

استانداردهایی برای آزمون‌گیری آموزشی و روان‌شناختی

اثری از

انجمن تحقیقات آموزشی آمریکا

انجمن روان‌شناسی آمریکا

شورای ملی اندازه‌گیری در آموزش آمریکا

مترجمان

دکتر علی مقدم‌زاده (عضو هیأت علمی دانشگاه تهران)

دکتر کیوان صالحی (عضو هیأت علمی دانشگاه تهران)

مریم کاظمی تبار

مهدی قربانخانی



شماره انتشار ۴۰۶۱

شماره مسلسل ۹۹۵۸

انتشارات دانشگاه تهران

عنوان و نام پدیدآور	: استانداردهایی برای آزمون‌گیری آموزشی و روانشناختی / اثری از انجمن تحقیقات آموزشی آمریکا، انجمن روانشناسی آمریکا، شورای ملی اندازه‌گیری در آموزش آمریکا؛ ترجمه علی مقدم‌زاده... [و دیگران].
مشخصات نشر	: تهران: دانشگاه تهران، مؤسسه انتشارات، ۱۳۹۸.
مشخصات ظاهری	: ۴۲۰ص: مصور، جدول، نمودار.
فروست	: انتشارات دانشگاه تهران؛ شماره انتشار ۴۰۶۱.
شابک	: 978-964-03-7309-5
وضعیت فهرست‌نویسی	: فیپا
یادداشت	: عنوان اصلی: Standards For Educational And Psychological Testing.
یادداشت	: ترجمه علی مقدم‌زاده، کیوان صالحی، مریم کاظمی‌تبار، مهدی قربانخانی.
یادداشت	: کتابنامه.
موضوع	: آموزش و پرورش -- آزمون‌ها -- استانداردها
موضوع	: روان‌شناسی -- آزمون‌ها -- استانداردها
شناسه افزوده	: مقدم‌زاده، علی، ۱۳۵۸- مترجم
شناسه افزوده	: انجمن تحقیق و آموزش آمریکا
شناسه افزوده	: انجمن روانشناسی آمریکا
شناسه افزوده	: شورای ملی اندازه‌گیری در آموزش
شناسه افزوده	: دانشگاه تهران. مؤسسه انتشارات
رده‌بندی کنگره	: ۱۳۹۸ الف/۵ LB۳۰۵۱
رده‌بندی دیویی	: ۳۷۱/۲۶
شماره کتابشناسی ملی	: ۵۶۲۹۸۶۸

این کتاب مشمول قانون حمایت از حقوق مؤلفان و مصنفان است. تکثیر کتاب به هر روش اعم از فتوکپی، ریسوگرافی، تهیه فایل‌های pdf، لوح فشرده، بازنویسی در وبلاگ‌ها، سایت‌ها، مجله‌ها و کتاب، بدون اجازه کتبی ناشر مجاز نیست و موجب پیگرد قانونی می‌شود و تمامی حقوق برای ناشر محفوظ است.



عنوان: استانداردهایی برای آزمون‌گیری آموزشی و روانشناختی
 تألیف: انجمن تحقیقات آموزشی آمریکا- انجمن روانشناسی آمریکا-
 شورای ملی اندازه‌گیری در آموزش آمریکا
 ترجمه: علی مقدم‌زاده- دکتر کیوان صالحی، مریم کاظمی‌تبار- مهدی قربانخانی
 نوبت چاپ: اول
 تاریخ انتشار: ۱۳۹۸
 شمارگان: ۲۰۰ نسخه
 ناشر: مؤسسه انتشارات دانشگاه تهران
 چاپ و صحافی: مؤسسه انتشارات دانشگاه تهران

«مسئولیت صحت مطالب کتاب با مترجمان است»

بها: ۵۲۰۰۰۰ ریال

خیابان کارگر شمالی - خیابان شهید فرش فرشی مقدم - مؤسسه انتشارات دانشگاه تهران
 پست الکترونیک: press@ut.ac.ir - تارنما: <http://press.ut.ac.ir>
 پخش و فروش: تلفکس ۸۸۳۳۸۷۱۲

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ
فهرست مطالب

مقدمه	ز
بخش اول - مبانی	۱
فصل اول - روایی	۳
مقدمه	۳
منابع شواهد روایی	۸
شواهد مبتنی بر محتوای آزمون	۹
شواهد مبتنی بر فرایندهای پاسخ	۱۱
شواهد مبتنی بر ساختار داخلی	۱۳
شواهد مبتنی بر روابط با سایر متغیرها	۱۴
شواهد همگرا و واگرا	۱۴
روابط آزمون- ملاک	۱۵
تعمیم روایی	۱۷
شواهدی برای روایی و پیامدهای آزمون‌گیری	۱۹
تفسیر و کاربردهای نمرات آزمون که مورد نظر توسعه‌دهندگان آزمون هستند	۲۰
ادعاهای مربوط به کاربرد آزمون که مستقیماً مبتنی بر تفسیر نمره آزمون نیستند	۲۱
پیامدهای ناخواسته	۲۲
یکپارچه‌سازی شواهد روایی	۲۴
استانداردهای روایی	۲۵
فصل دوم - اعتبار / دقت و خطاهای اندازه‌گیری	۴۳
مقدمه	۴۳
دلالت‌هایی برای روایی (مفاهیم روایی)	۴۶
ویژگی‌های تکرارهای شیوه آزمون‌گیری	۴۶
ارزیابی اعتبار/دقت	۵۰

ث □ استانداردهایی برای آزمون‌گیری آموزشی و روان‌شناختی

۵۱	ضرایب اعتبار / تعمیم‌پذیری
۵۳	عوامل مؤثر بر اعتبار/دقت
۵۵	خطاهای استاندارد اندازه‌گیری
۵۶	سازگاری تصمیم
۵۷	اعتبار/ دقت میانگین‌های گروه
۵۸	مستندسازی اعتبار/دقت
۵۹	استانداردهای اعتبار/دقت
۷۱	فصل سوم - عدالت در آزمون‌گیری
۷۱	مقدمه
۷۴	دیدگاه‌های عمومی در مورد عدالت
۷۵	عدالت در رفتار در طول فرایند آزمون‌گیری
۷۶	عدالت به منزله عدم سوگیری در اندازه‌گیری
۷۷	عدالت در دسترسی به سازه(های) مورد اندازه‌گیری
۷۹	عدالت به منزله روایی تفسیر نمره آزمون انفرادی برای کاربردهای مورد نظر
۸۱	تهدیدهایی برای تفسیر معتبر و عادلانه نمرات آزمون
۸۲	محتوای آزمون
۸۳	زمینه آزمون
۸۴	پاسخ‌دهی به آزمون
۸۵	فرصت یادگیری
۸۷	به حداقل رساندن مؤلفه‌های نامربوط به سازه از طریق طراحی آزمون و انطباق‌های آزمون‌گیری
۸۹	طیف انطباق‌های آزمون
۹۱	تطابق‌های آزمون: سنجه‌های قابل‌قیاسی که سازه مورد نظر را حفظ می‌کنند
۹۳	اصلاحات آزمون: سنجه‌های غیر قابل‌قیاسی که سازه مورد نظر را تغییر می‌دهند
۹۴	گزارش نمرات حاصل از آزمون‌های تطابق‌یافته و اصلاح‌شده
۹۵	استفاده مناسب از تطابق‌ها یا اصلاحات
۹۶	استانداردهای عدالت

