یادگیری
 ج) فعالیت یادگیری
 د) بازده نهایی
۵. کدام یک از پیامدهای زیر، غیرقابل مشاهده است؟
 الف) ساختن
 ب) کشیدن
 ج) دانستن
 د) فهرست کردن
۶. هر نوع شرایط خاصی که یادگیری در آن به وقوع خواهد پیوست، چه نامیده می‌شود؟
 الف) هدف آموزشی
 ب) بیان ملاک
 ج) بیان عملکرد
 د) ملاک عملکرد
۷. قدیمی‌ترین طبقه‌بندی هدف‌های آموزشی، مربوط به کیست؟
 الف) بلوم و همکاران
 ب) بنجامین
 ج) اندرسون و کراتول
 د) مارزانو و کندال
۸. کدام گزینه، مربوط به هدف‌های حوزه شناختی بلوم و همکاران نیست؟
 الف) دانش
 ب) آفریدن
 ج) کار بستن
 د) ارزشیابی
۹. در طبقه‌بندی حوزه شناختی اندرسون و همکاران، کدام طبقه به طبقه‌بندی بلوم همکاران اضافه شد؟
 الف) انواع کاریست
 ب) انواع فهمیدن
 ج) انواع دانش
 د) تحلیل و ترکیب
۱۰. در طبقه‌بندی تجدیدنظرشده اندرسون و کراتول، به فعل تغییر شکل داده است؟
 الف) دانش، آفریدن
 ب) ترکیب، فرایندهای تفکر
 ج) به یاد آوردن، دانش
 د) درک، فهمیدن

۱۱. طبقه‌بندی آموزشی بلوم، برای کدام مقطع تحصیلی، بیشترین کاربرد را دارد؟
 الف) دبستان
 ب) راهنمایی
 ج) دبیرستان
 د) دانشگاهی
۱۲. کدام مورد، شامل استفاده از روش‌ها و روندها برای انجام تمرین‌ها یا حل کردن مسئله است؟
 الف) دانش
 ب) فهمیدن
 ج) به‌کار بستن
 د) ارزیابی کردن
۱۳. کدام گزینه، شامل توانایی قضاوت، بررسی و حتی نقد ارزشی مواد به‌منظور دستیابی به یک هدف معین است؟
 الف) کاربرد
 ب) تحلیل
 ج) ترکیب
 د) ارزیابی
۱۴. براساس ابعاد دانش اندرسون و کراتول، دانش عناصر مورد نیاز یادگیرنده برای آشناسدن با یک رشته علمی یا حل مسائل مربوط به آن، چه دانشی است؟
 الف) واقعی
 ب) مفهومی
 ج) روند
 د) فراشناختی
۱۵. کدام بعد از ابعاد فرایند شناختی اندرسون و کراتول، معادل ترجمه کردن است؟
 الف) به‌یاد آوردن
 ب) فهمیدن
 ج) تفسیر کردن
 د) طبقه‌بندی کردن
۱۶. واریسی کردن و نقد کردن، مربوط به کدام بعد از ابعاد شناختی اندرسون و کراتول هستند؟
 الف) فهمیدن
 ب) به‌کار بستن
 ج) تحلیل کردن
 د) ارزشیابی کردن
۱۷. کدام نظام، شامل چهارچوبی از اعتقادات و هدف‌های مرتبط به هم است که در مورد مداخله یک تکلیف جدید، مصلحت‌اندیشی می‌کند؟
 الف) نظام شخصی
 ب) نظام فراشناختی
 ج) الگوی دانش
 د) نظام شناختی

۱۸. در کدام حوزه، دانش‌آموز فعالانه به محرک‌ها توجه می‌کند و در فرایند یادگیری شرکت فعال دارد؟
 الف) ارزش‌گذاری
 ب) سازمان‌دادن
 ج) دریافت‌کردن
 د) پاسخ‌دادن
۱۹. کدام حوزه، شامل رفتارهای صحبت‌کردن، نوشتن و پرتاب‌کردن می‌شود؟
 الف) حوزه شناختی
 ب) حوزه روانی-حرکتی
 ج) حوزه فراشناختی
 د) حوزه عاطفی
۲۰. کدام گزینه در مورد جدول مشخصات، صحیح است؟
 الف) سؤال‌های آزمون فقط مربوط به هدف‌های سطح بالای شناختی هستند.
 ب) نقشه ساختمان، باعث غفلت از جزئیات خواهد شد.
 ج) فرایندهای شناختی ساده، مد نظرند.
 د) از کل محتوای کتاب، سؤال طرح خواهد شد.

خودآزمایی تشریحی فصل دوم

۱. منظور از هدف‌های آموزشی چیست و چه شرایطی را باید برای نوشتن آن در نظر گرفت؟
۲. پیامدهای یادگیری و فعالیت‌های یادگیری را با ارائه مثال شرح دهید؟
۳. منظور از پیامدهای یادگیری قابل مشاهده و قابل اندازه‌گیری چیست؟
۴. هر هدف آموزشی، چند جزء دارد؟ با بیان مثال، آن‌ها را توضیح دهید.
۵. طبقه‌بندی هدف‌های آموزشی بلوم (۱۹۵۶) را به اختصار در حوزه‌های شناختی عاطفی و روانی حرکتی با ارائه مثال شرح دهید؟
۶. طبقه‌بندی بلوم (۱۹۵۶) و اندرسون و کراتول (۲۰۰۱) را در حوزه شناختی همدیگر مقایسه کنید؟
۷. ابعاد دانش اندرسون و کراتول (۲۰۰۱) را نام ببرید و ۲ مورد را به اختصار توضیح دهید.
۸. ابعاد فرایند شناختی اندرسون و کراتول (۲۰۰۱) را نام ببرید؟
۹. نظام‌های موجود در طبقه‌بندی مارزانو و کندال (۲۰۰۷) را نام ببرید و به اختصار شرح دهید؟
۱۰. منظور از نقشه آزمون یا جدول مشخصات چیست؟ کاربردهای آن را بنویسید.

فصل سوم

آزمون‌های استاندارد شدهٔ هنجاری، ملاکی و آزمون‌های معلم ساخته

هدف کلی

تهیه و استاندارد کردن آزمون‌های هنجاری، ملاکی و معلم ساخته و مقایسهٔ آن‌ها با همدیگر

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعهٔ این فصل بتواند:

۱. هدف از تهیهٔ آزمون‌های هنجاری، ملاکی و معلم ساخته را بیان کند.
۲. مراحل استاندارد کردن آزمون‌های هنجاری را بنویسد.
۳. آزمون‌های هنجاری و ملاکی را با هم مقایسه کند.
۴. آزمون‌های استاندارد شده و کلاسی را با هم مقایسه کند.
۵. مزایای آزمون‌های معلم ساخته را بنویسد.

مقدمه

در این فصل، آزمون‌های هنجار مرجع، ملاک مرجع و معلم ساخته، مورد بحث قرار می‌گیرند. این آزمون‌ها اطلاعات متفاوتی به ما می‌دهند. هر سه آزمون، اطلاعاتی دربارهٔ عملکرد آزمودنی در اختیار ما می‌گذارند. پیش از انتخاب آزمون، آزمونگر باید تصمیم بگیرد که آیا تصمیم به استفاده از آزمون‌های هنجاری، ملاکی و معلم ساخته دارد یا ترکیبی از همهٔ این آزمون‌ها.

در گذشته بیشتر از آزمون‌های ملاکی و معلم‌ساخته استفاده می‌شد، ولی در حال حاضر استفاده از آزمون‌های هنجاری در مدارس عمومیت یافته‌است. آزمون‌های هنجاری، حوزه وسیعی از دانش را اندازه‌گیری می‌کنند و وضعیت پیشرفت عمومی یک فرد معین را در رابطه با عملکرد یک گروه شناخته‌شده از افراد نشان می‌دهند.

۳-۱ آزمون‌های استاندارد شده و هدف‌های آن‌ها

آزمون‌های استاندارد شده، روش‌های یکسانی برای اجرا و نمره‌گذاری دارند و اغلب این امکان را فراهم می‌سازند که عملکرد دانش‌آموز، با عملکرد دیگر دانش‌آموزان هم‌سن و هم‌پایه، در یک مقیاس ملی مقایسه شود (سانتراک^۱، ۲۰۱۱).

سانتراک (۲۰۱۱)، هدف‌هایی را برای آزمون‌های استاندارد برمی‌شمرد:

- درباره پیشرفت تحصیلی دانش‌آموزان، اطلاعاتی فراهم می‌کنند. آزمون‌های استاندارد، مجموعه‌ای از اطلاعات درباره چگونگی عملکرد دانش‌آموزان فراهم می‌کنند. ممکن است دانش‌آموزی در یک درس، نمره الف بگیرد، اما در یک آزمون استاندارد ملی، در سطح متوسط عمل کند. دانش‌آموز دیگری ممکن است در کلاس، نمره «ب» بگیرد، ولی در همان آزمون، عملکرد بسیار خوبی داشته باشد. بدون وجود یک معیار بیرونی مثل یک آزمون استاندارد، معلمان یک کلاس برای آگاهی از نحوه عملکرد دانش‌آموزانشان در مقایسه با دانش‌آموزان مناطق دیگر شهر، استان یا کشور، مشکل خواهند داشت.
- نقاط قوت یا ضعف دانش‌آموزان را مشخص می‌کنند. آزمون‌های استاندارد می‌توانند درباره شدت یا ضعف یادگیری دانش‌آموزان نیز اطلاعاتی فراهم کنند. با اجرای یک یا چند آزمون استاندارد می‌توان نقطه ضعف یادگیری دانش‌آموزان را در یک کلاس، به دقت تعیین کرد. هنگامی که از آزمون‌های استاندارد به منظور تشخیص استفاده می‌شود، اغلب به صورت انفرادی اجرا می‌شوند.
- ملاکی برای جایابی دانش‌آموزان در برنامه‌های خاص فراهم می‌کنند. آزمون‌های استاندارد می‌توانند برای تصمیم‌گیری درباره مجوز ورود یا عدم ورود دانش‌آموزی به یک دوره خاص، مورد استفاده قرار گیرند. در دبستان می‌توان اطلاعات لازم

آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم‌ساخته ۵۳

برای قرارداد دادن دانش‌آموزان در گروه‌های مختلف خواندن، با استفاده از یک آزمون استاندارد به دست آورد. در دبیرستان، برای تعیین اینکه یک دانش‌آموز در کدام کلاس ریاضی باید شرکت کند، می‌توان از یک آزمون استاندارد استفاده کرد. در بعضی موارد، برای ارزیابی اینکه آیا یک دانش‌آموز، مجاز به جهش از کلاسی به کلاس بالاتر یا فارغ‌التحصیل شدن است یا نه، از آزمون‌های استاندارد همراه با اطلاعات دیگر استفاده می‌شود. همچنین ممکن است برای تعیین مناسب بودن دانش‌آموزان برای مشاغل خاص، از آن‌ها آزمون استاندارد به عمل آید.

یادآوری ۱. آزمون‌های استاندارد، مجموعه‌ای از اطلاعات درباره چگونگی عملکرد دانش‌آموزان فراهم می‌کنند.

یادآوری ۲. آزمون‌های استاندارد می‌توانند درباره شدت یا ضعف یادگیری دانش‌آموزان نیز اطلاعاتی فراهم کنند.

یادآوری ۳. آزمون‌های استاندارد می‌توانند برای تصمیم‌گیری درباره اینکه آیا دانش‌آموزی، مجاز به ورود به یک دوره خاص است یا نه، مورد استفاده قرار گیرند.

۲-۳ تعریف آزمون‌های هنجاری و ملاکی

آزمون‌های استاندارد شده می‌توانند هنجار مرجع یا ملاک مرجع باشند. آزمون هنجاری یا هنجار مرجع، به آزمونی گفته می‌شود که در آن، عملکرد آزمودنی با عملکرد گروهی که از لحاظ صفت مورد اندازه‌گیری، شبیه آن هستند، مقایسه می‌شود (کوییزین و بوریچ، ۲۰۱۳). این آزمون‌ها، هم در اجرا و هم در نمره‌گذاری، شیوه عمل یکسانی دارند.

آزمون‌های هنجاری اغلب این امکان را فراهم می‌کنند که عملکرد یک دانش‌آموز با همسالان یا هم‌کلاسی‌هایش، بر مبنای یک ملاک ملی یا محلی مقایسه شود. برای مثال، یک آزمون استاندارد شده برای سنجش دانش و مهارت‌های علوم چهارم ابتدایی، باید به یک نمونه ملی از دانش‌آموزان چهارم ابتدایی داده شود. نمره‌های این نمونه معرفی (متشکل از هزاران دانش‌آموز پایه چهارم)، مبنایی برای مقایسه می‌شود. یک گروه هنجار باید شامل دانش‌آموزانی از شهرها، شهرک‌ها و روستاها، نواحی جغرافیایی

مختلف، مدارس دولتی و خصوصی، پسران و دختران و گروه‌های قومی و نژادی مختلف باشد. یک معلم براساس نمره دانش‌آموزی خاص در آزمون استاندارد شده علوم، می‌تواند تعیین کند که عملکرد دانش‌آموز، چه مقدار بالاتر، در حد یا پایین‌تر از نرم (هنجار) ملی است (سانتراک، ۲۰۱۱).

در این نوع آزمون‌ها ملاک نسبی اندازه‌گیری، مورد استفاده قرار می‌گیرد. در تفسیر آزمون‌های هنجاری، گروهی که مورد مقایسه با نمره فرد قرار می‌گیرند، گروه هنجار نامیده می‌شود. آزمون‌های هنجاری در سنجش هوش، استعداد، پیشرفت تحصیلی، رغبت، نگرش و نیز در سنجش شخصیت و سایر جنبه‌های عاطفی به کار می‌روند. در این آزمون‌ها هیچ‌گونه ملاک از پیش تعیین شده‌ای مورد استفاده واقع نمی‌شوند، بلکه در آن‌ها عملکردهای آزمون‌شوندگان با یکدیگر مقایسه می‌شوند. نمره‌های آزمون‌های هنجاری، به صورت توزیع طبیعی پراکنده می‌شوند. در این توزیع، بیشتر نمره‌ها در اطراف میانگین، تعداد کمی از آن‌ها بسیار بالاتر از میانگین و به همان نسبت هم تعداد کمی از نمره‌ها بسیار پایین‌تر از میانگین قرار دارند.

مثال. آزمون‌های کتکور سراسری دانشگاه‌ها و مسابقات علمی، نمونه‌هایی از آزمون‌های ملاک نسبی (هنجاری) به شمار می‌روند.

۳-۳ استاندارد کردن آزمون‌های هنجاری

استاندارد کردن آزمون، در آزمون‌های هنجاری، یک جریان وقت‌گیر، پرهزینه و مستلزم سه اقدام گسترده است:

۱. در نخستین گام، انتخاب سؤال‌های بیانگر حیطه مورد اندازه‌گیری از سوی آزمونگر، اجرای آن روی گروه وسیعی از آزمودنی‌ها و گنجاندن سؤال‌های مناسب، در آزمون نهایی (برای اطلاع از مناسب بودن سؤال‌های آزمون، به فصل ۲ مراجعه کنید).
۲. در دومین گام، انتخاب چندین هزار نفر از افرادی که باید نمونه مناسب و نمایانگر تمام ویژگی‌های جمعیت مورد نظر (برای مثال برحسب آخرین سرشماری کشوری از جمعیت دانش‌آموزی) باشند به منظور اجرای آزمون روی آن‌ها.
۳. در سومین گام، تحلیل داده‌های به دست آمده از اجرای آزمون و تبدیل آن‌ها به نمره‌های استاندارد.

۳-۴ آزمون‌های ملاکی

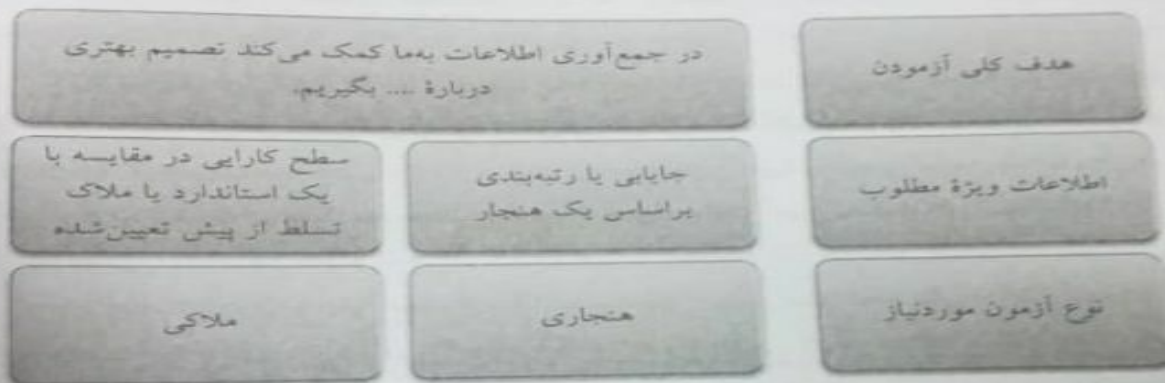
برخلاف آزمون‌های هنجار مرجع، آزمون‌های ملاکی یا ملاک مرجع^۱، آزمون‌های استاندارد هستند که در آن‌ها عملکرد دانش‌آموز با معیارها یا استانداردهای تعریف شده مقایسه می‌شود (مک میلان^۲، ۲۰۱۱). در این آزمون‌ها معیار عملکرد یا تسلط، از قبل تعیین شده است و عملکرد آزمودنی، با این معیار از پیش تعیین شده مقایسه می‌شود تا مشخص شود آزمودنی به هدف دست یافته است یا خیر؛ بنابراین از قبل، ملاک مطلق^۳ تعیین شده است و از آزمودنی انتظار می‌رود به آن هدف‌ها دست یابد.

آزمون‌های ملاکی، اغلب حوزه محدودی از هدف‌هایی را که به صورت ویژه مربوط به مواد تدریس شده‌اند، اندازه‌گیری می‌کنند (لطف‌آبادی، ۱۳۹۳). به این آزمون‌ها، آزمون‌های معطوف به هدف^۴ یا آزمون‌های حیطه‌ای^۵ نیز گفته می‌شود. با یک آزمون ملاکی می‌توان بدون مقایسه دانش‌آموز با سایر دانش‌آموزان، میزان تسلط او را در مهارت معینی تشخیص داد (کوهن^۶، ۲۰۰۹).

در این نوع آزمون‌ها معلمان برای سنجش پیشرفت تحصیلی دانش‌آموزان در طول آموزش یا پایان دوره، آن را به اجرا درمی‌آورند و در آن‌ها ملاک مطلق اندازه‌گیری به کار می‌رود. آزمون‌های ملاکی، برای سنجش مهارت و دانش دانش‌آموز در یک حوزه خاص مانند زبان انگلیسی، ریاضی و علوم طراحی می‌شوند.

درباره آزمون‌های هنجاری و ملاکی باید قبل از اجرای آزمون، نوع اطلاعات مورد نیاز را مشخص سازید. اگر قادر به انجام این کار نباشید، ممکن است با وجود در اختیار داشتن داده لازم، چگونگی استفاده از آن‌ها را ندانید. خیلی از معلمان، والدین و دیگران گزارش می‌کنند که بعد از اجرای آزمون در مقایسه با قبل از آن، چیزی به اطلاعات آن‌ها درباره دانش‌آموز اضافه نشده است. این اتفاق، زمانی رخ می‌دهد که تصمیم‌گیرندگان، هدف خود را قبل از انتخاب و اجرای یک آزمون، به روشنی مشخص نساخته‌اند.

در برخی مواقع معلمان به‌آگاهی از چگونگی عملکرد دانش‌آموزان در مقایسه با دیگران نیاز دارند. این اطلاعات، برای ارائه تصویر واقع‌بینانه‌ای از عملکرد دانش‌آموز به والدین و خود دانش‌آموزان اهمیت دارد. برای مثال، دانش‌آموزانی که مهارت‌های برجسته‌ای در علوم دارند، هم در موقعیت کلاس و مدرسه و هم در سطح ملی، باید از استثنایی بودن خود اطلاع داشته باشند. به‌طور کلی دانش‌آموزان، نیازمند داشتن برداشت دقیقی از توانمندی‌ها و ضعف‌های خود هستند تا بتوانند برای آینده، تصمیم‌گیری کنند.



شکل ۳-۱. روابط میان هدف آزمودن، اطلاعات مطلوب و نوع آزمون مورد نیاز (اقتباس از بوریج و کوییزین، ۲۰۱۳)

۳-۵ مقایسه آزمون‌های ملاکی و هنجاری

آزمون‌های ملاکی، اگر اطلاعاتی درباره مهارت‌ها به‌دست می‌دهند، باید خیلی اختصاصی باشند. این کار، مزایا و معایبی دارد. استفاده از آزمون خیلی اختصاصی باعث توانایی شما در تشخیص میزان تسلط دانش‌آموزانتان در مهارت خاص مورد سؤال می‌شود. ایراد عمده آزمون‌های ملاکی این است که بسیاری از چنین آزمون‌هایی لزوماً درباره تعداد زیادی از مهارت‌هایی تصمیم‌گیری می‌کنند که به‌نوعی در یک کلاس متوسط، آموزش داده می‌شوند. در مقابل، آزمون‌های هنجاری، گرایش به کلی بودن دارند. این آزمون‌ها مهارت‌های کلی و اختصاصی متنوعی را یکجا اندازه‌گیری می‌کنند، ولی در اندازه‌گیری آن‌ها به‌طور کامل ناتوان‌اند.

آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم‌ساخته ۵۷

در آزمون‌های هنجاری، تمرکز بر چگونگی عملکرد نسبی فرد یا گروه مورد مقایسه است، ولی در آزمون‌های ملاکی، تمرکز بر عملکرد آزمودنی‌هاست؛ اینکه آزمودنی‌ها چه کاری را می‌توانند و چه چیزی را نمی‌توانند انجام دهند؟ چه چیزی یاد گرفته‌اند و چه چیزی یاد نگرفته‌اند؟ آیا آزمودنی، به معیار مد نظر دست یافته یا دست نیافته است و مانند آن.

جدول ۳-۱. مقایسه آزمون‌های ملاکی و هنجاری

ملاکی	هنجاری	بعد
۸۰٪	۵۰٪	تعداد متوسط دانش‌آموزانی که به سؤال، پاسخ صحیح می‌دهند.
معیارهای نشانگر تسلط	عملکرد دانش‌آموزان دیگر	عملکرد دانش‌آموز را مقایسه می‌کند با
محدود، چند هدف را دربرمی‌گیرد.	گسترده، بسیاری از هدف‌ها را دربرمی‌گیرد	وسعت محتوای نمونه‌گیری شده
جامع، معمولاً سه سؤال یا بیش از آن برای هر هدف	سطحی، معمولاً یک یا دو سؤال برای هر هدف	جامعیت محتوای نمونه‌گیری شده
معنای نمره، وابسته به نمره‌های دیگر نیست؛ این نمره مستقیماً از سؤال‌ها و ملاک نشئت می‌گیرد. تغییرپذیری، حداقل است.	از آنجا که معناداری نمره هنجاری، اساساً وابسته به وضعیت نسبی نمره در مقایسه با دیگر نمره‌ها بوده، بهتر است.	تغییرپذیری
سؤال‌هایی انتخاب می‌شوند که رفتار ملاک را انعکاس دهند. تأکید بر حوزه شناسایی پاسخ‌های مرتبط است.	سؤال‌هایی انتخاب می‌شوند که پراکندگی را افزایش یا گسترش دهند. از سؤال‌هایی که «خیلی آسان» یا «خیلی سخت» هستند، خودداری می‌شود. یک هدف این است که «گزینه‌های انحرافی» خوب تهیه شود.	ساخت سؤال
	رتبه درصدی و نمره‌های استاندارد استفاده می‌شوند (رتبه‌های نسبی).	گزارش و تفسیر امور

دشواری سؤال‌های آزمون‌های هنجاری و ملاکی، متفاوت است. سؤال‌های آزمون‌های هنجاری، از بسیار دشوار که هیچ‌کس قادر به پاسخگویی آن‌ها نیست تا سؤال‌های بسیار ساده که همه می‌توانند پاسخ دهند، متفاوت هستند. سؤال‌های

آزمون‌های ملاکی، از نظر سطح دشواری، اغلب شبیه هم هستند. از آنجا که سؤال‌های ملاکی، معمولاً بعد از یک دوره آموزش به دانش‌آموزان داده می‌شود، برای دانش‌آموزان ساده به نظر می‌رسد. تعداد متوسط دانش‌آموزانی که در آزمون‌های ملاکی، به سؤال پاسخ صحیح می‌دهند، ۸۰٪، ولی در آزمون‌های هنجاری، ۵۰٪ است (کوبیزین و بوریچ، ۲۰۱۳).

آزمون‌های هنجاری، معمولاً در شرایط استاندارد شده یکسان، به هزاران دانش‌آموزی داده می‌شود که شبیه کسانی هستند که آزمون برای آن‌ها طراحی شده است. این شرایط، به سازنده آزمون این امکان را می‌دهد تا هنجارهایی برای مقایسه نمره هر نفر با آن تعیین کند. در جدول ۱-۳، دو آزمون با هم مقایسه شده‌اند.

۳-۶ آزمون‌های معلم‌ساخته

آزمون‌های معلم‌ساخته، آزمون‌هایی هستند که معلمان برای سنجش پیشرفت تحصیلی دانش‌آموزان خود تهیه می‌کنند و در طول دوره آموزش یا در پایان دوره، به اجرا درمی‌آورند. این نوع آزمون‌ها اغلب برای تعیین میزان موفقیت یادگیرندگان در رسیدن به هدف‌های دقیق آموزشی طرح‌ریزی می‌شوند و به این دلیل، معمولاً در آن‌ها ملاکی مطلق اندازه‌گیری به کار گرفته می‌شود (سیف، ۱۳۹۶).

آزمون‌های معلم‌ساخته، به وسیله معلم و برای تعداد محدودی دانش‌آموز ساخته می‌شود و هر معلم نیز سؤال‌ها را به روش خود نمره‌گذاری می‌کند. (بوئم^۱ ۱۹۹۲؛ به نقل از گولو^۲، ۲۰۰۹) دو مزیت برای آزمون‌های معلم‌ساخته و ملاکی، عبارت‌اند از:

۱. این نوع سنجش، با تسلط دانش‌آموز بر مهارت‌ها، دانش، یا فرایندها سروکار دارد و علاقه‌ای به مقایسه دانش‌آموزان با یکدیگر ندارد.
۲. نتایج آزمون‌های ملاکی و معلم‌ساخته، ساده‌تر به هدف‌های آموزشی تبدیل می‌شوند. برخلاف آزمون‌های هنجاری که نمره عملکرد، نشان‌دهنده چگونگی دستیابی دانش‌آموز به آن نمره نیست، آزمون ملاکی به‌ویژه، علاقه‌مند به مرتبط ساختن نمره عملکرد به توالی آموزشی است.

آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم‌ساخته ۵۹

یادآوری ۱. آزمون‌های هنجاری، در اجرا و نمره‌گذاری، شیوه عمل یکسانی دارند.

یادآوری ۲. در آزمون‌های ملاکی، عملکرد یا نمره دانش‌آموز، با یک ملاک از پیش تعیین شده مقایسه می‌شود.

یادآوری ۳. آزمون‌های معلم‌ساخته، اغلب برای تعیین میزان موفقیت یادگیرندگان در رسیدن به هدف‌های دقیق آموزشی طرح‌ریزی می‌شوند.

۷-۳ مقایسه آزمون‌های استاندارد شده و کلاسی

آزمون‌های پیشرفت استاندارد شده و آزمون‌های کلاسی به دقت تهیه شده، تشابه زیادی با هم دارند. هر دو بر مبنای جدول مشخصات (نقشه آزمون) تهیه می‌شوند و راهنمایی دقیقی برای دانش‌آموزان دارند. تفاوت‌های عمده بین این دو آزمون عبارت‌اند از:

۱. ماهیت پیامدهای یادگیری و محتوای اندازه‌گیری شده

۲. کیفیت سؤال‌های آزمون

۳. اعتبار آزمون‌ها

۴. روش‌های اجرا و نمره‌گذاری

۵. تفسیر نمره‌ها

مزایای تطبیقی آزمون‌های پیشرفت استاندارد شده و آزمون‌های غیررسمی کلاسی، در جدول ۲-۳ نشان داده شده است. مرور مزایای تطبیقی دو آزمون نشان می‌دهد که هر کدام، مناسب برای هدفی و نامناسب برای هدف دیگری است. به دلیل دامنه شمول گسترده‌تر آزمون‌های استاندارد شده، روش‌های کنترل سخت‌گیرانه‌تری بر اجرا و نمره‌گذاری آن حاکم است و دسترسی به هنجارهای ارزیابی نمره‌ها، آن را به‌ویژه برای هدف‌های آموزشی زیر، مفید می‌سازد:

۱. ارزیابی رشد مهارت کلی دانش‌آموز، در آن دسته از پیامدهای یادگیری که در بسیاری از دوره‌های تحصیلی مشترک‌اند.

۲. ارزیابی پیشرفت دانش‌آموز در سال تحصیلی یا یک دوره چندساله.

۶. سنجش و اندازه‌گیری

۳. تعیین نقاط قوت و ضعف دانش‌آموز در یک ماده درسی یا مهارت.
۴. مقایسه سطح کلی پیشرفت دانش‌آموزان با استعداد تحصیلی آن‌ها.
- عدم انعطاف‌پذیری آزمون‌های استاندارد شده در مقایسه با آزمون کلاسی غیررسمی، باعث کارآمدتر بودن آزمون‌های کلاسی در سنجش آن هدف‌ها می‌شود.

۳-۷-۱ مزایای آزمون‌های استاندارد

۱. ارزیابی پیامدهای یادگیری و محتوای اختصاصی یک کلاس یا مدرسه خاص.
 ۲. ارزیابی پیشرفت روزانه دانش‌آموزان و پیشرفت آن‌ها در واحدهای کاری با اندازه‌های مختلف.
 ۳. ارزیابی دانش تحولات جاری در حوزه‌های محتوایی نظیر علوم و مطالعات اجتماعی که به سرعت، در حال تغییرند.
- کارکردهای مکمل این دو نوع آزمون نشان می‌دهد که هر دو می‌توانند در یک برنامه آموزشی دقیق سهیم باشند. هر کدام با توجه به پیشرفت آموزشی دانش‌آموز، اطلاعات خاصی به دست می‌دهند. با وجود این، در هر دو مورد، ارزش اطلاعات به گستره‌ای است که آزمون را به هدف‌های آموزشی مدرسه مرتبط می‌سازد و همچنین به نحوه استفاده از نتایج آزمون وابسته است. آزمون‌های استاندارد شده پیشرفت مانند آزمون‌های کلاسی غیررسمی، می‌توانند هدف‌های آموزشی ارزشمند زیادی را که به آن‌ها نسبت داده شده است، به خدمت گیرند. محتوا فقط هنگام اندازه‌گیری پیامدهای خاصی به واسطه آزمون‌ها و از نظر کسانی که مسئول برنامه آموزشی‌اند، مهم تلقی می‌شود. به علاوه نتایج آزمون باید به طور مناسب، مورد استفاده قرار گیرد (میلر، لین و گراندلند، ۲۰۰۹).

آزمون‌های معلم ساخته به دلیل انعطاف‌پذیری و به موقع بودنشان، اغلب برای اندازه‌گیری هدف‌های آموزشی در یک دوره خاص، مناسب‌تر از آزمون‌های پیشرفت استاندارد شده‌اند. با وجود این، هنگامی که آزمون‌های استاندارد شده به دقت مطابق محتوای دوره و پیامدهای یادگیری انتخاب می‌شوند، می‌توانند اطلاعات ارزشمندی در اختیار معلم بگذارند.

آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم ساخته ۶۹

جدول ۳-۴. مقایسه مزایای تطبیقی آزمون‌های پیشرفت استاندارد شده و کلاسی

ویژگی	آزمون‌های پیشرفت استاندارد شده	آزمون‌های پیشرفت غیررسمی
پیامدهای یادگیری و محتوای اندازه‌گیری شده	پیامدها و محتوای مشترک بیشتر مدارس را اندازه‌گیری می‌کنند. آزمون مهارت‌های اساسی و پیامدهای پیچیده، با بسیاری از موقعیت‌های محلی سازگاری دارد. آزمون‌های معطوف به محتوا، به انعکاس تأکید یا به موقع بودن برنامه درسی محلی نمی‌پردازند.	با پیامدها و محتوای برنامه درسی، به خوبی سازگارند. انعطاف‌پذیری، تطابق پیوسته اندازه‌گیری را با مواد جدید و تغییرات در روش فراهم می‌سازد. با واحدهای کاری اندازه- متفاوت سازگار است. اغلب پیامدهای یادگیری پیچیده را نادیده می‌گیرد.
کیفیت سؤال‌های آزمون	کیفیت کلی سؤال‌ها بالاست. به وسیله متخصصان نوشته شده‌اند، پیش‌آزمون شده‌اند. بر مبنای اثربخشی انتخاب شده‌اند.	کیفیت سؤال‌ها نامعلوم است، مگر اینکه فایل سؤال‌ها نگه داشته شود و مورد استفاده قرار گیرد. کیفیت آن‌ها پایین‌تر از آزمون‌های استاندارد شده است، زیرا معلم فرصت کمی دارد و زمان کافی برای پیش‌آزمون سؤال‌هایی در اختیار ندارد.
اعتبار	اعتبار بالاست؛ معمولاً بین ۰/۸۰ تا ۰/۹۵؛ اغلب با میانۀ ۰/۹۰.	اعتبار، معمولاً نامعلوم است؛ اگر به دقت تهیه شوند، می‌تواند بالا باشد.
اجرا و نمره‌گذاری	روش‌ها استاندارد شده است؛ آموزش‌های اختصاصی فراهم شده است.	تأکید بر روش‌های یکسان است، اما ممکن است انعطاف‌پذیر باشد.
تفسیر	نمره‌ها با هنجار گروهی مقایسه می‌شود. راهنمای آزمون و دیگر راهنمایی‌ها، به تفسیر و کاربرد آزمون کمک می‌کنند.	مقایسه و تفسیر نمره‌ها، محدود به شرایط مدرسه محلی است و بر حسب تاریخچه آموزشی شناخته شده، قابل تفسیر است.

۳-۸ تفاوت‌های آزمون‌های استاندارد با آزمون‌های معلم ساخته

- تهیه آزمون‌های استاندارد در مقایسه با آزمون‌های معلم ساخته، به زمان، تخصص و هزینه‌های بیشتری نیاز دارد.
- از آنجا که آزمون‌های استاندارد برای استفاده در مناطق مختلف و مقایسه دانش‌آموزان مناطق گوناگون با یکدیگر تهیه می‌شوند، باید هدف‌های مختلف آموزشی مناطق مختلف را در آن‌ها منظور کرد، اما آزمون‌های معلم ساخته، با توجه به محتوا و هدف‌های ویژه درس‌های خاص تهیه می‌شوند و برای میزان پیشرفت دانش‌آموزان مناطق مختلف، چندان مفید نیستند.
- اصطلاح آزمون‌های استاندارد، به وجود داده‌های هنجاری اشاره می‌کند، ولی آزمون‌های معلم ساخته، نیازی به هنجار و هنجاریابی ندارند (سیف، ۱۳۹۶).

- ۳-۹ استفاده از آزمون‌های استاندارد و معلم‌ساخته در تصمیم‌گیری‌های آموزشی
- تصمیم آموزشی مربوط به کم‌وکیف جریان آموزش، باید بر مبنای آزمون‌های معلم‌ساخته اتخاذ شود، زیرا باعث بهبود جریان آموزش و بالا بردن سطح یادگیری دانش‌آموزان و دانشجویان است. همچنین تصمیمات مربوط به ارتقای دانش‌آموزان و دانشجویان نیز باید بر مبنای آزمون‌های معلم‌ساخته باشد.
 - تصمیمات تشخیصی مربوط به مشکلات یادگیرندگان و رفع نواقص یادگیری و آموزش را بر مبنای هر دو نوع آزمون می‌توان اتخاذ کرد.
 - برای تصمیمات مربوط به جاییابی یا پایه‌گزینی افراد، باید از آزمون‌های استاندارد استفاده کرد.
 - از آنجا که تصمیمات مربوط به راهنمایی و مشاوره دانش‌آموز، مربوط به مقایسه‌های هنجاری است، بنابراین بهتر است از آزمون‌های استاندارد استفاده شود.
 - در مورد تصمیمات مربوط به برنامه‌ریزی درسی که به انتخاب برنامه از میان برنامه‌های مختلف منجر می‌شود و نیازمند مقایسه است، از هر دو آزمون استاندارد و معلم‌ساخته می‌توان کمک گرفت.
 - برای تصمیمات مربوط به ارزشیابی از میزان موفقیت مدارس مختلف نیز بهتر است از آزمون‌های استاندارد استفاده شود (سیف، ۱۳۹۶).

خلاصه فصل سوم

در این فصل، آزمون‌های هنجار مرجع، ملاک مرجع و معلم‌ساخته مورد بحث قرار گرفت. آزمون‌های استاندارد شده، روش‌های یکسانی برای اجرا و نمره‌گذاری دارند و اغلب این امکان را فراهم می‌سازند که عملکرد دانش‌آموز با عملکرد دیگر دانش‌آموزان هم‌سن و هم‌پایه در یک مقیاس ملی مقایسه شود.

هدف‌های آزمون‌های استاندارد، عبارت بودند از: فراهم کردن اطلاعاتی درباره پیشرفت تحصیلی دانش‌آموزان، مشخص کردن نقاط قوت یا ضعف دانش‌آموزان، فراهم‌سازی ملاکی برای جاییابی دانش‌آموزان در برنامه‌های خاص.

آزمون‌های هنجاری اغلب امکان مقایسه عملکرد یک دانش‌آموز با همسالان یا هم‌کلاسی‌هایش بر مبنای یک ملاک ملی یا محلی را فراهم می‌کند. در تفسیر نمره‌های

آزمون‌های استاندارد شده هنجاری، ملاکی و آزمون‌های معلم‌ساخته ۶۳

این آزمون‌ها گروهی که مورد مقایسه با نمره فرد قرار می‌گیرد، گروه هنجار نامیده می‌شود.

آزمون‌های ملاکی معمولاً حوزه محدودی از هدف‌هایی را که به‌طور خاص مربوط به مواد تدریس شده است، اندازه‌گیری می‌کنند. به این آزمون‌ها آزمون‌های معطوف به هدف یا آزمون‌های حیطه‌ای نیز گفته می‌شود. آزمون‌های ملاکی، برای سنجش مهارت و دانش دانش‌آموز در یک حوزه خاص مانند زبان انگلیسی، ریاضی و علوم طراحی می‌شوند.

در آزمون‌های هنجاری، تمرکز بر چگونگی عملکرد نسبی فرد با گروه مورد مقایسه است، ولی در آزمون‌های ملاکی، تمرکز بر عملکرد آزمودنی‌هاست. آزمون‌های معلم‌ساخته نیز آزمون‌هایی هستند که معلمان برای سنجش پیشرفت تحصیلی دانش‌آموزان خود تهیه می‌کنند و در طول دوره آموزش یا در پایان دوره، به اجرا درمی‌آورند. این نوع آزمون‌ها اغلب برای تعیین میزان موفقیت یادگیرندگان در رسیدن به هدف‌های دقیق آموزشی طرح‌ریزی می‌شوند و به این دلیل معمولاً در آن‌ها ملاک مطلق اندازه‌گیری، به کار گرفته می‌شود. کارکردهای مکمل این دو نوع آزمون نشان می‌دهد که هر دو آزمون استاندارد شده و معلم‌ساخته می‌توانند در یک برنامه آموزشی دقیق سهیم باشند.

به‌طور کلی آزمون‌های معلم‌ساخته به دلیل انعطاف‌پذیری و به‌موقع بودن، اغلب برای اندازه‌گیری هدف‌های آموزشی در یک دوره خاص، مناسب‌تر از آزمون‌های پیشرفت استاندارد شده‌اند. با وجود این، هنگامی که آزمون‌های استاندارد شده به‌دقت مطابق محتوای دوره و پیامدهای یادگیری انتخاب می‌شوند، می‌توانند اطلاعات ارزشمندی در اختیار معلم بگذارند.

خودآزمایی چهارگزینه‌ای فصل سوم

۱. با استفاده از کدام نوع از آزمون‌ها می‌توان عملکرد دانش‌آموز را با عملکرد

دانش‌آموزان شهر یا استان، مورد مقایسه قرار داد؟

الف) استاندارد شده

ب) معلم‌ساخته

ج) ملاکی

د) ترکیبی

۲. برای تصمیم‌گیری درباره مجاز بودن دانش‌آموز برای یک دوره خاص، از کدام آزمون استفاده می‌شود؟
 الف) ملاکی
 ب) استاندارد
 ج) معلم‌ساخته
 د) کلاسی
۳. آزمون‌های کنکور سراسری دانشگاه‌ها و مراکز علمی، نمونه‌هایی از آزمون‌های هستند.
 الف) گروه ملاک
 ب) ملاک نسبی
 ج) شفاهی
 د) جورکردنی
۴. در کدام آزمون، عملکرد آزمودنی با عملکرد گروهی که از لحاظ صفت مورد اندازه‌گیری، شبیه وی هستند، مقایسه می‌شود؟
 الف) شفاهی
 ب) فردی
 ج) ملاکی
 د) هنجاری
۵. با استفاده از کدام آزمون می‌توان تسلط دانش‌آموز را بر یک مهارت، مورد سنجش قرار داد؟
 الف) ملاکی
 ب) هنجاری
 ج) معلم‌ساخته
 د) فردی
۶. نخستین گام در استاندارد کردن آزمون‌های هنجاری چیست؟
 الف) اجرای سؤال‌ها روی نمونه معرف
 ب) نمره‌گذاری آزمون اجرا شده روی نمونه
 ج) مقایسه نمره‌های آزمودنی‌ها با میانگین گروه
 د) انتخاب سؤال‌های مناسب با حیطه مورد اندازه‌گیری
۷. در مورد مقایسه آزمون‌های ملاکی و هنجاری، کدام گزینه صحیح است؟
 الف) در آزمون‌های هنجاری، تمرکز بر عملکرد آزمودنی است.
 ب) آزمون‌های هنجاری، نمره فرد را با ملاک مطلق، مورد مقایسه قرار می‌دهند.
 ج) در آزمون ملاکی، عملکرد فرد با نمره خاصی مورد مقایسه قرار می‌گیرد.
 د) در آزمون ملاکی، نمره فرد با ملاک نسبی مورد مقایسه قرار می‌گیرد.
۸. در کدام آزمون، عملکرد آزمون‌شوندگان با همدیگر مورد مقایسه قرار می‌گیرد؟
 الف) کلاسی
 ب) معلم‌ساخته
 ج) ملاکی
 د) هنجاری

۹. در کدام نوع از آزمون‌ها ملاک مطلق به‌کار رفته‌است؟
 الف) هنجاری
 ب) ملاکی
 ج) معلم‌ساخته
 د) انشایی
۱۰. تهیهٔ کدام آزمون‌ها نیاز به‌هزینه و تخصص بیشتری نسبت به‌سایر آزمون‌ها دارد؟
 الف) ملاکی
 ب) هنجاری
 ج) معلم‌ساخته
 د) انشایی
۱۱. در مورد مقایسهٔ آزمون‌های هنجاری و معلم‌ساخته، کدام گزینه درست است؟
 الف) در تصمیمات مربوط به‌ارتقای دانش‌آموزان و دانشجویان، از آزمون‌های استاندارد استفاده می‌شود.
 ب) در تصمیمات تشخیصی مربوط به‌مشکلات یادگیرندگان و رفع نواقص یادگیری، فقط از آزمون‌های معلم‌ساخته استفاده می‌شود.
 ج) برای تصمیمات مربوط به‌جایابی یا پایه‌گزینی افراد، از آزمون‌های استاندارد استفاده می‌شود.
 د) در تصمیمات مربوط به‌راهنمایی و مشاوره، از آزمون‌های معلم‌ساخته استفاده می‌شود.
۱۲. برای اتخاذ تصمیم در مورد میزان موفقیت مدارس، از کدام آزمون می‌توان استفاده کرد؟
 الف) استاندارد
 ب) معلم‌ساخته
 ج) ملاک نسبی
 د) ملاکی

خودآزمایی تشریحی فصل سوم

- هدف از تهیهٔ آزمون‌های استاندارد شده را بنویسید؟
- آزمون هنجاری را تعریف کنید؟ این آزمون، در چه مواردی به‌کاربرده می‌شود؟
- مراحل استاندارد کردن آزمون‌های هنجاری را بنویسید؟
- کاربردهای آزمون‌های ملاکی را با بیان مثال شرح دهید.
- دو مورد از نقاط قوت و ضعف آزمون‌های ملاکی و هنجاری را بنویسید؟
- مزایای آزمون‌های معلم‌ساخته از نظر بوئم (۱۹۹۲) را بنویسید؟
- تفاوت‌ها و شباهت‌های آزمون‌های استاندارد شده و کلاسی را بنویسید؟

فصل چهارم

آزمون‌های کوتاه پاسخ و عینی

هدف کلی

تهیه انواع آزمون‌های عینی و مزایا و معایب آن‌ها

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. آزمون‌های عینی را تقسیم‌بندی کند.
۲. محاسن و معایب آزمون‌های صحیح-غلط را برشمارد.
۳. قواعد تهیه سؤال‌های جورکردنی را بنویسد.
۴. هدف از تهیه سؤال کامل‌کردنی را توضیح دهد.
۵. مزایا و محدودیت‌های سؤال‌های کوتاه پاسخ را شرح دهد.
۶. قواعد تهیه سؤال‌های عددی را بیان کند.
۷. هدف از تهیه آزمون‌های چندگزینه‌ای و قواعد تهیه آن‌ها را توضیح دهد.