۱۱۵.....	بخش دوم - اجرا.....
۱۱۷.....	فصل چهارم - طراحی و توسعه آزمون.....
۱۱۷.....	مقدمه.....
۱۱۹.....	مشخصات آزمون.....
۱۱۹.....	ملاحظات عمومی.....
۱۱۹.....	بیانیه هدف و کاربردهای مورد نظر.....
۱۲۰.....	مشخصات محتوا.....
۱۲۱.....	مشخصات قالب آزمون.....
۱۲۲.....	دسترس پذیری قالب‌های گویه.....
۱۲۲.....	قالب‌های پیچیده گویه‌ها.....
۱۲۴.....	طول مدت آزمون.....
۱۲۵.....	مشخصات روان‌سنجی.....
۱۲۵.....	مشخصات نمره‌دهی.....
۱۲۷.....	مشخصات اجرای آزمون.....
۱۲۷.....	بهبودی مشخصات آزمون.....
۱۲۸.....	ملاحظات آزمون‌گیری انطباقی.....
۱۳۰.....	سیستم‌های پشتیبان تهیه گویه و آزمون.....
۱۳۰.....	طراحی و بازبینی گویه‌ها.....
۱۳۲.....	گردآوری و ارزیابی فرم‌های آزمون.....
۱۳۳.....	تهیه شیوه‌ها و مواد مربوط به اجرا و نمره‌دهی.....
۱۳۴.....	بازنویسی‌های آزمون.....
۱۳۴.....	استانداردهای طراحی و توسعه آزمون.....
۱۵۳.....	فصل پنجم - نمرات، مقیاس‌ها، هنجارها، مرتبط‌سازی نمرات و نمرات برش.....
۱۵۳.....	مقدمه.....
۱۵۴.....	تفاسیر نمرات.....
۱۵۷.....	هنجارها.....
۱۵۸.....	مرتبط‌سازی نمرات (پیونددهی نمرات).....
۱۶۳.....	نمرات برش.....

ح □ استانداردهایی برای آزمون‌گیری آموزشی و روان‌شناختی

استانداردهای نمرات، مقیاس‌ها، هنجارها، مرتبط‌سازی نمرات و نمرات برش	۱۶۵
فصل ششم - اجرا، نمره‌دهی، گزارش‌دهی و تفسیر آزمون	۱۸۱
مقدمه	۱۸۱
استانداردهای اجرا، نمره‌دهی، گزارش‌دهی و تفسیر آزمون	۱۸۵
فصل هفتم - مستندسازی برای پشتیبانی از آزمون‌ها	۱۹۹
مقدمه	۱۹۹
استانداردهای مستندسازی برای پشتیبانی از آزمون‌ها	۲۰۲
فصل هشتم - حقوق و مسئولیت‌های آزمون‌دهندگان	۲۱۱
مقدمه	۲۱۱
استانداردهای مربوط به حقوق و مسئولیت‌های آزمون‌دهندگان	۲۱۴
فصل نهم - حقوق و مسئولیت‌های کاربران آزمون	۲۲۵
مقدمه	۲۲۵
استانداردهای مربوط به حقوق و مسئولیت‌های کاربران	۲۳۰
بخش سوم - کاربردهای آزمون‌گیری	۲۴۳
فصل دهم - سنجش و آزمون‌گیری روان‌شناختی	۲۴۵
مقدمه	۲۴۵
انتخاب و اجرای آزمون	۲۴۷
تفسیر نمرات آزمون	۲۵۰
اطلاعات جنبی مورد استفاده در سنجش‌ها و آزمون‌گیری‌های روان‌شناختی	۲۵۲
انواع سنجش و آزمون‌گیری روان‌شناختی	۲۵۳
سنجش و آزمون‌گیری شناختی و عصب‌روان‌شناختی	۲۵۳
استدلال انتزاعی و تفکر طبقه‌بندی‌گرا	۲۵۴
پیشرفت تحصیلی	۲۵۴
توجه	۲۵۴
توانایی شناختی	۲۵۵

فهرست مطالب □ خ

۲۵۵	کارکرد اجرایی
۲۵۵	زبان
۲۵۶	یادگیری و حافظه
۲۵۶	کارکردهای حرکتی، کارکردهای حسی حرکتی و ترجیحات جانبی
۲۵۷	ادراک و سازمان‌دهی/یکپارچگی ادراکی
۲۵۷	آزمون‌گیری و سنجش رفتارهای مشکل‌آفرین
۲۵۷	آزمون‌گیری و سنجش خانواده و زوجین
۲۵۸	آزمون‌گیری و سنجش رفتارهای انطباقی و اجتماعی
۲۵۸	آزمون‌گیری و سنجش شخصیت
۲۵۹	آزمون‌گیری و سنجش شغلی
۲۵۹	پرسشنامه‌های علاقه‌سنجی (رغبت‌سنجی)
۲۶۰	پرسشنامه ارزش‌های شغلی
۲۶۰	سنجه‌های پیشرفت، بلوغ و بی‌تصمیمی در مسیر شغلی
۲۶۱	اهداف آزمون‌گیری و سنجش روان‌شناختی
۲۶۱	آزمون‌گیری برای تشخیص
۲۶۳	آزمون‌گیری برای ارزیابی‌های عصب‌روان‌شناختی
۲۶۴	آزمون‌گیری برای برنامه‌ریزی مداخلات و ارزیابی بازده (خروجی)
۲۶۴	آزمون‌گیری برای اتخاذ تصمیمات قضایی و حکومتی
۲۶۶	آزمون آگاهی شخصی، هویت اجتماعی و سلامت، رشد و عمل روان‌شناختی
۲۶۷	جمع‌بندی
۲۶۸	استانداردهای آزمون‌گیری و سنجش روان‌شناختی
۲۷۷	فصل یازدهم - آزمون‌گیری ضمن خدمت و اعطای گواهی
۲۷۷	مقدمه
۲۷۹	آزمون‌گیری استخدامی
۲۷۹	اثر زمینه بر کاربرد آزمون
۲۷۹	داوطلب درون‌سازمانی در مقابل داوطلب برون‌سازمانی
۲۷۹	داوطلبان آموزش‌دیده در مقابل داوطلبان آموزش‌ندیده
۲۸۰	تمرکزهای کوتاه‌مدت در مقابل تمرکزهای بلندمدت
۲۸۰	غربال‌گری ورودی در مقابل غربال‌گری خروجی

۲۸۰	تصمیم‌گیری ماشینی در مقابل تصمیم‌گیری قضاوتی
۲۸۰	استفاده مداوم در مقابل یک‌بار استفاده از آزمون
۲۸۱	عدم تغییر خزانة داوطلبان در مقابل تغییر مداوم آن
۲۸۱	حجم نمونه کم در مقابل حجم نمونه زیاد
۲۸۲	یک شغل جدید
۲۸۲	تعداد خزانة داوطلبان نسبت به تعداد فرصت‌های شغلی
۲۸۲	فرایند روایی در آزمون‌گیری استخدامی
۲۸۷	مبانی ارزیابی کاربرد آزمون استخدامی
۲۸۸	آزمون‌گیری برای اعطای گواهی‌های حرفه‌ای و شغلی
۲۹۳	استانداردهای آزمون‌گیری شغلی و اعطای گواهی
۳۰۳	فصل دوازدهم - سنجش و آزمون‌گیری آموزشی
۳۰۳	مقدمه
۳۰۵	طراحی و توسعه سنجش‌های آموزشی
۳۰۵	هدف‌های اصلی آزمون‌گیری آموزشی
۳۰۵	فراهم‌سازی اطلاعات برای یاددهی و یادگیری
۳۰۷	سنجش خروجی‌های دانش‌آموزان
۳۱۰	ارائه اطلاعات برای تصمیم‌گیری درباره دانش‌آموزان
۳۱۱	توسعه‌ی آزمون‌های آموزشی
۳۱۳	سنجش‌های چند هدفی
۳۱۴	کاربرد و تفسیر سنجش‌های آموزشی
۳۱۴	مخاطرات و پیامدهای سنجش‌ها
۳۱۵	سنجش‌های مربوط به افراد استثنایی و افراد دارای نیازهای خاص
۳۱۶	سنجش‌های جایگزین
۳۱۷	تطابق‌ها و اصلاحات
۳۱۸	سنجش بسندگی زبان انگلیسی
۳۲۰	سنجش‌های انفرادی
۳۲۱	میزان سواد سنجش و توسعه حرفه‌ای
۳۲۲	اجراء، نمره‌دهی و گزارش‌دهی سنجش‌های آموزشی
۳۲۲	اجرای آزمون‌های آموزشی

فهرست مطالب □ ذ

۳۲۳	نمره‌دهی وزنی و مرکب
۳۲۴	گزارش‌دهی نمرات
۳۲۶	استانداردهای آزمون‌گیری و ارزیابی آموزشی
فصل سیزدهم - کاربرد آزمون برای ارزشیابی برنامه‌ها، مطالعات سیاست‌گذاری و پاسخگویی ۳۳۹	
۳۳۹	مقدمه
۳۴۰	ارزشیابی برنامه‌ها و طرح‌های حمایتی سیاست‌گذاری
۳۴۳	نظام‌های پاسخگویی بر اساس آزمون
۳۴۵	موضوعات مطرح در ارزیابی برنامه‌ها و سیاست‌گذاری و پاسخگویی
۳۴۷	سایر ملاحظات
۳۵۷	واژه‌نامه
۳۸۹	نمایه

مقدمه

در جامعه‌ی ما سنجش و آزمون‌گیری آموزشی و روان‌شناختی، سهم بسیار زیادی در علوم شناختی و رفتاری دارند و منابع اطلاعاتی اساسی و قابل‌توجهی در مورد افراد و گروه‌ها فراهم می‌آورند. تمامی آزمون‌ها، به‌خوبی تدوین نشده‌اند و روش‌های آزمون‌گیری نیز، همگی مفید یا هوشمندانه نیستند، اما شواهد و مدارک زیادی وجود دارد که نشان می‌دهد آزمون‌های خوش‌ساختار و آن‌هایی که به‌خوبی تفسیر می‌شوند، سودمند هستند. آن‌دسته از آزمون‌های خوش‌ساختار که اعتبار لازم را برای دستیابی به هدف‌های مورد نظر خود دارند، مزایای قابل‌توجهی برای آزمون‌دهندگان و کاربران آزمون، می‌توانند به همراه داشته باشند. استفاده درست از آزمون‌ها نیز می‌تواند زمینه تصمیم‌گیری بهتری در مورد افراد و برنامه‌ها فراهم کند و راه دسترسی وسیع‌تر و عادلانه‌تری در راستای آموزش و استخدام ایجاد می‌کند. از سوی دیگر، استفاده نادرست از آزمون‌ها می‌تواند به آزمون‌دهندگان و افراد دیگری که تحت تأثیر تصمیمات مبتنی بر آزمون قرار دارند آسیب‌های قابل‌توجهی وارد کند. هدف از تدوین *استانداردهای آزمون‌گیری آموزشی و روان‌شناختی* این است که روش‌های بی‌عیب و نقصی برای آزمون‌گیری مهیا و مبنایی برای ارزیابی کیفیت آن روش‌ها فراهم شود. این استانداردها هم برای افرادی در نظر گرفته شده‌اند که به‌طور حرفه‌ای آزمون‌هایی را مشخص‌سازی، تهیه یا انتخاب می‌کنند و هم برای کسانی که کیفیت فنی نتایج آزمون را تفسیر یا ارزیابی می‌کنند.

هدف استانداردهای آزمون‌گیری آموزشی و روان‌شناختی

این استانداردها، با هدف ارائه ملاک‌هایی برای تهیه و ارزیابی آزمون‌ها و روش‌های آزمون‌گیری تدوین می‌شوند و در آن دستورالعمل‌هایی برای ارزیابی اعتبار تفسیر نمرات آزمون فراهم شده است. هرچند برای انجام این‌گونه ارزیابی‌ها، باید تا حد زیادی از قضاوت حرفه‌ای بهره‌گیری شود، اما این استانداردها، برای حصول اطمینان در رابطه با توجه به مسائل مربوطه، چارچوب مرجع^۱ مناسبی خواهند بود. تمامی کاربران، ناشران، حامیان و توسعه‌دهندگان حرفه‌ای آزمون، باید در جهت تأمین این استانداردها و پیروی از آن‌ها تلاش‌های معقولانه‌ای به‌عمل آورند و دیگران را نیز به انجام این کار تشویق کنند. تمامی استانداردهای قابل‌اجرا، باید در همه آزمون‌ها و در تمامی کاربردهای این آزمون‌ها رعایت شوند، مگر این‌که با دلیل حرفه‌ای واضحی، نشان دهیم که یک استاندارد، نامرتب است یا از نظر فنی در یک مورد خاص، قابل اجرا نمی‌باشد.

در این استانداردها تلاش نمی‌شود که به پرسش‌های مربوط به سیاست عمومی استفاده از آزمون‌ها، پاسخ‌های کمی روان‌شناختی داده شود. به‌طورکلی، در این استانداردها سعی بر این است که حتی‌الامکان، اطلاعات فنی مربوطه، در دسترس قرار گیرند تا افراد دخیل در تصمیمات مرتبط با سیاست‌گذاری‌ها، اطلاعات کاملی در اختیار داشته باشند.

سلب مسئولیت حقوقی

این استانداردها، به مثابه احکام و الزامات قانونی نیستند و منطبق‌شدن با آن‌ها به معنی جایگزین مشاوره حقوقی نمی‌باشد. تصمیمات زیادی از سوی نهادهای فدرال، ایالتی و محلی ابلاغ می‌شود که به برخی از جنبه‌های استفاده، تولید، نگهداری و توسعه آزمون‌ها و نتایج آزمون ارتباط دارند و استانداردهایی را بر می‌تابند که ممکن است در انواع مختلفی از آزمون‌گیری، تفاوت‌هایی به‌وجود آورند. بررسی این مسائل حقوقی فراتر از محدوده این استانداردهاست و هدف خاص آن این است که ملاک‌هایی برای روش‌های بی‌نقص آزمون‌گیری از دیدگاه متخصصان علوم رفتاری و شناختی ارائه کند. در مواردی که یک یا چند استاندارد به مسئله‌ای می‌پردازند که الزامات حقوقی مسلمی در خصوص آن وجود داشته باشد؛ ممکن است در خود استاندارد و یا در قالب توضیحات یا متن مقدماتی، مطالبی در این باره ذکر شده باشد. با این حال، اگر در رابطه با الزامات حقوقی خاصی صحبت نشده است، دلیلی بر این نیست که الزام حقوقی وجود ندارد. همچنین در صورتی که این استانداردها در سایر مرزهای بین‌المللی مورد استفاده قرار گیرند، تفاوت‌های حقوقی موجود، ممکن است زمینه‌ساز ایجاد مسائل دیگری شده یا لازم باشد تصمیمات متفاوتی درباره آن مسائل اتخاذ شود.

در برخی حوزه‌ها مانند جمع‌آوری، تحلیل و استفاده از نتایج و داده‌های آزمون برای زیرگروه‌های مختلف، ممکن است قانون، برخی از افراد مشارکت‌کننده در فرایند آزمون‌گیری را ملزم کند که اقدامات خاصی انجام دهند و از طرفی هم ممکن است مانع از انجام برخی اقدامات دیگر از سوی همان شرکت‌کنندگان شود. افزون بر این، چون علم آزمون‌گیری، رشته‌ای در حال تکامل است، بازنگری‌های اخیر در استانداردها، ممکن است هنوز در اختیارات قانونی موجود (ازجمله تصمیمات قضایی و دستورالعمل‌های سازمانی) منعکس نشده باشند. مشارکت‌کنندگان در فرایند آزمون‌گیری، در تمام موقعیت‌ها، باید در مورد الزامات حقوقی قابل اجرا، با مشاور رایزنی کنند.