مقدمه

در قسمت‌های گذشته، چگونگی تدوین هدف‌های آموزشی کلی و دقیق را آموختید. همچنین با جدول مشخصات و نحوه تهیه آن آشنا شدید. در این فصل، شیوه تهیه سؤال‌های آزمون‌های عینی بررسی می‌شود. آزمون‌های عینی شامل: آزمون‌های صحیح-غلط، جورکردنی، کامل‌کردنی و چندگزینه‌ای هستند.

۴-۱ تقسیم‌بندی آزمون‌های عینی

آزمون‌های سستی سنجش اغلب شامل انواع مختلفی از سؤال هستند؛ از جمله: چندگزینه‌ای^۱، پرکردن جای خالی^۲، کامل کردن جمله^۳، صحیح - غلط^۴، جورکردنی^۵ و انشایی^۶. آزمون‌های سستی، تنها روش یا لزوماً بهترین شیوه برای سنجش یادگیری دانش‌آموز نیستند. با وجود این، متداول‌ترین آزمون‌ها برای تشخیص میزان یادگیری یادگیرنده از یک مطلب‌اند. درواقع آزمون‌های سستی خوب طراحی‌شده، اطلاعات ارزشمندی درباره یادگیری دانش‌آموز به دست می‌دهند (استرنبرگ^۷ و ویلیامز^۸، ۲۰۰۹). در این بخش، فرایند طراحی و نوشتن سؤال‌های عینی مرور می‌شود. هنگامی که به مرحله نوشتن سؤال‌های آزمون رسیدید، ممکن است انتخاب یک چهارچوب یا ترکیبی از چهارچوب‌ها در آزمونتان لازم باشد. هرچند که انتخاب چهارچوب، تا حدی اختیاری است، ولی همیشه این‌گونه نیست. اغلب از قبل، تصمیم‌ها برای شما گرفته شده‌است یا به بیان دقیق‌تر، فقط تا حدودی در انتخاب چهارچوب طرح سؤال آزاد هستید. روال رایج‌تر، به شرح زیر است:

۱. صحیح - غلط
۲. جورکردنی
۳. چندگزینه‌ای
۴. کامل کردنی

۴-۲ چند نکته برای آماده‌سازی سؤال‌ها در طول دوره تحصیلی

- سؤال‌های آزمون را در حال آماده‌شدن برای تدریس تهیه کنید.
- از سؤال‌هایی که دانش‌آموزان همواره در کلاس می‌پرسند، یادداشت بردارید.
- از بدفهمی‌های متداول دانش‌آموزان طی کلاس یا طی انجام تکالیف منزل، یادداشت برداری کنید.
- دانش‌آموزان را برای آشنایی با سؤال‌ها در آخر کلاس یا زمان‌های دیگر، آماده‌سازید.

۳-۴ سؤال‌های صحیح - غلط

در آزمون صحیح - غلط، تعدادی سؤال یا جمله، در اختیار آزمون‌شونده گذاشته می‌شود و او صحیح یا غلط، درست یا نادرست، بله یا نه، واقعیت یا عقیده، موافق یا مخالف و مانند آن را تعیین می‌کند (میلر، لین و گراندلند، ۲۰۰۹). بدیهی است که آزمودنی تنها با دو پاسخ احتمالی، ۵۰ درصد شانس حدس‌زدن دارد. همچنین ماهیت جمله‌ای که باید به صورت صحیح یا غلط داوری شود، معمولاً ساده و واقعی است.

سؤال صحیح - غلط اغلب به صورت یک جمله خبری مثبت نوشته می‌شود و آزمودنی باید صحیح یا غلط بودن آن را تشخیص دهد. سؤال‌های صحیح - غلط، تا اندازه‌ای متداول‌اند، زیرا می‌توان آن‌ها را سریع و راحت نوشت یا حداقل این‌طور به نظر می‌رسد. این گونه سؤال‌ها برای کودکان کوچک که هنوز قادر به نوشتن پاسخ نیستند، مناسب است (استرترگ و ویلیامز، ۲۰۰۹). در واقع نوشتن سؤال‌های صحیح - غلط در مقایسه با سؤال‌های چندگزینه‌ای، زمان کمتری می‌طلبد، اما نوشتن سؤال‌های صحیح - غلط خوب، چندان راحت نیست.

احتمالاً متداول‌ترین استفاده از سؤال‌های صحیح - غلط، اندازه‌گیری توانایی تشخیص صحت جمله‌های مربوط به واقعیت‌ها، تعریف اصطلاحات، جمله‌هایی درباره اصول و مانند آن است. برای اندازه‌گیری چنین پیامدهای نسبتاً ساده‌ای، یک جمله خبری واحد با یکی از چندین روش پاسخگویی، مورد استفاده قرار می‌گیرد.

مثال‌ها.

راهنمایی. جمله‌های زیر را بخوانید. در صورت صحیح بودن جمله، دور «ص» و اگر غلط است، دور «غ» دایره بکشید.

۱. مواد سبزرنگ در برگ گیاه، کلروفیل نامیده می‌شوند. ص غ
۲. جام گل، شامل گلبرگ و کاسبرگ است. ص غ
۳. فتوسنتز فرایندی است که به واسطه آن، برگ‌ها غذای گیاه را می‌سازند. ص غ

پاسخ‌ها. (۱) ص (۲) غ (۳) ص

راهنمایی. هر کدام از سؤال‌های زیر را بخوانید. برای پاسخ بله، دور «ب» و برای خیر، دور «خ» دایره بکشید.

۷۰. سنجش و اندازه‌گیری

۱. آیا 50% عدد ۳۸، بیشتر از ۱۸ است؟
 ۲. آیا ۵۰ درصد $\frac{4}{10}$ برابر $\frac{2}{5}$ است؟
 ۳. اگر 60% عددی ۹ باشد، آن عدد، کوچک‌تر از ۹ است؟
 ۴. ۲۵ درصد ۴۴، کمتر از ۱۲ است؟
- پاسخ‌ها. ۱) ب ۲) ب ۳) خ ۴) ب

نکته. یکی از کارکردهای سودمند سؤال صحیح- غلط، اندازه‌گیری توانایی تشخیص واقعیت از ایده به وسیله دانش‌آموز است.

مثال‌ها.

راهنمایی. جمله‌های زیر را بخوانید. اگر جمله، یک واقعیت است، دور «و» و اگر یک ایده باشد، دور «ا» دایره بکشید.

۱. قانون اساسی، بالاترین قانون در ایران است. و
 ۲. اولین اصلاحیه قانون اساسی، مهم‌ترین اصلاحیه است. و
 ۳. در ایران، رانندگی برای افراد زیر ۱۸ سال ممنوع است. و

پاسخ‌ها. ۱) و ۲) ا ۳) و

راهنمایی. هریک از جمله‌های زیر را بخوانید. اگر جمله، صحیح است، دور «ص» و اگر غلط باشد، دور «غ» و اگر ایده است، دور «ا» دایره بکشید.

۱. زمین، سیاره است. ص غ
 ۲. زمین، دور ماه می‌چرخد. ص غ
 ۳. زندگی هوشمند. ص غ

پاسخ‌ها. ۱) ص ۲) غ ۳) ا

این سؤال‌ها پیامد مهم یادگیری را برای همه دروس می‌سنجند. اگر ک منتقدانه درباره موضوعی فکرمی‌کند، باید اول قادر به تشخیص ایده از واقع باشد.

مثال‌ها.

راهنمایی. جمله‌های زیر را بخوانید. اگر جمله، صحیح است، دور «ص»؛ اگر غلط باشد، دور «غ». اگر معکوس جمله نیز صحیح است، دور «م ص» و اگر معکوس جمله، غلط باشد، دور «م غ» دایره بکشید. به هر سؤال، دو پاسخ بدهید.

۱. همه درختان، گیاه هستند.
۲. تمام انگل‌ها حیوان هستند.

ص غ م ص م ص م غ
ص غ م ص م ص م غ

پاسخ‌ها. (۱) ص، م غ (۲) غ، م غ

یکی از ایرادهایی وارده به سؤال‌های صحیح- غلط این است که امکان دارد دانش‌آموز، جمله غلط را به‌عنوان غلط تشخیص دهد، اما درکی از جمله صحیح نداشته باشد. برای مثال، هنگامی که دانش‌آموز، جمله زیر را به‌عنوان جمله غلط تشخیص دهد، ممکن است نداند صحیح آن چیست.

مثال. خاموشی، باعث افزایش رفتار می‌شود. ص غ

پاسخ. غ. شاید دانش‌آموز، غلط بودن جمله را تشخیص دهد، اما نداند که پاسخ صحیح، تقویت است.

۱-۳-۴ انواع سؤال‌های صحیح- غلط

سیف (۱۳۹۶) در کتاب اندازه‌گیری، سنجش و ارزشیابی آموزشی، انواع سؤال‌های صحیح- غلط را به شرح زیر توضیح می‌دهد:

- نوع صحیح- غلط. در این نوع سؤال، یک جمله خبری در اختیار آزمودنی قرار می‌گیرد و از او خواسته می‌شود صحیح یا غلط بودن آن را تشخیص دهد.

مثال. پایتخت ایران، تهران است. ص غ

پاسخ. ص

- نوع بله- نه. در این نوع سؤال، یک جمله سؤالی در اختیار آزمودنی قرار می‌گیرد و او باید جواب بله یا نه را مشخص کند.

پله نه

مثال. مرکز استان کردستان، شهر ارومیه است؟

پاسخ. نه

- نوع اصلاحی. در این نوع سؤال، به آزمودنی جمله‌ای داده می‌شود که در آن، غلطی وجود دارد و او باید آن را تشخیص دهد، آن را خط بزند و درستش را بنویسد.

مثال. جذر عدد ۶۴، ۹ است.

دانش آموز باید عدد ۹ را خط بزند و به جای آن، عدد ۸ را قرار دهد.

این تنها سؤال صحیح - غلط است که در آن، افزون بر بازشناسی، یادآوری نیز صورت می‌گیرد، زیرا آزمون‌شونده برای پاسخ دادن به سؤال، هم باید جزء غلط آن را تشخیص دهد (بازشناسی) و هم درست آن را به یاد آورد و بنویسد (یادآوری). درواقع این نوع سؤال ترکیبی، از نوع صحیح - غلط و کوتاه پاسخ است.

در نوع اصلاحی، بهتر است در زیر بخشی از جمله که قرار است مورد قضاوت آزمودنی قرار گیرد، خط کشیده شود. در غیر این صورت، امکان دارد او بخش دیگری از جمله را که هدف آموزشی معلم نبوده است، مورد قضاوت قرار دهد. برای مثال، اگر در سؤال بالا زیر ۹ خط کشیده نشود، ممکن است دانش آموز، ۶۴ را خط بزند و به جای آن، ۸۱ را بنویسد.

- نوع خوشه‌ای. در این نوع سؤال که به آن سؤال چند صحیح - غلط نیز می‌گویند، تنه سؤال به صورت یک جمله ناتمام نوشته می‌شود و به دنبال آن، تعدادی جواب قرار می‌گیرند. در این سؤال‌ها، هریک از جواب‌های پیشنهادی باید به عنوان صحیح یا غلط، مورد قضاوت قرار گیرد.

مثال.

میانگین حسابی.

ص غ
ص غ
ص غ

۱. یکی از اندازه‌های گرایش مرکزی است.

۲. تحت تأثیر همه نمره‌هاست.

۳. همبستگی بین دو متغیر را نشان می‌دهد.

پاسخ‌ها. ۱) ص ۲) ص ۳) غ

این نوع سؤال‌ها به معلم اجازه می‌دهند در هر سؤال به جای یک مطلب، چندین مطلب بگنجاند و این گونه در فضای نوشتاری آزمون و زمان خواندن آن به وسیله آزمودنی، صرفه جویی کند.

جدول ۴-۱. مثال‌هایی از سؤال‌های صحیح - غلط خوب و ضعیف

ضعیف: ۱. رشد اخلاقی، شش مرحله دارد. صحیح - غلط چرا ضعیف است؟ سؤال، مشخص نکرده که نظریه به وسیله چه کسی ارائه شده است، بنابراین ممکن است فرد تصور کند سؤال می‌تواند صحیح یا غلط باشد. برای مثال، نظریه رشد اخلاقی پیاز، شش مرحله ندارد.	خوب: ۱. مطابق نظریه کلبه‌گر، رشد اخلاقی شش مرحله دارد. صحیح - غلط
ضعیف: ۲. ماشین، قبل از دوربین عکاسی و چرخ خیاطی اختراع شد. صحیح - غلط چرا ضعیف است؟ برای روشنی بیشتر باید یک مقایسه انجام می‌شد، نه دو مقایسه.	خوب: ۲. ماشین، قبل از دوربین عکاسی اختراع شد. صحیح - غلط
ضعیف: ۳. در بازی بیسبال، ۱۲ دور وجود دارد و بازی فوتبال، چهار ۱۵ دقیقه‌ای است، اما در بازی بسکتبال، فقط ۲ دور وجود دارد. صحیح - غلط چرا ضعیف است؟ در این سؤال، پاسخ صحیح و غلط، هر دو وجود دارد.	خوب: ۳. در یک بازی بیسبال، ۱۲ دور وجود دارد. صحیح - غلط یک بازی فوتبال، چهار دور ۱۵ دقیقه‌ای است. صحیح - غلط یک بازی بسکتبال، ۲ دور دارد. صحیح - غلط
ضعیف: ۴. لیتکلن، بزرگ‌تر از گرنت بود. صحیح - غلط چرا ضعیف است؟ بدون تعیین بعد و ایست که قد است، ممکن است آزمودنی تصور کند ابعاد دیگر، آزمون می‌شوند (مانند محبوبیت).	خوب: ۴. از نظر وزن، لیتکلن بزرگ‌تر از گرنت بود. صحیح - غلط
ضعیف: ۵. صفحه کلید، ۸۸ کلید دارد. صحیح - غلط چرا ضعیف است؟ صفحه کلیدهای ابزارهای مختلف، متفاوت هستند.	خوب: ۵. صفحه کلید پیانوهای جدید، ۸۸ کلید دارد. صحیح - غلط
ضعیف: ۶. موزارت، نابغه بود. صحیح - غلط چرا ضعیف است؟ بدون در نظر گرفتن توانایی موسیقی، آزمودنی ممکن است موزارت را نابغه نداند.	خوب: ۶. بسیاری از مردم، موزارت را با توجه به توانایی‌های موسیقی‌اش نابغه می‌دانند. صحیح - غلط

۲-۳-۴ قواعد تهیه سؤال‌های صحیح- غلط

کویزین و بورج (۲۰۱۳) برای تهیه سؤال‌های صحیح- غلط، موارد زیر را پیشنهاد می‌دهند:

۱. قبل از شروع آزمون، روش صحیح علامت‌گذاری صحیح- غلط باید به‌روشنی تبیین شود.

۲. سؤال‌ها را تا حد امکان مختصر، ساده و روشن بنویسید.

۳. جمله‌هایی به‌کاربرید که به‌وضوح، درست یا غلط باشند.

۴. از نوشتن جمله‌های متنی مضاعف^۱ پرهیزید. درک این جمله‌ها به‌زمان بیشتری نیاز دارد و تفسیر آن‌ها دشوارتر است.

۵. سعی کنید طول سؤال‌های صحیح و سؤال‌های غلط، هم‌اندازه باشد.

۶. سعی کنید تعداد سؤال‌های صحیح و غلط، کم‌ویش به یک اندازه باشد.

۷. از موارد زیر خودداری کنید:

از به‌کاربردن کلماتی چون همه، بعضی وقت‌ها، غالباً، هرگز و از این قبیل خودداری کنید.

از نوشتن سؤال‌های صحیح- غلط از روی جمله‌های کتاب پرهیزید.

از سرنخ‌های کلامی و جمله‌های پیچیده استفاده نکنید.

از نوشتن جواب سؤال‌ها براساس نظم خاصی (مانند: ص ص غ غ؛ یا ص غ ص

غ و مانند آن) پرهیزید.

۸. سعی کنید هر سؤال، به یک هدف آموزشی مهم مربوط باشد و از گنجاندن مطالب

بی‌اهمیت در سؤال پرهیز کنید.

۲-۳-۴ محاسن و معایب آزمون‌های صحیح- غلط

محاسن. آزمون‌های صحیح- غلط به دلیل سادگی تهیه، عینی بودن تصحیح، تعدد

مناسب سؤال و بیان موضوعات درسی در قالب عبارات ساده، از ارزش خاص

برخوردارند. آزمون‌های صحیح- غلط مؤثرند.

معیار.

۱. این آزمون‌ها اغلب برای سنجش اطلاعات جزئی و کم‌اهمیت به کار می‌روند. کاربرد زیاد آن‌ها موجب حفظ طوطی‌وار اطلاعات کم‌اهمیت می‌شود و همچنین برای سنجش هدف‌های سطح پایین به کار می‌روند. در عین حال می‌توان از این نوع پرسش‌ها برای آزمون هدف‌های سطح بالا نیز استفاده کرد و چنانچه مشکلی در سؤال‌های صحیح-غلط وجود داشته باشد، بیشتر به طرح نامناسب سؤال مربوط است تا خود سؤال.

۲. حدس زدن کورکورانه در آن‌ها بسیار زیاد است. صرف نظر از نحوه تهیه سؤال‌های صحیح-غلط، احتمال حدس زدن پاسخ صحیح، بدون اینکه آزمودنی حتی سؤال را خوانده باشد، ۵۰٪ است؛ یعنی اگر آزمونی، ۵۰ سؤال صحیح-غلط داشته باشد، احتمال دارد فردی که با محتوای مورد آزمون آشنایی ندارد، به ۲۵ سؤال پاسخ صحیح دهد. البته این بدان معنا نیست که به طور کلی نباید از این آزمون‌ها استفاده کرد، زیرا آن‌ها گاهی سودمندند. خوشبختانه راه‌هایی وجود دارد که می‌توان به کمک آن، حدس زدن را کاهش داد.

۳. در صورتی که دانش‌آموزان، پاسخ سؤالی را نمی‌دانند، آن‌ها را تشویق کنید پاسخ صحیح را حدس بزنند. از آنجا که نمی‌توان از حدس زدن دانش‌آموزان خاصی ممانعت به عمل آورد، همه دانش‌آموزان را به حدس زدن تشویق کنید تا به تأثیرات آن پی ببرند. این امر سبب می‌شود «عامل حدس» در نمره همه آزمودنی‌ها به شیوه کم و بیش یکسانی منعکس شود. این کار، به نفع کسانی است که محافظه کارند و نمی‌خواهند به سؤال‌هایی که نمی‌دانند، پاسخ دهند.

۴-۴ سؤال‌های جورکردنی

سؤال‌های جورکردنی، شبیه سؤال‌های چندگزینه‌ای هستند، با این تفاوت که به جای چهار گزینه، دو ستون دارند که آزمودنی باید مطالب آن‌ها را به همدیگر پیوند دهد. در این نوع آزمون‌ها معمولاً دو ستون، یکی ستون پرسش و دیگری پاسخ وجود دارد که آزمون‌شونده باید پاسخ را با پرسش مربوط به آن جور کند. تعداد پرسش‌ها و پاسخ‌ها می‌تواند یکسان یا متفاوت باشد (میلر، لین و گرانتلند، ۲۰۰۹). در نوع دیگری از

پرسش‌های جورکردنی، می‌توان نقشه یا تصویری ارائه کرد که بخش‌های مختلف آن شماره‌گذاری می‌شود و نام هر بخش باید یا شماره مربوط به آن جورشود (جلوی شماره، نام بخش موردنظر نوشته‌شود).

بهترین سؤال‌های جورکردنی آن‌هایی هستند که پاسخ یکسانی بارها مورد استفاده قرار می‌گیرد و ستون پرسش‌ها و پاسخ‌ها همگن و هماهنگ است (استرترگ و ویلیامز، ۲۰۰۹). منظور از اینکه می‌گوییم پاسخی باید بارها مورد استفاده قرارگیرد، این است که وقتی دانش‌آموزان، ستون پرسش‌ها را با ستون پاسخ‌ها جور می‌کنند، باید پاسخ‌هایی وجود داشته‌باشد که برای بیش از یک سؤال ستون پرسش‌ها درست‌اند. در این حالت، ستون پاسخ‌ها کوتاه‌تر از ستون پرسش‌هاست. در چنین حالتی، دانش‌آموزان نمی‌توانند فقط با فرایند حذف پاسخ‌های از پیش انتخاب‌شده، پاسخ صحیح بدهند.

مهم‌ترین مواردی که در سؤال‌های جورکردنی مطرح می‌شود، به شرح زیر است:

اشخاص	_____	موفقیت‌ها
تاریخ‌ها	_____	رویدادهای تاریخی
اصطلاحات	_____	تعاریف
قواعد	_____	مثال‌ها
نمادها	_____	مفاهیم
نویسندگان	_____	عناوین کتاب‌ها
کلمات خارجی	_____	معادل‌های فارسی
ماشین‌ها	_____	کاربردها
گیاهان یا حیوانات	_____	طبقه‌بندی
اصول	_____	تبیین‌ها
اشیاء	_____	نام اشیاء
بخش‌ها	_____	کارکردها

۴-۴-۱ قواعد تهیه سؤال‌های جورکردنی

۱. پرسش‌ها و پاسخ‌های متجانس یا همگون انتخاب کنید. این تجانس باید هم از لحاظ

- جمله‌بندی و هم از لحاظ موضوع و محتوا وجود داشته باشد. مثال: همگی شامل مکان یا اشخاص باشد.
۲. فهرست پرسش‌ها باید عبارات یا جمله‌های طولانی‌تری داشته باشد و در ستون سمت چپ نوشته شود و پاسخ‌های ستون سمت راست باید مرکب از عبارات، کلمات، یا نمادهای کوتاه باشد.
۳. فهرست پرسش‌ها تقریباً بیشتر از ده مورد نباشد.
۴. طول و فهرست پرسش و پاسخ را کوتاه انتخاب کنید و طول پاسخ‌ها کوتاه‌تر از طول پرسش‌ها انتخاب شود. خواندن و فهم مطالب در فهرست‌های طولانی، کار پاسخ‌دهی را دشوار می‌کند.
۵. همه پرسش‌ها و پاسخ‌های یک‌سؤالی را در یک صفحه قرار دهید. در صورت امکان، فهرست پاسخ‌ها را به‌طور منطقی مرتب کنید. اگر این امکان وجود نداشته باشد، لیست‌های جداگانه از پرسش و پاسخ تهیه کنید.
۶. در راهنمای سؤال، اطلاعات لازم را درباره چگونگی جورکردن پرسش‌ها و پاسخ‌ها ارائه کنید. توضیحات، مختصر و مفید باشد. یک پرسش نمونه هم می‌تواند راهنمای خوبی باشد. اگر قرار است پاسخی بیش از یک‌بار مورد استفاده قرار گیرد، در راهنمای آزمون توضیح داده شود.
۷. در صورت امکان، فهرست پاسخ‌ها را به‌طور منطقی مرتب کنید. این کار می‌تواند به ترتیب الفبایی، از کمتر به بیشتر یا برعکس و ... باشد. چنین عملی موجب پیدا کردن سریع‌تر پاسخ می‌شود.
۸. به منظور افزایش همگونی آزمون، مطمئن شوید هریک از پاسخ‌ها برای تمام پرسش‌های یک سؤال، درست جلوه کند.
۹. پرسش‌ها را با شماره و پاسخ‌ها را با حروف، مشخص کنید.
۱۰. در فهرست پرسش‌های سؤال‌های جورکردنی، از به کار بردن جمله‌های نیمه تمام خودداری کنید.
۱۱. از نوشتن سؤال‌هایی که پرسش‌ها و پاسخ‌ها با هم جور شوند، پرهیزید. پرسش‌ها و پاسخ‌ها طوری طراحی شوند که هر کدام مستقل از دیگری باشند و پاسخ به آن‌ها، مستلزم تسلط و آگاهی باشد و آزمون‌شوندگان بی‌اطلاع، نتوانند به پرسش‌هایی که جوابش را نمی‌دانند، پاسخ دهند (کویزین و بوریج، ۲۰۱۳).

جدول ۲-۲. مثال‌هایی از سؤال‌های جورکردنی ضعیف و خوب

خوب: ۱. سرمایه ایالت در ستون B را با ایالت مناسب در ستون A، کنار هم قرار دهید. <table border="1"> <thead> <tr> <th>ستون A</th> <th>ستون B</th> </tr> </thead> <tbody> <tr><td>۱. نیویورک</td><td>الف) جونو</td></tr> <tr><td>۲. آلاسکا</td><td>ب) لینکلن</td></tr> <tr><td>۳. فلوریدا</td><td>ج) مادیسون</td></tr> <tr><td>۴. نبراسکا</td><td>د) تالاهاسی</td></tr> <tr><td>۵. نیوجرسی</td><td>ه) ساکرامنتو</td></tr> <tr><td>۶. ویسکانسین</td><td>و) ترنتون</td></tr> <tr><td>۷. کالیفرنیا</td><td>ز) آلیانی</td></tr> </tbody> </table> * جونو، زن ژوپیتر است.	ستون A	ستون B	۱. نیویورک	الف) جونو	۲. آلاسکا	ب) لینکلن	۳. فلوریدا	ج) مادیسون	۴. نبراسکا	د) تالاهاسی	۵. نیوجرسی	ه) ساکرامنتو	۶. ویسکانسین	و) ترنتون	۷. کالیفرنیا	ز) آلیانی	ضعیف: ۱. ایالت ستون A را با منشأ لروئش در ستون B، کنار هم قرار دهید. <table border="1"> <thead> <tr> <th>ستون A</th> <th>ستون B</th> </tr> </thead> <tbody> <tr><td>۱. نیویورک</td><td>الف) کشاورزی</td></tr> <tr><td>۲. آلاسکا</td><td>ب) معادن</td></tr> <tr><td>۳. فلوریدا</td><td>ج) صنایع</td></tr> <tr><td>۴. نبراسکا</td><td>د) پردازش اطلاعات</td></tr> <tr><td>۵. نیوجرسی</td><td>ه) تولید</td></tr> <tr><td>۶. ویسکانسین</td><td>و) حمل و نقل</td></tr> <tr><td>۷. کالیفرنیا</td><td>ز) بیمه</td></tr> </tbody> </table> ح) آموزش ط) موقوفه‌ها چرا ضعیف است؟ سؤال و گزینه‌ها مبهم و گیج کننده‌اند.	ستون A	ستون B	۱. نیویورک	الف) کشاورزی	۲. آلاسکا	ب) معادن	۳. فلوریدا	ج) صنایع	۴. نبراسکا	د) پردازش اطلاعات	۵. نیوجرسی	ه) تولید	۶. ویسکانسین	و) حمل و نقل	۷. کالیفرنیا	ز) بیمه
ستون A	ستون B																																
۱. نیویورک	الف) جونو																																
۲. آلاسکا	ب) لینکلن																																
۳. فلوریدا	ج) مادیسون																																
۴. نبراسکا	د) تالاهاسی																																
۵. نیوجرسی	ه) ساکرامنتو																																
۶. ویسکانسین	و) ترنتون																																
۷. کالیفرنیا	ز) آلیانی																																
ستون A	ستون B																																
۱. نیویورک	الف) کشاورزی																																
۲. آلاسکا	ب) معادن																																
۳. فلوریدا	ج) صنایع																																
۴. نبراسکا	د) پردازش اطلاعات																																
۵. نیوجرسی	ه) تولید																																
۶. ویسکانسین	و) حمل و نقل																																
۷. کالیفرنیا	ز) بیمه																																
خوب: ۲. آهنگساز را در ستون B با کارش در ستون A تطبیق دهید. <table border="1"> <thead> <tr> <th>ستون A</th> <th>ستون B</th> </tr> </thead> <tbody> <tr><td>۱. دختر هنگ</td><td>الف) روسینی</td></tr> <tr><td>۲. ازدواج فیگارو</td><td>ب) گرشوین</td></tr> <tr><td>۳. پارگی و بس</td><td>ج) دونیزتی</td></tr> <tr><td>۴. آیدا</td><td>د) موزارت</td></tr> <tr><td>۵. آرایشگر سویل</td><td>ه) وردی</td></tr> </tbody> </table>	ستون A	ستون B	۱. دختر هنگ	الف) روسینی	۲. ازدواج فیگارو	ب) گرشوین	۳. پارگی و بس	ج) دونیزتی	۴. آیدا	د) موزارت	۵. آرایشگر سویل	ه) وردی	ضعیف: ۲. دو ستون را به گونه‌ای تطبیق دهید که آهنگساز با کاری که او سروده است، منطبق باشد. <table border="1"> <thead> <tr> <th>ستون A</th> <th>ستون B</th> </tr> </thead> <tbody> <tr><td>۱. موزارت</td><td>الف) وردی</td></tr> <tr><td>۲. دختر هنگ</td><td>ب) پارگی و بس</td></tr> <tr><td>۳. آرایشگر سویل</td><td>ج) دونیزتی</td></tr> <tr><td>۴. گرشوین</td><td>د) ازدواج فیگارو</td></tr> <tr><td>۵. آیدا</td><td>ه) روسینی</td></tr> </tbody> </table> چرا ضعیف است؟ زیرا آهنگساز و کارهایش، در همان ستون هستند و این، باعث سردرگمی می‌شود.	ستون A	ستون B	۱. موزارت	الف) وردی	۲. دختر هنگ	ب) پارگی و بس	۳. آرایشگر سویل	ج) دونیزتی	۴. گرشوین	د) ازدواج فیگارو	۵. آیدا	ه) روسینی								
ستون A	ستون B																																
۱. دختر هنگ	الف) روسینی																																
۲. ازدواج فیگارو	ب) گرشوین																																
۳. پارگی و بس	ج) دونیزتی																																
۴. آیدا	د) موزارت																																
۵. آرایشگر سویل	ه) وردی																																
ستون A	ستون B																																
۱. موزارت	الف) وردی																																
۲. دختر هنگ	ب) پارگی و بس																																
۳. آرایشگر سویل	ج) دونیزتی																																
۴. گرشوین	د) ازدواج فیگارو																																
۵. آیدا	ه) روسینی																																
خوب: ۳. عدد متداعی با مرحله شناخت رشد پیاژه را با پیشرفت رشدی ارائه شده در پایین، کنار هم قرار دهید. ۱. پایداری شیء حاصل می‌شود. الف) حسی - حرکتی ۲. مسائل انتزاعی حل می‌شود. ب) پیش عملیاتی ۳. بازی نمادی انجام می‌شود. ج) عملیات عینی ۴. برگشت پذیری حاصل می‌شود. د) عملیات صوری	ضعیف: ۳. عدد مرحله رشد شناختی مطرح شده به وسیله پیاژه را با پیشرفت رشدی ارائه شده در زیر، کنار هم قرار دهید. ۱. طی مرحله، پایداری شیء حاصل می‌شود. ۲. یک کودک هنگامی که به سن می‌رسد، قادر به حل مسائل انتزاعی است. ۳. کودک، برای اولین بار در سن قادر به بازی نمادی است. ۴. مرحله وقتی است که اصل برگشت پذیری حاصل می‌شود. الف) عملیات صوری ب) حسی - حرکتی ج) پیش عملیاتی د) عملیات عینی چرا ضعیف است؟ تعداد بسیار زیادی لغت و ساختاری گیج کننده دارد.																																

۲-۴-۴ محاسن آزمون‌های جورکردنی

۱. سؤال‌های آزمون‌های جورکردنی، برای آزمون دانش اصطلاحات، تعاریف، تاریخ‌ها، رویدادها و سایر موارد بسیار مناسب است.
۲. از آنجا که خواندن این سؤال‌ها زمان زیادی لازم ندارد، تعداد زیادی از آن‌ها را می‌توان در یک زمان معین، مورد استفاده قرار داد.
۳. تصحیح سؤال‌های جورکردنی، به سرعت و سهولت امکان‌پذیر است و نیازی به مصحح مشخص ندارد.

۳-۴-۴ معایب آزمون‌های جورکردنی

۱. در یادگیری تداعی بین امور، حفظ طوطی‌وار، دخالت زیادی دارد و معلمان اغلب در تهیه این نوع سؤال‌ها، بر روابط حفظی بین امور تأکید می‌کنند.
۲. بعضی وقت‌ها پیدا کردن مجموعه‌ای از پرسش‌ها که به اندازه کافی شبیه به هم باشند تا بتوان برای آن‌ها مجموعه پاسخ‌های همگونی پیدا کرد، کار دشواری است.

یادآوری ۱. آزمون‌های سستی، متداول‌ترین آزمون‌هایی هستند که به کمک آن‌ها می‌توان به میزان درک یادگیرنده از مطلب پی برد.

یادآوری ۲. در سؤال‌های صحیح-غلط، آزمودنی، ۵۰ درصد شانس حدس زدن دارد.

یادآوری ۳. سؤال‌های اصلاحی تنها نوع سؤال‌هایی هستند که در آن، افزون‌بر یازشناسی، یادآوری نیز صورت می‌گیرد.

۵-۴ سؤال‌های کامل کردنی

سؤال‌های کوتاه پاسخ و کامل کردنی، هر دو جزء آزمون‌های تدارک پاسخ^۱ هستند که بایستی به وسیله یک کلمه، عبارت، عدد یا نماد، به آن‌ها پاسخ داده شود (میلر، لین و گرائلند، ۲۰۰۹). این دو آزمون، از نظر ماهیت یکسان‌اند و تنها از نظر نحوه ارائه مسئله،

تفاوت دارند. سؤال کوتاه‌پاسخ، از روش منظم استفاده می‌کند، درحالی‌که سؤال کامل‌کردنی، جمله ناقصی ارائه می‌دهد که باید کامل شود.

بمیان دیگر، در سؤال‌های پرکردن جای خالی و کوتاه‌پاسخ (که کامل‌کردنی نامیده می‌شوند)، لازم است آزمودنی، جمله‌ای را با پرکردن اطلاعات یک جای خالی کامل کند یا یک پاسخ کوتاه به سؤال بدهد. معلمان، این‌گونه سؤال‌ها را ترجیح می‌دهند، زیرا نوشتن آن‌ها راحت است. سؤال‌های آزمون کوتاه‌پاسخ، از آزمودنی می‌خواهند کلمه، عبارت، جمله، عدد یا علامتی را در پاسخ به یک سؤال یا تکمیل یک جمله بنویسند.

آزمون‌های کوتاه‌پاسخ، برای اندازه‌گیری هدف‌های شناختی سطوح پایین به‌ویژه دانش یا معلومات مفیدند.

مثال کوتاه‌پاسخ. نام مخترع کشتی بخار چیست؟ (رابرت فالتون)

مثال کامل‌کردنی: نام مخترع کشتی بخار،..... است. (رابرت فالتون)

۴-۶ استفاده از سؤال‌های کوتاه‌پاسخ

سؤال کوتاه‌پاسخ، برای اندازه‌گیری پیامدهای نسبتاً ساده یادگیری، مناسب است. سؤال‌ها و آزمون زیر، بعضی از استفاده‌های متداول آن را نشان می‌دهد.

مثال‌های دانش واژه‌شناسی. مطالعه علمی رفتار و فرایندهای ذهنی،..... نامیده می‌شود. (روان‌شناسی)

دانش واقعیت‌های خاص. یک نماینده مجلس، برای مدت سال انتخاب می‌شود. (۴)

دانش اصول. سرنوشت رفتاری که پیامد نامطلوبی به دنبال دارد، چه خواهد شد (کاهش می‌یابد)

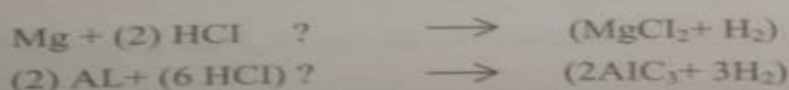
دانش روش یا روند^۱. برای دانستن مثبت یا منفی بودن جریان سیم برق، به چه وسیله نیاز داریم؟ (فازمتر)

- تفسیر ساده داده‌ها. کلمه ایران، چند بخش است؟ (۲)
 در عدد ۹۱۲، ارزش عدد ۹ چند است؟ (۶۰۰)
 در مثلث پایین، هر زاویه چند درجه است؟ (۶۰)



- اگر هواپیمایی از شمال شرقی پرواز کند و یک چرخش ۱۸۰ درجه‌ای داشته باشد، کدام جهت در رأس خواهد بود؟ (جنوب شرقی)
 هنگامی که برای اندازه‌گیری توانایی تفسیر نمودارها، چارت‌ها، شکل‌ها و داده‌های تصویری، از آزمون‌های کوتاه پاسخ استفاده می‌شود. تفسیر پیچیده‌تر عملی است.
 مثال‌هایی از حل مسائل عددی. یک کیلو شیر، ۱۵۰۰ تومان و یک سطل ۴ کیلویی، ۵۲۰۰ تومان است. اگر شما یک سطل شیر بخرید، در مقایسه با خرید یک کیلو، چقدر سود کرده‌اید؟ (۸۰۰)
 مهارت در دست‌کاری نمادهای ریاضی. اگر $x/b = 3/b - 1$ باشد، در نتیجه $x = ?$ ($3b/b - 1$)

توانایی کامل کردن و متعادل کردن فرمول‌های شیمی



برای مثال‌های بالا، یک سؤال کوتاه پاسخ، خیلی بهتر از سؤالی است که آزمودنی باید از میان آن‌ها پاسخ صحیح را پیدا کند.

۴-۶-۱ مزایا و محدودیت‌های سؤال‌های کوتاه پاسخ

تهیه سؤال‌های آزمون کوتاه پاسخ، بسیار راحت است، تا حدی به این دلیل که معمولاً پیامدهای نسبتاً ساده یادگیری را اندازه می‌گیرند. آزمون‌های کوتاه پاسخ به غیر از مواردی که برای حل مسائل ریاضی و علوم به کار می‌رود، کم‌وبیش به طور انحصاری برای اندازه‌گیری یادآوری اطلاعات حفظی به کار می‌رود.

مهم‌ترین مزیت آزمون کوتاه‌پاسخ این است که تدارک پاسخ باید به وسیله آزمون‌دهی انجام شود. این کار باعث کاهش احتمال یافتن پاسخ سؤال به وسیله آزمون‌دهی از روی حدس است. آن‌ها باید اطلاعات درخواست‌شده را به خاطر بیاورند و محاسبات لازم را برای حل مسئله ارائه‌شده انجام دهند.

در محدودیت عمده، استفاده از آزمون کوتاه‌پاسخ را با مشکل مواجه می‌سازد. یکی نامناسب بودن برای اندازه‌گیری پیامدهای یادگیری پیچیده و دیگری دشواری تمره‌گذاری.

۴-۶-۲ پیشنهادهایی برای نوشتن سؤال‌های کوتاه‌پاسخ

۱. سؤال را طوری طرح کنید که به پاسخی مختصر و اختصاصی نیاز داشته باشد. پاسخ شامل یک کلمه، عبارت، عدد یا نماد باشد.

مثال:

ضعیف: حیوانی که گوشت حیوانات دیگر را می‌خورد، است. (گوشت‌خوار)

بهبتر: حیوانی که گوشت حیوانات دیگر را می‌خورد، به عنوان طبقه‌بندی می‌شود. (گوشت‌خوار)

آزمون‌دهی در پاسخ به سؤال اول ممکن است گریز، شیر یا حتی گریسته را بنویسد. وقتی از دانش‌آموز خواسته می‌شود این نوع حیوان را دسته‌بندی کند (سؤال دوم)، مسئله بهتر شکل داده می‌شود و نوع پاسخ لازم را تعریف می‌کند.

۲. صورت سؤال را با بیان مشابه جمله‌های کتاب بنویسید. نوشتن سؤال از روی جمله‌ای کتاب، باعث تشویق دانش‌آموزان به حفظ کتاب می‌شود.

۳. هر سؤال باید موضوع مهمی را شامل شود.

۴. صورت سؤال باید نیازمند پاسخ مشخص و واحدی باشد؛ یعنی سؤال، مبهم و چندپهلوی نباشد. نام علمی زنگ آهن چیست؟

۵. در سؤال‌هایی که پاسخ‌های آن‌ها اعداد است، واحد مقیاس و میزان باید مشخص شود.

۶. در سؤال‌های کوتاه‌پاسخ کامل‌کردنی، تنها کلمات و عبارات مهم را حذف کنید.

بزرگ‌ترین قله ایران، قله است.

۷. در سؤال‌های کامل‌کردنی، تعداد زیادی جای خالی منظور نکنید.

مثال. جانوران، به دو دسته و تقسیم می‌شوند که در مناطق زندگی می‌کنند.

۸. جای خالی سؤال‌های کامل‌کردنی را تا حد امکان در قسمت پایانی سؤال قرار دهید تا شاگردان قبل از رسیدن به جواب، با خواندن سؤال، از موضوع اطلاع حاصل کنند.

مثال. بزرگ‌ترین شهر هر استان که با شهرهای دیگر در ارتباط است، آن استان نام دارد.

۹. تا حد امکان به جای سؤال‌های کامل‌کردنی، از سؤال‌های پرسشی استفاده کنید، زیرا تهیه آن آسان‌تر و دارای ابهام کمتری است، پاسخ‌دادن به سؤال پرسشی برای شاگرد آسان‌تر خواهد بود.

مثال. پایتخت عراق، شهر است.

مثال پرسشی. نام پایتخت کشور عراق چیست؟

۱۰. در سؤال کامل‌کردنی، از کاربرد اشاره‌های دستوری و موارد دیگر که مشخص‌کننده جواب است، خودداری کنید. مثال: (را) برای مشخص کردن مفعول

۴-۷ قواعد تهیه سؤال‌های عددی

۱. تا حد امکان از اعداد ساده استفاده کنید. هدف این سؤال‌ها آزمایش فهم دانش‌آموزان از جریان یا فرایند محاسبات است، نه دقت در محاسبات عددی (یعنی از اعداد دارای اعشار، کمتر استفاده شود).

۲. تا حد امکان مسائلی را مطرح کنید که دارای جواب صحیح باشند.

۳. میزان دقت موردانتظار را تعیین کنید. مثال: عدد را به مخلوط تبدیل کنید، ساده کنید، به صورت اعشاری بنویسید یا به عدد صحیح تبدیل کنید.

۴. اگر نوشتن واحد در پاسخ شاگردان الزامی است، به اطلاع شاگردان برسانید. مثل نوشتن سانتی‌متر مربع، هکتار، لیتر یا کیلومتر

یادآوری ۱. سؤال‌های کوتاه پاسخ و کامل‌کردنی، هر دو جزء آزمون‌های تدارک پاسخ هستند که بایستی به وسیله یک کلمه، عبارت، عدد یا نماد، به آن‌ها پاسخ داده شود.

یادآوری ۲. آزمون‌های کوتاه‌پاسخ، برای اندازه‌گیری هدف‌های شناختی سطوح پایین، به‌ویژه دانش یا معلومات مفیدند.