علاوه بر این، هرچند سازمان‌های حامی، این استانداردها را لازم‌الاجرا نمی‌دانند، اما آن دسته از استانداردهای حرفه‌ای که به‌طورکلی، پذیرفته شده‌اند و از سوی توسعه‌دهندگان و کاربران آزمون‌ها و سایر شیوه‌های گزینش، رعایت می‌شوند، بارها از سوی مراجع قانونی و دادگاه‌ها به رسمیت شناخته

شده و مورد استناد قرار گرفته‌اند. پیروی^۱ یا عدم پیروی^۲ از استانداردها را می‌توان به‌عنوان مدارک مرتبط با مسئولیت قانونی در دادرسی‌های قضایی و نظارتی، مورد استفاده قرار داد. بنابراین شایسته است که تمامی افراد مشارکت‌کننده در فرایند آزمون‌گیری، توجه دقیقی به استانداردها داشته باشند. هیچ بخشی از این استانداردها، جای مشاوره حقوقی را پُر نمی‌کند. ناشران نیز هیچ‌گونه مسئولیتی را در رابطه با مشارکت افراد در فرایند آزمون‌گیری نمی‌پذیرند.

آزمون‌ها و مواردی از به‌کارگیری آزمون‌ها که این استانداردها در مورد آن‌ها کاربرد دارند

آزمون^۳، ابزار یا رویه‌ای است که در آن، از رفتار آزمون‌دهنده در حیطه^۴ مشخصی، نمونه‌گیری به عمل می‌آید و سپس، آن نمونه با استفاده از فرایندی استاندارد، ارزیابی و نمره‌دهی می‌شود. گاهی اوقات برای اشاره به ابزارهای ارزیابی صحت یا کیفیت پاسخ‌ها نیز، از واژه آزمون استفاده می‌شود و برای اشاره به ابزارهای اندازه‌گیری نگرش‌ها، علایق و منش‌ها^۵ از واژه‌هایی همچون مقیاس^۶ و سیاهه رفتاری^۷ به کار می‌رود اما در این استانداردها، برای اشاره به تمامی انواع ابزارهای ارزیابی، تنها از واژه آزمون استفاده می‌شود.

برخی اوقات مفهوم آزمون با مفهوم سنجش^۸ تفاوت دارد. سنجش، اصطلاحی گسترده‌تر از آزمون است و معمولاً به فرایندی اشاره دارد که یکپارچه‌سازی اطلاعات آزمون با اطلاعات سایر منابع (مثل اطلاعات سایر آزمون‌ها، سیاهه‌های رفتاری و مصاحبه‌ها، یا تاریخچه اجتماعی، آموزشی، استخدامی، بهداشتی، یا روانی افراد) را شامل می‌شود. قابلیت به‌کارگیری استانداردها در یک روش یا دستگاه ارزیابی^۹، به‌طور ذاتی تعیین می‌شود و قرار نیست با برجستگی که به آن زده می‌شود (مثل آزمون، ارزیابی، مقیاس یا سیاهه رفتاری)، تغییری در آن ایجاد شود. همان‌طور که در پایان این فصل، در بخش «ملاحظات، هنگام استفاده از استانداردها» تأکید شده است، از این استانداردها نباید به‌عنوان فهرست واریسی^{۱۰} استفاده کرد.

1- Compliance

2- Noncompliance

3- Test

4- Domain

5- Dispositions

6- Scale

7- Inventory

8- Assessment

9- Evaluation device

10- Checklist

آزمون‌ها در برخی از ابعاد، با هم تفاوت دارند از جمله: روش ارائه مواد آزمون (مثل آزمون‌های مدادکاغذی^۱، شفاهی، یا آزمون‌های رایانه‌ای)، میزان استانداردسازی مواد محرک آزمون، نوع قالب‌بندی پاسخ‌ها (انتخاب یک پاسخ از بین چند پاسخ، ارائه یک پاسخ آزاد) و نسبت اندازه‌ای که مواد آزمون برای انعکاس یا شبیه‌سازی یک زمینه خاص، طراحی شده‌است. با این حال در تمام موارد، فرایند ارزیابی و نمره‌دهی پاسخ‌های آزمون‌دهندگان به مواد آزمون، استانداردسازی می‌شود. همان‌طور که در نسخه‌های قبلی استانداردها ذکر شد، برای قضاوت در مورد صحت نتایج حاصل از کاربرد انواع مختلف آزمون‌ها، به انواع مشابهی از اطلاعات کلی، نیاز است.

مرزبندی دقیق درباره این که ابزارهای اندازه‌گیری^۲ مورد استفاده در حوزه‌های آزمون‌گیری آموزشی و روان‌شناختی، در حوزه استانداردها قرار می‌گیرند یا نه، کار دشواری است. استانداردها غالباً به‌طور مستقیم روی سنجش‌های استاندارد^۳ که عموماً تحت عنوان «آزمون‌ها» شناخته می‌شوند، اعمال می‌شوند؛ مثل سنجش‌های توانایی، استعداد، دستاورد، نگرش، علاقه، شخصیت، کارکردهای شناختی^۴ و سلامت روان. اما با این حال، استانداردها را می‌توان به شکل مفید و با درجات مختلفی روی طیف وسیعی از فنون سنجش^۵ که رسمیت کمتری دارند نیز اعمال نمود. امکان استفاده دقیق از استانداردها در برخی از موارد، به‌طور کلی وجود ندارد، از جمله برای ارزیابی‌های غیراستاندارد (مثل برخی مصاحبه‌های شغلی). یا برای طیف وسیعی از نمونه‌های رفتاری ساختارنیافته که در قالب برخی از ارزیابی‌های روان‌شناختی مدرسه‌ای یا بالینی مورد استفاده قرار می‌گیرند (مثل مصاحبه ورودی برای پذیرش) و یا برای آزمون‌های معلم‌ساخته که به‌منظور سنجش عملکرد دانش‌آموزان در امر آموزش و پرورش به کار می‌روند.

برایمان سودمند خواهد بود، اگر بتوانیم بین ابزارهایی که در زمینه مفاهیم و فنون حوزه آزمون‌گیری آموزشی و روان‌شناختی حرفی برای گفتن دارند و ابزارهایی که برای اتخاذ تصمیمات ارزیابی روزمره، مساعدت‌های غیراستاندارد یا دارای استاندارد پایین ارائه می‌دهند، تمایز قائل شویم. هرچند می‌توان اصول و مفاهیم زیربنایی این استانداردها را به‌شکلی سازنده در تصمیم‌گیری‌های روزمره به کار بست (مثل زمانی که صاحب یک کسب‌وکار با یک متقاضی شغل مصاحبه می‌کند، یک مدیر، عملکرد زیردستان را ارزیابی می‌کند، یک معلم کلاس خود را ارزیابی می‌کند تا بر پیشرفت دانش‌آموزان برای رسیدن به هدف آموزشی نظارت کند و یا زمانی که یک مربی، ورزشکار آینده‌داری را

1- Paper-and-pencil
2 Measurement devices
3 Standardized measures
4- Cognitive functioning
5- Assessment techniques

ارزیابی می‌کند)، اما انتظار بالایی است که بخواهیم استانداردهای حوزه آزمون‌گیری آموزشی و روان‌شناختی، در این‌گونه تصمیمات روزمره نیز رعایت شوند. در مقابل، سیستم ساختاریافته‌ای برای انجام مصاحبه، در قلمرو این استانداردها دیده شده‌است که توسط یک روان‌شناس تهیه شده و ادعاهایی در رابطه با پیش‌بینی عملکرد شغلی در انواع محیط‌های دیگر، با آن همراه شده است. با توجه به افزایش تعداد آزمون‌دهندگان و نیاز به حفاظت از حقوق آن‌ها، پایبندی به رعایت استانداردها اهمیت بیشتری می‌یابد.

مشارکت‌کنندگان در فرایند آزمون‌گیری

ارزیابی و آزمون‌گیری آموزشی و روانشناختی، افراد، نهادها و جامعه را به‌عنوان یک کل، درگیر کرده و تأثیر قابل‌توجهی بر آن‌ها دارد. افرادی که تحت تأثیر این امر قرار دارند دانش‌آموزان، والدین، خانواده‌ها، معلمان، مدیران آموزشی، متقاضیان شغل، کارمندان، مراجعان، بیماران، سرپرستان، مدیران اجرایی و ارزیابان هستند. نهادهای تحت تأثیر این امر نیز عبارتند از: مدارس، دانشگاه‌ها، کسب‌وکارها، صنایع، کلینیک‌های روانی و سازمان‌های دولتی. آزمون‌گیری، زمانی به سود افراد و نهادهاست که به آن‌ها در دستیابی به هدف‌هایشان کمک کند. جامعه نیز به‌نوبه خود زمانی از آزمون‌گیری بهره‌مند می‌شود که آزمون‌گیری بتواند در دستیابی به هدف‌های فردی و سازمانی به جامعه کمک کند.