۴-۸ آزمون‌های چندگزینه‌ای

آزمون‌های چندگزینه‌ای، بهترین نوع آزمون عینی هستند. به‌کمک این آزمون می‌توان اطلاعات بسیار زیادی را در یک جلسه آزمود. همچنین آنچه را که سؤال‌های صحیح-غلط، جورکردنی و کوتاه‌پاسخ اندازه‌گیری می‌گیرند، به‌راحتی و به‌طور مؤثر سنجید. سؤال‌های چندگزینه‌ای، برای اندازه‌گیری سطوح دانش و دیگر پیامدهای یادگیری، مورد استفاده قرار می‌گیرند. آن‌ها اغلب برای اندازه‌گیری سطوح دانش، فهمیدن و به‌کار بستن، مناسب‌اند. آزمون چندگزینه‌ای، شامل تعدادی سؤال است که هریک از آن‌ها از یک قسمت اصلی که تنه سؤال^۱ نامیده می‌شود (که می‌تواند یک سؤال یا جمله ناقص باشد) و تعدادی گزینه (پاسخ) تشکیل می‌شود و آزمون‌شونده از میان گزینه‌های پیشنهادی، گزینه صحیح (پاسخ سؤال) را انتخاب می‌کند (در هر سؤال، یک گزینه صحیح وجود دارد).

هر سؤال چندگزینه‌ای، شامل بخش‌های زیر است:

۱. قسمت اصلی یا تنه سؤال: متن اصلی سؤال که مسئله یا موضوع مورد اندازه‌گیری را دربرمی‌گیرد.

۲. گزینه درست یا پاسخ: گزینه کلید هم به آن گفته می‌شود.

۳. گزینه‌های انحرافی: گزینه‌هایی غیر از گزینه‌های درست.

تعداد گزینه‌های پرسش‌های چندگزینه‌ای، از ۲ تا ۵ متغیر است و به‌سبب مشکل در پیدا کردن پنج گزینه و نیز کاهش اثر حدس و گمان در پرسش‌های دو و سه‌گزینه‌ای، بهترین نوع پرسش‌های چندگزینه‌ای، پرسش چهارگزینه‌ای است.

تنه سؤال، خواه از سؤال مستقیم استفاده شود یا جمله ناقص، به‌چندین عامل بستگی دارد. فرم سؤال مستقیم، برای نوشتن راحت‌تر است، برای کودکان کوچک‌تر، طبیعی‌تر جلوه کرده، احتمالاً مسئله را واضح‌تر بیان می‌کند. از طرف دیگر، جمله ناقص، دقیق‌تر است و در صورت طراحی ماهرانه، سؤال بهتری خواهد بود. این‌گونه مرسوم که سؤال آزمون را با جمله مستقیم نوشته‌شود و اگر دقت سؤال افزایش یابد، به‌جمله ناقص تبدیل شود.

مثال‌ها	فرم سؤال مستقیم
کدام یک از شهرهای زیر، مرکز استان خوزستان است؟	
الف) ایلام	
ب) اهواز	
ج) همدان	
د) کرمانشاه	
پاسخ: ب	
فرم جمله ناقص	
مرکز استان خوزستان، است.	
الف) ایلام	
ب) اهواز	
ج) همدان	
د) کرمانشاه	
پاسخ: ب	

شکل ۴-۱. مثال‌هایی از سؤال‌های چندگزینه‌ای نوع مستقیم و جمله ناقص

در این مثال‌ها، تنها یک پاسخ صحیح وجود دارد. مرکز استان خوزستان، اهواز است، نه جای دیگر. همه گزینه‌های دیگر، غلط هستند. به بیان دقیق‌تر، این نوع سؤال، جزء سؤال‌های نوع تنها گزینه درست یا نوع صحیح- پاسخ^۱ چندگزینه‌ای است. همه سؤال‌ها را نمی‌توان به این شکل نوشت، زیرا در بسیاری از مواقع، تنها یک پاسخ صحیح وجود ندارد؛ به خصوص هنگام فراتر رفتن از سطح دانش، سؤال‌هایی مطرح می‌شوند که از نظر صحت پاسخ، هریک از گزینه‌ها تا حدودی صحیح‌اند، اما یک گزینه، از بقیه صحیح‌تر است (برای مثال، سؤال‌هایی که با چرا، چه کسی، چگونه، چه چیزی و کجا شروع می‌شوند)، این‌ها نوع بهترین گزینه^۲ نام دارند. این نوع به‌ویژه برای اندازه‌گیری پیامدهای یادگیری که نیازمند درک، کاربرد یا تفسیر اطلاعات واقعی است، مناسب‌اند (میلر، لین و گرانتلد، ۲۰۰۹).

نوع بهترین پاسخ، سخت‌تر از نوع تنها پاسخ صحیح است، زیرا این نوع سؤال‌ها سطوح بالاتر شناختی را اندازه‌می‌گیرند.

این نوع آزمون‌ها، متداول‌ترین نوع آزمون‌های عینی هستند، به دلیل: یکتااختی برش‌ها، حساسیت کم در مقابل حدس زدن بدون آگاهی و آسانی تصحیح پاسخ‌نامه.

آزمونی دارای همه فاکتورهاست که به کمک آن بتوان مسائل ساده تا پیچیده را اندازه‌گرفت و با انواع موضوعات درسی، سازگار است. به دلیل این کاربرد گسترده، بیشتر از سؤال‌های چندگزینه‌ای در آزمون‌های استاندارد استفاده می‌شود. توصیه می‌شود در آزمون‌های کلاسی، از سؤال‌های چندگزینه‌ای، زیاد استفاده نشود، زیرا توانایی سازمان‌دهی و بیان ایده را به دانش‌آموز نمی‌دهد. برای چنین هدفی، نوعی از آزمون‌های عملکردی مناسب‌ترند.

مثال‌ها	نوع بهترین پاسخ
کدام یک از عوامل زیر، در انتخاب تهران به عنوان پایتخت، سهم بیشتری داشته است؟	
الف) بزرگراه خوب	
ب) آب‌وهوای مناسب	
ج) جمعیت زیاد	
د) زمین مرغوب	
پاسخ: د	

شکل ۴-۲. مثالی از سؤال چهارگزینه‌ای، نوع بهترین پاسخ

۴-۸-۱ محاسن آزمون‌های چندگزینه‌ای

۱. تحلیل سؤال‌ها نشان‌دهنده میزان دشواری هر سؤال است و اینکه تمایز بین افراد قوی و ضعیف در کلاس، چگونه ایجاد شود.
۲. عملکرد از کلاسی به کلاس دیگر و از سالی به سال دیگر، قابل مقایسه است.
۳. این آزمون‌ها از سایر آزمون‌ها انعطاف‌پذیرترند.
۴. در یک زمان محدود، می‌توان تعداد زیادی از هدف‌های آموزشی را اندازه‌گیری کرد.
۵. در مقایسه با آزمون‌های صحیح-غلط، امکان حدس زدن کورکورانه کمتری به آزمون‌شونده می‌دهند.
۶. پاسخ‌ها به سادگی و با عینیت کامل، قابل تصحیح‌اند.
۷. اگر گزینه‌های انحرافی به خوبی تهیه شوند، منبع بسیار مناسبی برای تشخیص مشکلات دانش‌آموزان خواهند بود.
۸. از سایر آزمون‌ها انعطاف‌پذیرترند و به وسیله آن‌ها می‌توان افزون‌بر دانش، توانایی استدلال، قضاوت و بسیاری از بازده‌های مهم یادگیری را نیز سنجید.

۴-۸-۲ معایب آزمون‌های چندگزینه‌ای

۱. تهیه سؤال‌های خوب، به‌زمان زیادی نیاز دارد و به‌همین دلیل، ساختن این آزمون‌ها و طراحی گزینه‌های انحرافی مناسب، دشوار است.
۲. برای اندازه‌گیری بعضی از انواع حل مسئله، توانایی سازمان‌دهی و بیان ایده‌ها، کارآمد نیست.
۳. حل مسائل در دنیای واقعی، متفاوت است. وقتی در دنیای واقعی، فرد با مسئله‌ای مواجه می‌شود، برای حل مسئله، چهار گزینه در اختیار وی قرار نمی‌گیرد.
۴. این‌گونه سؤال‌ها تا حد زیادی به‌توانایی خواندن دانش‌آموزان و توانایی نوشتن آموزش‌دهندگان وابسته‌اند.
۵. معلمان اغلب سؤال‌های چندگزینه‌ای را در حد سنجش اطلاعات جزئی و کم‌اهمیت می‌نویسند.
۶. مستلزم صرف وقت زیاد است.
۷. در صورت اعمال نمره منفی، دانش‌آموزانی که خطری می‌کنند، نمره بهتر از سایر دانش‌آموزان می‌گیرند.
۸. گاهی بیشتر از یک پاسخ «صحیح» قابل دفاع، وجود دارد.
۹. اغلب روی آزمودن اطلاعات واقعی متمرکز می‌شوند و کمتر به‌آزمون سطوح عالی‌تر تفکر شناختی می‌پردازند. به‌عبارت دیگر، اغلب این آزمون‌ها برای سنجش اطلاعات جزئی و کم‌اهمیت مورداستفاده قرار می‌گیرند. البته این نقص، به‌طراح سؤال برمی‌گردد، نه به‌آزمون.
۱۰. توانایی نوشتن را اندازه نمی‌گیرند.
۱۱. دانش‌آموزان قوی به‌سبب نظرات خارج از گزینه‌های مطرح‌شده و غیرمعمولی، نه‌تنها تشویق نمی‌شوند بلکه در محدودیت نیز قرار می‌گیرند.
۱۲. مشوق حدس‌زدن است (زیمارو، ۲۰۱۰).

۴-۸-۳ مهم‌ترین قواعد تهیه سؤال‌های چندگزینه‌ای

۱. نوشتن سؤال‌ها را به‌گونه‌ای شروع کنید که برای ویرایش آن‌ها فرصت کافی داشته باشید.

۲. تنه سؤال باید به روشنی، مسئله‌ای را بیان کند. تا حد امکان، گزینه‌ها کوتاه باشد. تنها مواردی را در سؤال بگنجانید که به شفاف‌تر و اختصاصی‌تر شدن سؤال کمک می‌کند. مطمئن شوید که مسئله مطرح شده، روشن و بدون ابهام است.

۳. سعی کنید سؤال‌هایی بنویسید که در آن‌ها یک پاسخ و تنها یک پاسخ صحیح وجود داشته باشد یا مشخصاً بهترین پاسخ وجود داشته باشد.

سؤال ضعیف. یک پسر کوررنگ، کوررنگی خود را از چه کسی به ارث می‌برد؟

- (الف) پدر
(ب) مادر
(ج) پدر مادر
(د) پدر پدر

سؤال بهتر. یک پسر کوررنگ به احتمال زیاد، کوررنگی خود را از چه کسی به ارث می‌برد؟

- (الف) پدر
(ب) مادر
(ج) پدر مادر
(د) پدر پدر

اشاره‌هایی مانند زمین، افعال، مفرد و جمع بودن و شباهت تنه سؤال با گزینه‌های خاص می‌تواند راهنمای خوبی برای افراد بی‌اطلاع از پاسخ درست باشد.

مثال. اسناد مالکیت املاک مردم، در کجا ثبت می‌شود؟

- (الف) ثبت اسناد و املاک
(ب) ثبت احوال
(ج) اقتصاد و دارایی
(د) بانک مرکزی

۴. در هر سؤالی تنها یک ایده اصلی گنجانده شود.

۵. مطمئن باشید گزینه‌های غلط (انحرافی)، حداقل مقبولیت را دارند. گزینه‌های انحرافی طوری نوشته شود که آزمون‌شوندگان بی‌اطلاع را به خود جلب کنند. گزینه‌ها ظاهری منطقی داشته باشند و در یک بررسی سطحی، درست جلوه کنند. بدین منظور معلم می‌تواند در انتخاب گزینه‌های انحرافی، از اشتباهات معمول شاگردان استفاده کند.

۶. موضوع اصلی را در تنه سؤال و مطالب اصلی سؤال را به‌طور کامل در متن سؤال بنویسید. تنه سؤال طوری نوشته شود که با پنهان کردن گزینه‌ها، خواسته مورد نظر سؤال به‌طور کامل مطرح شده باشد و ناقص نباشد. در غیراین صورت باید تجدیدنظر شود.

۷. در استفاده از عبارات منفی یا منقی مضاعف^۱ افراط نکنید.
۸. در هر سؤال، طول همه گزینه‌ها باید تقریباً یکسان و (بهتر است) کوتاه باشد.
۹. نباید هیچ‌گونه راهنمای گرامری برای اشاره به پاسخ صحیح وجود داشته باشد.
۱۰. گزینه‌ها را عمودی بچینید، نه افقی (هر گزینه، روی یک سطر باشد).
۱۱. به‌وفور از عباراتی مانند «همه موارد بالا» یا «هیچ کدام» استفاده نکنید.
۱۲. برای افزایش سطح دشواری سؤال، شباهت گزینه‌ها را به‌همدیگر زیاد کنید.
۱۳. مطالب تکراری گزینه‌ها را در تنه سؤال قرار دهید. رعایت این قاعده، موجب صرفه‌جویی در نوشتن، وقت خواندن، تکثیر سؤال و ارتباط بهتر با آزمون‌شونده خواهد شد.
۱۴. مطمئن شوید هر سؤال، مستقل از سؤال‌های دیگر است. پاسخ به یک سؤال، نباید راهگشای پاسخ دادن به سؤال بعدی باشد.
۱۵. سؤال‌ها را با پیامدهای موردنظر، هماهنگ‌سازید و سؤال‌های با سطح دشواری متوسط انتخاب کنید تا ارزیابی معتبری از هدف‌های آموزشی به‌عمل آورید.
۱۶. طول گزینه درست را در سؤال‌های مختلف تغییر دهید. به‌طور معمول گزینه کلید، طولانی‌تر از گزینه‌های انحرافی نوشته می‌شود و این خود راهنمای خوبی برای آزمون‌شوندگان بی‌اطلاع خواهد بود.
۱۷. هر سؤال، یک موضوع مهم را اندازه بگیرد.
۱۸. تعداد گزینه‌ها را از ۳ تا ۵ مورد انتخاب کنید. متخصصان سنجش عقیده دارند دانش‌آموزان کلاس دوم، ۳ گزینه، سوم و چهارم ۳ تا ۴ گزینه و ششم و بالاتر، ۴ تا ۵ گزینه را می‌توانند پاسخ دهند.
۱۹. گزینه‌های یک سؤال، با هم تجانس داشته باشد.
۲۰. محل گزینه درست را به‌طور تصادفی انتخاب کنید. دادن شماره به‌طور تصادفی یا ترتیب الفبایی.
۲۱. در سؤال‌های دارای تنه ناتمام، جای خالی را در انتها قرار دهید.
۲۲. تنها در صورت لزوم، از آزمون‌های چندگزینه‌ای استفاده کنید. برای مسائل محاسباتی ریاضی و علوم، از «کوتاه پاسخ»، سنجش واقعیت‌ها و عقاید از «صحیح-غلط» و زمانی که انتخاب گزینه انحرافی دشوار است و تعدادی مطالب متجانس یافت می‌شود، از «چورکردنی» استفاده کنید.

یادآوری ۱. هر سؤال چندگزینه‌ای، از قسمت اصلی یا تنه سؤال، گزینه درست یا پاسخ و گزینه‌های انحرافی تشکیل شده است.

یادآوری ۲. سؤال‌هایی که با چرا، چه کسی، چگونه، چه چیزی و کجا شروع می‌شوند، «بهترین نوع گزینه» نام دارند.

یادآوری ۳. یکی از قواعد تهیه سؤال‌های چندگزینه‌ای این است که برای افزایش سطح دشواری سؤال، شباهت گزینه‌ها را به هم زیاد کنید.

در خاتمه یادآوری می‌شود که یکی از اصول مهم آزمون‌سازی به کارگیری آن نوع سؤالی است که بهتر از بقیه بتواند هدف شما را به طور مستقیم و مؤثر سنجش کند. از آنجا که هیچ‌یک از آزمون‌ها به تنهایی نمی‌توانند به خوبی از عهده سنجش تمام هدف‌های سطوح مختلف یادگیری برآیند، بهتر است از ترکیبی از آزمون‌ها و نیز انواع گوناگون پرسش‌ها استفاده شود.

خلاصه فصل چهارم

در این فصل، درباره نحوه تهیه سؤال‌های آزمون‌های عینی، شامل آزمون‌های صحیح-غلط، جورکردنی، کامل‌کردنی و چندگزینه‌ای بحث شد. آزمون‌های سستی سنجش معمولاً شامل انواع مختلفی از جمله: چندگزینه‌ای، پرکردن جای خالی، کامل کردن جمله، صحیح-غلط، جورکردنی و انشایی هستند.

سؤال‌های صحیح-غلط. این نوع سؤال‌ها معمولاً به صورت یک جمله خبری مثبت نوشته می‌شوند و آزمودنی باید صحیح یا غلط بودن آن را تشخیص دهد. متداول‌ترین استفاده از سؤال‌های صحیح-غلط، اندازه‌گیری توانایی تشخیص صحت جمله‌های مربوط به واقعیت‌ها، تعریف اصطلاحات، جمله‌هایی درباره اصول و مانند آن است.

سؤال‌های بله-نه. در این نوع سؤال‌ها، یک جمله سؤالی در اختیار آزمودنی قرار می‌گیرد و او باید جواب بله یا نه را مشخص کند.

سؤال‌های از نوع اصلاحی. در این نوع سؤال‌ها، به آزمودنی جمله‌ای داده می‌شود که در آن غلطی وجود دارد و او باید آن را تشخیص دهد، آن را خط بزند و درستی را بنویسد. در نوع اصلاحی، بهتر است در زیر بخشی از جمله که قرار است مورد قضاوت آزمودنی قرار گیرد، خط کشیده شود.

سؤال‌های از نوع خوشه‌ای. در این نوع سؤال‌ها که به آن سؤال چند صحیح - غلط نیز می‌گویند، تنه سؤال به صورت یک جمله ناتمام نوشته می‌شود و به دنبال آن، تعدادی جواب قرار می‌گیرد. در این سؤال‌ها، هریک از جواب‌های پیشنهادی باید به عنوان صحیح یا غلط، مورد قضاوت قرار گیرد. سپس به قواعد تهیه سؤال‌های صحیح - غلط، محاسن و معایب آن‌ها اشاره شد.

سؤال‌های جورکردنی. این سؤال‌ها شبیه سؤال‌های چندگزینه‌ای هستند، با این تفاوت که به جای چهار گزینه، دو ستون دارند که آزمودنی باید آن‌ها را به همدیگر پیوند دهد. در این نوع آزمون‌ها معمولاً دو ستون، یکی ستون پرسش و دیگری پاسخ وجود دارد که آزمون‌شونده باید پاسخ را با پرسش مربوط به آن جور کند. همچنین در ادامه مطالب، به قواعد تهیه سؤال‌های جورکردنی، محاسن و معایب آن‌ها اشاره شد.

سؤال‌های کامل کردنی. سؤال‌های کوتاه پاسخ و کامل کردنی، هر دو جزء آزمون‌های تدارک پاسخ هستند که بایستی به وسیله یک کلمه، عبارت، عدد یا نماد، به آن‌ها پاسخ داده شود. سؤال کامل کردنی، جمله ناقصی را ارائه می‌دهد که باید کامل شود. آزمون‌های کوتاه پاسخ، برای اندازه‌گیری هدف‌های شناختی سطوح پایین، به ویژه دانش یا معلومات مفیدند.

آزمون‌های چندگزینه‌ای. آزمون‌های چندگزینه‌ای، بهترین نوع آزمون عینی هستند. به کمک این آزمون می‌توان اطلاعات بسیار زیادی را در یک جلسه آزمود. همچنین آنچه که سؤال‌های صحیح - غلط، جورکردنی و کوتاه پاسخ اندازه‌گیری می‌گیرند، به راحتی و به طور مؤثر مورد سنجش قرار داد. سؤال‌های چندگزینه‌ای، برای اندازه‌گیری سطوح دانش و دیگر پیامدهای یادگیری مورد استفاده قرار می‌گیرد. همچنین در ادامه مطالب، به قواعد تهیه سؤال‌های چندگزینه‌ای، محاسن و معایب آن‌ها اشاره شد.

آزمونی دارای همه شرایط است که به کمک آن بتوان مسائل ساده تا پیچیده را اندازه گرفت و با انواع موضوعات درسی سازگار است. از آنجا که هیچ‌یک از آزمون‌ها

به‌تنهایی نمی‌توانند از عهده سنجش تمام هدف‌های سطوح مختلف یادگیری برآیند، بهتر است از ترکیبی از آزمون‌ها و نیز انواع گوناگون پرسش‌ها استفاده شود.

خودآزمایی چهارگزینه‌ای فصل چهارم

۱. احتمال حدس پاسخ صحیح در آزمون‌های عینی چند است؟
 الف) مشخص نیست
 ب) بیشتر از ۱
 ج) کمتر از ۵۰ درصد
 د) بیشتر از ۵۰ درصد
۲. کدام نوع سؤال، برای کودکان کوچک که هنوز نوشتن پاسخ را یادنگرفته‌اند، مفید است؟
 الف) دوگزینه‌ای
 ب) سه‌گزینه‌ای
 ج) چهارگزینه‌ای
 د) کوتاه‌پاسخ
۳. در کدام نوع از آزمون‌های دوگزینه‌ای، جمله‌ای در اختیار آزمودنی قرار می‌گیرد و او باید غلط موجود در آن را تشخیص دهد و درستش را بنویسد؟
 الف) صحیح - غلط
 ب) بله - خیر
 ج) خوشه‌ای
 د) اصلاحی
۴. در کدام نوع از آزمون‌های زیر، افزون‌بر بازشناسی، یادآوری هم صورت می‌گیرد؟
 الف) صحیح - غلط
 ب) بله - خیر
 ج) خوشه‌ای
 د) اصلاحی
۵. کدام گزینه، جزء قواعد تهیه سؤال‌های صحیح - غلط است؟
 الف) سؤال‌ها باید تا حد امکان دشوار باشند.
 ب) از نوشتن جمله‌های منفی مضاعف پرهیزید.
 ج) طول سؤال‌های صحیح، بیشتر از سؤال‌های غلط باشد.
 د) در هر سؤال، حداقل دو هدف آموزشی مطرح شود.
۶. بهترین سؤال در آزمون‌های عینی، سؤالی است که هدف آموزشی را مد نظر قرار دهد؟
 الف) ۱
 ب) ۲
 ج) ۳
 د) ۴

۷. علل باارزش بودن آزمون‌های صحیح - غلط چیست؟
 الف) نوشتن تعداد زیادی سؤال در زمان کم
 ب) وجود تعداد زیاد پاسخ
 ج) عینی بودن تصحیح سؤال
 د) پیچیده بودن مراحل تهیه سؤال
۸. احتمال حدس پاسخ درست در سؤال‌های کوتاه پاسخ چقدر است؟
 الف) ۰ درصد
 ب) ۲۵ درصد
 ج) ۵۰ درصد
 د) ۷۵/۰ درصد
۹. کدام مورد، جزء معایب آزمون‌های صحیح - غلط نیست؟
 الف) اغلب برای سنجش اطلاعات جزئی به کار برده می‌شوند.
 ب) آزمون‌های صحیح - غلط، به راحتی تهیه می‌شوند.
 ج) حدس زدن کورکورانه در آن‌ها خیلی زیاد است.
 د) برای سنجش هدف‌های آموزشی کم اهمیت به کار برده می‌شوند.
۱۰. در تهیه سؤال‌های جورکردنی، کدام گزینه به طور صحیح مطرح نشده است؟
 الف) اشخاص، عناوین کتاب‌ها
 ب) قواعد، مفاهیم
 ج) ماشین‌ها، کاربردها
 د) اصطلاحات، تبیین‌ها
۱۱. در مورد قواعد تهیه سؤال‌های جورکردنی، کدام گزینه صحیح نیست؟
 الف) پرسش‌ها و پاسخ‌های متجانس انتخاب کنید.
 ب) در لیست پرسش‌ها، تقریباً بیش از ده پرسش نباشد.
 ج) در صورت امکان، فهرست پاسخ‌ها را به طور منطقی مرتب کنید.
 د) پرسش‌ها را با حروف و پاسخ‌ها را با عدد مشخص کنید.
۱۲. نمره گذاری کدام نوع از آزمون‌ها، از بقیه دشوارتر است؟
 الف) تستی
 ب) دوگزینه‌ای
 ج) کوتاه پاسخ
 د) اصلاحی
۱۳. در کدام نوع از آزمون‌های زیر، احتمال حدس زدن به صفر می‌رسد؟
 الف) صحیح - غلط
 ب) کوتاه پاسخ
 ج) جورکردنی
 د) چندگزینه‌ای

۱۴. کدام مورد، جزء پیشنهادهای مربوط به سؤال‌های کوتاه‌پاسخ نیست؟
 الف) تا حد امکان، به جای سؤال‌های پرشی، از سؤال‌های کامل‌کردنی استفاده کنید.
 ب) هر سؤال باید موضوع مهمی را شامل شود.
 ج) در سؤال‌های کامل‌کردنی، تعداد زیادی جای خالی منظور نکنید.
 د) صورت سؤال را با بیان مطالب شبیه جمله‌های کتاب ننویسید.

۱۵. بهترین نوع پرسش‌های چندگزینه‌ای، گزینه‌ای است.

الف) ۲

ب) ۳

ج) ۴

د) ۵

۱۶. کدام مورد، جزء محاسن استفاده از آزمون‌های چندگزینه‌ای است؟
 الف) از سایر آزمون‌ها انعطاف‌پذیرترند.
 ب) ساختن این آزمون‌ها دشوار است.
 ج) مستلزم صرف وقت زیاد است.
 د) توانایی نوشتن را اندازه‌ نمی‌گیرند.

۱۷. کدام مورد، جزء معایب استفاده از آزمون‌های چندگزینه‌ای است؟

- الف) پاسخ‌ها به صورت عینی، قابل تصحیح‌اند.
 ب) برای سنجش دانش، توانایی و استدلال مفیدند.
 ج) حدس زدن را تشویق می‌کنند.
 د) عملکرد، از سالی به سال دیگر، قابل مقایسه است.

۱۸. احتمال حدس پاسخ صحیح به صورت کورکورانه، در کدام یک از آزمون‌های ز کمتر است؟

الف) دوگزینه‌ای

ب) سه‌گزینه‌ای

ج) چهارگزینه‌ای

د) پنج‌گزینه‌ای

۱۹. کدام مورد، جزء قواعد تهیه سؤال‌های چندگزینه‌ای نیست؟

- الف) هر سؤال، موضوع موردنظر را به وضوح بیان کند.
 ب) در هر سؤال، فقط یک ایده اصلی گنجانده شود.
 ج) به طور افراطی، از عبارت منفی مضاعف استفاده شود.
 د) گزینه‌ها را عمودی بچینید، نه افقی.

۲۰. اگر یک آزمودنی، به یک آزمون ۲۰ سؤالی پاسخ دهد و از این ۲۰ سؤال، پاسخ صحیح ۵ سؤال را بنویسد، بدون در نظر گرفتن نمره منفی، نمره اش چند خواهد شد؟
- الف) ۰
ب) ۵
ج) ۱۰
د) ۱۵

خودآزمایی تشریحی فصل چهارم

۱. نکات لازم برای آماده سازی سؤال‌ها در طول دوره تحصیلی کدام‌اند؟ (دو مورد)
۲. انواع سؤال‌های صحیح-غلط را نام ببرید و برای هر کدام مثال بزنید؟
۳. چهار مورد از قواعد تهیه سؤال‌های صحیح-غلط را بنویسید؟
۴. معایب آزمون‌های صحیح-غلط را بنویسید؟
۵. از قواعد تهیه سؤال‌های جورکردنی، به چهار مورد اشاره کنید؟
۶. محاسن و معایب آزمون‌های جورکردنی (هر کدام دو مورد) را بنویسید؟
۷. مزایا و محدودیت‌های سؤال‌های کوتاه پاسخ را بنویسید؟
۸. سه مورد از قواعد تهیه سؤال‌های عددی را بنویسید؟
۹. منظور از آزمون چندگزینه‌ای چیست و چه کاربردی می‌توانند داشته باشند؟
۱۰. چرا توصیه می‌شود در آزمون‌های کلاسی، از سؤال‌های چندگزینه‌ای، به وفور استفاده شود؟
۱۱. برای هریک از محاسن و معایب آزمون‌های چندگزینه‌ای، سه مورد بنویسید؟
۱۲. چهار مورد از مهم‌ترین قواعد تهیه سؤال‌های چندگزینه‌ای را بیان کنید؟
۱۳. چگونه می‌توان سطح دشواری سؤال‌ها را افزایش داد؟
۱۴. به کمک کدام آزمون‌ها می‌توان به میزان یادگیری مطالب از سوی یادگیرنده پی برد؟
۱۵. آیا آزمون‌های صحیح-غلط، به صورت یک جمله خبری نوشته می‌شوند؟
۱۶. احتمال حدس پاسخ درست در سؤال سه‌گزینه‌ای، بدون خواندن پاسخ، چند درصد است؟

فصل پنجم

نوشتن سؤال‌های انشایی

هدف کلی

تهیه انواع آزمون‌های انشایی و قواعد نمره‌گذاری آنها

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. هدف از تهیه آزمون‌های انشایی را توضیح دهد.
۲. انواع آزمون‌های انشایی را نام ببرد.
۳. آزمون‌های انشایی محدود پاسخ و گسترده پاسخ را شرح دهد.
۴. قواعد تهیه سؤال‌های انشایی را بازگو کند.
۵. قواعد نمره‌گذاری سؤال‌های انشایی را بیان کند.
۶. محاسن و معایب آزمون‌های انشایی را بنویسد.

مقدمه

در فصل قبل، آزمون‌های عینی بررسی شدند. اشاره شد که این آزمون‌ها می‌توانند پیامدهای ساده و پیچیده یادگیری را اندازه‌گیری کنند. با این حال، این آزمون‌ها قادر به سنجش بعضی پیامدهای یادگیری نیستند. برای سنجش این پیامدها آزمون‌های انشایی مناسب‌ترند. در این فصل، درباره آزمون‌های انشایی، انواع آن، مزایا و معایب چنین آزمون‌هایی و قواعد تهیه آنها بحث می‌شود.

۵-۱ آزمون‌های انشایی

بعضی پیامدهای مهم یادگیری ممکن است به وسیله سؤال‌های انشایی بازپاسخ یا دیگر آزمون‌های عملکردی، بهتر اندازه‌گیری شوند. این سؤال‌ها به آزمودنی کمک می‌کنند توانایی سازمان‌دهی، یکپارچه‌سازی و ارزیابی عقاید و اطلاعات را به‌طور روشن نشان‌دهند و دانش و مهارت‌هایش را به‌کارگیرد (سانتراک، ۲۰۱۱).

آزمون‌های انشایی یا گسترده‌پاسخ^۱، آزمون‌هایی هستند که در آن‌ها سؤال‌ها در اختیار آزمون‌شونده گذاشته می‌شود و او جواب سؤال‌ها را خود آماده می‌کند و در برگه امتحانی می‌نویسد (سانتراک، ۲۰۱۱). بسیاری از معلمان، از آزمون‌های انشایی فقط برای اندازه‌گیری اطلاعات واقعی استفاده می‌کنند. این کار، برای سنجش توانایی خلاصه‌کردن یک داستان به‌زبان خود، بیان علل یک رویداد تاریخی یا گام‌های یک فرایند علمی که می‌توان آن‌ها را به‌کمک نوشتن نشان‌داد، قطعاً مفید است. هرچند اندازه‌گیری دانش اطلاعات واقعی به‌کمک آزمون‌های انشایی، مفید و معتبر است، ولی کارایی آزمون‌های انشایی، خیلی بیشتر از این‌هاست (سانتراک، ۲۰۱۱).

۵-۲ انواع آزمون‌های انشایی

درکل، آزمون‌های انشایی را با توجه به آزادی عمل آزمون‌شونده در پاسخ‌دادن به سؤال‌های آزمون، به دو دسته گسترده‌پاسخ و محدودپاسخ^۲ تقسیم می‌کنند. نمونه‌ای از امتحان‌های بازپاسخ، عبارت‌اند از: امتحان‌های بازپاسخ^۳ آخر ترم، مقاله‌های ترمی^۴ یا آزمون‌های کتاب‌باز^۵ که از نظر تعداد صفحات، انعطاف‌پذیرند (برای مثال، ۱۰-۱۲ صفحه یا کمتر از ۲۰ صفحه) آزمون گسترده‌پاسخ و یک‌صفحه‌ای یا کمتر از آن، آزمون‌های محدودپاسخ^۶ نامیده می‌شوند.

۵-۳ آزمون‌های انشایی گسترده‌پاسخ

آزمون‌های انشایی گسترده‌پاسخ، برای سنجش توانایی ارزیابی‌کردن، تحلیل‌کردن،

سازمان‌دادن، آفریدن و انتخاب دیدگاه‌ها مناسب‌ترند. این آزمون‌ها به دانش‌آموز اجازه می‌دهند طول و پیچیدگی پاسخ را خودش تعیین کند. در آزمون گسترده پاسخ، برای پاسخ آزمون‌شوندگان هیچ محدودیتی وجود ندارد و او آزاد است بنا بر خواست خودش، پاسخ خود را به‌پروراند و سازمان‌دهد.

در واقع مشخصه بارز آزمون انشایی، آزادی عمل در پاسخ‌دهی است. دانش‌آموزان در نوشتن، مرتبط‌ساختن و ارائه عقاید به‌زبان خود آزادند (کوهن، ۲۰۰۹). این آزادی عمل، ارزش سؤال‌های انشایی را در اندازه‌گیری سطوح پیچیده شناختی بالایی برد، اما دشواری نمره‌گذاری، آن را تا حد ابزاری برای اندازه‌گیری دانش امور واقعی پایین می‌آورد. بهتر است دانش امور واقعی، به وسیله آزمون‌های عینی اندازه‌گیری شود و آزمون‌های انشایی برای اندازه‌گیری آن دسته از توانایی‌ها مورد استفاده قرارگیرد که به کمک دیگر آزمون‌ها به راحتی قابل اندازه‌گیری نیستند (میلر، لین و گرانلند، ۲۰۰۹).

به عبارت دیگر، همانند آزمون‌های عینی، آزمون‌های انشایی نیز ممکن است از کیفیت بالا یا پایینی برخوردار باشند. سؤال‌های خوب، توانایی‌های شناختی پیچیده را که مستلزم سازمان‌دهی، یکپارچه‌سازی و ترکیب دانش برای حل مسائل جدید است، اندازه‌گیری‌کنند، اما سؤال‌های ضعیف، فقط به اندازه‌گیری اطلاعات ساده در سطح به یاد آوردن می‌پردازند (کوبیزین و بوریچ، ۲۰۱۳).

۴-۵ آزمون‌های انشایی محدود پاسخ

یک آزمون انشایی که مسئله خاصی را مطرح می‌کند و دانش‌آموز باید اطلاعات مناسب را به یاد آورد، آن را به شیوه‌ای مناسب سازمان‌دهد، نتیجه‌گیری قابل دفاعی ارائه دهد و آن را در قالب محدودیتی که از نظر تعداد صفحات یا زمان مطرح شده است، پاسخ دهد، آزمون محدود پاسخ نامیده می‌شود. در آزمون‌های انشایی محدود پاسخ، آزمون‌شونده در پاسخ دادن به سؤال‌ها آزادی کامل ندارد، بلکه صورت سؤال، او را ملزم می‌سازد تا پاسخ خود را در چهارچوب شرایطی خاص محدود کند.

سؤال‌های آزمون محدود پاسخ، احتمالاً بیشتر برای سنجش به یاد آوردن، فهمیدن و پیامدهای یادگیری سطح به کار بستن به کار می‌روند. به عبارت دیگر همان‌طور که مطرح شد، در آزمون‌های محدود پاسخ، آزمون‌شونده در دادن پاسخ به سؤال‌ها آزادی

کامل ندارد و سازندهٔ آزمون، پاسخ‌دهنده را ملزم می‌کند پاسخ‌های خود را در چهارچوب و شرایطی محدود سازد.

مثال. شباهت‌ها و تفاوت‌های عمدهٔ نظریهٔ پاولف و اسکینر را در یک پاراگراف توضیح دهید. نمرهٔ شما بستگی به صحت، سازمان‌دهی و اختصار دارد.

مثال. «در یک پاراگراف، به‌طور خلاصه توضیح دهید چرا فشارسنج، بهترین ابزار پیش‌بینی وضع هواست؟»

از آنجا که آزمون‌های محدود پاسخ، ساختاریافته‌تر از آزمون‌های گسترده پاسخ‌اند، برای اندازه‌گیری پیامدهای یادگیری که مستلزم تفسیر و کاربرد داده‌ها در یک حوزهٔ خاص است، مناسب‌ترند. استفاده از این سؤال‌ها، متداول‌تر از سؤال‌های گسترده پاسخ است.

۵-۵ مقایسهٔ پیامدهای یادگیری اندازه‌گیری‌شدهٔ آزمون‌های محدود پاسخ

در پایین، بعضی از شرایط مناسب‌تر بودن آزمون‌های محدود پاسخ، توصیف شده‌اند:

- در آزمون‌های محدود پاسخ، هدف‌های آموزشی، مستلزم تدارک اطلاعات هستند تا فقط تشخیص اطلاعات. این فرایندها اغلب با آزمون‌های عینی، قابل اندازه‌گیری نیستند.
- بخش کوچکی از محتوا اندازه‌گیری می‌شود. اگر شما ۳۰ دانش‌آموز دارید و آزمونی با شش سؤال محدود پاسخ طرح کنید، زمان زیادی را صرف نمره‌گذاری آزمون خواهید کرد. هنگامی که تعداد دانش‌آموزان کم است، از آزمون‌های محدود پاسخ استفاده کنید یا از آن در ترکیب با آزمون‌های عینی استفاده شود.
- امنیت آزمون، قابل توجه است. اگر نگران تقلب در آزمون‌های چهارگزینه‌ای هستید، بهتر است از یک آزمون محدود پاسخ استفاده کنید. در کل، یک آزمون انشایی محدود پاسخ خوب، زمان کمتری از آزمون‌های عینی می‌طلبد.
- بعضی از پیامدهای یادگیری که نشان می‌دهد برای اندازه‌گیری، بهتر است آزمون‌های محدود پاسخ استفاده شود، به شرح زیر هستند:

• تحلیل رابطه

مثال. رنگ‌های آبی و خاکستری، با دمای سرد رابطه دارند. رنگ‌های دیگر با چه دمای رابطه دارند؟ این رنگ‌ها چه تأثیری روی نقاشی‌ای که ترسیم خواهید کرد، دارد؟

- مقایسه و مقابله وضعیت‌ها

مثال. دو شخصیت را از داستان‌هایی که خوانده‌اید، با هم مقایسه کنید تا بفهمید که آن‌ها در داستان چگونه به صورت متفاوت، به وضعیت‌ها پاسخ داده‌اند.

- بیان فرض‌های ضروری

مثال. هنگامی که کریستف کلمب وارد سان‌سالوادور شد، درباره سرزمینی که کشف کرده بود، چه تصویری داشت؟ کدام فرض‌هایش درست بودند؟

- تشخیص نتیجه‌گیری‌های مناسب

مثال. بعضی دلایل موافق و مخالف ساخت محل‌های دفن زیاده در نزدیکی خانه‌ها چیست؟

- تبیین رابطه علت و معلولی

مثال. چه عاملی باعث شد آمریکایی‌های اولیه، در دهه ۱۷۸۰ به غرب سفر کنند؟

- تدوین فرضیه

مثال. با مشاهده ابرها، پیش‌بینی شما در مورد طوفان‌های آینده چیست؟ توضیح دهید چه عاملی کمک می‌کند با دیدن ابر، باران را پیش‌بینی کنید.

- سازمان‌دهی داده‌ها به منظور حمایت از یک دیدگاه

مثال. روی تخته، تعداد خانه‌های تازه‌ای که ساخته خواهند شد و تعداد اتومبیل‌هایی که در سال گذشته در هر ماه خریداری شده‌اند را خواهید یافت. این داده‌ها را به منظور حمایت از این ایده که اقتصاد ما رشد فزاینده یا کاهنده دارد، مورد استفاده قرار دهید.

- اشاره به نقاط قوت و ضعف

مثال. برای یک گروه کوهنوردی، نقاط قوت یا ضعف ابزارهای زیر چیست؟ کفش، جنگک، چادر، طناب

۵-۶ قواعد تهیه سؤال‌های انشایی

بزرگ‌ترین امتیاز آزمون‌های انشایی این است که آزمون‌شوندگان را وادار می‌کند اندیشه‌های خود را به‌طور منطقی، منسجم و سازمان‌یافته بیان کنند و در این کار، خلاقیت خود را نشان دهند. قواعد زیر به‌منظور کاستن از معایب این سؤال‌ها پیشنهاد شده‌اند (سیف، ۱۳۹۶).

۱. در نوشتن صورت سؤال‌ها، با رسم جدول مشخصات آن درس، سؤال‌ها به‌طور مستقیم به‌هدف‌های آموزشی مربوط شوند.
 ۲. محدودیت‌ها را مشخص کنید. طول پاسخ درست و وزنی را که در تعیین نمره‌ها یا قضاوت‌ها به‌هر سؤال اختصاص می‌یابد، مشخص سازید.
 ۳. سؤال‌های انشایی را تنها برای اندازه‌گیری هدف‌هایی اختصاص دهید که به‌کمک سایر آزمون‌های عینی، قابل اندازه‌گیری نیستند. در صورتی که هدف به‌وسیله آزمون عینی قابل اندازه‌گیری است، از آن‌ها استفاده کنید.
 ۴. سؤال‌های انشایی را با کلماتی مانند مقایسه کنید، مقابله کنید، شواهدی ارائه دهید، پیش‌بینی کنید چه اتفاقی خواهد افتاد، اگر و مانند آن شروع کنید. از کاربرد کلمات «چه کسی»، «چه» و «کجا» بپرهیزید، زیرا جواب این سؤال‌ها کوتاه‌پاسخ هستند و فقط نیازمند یادآوری اطلاعات‌اند.
 ۵. سؤال‌هایی طرح کنید که نیازمند مهارت‌های ویژه‌ای و مطابق استاندارهای یادگیری باشد. سؤال‌های گسترده‌پاسخ بدون محدودیت، خیلی مناسب نیستند.
- مثال. یک گزارش دو صفحه‌ای بنویسید که از اهمیت کنسرو کردن منابع طبیعی، دفاع کند (پاسخ شما با توجه به سازمان، جامعیت و ارتباط آن با مباحث ارائه‌شده ارزیابی خواهد شد).
۶. سؤال‌های مربوط به موضوعات و عقاید بحث‌انگیز طوری طرح شوند آزمون‌شونده بتواند برای مستند کردن عقیده انتخابی خود، شواهد لازم گردآوری کند.
 ۷. به آزمون‌شونده حق انتخاب چند سؤال از میان تعدادی سؤال را ندهید. مثال: از سؤال زیر، به ۱۰ سؤال به انتخاب خود جواب بدهید. این کار باعث می‌شود نمره‌گذاری متفاوت باشد، زیرا همه دانش‌آموزان به سؤال‌های یکسانی پاسخ

نمی‌دهند. این کار معمولاً به‌ضرر دانش‌آموزان قوی خواهد بود، زیرا اغلب سؤال‌های

دانش‌آموزان را انتخاب می‌کنند. برای پاسخ دادن به سؤال، زمان کافی و زمان هر سؤال را نیز جداگانه در نظر بگیرید؛ برای سؤال‌های ساده، زمان کمتر و برای سؤال‌های دشوار، زمان بیشتر. زمانی که معلم سؤال را طرح می‌کند، زمان مناسبی برای اختصاص زمان به هر سؤال است. در اختصاص زمان، دانش‌آموزان کنندنویس را مد نظر داشته باشید.

با نوشتن سؤال‌هایی که به جواب کوتاه نیاز دارند، تعداد سؤال‌ها را افزایش دهید و از ضعف نمونه‌گیری بکاهید.

۱۱. عواملی را که در ارزشیابی آزمون انشایی دخالت خواهید داد، از پیش مشخص کنید و به اطلاع آزمون‌شوندگان برسانید. مثل املائی درست کلمات، رعایت قواعد درست کلمات، رعایت قواعد نگارش و خوش خط نویسی.

۱۲. صورت سؤال‌های انشایی را با عبارات و کلمات واضح بنویسید و از کلی‌گویی پرهیزید. صورت سؤال باید روشن باشد تا افراد متفاوت، آن را یکسان درک کنند.

۱۳. برای کودکان کوچک، سؤال‌های انشایی را خیلی طولانی ننویسید. همچنین سؤال نباید به پاسخ طولانی نیاز داشته باشد (استرنبرگ و ویلیامز، ۲۰۰۹).

یادآوری ۱. در آزمون‌های انشایی یا گسترده پاسخ، سؤال‌ها در اختیار آزمون‌شونده گذاشته می‌شود و او جواب سؤال‌ها را خود آماده می‌کند و در برگه امتحانی می‌نویسد.

یادآوری ۲. بهتر است دانش‌آموز امور واقعی به وسیله آزمون‌های عینی و اندازه‌گیری طولانی‌ها به وسیله آزمون‌های انشایی اندازه‌گیری شود.

یادآوری ۳. سؤال‌های آزمون پاسخ محدود، احتمالاً بیشتر برای سنجش به یاد آوردن، فهمیدن و پیامدهای یادگیری سطح به کار بستن به کار می‌روند.

۷-۵ قواعد نمره‌گذاری سؤال‌های انشایی

تصحیح سؤال‌های انشایی، از رأی و نظر مصحح تأثیر می‌پذیرد، به همین دلیل به منظور کاهش یا حذف سوگیری معلم، قواعدی برای نمره‌گذاری پیشنهاد شده است که رعایت

آن، به اعتبار نمره‌گذاری کمک می‌کند. برای تصحیح برگه‌های امتحانی، دو روش کلی وجود دارد: تحلیلی^۱ و کلی^۲. (کویزین و بورج، ۲۰۱۳)

۱-۷-۵ روش نمره‌گذاری تحلیلی
در روش تحلیلی نمره‌گذاری که به آن، روش امتیازبندی نیز گفته می‌شود، پاسخ نمونه به اجزاء کوچک‌تری تقسیم می‌شود و برای هر جزء مشخص، نمره یا امتیاز جداگانه‌ای در نظر گرفته می‌شود. همچنین در این روش، عواملی مانند «قدرت بیان»، «ساختار منطقی پاسخ» و «بیان دلایل»، امتیازهای جداگانه‌ای دریافت می‌کنند (سیف، ۱۳۹۶).

۲-۷-۵ روش نمره‌گذاری کلی
در روش کلی، به کل پاسخ فقط یک نمره تعلق می‌گیرد. در این روش، مصحح پاسخ آزمودنی را می‌خواند و برداشتی کلی از آن به دست می‌آورد و به این برداشت، یک نمره اختصاص می‌دهد. این روش نمره‌گذاری، سریع‌تر از روش تحلیلی است. ایراد این روش نمره‌گذاری این است که اگر نمره‌ای از دانش‌آموز کسر شود، نمی‌داند بابت چه چیزی بوده است و بازخورد دقیقی به وی نمی‌دهد. البته معلم می‌تواند با نوشتن یادداشت‌هایی در حاشیه برگه، به آزمودنی بازخورد بدهد. این کار تا حدودی ابهام را کاهش می‌دهد.

۵-۸ قواعد تصحیح سؤال‌های انشایی

همان‌طور که گفته شد، نمره‌گذاری سؤال‌های محدود پاسخ، دشوار است و از مصحح تأثیری پذیرد؛ یعنی یک نفر ممکن است به سؤال یکسانی، نمره ۵ و دیگری ۳ یا ۲ بدهد. نمره یک مصحح، در زمان‌های مختلف یکسان نیست، ممکن است به همان سؤال نمره ۵، ۳ یا ۲ بدهد.

به منظور درک دشواری‌های اعتبار نمره‌گذاری سؤال‌های انشایی، لازم است به دشواری تهیه سؤال‌های انشایی خوب توجه کنیم. همان‌طور که در قسمت‌های قب اشاره شد، هرچه هدف‌های آموزشی روشن‌تر باشند، تهیه سؤال‌های انشایی ساده‌تر است. هرچه سؤال روشن‌تر باشد، نمره‌گذاری آن نیز ساده‌تر خواهد بود (کویزین و بورج، ۲۰۱۳).

جدول ۵-۱. مثال مقیاس تحلیلی برای انشاهای توضیحی یا خلاصه‌های توصیفی

نمره	برداشت کلی	تمرکز / سازمان	زبان	بسط	سازوکار
۵	پیشرفت انشایی	- ایده اصلی، به روشنی بیان شده - تمرکز و سازمان، یکی شده - به‌طور مؤثر، به‌خواننده جهت می‌دهد	- اختصاصی و عینی - جزئیات با هدف، همخوانی دارد - جزئیات، تصویر واضح و روشنی به‌دست می‌دهد	- یک - ایده اصلی به‌طور گسترده بسط داده شده	- یک یا دو خطای جزئی - بدون خطای عمده
۵	پیشرفت قابل توجه	- ایده اصلی، بیان شده یا به‌طور ضمنی بیان شده - متمرکز و سازمان‌یافته - به‌طور مؤثر به‌خواننده جهت می‌دهد	- جزئیات حسی ویژه - اکثر جزئیات، با هدف همخوان هستند	- بسط کامل یک نکته اصلی	- خطاهای جزئی کم - کمتر از یک خطای عمده
۴	پیشرفت کافی	- ایده اصلی، وجود دارد، اما ممکن است تمرکز همخوان نباشد - تا حدودی به‌خواننده جهت می‌دهد	- بعضی جزئیات ویژه - جزئیات معمولاً روشن است - به‌طور کلی تصویر واضح است	- بسط متوسط نکته اصلی	- بعضی خطاهای جزئی - یک یا دو خطای عمده - خطاها باعث نمی‌شود خواننده سردرگم شود
۳	شواهدی مبتنی بر پیشرفت وجود دارد	- ایده اصلی، واضح نیست - ایده اصلی، وجود دارد، اما سردرگمی‌هایی دارد	- جزئیات، کم یا ناهمخوان است - جزئیاتی وجود دارد، اما همه نه	- بحث نکته اصلی محدود است	- خطاهای جزئی و عمده وجود دارد - خواننده، مقداری سردرگم است
۲	شواهدی پیشرفت اندک است	- نشانه مبهم از ایده اصلی یا تمرکز - انحراف از موضوع، معنادار است	- زبان، چندان عینی نیست	- بسط اندک نکته اصلی	- خطاهای عمده و خرد زیادی وجود دارد - خطاها درک خواننده را مختل می‌سازد
۱	شواهدی پیشرفت اندک است	- ایده اصلی آشکار نیست	- زبان، عینی نیست	- بسط ایده اصلی وجود ندارد	- خطاهای عمده مکرر، باعث سردرگمی خواننده است

رعایت نکات زیر، به کیفیت نمره گذاری سؤال های انشایی کمک خواهد کرد:

۱. سؤال های انشایی خوب بنویسید. سؤال های ضعیف، یکی از منابع ناهمخوانی نمره گذاری بین مصححان مختلف است. نمره گذاری سؤال های پاسخ بلند (برای مثال، سه صفحه ای)، دشوارتر از سؤال های کوتاه پاسخ تر (برای مثال، یک صفحه ای) است. طولانی بودن پاسخ، باعث خستگی دانش آموز می شود و همین امر، بی دقتی و کاهش انگیزه را به دنبال خواهد داشت و بر نمره تأثیر می گذارد.

۲. قبل از تصحیح پاسخ ها، یک پاسخ نمونه تهیه کنید. این کار سبب جلوگیری از دخالت عوامل نامربوط می شود. اهمیت تهیه این پاسخ نمونه از این لحاظ است که در آزمون های انشایی، نظر مصحح دخیل است و این کار باعث کاهش تأثیر پذیری نمره فرد از سلیقه شخصی مصحح می شود. بهتر است پاسخ هر سؤال را به چند قسمت تقسیم کنید و برای هر قسمت، نمره جداگانه ای در نظر بگیرید. برای مثال به منظور سازمان دهی روابط بین اجزاء، مثال های داده شده و مواردی از این قبیل، نمره در نظر بگیرید.

مثال. اگر معلم از دانش آموز بخواهد سه علت شروع جنگ ایران و عراق را بنویسد، ممکن است معلم فهرستی از پاسخ های قابل قبول تهیه کند و به هر دلیل، فقط یک امتیاز و حداکثر سه امتیاز بدهد.

۳. سربرگ سؤال ها را جدا کنید یا به نام دانش آموز توجه نکنید. هدف از جدا کردن سربرگ سؤال ها این است که دیدن نام دانش آموز، بر نمره وی تأثیر نگذارد. به هم زدن برگه ها قبل از تصحیح نیز روش مفیدی است، زیرا باعث کاهش سوگیری می شود. از آنجا که دانش آموزان زرنگ، معمولاً ذهن سازمان یافته تری دارند، سریع تر پاسخ داده، برگه امتحانی خود را زودتر تحویل می دهند. از دانش آموزان بخواهید نامشان را پشت برگه بنویسند. اگر هم دستخط دانش آموز آشناست، بهتر است برای جلوگیری از جهت گیری، تلاش آگاهانه ای به خرج دهید. روش های ابتکاری دیگر نیز به منظور جلوگیری از شناسایی نام دانش آموز از سوی معلم، مفیدند.

۴. از یک قاعده نمره گذاری که مناسب تر است، استفاده کنید. همان طور که در بالا اشاره شد، در نمره گذاری سؤال های انشایی، معمولاً دو روش نمره گذاری وجود دارد: تحلیلی و کلی. روش تحلیلی، در یک زمان روی یک ویژگی توجه دارد و به ویژه در

فراهم کردن بازخورد خاص در مورد جنبه‌هایی از کار دانش‌آموز، مفید است. روش کلی نمره‌گذاری، زمانی مفید است که تمرکز سنجش، بر درک دانش‌آموز از محتوای کلی است تا مهارت نوشتن.

۵. تأثیر عوامل نامربوط بر نمره‌گذاری را حذف کنید. چندین عامل که ربطی به هدف‌های اندازه‌گیری ندارند، به‌طور غیرمستقیم بر ارزیابی نمره دانش‌آموز تأثیری گذارند. برجسته‌ترین آن‌ها عبارت‌اند از: خوانانوشتن، املا، درستی کلمات، جمله‌بندی درست، رعایت قواعد نقطه‌گذاری و پاکیزگی. این نکات را تنها زمانی که امتحان مربوط به خوش‌نویسی، املا و دستور زبان است، منظور کنید. در صورت دخالت دادن این موارد، برای هر کدام نمره جداگانه‌ای در نظر بگیرید.

دخالت دادن این عوامل، گاهی هم مفید است و باعث می‌شود دانش‌آموز دقت کند و هر چیز نامربوطی را ننویسد. از طرفی اگر پاسخ سؤال را هم کاملاً صحیح نوشته باشد، ممکن است کم کردن نمره، ناعادلانه به نظر برسد. به هر حال، نمره کمی اختصاص دهید و از قبل به دانش‌آموز در مورد کم کردن نمره هشدار دهید (کوپیزین و بوریچ، ۲۰۱۳).

۶. تمام پاسخ‌های آزمون‌شوندگان به یک سؤال را تصحیح کنید، سپس به سؤال بعدی بروید. بدین معنا که اول پاسخ سؤال اول همه دانش‌آموزان را تصحیح کنید، سپس سؤال دوم، سوم تا آخر. این روش، امکان مقایسه پاسخ‌های آزمون‌شوندگان مختلف به یک سؤال را فراهم می‌کند. کار نمره‌گذاری به آن سؤال را آسان و امکان کاربرد یک ملاک واحد را به صورت هماهنگ برای همه افراد، فراهم می‌کند. در واقع یکی از عواملی که باعث ناپایاب شدن نمره‌گذاری آزمون‌های انشایی می‌شود، وجود تفاوت در قواعد نمره‌گذاری از یک برگه به برگه دیگر است. وقتی پاسخ‌ها برگه به برگه تصحیح می‌شوند و نه سؤال به سؤال، این مشکل پیش می‌آید که فردی که برگه‌اش پشت سر برگه یک فرد قوی واقع شده است، نمره کمتری بگیرد و برعکس. یک راه برای رفع این مشکل، تصحیح سؤال به سؤال و هم‌زدن برگه‌ها بعد از تصحیح هر سؤال است تا جای برگه‌ها عوض شود و فرد خاصی همیشه پشت سر دیگری قرار نگیرد.

تصحیح برگه به برگه، این مشکل را ایجاد می‌کند که اگر پاسخ دانش‌آموز به چند سؤال اول خوب بود، این ذهنیت در نمره دادن ایجاد می‌شود که فرد، قوی است و

در نتیجه از کاستی‌های وی چشم‌پوشی می‌شود و برعکس. به این خطا، «اثر هالو» یا اثر هاله» گفته می‌شود.

۷. اگر تصمیم مهم است، باید افراد مستقل دوباره برگه‌ها را نمره‌گذاری کنند. گاهی از سؤال‌های انشایی، برای اعطای بورس تحصیلی، اهدای جوایز، آموزش ویژه و مانند آن استفاده می‌شود. در چنین مواردی، برگه‌ها بایستی به وسیله دو نفر مستقل نمره‌گذاری شوند و سپس نمره آن‌ها با همدیگر مقایسه شود. اگر تفاوت نمره‌ها زیاد باشد، باید نفر سومی دوباره برگه‌ها را تصحیح کند (برای انجام این کار، قبل از نمره‌گذاری، از برگه‌ها کپی تهیه می‌شود) و میانگین سه نمره، نمره نهایی فرد را تعیین می‌کند.

یادآوری ۱. به روش تحلیلی نمره‌گذاری، روش امتیازبندی گفته می‌شود.

یادآوری ۲. در روش کلی، به کل پاسخ تنها یک نمره تعلق می‌گیرد.

یادآوری ۳. هرچه سؤال روشن‌تر باشد، نمره‌گذاری آن نیز ساده‌تر خواهد بود.

یادآوری ۴. هرچه هدف‌های آموزشی روشن‌تر باشند، تهیه سؤال‌های انشایی، ساده‌تر خواهد بود.

۵-۹ محاسن و معایب آزمون‌های انشایی

محاسن.

۱. در سنجش پیامدهای یادگیری پیچیده، کاراترند. آزمون‌های انشایی، در اندازه‌گیری توانایی حل مسئله، تحلیل و ارزیابی اطلاعات یا انجام دیگر مهارت‌های شناختی سطح بالا، ابزارهای سنجشی کاراتری هستند. به عبارت دیگر، بالاترین سطح پیامدهای یادگیری (تحلیل کردن، ارزیابی کردن و آفریدن) را اندازه می‌گیرند.

و تأکید بر اقدام و به‌کار بستن مهارت‌های تفکر و حل مسئله.
 «ارزیابی مستقیم مهارت‌های نوشتن. در بعضی موارد، ارزیابی مهارت‌های نوشتن ویژه ممکن است با سنجش دانش و درک مادهٔ درسی ترکیب شود (برای مثال، ارتباط اصول ریاضی یا علمی، ایده‌ها و مفاهیم). گاهی نیز سنجش مهارت‌های نوشتن ممکن است هدف انحصاری یا اولیه باشد (برای مثال، مهارت در گسترش شخصیت‌ها در یک داستان روایی یا فوت‌وفن نوشتن).

و تهیه آن‌ها نسبتاً آسان است. اگرچه آزمون‌های انشایی، نسبتاً آسان تهیه می‌شوند، ولی سؤال‌ها نباید اتفاقی تهیه شوند. از جدول مشخصات کمک‌بگیرید، تنها موضوعات و هدف‌هایی را انتخاب کنید که به کمک آزمون‌های انشایی، بهتر سنجیده می‌شوند.

۵. توانایی پاسخ دادن به سؤال‌ها را می‌سنجد، نه توانایی انتخاب پاسخ‌ها را.
۶. موقعیت‌های واقعی تری را از آزمون‌های عینی به آزمون‌شوندگان عرضه می‌کنند.
۷. حدس زدن حذف می‌شود. از آنجا که هیچ گزینه‌ای ارائه نشده است، دانش آموز به جای انتخاب گزینه مناسب، پاسخ خود را می‌نویسد و این عامل، حدس را حذف می‌کند.
۸. نقش زیاد و مستقیمی در یادگیری دانش آموز دارند. فرایند آماده سازی پاسخ به یک سؤال انشایی گسترده پاسخ، ممکن است تمرین یادگیری مؤثری باشد.

معیار:

۱. متداول‌ترین محدودیت سؤال‌های انشایی، پایان‌بودن نمره‌گذاری است. طی سال‌ها، مطالعات مختلف نشان داده است که معلمان متفاوت، به یک سؤال نمره متفاوتی داده‌اند. حتی یک معلم خاص هم در دو زمان متفاوت، به یک سؤال، دو نمره متفاوت داده است.
۲. نمونه‌گیری محدود. در این آزمون‌ها به دلیل وقت‌گیر بودن، تعداد سؤال کمی طرح می‌شود و نمونه کوچکی از محتوای درس و هدف‌های آموزش را اندازه می‌گیرند. این نمونه‌گیری محدود، آزمون‌های انشایی را به‌ویژه در اندازه‌گیری دانش اطلاعات واقعی، ناکارآمد می‌کند.
۳. دشواری مرتبط ساختن پاسخ با پیامد. به دلیل آزادی انتخاب، سازمان‌دهی و بیان ایده‌ها و مرتبط ساختن پاسخ‌ها با پیامدهای یادگیری مد نظر دشوار است.

جدول ۵-۴. انواع پیامدهای یادگیری اندازه‌گیری شده به وسیله سؤالات انشایی

نوع سؤال	مثال‌هایی از پیامدهای یادگیری مورد اندازه‌گیری
سؤال انشایی محدود پاسخ	توانایی: - تبیین روابط علت و معلولی - توصیف کاربرد اصول - ارائه بحث‌های مرتبط - بیان فرض‌های قابل دفاع - بیان نتیجه‌گیری‌های معتبر - بیان فرض‌های ضروری - توصیف محدودیت داده‌ها (و پیامدهای مشابه مبتنی بر توانایی شاگرد برای به‌دست‌آوردن پاسخ)
سؤال انشایی گسترده پاسخ	توانایی: - ایجاد، سازمان‌دهی و بیان عقاید - درهم‌آمیختن یادگیری‌ها در حوزه‌های مختلف - ایجاد فرم‌های اصلی (برای مثال، طراحی یک آزمایش) - خلاصه‌کردن (برای مثال، نوشتن خلاصه‌ای از یک داستان) - نوشتن داستان‌های خلاق (برای مثال، انشاهای داستانی) - توضیح مفاهیم یا اصول (برای مثال، انشای تفسیری) - قانع‌کردن یک خواننده (برای مثال، انشای قانع‌کننده) (و پیامدهای مشابه مبتنی بر توانایی شاگرد به‌منظور نوشتن انشایی برای یک هدف معین)

۴. ذهنی‌بودن. مهارت نوشتن و پروردن مطالب، باعث افزایش نمره دانش‌آموز می‌شود و ضعف در نوشتن، غلط‌های املایی و بدخطی، نمره او را کاهش می‌دهد.
۵. لاف‌زدن. بدیهی است که پاسخ‌های طولانی‌تر، صرف‌نظر از محتوا، ممکن است نمره بیشتری به‌دست‌آورند. به همین دلیل ممکن است دانش‌آموز لاف‌بزند.
۶. تصحیح برگه‌های امتحانی این آزمون‌ها بسیار وقت‌گیر است. در صورت زیادبودن تعداد سؤال‌ها و دانش‌آموزان، زمان زیادی برای تصحیح و نمره‌گذاری اوراق لازم است و این امر باعث می‌شود نتوان نمره دانش‌آموزان را به‌سرعت اعلام کرد.

خلاصه فصل پنجم

در این فصل، درباره آزمون‌های انشایی، انواع آن، مزایا و معایب آن‌ها بحث شد. به

پایه‌های مهم یادگیری ممکن است به وسیله سؤال‌های انشایی بازپاسخ یا دیگر آزمون‌های عملکردی، بهتر اندازه‌گیری شوند. این سؤال‌ها به آزمودنی کمک می‌کنند توانایی سازمان‌دهی، یکپارچه‌سازی و ارزیابی عقاید و اطلاعات را به وضوح نشان دهد و دانش و مهارت‌هایش را به کار گیرد.

در کل، آزمون‌های انشایی را با توجه به آزادی عمل آزمون‌شونده در پاسخ دادن و به سؤال‌های آزمون، به دو دسته گسترده پاسخ و محدود پاسخ تقسیم می‌کنند. در آزمون گسترده پاسخ، برای پاسخ آزمون‌شوندگان هیچ محدودیتی وجود ندارد و آن‌ها آزادند طبق میل خود، پاسخ خود را بیروارند و سازمان‌دهی کنند، در حالی که در آزمون‌های انشایی محدود پاسخ، آزمون‌شونده در پاسخ دادن به سؤال‌ها آزادی کامل ندارد، بلکه صورت سؤال، او را ملزم می‌سازد پاسخ خود را در چهارچوب شرایط خاصی محدود کند. استفاده از سؤال‌های محدود پاسخ، متداول‌تر از سؤال‌های گسترده پاسخ است.

بزرگ‌ترین امتیاز آزمون‌های انشایی این است که آزمون‌شوندگان را وادار به بیان اندیشه‌های خود به طور منطقی، منسجم و سازمان‌یافته می‌کنند تا در این کار، خلاقیت خود را نشان دهند. قواعدی به منظور کاستن از معایب این سؤال‌ها پیشنهاد شده است. تصحیح سؤال‌های انشایی، از رأی و نظر مصحح تأثیر می‌پذیرد. بدین منظور برای تصحیح برگه‌های امتحانی، دو روش کلی وجود دارد که عبارت‌اند از: تحلیلی و کلی. در روش تحلیلی نمره‌گذاری که به آن روش امتیازبندی نیز گفته می‌شود، پاسخ نمونه، به اجزاء کوچک‌تری تقسیم می‌شود و برای هر جزء مشخص، نمره یا امتیاز جداگانه‌ای در نظر گرفته می‌شود. در روش کلی، به کل پاسخ تنها یک نمره تعلق می‌گیرد. در این روش، مصحح پاسخ آزمودنی را می‌خواند و یک برداشت کلی از آن به دست می‌آورد و به این برداشت، یک نمره اختصاص می‌دهد.

رعایت نکات زیر، به کیفیت نمره‌گذاری سؤال‌های انشایی کمک خواهد کرد:

۱. سؤال‌های انشایی خوب بنویسید.
۲. قبل از تصحیح پاسخ‌ها، یک پاسخ نمونه تهیه کنید.
۳. سربرگ سؤال‌ها را جدا کنید یا به نام دانش آموز نگاه نکنید.
۴. از یک قاعده نمره‌گذاری که مناسب‌تر است، استفاده کنید.
۵. تأثیر عوامل نامربوط بر نمره‌گذاری را حذف کنید.

۶. تمام پاسخ های آزمون شوندگان به یک سؤال را تصحیح کنید، سپس به سؤال بعدی بروید.
۷. اگر تصمیم مهم است، باید افراد مستقل، دوباره برگه ها را نمره گذاری کنند. در ادامه، نکاتی در مورد محاسن و معایب آزمون های انشایی گفته شد.

خود آزمایی چهارگزینه ای فصل پنجم

۱. برای سنجش بیان علل یک رویداد تاریخی و اطلاعات واقعی، استفاده از کدام آزمون ها بهتر است؟
 - الف) انشایی
 - ب) چندگزینه ای
 - ج) تکمیلی
 - د) محدود پاسخ
۲. برای اندازه گیری توانایی دانش آموز، استفاده از کدام آزمون بهتر است؟
 - الف) چندگزینه ای
 - ب) تکمیلی
 - ج) انشایی
 - د) صحیح - غلط
۳. آزادی عمل در پاسخگویی به کدام سؤال ها بیشتر است؟
 - الف) چندگزینه ای
 - ب) تکمیلی
 - ج) انشایی
 - د) صحیح - غلط
۴. برای اندازه گیری پیامدهای یادگیری بهتر است از آزمون های استفاده شود.
 - الف) دوگزینه ای
 - ب) تستی
 - ج) گسترده پاسخ
 - د) محدود پاسخ
۵. این سؤال، مربوط به کدام نوع از آزمون هاست؟

«به طور خلاصه توضیح دهید که چرا فشارسنج، بهترین ابزار برای پیش بینی وضع هواست؟»

 - الف) گسترده پاسخ
 - ب) محدود پاسخ
 - ج) تستی
 - د) اصلاحی
۶. در مثال «با مشاهده ابرها، پیش بینی شما در مورد طوفان های آینده چیست؟» از کدام پیامد اندازه گیری آزمون های محدود پاسخ استفاده می شود.
 - الف) تدوین فرضیه
 - ب) تحلیل رابطه
 - ج) تشخیص نتیجه گیری
 - د) مقابله وضعیت ها

۷. در کدام نوع از آزمون‌ها، آزمودنی اندیشه‌های خود را به‌طور منطقی، منسجم و سازمان‌یافته بیان می‌کند؟

- (الف) اصلاحی
(ب) انشایی
(ج) تستی
(د) تکمیلی

۸. کدام مورد، جزء قواعد تهیه سؤال‌های انشایی نیست؟

- (الف) سؤال‌های گسترده پاسخ بدون محدودیت، خیلی مناسب‌اند.
(ب) در مورد تهیه سؤال برای موضوعات و مستند کردن عقاید لازم نیست.
(ج) سؤال نباید پاسخ طولانی داشته باشد.
(د) برای سؤال‌های ساده، همانند سایر سؤال‌ها زمان یکسان در نظر گرفته شود.

۹. اگر در یک آزمون، حق انتخاب چند سؤال از میان تعدادی سؤال انشایی به دانش‌آموزان داده شود، این کار معمولاً به‌ضرر کدام گروه از دانش‌آموزان خواهد بود؟

- (الف) ضعیف
(ب) متوسط
(ج) قوی
(د) متوسط به پایین

۱۰. در روش نمره‌گذاری تحلیل آزمون‌های انشایی، کدام گزینه امتیاز دریافت نمی‌کند؟

- (الف) قدرت بیان
(ب) وضعیت ظاهری آزمودنی
(ج) ساختمان منطقی پاسخ
(د) بیان دلایل

۱۱. در کدام روش نمره‌گذاری، نظر مصحح بیشترین تأثیر را دارد؟

- (الف) تستی
(ب) اصلاحی
(ج) تکمیلی
(د) انشایی

۱۲. کدام روش، باعث کاهش سوگیری در نمره‌گذاری می‌شود؟

- (الف) اصلاح برگه‌های هر فرد پس از فرد دیگر
(ب) به هم زدن برگه‌ها قبل از تصحیح
(ج) نگاه کردن به مشخصات برگه سؤال
(د) اصلاح سؤال به سؤال برگه‌ها و به خاطر سپردن نام دانش‌آموز

۱۳. در روش کلی نمره‌گذاری سؤال‌های انشایی، بر کدام مورد تأکید می‌شود؟

- (الف) درک دانش‌آموز
(ب) محتوای کلی
(ج) تعداد پاسخ‌ها
(د) کمیت پاسخ

۱۴. وقتی در سؤال‌های انشایی، سه مصحح برگه‌ای را تصحیح می‌کنند، نمره نهایی فرد برابر است با

- الف) نمره مصحح اول
ب) نمره مصحح دوم
ج) نمره مصحح سوم
د) میانگین نمره سه مصحح

۱۵. کدام گزینه، جزء معایب آزمون‌های انشایی است؟

- الف) سنجش توانایی آزمودنی
ب) ذهنی بودن نمره‌گذاری
ج) ایجاد موقعیت واقعی‌تر
د) سنجش توانایی آزمودنی

۱۶. کدام نوع از آزمون‌های زیر، نمونه‌های کمتری از هدف‌های آموزشی کتاب را مورد سنجش قرار می‌دهد؟

- الف) کارپوشه
ب) چهارگزینه‌ای
ج) انشایی
د) کامل‌کردنی

۱۷. وقتی دانش‌آموز در پاسخ به سؤال‌ها، صرف‌نظر از محتوای سؤال، پاسخ‌های طولانی‌تری می‌نویسد، از کدام روش استفاده کرده‌است؟

- الف) اثر هاله‌ای
ب) لاف‌زدن
ج) نمونه‌گیری گسترده
د) حدس‌زدن

۱۸. کدام گزینه، جزء محاسن آزمون‌های انشایی است؟

- الف) لاف‌زدن آزمودنی
ب) وقت‌گیر بودن آزمون
ج) نمونه‌گیری محدود از محتوا
د) سنجش توانایی پاسخ‌دادن

خودآزمایی تشریحی فصل پنجم

۱. انواع آزمون‌های انشایی را نام ببرید و هرکدام از آن‌ها را با بیان مثال شرح دهید؟
۲. در چه شرایطی سؤال‌های محدودپاسخ در مقایسه با سؤال‌های گسترده‌پاسخ مناسب‌ترند؟
۳. چهار مورد از پیامدهای یادگیری را که به وسیله آزمون‌های محدودپاسخ اندازه‌گیری می‌شوند، با ارائه مثال نام ببرید؟
۴. هدف از تهیه آزمون‌های انشایی چیست؟ معایب و محاسن آن‌ها را بنویسید.
۵. از قواعد تهیه سؤال‌های انشایی، ۴ مورد را نام ببرید؟

نوشتن سؤال‌های انشایی ۱۱۵

۶. چرا در آزمون‌های انشایی، آزمودنی نباید حق انتخاب چند سؤال از میان تعدادی سؤال داشته باشد؟
۷. در روش نمره‌گذاری سؤال‌های انشایی، روش نمره‌گذاری تحلیلی را با بیان مثال شرح دهید؟
۸. در روش نمره‌گذاری سؤال‌های انشایی، روش نمره‌گذاری کلی را با ارائه مثال شرح دهید؟
۹. ایراد وارد بر روش نمره‌گذاری کلی در آزمون‌های انشایی چیست؟
۱۰. کدام یک از روش‌های نمره‌گذاری کلی یا تحلیلی، سریع‌تر است؟
۱۱. رعایت چه نکاتی به کیفیت نمره‌گذاری سؤال‌های انشایی کمک خواهد کرد؟
۱۲. علت تهیه پاسخ نمونه برای نمره‌گذاری آزمون‌های انشایی چیست؟
۱۳. منظور از اثر هاله‌ای در نمره‌گذاری سؤال‌های انشایی چیست؟
۱۴. دو مورد از محاسن تهیه سؤال‌های انشایی را بیان کنید و توضیح دهید؟
۱۵. از معایب سؤال‌های تشریحی، دو مورد را نام ببرید و شرح دهید؟
۱۶. کدام روش نمره‌گذاری سؤال‌های تشریحی، روش امتیازبندی نامیده می‌شود؟

فصل ششم

سنجش کارپوشه

هدف کلی

سنجش فعالیت آموزشی دانش آموزان با استفاده از کارپوشه و تأمین روایی آن

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

1. سنجش کارپوشه را تعریف کند.
2. هدف از تهیه سنجش کارپوشه را بیان کند.
3. روایی کارپوشه را توضیح دهد.
4. مراحل تأمین روایی کارپوشه را بتواند.

مقدمه

سنجش کارپوشه، یکی از روش‌های ارزشیابی توصیفی است و جزء روش‌های سنجش عملکردی به‌شمار می‌آید. در میان آزمون‌های عملکردی، این شیوه متداول‌تر است. یک شیوه سنجش جدید است که روند پیشرفت یادگیرنده را نشان می‌دهد. در این فصل، نخست سنجش کارپوشه معرفی می‌شود. سپس طرح‌های معنادار قابل توجه مانند تصمیم‌گیری درباره هدف کارپوشه، پیامدهای شناختی¹ که باید سنجیده‌شود، اینکه چه کسی آن را طراحی خواهد کرد، معیارهایی برای سنجش پیامدها،

داده‌هایی که باید جمع‌آوری شود تا پیشرفت و تلاش را مستند سازد و اینکه بازخورد مشارکتی^۱ چگونه داده می‌شود، بررسی می‌شوند.

۶-۱ ارزشیابی توصیفی

ارزشیابی توصیفی، شکلی از ارزشیابی تحصیلی- تربیتی است که در آن، معلم با مشارکت فعال دانش‌آموز و اولیای ایشان، با استفاده از ابزارهای مختلف، به جمع‌آوری اطلاعات در زمینه تلاش‌ها، پیشرفت‌ها و موفقیت‌های دانش‌آموزان می‌پردازد و با طبقه‌بندی، تحلیل و تفسیر اطلاعات به آن‌ها کمک می‌کند بهتر یادگیرند و مشکلات یادگیری خود را به کمک اولیا و معلمان برطرف کنند.

«ارزشیابی توصیفی»، الگوی جدیدی است که تلاش می‌کند زمینه‌ای فراهم سازد که دانش‌آموزان در کلاس درس، با شادابی و نشاط بهتر و بیشتر و عمیق‌تر، مطالب درسی را یادگیرند؛ بنابراین به جای توجه افراطی به آزمون‌های پایانی و نمره، روند یاددهی و یادگیری را در طول سال تحصیلی، مورد توجه قرار می‌دهد و در راستای این توجه، رشد عاطفی و اجتماعی و حتی جسمانی دانش‌آموز را نیز در نظر می‌گیرد. این طرح، با هدف بهبود کیفیت یادگیری و یاددهی و افزایش بهداشت روانی محیط یاددهی و یادگیری، تغییرات اساسی را در نظام ارزشیابی فعلی پیشنهاد کرده است. از جمله تغییر مقیاس فاصله‌ای (۰ تا ۲۰) به مقیاس ترتیبی، تغییر ساختار کارنامه و تبدیل آن به کارنامه توصیفی با عنوان گزارش پیشرفت تحصیلی، تغییر و تنوع بخشی در ابزارها و روش‌های جمع‌آوری اطلاعات از وضعیت تحصیلی دانش‌آموز مانند کارپوشه، آزمون‌های عملکردی، فهرست واری و برگ ثبت مشاهدات، تغییر در رویکرد کلی ارزشیابی از ارزشیابی پایانی به ارزشیابی تکوینی و فرایندی، تغییر در مرجع تصمیم‌گیرنده در ارتقای دانش‌آموزان.

۶-۲ سنجش کارپوشه

سنجش کارپوشه^۲، بر این ایده تأکید دارد که مجموعه‌ای از کارهای یادگیرنده در طول سال تحصیلی، یکی از بهترین شیوه‌ها برای نشان دادن پیشرفت دانش‌آموز و تلاش و برای رسیدن به این پیشرفت است.

کارپوشه کلاسی، چنین هدفی را دنبال می‌کند. آن‌ها بهترین نوشته، کار هنری، پروژه‌های علمی، تفکر تاریخی یا پیشرفت ریاضی یادگیرنده را به‌او نشان می‌دهند (کوئن، ۲۰۰۹). همچنین گام‌هایی را نیز که یادگیرنده برداشته‌است تا به‌اینجا برسد، نشان می‌دهند. بهترین کارهای یادگیرنده را گردآوری می‌کنند، اما کارهای حین پیشرفت را نیز شامل می‌شوند: پیش‌نویس‌های اولیه، آزمون‌های اولیه، مطالعات مقدماتی یا کوشش‌های مقدماتی؛ بنابراین آن‌ها شیوه‌ای ایدئال برای سنجش تسلط نهایی، تلاش، بازتاب و رشد در یادگیری‌اند که «داستان» پیشرفت یادگیرنده را می‌گویند (لارکین^۱ و همکاران، ۲۰۰۲).
 ایده سنجش کارپوشه کلاسی^۲، حمایت قابل توجهی کسب کرده و شتاب سریعی داشته‌است. بسیاری از مناطق آموزشی، از کارپوشه و دیگر انواع نمایشگاه‌ها بهره‌می‌گیرند تا تلاش دانش‌آموز را برانگیزند و رشد و پیشرفت در یادگیری را نشان‌دهند. هرچند اعتبار (پایایی) و روایی قضاوت‌های کلاسی معلم، همیشه جای نگرانی بوده‌است، اما زمانی که معلم، فرصت‌های متعددی برای تعامل با یادگیرنده، مشاهده کارش و تصحیح قضاوت‌های خود درباره توانایی دانش‌آموزان دارد، این مشکل کم‌تر می‌شود (میلر، لین و گراندلند، ۲۰۰۹).

۳-۶ منطق کارپوشه

می‌توان گفت بزرگ‌ترین کارکرد کارپوشه این است که در مقایسه با آزمون‌های مداد-کاغذی، مجموعه غنی‌تری از آنچه دانش‌آموز می‌داند و می‌تواند انجام‌دهد، به‌معلم، والدین و یادگیرندگان نشان می‌دهد. اگر کارپوشه به‌شکل مناسب طراحی شود، می‌تواند توانایی تفکر^۳، حل مسئله^۴، به‌کارگیری راهبردها و مهارت‌های روندی^۵ و ساخت دانش به‌وسیله یادگیرنده را نشان‌دهد. افزون‌براین، آن‌ها همچنین در رابطه با مقاومت، تلاش، میل به تغییر، مهارت بازیابی و نظارت بر یادگیری و توانایی خودبازخوردی^۶، اطلاعاتی در اختیار یادگیرنده قرار می‌دهند. یک هدف کارپوشه، دادن اطلاعاتی درباره یادگیرنده به‌معلم است که با روش‌های دیگر میسر نیست (میلر، لین و گراندلند، ۲۰۰۹).

1. Larkin

2. Classroom Portfolio

دلایل دیگری هم برای استفاده از کارپوشه وجود دارد. کارپوشه‌ها همچنین وسیله‌ای برای برقراری ارتباط با والدین و دیگر معلمان‌اند تا سطح پیشرفتی که یادگیرنده به آن رسیده‌است، نشان داده‌شود. کارنامه یادگیرنده، اطلاعات محدودی در اختیار ما می‌گذارد، اما کارپوشه‌ها با نشان دادن شواهد حمایتی به والدین، معلم و یادگیرنده، این نقص را جبران می‌کنند (میلر، لین و گراندلند، ۲۰۰۹).

کارپوشه‌ها جایگزینی برای آزمون‌های مداد-کاغذی، آزمون‌های انشایی یا آزمون‌های عملکردی نیستند. هرکدام از این ابزارها برای هدفی طراحی شده‌اند که به وسیله ابزارهای دیگر برآورده نمی‌شود. اگر بخواهید دانش پایه واقعی یادگیرنده را بسنجید، آزمون‌های عینی مناسب‌اند. کارپوشه، برای نشان دادن میزان مهارت یادگیرنده در به کارگیری یک مهارت شناختی، مناسب نیست، ولی برای سنجش پیشرفت و رشد در یک بافت واقعی، کارپوشه ابزاری است که باید مد نظر قرار گیرد.

سرانجام کارپوشه‌ها راهی برای افزایش انگیزه تلاش در یادگیرنده با هدف رسیدن به سطوح بالاتر هستند. آن‌ها ارتباط مستقیمی بین آموزش و سنجش کلاسی، به شیوه‌ای هماهنگ با نظریه‌های شناختی یادگیری و آموزش مدرن، تدارک می‌بیتند.

۳-۶ تأمین روایی کارپوشه

یکی از هدف‌های مد نظر شما برای کارپوشه این است که یادگیرنده تا چه اندازه می‌تواند به خوبی با مخاطبان متنوع ارتباط برقرار کند. با وجود این، در یک نشریه ادبی، فقط باید نمونه‌های نوشتاری را که می‌خواهید نشریه ادبی پذیرش کند، جمع‌آوری کنید و در کارپوشه ریاضی، تنها رشد توانایی حل مسئله را بگنجانید. با این حال، معیار ارزیابی شما تأکید بسپا زیادی روی راه حل نهایی خواهد داشت. این‌ها بعضی از دشواری‌های تضعیف‌کننده روایی کارپوشه‌اند. درکل، سه چالش فراوی روایی کارپوشه قرار دارد که باید موردتوجه قرار گیرند: بیانگری^۱، راهنمایی^۲ و ارتباط^۳ (کویزین و بوریچ، ۲۰۱۳).

بیانگری. بهترین راه اطمینان از بیانگری این است که در ابتدای کار، روشن‌ترین مهارت‌های شناختی یادگیری و ترتیبی است را که می‌خواهید بسنجید و در صوا

لرزم، گستره‌ای از بازده‌هایی که این را انعکاس می‌دهند، گنجانده شوند. شما نیازمند نمونه‌هایی از نوشتن، تفکر علمی، حل مسائل ریاضی، مهارت‌های روندی یا موقعیتی هستید که می‌خواهید کارپوشه اندازه بگیرد.

راهنمایی. درباره نوشتن راهنمایی، از قبل مطالعاتی داشته‌اید (نوشتن راهنمایی سؤال‌های عینی و انشایی). همان ملاحظات، به منظور طراحی معیارهای روشن با هدف سنجش عملکردهای پیچیده یا توضیحات برای سنجش کارپوشه‌ها نیز به کار گرفته می‌شود. شما به معیارهایی برای سنجش ثبت فردی و کارپوشه به عنوان یک کل نیاز دارید و می‌توانید این امر را با توسعه یک سیستم نمره‌گذاری روشن و دقیق که راهنمایی نامیده می‌شود، تحقق بخشید. با داشتن راهنمایی دقیق، می‌توانید یک سیستم نمره‌گذاری به دست آورید که قضاوت‌های مطلوب نظر شما را به حداقل می‌رساند و معیارهای بالای یادگیرندگان برای پیشرفت را مد نظر قرار دهد.

ارتباط. تهیه کارپوشه نباید نیازمند توانایی‌هایی از یادگیرنده باشد که با موضوع مورد سنجش مرتبط نباشد. کارپوشه جغرافیای یک کلاس پنجمی که هدفش نشان دادن توانایی نقشه‌کشی یادگیرنده است، نباید مهارت‌های ظریف موتور را بطلبد که انتظار می‌رود یک دانش‌آموز دبیرستانی از عهده آن برآید. ابزارهای اندازه‌گیری اغلب در اندازه‌گیری چیزهای مورد انتظار از آن‌ها ناتوان‌اند (یعنی فقدان روایی)، زیرا آن‌ها مهارتی از یادگیرنده می‌خواهند که ربطی به آموخته‌های یادگیرنده ندارد.

یادآوری ۱. کارپوشه، ابزاری است که برای سنجش پیشرفت و رشد دانش‌آموز در یک بافت واقعی به کار برده می‌شود.

یادآوری ۲. سه چالش فراروی روایی کارپوشه عبارت‌اند از: بیانگری، راهنمایی و ارتباط.

۴-۶ گام‌های ساخت کارپوشه روا

گام‌هایی که برای ساخت یک کارپوشه روا لازم است، عبارت‌اند از:

۱. تصمیم‌گیری در مورد هدف‌های کارپوشه. آیا یادگیرندگان شما درباره هدفشان از جمع‌آوری کارپوشه فکرمی‌کنند. یادگیرندگان، خود تشخیص می‌دهند هدف

کارپوشه، شیوه‌ای برای افزایش اعتبار کار است. پیشنهاد می‌شود این را به‌عنوان بخشی از راهبرد تدریس خود، به‌کارگیرید. با وجود این، هدف‌های یادگیرنده شما برای کارپوشه (برای مثال، به‌دست‌آوردن کار در یک شرکت محلی) لزوماً با هدف‌های شما مطابقت ندارد (ارزیابی تدریس شما). در این بخش، درباره روشن‌سازی هدف‌های طرح کارپوشه بحث می‌شود:

- واریسی پیشرفت دانش‌آموز
- انتقال آموخته‌های دانش‌آموز به والدین
- رساندن اطلاعات به معلمان بعدی
- ارزیابی چگونگی آموختن چیزی
- برجسته کردن آنچه تحقق یافته است
- تعیین پایه تحصیلی

۲. شناسایی مهارت‌های شناختی و گرایش‌ها. کارپوشه، همچون سنجش‌های عملکردی،

درک عمیق و پیشرفت واقعی را اندازه‌گیری می‌گیرد. آن‌ها می‌توانند رشد و پیشرفت شایستگی را در حوزه‌هایی مانند ساخت دانش^۱ (برای مثال، سازمان‌دهی دانش)، راهبردهای شناختی^۲ (تحلیل، تفسیر، طراحی، سازمان‌دهی و تجدیدنظر)، مهارت‌های روندی (برای مثال، ارتباط روشن، ویرایش، ترسیم، صحبت کردن و ساختن)، فراشناخت^۳ (برای مثال، خودنظارتی و خودبازخوردی) و همچنین گرایش‌های معین (یا عادات ذهنی) نظیر انعطاف‌پذیری^۴، انطباق‌پذیری^۵، انتقادپذیری^۶، مقاومت^۷، مشارکت و میل به تسلط، اندازه‌گیری کنند (لین، میلر و گراندلند، ۲۰۰۹).