مشارکت‌کنندگان زیادی در فرایند آزمون‌گیری وجود دارند که برخی از آنان عبارتند از:

- ۱- کسانی که آزمون را آماده کرده و فرایند توسعه آزمون را انجام می‌دهند؛ ۲- کسانی که آزمون را منتشر و روانه بازار می‌کنند؛ ۳- کسانی که آزمون را اجرا و نمره‌دهی می‌کنند؛ ۴- کسانی که نتایج آزمون را برای مراجعان تفسیر می‌کنند؛ ۵- کسانی که از نتایج آزمون برای برخی هدف‌های تصمیم‌گیری استفاده می‌کنند مثل سیاست‌گذاران و کسانی که از داده‌ها برای اطلاع‌رسانی سیاست اجتماعی استفاده می‌کنند؛ ۶- کسانی که به انتخاب خود یا برحسب دستور و یا بنا بر ضرورت، آزمون می‌دهند؛ ۷- کسانی که حامی مالی آزمون‌ها هستند مثل هیئت‌های نمایندگی نهادها یا سازمان‌های دولتی که برای دریافت خدمات یا ابزار خاصی با یک توسعه‌دهنده آزمون، قرارداد می‌بندند؛ و ۸- کسانی که آزمون‌ها را بازنگری کرده و از میان آن‌ها انتخاب‌هایی انجام می‌دهند و مزیت‌های مقایسه‌ای خود یا مناسب بودن آزمون‌ها برای کاربردهای پیشنهادی را ارزیابی می‌نمایند. به‌طور کلی، کسانی که در فرایند آزمون‌گیری شرکت می‌کنند باید دانش مناسبی در زمینه آزمون‌ها و ارزیابی‌ها داشته باشند تا بتوانند به خوبی تصمیم بگیرند که از کدام آزمون‌ها استفاده نمایند و نتایج آزمون را چگونه تفسیر کنند.

علائق گروه‌های مختلف درگیر در فرایند آزمون‌گیری ممکن است با هم متناسب باشند یا این‌که تناسبی با هم نداشته باشند. به‌عنوان مثال، در آزمونی که با هدف مشاوره یا تعیین سطح شغلی گرفته می‌شود، علایق فرد و سازمان اغلب متناسب و با هم سازگار هستند. در مقابل، وقتی در یک شغل بسیار رقابتی یا برای ورود به یک برنامه آموزشی، آزمونی برای گزینش یک نفر از بین جمعیت کثیری از افراد، برگزار می‌شود، اولویت‌های متقاضی آزمون ممکن است با اولویت‌های کارفرما یا مسئول پذیرش تناقض داشته باشند. همین‌طور، وقتی آزمون‌گیری توسط دادگاه اجباری می‌شود، علایق آزمون‌دهنده ممکن است با علایق کسی که از دادگاه درخواست صدور حکم کرده است، متفاوت باشد.

برخی از افراد یا نهادها ممکن است به‌طور همزمان چندین نقش در فرایند آزمون‌گیری داشته باشند. به‌عنوان مثال در کلینیک‌های پزشکی، کسی که مورد آزمون یا همان *آزمایش پزشکی* قرار می‌گیرد معمولاً از نتایج این آزمون نیز خودش بهره می‌برد. در برخی موارد، مجری آزمون، خود، وظیفه تهیه‌کنندگی آزمون را برعهده دارد و گاهی اوقات مجری آزمون، کاربر آزمون‌ها^۱ نیز هست. وقتی یک سازمان، آزمون‌های استخدامی خود را آماده می‌کند، هم توسعه‌دهنده است و هم کاربر؛ در برخی اوقات نیز یک آزمون توسط فردی ترجمه یا تالیف می‌شود که مؤلف آزمون^۲ نامیده می‌شود، اما توسط یک ناشر مستقل، منتشر، روانه بازار و توزیع می‌شود، هرچند امکان دارد که ناشر در فرایند توسعه آزمون نیز نقش فعالی داشته باشد. در ضمن، نقش‌های موجود را می‌توان به زیردسته‌های بیشتری تقسیم کرد. به‌عنوان مثال، سازمان‌های ارزیاب و ارزیابان حرفه‌ای نیز هر دو می‌توانند در تهیه و تدارک یک کانون ارزیابی نقش داشته باشند. با توجه به این آمیختگی نقش‌ها در فرایند آزمون‌گیری، اغلب به‌سختی می‌توان مسئولیت دقیقی را برای بررسی استانداردهای مختلف، به مشارکت‌کنندگان خاص اختصاص داد. کاربردهای آزمون‌ها و روش‌های آزمون‌گیری تا حدی بهبود می‌یابند که افراد درگیر در آن‌ها، از سواد ارزیابی در سطح مناسبی برخوردار باشند.

آزمون‌ها با روش‌های بسیار متنوعی، طراحی، تهیه و به‌کار گرفته می‌شوند. در برخی موارد، برای استفاده در خارج از سازمانی که این آزمون‌ها را تولید کرده است، تهیه شده و «منتشر» می‌شوند. در مواردی دیگر، از جمله سنجش‌های آموزشی استانی، آزمون‌ها توسط سازمان آموزش و پرورش استان طراحی شده و توسط پیمانکاران برای استفاده انحصاری و اغلب یک‌بار مصرف آن سازمان دولتی تهیه می‌شوند و درواقع به مرحله «انتشار» نمی‌رسند. ما در تمامی این استانداردها به جای واژه خاص *ناشر*

1- Test user

(کاربران آزمون، افراد یا گروه‌هایی هستند که از نتایج نهایی و ارزیابی‌های آزمون برای رسیدن به هدفی خاص استفاده می‌کنند)

2- Test author

آزمون^۱، از واژه کلی توسعه‌دهنده آزمون^۲ استفاده می‌کنیم تا بدین طریق به تمامی افرادی که در طراحی و توسعه آزمون‌ها، در طیف کامل سناریوهای توسعه آزمون دخالت دارند، اشاره کنیم. این استانداردها، بر این فرض استوار است که سنجش و آزمون‌گیری اثربخش، مستلزم این است که تمام متخصصان در فرایند آزمون‌گیری، از دانش، مهارت و توانایی لازم برای تحقق نقش‌های خود برخوردارند و همچنین نسبت به عوامل زمینه‌ای و شخصی (که ممکن است بر فرایند آزمون‌گیری تأثیر بگذارند) آگاهی دارند. به‌عنوان مثال، توسعه‌دهندگان آزمون و کسانی که آزمون‌ها را انتخاب کرده و نتایج آزمون را تفسیر می‌کنند، به دانش کافی در مورد اصول روان‌شناسی مثل اعتبار^۳ و روایی^۴ نیاز دارند. آن‌ها همچنین باید هرگونه تجربه نظارتی متناسب و گواهی‌نامه‌های عملی مورد نیاز (که اجبار قانونی دارند) را دریافت کنند تا جنبه‌هایی از فرایند آزمون‌گیری را که در آن‌ها نقش دارند، با شایستگی به انجام برسانند. تمام متخصصان فرایند آزمون‌گیری باید از دستورالعمل‌های اخلاقی حرفه‌ای خود، پیروی کنند.