۳. تصمیم‌گیری درباره اینکه چه کسی کارپوشه را تهیه می‌کند. هنگامی که تصمیم

می‌گیرید چه کسی کارپوشه را تهیه خواهد کرد، به‌طور مثال مشخص کنید برای مسابقه ژیمناستیک یا اسکی، چه کسی درگیر است، والدین برای آموزش فرزندان مربی می‌گیرند. مربی، شاگرد و والدین با هم، حرکات، لباس‌ها، زمان‌های تمرین و مانند آن را تدارک می‌بینند. آن‌ها تیمی هستند که تنها هدفشان نمایش بهترین

عملکرد ممکن است. آیا ژیمناستیک کار یا اسکی باز می‌خواهد بهترین عملکردش را به نمایش بگذارد یا اینکه فقط می‌خواهد والدین و مربی‌اش لذت ببرند و انتظارات آنها را برآورده سازد؟ آنچه مهم است این است که جو بازی با هیجان، از خودگذشتگی و تعهد به تلاش واقعی همه اعضای گروه، نیرومی‌گیرد.

هنگام استفاده از کارپوشه، فضایی که شما می‌خواهید ایجاد کنید، این است که تکلیف و وظیفه هرکسی کاملاً مشخص باشد. شما، یادگیرنده و والدین، تیمی هستید که به دانش‌آموز کمک می‌کنید نوشتن، استدلال ریاضی یا تفکر علمی را بهبود بخشید و مثال‌هایی از رشد این شایستگی جمع‌آوری کنید. یادگیرندگان می‌خواهند توانایی خود را در انجام کارها نشان دهند و اعتماد و اطمینانی را که شما و خانواده‌شان به آنها دارند، اثبات کنند. این کارپوشه، اجرا، دوره مسابقه یا رقابت آنهاست.

۴. **تصمیم‌گیری در مورد نوع بازده‌های موجود در کارپوشه و اینکه از هر بازده، چند نمونه باشد.** دو تصمیم کلیدی وجود دارد که باید مد نظر قرار گیرند: مالکیت^۱ و ارتباط کارپوشه با آموزش^۲ شما. مالکیت به درک یادگیرندگان شما اشاره دارد و اینکه کارپوشه می‌تواند شامل آن چیزی باشد که مایل‌اند در کارپوشه باشد (این را در بند ۳ ملاحظه کردید). با درگیر کردن یادگیرندگان و والدین آنها در فرایند برنامه‌ریزی، حس مالکیت آنها را بالامی‌برید. همچنین با گنجاندن کارهای مد نظر آنها در کارپوشه، این حس را انتقال می‌دهید.

یادگیرندگان و والدین آنها باید بدانند که آموزش کلاس شما، بر تدریس مهارت‌های ضروری برای شکل دادن به محتوای کارپوشه متمرکز است. لازم نیست بازده‌هایی در ریاضی، علوم یا مطالعات اجتماعی که یادگیرندگان را برای آن آماده نکرده‌اید، در برنامه بگنجانید. اگر بازده کارپوشه نوشتن است، پس هدف‌های آموزشی شما باید تدریس مهارت‌های نوشتن اشعار، انشا، ویراستاری یا هر چیزی را دربرگیرد که برنامه شما ایجاد می‌کند. برای کارپوشه علوم، ریاضی، جغرافیا یا تاریخ نیز باید به همین روش عمل شود.

بهترین شیوه برای ارضای حس مالکیت یادگیرنده و نیاز معلم برای اندازه‌گیری آنچه آموزش می‌دهد، این است که معلم، حیطه‌های هماهنگ با هدف‌های آموزشی خود را مشخص کند تا والدین و دانش‌آموزان بتوانند از بین آنها نمونه‌هایی را در هر حیطه انتخاب کنند.

فرم درجه‌بندی کارپوشه دانش	
پیش نویس اولیه	
پیش نویس دوم	
پیش نویس نهایی	
پروپوزال دانش آموز تکمیل شود	
۱. تاریخ نوشتن	
۲. به طور خلاصه تئوری کلیه این اشیا درباره شما چه چیزی می‌گوید؟	
۳. در مورد این بخش از نوشتن می‌خواهید چه چیزی بهترین باشد؟	
۴. در پیش نویس بعدی می‌خواهید چه چیزی را اصلاح کنید؟	
۵. اگر این پیش نویس نهایی شماست، این را در کارپوشه دانش فراموشی و چیزی؟	
پروپوزال معلم تکمیل شود	
تجلیت الزامی	
توصیف	
درجه‌بندی	
۵	خیلی واضح بیان کنید درباره اشیا دانش آموز چه چیزی را بیشتر و کمتر دوست دارید. جزئیات بیشتری درباره چگونگی اصلاح کار بررسی کنید.
۴	خیلی واضح بیان کنید درباره اشیا دانش آموز، چه چیزی را دوست دارید و چه چیزی را دوست ندارید. جزئیات بیشتری درباره چگونگی اصلاح کار ارائه دهید.
۳	دوست داشتن ها و دوست نداشتن هایش را بیان کنید، اما با جزئیات بیشتر.
۲	با چیز مهمی درباره دوست داشتن ها و دوست نداشتن ها وجود دارد. جزئیات کمی درباره چگونگی اصلاح اشیا، ارائه دهید.
۱	مدرکی از هیچ بازتابی در کار نیست.
۱. آداب نوشتن	
توصیف	
درجه‌بندی	
۵	استفاده از آداب نوشتن، خیلی چشمگیر است. هیچ خطای آشکاری وجود ندارد. این آداب، سیال و پیچیده‌اند. نقطه گذاری، نحوه کاربرد گر امر، ساختار جمله، همجی.
۴	استفاده از آداب نوشتن، چشمگیر است. خطای کمی وجود دارد.
۳	این آداب، تقریباً همه چشمگیرند. نقطه گذاری، نحوه کاربرد گر امر، ساختار جمله، همجی.
۲	استفاده از آداب نوشتن، تا حدی چشمگیر است. خطاها در معنا ملاحظه می‌کنند. این آداب، تا حدی چشمگیرند. نقطه گذاری، نحوه کاربرد گر امر، ساختار جمله، همجی.
۱	خطاهای استفاده از آداب نوشتن، در معنا ملاحظه می‌کنند. این خطاها محدود و متفاوت‌اند. نقطه گذاری، نحوه کاربرد گر امر، ساختار جمله، همجی.
۱	خطاهای صمیمی در استفاده از آداب نوشتن وجود دارد که معنا را سهم می‌سازد. فاقد درک نقطه گذاری، نحوه کاربرد گر امر، ساختار جمله، همجی.

ردیف	توضیح
۱	در صورتی که...
۲	در صورتی که...
۳	در صورتی که...
۴	در صورتی که...
۵	در صورتی که...
۶	در صورتی که...
۷	در صورتی که...
۸	در صورتی که...
۹	در صورتی که...
۱۰	در صورتی که...
۱۱	در صورتی که...
۱۲	در صورتی که...
۱۳	در صورتی که...
۱۴	در صورتی که...
۱۵	در صورتی که...
۱۶	در صورتی که...
۱۷	در صورتی که...
۱۸	در صورتی که...
۱۹	در صورتی که...
۲۰	در صورتی که...
۲۱	در صورتی که...
۲۲	در صورتی که...
۲۳	در صورتی که...
۲۴	در صورتی که...
۲۵	در صورتی که...
۲۶	در صورتی که...
۲۷	در صورتی که...
۲۸	در صورتی که...
۲۹	در صورتی که...
۳۰	در صورتی که...
۳۱	در صورتی که...
۳۲	در صورتی که...
۳۳	در صورتی که...
۳۴	در صورتی که...
۳۵	در صورتی که...
۳۶	در صورتی که...
۳۷	در صورتی که...
۳۸	در صورتی که...
۳۹	در صورتی که...
۴۰	در صورتی که...
۴۱	در صورتی که...
۴۲	در صورتی که...
۴۳	در صورتی که...
۴۴	در صورتی که...
۴۵	در صورتی که...
۴۶	در صورتی که...
۴۷	در صورتی که...
۴۸	در صورتی که...
۴۹	در صورتی که...
۵۰	در صورتی که...
۵۱	در صورتی که...
۵۲	در صورتی که...
۵۳	در صورتی که...
۵۴	در صورتی که...
۵۵	در صورتی که...
۵۶	در صورتی که...
۵۷	در صورتی که...
۵۸	در صورتی که...
۵۹	در صورتی که...
۶۰	در صورتی که...
۶۱	در صورتی که...
۶۲	در صورتی که...
۶۳	در صورتی که...
۶۴	در صورتی که...
۶۵	در صورتی که...
۶۶	در صورتی که...
۶۷	در صورتی که...
۶۸	در صورتی که...
۶۹	در صورتی که...
۷۰	در صورتی که...
۷۱	در صورتی که...
۷۲	در صورتی که...
۷۳	در صورتی که...
۷۴	در صورتی که...
۷۵	در صورتی که...
۷۶	در صورتی که...
۷۷	در صورتی که...
۷۸	در صورتی که...
۷۹	در صورتی که...
۸۰	در صورتی که...
۸۱	در صورتی که...
۸۲	در صورتی که...
۸۳	در صورتی که...
۸۴	در صورتی که...
۸۵	در صورتی که...
۸۶	در صورتی که...
۸۷	در صورتی که...
۸۸	در صورتی که...
۸۹	در صورتی که...
۹۰	در صورتی که...
۹۱	در صورتی که...
۹۲	در صورتی که...
۹۳	در صورتی که...
۹۴	در صورتی که...
۹۵	در صورتی که...
۹۶	در صورتی که...
۹۷	در صورتی که...
۹۸	در صورتی که...
۹۹	در صورتی که...
۱۰۰	در صورتی که...

شکل ۱-۹. فرم درجه بندی کاربرده نشا

۵. پی‌ریزی راهنمایی‌های کارپوشه. در دومین گام، مهارت‌های شناختی عمده و گرایش‌هایی مورد نظر خود برای اندازه‌گیری با کارپوشه را مشخص کنید. در چهارمین گام، شما حیطه‌های محتوایی را که در کارپوشه گنجانده می‌شود، مشخص کردید. حالا باید عملکرد مناسب را برای وارد کردن در کارپوشه انتخاب کنید تا کارپوشه به‌عنوان یک کل، خوب، متوسط و ضعیف به‌نظر برسد. برای هر بازده شناختی یادگیری و هر حیطه از محتوای کارپوشه، صفات یا ویژگی‌های اولیه مهم را از نظر خود لیست کنید. سپس یک مقیاس درجه‌بندی برای توصیف محدوده عملکرد دانش‌آموز، درست کنید.

فرم درجه‌بندی تراکمی انشا

انشای کامل‌شده را ضمیمه کنید

نمونه انشای ۱		نمونه انشای ۲		پیش‌نویس ۱		پیش‌نویس ۲		پیش‌نویس ۳	
نمره	معیار	نمره	معیار	نمره	معیار	نمره	معیار	نمره	معیار
۳	تأمل	۴	تأمل	۳	تأمل	۴	تأمل	۳	تأمل
۴	آداب	۴	آداب	۳	آداب	۴	آداب	۳	آداب
۵	سازمان	۵	سازمان	۴	سازمان	۵	سازمان	۴	سازمان
	طرح		طرح		طرح		طرح		طرح
	۳/۷۵ میانگین		۲/۲۵		۳/۵		۳/۵		۳/۵
معلم: اظهارنظر در مورد رشد انشای نهایی:									
دانش‌آموز: اظهارنظر در مورد رشد انشای نهایی:									
والدین: اظهارنظر در مورد رشد انشای نهایی:									
در کارپوشه، قرار داده‌شود: پله.....									
خیر.....									

شکل ۶-۲. فرم درجه‌بندی تراکمی کارپوشه انشا

۵-۶ مزایای کارپوشه

- از آنجا که کارپوشه، شامل بازده‌هایی می‌شود که حاصل آموزش کلاسی است، به‌راحتی قابل ادغام در آموزش کلاسی خواهد بود.

- کارپوشه، به دانش آموزان فرصت‌هایی می‌دهد تا به آن‌ها نشان دهد چه کارهایی می‌توانند انجام دهند.
- کارپوشه، دانش آموزان را تشویق می‌کند تا یادگیرنده‌ای تأملی باشند و مهارت‌های ارزیابی نقاط قوت و ضعف کارشان را رشد دهند.
- کارپوشه، به دانش آموزان کمک می‌کند برای تعیین هدف و ارزیابی پیشرفت خود، مسئولیت‌پذیر باشند.
- کارپوشه، برای معلمان و دانش آموزان فرصت‌هایی فراهم می‌سازد تا در پیشرفت دانش آموز سهیم باشند و آن پیشرفت را به‌ویژه انعکاس دهند.
- کارپوشه به‌واسطه نشان دادن مثال‌های ملموس کار دانش آموز و روشن ساختن پیشرفت او، شیوه‌ای مؤثر برای برقراری ارتباط با والدین است.
- کارپوشه می‌تواند مثال‌های ملموسی از پیشرفت دانش آموز در طول زمان و مهارت‌های فعلی او به والدین نشان دهد.

خلاصه فصل ششم

سنجش کارپوشه، جزء روش‌های سنجش عملکردی به‌شمار می‌آید و مبتنی بر این ایده است که مجموعه‌ای از کارهای یادگیرنده را در طول سال تحصیلی نشان می‌دهد و یکی از بهترین شیوه‌ها برای نشان دادن پیشرفت دانش آموز و تلاش وی برای رسیدن به این پیشرفت است. یک کارپوشه، شیوه معلمان برای نشان دادن این نکته است که واقعاً دانش آموزان، قادر به انجام چه کارهایی هستند. ایده سنجش کارپوشه کلاسی، حمایت قابل توجهی کسب کرده و شتاب سریعی داشته است.

بزرگ‌ترین کارکرد کارپوشه این است که در مقایسه با آزمون‌های مداد-کاغذی، به معلمان، والدین و یادگیرندگان، مجموعه غنی‌تری از سطح دانش آموز و میزان توانایی‌اش را نشان می‌دهد. اگر کارپوشه به‌شکل مناسب طراحی شود، می‌تواند توانایی تفکر، حل مسئله، به‌کارگیری راهبردها و مهارت‌های روندی و ساخت دانش به‌وسیله یادگیرنده را نشان دهد.

سه چالش فراوانی روایی کارپوشه، عبارت‌اند از: بیانگری، راهنمایی و ارتباط. همچنین گام‌هایی که برای ساخت یک کارپوشه روا لازم است، عبارت‌اند از: تصمیم‌گیری در مورد هدف‌های کارپوشه، شناسایی مهارت‌های شناختی و گرایش‌ها، تصمیم‌گیری درباره تهیه‌کننده کارپوشه، تصمیم‌گیری در مورد نوع بازده‌های موجود در

کارپوشه و تعداد نمونه‌های هر بازده، پی‌ریزی راهنمایی‌های کارپوشه و رشد یک شیوه برای جمع کردن تمام درجه‌بندی‌های کارپوشه. بعضی از ویژگی‌هایی که در سازوکار نمره‌گذاری کارپوشه باید مدنظر قرار بگیرند، عبارت‌اند از: نظم و دقت، تنوع، رشد و پیشرفت، کیفیت کلی، خودبازخوردی، انعطاف‌پذیری، سازمان و ظاهر. برای هر مقوله محتوا که در کارپوشه در نظر گرفته‌اید، یادگیرندگان برای هر پیش‌نویس یک نمره و برای محصول نهایی نیز یک نمره خواهند گرفت.

خودآزمایی چهارگزینه‌ای فصل ششم

۱. کدام روش، بهترین شیوه برای نشان دادن پیشرفت دانش‌آموز در طول دوره تحصیلی است؟
 - (الف) آزمون‌های انشایی
 - (ب) آزمون‌های تستی
 - (ج) سنجش کارپوشه
 - (د) آزمون‌های اصلاحی
۲. در کدام یک از روش‌های زیر می‌توان توانایی بازخوردی به دانش‌آموز ارائه داد؟
 - (الف) کارپوشه
 - (ب) آزمون‌های کتبی
 - (ج) آزمون‌های انشایی
 - (د) اعتبار آزمون
۳. کدام گزینه، اطلاعاتی درباره دانش‌آموز در اختیار معلم می‌گذارد که به واسطه سایر روش‌ها امکان‌پذیر نیست؟
 - (الف) آزمون انشایی
 - (ب) کارپوشه
 - (ج) آزمون چندگزینه‌ای
 - (د) آزمون تکمیلی
۴. با استفاده از کدام روش می‌توان اطلاعات حمایتی در مورد دانش‌آموزان به معلمان دانش‌آموزان ارائه داد؟
 - (الف) کارپوشه
 - (ب) آزمون انشایی
 - (ج) آزمون چندگزینه‌ای
 - (د) آزمون تکمیلی
۵. کدام مورد، ارتباط مستقیمی بین دانش‌آموز و سنجش کلاسی به شیوه‌ای هماهنگ نظریه‌های شناختی تدارک می‌بیند؟
 - (الف) آزمون انشایی
 - (ب) کارپوشه
 - (ج) آزمون اصلاحی
 - (د) نحوه تهیه سؤال

۶. کدام روش، جزء تأمین روایی کارپوشه نیست؟
 الف) بیانگری
 ب) راهنمایی
 ج) هنجار
 د) ارتباط
۷. برای سنجش پیشرفت و رشد در یک بافت واقعی، استفاده از کدام روش مناسبتر است؟
 الف) آزمون انشایی
 ب) کارپوشه
 ج) آزمون چندگزینه‌ای
 د) آزمون مداد-کاغذی
۸. اولین گام در تهیه کارپوشه روا چیست؟
 الف) تصمیم‌گیری در مورد تهیه کارپوشه به وسیله چه افرادی
 ب) شناسایی مهارت‌های شناختی
 ج) تصمیم‌گیری در مورد هدف‌های کارپوشه
 د) پی‌ریزی راهنمایی کارپوشه
۹. تحلیل، تفسیر، طراحی و سازمان‌دهی، جزء کدام یک از مهارت‌های شناختی و گرایش‌هاست؟
 الف) راهبردهای شناختی
 ب) فراشناخت
 ج) مهارت‌های روندی
 د) ساخت دانش
۱۰. وزن نهایی پیش‌نویس‌ها و محصول نهایی در کارپوشه، باید برابر با چند باشد؟
 الف) ۰
 ب) ۰/۵۰
 ج) ۱
 د) ۱۰۰

خودآزمایی تشریحی فصل ششم

۱. سنجش کارپوشه، مبتنی بر چه ایده‌ای است؟
۲. منظور از تأمین روایی کارپوشه چیست؟
۳. چالش‌های فراروی کارپوشه را نام ببرید و شرح دهید.
۴. مراحل لازم برای سنجش کارپوشه روا را به ترتیب نام ببرید؟
۵. از مراحل تهیه کارپوشه روا، شناسایی مهارت‌های شناختی و گرایش‌ها را شرح دهید؟
۶. از دلایل روشن‌سازی هدف‌های کارپوشه، به ۴ مورد اشاره کنید؟

۷. در تهیه کارپوشه، مالکیت و ارتباط به چه مواردی اشاره دارند؟
۸. مواردی ضروری در سازوکار نمره‌گذاری کارپوشه را نام ببرید؟
۹. منظور از فرم درجه‌بندی تراکمی حل مسئله ریاضی در تدوین کارپوشه چیست؟
۱۰. پنج حوزه محتوایی را که در نوشتن کارپوشه لازم است، فقط نام ببرید؟
۱۱. منظور از وزن‌دادن در نمره‌گذاری کارپوشه چیست؟

فصل هفتم

هنجار آزمون

هدف کلی

بحث درباره روش‌های تفسیر نمره‌های خام و تبدیل آن‌ها به نمره‌های استاندارد شده

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

1. هدف از تفسیر نمره‌ها و هنجارهای مرتبط را بیان کند.
2. روش تفسیر نمره‌های خام را توضیح دهد.
3. تفسیرهای ملاکی و هنجاری را تعریف کند و شرایط لازم برای تفسیر آن‌ها را شرح دهد.
4. هنجار را تعریف کند و انواع آن را نام ببرد.
5. انواع نمره‌های ترازشده را نام ببرد و کاربرد آن‌ها را برشمارد.
6. نیم‌رخ را تعریف کند و کاربرد آن را توضیح دهد.

مقدمه

در این فصل ابتدا در مورد تفسیر نمره‌ها و هنجارهای آزمون بحث می‌شود. سپس روش‌های تبدیل نمره‌های خام به نمره‌های استاندارد، توضیح داده می‌شود. در ادامه به آزمون‌های ملاکی، هنجاری و انواع هنجارها پرداخته می‌شود. نمره‌های استاندارد و انواع بر کاربرد آن توضیح داده می‌شوند و در آخر، مطالبی درباره نیم‌رخ روانی ارائه خواهد شد.

۷-۱ تفسیر نمره‌ها و هنجارهای آزمون

اگر همانند نمره مقیاس‌هایی که برای اندازه‌گیری فیزیکی (متر و مانند آن برای اندازه‌گیری طول، ترازو و مانند آن برای اندازه‌گیری وزن) به‌کار می‌روند، تبدیل نمره آزمون‌ها امکان‌پذیر بود، تفسیر نمره آزمون‌ها خیلی راحت می‌شد. ۲۵ سانتی‌متر، یک‌چهارم یک متر، ۷۵۰ کیلومتر سه برابر ۲۵۰ کیلومتر و یک اینچ، ۲/۵۴ سانتی‌متر است. یک اونس، معادل ۳۱/۱ گرم و ۴۰ کیلو، دو برابر ۲۰ کیلوست و مانند آن. این قابلیت مقایسه اندازه‌گیری‌ها از یک وضعیت به وضعیت دیگر مانند «یک‌چهارم»، «سه برابر» و همانند آن، به این دلیل است که این مقیاس‌ها «صفر واقعی»^۱ دارند و فاصله بین واحدهای آن‌ها «برابر» است.

نقطه صفر واقعی^۳ (یعنی نقطه‌ای که به هیچ وجه، هیچ وزن و طولی وجود ندارد) به‌طور دقیق نقطه شروع اندازه‌گیری را نشان می‌دهد. واحدهای مساوی (برای مثال، متر، کیلو یا دقیقه)، از یک وضعیت به وضعیت دیگر و از یک بخش از مقیاس به بخش دیگر، معنای یکسانی دارند. ده کیلو در نزد دکتر، بقال، کشاورز و بنا، یک وزن را نشان می‌دهد. تفاوت بین ۲۵ و ۳۵ کیلو، دقیقاً مشابه تفاوت بین ۱۵۰ و ۱۶۰ کیلو است.

ویژگی‌های مقیاس‌های اندازه‌گیری فیزیکی که همه با آن‌ها آشنا هستند، متأسفانه معمولاً در اندازه‌گیری‌های تربیتی وجود ندارد. صفرگرفتن یک نفر در امتحان املا، به معنای ناتوانی مطلق او در نوشتن یک کلمه درست نیست. بسیاری از کلمات هستند که شخص، قادر به نوشتن آن‌هاست، ولی در املا نیامده و به‌طور قراردادی، ۲۰ غلط معادل نمره صفر خواهد بود. اینجا مقیاس، دارای صفر مطلق نیست (مثل مقیاس‌های اندازه‌گیری طول و وزن)، بلکه صفر، مقیاس «قراردادی»^۲ یا «دلبخواهی»^۵ است. این حالت قراردادی یا دلبخواهی بودن، مانع این اظهار نظر می‌شود که دانایی شخص دارای نمره ۲۰، دقیقاً دو برابر کسی است که نمره ۱۰ می‌گیرد، زیرا ممکن است در املا دیگری، نمره‌ها تغییر کنند و شخص دارنده ۲۰، نمره ۱۸ بگیرد و کسی که نمره صفر

گرفته بود، ۸ بگیرد. در این صورت، قاعده به هم می خورد، ولی در مقیاس هایی که صفر واقعی دارند، ۲۰ همیشه دو برابر ۱۰ است و تغییر نمی کند. از آنجا که مکان صفر واقعی آزمون های پیشرفت مشخص نیست، تفسیر نمره آزمون ها به همان شیوه اندازه گیری های فیزیکی، امکان پذیر نیست.

تفسیر نتایج آزمون ها، از یک عامل دیگر نیز تأثیر می پذیرد و آن، نابرابری واحدهای اندازه گیری است. پاسخ دادن به ۲۰ سؤال آسان در یک آزمون ریاضی، معنایی همانند پاسخ دادن به ۲۰ سؤال دشوار در همان درس ندارد و نمره ۱۵ در دروس ریاضی، علوم، فارسی و املا، دارای مفهوم مشابهی نیستند. نمره ۱۵ برای درس ریاضی ممکن است نمره بالایی محسوب شود، برای دروس علوم و فارسی، نمره متوسطی باشد و برای درس املا، نمره پایینی تلقی شود.

به منظور یکسان سازی تفسیر نمره ها در اندازه گیری های تربیتی، روش های متعددی ابداع شده اند؛ بنابراین بهتر است معلمان، با روش های متعدد گزارش نمره آشنا شوند. این کار فقط برای خود معلمان نیست، بلکه به منظور کمک معلمان به والدین در دانستن جایگاه تحصیلی فرزندشان است.

۷-۲ روش های تفسیر نمره های خام آزمون

اگر دانش آموزی به ۳۲ سؤال یک آزمون چهارگزینه ای که اعتبار هر سؤال آن، نیم نمره است، پاسخ درست داده باشد، نمره خام وی ۱۶ خواهد بود؛ بنابراین نمره خام^۱، فقط بیانگر امتیازاتی است که دانش آموز از آزمونی می گیرد که برابر دستورالعمل، نمره گذاری شده است. دادن نیم نمره به سؤال، کسر نمره برای حدس زدن یا به کارگیری هر قاعده دیگری در تعیین نمره مهم نیست، به هر حال نمره به دست آمده، نمره خام نام دارد. به عبارت دیگر نمره خام، تعداد سؤال هایی است که به درستی پاسخ داده شده اند (گولو، ۲۰۰۹).

هرچند نمره خام، خلاصه ای عددی از عملکرد آزمون دانش آموز است، ولی بدون اطلاعات بعدی، خیلی معنادار نیست. اگر ما بگوییم نرگس به ۲۸ سؤال یک آزمون زبان انگلیسی پاسخ صحیح داده است، بنابراین نمره خام وی ۲۸ است. پاسخ ما احتمالاً سؤالی شبیه این به دنبال خواهد داشت: «معنی ۲۸ چیست؟» یا «آیا نمره خوبی

است؟». ممکن است سؤال‌های دقیق‌تری هم مطرح شود: «آزمون، چند سؤال داشت؟»، «آزمون، چقدر دشوار بود؟»، «نمره ۲۸ نرگس در مقایسه با نمره دانش‌آموزان دیگر کلاس، در چه وضعیتی است؟» پاسخ به این سؤال‌ها یا سؤال‌هایی شبیه به این، نیازمند این است که نمره خام، معنادار و قابل فهم شود. برای اینکه این نمره، قابل فهم‌تر شود، لازم است نمره به چیزی تبدیل (برای مثال نمره ۲۸، از ۸۵٪ نمره افراد کلاس، بالاتر است) یا با یک استاندارد مقایسه شود (مثال: در این آزمون، میانگین کشوری ۳۴ است) که افراد کم‌اطلاع به راحتی معنی آن را بفهمند.

۷-۳ تفسیرهای ملاکی و هنجاری

۷-۳-۱ آزمون‌های ملاکی

آزمون ملاکی یا ملاک مرجع، به آزمونی گفته می‌شود که در آن، معیار یا ملاک ارزشیابی، از پیش تعیین و عملکرد یادگیرنده در آزمون، با توجه به آن ملاک سنجش می‌شود. تفسیر آزمون ملاکی به‌ما اجازه تفسیر عملکرد فرد را بدون نیاز به مقایسه عملکرد وی با دیگران می‌دهد. در این‌گونه آزمون‌ها عملکرد فرد، با یک ملاک از پیش تعیین‌شده که همان ملاک عملکرد لحاظ‌شده در هدف‌های آموزشی است، مقایسه می‌شود؛ بنابراین ممکن است عملکرد دانش‌آموز، با توجه به سرعت انجام عملکرد (تایپ ۴۰ کلمه در دقیقه)، میزان دقت عملکرد (اندازه‌گیری طول یک خط با یک‌دهم میلی‌متر خطا)، درصد ماده‌های صحیح تکالیف یادگیری (معنی کردن ۹۰٪ کلمات به‌طور صحیح) و مانند آن سنجیده شود (میلر، لین و گرانلند، ۲۰۰۹). از میان آزمون‌های ملاکی می‌توان به آزمون‌های کلاسی اشاره کرد. در مثال‌های زیر، ملاک، سن است: دریافت گواهینامه رانندگی که ملاکش سن ۱۸ سال است، رفتن به سریازی، سن اولین رأی و ثبت‌نام در پایه اول دبستان.

۷-۳-۲ نمره‌های هنجاری

نمره‌های هنجاری^۱، نمره‌های استاندارد به‌ما اجازه می‌دهند عملکرد یک دانش‌آموز را د

یک آزمون با عملکرد وی در آزمونی دیگر مقایسه کنیم. نمره استاندارد به این صورت به دست می آید که ابتدا از گروه زیادی از افراد که شبیه هم هستند (برای مثال، سن، پایه تحصیلی، توانایی ورزشی و مانند آن)، آزمونی به عمل می آید. سپس میانگین و انحراف استاندارد نمره های این گروه محاسبه می شود. آن گاه مشخص می شود که نمره فرد، پایین تر از میانگین یا بالاتر از آن قرار گرفته است. به عبارت دیگر، اگر برای تفسیر نمره یک دانش آموز، نمره او با نمره یک گروه مرجع مقایسه شود، به آن گروه مرجع، گروه هتجار گویند. گروه مرجع، از کسانی تشکیل می شود که به گونه ای شبیه به دانش آموز مورد نظرند.

به طور خلاصه، هدف عمده در استفاده از نمره های هتجاری، تعیین وضعیت یک دانش آموز در مقایسه با دانش آموز دیگر است. آزمون های هوشی، شخصیتی، کنکورهای سراسری و آزمون های استاندارد پیشرفت تحصیلی هتجاری اند.

نمره های مشتق. هر چند رتبه بندی نمره های خام برای گزارش نتایج یک آزمون کلاسی مفید است، ولی خارج از کلاس، ارزش چندانی ندارد، زیرا رتبه داده شده، به ترکیب گروه و تعداد اعضای گروه بستگی دارد. به منظور به دست آوردن یک چهارچوب کلی تر برای تفسیر هتجاری، نمره های خام، به نوعی نمره مشتق^۱ تبدیل می شوند. نمره های هتجار که به آن ها نمره های مشتق هم گفته می شود، دارای انواع مختلفی هستند که معروف ترین آن ها عبارت اند از: هتجارهای سنی و کلاس معادل^۲، رتبه های درصدی^۳ و نمره های استاندارد. در نمره های رتبه معادل، عملکرد دانش آموز، با عملکرد یک گروه مرجع مقایسه می شود. رتبه های درصدی و نمره های استاندارد، موقعیت نسبی دانش آموز را در گروهی که خود عضوی از آن است، نشان می دهد (جدول ۷-۱). تبدیل نمره های خام به نمره های مشتق، با آزمون های هتجاری ساده است (میلر، لین و گرانلند، ۲۰۰۹). ناشر آزمون هتجاری قبل از گزارش نمره های خام، تبدیل را انجام می دهد.

شیوه های متعددی برای بیان نمره های استاندارد وجود دارد (برای مثال، نمره T، نمره های معادل منحنی نرمال و نمره های سنی هتجار)، اما آن ها مبتنی بر یک منطق مشترک اند. این نمره ها تنها بر حسب ارزش های عددی و اینکه مبتنی بر توزیع نرمال اند یا خیر، متفاوت هستند.

جدول ۷-۱. متداول‌ترین انواع هنجارهای آزمون

نوع هنجار آزمون	نام نمره مشتق	معنا یا توجه به عملکرد آزمون
هنجارهای کلاسی	معادل کلاسی	نمره میانی، یک آزمون است که دانش‌آموزان پایه تحصیلی معینی، نمره‌ای برابر آن گرفته‌اند.
هنجارهای درصدی	رتبه‌های درصدی یا نمره‌های درصدی	درصد دانش‌آموزانی که در گروه مرجع، نمره خامی پایین‌تر از نمره دانش‌آموز گرفته‌اند.
هنجارهای نمره استاندارد	نمره‌های استاندارد	فاصله نمره خام دانش‌آموز بالا یا پایین میانگین گروه مرجع، برحسب واحدهای انحراف استاندارد

۴-۷ هنجارها

جدول‌های هنجاری در دفترچه راهنمای آزمون، فقط نمره‌هایی را ارائه می‌دهند که به وسیله دانش‌آموزان در گروه‌های مرجع به‌دقت تعریف‌شده، کسب شده‌است (میلر، لین و گراندلند، ۲۰۰۹). نمره‌های خام و مشتق، در دو ستون کنار یکدیگر قرار داده شده‌اند تا تبدیل آن‌ها به راحتی صورت گیرد. نمره‌های مشتق، مربوط به گروهی است که آزمون برای آن‌ها هنجار شده‌است. این هنجارها توانایی پاسخ‌دادن به سؤال‌های زیر را به ما می‌دهند:

۱. عملکرد دانش‌آموز، در مقایسه با دیگر دانش‌آموزان چگونه است؟
 ۲. عملکرد آزمودنی در یک آزمون (یا خرده‌آزمون)، در مقایسه با عملکرد وی در آزمون (یا خرده‌آزمون) دیگر چگونه است؟
 ۳. عملکرد یک دانش‌آموز در یک فرم آزمون، در قیاس با عملکرد او در فرم دیگر آزمون که در گذشته اجرا شده، چگونه است؟
- مقایسه‌های نمره‌های آزمون در موارد زیر به ما کمک می‌کند: پیش‌بینی موفقیت احتمالی دانش‌آموز در زمینه‌های مختلف، تشخیص نقاط قوت و ضعف عملکرد وی، اندازه‌گیری رشد آموزشی وی (میلر، لین و گراندلند، ۲۰۰۹).

۱-۴-۷ هنجارهای سنی و کلاسی

هنجارهای سنی و کلاسی، از جمله هنجارهایی هستند که مصرف‌کنندگان آزمون‌ها با آن‌ها آشنایی دارند. هنجار سنی (هنجار معادل سن تقویمی)، نمره میانی یک آزمون است که افراد گروه سنی معینی، نمره‌ای برابر آن گرفته‌اند. هنجارهای سنی، برحسب سال و ماه و با فاصله‌های یک ماه تدوین می‌شوند. برای مثال، هنجار سنی برای کودکان ۵ ساله، به صورت

زیر طبقه‌بندی می‌شود که در آن‌ها اعداد سمت چپ، نماینده سال و اعداد سمت راست بعد از اعشار، نماینده ماه است (شریفی، ۱۳۹۳). هنگام سنی، برای سنجش ویژگی‌هایی به‌کار می‌رود که با افزایش سن، رابطه دارند؛ مانند هوش و بعضی توانایی‌های شناختی دیگر.

۵/۱۱، ۵/۱۰، ۵/۹، ۵/۸، ۵/۷، ۵/۶، ۵/۵، ۵/۴، ۵/۳، ۵/۲، ۵/۱، ۵/۰ سال

با توجه به مطالب قبل، هنگام کلاسی (معادل کلاسی)، نمره میانی یک آزمون است که دانش‌آموزان پایه تحصیلی معینی، نمره‌ای برابر آن گرفته‌اند. هنگام‌های کلاسی، به‌طور گسترده با آزمون‌های پیشرفت تحصیلی استاندارد به‌ویژه در مقطع ابتدایی، به‌کار گرفته شده‌اند. در هنگام‌های کلاسی، از گروه‌های کلاسی به‌عنوان گروه‌های مرجع استفاده می‌کنند. هنگام‌های کلاسی، بیانگر عملکرد متوسط دانش‌آموزان کلاس‌ها هستند. هنگام‌های کلاسی نیز مثل هنگام‌های سنی، برحسب پایه و ماه‌های سال تحصیلی با فاصله یک ماه و به‌صورت زیر طبقه‌بندی می‌شوند که در آن‌ها اعداد سمت چپ، پایه تحصیلی و اعداد سمت راست، ماه‌های سال تحصیلی را نشان می‌دهد. طبقه‌بندی هنگام کلاسی به‌فاصله یک ماه، بر این فرض استوار است که با گذشت ماه‌های سال تحصیلی، به‌تدریج و با آهنگ یکسان، سطح پیشرفت دانش‌آموزان افزایش می‌یابد. هنگام کلاسی، برای سنجش پیشرفت تحصیلی که با افزایش پایه تحصیلی افزایش می‌یابد، به‌کار می‌رود. برای مثال، اگر آزمون استاندارد ریاضی، در ماه‌های مهر و اردیبهشت، از دانش‌آموزان کلاس‌های چهارم و پنجم به‌عمل آید، انتظار می‌رود دانش‌آموزان پنجم در ماه مهر، نمره‌های بالاتر از دانش‌آموزان چهارم در اردیبهشت‌ماه بگیرند.

با توجه به جدول ۷-۲، این انتظار محقق شده است. ۵۰٪ دانش‌آموزان کلاس چهارم در مهرماه، در درس ریاضی نمره خام کمتر از ۵/۷۵ (میانگین) گرفته‌اند، اما این نمره برای کلاس چهارم در اردیبهشت‌ماه، به ۷/۵۵ افزایش یافته است. این نمره برای کلاس پنجم، در مهرماه به ۸ و در اردیبهشت‌ماه در کلاس پنجم، به ۹ رسیده است؛ بنابراین اگر قرار است نمره‌های بیشتر از میانگین نمره دانش‌آموزان کلاس پنجم در اردیبهشت‌ماه بگیرد، باید نمره خام وی، بالاتر از ۹ باشد تا به معادل کلاسی ۵/۸ برسد (معادل کلاسی، به‌صورت دورقمی نشان داده می‌شود. رقم اول، برای سال و رقم بعد از اعشار، برای ماه). پس به نمره خام ۹، معادل کلاسی ۵/۸ داده می‌شود. رقم دوم، معادل کلاسی تا ۹ را شامل می‌شود، زیرا سال تحصیلی، ۹ ماه است.

جدول ۷-۲. نمایش چگونگی ایجاد نمره‌های معادل کلاسی (اقتباس از میلر، لین و گرانلند، ۲۰۰۹)

کلاسی	زمان اجرا (ماه)	میانۀ نمره‌های خام	معادل کلاسی
۴	مهر (۰/۱)	۵/۷۵	۴/۱
۴	اردیبهشت (۰/۸)	۷/۵	۴/۸
۵	مهر (۰/۱)	۸	۵/۱
۵	اردیبهشت (۰/۸)	۹	۵/۸

• پنجاه درصد دانش‌آموزان در گروه هنجار، در بالای میانۀ نمره خام و ۵۰٪ در پایین میانۀ واقع شده‌اند.

فرض کنید نمره‌های زیر برای ابراهیم، دانش‌آموزی در میانۀ کلاس پنجم، به دست آمده‌است:

ریاضی	۵/۵
اجتماعی	۶/۵
فارسی	۹/۰

در بررسی این نمره‌ها، معلمان، والدین و دانش‌آموزان به‌طور مشابه تشخیص می‌دهند که ابراهیم در ریاضی، دقیقاً در سطح متوسط قرار دارد، در اجتماعی، یک سال جلوتر و در فارسی نیز سه سال و نیم از کلاس خودش بالاتر است. هنجارهای سنی و کلاسی، ایرادهایی دارند:

۱. واحدهای مقیاس این دو نوع هنجار، در طول مقیاس برابر نیستند.
۲. افراد ناآشنا با محدودیت‌های هنجار کلاسی ممکن است آن را اشتباه بفهمند یا تفسیر و برداشت نادرستی از آن داشته باشند.
۳. مقدار پیشرفت در یک سال، در تمام دوران تحصیل ثابت نیست و میزان پیشرفت، از درسی به درس دیگر فرق می‌کند.

تفسیرهای نامناسب از معادل‌های کلاسی، قابل توجه‌اند:

۱. فرض شود که هنجارها استانداردهایی از آن چیزی هستند که باید باشند.
۲. فرض شود که معادل‌های کلاسی، برای جایابی دانش‌آموز در یک کلاس مناسب‌اند.
۳. فرض شود که از همه دانش‌آموزان این انتظار باید باشد که هر سال، یک واحد معادل کلاسی ارتقا یابند.
۴. فرض شود که واحدها در سراسر دامنه نمره‌ها برابرند.
۵. فرض شود که معادل‌های کلاسی دروس مختلف، قابل مقایسه با یکدیگرند.

میلر، لین و گراند (۲۰۰۹)، هرکدام از این فرض‌های ناقص و نتایج بدفهمی‌های معادل کلاسی را به شرح زیر خلاصه کرده‌اند:

۱. هتجارها را با استانداردهایی که باید باشند، اشتباه‌نگیرید. در هر معادل کلاسی، ۵۰٪ دانش‌آموزان، بالای آن و ۵۰٪، پایین آن قرار می‌گیرند. در نتیجه معادل کلاسی نباید نمره‌ای فرض شود که همه دانش‌آموزان یک کلاس خاص باید آن را کسب کنند. خوب یا بد بودن این موضوع، به چندین عامل بستگی دارد؛ مانند توانایی دانش‌آموزان ما، گستره‌ای که آزمون ما پیامدهای یادگیری را اندازه می‌گیرد و کیفیت تسهیلات آموزشی در اختیار ما. اگر ما به دانش‌آموزانی که توانایی آن‌ها بالاتر از متوسط است، تحت شرایط مشابهی با مدارس گروه هتجار درس می‌دهیم، تنها خودمان را با هتجار مطابقت داده‌ایم. از طرف دیگر، اگر دانش‌آموزان ما در پایه‌های قبلی، سابقه ضعف تحصیلی دارند، رسیدن به هتجار شاید سبب مباحث زیادی شود.
۲. از معادل کلاسی، برای جایی پایت تحصیل دانش‌آموز استفاده نکنید. یکی از مواردی که باعث سوءتعبیر هتجارهای کلاسی می‌شود، این تصور است که دانش‌آموزانی که نمره‌های معادل کلاسی خاصی را کسب کرده‌اند، برای تحصیل در آن پایه و کلاس، آمادگی دارند. برای مثال، ممکن است تصور کنیم دانش‌آموز کلاس چهارمی که در آزمون مهارت‌های زبانی، معادل کلاسی ۶۰٪ کسب کرده است، باید بتواند مانند یک دانش‌آموز کلاس ششمی، مهارت‌های زبانی را بفهمد. این تنها به این معناست که این دانش‌آموز، برای آموزشی متفاوت از همکلاسی‌های خودش آمادگی دارد.
۳. انتظار نداشته باشید همه دانش‌آموزان، هر سال ۱۰٪ معادل کلاسی را کسب کنند. اگر نمره معادل کلاسی احمد و محمود در پایان کلاس سوم، به ترتیب ۴۷٪ و ۲۵٪ باشد، طبیعی است انتظار داشته باشیم در پایان سال چهارم، ۱۰٪ واحد به معادل کلاسشان اضافه شود و به ۵۷٪ و ۳۵٪ برسند. این برداشت درست نیست. نخست اینکه هتجارها نباید با استانداردها اشتباه گرفته شوند و دوم اینکه ۱۰٪ واحد ارتقا، ممکن است برای دانش‌آموزانی صدق کند که متوسط‌اند. دانش‌آموزانی که خیلی بالاتر از متوسط‌اند، ممکن است ۱۰٪ معادل کلاسی بیشتر ارتقا یابند و کسانی که خیلی پایین‌تر از متوسط کلاسی هستند، کمتر از ۱۰٪ کلاسی ارتقا یابند.

۴. فرض نکنید واحدها در بخش‌های مختلف مقیاس برابرند. برای مثال، یک سال پیشرفت ریاضی از کلاس ۲/۰ به ۳/۰ ممکن است بیشتر از یک سال پیشرفت از کلاس ۴/۰ تا ۵/۰ یا ۸/۰ تا ۹/۰ باشد؛ بنابراین از آنجا که پیشرفت در موضوعات درسی به‌طور نامساوی صورت می‌گیرد و میزان پیشرفت، از درسی به‌درس دیگر فرق می‌کند، پیشرفت در موضوعات مختلف، معنای متفاوتی خواهد داشت. دانش‌آموزانی که در چندین پایه، معادل کلاسی بالاتر از پایه تحصیلی خود کسب می‌کنند یا پیشرفت آن‌ها خیلی بالاتر از دیگران یا عملکردشان فقط کمی بالاتر از حد متوسط خواهد بود، پیشرفتشان بیشتر از دانش‌آموزان دیگر است، ولی در طول پایه‌های مختلف برابر نیست.

۵. تصور نکنید نمره‌ها در آزمون‌های مختلف، قابل مقایسه‌اند. کسب یک معادل کلاسی ۸/۰ در یک آزمون درک خواندن پایه پنجم و کسب همان نمره در آزمون درک خواندن دیگری که مختص پایه پنجم است، معنای یکسانی ندارند. نمره ۸/۰ در هر دو آزمون، خیلی بالاتر از متوسط پایه پنجم است. ممکن است نمره ۸/۰ در یک آزمون، بالاترین نمره ممکن باشد، ولی در آزمون دیگر که به‌وسیله ناشر دیگری عرضه شده است، نمره بالا به‌شمار آید و گرفتن نمره بیشتر از ۸/۰ نیز امکان‌پذیر باشد.

یادآوری ۱. نمره خام، تعداد امتیازاتی است که دانش‌آموز از آزمونی برابر دستورالعمل نمره‌گذاری شده کسب می‌کند.

یادآوری ۲. در تفسیر آزمون ملاکی، عملکرد فرد را بدون نیاز به مقایسه عملکرد وی با دیگران تفسیر کنیم.

یادآوری ۳. اگر برای تفسیر نمره یک دانش‌آموز، نمره او را با نمره یک گروه مرجع مقایسه کنیم، به آن گروه مرجع گروه هنجار گویند.

یادآوری ۴. هنجارهای سنی، برحسب سال و ماه و با فاصله‌های یک تدریس می‌شوند.

۲-۴-۷ رتبه درصدی

یکی از روش‌هایی که به‌طور گسترده برای توصیف عملکرد آزمون، استفاده و به‌راحتی فهمیده می‌شود، رتبه درصدی است. رتبه درصدی یا (نمره درصدی^۱)، وضعیت نسبی فرد را در گروه برحسب کسانی که نمره پایین‌تر از او گرفته‌اند، مشخص می‌کند. این نمره، نشان می‌دهد چند درصد از افراد گروه مورد مقایسه، نمره بالاتر یا پایین از نمره دانش‌آموز گرفته‌اند. برای مثال، اگر دانش‌آموزی نمره درصدی ۷۲ بگیرد، نشان می‌دهد که این دانش‌آموز خاص، از ۷۲٪ دانش‌آموزانی که وی با آن‌ها مقایسه شده، بهتر عمل کرده‌است.

رتبه درصدی، موقعیت فرد در گروه یا توزیع نمره‌ها را به‌صورت نسبی مشخص می‌کند، نه به‌صورت مطلق (امین‌پور، ۱۳۹۴).

۳-۴-۷ محاسبه رتبه درصدی

برای محاسبه رتبه درصدی، از فرمول ۱-۷ استفاده می‌شود:

$$Pr = \frac{Cf + \frac{F}{2}}{N} \times 100$$

فرمول ۱-۷

Pr رتبه درصدی

Cf فراوانی تراکمی تا طبقه‌ای که قصد محاسبه رتبه درصدی آن را داریم.

F فراوانی طبقه‌ای که قصد محاسبه رتبه درصدی آن را داریم.

تمرین. داده‌های زیر، مربوط به نمره ۲۰ دانش‌آموز در درس قرآن است. رتبه درصدی نمره ۱۷ را در این توزیع مشخص کنید.

۱۴-۱۹-۲۰-۱۴-۱۹-۱۴-۲۰-۱۵-۱۹-۲۰-۲۰-۱۸-۲۰-۱۷-۱۶-۱۶-۱۵

۱۵-۱۷-۱۴

در اولین مرحله، داده‌ها را از کوچک به‌بزرگ مرتب می‌کنیم و آن‌ها را در جدول توزیع فراوانی قرار می‌دهیم.

X	F	Cf
۲۰	۵	۲۰
۱۹	۳	۱۵
۱۸	۱	۱۲
۱۷	۲	۱۱
۱۶	۲	۹
۱۵	۳	۷
۱۴	۴	۴
i=1	$\Sigma F = 16$	-----

سپس داده‌ها را در فرمول رتبه درصدی قرار می‌دهیم.

$$Pr = \frac{9 + \frac{2}{2}}{20} \times 100 = \frac{10}{20} \times 100 = 50$$

رتبه درصدی نمره ۱۷، برابر با ۵۰ است.

یک مشخصه مطلوب هنجارهای درصدی، امکان تفسیر عملکرد دانش‌آموز برحسب گروه‌های گوناگون است. متداول‌تر اینکه عملکرد دانش‌آموز، براساس وضعیت نسبی وی در کلاس یا گروه سنی خودش گزارش می‌شود. استفاده گسترده از هنجارهای درصدی، بدون نقص نیست. هنگام تفسیر رتبه درصدی، همیشه باید به گروه هنجاری که رتبه، مبتنی بر آن است، اشاره شود. نمی‌گوییم دانش‌آموز رتبه درصدی ۸۰ دارد، بلکه می‌گوییم رتبه درصدی ۸۰ را در گروه خاصی دارد. نمره خام در یک آزمون استعداد تحصیلی ممکن است رتبه درصدی برای دانش‌آموزان دبیرستانی در گروه کلی دانش‌آموزان دبیرستانی ۸۰ باشد. برای یک دانش‌آموز پیش‌دانشگاهی، ۶۳ و برای یک دانشجوی ورودی جدید یک دانشگاه ممتاز، ۲۵ باشد. برحسب اینکه فرد عضو چه گروهی باشد، وضعیت نسبی‌اش فرق می‌کند. محدودیت اصلی هنجار درصدی این است که واحدهای آن در طول مقیاس یکسان نیستند. تفاوت صدک‌های وسط مقیاس (برای مثال، ۴۵-۵۵)، خیلی کمتر از تفاوت صدک‌های دو انتهای مقیاس (برای مثال، ۸۵-۹۵ یا ۵-۱۵) است، زیرا تعداد زیادی از افراد هر گروه، نمره‌هایی به دست می‌آورند که در وسط مقیاس قرار دارند درحالی‌که نمره تعداد کمی از افراد، خیلی کم یا خیلی زیاد است. برای مثال، اگر نمره کسی که ۱۴ گرفته است (برای مثال، رتبه درصدی این نمره، ۴۷ است)، نیم نمره

تغیر کنند، احتمالاً رتبه درصدی او تغییر زیادی خواهد داشت (برای مثال ۱۰ درصد و رتبه درصدی او، به ۵۷ خواهد رسید)، اما اگر کسی نمره ۳ گرفته باشد (برای مثال رتبه درصدی او ۵ باشد)، باید تغییر زیادی در نمره اش ایجاد شود تا رتبه درصدی او ۱۰ درصد افزایش یابد (برای مثال نمره اش از ۳ به ۸ برسد تا رتبه درصدی اش ۱۵ شود).

رتبه‌های درصدی، از نوع مقیاس رتبه‌ای هستند، نه مقیاس فاصله‌ای؛ و به همین دلیل واحدهای رتبه‌های درصدی در طول مقیاس، برابر نیستند. این نابرابری واحدها در طول مقیاس ایجاد می‌کند رتبه‌های درصدی، با احتیاط تفسیر شوند. نخست اینکه تفاوت چند درصدی در کناره‌های توزیع، در مقایسه با وسط توزیع، بایستی وزن بیشتری داشته باشد. در واقع تغییرات جزئی وسط توزیع را معمولاً می‌توان نادیده گرفت. دوم اینکه نمی‌توان برای رتبه‌های درصدی، میانگین محاسبه کرد. میانگین مناسب برای رتبه‌های درصدی، رتبه ۵۰ درصدی است. این رتبه، معرف نقطه وسط توزیع است و میانه نام دارد.

یادآوری ۱. رتبه درصدی نشان می‌دهد چند درصد از افراد گروه مورد مقایسه، نمره بالاتر یا پایین‌تر از نمره دانش‌آموز گرفته‌اند.

یادآوری ۲. رتبه‌های درصدی، از نوع مقیاس رتبه‌ای هستند.

۷-۴-۴ نمره‌های استاندارد

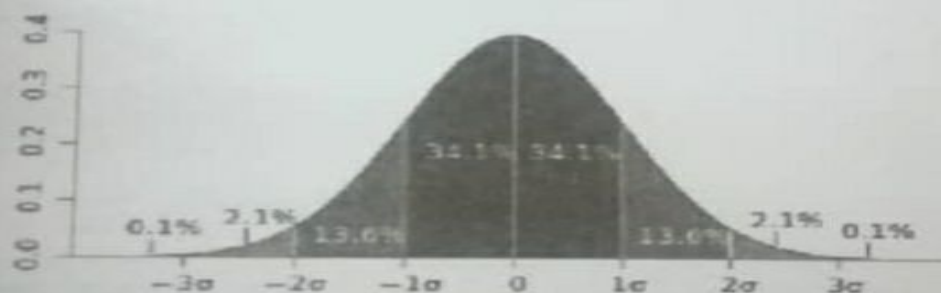
با توجه به مطالب قبل، واحدهای هتجارهای ستی، کلاسی و درصدی، نامساوی‌اند و این عدم تساوی، مقایسه افراد مختلف را از لحاظ عملکرد با یکدیگر، با مشکل مواجه می‌سازد. برای رفع این مشکل، از نمره‌های استاندارد استفاده می‌شود. نمره‌های استاندارد، از نوع مقیاس فاصله‌ای هستند که اندازه واحدهای آن‌ها در سراسر مقیاس برابر است. خاصیت نمره‌های استاندارد این است که بر پایه میانگین و انحراف استاندارد، نمره‌هایی به دست می‌آیند. در این روش، از طریق یک فرمول (Z)، نمره خام به نمره استاندارد تبدیل و سپس مشخص می‌شود این نمره، چقدر بالاتر یا پایین‌تر از میانگین قرار دارد. میانگین، این‌گونه به دست می‌آید که تمام نمره‌های خام یا هم جمع و سپس بر تعداد آن‌ها تقسیم می‌شوند. انحراف استاندارد (S)، پراکندگی نمره‌ها را از میانگین نشان می‌دهد (نحوه محاسبه انحراف استاندارد، اینجا توضیح داده نمی‌شود. برای

کسب اطلاع بیشتر، به کتاب‌های آمار توصیفی به‌ویژه کتاب آمار توصیفی در روان‌شناسی و علوم تربیتی تألیف امین‌پور چاپ پیام نور مراجعه کنید).
به‌منظور درک بیشتر موضوع، به‌اختصار دربارهٔ توزیع نرمال بحث می‌شود.

۵-۷ توزیع نرمال

توزیع نرمال، نوع ویژه‌ای از توزیع متقارن است که ویژگی‌های آماری خاصی را به‌نمایش می‌گذارد. این توزیع، برای ما خیلی مهم است، زیرا الگوی موردنیاز برای مقایسهٔ نمره‌ها و دیگر تصمیم‌های آماری است. مهم است بدانیم که توزیع نرمال، یک توزیع فرضی است.

از منظر تفسیر آزمون، یکی از مفیدترین کاربردهایش این است که هنگام تبدیل شدن به‌واحد انحراف استاندارد، در زیر هر بخش آن، درصد ثابتی از افراد قرار می‌گیرند. درصدهای هر بخش منحنی نرمال، به‌شرح زیر هستند:



شکل ۷-۱. منحنی نرمال

۳۴٪، بین میانگین و $+1$ انحراف استاندارد، ۱۴٪، بین $+1$ و $+2$ انحراف استاندارد و ۲٪، بین $+2$ و $+3$ انحراف استاندارد قرار می‌گیرند (همهٔ درصدها گرد شده‌اند). البته همین توزیع برای انحراف استاندارهای پایین‌تر از میانگین نیز صدق می‌کند. تنها ۰/۱۳٪ زیر -3 یا $+3$ انحراف استاندارد قرار می‌گیرند.

۶-۷ انواع نمره‌های استاندارد

نمره‌های استاندارد زیادی وجود دارند که در آزمون، مورد استفاده قرار می‌گیرند. همه

نمره‌های استاندارد تبدیل خطی ساده، نمره‌های خام‌اند. میانگین و انحراف استاندارد توزیع نمره‌های تبدیل‌شده، با میانگین و انحراف استاندارد نمره‌های خام تفاوت دارد، اما شکل توزیع در هر دو مورد، یکسان است، زیرا همه نمره‌های استاندارد، مبتنی بر اصول مشابهی هستند و به‌نوعی به‌شیوه مشابهی تفسیر می‌شوند. در این قسمت، فقط متداول‌ترین آن‌ها بررسی می‌شوند.

۶-۷-۱ نمره Z

ساده‌ترین نمره استاندارد که سایر نمره‌ها هم مبتنی بر آن هستند، نمره Z است. نمره Z به‌ما نشان می‌دهد یک نمره خام، به‌اندازه چند واحد انحراف استاندارد، بالاتر یا پایین‌تر از میانگین است. برای محاسبه نمره Z، از فرمول استفاده می‌شود.

$$Z = \frac{X - \bar{X}}{S}$$

فرمول ۷-۲

X نمره خام

 $\bar{X}$ میانگین حسابی نمره‌های خام

S انحراف استاندارد نمره‌های خام

فرمول Z فرمول ساده‌ای است و با چند تمرین ساده، به‌راحتی می‌توان بر آن تسلط پیدا کرد. فرض کنید دو نفر در یک آزمون هوشی، به‌ترتیب نمره‌های خام ۸۵ و ۱۱۵ می‌گیرند. انحراف استاندارد آزمون، ۱۵ و میانگین، نمره ۱۰۰ است. نمره صفر، برابر فرمول Z زمانی به‌دست می‌آید که نمره آزمودنی، برابر میانگین نمره گروه باشد.

مثال.

$$Z = \frac{X - \bar{X}}{S} = \frac{100 - 100}{15} = 0$$

$$Z = \frac{X - \bar{X}}{S} = \frac{85 - 100}{15} = \frac{-15}{15} = -1$$

نمره Z آزمودنی اول، برابر -۱ و به‌این معناست که این نمره، ۱ انحراف استاندارد پایین‌تر از میانگین، واقع شده‌است و هوش‌یهر آزمودنی موردنظر، از ۱۶٪ افراد مورد مقایسه بالاتر است. هر انحراف استاندارد در این آزمون، برابر ۱۵ نمره است؛ بنابراین یک انحراف استاندارد پایین‌تر از میانگین (۱۰۰) یعنی ۸۵ = ۱۰۰ - ۱۵. به‌عبارت دیگر،

کسی که هوش بهرش یک انحراف استاندارد پایین‌تر از میانگین است (در مثال بالا)، یعنی هوش بهرش ۸۵ است.

$$Z = \frac{X - \bar{X}}{S} = \frac{130 - 100}{15} = \frac{+30}{2} = +2$$

نمره Z آزمودنی دوم، برابر +۲ و به این معناست که این نمره، +۲ انحراف استاندارد بالاتر از میانگین، واقع شده است و آزمودنی مورد نظر، هوش بهرش از ۹۵٪ افراد مورد مقایسه بالاتر است. دو انحراف استاندارد بالاتر از میانگین، یعنی ۱۳۰ = ۱۰۰ + ۳۰. هر انحراف استاندارد، معادل ۱۵ نمره است و ۲ انحراف استاندارد، یعنی ۱۵ + ۱۵ که می‌شود ۳۰. به عبارت دیگر، کسی که هوش بهرش، دو انحراف استاندارد بالاتر از میانگین بوده، یعنی هوش بهرش ۱۳۰ است.

میانگین نمره Z همیشه «صفر» و انحراف استاندارد آن، مساوی «یک» است. یکی از معایب عمده نمره Z، منفی بودن نیمی از نمره‌های آن است. اگر نمره خام، کوچک‌تر از میانگین باشد، نمره Z همیشه منفی است. بیشتر افراد ترجیح می‌دهند با نمره‌های منفی کار نکنند، زیرا در محاسبه این‌گونه نمره‌ها احتمال خطا بیشتر است. فراموش کردن علامت منفی ممکن است باعث خطای جدی در تفسیر نمره‌های آزمون شود. به همین دلیل است که در هنجارهای آزمون‌ها به ندرت به‌طور مستقیم از نمره‌های Z استفاده می‌شود و نمره Z اغلب به نمره استاندارد تبدیل خواهد شد که همیشه از اعداد مثبت استفاده می‌کند.

۷-۶-۲ نمره T

نمره‌های معیار دیگری که به نمره‌های T شهرت دارند، برای رفع مشکلات نمره‌های Z ابداع شده‌اند. نمره T، دارای میانگین ۵۰ و انحراف استاندارد ۱۰ است. مقیاس نمره‌های T، با ضرب نمره‌های Z در ۱۰ و جمع کردن نمره‌های حاصل با ۵۰ به دست می‌آید. نمره Z به این دلیل در ۱۰ ضرب می‌شود که علامت اعشار آن از بین برود و برای از بین بردن علامت منفی، با ۵۰ جمع می‌شود (در مواردی که نمره Z بالاتر از -۵ است، این مورد صدق نمی‌کند). برای محاسبه نمره‌های T می‌توان از فرمول ۷-۳ استفاده کرد.

$$= 10(Z) + 50$$

فرمول ۷-۳

مثال. اگر نمره‌های دو دانش‌آموز در مقیاس Z برابر $۰/۵$ و $-۱/۵$ باشد، نمره‌های آن‌ها در مقیاس T به شرح زیر خواهد بود:

$$T = 10(-1/5) + 50 \qquad T = 10(0/50) + 50$$

$$T = -15 + 50 = 35 \qquad T = 5 + 50 = 55$$

از آنجا که همه نمره‌های T دارای میانگین ۵۰ و انحراف استاندارد ۱۰ هستند، هر نمره T به تنهایی و به طور مستقیم، قابل تفسیر و استفاده است. برای مثال، نمره $T = ۵۵$ حاکی از این است که آن نمره، به اندازه نصف انحراف استاندارد، بالای میانگین است و نمره $T = ۳۵$ نشان می‌دهد که آن نمره، به اندازه یک و نیم انحراف استاندارد، پایین‌تر از میانگین قرار دارد تا آخر (میلر، لین و گراندلند، ۲۰۰۹).

یک دلیل برای ترجیح نمره T به نمره Z این است که نمره T ، فقط عدد صحیح به دست می‌دهد. برای مثال، گزارش نمره آزمون یک فرد به صورت ۳۳ ، ارجح‌تر است تا اینکه آن را به صورت نمره $Z = -۱/۷$ نشان دهیم. با وجود این، دو نمره با هم معادل‌اند. از آنجا که نمره‌های T همیشه میانگین ۵۰ و انحراف استاندارد ۱۰ دارند، هر نمره T مستقیماً قابل تفسیر است.

نمره‌های بهنجار شده T به وسیله:

- تبدیل توزیع نمره‌های خام به رتبه‌های درصدی
- جست‌وجوی نمره Z هر رتبه درصدی در توزیع نرمال و اختصاص این نمره‌ها Z به نمره‌های خام مربوطه
- تبدیل نمره‌های Z به نمره‌های T با استفاده از فرمول قبل

نمره‌های تراز شده بهنجار شده بسیار دیگری مانند IQ ، SAT ، SAS و امثال آن وجود دارند که به روش نمره‌های T محاسبه می‌شوند، اما میانگین و انحراف استانداردشان متفاوت است. برای مثال، کسی که نمره هوش‌بهرش ۱۱۵ است که یک انحراف استاندارد بالاتر از میانگین است، معنایی مشابه $T = ۶۰$ دارد (در آزمون‌های هوشی، معمولاً میانگین ۱۰۰ و انحراف استاندارد ۱۵ است). به منظور کسب اطلاعات بیشتر در این زمینه، به کتاب آمار توصیفی در روان‌شناسی و علوم تربیتی تألیف امین‌پور چاپ پیام نور مراجعه کنید.

۷-۶-۳ نمره‌های ۹ بخشی

نوع دیگر هنجار به نام نمره‌های ۹ بخشی وجود دارد که به صورت نمره‌های استاندارد یک‌رقمی از ۱ تا ۹ نشان داده می‌شود. علت نام‌گذاری این مقیاس نمره‌ها به ۹ بخشی این است که توزیع این نمره‌ها، به ۹ بخش محدود می‌شود. این نمره‌ها ابتدا برای درجه‌بندی افراد در نیروی هوایی آمریکا و بعدها برای سایر مقاصد آموزشی و ارزشیابی، موارد استفاده فراوان پیدا کرد. میانگین این نمره‌ها ۵ و انحراف استاندارد آن‌ها تقریباً ۲ است. عدد ۵ در مقیاس ۹ بخشی، دقیقاً در وسط توزیع واقع می‌شود. در توزیع نرمال، رتبه‌های درصدی ۴۰ تا ۵۹، مطابق نمره ۵ مقیاس ۹ بخشی است (میلر، لین و گرانلند، ۲۰۰۹). برای محاسبه نمره‌های ۹ بخشی، از فرمول ۷-۴ استفاده می‌شود.

$$S_9 = 2(z) + 5$$

فرمول ۷-۴

تمرین. اگر نمره استاندارد دانشجویی در درس اندیشه اسلامی، برابر با $1/3$ باشد، آن را به نمره ۹ بخشی تبدیل کنید؟

$$S_9 = 2(1/3) + 5 = 7/6$$

مزیت عمده استفاده از نمره‌های ۹ بخشی طبیعی بودن تقریبی توزیع آن‌هاست و هریک از قسمت‌های نه‌گانه، مربوط به یک نمره است. نمره‌های ۹ بخشی، برای انتخاب آزمودنی‌ها به صورت کلی مفیدند، اما از آنجا که دارای ۹ نمره مختلف است، با استفاده از آن‌ها نمی‌توان آزمودنی‌ها را به دقت از هم تشخیص داد. دامنه رتبه‌های درصدی برای هریک از نمره‌های ۹ بخشی، در جدول ۷-۳ نشان داده شده است.

جدول ۷-۳. دامنه رتبه‌های درصدی معادل نمره‌های ۹ بخشی

نمره ۹ بخشی	دامنه رتبه‌های درصدی
۹	۹۶ تا ۱۰۰
۸	۸۹ تا ۹۵
۷	۷۷ تا ۸۸
۶	۶۰ تا ۷۶
۵	۴۰ تا ۵۹
۴	۲۳ تا ۳۹
۳	۱۱ تا ۲۲
۲	۴ تا ۱۰
۱	۰ تا ۳

میلر، لین و گرانلند (۲۰۰۹)، نقاط قوت نمره‌های ۹ بخشی را به شرح زیر، خلاصه کرده‌اند:

۱. سیستم ۹ بخشی، از یک مقیاس ۹ نقطه‌ای بهره‌می‌گیرد که در آن، ۹ معادل بالا، ۱ پایین و ۵ معادل میانگین است؛ بنابراین سیستم را می‌توان به راحتی برای والدین و دانش‌آموزان توضیح داد.
۲. نمره‌های ۹ بخشی، نمره‌های تراز شده بهنجار شده‌ای هستند که امکان مقایسه عملکرد دانش‌آموزان را در آزمون‌های مختلف فراهم می‌سازد (اگر براساس یک گروه مشترک باشند). یک تفاوت دو نمره‌ای در مقیاس ۹ بخشی، نشانگر تفاوت معناداری در عملکرد است. برای مثال اگر نمره ۹ بخشی دانش‌آموزی در ریاضی، ۷ و در هجی کردن ۵ باشد، احتمالاً عملکرد بهتری در ریاضی دارد.
۳. نظام ۹ بخشی، امکان ترکیب انواع مختلف داده‌ها را ساده می‌کند (برای مثال نمره‌های آزمون‌ها، مقیاس‌های درجه‌بندی و داده‌های رتبه‌بندی شده)، زیرا نمره‌های ۹ بخشی، مثل رتبه‌های درصدی محاسبه شده، اما در قالب نمره‌های استاندارد بیان می‌شوند؛ بنابراین تبدیل به نمره‌های ۹ بخشی ساده است و از آنجا که ویژگی نمره‌های استاندارد را دارند، این امکان را ایجاد می‌کنند که نمره‌های ۹ بخشی را که از اندازه‌گیری‌های مختلف به دست می‌آید، با هم جمع بزنیم و نمره‌ای ترکیبی به دست آوریم.
۴. با توجه به اینکه نظام ۹ بخشی از یک نمره تک‌رقمی استفاده می‌کند، به راحتی ثبت می‌شود و فضای کمتری از سایر نمره‌ها اشغال می‌کند.

یادآوری ۱. میانگین و انحراف استاندارد نمره Z ، به ترتیب ۰ و ۱ است.

یادآوری ۲. میانگین و انحراف استاندارد نمره T ، به ترتیب ۵۰ و ۱۰ است.

یادآوری ۳. میانگین و انحراف استاندارد نمره ۹ بخشی، به ترتیب ۵ و ۲ است.

۷-۷ نیم‌رخ

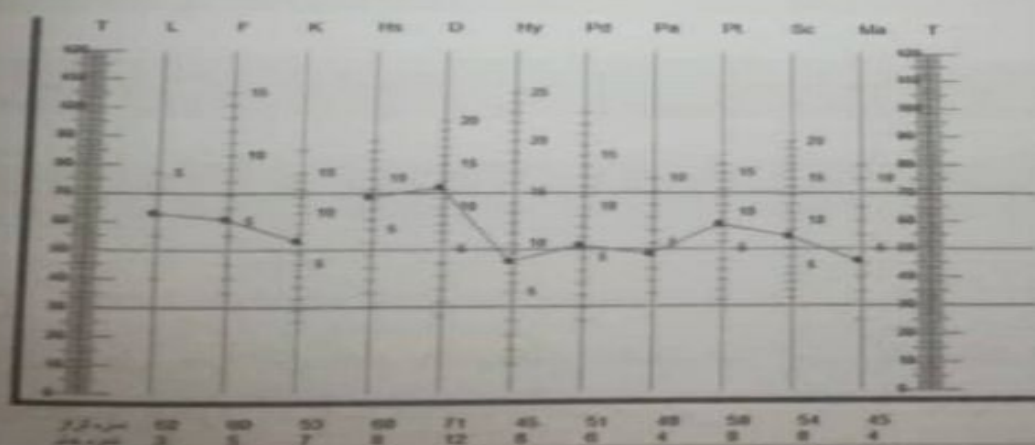
یکی از مزایای تبدیل نمره‌های خام به نمره‌های مشتق این است که عملکرد آزمودنی در آزمون‌های مختلف، به طور مستقیم مقایسه می‌شوند. این کار اغلب به کمک یک نیم‌رخ^۱

آزمون انجام می‌شود. در نیم‌رخ آزمون‌ها، نمره‌های آزمودنی در آزمون‌های مختلف یا در خرده‌مقیاس‌های یک آزمون، ممکن است به صورت رتبه‌های درصدی، نمره‌های تراز شده یا هوش‌بهر نشان داده شود. نیم‌رخ را می‌توان به صورت نمودار ستونی یا نمودار چندضلعی ترسیم کرد. برای رسم نیم‌رخ، ابتدا دو محور عمود بر هم رسم می‌شود. سپس روی محور افقی، نام آزمون یا ویژگی مورد اندازه‌گیری و روی محور عمودی، نمره تراز شده یا نمره هنجاری خصایص اندازه‌گیری شده مشخص می‌شود.

با نگاهی به نیم‌رخ‌ها به راحتی می‌توان نقاط قوت و ضعف آزمودنی را مشخص کرد. نیم‌رخ، وسیله‌ای است که به کمک آن می‌توان نمره‌هایی را که دانش‌آموز (یا گروهی از دانش‌آموزان) در دو یا چند درس می‌گیرد، با هم مقایسه کرد. این مقایسه به کمک اعداد خام، درست نخواهد بود. هدف از ترسیم نیم‌رخ، نشان دادن نتایج به صورت تصویر یا نمودار است.

درواقع نیم‌رخ یعنی بیان تصویری یا نموداری نتایج. نیم‌رخ، نقاط قوت و ضعف عملکرد یک فرد را نشان می‌دهد. درست همان‌طور که تصویر یک فرد، بیانگر وضع ظاهری او است، نیم‌رخ نیز نمره‌های وضع تحصیلی و روانی او را نشان خواهد داد.

مثال. برای مقایسه نمره یک دانش‌آموز در آزمون درس علوم که تنها به ۱۰ سؤال آن جواب داده، با نمره همین دانش‌آموز در آزمون درسی ریاضی که به ۱۶ سؤال آن پاسخ گفته است، باید این دو نمره را به نوعی هنجار کلاسی، سنی یا درصدی تبدیل کنیم تا مقایسه آن‌ها امکان‌پذیر شود.



شکل ۷-۲. نمونه‌ای از نیم‌رخ آزمودنی در آزمون MMPI

برای به دست آوردن نیم رخ عملکرد دانش آموز، می توان نمره های آزمون های مختلف یا خورده آزمون های یک آزمون را به یک هتجار تبدیل و سپس آن ها را روی یک نمودار رسم کرد.
باید توجه داشت:

تفاوت نمره های یک دانش آموز در آزمون های مختلف ممکن است ناشی از خطای اندازه گیری آزمون ها باشد، نه یک تفاوت واقعی؛ بنابراین تفاوت میان نمره های آزمون ها را باید با احتیاط تفسیر کرد.

روش دیگر ترسیم نیم رخ نمره هایی که برای رفع مشکل خطای اندازه گیری پیشنهاد شده این است که به جای نشان دادن نمره های آزمون به صورت نقطه هایی روی نیم رخ نمره ها، عملکرد دانش آموز به صورت توارهایی ترسیم شود که هر توار، به اندازه یک خطای معیار اندازه گیری بالای نمره دانش آموز و یک خطای اندازه گیری پایین آن نمره را شامل شود.

یادآوری. هدف ترسیم نیم رخ، نشان دادن نتایج به صورت تصویر یا نمودار است.

خلاصه فصل هفتم

در این فصل، تفسیر نمره ها و هتجارهای آزمون، مورد بررسی قرار گرفت. به منظور یکسان سازی تفسیر نمره ها در اندازه گیری های تربیتی، روش های متعددی ابداع شده اند.

نمره خام. خلاصه ای عددی از عملکرد آزمون دانش آموز بوده و بدون اطلاعات بعدی، خیلی معنادار نیست. برای تفسیر نمره های خام، دو نوع تفسیر ملاکی و هتجاری مورد استفاده قرار می گیرد. آزمون ملاکی یا ملاک مرجع، به آزمونی گفته می شود که در آن، معیار یا ملاک ارزشیابی، از پیش تعیین شده است و عملکرد یادگیرنده در آزمون، با توجه به آن ملاک سنجش می شود.

نمره های هتجاری. نمره های استاندارد به ما اجازه می دهند عملکرد یک دانش آموز را در یک آزمون، با عملکرد وی در آزمونی دیگر مقایسه کنیم. به طور خلاصه هدف عمده در استفاده از نمره های هتجاری، تعیین وضعیت یک دانش آموز در مقایسه با دانش آموز دیگر است.

به منظور به دست آوردن یک چهارچوب کلی‌تر برای تفسیر هنجاری، نمره‌های خام به نوعی نمره مشتق تبدیل می‌شوند. این نمره‌ها انواع مختلفی دارند که معروف‌ترین آن‌ها عبارت‌اند از: هنجارهای سنی و کلاس معادل، رتبه‌های درصدی و نمره‌های استاندارد.

برای مقایسه نمره‌های آزمودنی، انواع هنجار وجود دارد که عبارت‌اند از: هنجارهای سنی و کلاسی. هنجار سنی (هنجار معادل سن تقویمی)، نمره میانی یک آزمون است که افراد گروه سنی معینی، نمره‌ای برابر آن گرفته‌اند. هنجارهای سنی، برحسب سال و ماه و یا فاصله‌های یک ماه تدوین می‌شوند. همچنین میلر، لین و گرانلند (۲۰۰۹)، فرض‌های ناقص و نتایج بدفهمی‌های این هنجار را خلاصه کرده، مورد مطالعه قرار داده‌اند.

رتبه درصدی. رتبه درصدی (نمره درصدی) وضعیت نسبی فرد را برحسب کسانی در گروه که نمره پایین‌تر از او گرفته‌اند، مشخص می‌کند.

نمره‌های استاندارد. نمره‌های استاندارد، از نوع مقیاس فاصله‌ای هستند که اندازه واحدهای آن‌ها در سراسر مقیاس برابر است. خاصیت نمره‌های استاندارد این است که برحسب میانگین و انحراف استاندارد قرار دارند.

توزیع نرمال. توزیع نرمال، الگویی است که برای مقایسه نمره‌ها و دیگر تصمیم‌های آماری، مورد نیاز است. انواع نمره‌های استاندارد نیز مورد بحث قرار گرفت. این نمره عبارت‌اند از: نمره Z (میانگین نمره Z ، صفر و انحراف استاندارد آن، یک است. این نمره نشان می‌دهد یک نمره خام، به اندازه چند واحد انحراف استاندارد، بالاتر یا پایین‌تر از میانگین است) و نمره T (نمره T ، دارای میانگین ۵۰ و انحراف استاندارد ۱۰ است). نمره‌های تراز شده بهنجار شده بسیار دیگری مانند IQ ، SAT ، SAS و نه بخشی و امثال آن وجود دارند که به روش نمره‌های T محاسبه می‌شوند، اما میانگین و انحراف استاندارد آن‌ها متفاوت است).

نیم‌رخ. نیم‌رخ یعنی بیان تصویری یا نموداری نتایج. نیم‌رخ، نقاط قوت و ضعف عملکرد یک فرد را نشان می‌دهد. درست همان‌گونه که تصویر یک فرد، نشان‌دهنده وضع ظاهری اوست، نیم‌رخ نمره‌ها نیز وضع تحصیلی و روانی او را نشان می‌دهد. ادامه، توزیع بهنجار و مشخصات آن مورد بحث قرار گرفت.

خود آزمایی چهارگزینه‌ای فصل هفتم

۱. تعداد سؤال‌هایی که آزمودنی در یک آزمون به‌درستی پاسخ می‌دهد، نمره نامیده می‌شود.

الف) خام	ب) استاندارد
ج) هتجار	د) تراز
۲. هنگام مقایسه عملکرد فرد با ملاک از پیش تعیین‌شده، از استفاده می‌کنند.

الف) نمره مشتق	ب) هتجار ملاکی
ج) نمره استاندارد	د) نیم‌رخ روانی
۳. کدام آزمون، جزء آزمون‌های هتجاری نیست؟

الف) هوشی	ب) شخصیتی
ج) کلاسی	د) کنکور
۴. کدام هتجار، مانند سن تقویمی در نظر گرفته می‌شود؟

الف) رتبه درصدی	ب) نمره استاندارد
ج) کلاسی	د) سنی
۵. علی در درس آمار توصیفی، نمره ۱۵ گرفته و رتبه درصدی او، معادل ۸۰ است، یعنی علی:

الف) از ۵۰ درصد دانشجویان، عملکرد بهتری دارد.
ب) از ۸۰ درصد دانشجویان، عملکرد بهتری دارد.
ج) از ۲۰ درصد دانشجویان، عملکرد بهتری دارد.
د) از ۸۵ درصد دانشجویان، عملکرد بهتری دارد.
۶. رتبه‌های درصدی، از نوع کدام مقیاس‌اند؟

الف) اسمی	ب) ترتیبی
ج) فاصله‌ای	د) نسبی
۷. اگر دانشجویی در درس آمار توصیفی، نمره ۱۵ کسب کند و میانگین و انحراف استاندارد نمره‌های دانشجویان در این کلاس، به‌ترتیب ۱۲ و ۳ باشد، نمره استانداردش چند خواهد شد؟

الف) -۱	ب) +۱
ج) +۲	د) -۲

۸. اگر عملکرد آزمودنی در یک آزمون، برابر با $Z=+1$ باشد، رتبه درصدی وی چند خواهد شد؟

- الف) ۱۴
ب) ۵۰
ج) ۶۸
د) ۸۴

۹. اگر نمره خام، برابر با میانگین عملکرد گروه باشد، نمره استاندارد Z چند خواهد شد؟

- الف) ۰
ب) ۱
ج) -۱
د) ۵۰

۱۰. میانگین و انحراف استاندارد نمره استاندارد Z به ترتیب برابر است با و

- الف) ۱ و ۰
ب) ۰ و ۲
ج) ۵۰ و ۱۰
د) ۱۰ و ۱

۱۱. میانگین و انحراف استاندارد نمره استاندارد T به ترتیب برابر است با و

- الف) ۲۰۰ و ۱۵
ب) ۱ و ۰
ج) ۵۰ و ۱۰
د) ۵۰ و ۱۰۰

۱۲. اگر نمره $T=67$ باشد، نمره IQ چند خواهد شد؟

- الف) ۱/۷
ب) ۵۷
ج) ۱۲۵
د) ۸

۱۳. اگر نمره $Z=1/5$ باشد، نمره T چند خواهد شد؟

- الف) ۳۵
ب) ۵۵
ج) ۶۵
د) ۸۰

۱۴. اگر نمره دانشجویی برابر با $T=80$ باشد، نمره Z چند خواهد شد؟

- الف) -۳
ب) ۱
ج) -۱
د) +۳

۱۵. اگر نمره فردی در آزمون هوشی برابر با ۱۱۵ باشد، نمره T چند خواهد شد؟

- الف) ۵۵
ب) ۶۰
ج) ۷۰
د) ۷۵

۱۶. میانگین و انحراف استاندارد کدام نمره، برابر با ۵ و ۲ است؟

- الف) ۹ بخشی
ب) نمره T
ج) نمره Z
د) نمره IQ

۱۷. بیان تصویری نتایج آزمون‌ها را می‌نامند.

الف) نمره T

ب) نیم‌رخ

د) نمره ۹ بخشی

ج) نمره استاندارد

۱۸. در توزیع بهنجار، چند درصد از نمره‌ها بین نمره T برابر با ۴۰ و ۶۰ قرار می‌گیرند؟

الف) ۹۵

ب) ۷۲

د) ۲۷

ج) ۶۸

خودآزمایی تشریحی فصل هفتم

۱. آیا نمره دانش‌آموز در درس علوم تجربی، مقیاس نسبی است؟

۲. منظور از کسب نمره خام در آزمون چیست؟

۳. منظور از تفسیر نمره‌های ملاکی و هنجاری چیست؟

۴. هنگام سنجش عملکرد دانش‌آموز با توجه به سرعت انجام عملکرد (تایپ ۴۰ کلمه در دقیقه)، از تفسیر استفاده شده است.

۵. معروف‌ترین نمره‌های مشتق برای تفسیر عملکرد آزمودنی را نام ببرید و شرح دهید؟

۶. ایرادهای وارد بر هتجار سنی و کلاسی را بنویسید؟

۷. فرض‌های ناقص و نتایج بدفهمی‌های معادل کلاسی را از نظر میلر و همکاران (۲۰۰۹) بیان کنید؟

۸. کاربرد نمره‌های استاندارد در سنجش را با بیان مثال شرح دهید؟

۹. انواع نمره‌های استاندارد را نام ببرید و با ارائه مثال، کاربردهای آن‌ها را بنویسید؟

۱۰. اگر در توزیع بهنجار، نمره خام برابر با میانگین باشد، رتبه درصدی چند خواهد شد؟

۱۱. نقاط قوت نمره‌های ۹ بخشی را از نظر میلر، لین و گرانتلند (۲۰۰۹) بنویسید.

۱۲. منظور از نیم‌رخ روانی در تفسیر نمره‌های آزمودنی چیست؟

۱۳. ویژگی‌های منحنی بهنجار را بیان کنید.

فصل هشتم

اعتبار آزمون

هدف کلی

مطالعه اعتبار آزمون و روش‌های تعیین آن

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. اعتبار آزمون را تعریف کند.
۲. روش‌های تعیین اعتبار آزمون را نام ببرد.
۳. در تعیین اعتبار آزمون، روش بازآزمایی را تعریف کند.
۴. در تعیین اعتبار آزمون، روش فرم‌های موازی و هم‌تا را تعریف کند.
۵. در تعیین اعتبار آزمون، روش‌های برآورد همسانی درونی را نام ببرد و تعریف کند.
۶. خطای معیار اندازه‌گیری را تعریف کند.

مقدمه

هر ابزار اندازه‌گیری باید دو مشخصه داشته باشد: اعتبار^۱ و روایی^۲. به بیان دیگر، ابزار اندازه‌گیری هم باید دقیق باشد (اعتبار) و هم اینکه دقیقاً همان چیزی را اندازه بگیرد که برای اندازه‌گیری آن، ساخته شده (روایی) است. داشتن تنها یکی از این دو مشخصه، نشانه بی‌اعتباری وسیله اندازه‌گیری است. اگر برای اندازه‌گیری طول یک اتاق، چند

1. Reliability
2. Validity

وسیله اندازه‌گیری مانند وجب، کش لاستیکی، متر یا متر لیزری داشته باشید، به‌کدام یک بیشتر اعتماد خواهید داشت؟ طبیعی است که متر لیزری را بر بقیه ترجیح می‌دهید. دلیل این امر چیست؟ بدیهی است که به دلیل دقتش آن را انتخاب می‌کنید، زیرا اگر بارها ابعاد آن مکان را اندازه‌گیری کنید، دقیقاً به یک عدد خواهید رسید. ابزارهایی نظیر وجب و کش لاستیکی، کمتر قابل اعتمادند، زیرا اگر ابعاد یک چیز، چندین بار با آن اندازه‌گیری شود، نتایج یکسانی به دست نمی‌آید. شما برای اندازه‌گیری وزن خود، به ترازویی که هر روز وزن متفاوتی را نشان می‌دهد، اعتماد نمی‌کنید، بلکه دنبال ابزار دقیق‌تری خواهید گشت. این موضوع نشان می‌دهد که دقت، عامل مهمی است. غیر از موضوع دقت، از هر وسیله‌ای هم باید برای اندازه‌گیری آن چیزی استفاده کرد که با هدف اندازه‌گیری آن ساخته شده است. از ترازو نمی‌توان برای اندازه‌گیری طول استفاده کرد.

۸-۱ تعریف اعتبار

اعتبار به دقت اندازه‌گیری و ثبات و اعتبار یک وسیله اندازه‌گیری اشاره می‌کند و دو معنای متفاوت دارد (ولوتن^۱ و شوویرس^۲، ۲۰۰۵). یک معنای اعتبار، ثبات و اعتبار نمره‌های آزمون در طول زمان است. بدین معنا که اگر یک آزمون، چند بار درباره یک آزمودنی اجرا شود، نمره وی در همه موارد یکسان باشد. معنای دوم اعتبار، به همسانی درونی آن اشاره دارد و مفهوم آن این است که سؤال‌های آزمون، تا چه اندازه با یکدیگر همبستگی متقابل دارند. اگر ضریب همسانی درونی آزمون کم باشد، بدان معناست که سؤال‌های مختلف آزمون، متغیر واحدی را اندازه‌گیری نمی‌کنند. به عبارت دیگر، آزمون ممکن است دارای نوعی اشکال باشد؛ بنابراین آزمونی که بین سؤال‌های آن، تشابه و هماهنگی وجود داشته باشد، از همسانی درونی بالایی برخوردار است، اما باید دانست که گاهی بالا بودن همسانی درونی آزمون، به کاهش روایی منجر می‌شود.

رابطه بین روایی و اعتبار یک آزمون، این گونه است که یک آزمون باید پایا^۳ باشد تا بتواند روا^۴ باشد (کوهن، ۲۰۰۹). اگر از یک آزمون هوشی در هر بار اجرا روزه تعدادی افراد، نتایج مختلفی حاصل شود، یک آزمون پایا و قابل اعتماد نخواهد بود.

چیزی را به درستی اندازه نمی گیرد. بدیهی است از آنجا که مقیاس اندازه گیری آزمون های هوشی، نسبی نیست، دقت آن ها در حد متر دیجیتال یا متر معمولی نخواهد بود، بلکه به لحاظ دقت، بیشتر شبیه وجب و کش لاستیکی اند. از یک آزمون هوشی هم نمی توان برای اندازه گیری شخصیت استفاده کرد؛ بنابراین برای روایی یک آزمون، باید نخست پایا باشد؛ یعنی اعتبار، شرط روایی است، اما روایی برای اعتبار، ضروری نیست (سیف، ۱۳۹۶).

یادآوری ۱. اعتبار یک آزمون، به دقت اندازه گیری و ثبات یک وسیله اندازه گیری اشاره دارد.

یادآوری ۲. یک آزمون برای داشتن روایی، باید پایا باشد.

۸-۲ روش های تعیین اعتبار

برای محاسبه اعتبار آزمون، از روش های مختلفی استفاده خواهد شد. در ادامه، روش های بازآزمایی، فرم های موازی و همتا، دوتیمه کردن، آلفای کرونباخ و کودر ریچاردسون توضیح داده می شوند.