محدوده بازبینی^۵

در این جلد، استانداردهای آزمون‌گیری آموزشی و روان‌شناختی سال ۱۹۹۹ مورد بازنگری قرار گرفته‌است. فرایند بازنگری با انتصاب یک کمیته مدیریت آغاز شد که نمایندگانی از سوی سه سازمان حامی در آن حضور داشتند و مسئول نظارت بر مسیر کلی فعالیت‌ها بودند. این سه سازمان عبارتند از: انجمن تحقیقات آموزشی آمریکا (AERA)، انجمن روان‌شناسی آمریکا (APA) و شورای ملی اندازه‌گیری در آموزش (NCME). کمیته مدیریت برای راهبری بازبینی‌ها از اعضای سازمان‌های حامی درخواست کرد در مورد استانداردهای ۱۹۹۹ اظهارنظر کنند. سپس این نظرات را ترکیب کرد و در سال ۲۰۰۹ کمیته مشترکی برای بازبینی استانداردهای ۱۹۹۹ تشکیل داد تا بازبینی واقعی را انجام دهد. اعضای کمیته مشترک نیز از سازمان‌های حامی انتخاب شده بودند و از سوی کمیته مدیریت، مسئولیت داشتند که پنج حوزه عمده را بررسی کنند: در نظر گرفتن مسائل مربوط به پاسخگویی^۶ برای استفاده از آزمون‌ها در سیاست آموزشی؛ گسترش مفهوم دسترس‌پذیری^۷ آزمون‌ها برای تمام متقاضیان آزمون؛ نمایش جامع‌تر نقش آزمون‌ها در محل کار؛ گسترش نقش فناوری در آزمون‌گیری؛ و فراهم کردن یک ساختار سازمانی بهتر برای منتقل کردن استانداردها. برای پاسخگو بودن در برابر این مسئولیت، چندین اقدام صورت گرفت:

1- Test publisher
2- Test developer
3- Reliability
4- Validity
5- Scope of the revision
6- Accountability
7- Accessibility

- فصل‌های «سنجش و آزمون‌گیری آموزشی» و «آزمون‌گیری در ارزشیابی برنامه و سیاست‌گذاری عمومی» در نسخه سال ۱۹۹۹، در این نسخه، بازنویسی شدند تا به موضوعات مربوط به کاربردهای آزمون‌ها برای اهداف پاسخگویی آموزشی، بهتر پرداخته شود.

- یک فصل جدید تحت عنوان «عدالت در آزمون‌گیری»^۱ نوشته شد تا بر دسترس‌پذیری و عدالت، به‌عنوان موضوعات اساسی در آزمون‌گیری تأکید شود. نگرانی‌های خاص مربوط به موضوع عدالت، در تمام فصل‌های استانداردها مطرح شده‌اند.

- فصل «آزمون‌گیری در استخدام و اعطای گواهی» (با عنوان جدید «آزمون‌گیری حین کار و اعطای گواهی») مجدداً سازمان‌دهی شد تا دقیق‌تر مشخص شود که چه زمانی یک استاندارد به استخدام و/یا اعطای گواهی ارتباط دارد.

- تأثیر فناوری در سرتاسر این جلد در نظر گرفته شد. یکی از مسائل عمده فناوری که مورد شناسایی قرار گرفت، این بود که عدم امکان استفاده از الگوریتم‌های اختصاصی طراحان و از سویی نیاز مبرم کاربران آزمون به مهارت لازم برای ارزشیابی برنامه‌های رایانه‌ای پیچیده، باعث بروز تنش‌هایی می‌شد از جمله در حوزه‌هایی مثل نمره‌دهی خودکار نوشتارها، اجرا و نمره‌دهی انواع اقلام ابتکاری و جدید و آزمون‌گیری رایانه‌ای. این موضوعات در فصل «طراحی و توسعه آزمون» مطرح می‌شوند.

- برای تأمین دقت فنی و وضوح محتوایی هر فصل و همچنین حفظ انسجام زبانی در کل فصل‌ها از یک ویراستار حرفه‌ای کمک گرفته شد. همان‌طور که در زیر ذکر شده، فصل‌های بخش ۱ (مبانی) و بخش ۲ (عملیات‌ها) در حال حاضر دارای یک «استاندارد فراگیر» و همچنین مضامینی هستند که سایر استانداردهای تکی، به‌عنوان زیرمجموعه‌ای از آن‌ها سازمان‌دهی می‌شوند. علاوه بر این، واژه‌نامه استانداردهای ۱۹۹۹ برای آزمون‌گیری آموزشی و روان‌شناختی، به‌روزرسانی شده است. همان‌طور که ذکر شد، تغییر عمده‌ای که در ساختار این استاندارد رخ داده است، همان مفهومی‌سازی عدالت است که به این جلد اضافه شده است. در ویرایش سال ۱۹۹۹، تنها بخشی از استاندارد و آن هم در فصل‌های جداگانه‌ای با عنوان «عدالت در آزمون‌گیری و کاربرد آزمون»، «آزمون‌گیری از افرادی با پس‌زمینه‌های زبانی متنوع» و «آزمون‌گیری از افراد استثنائی»^۲ به این موضوع اختصاص یافته بود. در ویرایش کنونی، موضوعاتی که در آن فصل‌ها بررسی شده بودند در یک فصل واحد و جامع با هم ترکیب شده‌اند که این فصل در بخش (۱) این استاندارد قرار گرفته است. این تغییر ایجاد شد تا تأکید شود که موضوع عدالت، الزام می‌کند که با تمام آزمون‌دهندگان با عدالت و برابری برخورد شود. عدالت و دسترس‌پذیری، که

1- Fairness in testing

۲- در این کتاب برای ترجمه واژه معلولیت (disability) از معادل **ناتوانایی** (نه ناتوانی) و برای اشاره به افراد دارای ناتوانی‌های خاص (معلولان) از عبارت **افراد استثنائی** استفاده شده است.

مقدمه □ غ

فرصت بلامانعی برای تمام آزمون‌دهندگان فراهم می‌کند تا جایگاه^۱ خود را در رابطه با سازه(های) مورد آزمون نشان دهند، برای تمام افراد و زیرگروه‌های حاضر در جامعه آماری آزمون‌دهندگان مورد نظر و همچنین، به منظور دستیابی به تفاسیر معتبری در رابطه با نمرات، قابل اتکا می‌باشند. از آنجایی که موضوعات مربوط به عدالت در آزمون‌گیری، صرفاً به افرادی با پس‌زمینه‌های متنوع زبانی یا افراد استثنایی، محدود نمی‌شوند، در این فصل به‌طور گسترده‌ای به تجربیات آزمون‌گیری مناسب با انواع مختلف افراد پرداخته شده است. هرچند مثال‌های موجود در این فصل اغلب به افرادی با پس‌زمینه‌های متنوع زبانی و فرهنگی و افراد استثنایی اشاره دارد، اما مثال‌هایی درباره جنسیت و افراد مُسن، افرادی با پس‌زمینه‌های نژادی و قومی مختلف و کودکان خردسال نیز آورده شده است تا به تمامی آزمون‌دهندگان نشان داده شود که چه موانع بالقوه‌ای در مسیر سنجش بی‌طرفانه و عادلانه وجود دارد.

ساختار کتاب

بخش اول استانداردها (مبانی)، شامل استانداردهای رواسازی (فصل ۱)، اعتبار/ دقت و خطاهای اندازه‌گیری (فصل ۲) و عدالت در آزمون‌گیری (فصل ۳) است. در بخش دوم (عملیات‌ها)، طراحی و توسعه آزمون (فصل ۴)، نمرات، مقیاس‌ها، هنجارها، مرتبط‌سازی نمرات (پیوند نمرات) و نمرات برش (فصل ۵)، اجرای آزمون، نمره‌دهی، گزارش‌دهی و تفسیر (فصل ۶)، پشتیبانی از مستندات آزمون‌ها (فصل ۷)، حقوق و مسئولیت‌های آزمون‌دهندگان (فصل ۸) و حقوق و مسئولیت‌های کاربران آزمون (فصل ۹) بررسی می‌شوند. در بخش سوم (کاربردهای آزمون‌گیری)، با کاربردهای خاص در زمینه سنجش و آزمون‌گیری روان‌شناختی (فصل ۱۰)، آزمون‌گیری حین‌کار و اعطای گواهی (فصل ۱۱)، سنجش و آزمون‌گیری آموزشی (فصل ۱۲) و کاربردهای آزمون‌ها برای ارزشیابی برنامه، مطالعات سیاست‌گذاری و پاسخگویی (فصل ۱۳) سر و کار داریم. تعاریف اصطلاحات خاص مطرح شده در متن نیز در واژه‌نامه پایانی کتاب ارائه شده است.