۸-۲-۱ روش بازآزمایی

همان طور که از نامش پیداست، آزمون دو بار اجرا می شود. در این روش، آزمون را در دو نوبت زمانی متفاوت (برای مثال بعد از یک هفته، یک ماه یا بیشتر)، به گروه واحدی از آزمون شوندگان می دهند و ضریب همبستگی بین نمره های حاصل از دو بار اجرای آزمون محاسبه می شود. ضریب همبستگی^۱ بین نمره های حاصل از دو بار اجرای آزمون، ضریب اعتبار یا ثبات^۲ آزمون است. چنانچه نمره مشاهده شده ای که هریک از آزمودنی ها در زمان دوم دریافت می کنند، دقیقاً برابر نمره آن ها در بار اول آزمون باشد، اعتبار کامل است. به عبارت دیگر، اعتبار برآورد شده، برابر یک است (آلن و یین؛ ترجمه دلاور، ۱۳۹۲).

این روش در مورد اندازه‌گیری صفات و ویژگی‌هایی مناسب است که اغلب در طول زمان تغییر نمی‌کنند. هوش، یکی از ویژگی‌هایی است که معمولاً بعد از ۱۶ سالگی، در یک فاصله زمانی محدود (برای مثال چند ماه) تغییر چندانی نمی‌کند. اگر همبستگی بین نمره‌های دو بار اجرای آزمون، پایین باشد، به این معنا نیست که آزمودنی‌ها باهوش‌تر یا کم‌هوش‌تر شده‌اند، بلکه به معنای کم‌دقتی، بی‌دقتی یا خطای ابزار اندازه‌گیری (یعنی آزمون هوش مورد استفاده) است.

در استفاده از روش بازآزمایی، افزون‌بر ملاحظات کلی بالا، نکات دیگری را نیز که بر اعتبار تأثیر می‌گذارند، باید در نظر گرفت:

۱. اثر انتقال^۱: منظور از اثر انتقال آن است که آشنایی آزمودنی‌ها با سؤال‌های آزمون در بار اول، ممکن است بر نمره آزمون دوم آن‌ها تأثیر بگذارد. این امر هنگامی اتفاق می‌افتد که آزمودنی‌ها بعضی از پاسخ‌هایی را که در آزمایش قبل به سؤال‌ها داده‌اند، به یاد بیاورند و در آزمون بعدی نیز همان پاسخ‌ها را به سؤال‌ها بدهند، به ویژه اگر فاصله بین دو بار اجرای آزمون، کوتاه باشد. در آزمون‌های چالاک‌ی و بعضی آزمون‌های توانایی، اغلب این اتفاق رخ می‌دهد، زیرا افراد تمایل دارند رفتار خود را با اندازه‌گیری‌های مکرر اصلاح کنند.
۲. بعضی از متخصصان بر این باورند که چون نمونه سؤال‌های دو بار اجرای آزمون، کاملاً مشابه‌اند، اطلاع مفیدی به ما نمی‌دهند، زیرا اگر سؤال‌ها متفاوت بود، ممکن است نتیجه متفاوت باشد.
۳. اثر تمرین^۲: بعضی از مهارت‌ها، با تمرین افزایش می‌یابند. به این ترتیب هنگامی که آزمون برای بار دوم اجرا می‌شود، آزمودنی‌ها راحت‌تر به سؤال‌ها پاسخ داده یا فعالیت خواسته شده را انجام می‌دهد. همین امر باعث می‌شود نمره آزمون دوم، افزایش یابد و این مسئله هم ضریب اعتبار بازآزمایی را تحت تأثیر قرار می‌دهد.
۴. اگر فاصله زمانی بین دو بار اجرای آزمون کم باشد، به دلیل وجود حافظه، تمرین و تغییرات خلقی، موجب انتقال اثر می‌شود. اگر فاصله زمانی بین دو بار اجرای آزمون زیاد باشد، ممکن است خطای اندازه‌گیری، با تغییرات واقعی ایجاد شده در توانایی آزمودن‌شوندگان که در نتیجه یادگیری حاصل می‌شود، درهم می‌آمیزد.

۵. کاهش انگیزه^۱. اجرای مجدد یک آزمون برای تعیین اعتبار آن، مورد علاقه یادگیرندگان نیست و آن را کاری بیهوده می‌پندارند که باعث اتلاف وقت می‌شود. در نتیجه ممکن است نمره‌های آزمودنی‌ها در بار دوم، پایین‌تر از بار اول باشد.

۸-۲-۲ روش فرم‌های موازی و همتا

دو آزمون را در صورتی همتا^۲ گویند که هر دو برای اندازه‌گیری ویژگی واحدی در مورد جامعه واحدی تهیه شده باشند. در این روش، دو آزمون موازی یا همتا برای یک مطلب یا موضوع تهیه می‌کنند و آن‌ها را یکی پس از دیگری و بدون فاصله زمانی یا در فاصله زمانی کوتاهی، به گروه واحدی از آزمودنی‌گان می‌دهند.

دو فرم یک آزمون، در صورتی هم‌ارز یا موازی‌اند که میانگین و واریانس آن‌ها برابر باشند. اغلب در عمل امکان ندارد بتوان موازی بودن دو آزمون را تأیید کرد و بیشتر فرم‌های همتا، به جای فرم‌های موازی به کار برده می‌شوند (آلن^۳، ۲۰۰۸).

در روش فرم‌های همتا برخلاف روش بازآزمایی، واریانس خطای ناشی از نمره‌های مختلف سؤال‌ها مورد بررسی قرار می‌گیرد و این در صورتی است که هر دو آزمون، بدون فاصله زمانی و در یک جلسه اجرا شوند. در این مورد، برای اجتناب از تأثیر عامل تمرین، می‌توان ترتیب اجرای آزمون را در مورد گروه‌های مختلف تغییر داد. شاید مطلوب‌ترین روش برآورد اعتبار آزمون، اجرای فرم‌های همتای آزمون در زمان‌های مختلف باشد. بدین ترتیب هر دو نوع خطای اندازه‌گیری، یعنی خطای ناشی از نمونه‌های مختلف سؤال‌ها و خطای ناشی از نمونه‌گیری زمانی، در محاسبه ضریب اعتبار منظور می‌شوند (شریفی، ۱۳۹۳).

روش فرم‌های همتا نیز مثل روش بازآزمایی ایرادهایی دارد:

۱. از آنجا که شیوه پاسخ‌دهی به دو آزمون همتا مشابه است، آزمودنی در جواب دادن به سؤال‌های فرم اول آزمون، با اصولی آشنا می‌شود که عملکرد وی را در فرم دوم، تحت تأثیر قرار می‌دهد. آزمون‌های هوشی کتل، هرکدام دو فرم موازی‌اند.
۲. تهیه دو فرم همتا برای آزمون‌های پیشرفت تحصیلی، معمولاً سخت و دشوار است و چنین آزمون‌هایی اغلب در یک فرم واحد تهیه می‌شوند.

یادآوری ۱. ضریب همبستگی بین نمره‌های حاصل از دو بار اجرای آزمون، ضریب اعتبار یا ثبات آزمون است.

یادآوری ۲. منظور از اثر انتقال آن است که آشنایی آزمودنی‌ها با سؤال‌های آزمون در بار اول، ممکن است بر نمره آزمون دوم آن‌ها تأثیر بگذارد.

یادآوری ۳. دو فرم یک آزمون، در صورتی هم‌ارز یا موازی‌اند که میانگین و واریانس آن‌ها برابر باشد.

۸-۲-۳ روش‌های برآورد همسانی درونی

همسانی درونی^۱ به این معناست که سؤال‌ها یا مواد تشکیل‌دهنده آزمون، تا چه اندازه با همدیگر همخوانی و هماهنگی دارند. در روش‌های همسانی درونی، آزمون فقط یک‌بار روی گروه واحدی از آزمودنی‌ها اجرا می‌شود و سپس به کمک یکی از روش‌های تعیین همسانی درونی (مانند روش دو نیمه کردن، کودر-ریچاردسون و آلفای کرونباخ) ضریب اعتبار آن محاسبه خواهد شد.

۸-۲-۴ روش دو نیمه کردن

در روش دو نیمه کردن^۲، آزمون یک‌بار برای گروه نمونه‌ای از آزمودنی‌ها اجرا می‌شود. سپس پس از اجرای آزمون، آن به دو نیمه مساوی تقسیم می‌شود. برای دو نیمه کردن آزمون، روش‌های مختلفی وجود دارد. اگر تعداد سؤال‌های آزمون، زیاد و سطح دشواری سؤال‌ها در سراسر آزمون یکسان باشد و آزمون از نوع آزمون سرعت نباشد، به طوری که همه آزمودنی‌ها فرصت کافی برای پاسخ دادن به سؤال‌های انتهایی آزمون را داشته باشند، می‌توان همبستگی سؤال‌های نیمه اول را با نیمه دوم حساب کرد (شریفی، ۱۳۹۳). اما اگر سطح دشواری سؤال‌ها یکسان نباشد و سؤال‌های آخر آزمون، دشوارتر باشند، بهترین راه این است که همه سؤال‌های فرد را یک آزمون و همه سؤال‌های زوج را نیز آزمون دیگری به شمار آورید. ضریب همبستگی دو نیمه آزمون، شاخصی از اعتبار هریک از دو نیمه آن است.

اعتبار آزمون ۱۶۳

از آنجا که با افزایش تعداد سؤال‌های آزمون، ضریب اعتبار آن نیز افزایش می‌یابد، بنابراین ضریب همبستگی محاسبه‌شده، از ضریب همبستگی کل آزمون کمتر است. برای محاسبه ضریب اعتبار آزمون می‌توان از فرمول اسپیرمن - براون^۱ (فرمول ۸-۱) استفاده کرد. در این روش، برای محاسبه ضریب اعتبار کل آزمون، ضریب همبستگی بین نیمه‌ها به کمک فرمول اسپیرمن - براون قابل محاسبه است.

$$r_{11} = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1+r_{\frac{1}{2}\frac{1}{2}}}$$

فرمول ۸-۱

در این فرمول:

 r_{11} ضریب اعتبار کل آزمون $r_{\frac{1}{2}\frac{1}{2}}$ ضریب همبستگی نیمه‌های آزمون

برای مثال، اگر ضریب همبستگی بین دو نیمه آزمون، ۰/۵۵ باشد، ضریب اعتبار کل آزمون چنین خواهد بود:

$$r_{11} = \frac{2(0.55)}{1+0.55} = 0.71$$

همان‌گونه که ملاحظه می‌شود با استفاده از فرمول اسپیرمن - براون، ضریب اعتبار برآوردشده افزایش می‌یابد. هرچه ضریب همبستگی دو نیمه، به حد وسط (یعنی ۰/۵۰) نزدیک‌تر باشد، ضریب اعتبار کل بیشتر خواهد بود. اگر ضریب اعتبار دو نیمه، ۰/۱۰ باشد، ضریب اعتبار کل ۰/۱۸ خواهد بود؛ یعنی ضریب اعتبار کل، فقط ۰/۰۸ بیشتر از ضریب اعتبار دو نیمه است. اگر ضریب اعتبار دو نیمه، ۰/۹۰ باشد، ضریب اعتبار کل، ۰/۹۵ خواهد بود که در مقایسه با زمانی که ضریب اعتبار ۰/۵۵ است، افزایش خیلی کمی دارد. وقتی ضریب اعتبار دو نیمه، ۰/۵۵ باشد، ضریب اعتبار کل، ۰/۷۱ خواهد بود و ۰/۱۶ بیشتر از ضرب اعتبار دو نیمه است.

در صورتی که واریانس‌های نمره‌های نیمه‌ها مساوی نباشند یا نشانه‌های دیگری درباره موازی نبودن نیمه‌ها وجود داشته باشد، اعتبار کل آزمون را می‌توان با استفاده از ضریب آلفای کرونباخ برآورد کرد (آلن و ین؛ ترجمه دلاور، ۱۳۹۲). فرمول اسپیرمن - براون را با تغییر جزئی می‌توان پس از آنکه تعدادی سؤال به آن افزوده شد نیز برای محاسبه ضریب اعتبار یک آزمون مورد استفاده قرار داد. فرمول جدید، به قرار زیر است:

فرمول ۸-۲، محاسبه ضریب اعتبار آزمون پس از افزودن سؤال:

$$r'_{11} = \frac{kr_{11}}{1+(k-1)r_{11}}$$

در این فرمول:

r'_{11} ضریب اعتبار آزمون جدید

r_{11} ضریب اعتبار آزمون اولیه

k نسبت تعداد سؤال‌های آزمون جدید به تعداد سؤال‌های اولیه آزمون

۸-۲-۵ روش آلفای کرونباخ

یکی از روش‌هایی که برای ارزیابی اعتبار آزمون به‌کار می‌رود، استفاده از تحلیل واریانس^۱ است که به‌وسیله کرونباخ^۲ و همکاران وی طراحی شده است (لطف‌آبادی، ۱۳۹۳). روش آلفای کرونباخ^۳ نیز مبتنی بر همسانی درونی است. در این روش، اجزاء یا قسمت‌های آزمون، برای سنجش ضریب اعتبار آزمون به‌کار می‌روند. اگر قسمت‌های آزمون همان سؤال‌های آزمون باشند و سؤال‌ها به‌صورت درست (۱) و غلط (۰) تصحیح شوند، ضریب آلفا برابر خواهد بود با KR20. اگر آزمون از چندین خرده‌آزمون تشکیل شده باشد (مثل آزمون‌های هوشی کتل که چند خرده‌آزمون دارند)، برای آن‌ها در ضریب اعتبار کل، از فرمول ۸-۳ استفاده می‌شود.

$$r_{\alpha} = \frac{1}{j-1} \left(1 - \frac{\sum S^2}{S^2} \right) \quad \text{فرمول ۸-۳}$$

r_{α} ضریب اعتبار کل آزمون

j تعداد بخش‌های آزمون

S_j^2 واریانس نمره‌های هر بخش آزمون

اگر بخش‌های آزمون یا خرده‌آزمون‌هایی که از مجموع آن‌ها آزمون کلی تشکیل شده است، به‌طور جداگانه نمره‌گذاری شوند، در آن صورت ضریب آلفا مستلزم تصحیح سؤال‌های انفرادی به‌صورت درست و غلط نیست (سیف، ۱۳۹۶).

اعتبار آزمون ۱۶۵

مثال. آزمون هوشی کتل (فرم ب)، از چهار خرده‌مقیاس تشکیل شده‌است. فرض کنید واریانس این چهار خرده‌آزمون و واریانس کل، به شرح زیر باشد:

$$(S_1^2 = 6, S_2^2 = 4, S_3^2 = 5, S_4^2 = 6, S_T^2 = 25)$$

با توجه به اینکه $j=4$ است، ضریب آلفا به صورت زیر محاسبه می‌شود:

$$r_{\alpha} = \frac{4}{4-1} \left(1 - \frac{6+4+5+6}{25} \right)$$

$$r_{\alpha} = \frac{4}{3} \left(1 - \frac{21}{25} \right)$$

$$r_{\alpha} = 1.33(0.16) = 0.21$$

چنانچه واریانس‌های نمره‌های مشاهده‌شده دو نیمه آزمون مساوی باشند، فرمول اسپیرمن-براون و ضریب آلفا برابرند، اما اگر واریانس نمره‌های دو نیمه مساوی نباشد، فرمول اسپیرمن-براون و ضریب آلفا، اعتبار آزمون را کمتر از اندازه واقعی برآورد خواهند کرد (آلن و ین؛ ترجمه دلاور، ۱۳۹۲).

۸-۲-۶ روش کودر-ریچاردسون

سومین روش برآورد اعتبار درونی، روش کودر-ریچاردسون^۱ است. در این روش نیز آزمون یکبار اجرایی شود و همه سؤال‌های آزمون، مورد تجزیه و تحلیل قرار می‌گیرند. کودر و ریچاردسون، برای برآورد همسانی درونی، دو فرمول دارند: یکی KR20 و دیگری KR21. اینکه کدام فرمول، مورد استفاده قرار گیرد، به سطح دشواری سؤال‌های آزمون بستگی دارد. اگر همه سؤال‌های آزمون، دارای درجه دشواری مشابه باشند، KR21 مورد استفاده قرار می‌گیرد. دو فرمول، زمانی مساوی خواهند بود که دشواری همه سؤال‌ها مساوی باشد. در این صورت، KR21 کوچک‌تر از KR20 خواهد شد و ضریب اعتبار، کمتر از اندازه واقعی برآورد خواهد شد (آلن و ین؛ ترجمه دلاور، ۱۳۹۲). از لحاظ سطح دشواری، هرچه تفاوت میان سؤال‌های آزمون، بیشتر باشد، به همان نسبت، از ضریب اعتبار حاصل از فرمول KR21 کاسته می‌شود (سیف، ۱۳۹۶).

تنها زمانی می‌توان از فرمول‌های کودر-ریچاردسون استفاده کرد که پاسخ سؤال‌ها به صورت صفر و یک باشد. اگر پاسخ سؤال‌های آزمون، به صورت لیکرت باشد

(یعنی به صورت پیوستاری و از یک تا عدد خاصی مانند پنج یا بیشتر باشد)، نمی‌توان اعتبار را با فرمول کودر-ریچاردسون محاسبه کرد (لسی، ۲۰۱۴). در چنین شرایطی بهتر است از روش آلفای کرونباخ استفاده شود.

یکی از محاسن استفاده از روش کودر-ریچاردسون، امکان برآورد ضریب اعتبار آزمون با صرف وقت کمتری است (شریفی، ۱۳۹۳). جدول ۸-۱، پاسخ‌های یک گروه ۱۰ نفری از آزمودنی‌ها به یک آزمون عسوالی را نشان می‌دهد. در ردیف یک جدول، شماره افراد و در ستون یک، شماره سؤال‌ها و در خانه‌های جدول، پاسخ هر آزمودنی به هر یک از سؤال‌ها با نمره‌های ۱ و ۰ نشان داده شده است. نمره ۱ نشان می‌دهد که آزمودنی به سؤال موردنظر، پاسخ صحیح داده و نمره صفر، نشانگر پاسخ غلط آزمودنی به سؤال است.

جدول ۸-۱. داده‌های لازم برای استفاده از روش کودر-ریچاردسون در محاسبه ضریب اعتبار

شماره سؤال	۱	۲	۳	۴	۵	۶	نمره کل هر آزمودنی
شماره افراد							
۱	۱	۱	۱	۱	۱	۱	۶
۲	۱	۱	۱	۱	۱	۰	۵
۳	۰	۱	۱	۱	۱	۰	۴
۴	۱	۱	۱	۰	۱	۱	۵
۵	۱	۰	۱	۰	۰	۰	۲
۶	۰	۰	۰	۰	۰	۰	۰
۷	۱	۱	۰	۰	۰	۰	۲
۸	۱	۱	۰	۱	۰	۰	۳
۹	۱	۰	۱	۰	۱	۰	۳
۱۰	۰	۱	۱	۱	۱	۰	۴
تعداد کل آزمودنی‌هایی که به سؤال، پاسخ صحیح داده‌اند	۷	۷	۷	۵	۶	۲	۳۴
نسبت درست p	۰/۷	۰/۷	۰/۷	۰/۵	۰/۶	۰/۳۳	
نسبت غلط q	۰/۳	۰/۳	۰/۳	۰/۵	۰/۴	۰/۶۷	
pq	۰/۲۱	۰/۲۱	۰/۲۱	۰/۲۵	۰/۲۴	۰/۲۲	$\Sigma pq = 1/۳۴$

$$S^2 = 1/۶۸ \quad S = 1/۳ \quad \text{میانگین} = ۳۴ \div ۱۰ = ۳/۴$$

فرمول KR20 به صورت زیر است:

$$KR20 = r_{11} = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$$

فرمول ۴-۸

عناصر فرمول، عبارت‌اند از:

r_{11} ضریب اعتبار آزمون

n تعداد سؤال‌های آزمون

S^2 واریانس نمره‌های خام آزمون

P نسبت آزمودنی‌هایی که به سؤال، پاسخ درست داده‌اند.

Q نسبت آزمودنی‌هایی که به سؤال، پاسخ غلط داده‌اند.

$\sum pq$ حاصل جمع pq برای کل سؤال‌ها

$$KR20 = r_{11} = \frac{6}{6-1} \left(1 - \frac{1.34}{1.68} \right)$$

$$r_{11} = 1.2(1 - 0.80) = 0.24$$

با توجه به فرمول ۴-۸، ملاحظه می‌شود برای مثبت و معنادار بودن ضریب اعتبار آزمون باید واریانس نمره‌های خام S^2 به قدر کافی از مجموع واریانس سؤال‌های آزمون $\sum pq$ بزرگ‌تر باشد. اگر همبستگی بین سؤال‌ها زیاد باشد، گفته می‌شود بین آن‌ها کوواریانس^۱ (واریانس مشترک) وجود دارد. هرچه مقدار کوواریانس بزرگ‌تر باشد، اندازه $\sum pq$ کوچک‌تر می‌شود. چنانچه کوواریانس بین سؤال‌ها بالا باشد، می‌توان چنین فرض کرد که سؤال‌های آزمون، یک ویژگی کلی را اندازه می‌گیرند که در این صورت، ضریب اعتبار آزمون بالا خواهد بود. ضریب اعتبار آزمون، با تعداد سؤال‌های آن (K) نیز رابطه مستقیم دارد. هرچه تعداد سؤال‌ها بیشتر باشد، ضریب اعتبار آزمون بالاتر است (شریفی، ۱۳۹۳).

یادآوری. روش‌های برآورد اعتبار به شیوه همسانی درونی، عبارت‌اند از: آلفای کرونباخ، کودر-ریچاردسون و روش دوتیمه کردن.

۳-۸ خطای معیار اندازه‌گیری

خطای معیار اندازه‌گیری^۲ یعنی اینکه همه نمره‌های مشاهده‌شده (نه نمره‌های واقعی) یک

آزمودنی را در یک آزمون خاص یا آزمون‌های موازی که به صورت مکرر و مستقل اندازه‌گیری شده‌اند، در اختیار داشته باشیم و سپس انحراف استاندارد آن را محاسبه کنیم. به این انحراف استاندارد محاسبه شده، خطای معیار اندازه‌گیری گفته می‌شود. خطای معیار اندازه‌گیری وقتی اتفاق می‌افتد که ابزار اندازه‌گیری دقیق نیست. آزمون‌های ساخته شده در حوزه روان‌سنجی و اندازه‌گیری ویژگی‌های روانی، بسیار دقیق نیستند و اگر آزمون در یک فاصله زمانی کوتاه بارها روی یک فرد خاص اجرا شود، اغلب عدد ثابتی به دست نمی‌آید. این امر نشان می‌دهد که ابزار اندازه‌گیری دقیق نیست، زیرا اگر ابزار دقیق بود (مثل متر دیجیتالی)، باید هر بار همان عدد قبلی حاصل می‌شد. در ابزارهای دقیق اندازه‌گیری، خطای اندازه‌گیری وجود ندارد، بنابراین خطای معیار اندازه‌گیری وجود ندارد.

به طور معمول از آنجا که صفات روانی و ذهنی، مستقیماً قابل مشاهده و اندازه‌گیری نیستند و از رفتار آشکار استنباط می‌شوند، براساس نمره‌های مشاهده شده آزمودنی در مورد ویژگی روانی مورد اندازه‌گیری او قضاوت می‌شوند. همچنین بدیهی است که توانایی‌های فرد در طول یک زمان کوتاه، نسبتاً ثابت‌اند. اگر چند بار یک صفت اندازه‌گیری شود، مشخص خواهد شد که نمره واقعی فرد، پیرامون چه عددی قرار دارد. هدف اندازه‌گیری خطای معیار اندازه‌گیری هم این است که دقت اندازه‌گیری را افزایش دهد.

هرچند ضریب اعتبار، شاخص مفیدی از ثبات نمره‌ها در یک آزمون به دست می‌دهد، اما تصویر کاملی از مسئله ارائه نمی‌کند؛ یعنی به طور مستقیم میزان اشتباه موجود در نمره فرد را نشان نمی‌دهد. این اطلاعات، به کمک خطای معیار اندازه‌گیری به دست می‌آید. همان‌طور که در قسمت‌های گذشته بیان شد، واریانس نمره‌های مشاهده شده حاصل از اجرای آزمون، یعنی مقادیر (X) ، ممکن است برآمده از دو منبع، یعنی نمره‌های واقعی و نمره‌های اشتباه باشد.

به بیان دیگر، خطای معیار اندازه‌گیری نشان می‌دهد که نمره‌های فرد در یک آزمون، همان نیست و بنابراین نمره‌ای که یک فرد در یک آزمون به دست می‌آورد، فقط تخمینی از توانایی واقعی اوست.

اگر نمره مشاهده شده^۱ با X_o ، نمره واقعی^۲ با X_T و نمره خطا^۳ با X_e نمایش داده شود، تساوی زیر برقرار است:

اعتبار آزمون ۱۶۹

$$X_{ij} = X_i - X_{..}$$

واریانس نمره‌های مشاهده شده را نیز می‌توان با استفاده از فرمول زیر محاسبه کرد:

$$S_o^2 = S_e^2 + S_g^2$$

اگر اعتبار آزمون مشخص باشد، واریانس خطا را می‌توان به روش زیر به دست آورد:

$$S_e^2 = S_o^2 (1 - r_{11})$$

فرمول ۵-۸ برآورد خطای معیار اندازه‌گیری نیز به شرح زیر است:

$$SEM = S\sqrt{1 - r_{11}}$$

SEM خطای معیار اندازه‌گیری نمره‌های خام

S انحراف استاندارد نمره‌های خام

برای مثال، فرض کنید انحراف استاندارد و ضریب اعتبار یک آزمون به ترتیب برابر ۱۵ و ۰/۹۱ باشد. خطای معیار اندازه‌گیری آزمون برابر است با:

$$SEM = 15\sqrt{1 - 0.91}$$

$$15\sqrt{0.09}$$

$$(15)(0.3) = 4.5$$

۴-۸ فاصله اطمینان

همان‌طور که در بالا گفته شد، نمره‌های مشاهده شده یعنی نمره‌ای که فرد از یک آزمون خاص می‌گیرد، نمره واقعی او نیست و این نمره با مقداری خطا همراه است. خطای معیار اندازه‌گیری، به مشخص شدن مقدار خطا کمک می‌کند. فاصله اطمینان، آزمون دیگری است که به افزایش دقت کمک می‌کند. آنچه از فاصله اطمینان دریافت می‌شود این است که در واقع پی بردن به نمره واقعی فرد، هیچ‌گاه میسر نیست و تنها می‌توان حدود نمره واقعی را مشخص کرد. فرمول محاسبه فاصله اطمینان (فرمول ۶-۸) به شرح زیر است:

$$X \pm (Z)(SEM)$$

فرمول ۶-۸

X نمره مشاهده‌شده

Z نمره استاندارد

SEM خطای معیار اندازه‌گیری نمره‌های خام

خطای معیار اندازه‌گیری، دارای توزیع نرمال است، بنابراین با توجه به اندازه‌های سطح زیر منحنی نرمال و توزیع طبیعی Z می‌توان انتظار داشت نمره واقعی ۶۸ درصد آزمودنی‌هایی که نمره واحدی گرفته‌اند، در فاصله بین ± 1 خطای معیار اندازه‌گیری و نمره ۹۵ درصد آزمودنی‌ها در فاصله بین ± 2 خطای معیار اندازه‌گیری از نمره مشاهده شده باشد (برای کسب اطلاعات بیشتر، به کتاب آمار توصیفی تألیف امین‌پور، ۱۳۹۴ چاپ پیام نور مراجعه کنید).

براساس داده‌های بالا، فرض کنید یک آزمودنی در آزمون هوشی ریون، نمره ۱۲۰ گرفته باشد. انحراف استاندارد (معیار) آزمون ۱۵ و ضریب اعتبار آن، $r_{11} = 0.91$ است. حدود نمره واقعی او در سطح اطمینان ۹۵ درصد و ۹۹ درصد، به شرح زیر خواهد بود:

$$X \pm (Z)(SEM)$$

$$120 + (1/96)(4/5)$$

$$= 120 + 8/82$$

$$= 128/82$$

حد بالای فاصله اطمینان در سطح ۹۵ درصد

در آزمون‌های دو دامنه، نمره Z در سطح ۹۵ درصد، ۱/۹۶ است.

$$120 - (1/96)(4/5)$$

$$= 120 - 8/82$$

$$= 111/82$$

حد پایین فاصله اطمینان در سطح ۹۵ درصد

با ۹۵ درصد اطمینان می‌توان گفت نمره مشاهده‌شده آزمودنی، کمتر از ۱۱۱/۸۲ و بیشتر از ۱۲۸/۸۲ نخواهد بود.

$$120 - (2/58)(4/5)$$

$$= 120 + 11/61$$

$$= 131/61$$

حد بالای فاصله اطمینان در سطح ۹۹ درصد

در آزمون‌های دو دامنه، نمره Z در سطح ۹۹ درصد، ۲/۵۸ است.

$$120 - (2/58)(4/5)$$

$$= 120 - 11/61$$

$$= 108/39$$

حد پایین فاصله اطمینان در سطح ۹۹ درصد

با ۹۹ درصد اطمینان می‌توان گفت نمره مشاهده‌شده آزمودنی، کمتر از ۱۰۸/۳۹ و بیشتر از ۱۳۱/۶۱ نخواهد بود.

۵-۸ تفسیر خطای معیار اندازه‌گیری

ضریب اعتبار، بین صفر و یک قرار دارد. صفر یعنی آزمون، اصلاً پایا و قابل اعتماد نیست و یک یعنی اینکه آزمون، کاملاً پایا و قابل اعتماد است. در حوزه آزمون‌های روانی، معمولاً اعتبار آزمون بین صفر و یک قرار دارد و اعتبار آزمون، اغلب کامل نیست.

رابطه بین خطای معیار اندازه‌گیری و اعتبار، معکوس است. به بیان دیگر، خطای معیار اندازه‌گیری، برعکس اعتبار است. اگر خطای معیار، صفر باشد، یعنی آزمون پایاست و خطا ندارد. هرچه مقدار خطای معیار اندازه‌گیری بیشتر باشد، اعتبار آزمون کمتر است.

در تفسیر نمره آزمون‌ها باید به خطای معیار اندازه‌گیری توجه شود، زیرا اگر نمره آزمونی ناپایا باشد، دامنه‌های اطمینان برای نمره‌های واقعی، خیلی گسترده خواهند شد؛ به این معنا که نمره‌های مشاهده‌شده، نشانه‌های بسیار خوبی از نمره‌های واقعی نیستند. در آزمون‌های پایا انتظار می‌رود نمره‌های مشاهده‌شده، به نمره‌های واقعی نزدیک‌تر باشند.

۶-۸ عوامل مؤثر بر افزایش ضریب اعتبار آزمون

بر اساس نظریه‌های روان‌سنجی، روش‌هایی وجود دارند که با به کار بستن آن‌ها می‌توان اعتبار آزمون را افزایش داد. بعضی از این عوامل، عبارت‌اند از:

۱. افزایش تعداد سؤال‌ها. هرچه تعداد سؤال‌های آزمون بیشتر باشد، احتمال اینکه آزمون، معرف واقعی ویژگی مورد اندازه‌گیری باشد، بیشتر است.

۲. حذف سؤال‌های نامتجانس. سؤال‌های متجانس، با کل آزمون همبستگی بالایی دارند، اما همبستگی سؤال‌های نامتجانس با کل آزمون کم است. سؤال‌های خیلی آسان و خیلی دشوار، قدرت تشخیص بالایی ندارند و ضریب اعتبار آزمون را کاهش می‌دهند. در هر دو صورت، سؤال یا سؤال‌هایی که همبستگی کمی با آزمون دارند، باید از آزمون حذف شوند. این کار، اعتبار آزمون را افزایش می‌دهد.

۳. افزودن سؤال‌های دقیق. با افزودن تعدادی سؤال با سطح دشواری متوسط و قدرت تشخیص بالا به یک آزمون می‌توان ضریب اعتبار را افزایش داد.

یادآوری ۱. هدف اندازه‌گیری خطای معیار اندازه‌گیری هم این است که دقت اندازه‌گیری را افزایش دهد.

یادآوری ۲. رابطه بین خطای معیار اندازه‌گیری و اعتبار، معکوس است.

خلاصه فصل هشتم

هر ابزار اندازه‌گیری باید دو مشخصه داشته باشد: اعتبار و روایی.

اعتبار، به دقت اندازه‌گیری و ثبات و اعتبار یک وسیله اندازه‌گیری اشاره می‌کند. اعتبار، شرط روایی است، اما روایی برای اعتبار، ضروری نیست. روش‌های تعیین اعتبار، عبارت‌اند از: روش بازآزمایی: در این روش، آزمون در دو نوبت زمانی متفاوت، روی گروه واحدی از آزمودنی‌ها اجرا می‌شود. ضریب همبستگی بین نمره‌های حاصل از دو بار اجرای آزمون، ضریب اعتبار یا ثبات آزمون خواهد بود. در استفاده از روش بازآزمایی، اثر انتقال و کاهش انگیزه بر اعتبار تأثیر می‌گذارند.

روش فرم‌های موازی و هم‌تا. دو آزمون را در صورتی هم‌تا گویند که هر دو برای اندازه‌گیری ویژگی واحدی در مورد جامعه واحدی تهیه شده باشند.

روش‌های برآورد همسانی درونی عبارت‌اند از: روش دو نیمه کردن: در این روش آزمون یک‌بار برای گروه نمونه‌ای از آزمودنی‌ها اجرا می‌شود. سپس پس از اجرای آزمون، آن به دو نیمه مساوی تقسیم می‌شود. همبستگی بین نمره‌های سؤال‌های نیمه اول و نیمه دوم را اعتبار آزمون به روش دو نیمه کردن می‌نامند.

روش آلفای کرونباخ. روش آلفای کرونباخ نیز مبتنی بر همسانی درونی است. در این روش، اجزاء یا قسمت‌های آزمون، برای سنجش ضریب اعتبار آزمون به کار می‌روند.

روش کودر- ریچاردسون. در این روش نیز آزمون یکبار اجرا می‌شود و همه سؤالات آزمون، مورد تجزیه و تحلیل قرار می‌گیرند. تنها زمانی می‌توان از فرمول‌های کودر- ریچاردسون استفاده کرد که پاسخ سؤالات به صورت صفر و یک باشد.

خطای معیار اندازه‌گیری نشان می‌دهد که نمره‌های فرد در یک آزمون، همسان نیست و بنابراین نمره‌ای که یک فرد در یک آزمون به دست می‌آورد، فقط تخمینی از توانایی واقعی اوست؛ بنابراین در تفسیر نمره آزمون‌ها باید به خطای معیار اندازه‌گیری توجه شود، زیرا اگر نمره آزمونی ناپایا باشد، دامنه‌های اطمینان برای نمره‌های واقعی، خیلی گسترده خواهند شد؛ به این معنی که نمره‌های مشاهده شده، نشانه‌های بسیار خوبی از نمره‌های واقعی نیستند. بر اساس نظریه‌های روان‌سنجی، روش‌هایی وجود دارند که با به کار بستن آن‌ها می‌توان اعتبار آزمون را افزایش داد. بعضی از این عوامل، عبارت‌اند از: افزایش تعداد سؤالات، حذف سؤالات نامتجانس و افزودن سؤالات دقیق.

خودآزمایی چهارگزینه‌ای فصل هشتم

۱. در مورد اعتبار، کدام گزینه صحیح نیست؟
 - الف) معنای اعتبار، ثبات، شباهت و نزدیکی نمره‌های آزمون در طول زمان است.
 - ب) به همسانی درونی آزمون اشاره دارد.
 - ج) گاهی بالا بودن همسانی درونی آزمون، به بهای کاهش روایی انجام می‌گیرد.
 - د) روای بودن یک آزمون، مهم‌تر از داشتن اعتبار آزمون است.
۲. بر اساس کدام روش می‌توان اعتبار آزمون را مورد سنجش قرار داد؟
 - الف) میانگین
 - ب) انحراف استاندارد
 - ج) ضریب همبستگی
 - د) نمره استاندارد
۳. اگر همبستگی بین نمره‌های دو بار اجرای آزمون پایین باشد، نشان‌دهنده این است که آزمون ندارد.
 - الف) اعتبار
 - ب) روایی
 - ج) متجانس
 - د) همسانی

۴. اگر فاصله بین دو بار اجرای آزمون کم باشد، موجب ایجاد می‌شود.
 الف) تمرین (ب) انتقال
 ج) تکرار (د) کاهش انگیزه
۵. در مورد تهیه فرم‌های همتا، کدام گزینه نادرست است؟
 الف) دو فرم یک آزمون، در صورتی هم‌ارز یا مساوی‌اند که میانگین و واریانس آن‌ها برابر باشد.
 ب) بیشتر فرم‌های موازی، به جای فرم‌های همتا به کار برده می‌شوند.
 ج) دو فرم آزمون را زمانی همتا می‌نامند که هر دو برای اندازه‌گیری ویژگی واحدی تهیه شده باشند.
 د) هر دو آزمون، بدون فاصله زمانی و در یک جلسه اجرا شوند.
۶. در کدام روش به دست آوردن اعتبار، آزمون دو بار اجرا می‌شود؟
 الف) فرم‌های موازی (ب) آلفای کرونباخ
 ج) دو نیمه کردن (د) بازآزمایی
۷. کدام یک از روش‌های زیر، برای برآورد همسانی درونی آزمون به کار می‌رود؟
 الف) دو نیمه کردن (ب) بازآزمایی
 ج) فرم‌های همتا (د) فرم‌های ناهمتا
۸. کدام یک از روش‌های زیر، جزء روش‌های برآورد اعتبار به شیوه همسانی درونی است؟
 الف) فرم‌های موازی (ب) بازآزمایی
 ج) دو نیمه کردن (د) روایی
۹. کدام یک از روش‌های برآورد اعتبار، جزء روش‌های برآورد همسانی درونی نیست؟
 الف) دو نیمه کردن (ب) آلفای کرونباخ
 ج) کودر-ریچاردسون (د) فرم‌های موازی
۱۰. زمانی که اعتبار آزمون به روش دو نیمه کردن، برابر ۰/۷۰ باشد، اعتبار نهایی آزمون چند خواهد شد؟
 الف) ۰/۸۸ (ب) ۰/۸۲
 ج) ۰/۷۰ (د) ۰/۳۵

۱۱. در صورتی نتایج به دست آمده از فرمول‌های اسپیرمن براون و ضریب آلفا با هم برابر خواهند شد که
 الف) واریانس
 ب) تعداد نمونه
 ج) تعداد سؤال
 د) میانگین
۱۲. زمانی از فرمول کوردر ریچاردسون استفاده می‌شود که پاسخ سؤال‌ها به صورت باشد.

- الف) ۰ و ۱
 ب) ۱، ۲ و ۳
 ج) ۲، ۳، ۴ و ۵
 د) ۰، ۱ و ۲
۱۳. اگر انحراف استاندارد و ضریب اعتبار یک آزمون به ترتیب ۴ و ۰/۶۴ باشد، خطای معیار اندازه‌گیری چند خواهد شد؟
 الف) ۰/۸۰
 ب) ۰/۳۶
 ج) ۲/۴
 د) ۳/۲
۱۴. براساس توزیع نرمال، چند درصد نمره‌ها بین ± 2 انحراف استاندارد قرار می‌گیرند؟
 الف) ۶۸
 ب) ۸۶
 ج) ۹۵
 د) ۹۹

خودآزمایی تشریحی فصل هشتم

- از روش‌های تعیین اعتبار آزمون، روش بازآزمایی را با ارائه مثال تعریف کنید؟
- نکات مؤثر در اعتبار آزمون به روش بازآزمایی کدام‌اند؟
- روش‌های برآورد همسانی درونی برای برآورد اعتبار آزمون را نام ببرید و به اختصار شرح دهید؟
- ایرادهای فرم‌های همتا در مقایسه با روش‌های بازآزمایی کدام‌اند؟
- آیا براساس فرمول اسپیرمن براون، اعتبار کل آزمون افزایش پیدا خواهد کرد؟ بله - خیر
- محاسبه اعتبار آزمون به روش آلفای کرونباخ، مبتنی بر دو نیمه کردن است. صحیح - غلط
- اگر سطح دشواری همه سؤال‌های آزمون برابر باشد، از فرمول KR21 استفاده می‌شود. صحیح - غلط
- منظور از خطای معیار اندازه‌گیری چیست؟ با بیان مثال، آن را شرح دهید.

۹. در صورتی که میانگین و انحراف استاندارد نمره‌های یک آزمون به ترتیب ۱۵ و ۳ و ضریب اعتبار آن ۰/۶۴ باشد، خطای معیار اندازه‌گیری را حساب کنید و فاصله اطمینان آن را در سطح آلفا ۰/۰۵ به دست آورید؟
۱۰. اگر اعتبار دو نیمه آزمون، برابر با ۰/۵۰ باشد، اعتبار کل آزمون چند خواهد شد؟

فصل نهم

روایی آزمون

هدف کلی

تعریف روایی و انواع آن و عوامل تأثیرگذار بر روایی آزمون

هدف‌های یادگیری

از دانشجو انتظار می‌رود بعد از مطالعه این فصل بتواند:

۱. روایی را تعریف کند و ماهیت آن را توضیح دهد.

۲. انواع روایی را نام ببرد.

۳. روایی محتوایی و صوری را تعریف کند.

۴. روایی ملاکی و سازه را تعریف کند.

۵. روش‌های تعیین روایی سازه را نام ببرد.

۶. عوامل مؤثر بر روایی آزمون را بیان کند.

مقدمه

برای افرادی که برای بار اول با اصطلاحات اعتبار و روایی مواجه می‌شوند، ممکن است رابطه بین این دو گیج‌کننده باشد. برای به دست آوردن نتایج «روا»، اعتبار اندازه‌گیری لازم است، اما داشتن اعتبار بدون روایی نیز ممکن است؛ یعنی می‌توان نتایج هماهنگ و یکدستی در اختیار داشت که اطلاعات غلطی به دست می‌دهند یا به طور نامناسبی تفسیر می‌شوند. توضیح سیل شکل ۹-۱، این مفهوم را که «اعتبار لازم است، اما شرط کافی برای روایی نیست»، به تصویر می‌کشد (میلر، لین و گراندلند، ۲۰۰۹).

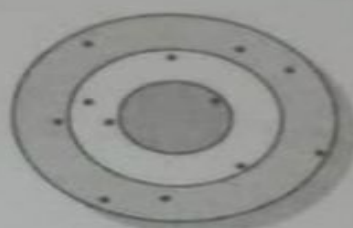
یک روش سنجش، افزون‌بر تدارک نتایج دارای روایی و اعتبار رضایت‌بخش، باید ملزومات عملی مشخصی نیز داشته و از نظر زمانی و اقتصادی به‌صرفه‌باشد. به‌آسانی اجرا و نمره‌گذاری‌شود و نتایجی را به‌بارآورد که به‌طور صحیح تفسیرشود و به‌وسیله کارکنان موجود در مدرسه قابل اجرا باشد (میلر، لین و گرانلند، ۲۰۰۹).



هدف ۳

شلیک وسط

(شلیک پایا و معتبر)



هدف ۲

شلیک پراکنده

(شلیک ناپایا و نامعتبر)



هدف ۱

شلیک سمت راست

(شلیک پایا اما نامعتبر)

شکل ۹-۱. توضیح رابطه اعتبار و روایی

۹-۱ ماهیت روایی

هنگام به‌کاربردن اصطلاح روایی در رابطه با آزمودن و سنجش، هشدارهای زیر را در نظر داشته‌باشید:

۱. روایی، به‌تناسب تفسیر و کاربرد نتایج یک روش سنجش برای گروه معینی از افراد اشاره‌دارد، نه فقط به‌تناسب آن برای روش سنجش. گاهی به‌منظور راحتی، از «روایی آزمون» صحبت می‌شود، اما صحیح‌تر است روایی تفسیر و استفاده‌ای که بایستی از نتایج به‌عمل‌آید، بررسی‌شود.
۲. روایی، یک موضوع درجه‌بندی است و سوگیری همه یا هیچ ندارد. در نتیجه باید از تفکر درباره نتایج سنجش به‌صورت «روا» یا «ناروا» اجتناب‌شود. به‌جای آن بهتر است روایی به‌صورت درجه‌ای مانند روایی بالا، متوسط یا پایین تفسیر‌شود.
۳. روایی همیشه برای بعضی مقاصد یا جمعیت خاصی از آزمودنی‌ها کاربرد دارد و قابل تفسیر است. هیچ سنجشی، برای تمام مقاصد روا نیست. برای مثال، آزمون ریاضی برای نشان‌دادن مهارت محاسباتی، روایی بالا، برای نشان‌دادن استدلال

ریاضی، روایی پایین و برای پیش‌بینی موفقیت در دوره‌های ریاضی در آینده، روایی متوسطی دارد و دراصل برای پیش‌بینی موفقیت در هنر یا موسیقی فاقد روایی است. وقتی مهارت محاسباتی نشان داده می‌شود، آزمون ریاضی ممکن است روایی بالایی برای دانش‌آموزان پایه سوم یا چهارم نیز داشته، اما برای دانش‌آموزان پایه دوم یا پنجم، دارای روایی پایینی باشد؛ بنابراین هنگام ارزیابی یا توصیف روایی، ضروری است تفسیر یا استفاده خاصی که از نتایج به عمل می‌آید، مد نظر قرار گیرد. روش‌های بسیاری برای گردآوری شواهد حمایتی یا به چالش کشیدن روایی یک تفسیر یا بهره‌برداری از نتایج سنجش وجود دارد.

یادآوری. هیچ سنجشی برای تمام مقاصد روا نیست.

۹-۲ تعریف روایی

روایی به لحاظ لغوی، از روابودن می‌آید. روابودن یعنی جایز بودن، یعنی آیا جایز است از یک ابزار خاص برای اندازه‌گیری یک صفت یا ویژگی خاص استفاده شود؟ در صورتی رواست که آزمونی برای اندازه‌گیری صفتی به کار گرفته شود که آزمون، برای اندازه‌گیری آن صفت ساخته شده باشد (ولوتن و شوویرس، ۲۰۰۵). به بیان دیگر، آزمونی دارای روایی است که برای اندازه‌گیری آنچه مورد نظر است، مناسب باشد. برای مثال، یک آزمون شخصیت، در صورتی برای اندازه‌گیری شخصیت رواست که فقط شخصیت را اندازه بگیرد، نه چیز دیگری را.

روایی اندازه‌گیری به معنای آن است که اندازه‌های به دست آمده در یک تحقیق، تا چه اندازه معرف وضعیت و اندازه‌های واقعی همان پدیده‌ای است که اندازه‌گیری آن مورد نظر بوده است؛ بنابراین روایی آزمون، به متناسب بودن استنباط‌هایی بستگی دارد که براساس نتایج آزمون به عمل می‌آید. روایی، هم می‌تواند معنایی بسیار عینی داشته باشد و هم تا حدودی معنای ذهنی^۱ (راجر^۲، ۱۹۹۹).

۹-۳ انواع روایی

روایی می‌تواند به روش‌های گوناگون ارزشیابی شود. این عمل، به نوع آزمون و کاربرد آن بستگی دارد. سه نوع روایی وجود دارد که عبارت‌اند از: روایی محتوایی^۱، روایی ملاکی^۲ و روایی سازه^۳. روایی ملاکی و سازه، به محاسبه آماری نیاز دارند، ولی روایی محتوایی نیازمند محاسبه آماری نیست (آلن و ین؛ ترجمه دلاور، ۱۳۹۲).

هرکدام از روایی‌های نام‌برده، زیرطبقه‌هایی دارند: روایی هم‌زمان^۴ و روایی پیش‌بینی^۵، زیرطبقه روایی ملاکی، روایی همگرا^۶ و روایی تمیزی^۷ نیز شواهدی از روایی سازه‌اند.

تا دهه ۱۹۶۰، مفهوم روایی سستی چنین بود: آیا همان چیزی را که قصد داریم، اندازه‌می‌گیریم؟ با این تصور که یک ملاک عینی مکنون برای اندازه‌گیری وجود دارد (مانند تصور از اندازه یک متر در تمام دنیا). در اصل ممکن است محقق، قادر به استدلال اینکه ابزار اندازه‌گیری‌اش چه چیزی را اندازه‌می‌گیرد یا چه ارتباطی بین مسئله مورد اندازه‌گیری و سایر متغیرها وجود دارد، نباشد. در زمینه گسترده‌ای مانند هوش، هوش چیزی بود که اندازه‌گیری آن، به وسیله ابزار اندازه‌گیری مناسب مانند آزمون‌ها صورت می‌گرفت، ولی آیا اگر ساینز کفش، با مهارت رهبری همبستگی داشت، ملاک خوبی برای رهبری بود؟ در دهه ۱۹۷۰، انجمن روان‌شناسی آمریکا استانداردهایی معرفی کرد. در این استانداردها برای تأیید روایی، به لزوم استنباط مناسب از نتایج به دست آمده تأکید شده بود و در نتیجه استدلال کردن به دلیل استنباط از نتایج، مهم تلقی شد. روایی سازه، مهم‌ترین مؤلفه مناسب برای روایی آزمون‌های تحصیلی شد که در سایر حوزه‌ها اهمیت کمتری داشت. بدون شک تفکر در مورد علت استفاده از موارد خاصی به عنوان ملاک انتخاب دانش‌آموزان، مهم‌تر از قبول بدون تفکر عمیق این موارد است.

۹-۳-۱ روایی محتوایی

به زبان ساده، وقتی از روایی محتوایی صحبت می‌شود، این موضوع به ذهن می‌رسد که سؤال‌هایی که از محتوای موضوع تهیه شده، تا چه اندازه معرف و نمونه کامل همه محتوا و هدف‌های موضوع مورد نظر است. یک آزمون روا، معرفتی از همه حیطه مهارت‌ها، توانایی‌ها، درک مفاهیم و دیگر رفتارهایی است که قرار است به وسیله آزمون، اندازه‌گیری شود. هر چه آزمون، از این لحاظ کامل‌تر باشد، دارای روایی بیشتری است. برای مثال، آزمونی که برای اندازه‌گیری اطلاعات دانشجویان از این کتاب تهیه می‌شود، در صورتی دارای روایی محتوایی است که:

- از تمام فصل‌های آن، سؤال طرح شده باشد.
- تعداد سؤال‌های طرح شده از هر فصل، آن قدر زیاد باشد که تمام موضوعات آن فصل را پوشش دهد.
- سؤال‌های طرح شده، همه طبقات مختلف حیطه شناختی (دانش، فهمیدن، به کار بستن و غیره) را دربرگیرد. به بیان دقیق‌تر، هم سؤال‌های ساده طرح شود و هم سؤال‌های متوسط و دشوار.
- به بیان دیگر، برای پی بردن به روایی محتوایی لازم است سه مسئله، مورد ارزیابی قرار گیرد:

۱. آیا ماده‌ها یا سؤال‌های آزمون، مناسب‌اند؟

۲. آیا نمونه سؤال‌ها کامل‌اند؟

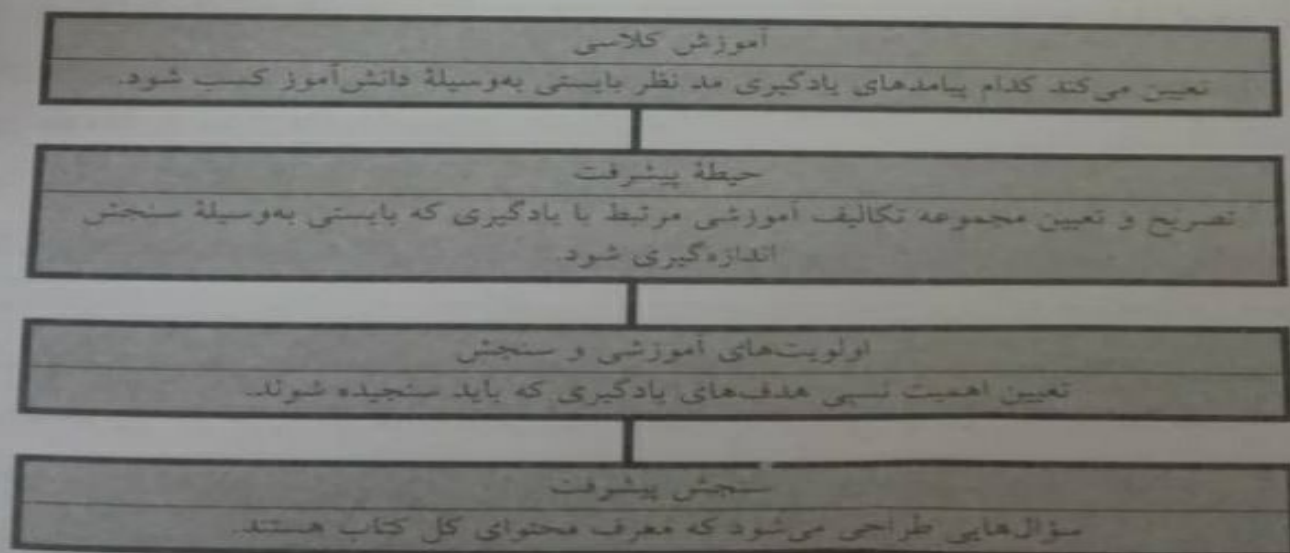
۳. آیا سؤال‌ها موضوع آزمون را به دقت می‌سنجند؟

سؤال‌ها باید به دقت، مشخص و به صورت عملیاتی تعریف شوند. سؤال‌های آزمون باید بر اساس هدف‌های رفتاری و محتوای مورد اندازه‌گیری تدوین شوند. اگر آزمون به نحوی طرح ریزی شده باشد، پاسخ‌های آزمودنی به سؤال‌های آن (یعنی نمره‌ای که آزمودنی از آزمون می‌گیرد)، معرف توانایی و مهارت او در حیطه‌ای است که سؤال‌های آزمون از آن نمونه‌گیری شده است. روایی محتوایی اغلب در آزمون‌های پیشرفت تحصیلی^۱ مورد استفاده قرار می‌گیرد. برای تحلیل روایی محتوایی آزمون پیشرفت تحصیلی که در مورد یک درس ساخته شده است، ابتدا باید جدول مشخصات

آزمونی که شامل محتوا و هدف‌های آموزشی است تهیه‌شود. ملاحظات محتوایی در سنجش پیشرفت کلاسی، در شکل ۹-۲ نشان داده شده‌است.

۹-۳-۲ روایی صوری

روایی صوری^۱ یکی از مشتقات روایی محتوایی است (سیف، ۱۳۹۶). این مفهوم در اصل، روایی به معنی فنی کلمه نیست، زیرا به آنچه آزمون در ظاهر می‌سنجد، اشاره دارد، نه آنچه آزمون در واقع می‌سنجد. روایی صوری گاهی روایی ذهنی^۲ نامیده می‌شود و هنگامی این روایی، مورد توجه قرار می‌گیرد که فردی پس از بررسی یک آزمون، نتیجه بگیرد که این آزمون، صفت مورد نظر را اندازه‌گیری می‌کند. فردی که آزمون را بررسی می‌کند، می‌تواند یک آزمودنی یا یک متخصص باشد. به عبارت دیگر، روایی صوری بیانگر این نکته است که آزمون در نظر افراد غیرمتخصص از قبیل آزمودنی‌ها و همچنین کارکنان اجرایی که درباره استفاده از آزمون تصمیم می‌گیرند، تا چه اندازه «به ظاهر معتبر» است (آناستازی، ۱۳۸۶). در واقع روایی صوری به این موضوع اشاره دارد که سؤال‌های آزمون، در ظاهر تا چه اندازه شبیه موضوعی هستند که برای اندازه‌گیری آن تهیه شده‌اند.



شکل ۹-۲. ملاحظات محتوایی در سنجش پیشرفت کلاسی

روایی صوری، گاهی لازم و گاهی هم غیر ضروری است. زمانی که می‌خواهیم آزمودنی برای پاسخ‌دهی به سؤال‌های آزمون انگیزه داشته باشد و بیشترین توانش را برای پاسخ‌دهی به کارگیرد، روایی صوری لازم است. برای مثال یک بانک خصوصی می‌خواهد افرادی را استخدام کند و یکی از آزمون‌های مورد نظر، محاسبه عددی است. اگر سؤال‌های آزمون، درباره وام و بهره بانکی باشد، انگیزه آزمودنی برای پاسخ‌دهی، بیشتر از زمانی است که سؤال‌ها مربوط به نرخ زاد و ولد باشد.

۳-۳-۹ روایی منطقی

روایی منطقی^۱ یا نمونه‌گیری، شکل پیچیده‌تری از روایی صوری است. این روش، شامل است بر تعریف دقیقی از حیطه رفتار، دانش نگارش، سایر حیطه‌های مورد اندازه‌گیری پرسش‌نامه و طرح منطقی سؤال‌هایی که همه بخش‌های عمده حیطه را می‌پوشاند. لازمه و شرط اعتبار یک آزمون از نظر محتوایی این است که معرف کامل حوزه معنای مفهوم اصلی مورد مطالعه باشد. به عبارت دیگر نمونه‌گیری دقیقی از حوزه‌های مختلف مفهوم مورد نظر صورت گرفته و براساس آن، ابزار اندازه‌گیری طراحی شده باشد. در این مرحله، افزون بر ارزیابی مربوط بودن آزمون‌ها (روایی صوری) به بررسی این نیز می‌پردازد که آیا دامنه تغییرات مفهوم یا مفاهیم مورد نظر، تحت پوشش قرار می‌گیرند. برای مثال گزارش‌های بخش حوادث روزنامه‌ها در مورد جرائم و آمار پلیس، هر دو دارای روایی صوری‌اند، اما اعتبار محتوایی گزارش‌های بخش حوادث روزنامه‌ها، از روایی محتوایی کمتری برخوردارند، زیرا معمولاً گزارش‌های روزنامه‌ها از بسیاری از جرائم غیر محسوس، چشم‌پوشی می‌کنند؛ بنابراین جامعیت لازم را ندارند. پژوهشگر برای بررسی روایی محتوایی یک پرسش‌نامه یا قرم مصاحبه، لازم است به گزاره‌های پژوهش (هدف‌ها و فرضیه‌ها) مراجعه کرده، سؤال‌های مورد نیاز را براساس هدف‌ها یا فرضیه‌ها و در نهایت تعاریف عملیاتی متغیرها تدوین کند. ارزیابی روایی محتوایی داده‌های عینی، چندان مشکل نیست، اما هر قدر مفاهیم مورد نظر انتزاعی‌تر می‌شوند، تعیین روایی محتوایی آن نیز مشکل‌تر می‌شود، زیرا حوزه محتوای چنین مفاهیمی اغلب مورد توافق همگان نیست. از طرف دیگر با قرض مشخص بودن

کامل حوزه محتوا، باز هم طراحی مجموعه‌ای از سؤال‌ها که برای نشان دادن هر یک از بخش‌های حوزه محتوا به اندازه کافی بزرگ باشد، بسیار دشوار است.

یادآوری ۱. سه نوع روایی وجود دارد که عبارت‌اند از: روایی محتوایی، روایی ملاکی و روایی سازه.

یادآوری ۲. روایی هم‌زمان و روایی پیش‌بینی، زیرطبقه روایی ملاکی و روایی همگرا و روایی تمیزی نیز شواهدی از روایی سازه‌اند.

یادآوری ۳. روایی محتوایی اغلب در آزمون‌های پیشرفت تحصیلی مورد استفاده قرار می‌گیرد.

یادآوری ۴. روایی صوری گاهی روایی ذهنی نامیده می‌شود.

یادآوری ۵. روایی منطقی یا نمونه‌گیری، شکل پیچیده‌تری از روایی صوری است.

۹-۳-۴ روایی ملاکی

منظور از روایی ملاکی، مقایسه نمره‌های آزمون یا مقیاس با یک یا چند ملاک یا معیار است و بدین ترتیب آزمون یا مقیاسی معتبر است که نمره‌های آن‌ها، با ملاک‌ها یا معیارهای مورد نظر، همبستگی بیشتری دارد. هنگامی که روایی ملاکی توجه می‌کنیم که در صدد شناسایی برخی از معیارهای قابل مشاهده به وسیله یک آزمون یا مقیاس هستیم. در روایی ملاکی، معمولاً با دو آزمون سروکار داریم. این دو آزمون، یا بلافاصله بعد از همدیگر اجرا می‌شوند، یا بین اجرای دو آزمون، فاصله زمانی وجود دارد. اگر آزمون‌ها بلافاصله بعد از یکدیگر اجرا شوند، هدف، جایگزین ساختن یکی به جای دیگری است. به این نوع روایی، روایی هم‌زمان گفته می‌شود. اگر نمره‌های آزمون دوم، پس از گذشت یک فاصله زمانی از آزمون اول گردآوری شود، هدف، پیش‌بینی است و به این نوع روایی، روایی پیش‌بینی گفته می‌شود.

به عبارت دیگر روایی ملاکی، به رابطه یا همبستگی میان نمره‌های آزمون و نمره‌های ابزار اندازه‌گیری دیگری که معرف یک ملاک از پیش تعیین شده است، مربوط می‌شود (لطف‌آبادی، ۱۳۹۳). به آزمون اول، آزمون پیش‌بین و به آزمون دوم، آزمون ملاک گفته می‌شود.

۹-۳-۵ روایی پیش‌بینی

روایی پیش‌بینی، عبارت است از بررسی رابطه نمره‌ها و نتایج حاصل از مقیاس آزمون با عملکرد و رفتار افراد در زمان آینده. در این حالت، این سؤال مطرح می‌شود که آیا ابزار اندازه‌گیری قادر است آنچه را که ادعا می‌کند، پیش‌بینی کند. برای مثال یکی از دلایل اجرای کنکور برای داوطلبان شرکت در دانشگاه، گزینش دانشجویان در رشته‌هایی است که بتوانند با توجه به استعدادشان، دوره آموزشی موردنظر را با موفقیت طی کنند؛ بنابراین اگر بین نمره‌های قبولی کنکور در یک رشته خاص و شاخص پیشرفت تحصیلی دانشجویان آن رشته (که معمولاً از معدل استفاده می‌شود)، همبستگی وجود داشته باشد، می‌توان گفت آزمون یادشده، روایی پیش‌بین بالایی دارد. همچنین سازمان‌هایی که برای استخدام پرسنل کارآمد، اقدام به گزینش و آزمون می‌کنند، از تعداد آزمون‌ها و مصاحبه‌ها بهره می‌برند و هدف آن‌ها جذب بهترین پرسنل برای افزایش کارایی سازمان است؛ بنابراین در صورت عدم روایی آزمون‌ها و مصاحبه‌ها، قادر به جذب بهترین پرسنل نخواهند بود.

به عبارت دیگر، یک آزمون هنگامی دارای روایی پیش‌بین است که نمره‌های آن، با معیاری که در آینده اندازه‌گیری می‌شود، همبستگی داشته باشد. به همین دلیل، ضریب همبستگی بین نمره‌های پیش‌بینی‌کننده و نمره‌های پیش‌بینی‌شده را ضریب روایی پیش‌بین می‌نامند. همان‌طور که اشاره شد، آزمون‌های ورودی دانشگاه‌ها، به منظور پیش‌بینی احتمال موفقیت تحصیلی داوطلبان در دانشگاه‌ها به کار می‌روند. در این مثال، سؤال‌های آزمون‌ها، متغیر پیش‌بین و نمره‌های دانشگاهی پذیرفته‌شدگان، متغیر ملاک‌اند. در یک آزمون استخدامی، سؤال‌های آزمونی که برای گزینش داوطلبان اجرایی‌شود، متغیر پیش‌بین و موفقیت و کارایی انتخاب‌شدگان در شغل آینده، متغیر ملاک است. در برخی مواقع، برای مثال هنگامی که تعداد کمی از افراد را می‌توان برای شغل معینی انتخاب کرد یا در مواقعی که هزینه‌های آموزشی در یک مورد معین، بسیار سنگین است و باید افراد مناسبی انتخاب شوند، استفاده از روایی پیش‌بینی، اهمیت زیادی پیدا می‌کند.

ضریب روایی پیش‌بین، به وسیله اجرای آزمون برای همه افراد موردنظر، انتظار برای یک مدت معقول، گردآوری نمره‌های ملاک و محاسبه ضریب روایی به دست می‌آید. به طور نمونه در مورد مثال قبل، روایی پیش‌بین برای یک آزمون

استخدام، به وسیله سنجش هریک از متقاضیان با کمک آزمون، استخدام عده‌ای از آنها برای چند هفته یا چند ماه تا زمانی که ملاک (عملکرد شغلی) را بتوان به‌طور پایا و منطقی ارزشیابی کرد، ادامه می‌یابد. محاسبه ضریب روایی، بین نمره‌های پیش‌بینی‌کننده (آزمون) و ملاک (عملکرد شغلی) برقرار خواهد شد. این شیوه درباره نحوه پیش‌بینی آزمون از رفتار آتی، اطلاعات خوبی به دست می‌دهد، اما این عمل، وقت‌گیر و مستلزم هزینه زیادی است (آلن و ین؛ ترجمه دلاور، ۱۳۹۲).

۹-۳-۶ روایی هم‌زمان

یکی دیگر از انواع روایی ملاکی، روایی هم‌زمان است. این نوع روایی، رابطه هم‌زمان بین آزمون و ملاک را مشخص می‌کند. در این نوع روایی، به جای مشخص شدن رابطه بین دو آزمون در یک فاصله زمانی، رابطه بین دو آزمون به‌طور هم‌زمان تعیین می‌شود. روایی هم‌زمان، در مورد آزمون‌هایی مناسب‌تر است که برای تشخیص وضع فعلی افراد به کار می‌روند، نه به منظور پیش‌بینی رویدادهای آینده. تمایز روایی هم‌زمان و پیش‌بین، در دو سؤال زیر آشکار است: «آیا احمد، فرد افسرده‌ای است؟» (روایی هم‌زمان). «آیا ممکن است علی در آینده، فرد افسرده‌ای شود؟» (روایی پیش‌بین).

با توجه به اینکه ملاک مربوط به روایی هم‌زمان، همیشه در زمان آزمایش با آزمون، در دسترس است، این سؤال پیش می‌آید که «پس در این مورد، نقش آزمون چیست؟». روایی هم‌زمان در مواقعی به کار می‌رود که قرار است یک اندازه‌گیری جدید، جایگزین یک اندازه‌گیری معمول و مورداستفاده در گذشته یا جایگزین مقیاسی شود که اجرای آن دشوار است. به عبارت دیگر، این قبیل آزمون‌ها جانشین‌های ساده، سریع و کم‌هزینه‌ای برای داده‌های ملاک محسوب می‌شوند. برای مثال، آزمون چندوجهی شخصیت مینه سوتا^۱ MMPI، یک آزمون بلند شخصیت است (۵۶۷ سؤال دارد). درمانگر می‌تواند فرم کوتاه و بلند آزمون را به تمام مراجعه‌کنندگان بدهد. در صورتی که نتایج دو آزمون، مشابه و قابل اعتماد بود، فرم کوتاه را جایگزین فرم بلند کند (آناستازی؛ ترجمه براهنی ۱۳۸۶). استفاده از فرم کوتاه، هم به لحاظ زمانی (آزمودنی، در زمان کوتاهی به سؤال‌ها پاسخ می‌دهد و درمانگر در مدت زمان کمی، پاسخ‌نامه را تصحیح می‌کند) و هم به لحاظ هزینه‌ای که درمان‌جو پرداخت می‌کند، به صرفه است.

اصل اول. ضریب روایی هم‌زمان، معمولاً بالاتر از ضریب روایی پیش‌بین است. با این حال، این بدان معنا نیست که آزمونی که ضریب روایی بالاتری دارد، همیشه برای یک هدف خاص مناسب‌تر است.

اصل دوم. تغییرپذیری گروه، بر اندازه ضریب روایی تأثیر می‌گذارد. ضریب روایی گروه‌های ناهمگن، بیشتر از ضریب روایی گروه‌های همگن است.

۹-۴ ضرایب روایی ملاکی

رابطه بین آزمون و متغیر ملاک، به صورت ضریب همبستگی بیان می‌شود که آن را ضریب روایی می‌نامند. برای تعیین ضریب روایی ملاکی آزمون، کافی است بین نمره‌های آزمون گروه نمونه‌ای از آزمودنی‌ها و نمره‌های ملاک آنان، همبستگی محاسبه شود. هرچه ضریب همبستگی بزرگ‌تر باشد، آزمون دارای روایی بیشتری خواهد بود و هرچه این آزمون کوچک‌تر باشد، آزمون موردنظر، روایی کمتری خواهد داشت.

البته همه ضرایب روایی، ارزش یکسان ندارند و برای تعیین میزان معناداری این ضرایب، قاعده‌ای قطعی وجود ندارد. در عمل، به تدریج ممکن است ضریب روایی یک آزمون، از ۰/۶۰ بالاتر باشد؛ بنابراین ضرایب روایی بین ۰/۳۰ و ۰/۴۰، ضرایب به نسبت بالایی به شمار می‌روند. برای دانستن اینکه رابطه بین دو آزمون، تصادفی است یا خیر، از آزمون معناداری آماری استفاده می‌شود. اگر مقدار خطا کمتر از ۰/۰۵ باشد، رابطه معنادار است و می‌توان گفت آزمون، برای سنجش صفت موردنظر روایی دارد.

جدول ۹-۱. مقایسه انواع روایی

نوع روایی	پاسخ‌گویی به سؤال	پرسیدن این سؤال
روایی محتوا	سؤال‌ها را یا هدف‌ها جور کنید.	آیا سؤال‌های آزمون، جور هستند و هدف‌ها را اندازه می‌گیرند؟
روایی هم‌زمان	آزمون جدید را با یک ملاک مقبول، مرتبط سازید؛ برای مثال یک آزمون استاندارد که همان رفتار را می‌سنجد.	عملکرد در آزمون جدید، تا چه اندازه با عملکرد در آزمون استاندارد، همخوانی دارد؟
روایی پیش‌بین	نمره‌های آزمون جدید را با عملکردهای آینده، همبسته سازید.	آیا آزمون، عملکرد بعدی را پیش‌بینی می‌کند؟ (برای مثال موفقیت یا شکست در پایه بالاتر)

به یاد داشته باشید که روایی پیش‌بین، نیازمند یک فاصله زمانی است. آزمون، اجرایی‌شود و بعد از یک دوره زمانی، رفتاری اندازه‌گیری خواهد شد که آزمون، قصد پیش‌بینی آن را دارد.
برای مثال:

معدل دوره دبیرستان

آیا معدل دوره دبیرستان
می‌تواند پیشرفت دانشگاهی
را پیش‌بینی کند؟

پیشرفت تحصیلی در
دانشگاه

توجه داشته باشید که روایی هم‌زمان، نیازمند یک فاصله زمانی نیست. آزمونی اجرایی‌شود و رابطه‌اش با یک آزمون جاافتاده که همان رفتار را اندازه‌گیری می‌کند، تعیین خواهد شد.
برای مثال:

آزمون جدید استعداد
دانشگاهی

آیا آزمون جدید با آزمون
قدیم رابطه دارد؟

آزمون قدیم

یادآوری ۱. منظور از روایی ملاکی، مقایسه نمره‌های آزمون یا مقیاس با یک یا چند ملاک یا معیار است.

یادآوری ۲. ضریب همبستگی بین نمره‌های پیش‌بینی‌کننده و نمره‌های پیش‌بینی‌شده را ضریب روایی پیش‌بین می‌نامند.

۹-۵ روایی سازه

سازه، مفهومی انتزاعی است که به‌طور مستقیم قابل مشاهده نیست، ولی از نظریه‌ها قابل

استباط است؛ بنابراین مقیاس‌ها و آزمون‌هایی که برای سنجش سازه‌ها درست می‌شوند، به‌طور مستقیم سازه‌ها را نمی‌سنجند، بلکه عملکرد یا رفتارهایی را می‌سنجند که گمان می‌رود منعکس‌کننده سازه‌ها باشند. در صورتی یک ابزار سنجش، دارای روایی سازه یا توافقی است که نمره‌های حاصل از اجرای آن، به‌مفاهیم یا سازه‌ها و نظریه‌های موردنظر و مرتبط با آن کار، مربوط باشد. روایی سازه، در مورد آزمون‌هایی لازم است که به‌منظور سنجش استعدادهای کلی و پدیده‌های دارای ابعاد مختلف طراحی می‌شوند. برای مثال هوش، پدیده‌ای ذهنی است که دارای ابعاد مختلفی است، از قبیل: سرعت انتقال، عکس‌العمل متناسب، تشخیص روابط بین پدیده‌ها و میزان سازگاری. پس مجموعه‌ای از عوامل که یک عامل را می‌سازند، سازه نامیده می‌شود. به‌بیان دیگر، روایی سازه بر این مطلب تأکید می‌کند که آزمون تا چه اندازه سازه نظری یا صفت موردنظر را اندازه می‌گیرد. این سنجش، مستلزم سه اقدام اساسی است: در ابتدا سازنده آزمون باید صفت موردنظر را به‌دقت تحلیل کند. در مرحله بعد، چگونگی ارتباط صفت با متغیرهای دیگر را موردتوجه قرار دهد و سپس به‌واسطه آزمایش معلوم کند که آیا این روابط فرضی، درواقع وجود دارند یا نه؟ دراصل، روایی سازه با این موضوع سروکار دارد که یک آزمون خاص، تا چه اندازه سازه یا صفت خاصی را اندازه‌گیری می‌کند (آناستازی؛ ترجمه براهنی ۱۳۸۶). روایی سازه، به‌همسانی سؤال‌های یک آزمون یا پرسش‌نامه و سازه‌های نظری پذیرفته‌شده موضوع مورد مطالعه اشاره دارد (راجر، ۱۹۹۹). بعضی از صفات و ویژگی‌ها، قابل مشاهده نیستند و فقط می‌توان به‌وسیله نشانه‌های آن‌ها به آن صفات پی برد. هوش، شخصیت، اضطراب، افسردگی، عشق، کنجکاوی، تخیل و سرعت پاسخ‌دهی، نمونه‌هایی از سازه‌ها هستند. برای هر کدام از سازه‌های نام‌برده، نظریه‌های گوناگونی از سوی نظریه‌پردازان، مطرح شده است. آزمون‌های ساخته‌شده، برای دانستن این موضوع طراحی می‌شوند که یک فرد خاص، چه مقدار از یک صفت یا ویژگی خاص را دارد. بعضی صفات مانند هوش، با گذشت زمان تغییر می‌کنند، بنابراین آزمون باید این را انعکاس دهد. آزمون باید بین افراد توانمند و کم‌توان در یک صفت، تفکیک قائل شود.

روایی سازه، بیشتر از روایی محتوایی و روایی پیش‌بین، جنبه نظری دارد. یک آزمون در صورتی دارای روایی سازه است که نمره‌های حاصل از اجرای آن، به‌مفاهیم یا سازه‌های نظریه موردنظر مربوط باشند (کرونیاخ، ۱۹۷۰؛ به‌تقل از سیف، ۱۳۹۶).

۹-۶ روش‌های تعیین روایی سازه

روایی سازه، مستلزم جمع‌آوری شواهدی است تا براساس آن‌ها معنای آزمون روشن شود. با به‌دست‌آمدن هر رابطه، یک معنای اضافی برای آزمون پیدامی‌شود. پس از پاره‌ای مطالعات، معنای آزمون به‌تدریج شکل می‌گیرد؛ بنابراین روایی سازه، فرایندی است که شاید در طول زمان، دستخوش تغییر و تحول شود.

روندی که برای تعیین روایی سازه موردنظر طی می‌شود، به‌این صورت است:

۱. ابتدا سازه براساس مفاهیم نظریه تعریف می‌شود.
 ۲. برای اندازه‌گیری سازه، آزمون طراحی می‌شود.
 ۳. نتایج آزمون طراحی‌شده، با پیش‌بینی‌ها مقایسه می‌شود.
- به‌کمک روش‌های زیر می‌توان روایی سازه را تعیین کرد:

۹-۶-۱ شواهد همگرا و واگرا

اینکه ادعا شود آزمونی برای اندازه‌گیری یک ویژگی مناسب است، کفایت نمی‌کند. برای دانستن اینکه آزمونی روا است یا خیر، باید آزمون طراحی‌شده، با آزمون‌های دیگری که از قبل برای اندازه‌گیری آن ویژگی طراحی‌شده‌اند و روایی آن‌ها به‌اثبات رسیده و مشخص شده‌است که برای سنجش ویژگی موردنظر مناسب‌اند، به‌طور هم‌زمان روی تعدادی از افراد اجرا شود. اگر همبستگی دو آزمون با یکدیگر زیاد باشد، نشانگر این است که آزمون روا است. برای مثال یک آزمون جدید شخصیت باید با آزمون‌های موجود مانند آزمون MMPI همبستگی بالایی داشته‌باشد. هرگاه همبستگی آزمون با آزمون‌های مشابه، بسیار زیاد باشد، بدون داشتن مزایایی از نوع کوتاهی یا سهولت اجرا، در این صورت آزمون جدید جز دوباره‌کاری، فایده‌ای نخواهد داشت. به‌این روایی، روایی همگرا گفته می‌شود.

از طرف دیگر، اگر این آزمون با آزمون‌های دیگری که سازه دیگری (مانند هوش) را اندازه‌می‌گیرند، همبستگی نداشته‌باشد، به‌این روایی، روایی واگرا گفته می‌شود. اگر آزمون، با آزمون‌هایی که ویژگی دیگری را اندازه‌گیری می‌کنند همبستگی قابل ملاحظه‌ای داشته‌باشد، نشانه نقص آزمون است.

۹-۶-۲ تفکیک سنی

راه دیگر تعیین روایی سازه، توجه به تغییرات زمان است. اگر یکی از ویژگی‌های انسان، با گذشت زمان تغییر کند، بنابراین می‌بایست آزمون مورد نظر، این تغییرات را با گذشت زمان منعکس کند. برای مثال، آزمونی که مهارت‌های ارتباط کلامی را اندازه‌گیری می‌کند، باید با بزرگ شدن کودکان یا حضور آن‌ها در کلاس‌های مربوطه، نمره‌های بالاتری را نشان دهد.

۹-۶-۳ تحلیل عاملی

یکی از روش‌های بررسی روایی سازه، تحلیل عاملی^۱ است. تحلیل عاملی، تکنیکی بسیار پیشرفته و فنی است که برای تعیین بار عاملی هریک از سؤال‌ها یا عبارات اندازه‌گیری مانند پرسش‌نامه استفاده می‌شود. به‌طور کلی در تکنیک تحلیل عاملی، اگر بیشتر سؤال‌هایی که برای سنجش مؤلفه خاصی در نظر گرفته شده‌اند، در تبیین آن، بار عاملی بالایی (معمولاً بیش از ۱) داشته باشند، مؤلفه‌های مربوط، دارای روایی عاملی خواهند بود. تحلیل عاملی، نشان‌دهنده شیوه‌های ریاضی گوناگون برای تحلیل همبستگی‌های درونی بین مجموعه‌ای از متغیرها و تبیین این همبستگی‌ها برحسب تعداد معدودی از متغیرهاست که عامل نامیده می‌شوند. عامل، یک متغیر فرضی است که بر یک یا چند متغیر مورد مشاهده تأثیر می‌گذارد. به عبارت دیگر هدف تحلیل عاملی، استخراج تعداد معدودی عامل از میان تعداد زیادی متغیرهای هم‌پوش به گونه‌ای است که:

۱. بین همه متغیرها مشترک باشد.
۲. جانشین تعداد زیادی متغیر شود.
۳. این عوامل با هم، همبستگی و تداخل نداشته باشند.
۴. سازه‌ای روشن و بامعنا تبیین کنند.

۹-۶-۳ همسانی درونی

آخرین روش تعیین روایی سازه، همسانی درونی است. در همسانی درونی، از نمره کل

آزمون برای بررسی همسانی درونی استفاده می‌شود. ویژگی اصلی این روش، این فرضیه است که ملاک موردنظر، چیزی جز نمره‌های کل آزمون نیست. یکی از روش‌های بررسی روایی درونی، بررسی سؤال‌های آزمون از لحاظ توانایی تشخیص افراد قوی و ضعیف را از همدیگر است. اگر سؤال یا سؤال‌هایی، قادر به این کار نباشند، فاقد اعتبار تلقی می‌شود و باید حذف شوند یا مورد تجدیدنظر قرار گیرند. می‌توان ضریب همبستگی هر سؤال را با کل آزمون نیز حساب و فقط سؤال‌هایی را انتخاب کرد که همبستگی بالا و معناداری با کل آزمون دارند. یک آزمون ممکن است از خرده‌مقیاس‌هایی تشکیل شده باشد، بنابراین می‌توان همبستگی خرده‌مقیاس‌ها را با کل آزمون حساب کرد. خرده‌آزمون‌ها باید دارای همبستگی کمی با هم باشند، ولی همبستگی آن‌ها با کل آزمون، زیاد باشد (روش‌های بررسی همسانی درونی، در مبحث اعتبار، مورد بحث قرار گرفتند).

اصطلاحات روایی محتوایی، روایی پیش‌بینی و روایی سازه، ابتدا فقط برای آزمون‌های وابسته به هنجار به کار می‌رفتند، اما کاربرد آن‌ها در مورد آزمون‌های وابسته به ملاک نیز کاملاً امکان‌پذیر است (سیف، ۱۳۹۶).

یادآوری ۱. روایی سازه، با این موضوع سروکار دارد که یک آزمون خاص، تا چه اندازه سازه یا صفت خاصی را اندازه‌گیری می‌کند.

خلاصه فصل نهم

در این فصل، در مورد روایی و انواع آن مطالبی گفته شد. روایی یعنی اینکه آزمون چیزی را اندازه‌بگیرد که برای اندازه‌گیری آن ساخته شده است. روایی می‌تواند به روش‌های گوناگون ارزشیابی شود. این عمل، به نوع آزمون و کاربرد آن بستگی دارد. سه نوع روایی وجود دارد که عبارت‌اند از: روایی محتوایی، روایی ملاکی و روایی سازه. هر کدام از این روایی‌ها، طبقه‌هایی دارند: روایی هم‌زمان و روایی پیش‌بینی، زیرطبقه روایی ملاکی و روایی همگرا و روایی تمیزی نیز شواهدی از روایی سازه‌اند.

روایی محتوایی. در روایی محتوایی، این بحث مطرح می‌شود که محتوای موضوع تهیه شده، تا چه اندازه معرف و نمونه کامل همه محتوا و هدف‌های موضوع موردنظر است. این روایی، بیشتر در آزمون‌های پیشرفت تحصیلی مورد استفاده قرار می‌گیرد.

روایی صوری. این روایی، یکی از مشتقات روایی محتوایی است و به آنچه آزمون در ظاهر می‌ستجد، اشاره دارد و گاهی روایی ذهنی نامیده می‌شود. روایی منطقی یا نمونه‌گیری، شکل پیچیده‌تری از روایی صوری است. این روش، شامل است بر تعریف دقیقی از حیطة رفتار، دانش نگارش و سایر حیطه‌های مورد اندازه‌گیری پرسش‌نامه.

روایی ملاکی. منظور از روایی ملاکی، مقایسه نمره‌های آزمون یا مقیاس با یک یا چند ملاک یا معیار است و بدین ترتیب آزمون یا مقیاسی معتبر است که نمره‌های آزمون یا مقیاس، با ملاک‌ها یا معیارهای مورد نظر همبستگی بیشتری داشته باشد.

روایی پیش‌بینی. یک آزمون، هنگامی دارای روایی پیش‌بین است که نمره‌های آن، با معیاری که در آینده اندازه‌گیری می‌شود، همبستگی داشته باشد. به همین علت، ضریب همبستگی بین نمره‌های پیش‌بینی‌کننده و نمره‌های پیش‌بینی‌شده را ضریب روایی پیش‌بین می‌نامند.

روایی هم‌زمان. یکی دیگر از انواع روایی ملاکی، روایی هم‌زمان است. این نوع روایی، رابطه هم‌زمان بین آزمون و ملاک را مشخص می‌کند. در این نوع روایی، رابطه بین دو آزمون به‌طور هم‌زمان تعیین می‌شود.

روایی سازه. در صورتی ابزار، دارای روایی سازه یا توافقی است که نمره‌های حاصل از اجرای آن، به‌مفاهیم یا سازه‌ها و نظریه‌های مورد نظر و مرتبط با آن کار، مربوط باشد. روایی سازه، در مورد آزمون‌هایی لازم است که به‌منظور سنجش استعدادهای کلی و پدیده‌های دارای ابعاد مختلف طراحی می‌شوند. به‌کمک روش‌های شواهد همگرا و واگرا، تفکیک سنی و تحلیل عاملی می‌توان روایی سازه را تعیین کرد.

خودآزمایی چهارگزینه‌ای فصل نهم

۱. در مورد روایی آزمون، کدام گزینه صحیح است؟
 - الف) روایی آزمون را نمی‌توان مورد سنجش قرار داد.
 - ب) روایی آزمون را نمی‌توان تفسیر کرد.
 - ج) روایی سازه، روایی ذهنی نامیده می‌شود.
 - د) روایی منطقی، شکل پیچیده‌تری از روایی صوری است.

۲. برای محاسبه کدام روایی، نیاز به محاسبه آماری نیست؟
 الف) محتوایی
 ب) سازه
 ج) ملاکی
 د) هم‌زمان
۳. زیر طبقه‌های روایی ملاکی عبارت‌اند از:
 الف) همگرا و تمیزی
 ب) هم‌زمان و پیش‌بین
 ج) همگرا و سازه
 د) محتوا و همگرا
۴. کدام نوع روایی، روایی ذهنی نامیده می‌شود؟
 الف) صوری
 ب) محتوایی
 ج) سازه
 د) ملاکی
۵. گزارش‌های مربوط به بخش حوادث روزنامه‌ها در مورد جرائم و آمار، دارای کدام نوع روایی‌اند؟
 الف) صوری
 ب) سازه
 ج) ملاکی
 د) پیش‌بین
۶. در محاسبه کدام نوع از روایی، داشتن دو نوع آزمون لازم است؟
 الف) صوری
 ب) محتوایی
 ج) سازه
 د) ملاکی
۷. ضریب همبستگی بین نمره‌های قبولی کنکور در یک رشته خاص و شاخص پیشرفت تحصیلی دانشجویان آن رشته، نشان‌دهنده کدام نوع روایی است؟
 الف) سازه
 ب) پیش‌بین
 ج) محتوایی
 د) هم‌زمان
۸. ضریب همبستگی بین نمره‌های پیش‌بینی‌کننده و پیش‌بینی‌شده را ضریب می‌نامند.
 الف) روایی
 ب) اعتبار
 ج) هنجار
 د) همبستگی
۹. احتمال موفقیت تحصیلی داوطلبان در دانشگاه‌ها با توجه به آزمون ورودی دانشگاه‌ها، نشان‌دهنده کدام نوع روایی است؟
 الف) سازه
 ب) پیش‌بینی
 ج) هم‌زمان
 د) منطق

۱۰. کدام نوع روایی بیشتر از سایر روایی‌ها جنبه نظری دارد؟
 الف) محتوایی
 ب) پیش‌بینی
 ج) سازه
 د) ملاکی
۱۱. اگر آزمونی با آزمون‌های دیگری که سازه دیگری را اندازه‌می‌گیرند، همبستگی نداشته باشد، نشان‌دهنده کدام نوع روایی است؟
 الف) پیش‌بینی
 ب) منطقی
 ج) همگرا
 د) واگرا
۱۲. تحلیل عاملی، یکی از روش‌های به دست آوردن کدام روایی است؟
 الف) منطقی
 ب) محتوایی
 ج) ملاکی
 د) سازه

خودآزمایی تشریحی فصل نهم

۱. در مورد هشدارهای لازم درباره روایی آزمون، به دو مورد اشاره کنید؟
۲. انواع روایی آزمون را نام ببرید؟
۳. روایی محتوایی را با ارائه مثال تعریف کنید؟
۴. برای پی بردن به روایی محتوایی، کدام مسائل باید مورد ارزیابی قرار گیرند؟
۵. آیا هرچه مفاهیم مورد نظر آزمون، انتزاعی تر شوند، تعیین روایی محتوایی، آسان تر خواهد شد؟
۶. روایی ملاکی را تعریف کنید و زیرمجموعه‌های آن را نام ببرید؟
۷. روایی هم‌زمان را تعریف کنید و برای آن مثالی بیاورید؟
۸. ضریب روایی ملاکی را تعریف کنید و برای آن مثالی بیاورید؟
۹. آیا ضریب روایی هم‌زمان معمولاً بالاتر از ضریب روایی پیش‌بین است؟ بله - خیر
۱۰. منظور از روایی سازه چیست؟
۱۱. روش‌های مربوط به تعیین روایی سازه کدام‌اند؟
۱۲. منظور از روایی همگرا و واگرا چیست؟
۱۳. هدف تحلیل عاملی چیست؟
۱۴. عوامل مؤثر در روایی آزمون را نام ببرید؟