هر فصل با یک متن مقدماتی آغاز می‌شود که در آن، پس‌زمینه‌های مربوط به استانداردهای آن فصل مطرح می‌شود. هر چند در برخی از موارد، دیدگاه‌هایی نیز در متون مقدماتی آورده شده است اما نباید آن‌ها را تحت عنوان استانداردهای مربوط به آن فصل تفسیر کرد.

دسته‌بندی استانداردها

متن هر استاندارد و توضیحات^۱ همراه آن، شرایطی را نشان می‌دهند که یک استاندارد تحت آن شرایط، مرتبط تلقی می‌شود. با توجه به زمینه و هدف تهیه یا کاربرد آزمون، حضور برخی استانداردها برجسته‌تر از سایر استانداردها خواهد بود. علاوه بر این، برخی استانداردها محدوده^۲ گسترده‌ای دارند و نگرانی‌ها و الزاماتی را مطرح می‌کنند که با تمامی انواع آزمون‌ها یا زمینه‌های آزمون‌گیری در ارتباط هستند و سایر استانداردها، محدوده کوچک‌تری دارند. با این حال، تمامی استانداردها در زمینه‌هایی که اعمال می‌شوند، دارای اهمیت هستند. هر نوع طبقه‌بندی که به‌ظاهر، اهمیت کلی بالاتری برای برخی از استانداردها، نسبت به سایر استانداردها قائل شود، ممکن است سبب غفلت از بررسی برخی استانداردهای مورد نیاز در شرایط خاص شود. در این ویرایش به جای متمایز ساختن استانداردها با استفاده از برچسب‌های اولویت‌بخش مثل «اصلی»، «ثانویه»، یا «مشروط» (همانند استانداردهای ۱۹۸۵)، تأکید می‌شود که تمام استانداردها باید تأمین شوند و همه آن‌ها ضرورتاً به استانداردهای «اصلی» در آن زمینه تبدیل می‌شوند، مگر این‌که یک استاندارد، برای کاربردی خاص، کاملاً بی‌ربط، نامناسب، یا از نظر فنی غیرممکن به نظر برسد.

به جز مواردی که در یک استاندارد یا توضیحات آن مشخص شده‌اند و با توجه به ملاحظات که در بخش پایانی این فصل بیان می‌شوند، تمامی استانداردها باید قبل از استفاده عملیاتی از آزمون، تأمین گردند. هر استاندارد باید به دقت مدنظر قرار گیرد تا امکان استفاده از آن در رابطه با زمینه آزمون‌گیری مورد نظر، مشخص شود. در موارد معینی، ممکن است به دلایلی کاملاً حرفه‌ای، تبعیت از استاندارد خاصی، نامناسب قلمداد شود. همچنین، گاهی ممکن است که امکان‌سنجی فنی، بر قابلیت تأمین یک استاندارد پیش از استفاده عملیاتی از آزمون، تأثیر بگذارد. به‌عنوان مثال، برخی استانداردها ممکن است مستلزم تحلیل داده‌هایی باشند که در ابتدای استفاده عملیاتی از آزمون هنوز در دسترس نیستند. یا در مواردی، به دلیل کوچک بودن حجم نمونه‌ها، ممکن است نتوان تحلیل‌های کمی را به‌روش سنتی انجام داد. با این حال ممکن است به‌منظور پشتیبانی از استاندارد، روش‌های دیگری برای جمع‌آوری داده‌ها وجود داشته باشد، مثل روش‌های نمونه کوچک، تحقیقات کیفی، گروه‌های کانونی و حتی تحلیل منطقی. در این موارد، کاربران و توسعه‌دهندگان آزمون باید صادقانه تلاش کنند تا انواع داده‌های موردنیاز در استاندارد را فراهم کنند تا به اعتبار تفسیر نتایج آزمون برای هدف‌های مورد نظر خود کمک نمایند. اگر به تشخیص توسعه‌دهندگان و کاربران آزمون، یا به تشخیص حامیان آزمون (در صورت وجود حامی برای آزمون)، یک استاندارد غیرکاربردی یا از نظر فنی غیرعملی در نظر گرفته

1- Commentary

2- Scope

شود. باید بتوانند - در مواقع لزوم - مبنای تصمیم خود را توضیح دهند. با این حال، انتظار نمی‌رود که در حالت عادی، مستندات مربوط به تمامی این‌گونه تصمیمات در دسترس باشد.

ارائه استانداردهای اختصاصی

استانداردهای اختصاصی^۱ اغلب بعد از آوردن یک متن مقدماتی که شامل برخی مفاهیم کلیدی برای تفسیر و اعمال استانداردها است، ارائه می‌شوند. در بسیاری از موارد، استانداردها خودشان با یک یا چند توضیح همراه هستند. هدف از این توضیحات، تفصیل، تصریح یا ارائه مثال‌هایی برای کمک به تفسیر معنی استانداردهاست. این استانداردها معمولاً یک توسعه‌دهنده یا کاربر آزمون را به سوی پیاده‌سازی اقدامات خاصی رهنمون می‌سازند. با توجه به نوع آزمون، گاهی اوقات در بیانیه یک استاندارد تصریح نشده است که این استاندارد چه کسی را مورد خطاب قرار می‌دهد. به‌عنوان مثال، در استاندارد ۱،۲ در فصل «روایی» آمده است که:

به‌منظور تفسیر نمرات آزمون در حوزه و کاربردی معین، ضرورت دارد تا دلیل و منطق آن به همراه خلاصه‌ای از نظریه و شواهدی که متکی بر تفسیر مورد نظر است، ارائه شود.

کسی که مسئول پیاده‌سازی این استاندارد است، گروه یا شخصی است که تفسیر توصیه‌شده مربوط به نمرات آزمون را بیان می‌نماید. این شخص ممکن است کاربر یا توسعه‌دهنده آزمون یا کسی باشد که به‌منظور استفاده از نمرات آزمون برای یک هدف خاص (مثل طبقه‌بندی یا تصمیمات مربوط به صدور مجوز) برنامه‌ریزی انجام می‌دهد. اغلب، این امکان وجود ندارد که در بیانیه یک استاندارد مسئول این اقدامات را مشخص کرد و بنابراین: گروه یا شخصی که اقدام مشخص شده در استاندارد را انجام می‌دهد به‌عنوان مسئول پابندی به استاندارد فرض می‌شود.

برخی از استانداردهای اختصاصی و متون مقدماتی، به گروه‌ها و زیرگروه‌ها اشاره دارند. واژه گروه^۲ به‌طور کلی برای اشاره به جامعه آماری کل آزمون‌دهندگان به کار می‌رود که با عناوین دیگری از جمله گروه آزمون‌شوندگان مورد نظر^۳، گروه آزمون‌دهندگان مورد نظر^۴، جامعه آماری آزمون‌شوندگان مورد نظر^۵، یا جمعیت^۶ (جامعه آماری) نیز مورد خطاب قرار می‌گیرند. یک زیرگروه^۷، متشکل از اعضای گروه بزرگ‌تر است که به شیوه‌ای مرتبط با استاندارد در حال اعمال، شناسایی می‌شوند. وقتی داده‌ها یا

1- Individual standards

2- Comments

3- Group

4- Intended examinee group

5- Intended test-taker group

6- Intended examinee population

7- Population

8- Subgroup

ک □ استانداردهایی برای آزمون‌گیری آموزشی و روان‌شناختی

تحلیل‌هایی برای زیرگروه‌های مختلف نشان داده می‌شوند، به‌طورکلی با عناوین مختلفی مثل «زیرگروه‌هایی از گروه آزمون‌دهندگان مورد نظر»، «زیرگروه‌هایی از جامعه آماری آزمون‌شوندگان مورد نظر»، یا «زیرگروه‌های مربوطه» به آن‌ها اشاره می‌شود.

در اعمال استانداردها، مهم است به یاد داشته باشیم که در رابطه با استانداردهای اختصاصی، زیرگروه‌های مورد نظری که به آن‌ها ارجاع صورت می‌گیرد، مختص زمینه^۱ هستند. به‌عنوان مثال، زیرگروه‌های ارجاعی قومی^۲ که در طول مرحله طراحی یک آزمون مدنظر قرار گرفته می‌شوند، به ترکیب قومیتی مورد انتظار در گروه مورد نظر آزمون، بستگی دارند. علاوه بر این، در مقایسه با استاندارد که به موضوع انطباق‌های قالب یک آزمون می‌پردازد، زیرگروه‌های بسیار بیشتری می‌تواند در ارتباط با استاندارد که به‌منظور طراحی پرسش‌های عادلانه در آزمون تدوین شده است، وجود داشته باشد. کاربران استانداردها، وقتی می‌خواهند تصمیم بگیرند کدام زیرگروه‌های خاص مربوط به کاربرد یک استاندارد خاص هستند، باید تلاش کنند قضاوت حرفه‌ای داشته باشند.

برای تصمیم‌گیری در مورد این‌که کدام زیرگروه‌ها، مربوط به یک استاندارد خاص هستند، عوامل زیر را می‌توان در نظر گرفت: مدرک موثقی که نشان‌دهنده امکان رویارویی یک گروه با موانع خاصی باشد که مربوط به سازه‌های آزمون نبوده و عملکرد آزمون را مختل می‌سازند، قوانین یا مقرراتی که نشان‌دهنده ارتباط یک گروه با تفسیر نمرات، یا با تعداد زیادی از افراد گروه در جامعه آماری کل باشد. با توجه به زمینه، زیرگروه‌های مربوطه ممکن است شامل افراد زیر باشند: مردان و زنان، افرادی با وضعیت اجتماعی-اقتصادی مختلف، افرادی با نژاد و/یا قومیت مختلف، افرادی با جهت‌گیری‌های جنسیتی مختلف، افرادی با پس‌زمینه‌های متنوع فرهنگی و زبانی (به‌ویژه وقتی آزمون‌گیری در مرزهای بین‌المللی گسترش می‌یابد)، افراد استثنایی، کودکان، یا افراد مسن.

در استانداردها مثال‌های زیادی برای روشن‌شدن نکات یا مثال‌هایی از نحوه اعمال یک استاندارد خاص ارائه می‌شوند. بیشتر مثال‌ها از مطالعاتی برگرفته شده است که با دانش‌آموزان استثنایی یا اشخاصی با گروه‌های فرهنگی یا زبانی متنوع انجام گرفته‌اند و برخی از مثال‌ها هم از مطالعاتی آورده شده است که با سایر گروه‌های قابل‌شناسایی مثل کودکان خردسال یا بزرگسالان انجام شده‌اند. همچنین به‌منظور ارائه مثال‌هایی در محیط‌های صنعتی، آموزشی و روان‌شناسی، تلاش‌های هدفمندی صورت گرفته است.

1- Context specific

2- Referent ethnic subgroups

مقدمه □ ل

استانداردهای موجود هر فصل از بخش‌های ۱ و ۲ (مبانی و عملیات‌ها) با استفاده از یک استاندارد فراگیر^۱ که برای انتقال هدف اصلی آن فصل طراحی شده است، معرفی می‌شوند. این استانداردهای غالب همیشه با آوردن رقم صفر پس از شماره‌ی فصل، شماره‌گذاری می‌شوند. به‌عنوان مثال، استاندارد فراگیر فصل ۱ به‌صورت ۱,۰ شماره‌گذاری می‌شود. استانداردهای غالب، اصول راهنما را که برای تمام آزمون‌ها و کاربردهای آزمون قابل استفاده هستند به‌طور خلاصه مطرح می‌کنند. علاوه بر این، مضامین^۲ و استانداردهای هر فصل به‌گونه‌ای مرتب می‌شوند که با توالی مطالب ذکرشده در متن مقدماتی فصل سازگار باشند. از آنجایی‌که برخی از کاربران استانداردها ممکن است فقط به فصل‌هایی که مستقیماً مربوط به کاربردی خاص هستند مراجعه کنند، استانداردهای خاص در فصل‌های مختلف، به‌خصوص در بخش ۳ (کاربردهای آزمون‌گیری) تکرار می‌شوند. در تمام این تکرارها، ماهیت استاندارد^۳، یکسان است، فقط جمله‌بندی، حوزه کاربرد، یا سطح جزئیات در توضیحات تغییر می‌کنند.

ملاحظات مهم استفاده از استانداردها

برای اجتناب از تفاسیر و کاربردهای اشتباه و سوء استفاده از استانداردها، علاوه بر سلب مسئولیت حقوقی که در بالا مطرح شد، در نظر گرفتن ملاحظات زیر نیز از اهمیت ویژه‌ای برخوردار است:

- ارزیابی مقبولیت^۴ یک آزمون یا مقبولیت کاربرد آزمون، متکی به تحقق ظاهری تک تک استانداردهای این کتاب نیست و آن را نمی‌توان با استفاده از یک فهرست واریسی تعیین کرد. شرایط خاص، بر اهمیت استانداردهای اختصاصی تأثیر می‌گذارند و استانداردهای اختصاصی نباید به‌طور مجزا در نظر گرفته شوند. بنابراین ارزیابی مقبولیت به موارد زیر بستگی دارد: ۱- قضاوت حرفه‌ای که مبتنی بر دانش علوم رفتاری، روان‌سنجی و استانداردهای مربوطه در حوزه تخصصی آزمون است؛ ۲- میزان برآورده شدن معیارهای استاندارد بودن توسط کاربر یا توسعه‌دهنده آزمون؛ ۳- ابزارهای اندازه‌گیری جایگزین که به‌راحتی در دسترس هستند؛ ۴- پژوهش و شواهد تجربی در مورد امکان‌سنجی تحقق استاندارد؛ ۵- قوانین و مقررات قابل اجرا.
- وقتی آزمون‌ها در محاکم قانونی و یا شرایطی که مستلزم شهادت یک شاهد متخصص است، مورد مناقشه باشند، مقتضی است که بر مبنای دانش پذیرفته‌شده در مورد تعیین استانداردهای خاص در یک وضعیت خاص، قضاوتی حرفه‌ای صورت گیرد. هدف این استانداردها، ارائه راهنمایی‌هایی برای این‌گونه قضاوت‌ها است.

1- Overarching standard

2- themes

3- The essence of the standard

4- Acceptability

- کاربران یا توسعه‌دهندگان آزمون، باید ادعاهای خود را درباره رعایت این استانداردها از سوی یک آزمون، کتابچه راهنما و یا یک رویه، با حفظ جوانب احتیاط مطرح نمایند. برای توسعه‌دهندگان یا کاربران، مناسب است که اعلام کنند، تلاش‌هایی برای رعایت استانداردها انجام شده و اسنادی ارائه کنند که آن تلاش‌ها را تشریح و از آن‌ها پشتیبانی کند. ادعاهای مبهم و غیرصریحی که شواهدی برای پشتیبانی از آن‌ها وجود ندارد، نباید مطرح شوند.
- استانداردها، از جمله حوزه‌هایی هستند که به‌سرعت در حال تحول است. در نتیجه لازم است بر تغییرات این حوزه به‌طور مستمر نظارت شود و با توسعه دانش، این سند نیز مورد تجدیدنظر قرار گیرد. استفاده از نسخه‌های قدیمی‌تر استانداردها ممکن است به زیان آزمون‌دهندگان و کاربران آزمون باشد.
- هدف استانداردها، ارائه الزاماتی برای استفاده از روش‌های فنی خاص نیست. به‌عنوان مثال، در جایی که الزامات گزارش‌دهی آماری خاصی بیان می‌شوند، باید اصطلاح مربوطه یا «معادل پذیرفته‌شده همگانی»^۱ آن همواره مورد ادراک قرار گیرد.