

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

مبانی نظریه پرسش-پاسخ

نویسندگان:

رونالد ک. همبلتون، اچ. سوامیناتان، اچ. جین راجرز

مترجم:

دکتر محمد رضا فلسفی نژاد



انتشارات دانشگاه علامه طباطبائی

چاپ اول ۱۳۸۹

سرشناسه	: همبلتون، رونالد کی- Hambleton, Ronald K
عنوان و پدیدآور	: مبانی نظریه پرسش - پاسخ / نویسندگان رونالد ک. همبلتون، اچ. سوامیناتان، اچ. جین راجرز؛ مترجم محمدرضا فلسفی نژاد.
مشخصات نشر	: تهران: دانشگاه علامه طباطبائی، ۱۳۸۹.
مشخصات ظاهری	: ۲۴۴ ص:، جدول، نمودار
فروست	: انتشارات دانشگاه علامه طباطبائی؛ شماره نشر ۲۵۸.
شابک	: ۹۷۸-۹۶۴-۲۱۷-۰۴۲-۵
یادداشت	: عنوان اصلی: Fundamentals of item response theory, C1991
موضوع	: نظریه سوال پاسخ
شناسه افزوده	: سوامیناتان، هاریهاران
شناسه افزوده	: Swaminathan, Hariharan
شناسه افزوده	: راجرز، اچ. جین
شناسه افزوده	: Rogers, H. Jane
شناسه افزوده	: فلسفی نژاد، محمدرضا، مترجم
شناسه افزوده	: دانشگاه علامه طباطبائی
رده‌بندی کنگره	: BF ۱۷۶/۵۸۲ ۱۳۸۹
رده‌بندی دیویی	: ۷۸۲/۰۵
شماره کتابخانه ملی	: ۲۰۶۸۹۰۹



انتشارات دانشگاه علامه طباطبائی

شماره نشر: ۲۵۸

ISBN 978-964-217-042-5

شابک ۹۷۸ - ۹۶۴ - ۲۱۷ - ۰۴۲ - ۵

مبانی نظریه پرسش - پاسخ

نویسندگان: رونالد ک. همبلتون، اچ. سوامیناتان، اچ. جین راجرز

مترجم: دکتر محمدرضا فلسفی نژاد

چاپ اول: ۱۳۸۹ شمارگان: ۱۰۰۰ ناظر فنی: مهندس رضا دنیوی

لیتوگرافی، چاپ و صحافی: مرکز چاپ و انتشارات دانشگاه علامه طباطبائی

حق چاپ برای دانشگاه علامه طباطبائی محفوظ است. قیمت: ۴۰۰۰۰ ریال

سازمان مرکزی: تهران دهکده المپیک، تقاطع شهید همت، جنب پارک فرهنگ، صندوق پستی: ۱۵۸۱۵/۳۴۸۷

فهرست مطالب

فصل اول. پیشینه.....	۱
فصل دوم. مفاهیم، مدل‌ها و ویژگی‌ها.....	۱۱
فصل سوم. برآورد پارامتر سوال و توانایی.....	۴۳
فصل چهارم. سنجش برازش داده‌ها با مدل.....	۶۹
فصل پنجم. مقیاس توانایی.....	۹۹
فصل ششم. آگاهی سوالات و آزمون و توابع کارایی.....	۱۱۵
فصل هفتم. ساخت آزمون.....	۱۲۵
فصل هشتم. شناسایی سوالاتی که به‌طور بالقوه سوگیری دارند.....	۱۳۹
فصل نهم. همترازسازی نمرات آزمون.....	۱۵۷
فصل دهم. سنجش انطباقی کامپیوتری.....	۱۸۵
فصل یازدهم. جهت‌گیری‌های آینده نظریه پرسش - پاسخ.....	۱۹۷
ضمائم.....	۲۰۱

پیشگفتار مترجم

از اوایل قرن بیستم، یعنی وقتی که مطالعات اسپیرمن به ایجاد روش‌های آماری و ریاضی برای تحلیل رفتارهای تستی آزمودنی‌ها انجامید، زمان زیادی می‌گذرد. در این رهگذر، حوزه سنجش خصایص روانی، تحولات عظیمی را به خود دیده و نظریه‌ها و رویکردهای جدیدی را به عرصه وجود نهاده‌اند. از همان آغاز، دغدغه اصلی متخصصین اندازه‌گیری در اختیار داشتن ابزار و وسایلی بوده که فارغ از آزمونگر و شرایطی که در آن آزمون برگزار می‌شود، بتواند توانایی آزمودنی‌ها را بسنجد.

اگر چه ایده اصلی ساخت چنین ابزارهایی به اواخر دهه دوم قرن بیستم باز می‌گردد، اما تا به بازار آمدن کامپیوترهای پرسرعت در در اواخر دهه ۱۹۸۰ طول کشید تا ایده‌های نظری به استراتژی‌های تجربی و عملی تبدیل شود و آرزوی داشتن ویژگی عینیت خاص در سنجش روانی تحقق یابد.

امروزه نظریه‌های جدید اندازه‌گیری در کشور های پیشرفته شناخته شده‌اند و در تهیه و تنظیم بسیاری از آزمون‌های روانی مورد استفاده قرار می‌گیرند. نرم افزارهایی نظیر BILOG و MULTILOG کار کاربران را برای دستیابی به منافع حاصل از به کارگیری این روش‌ها تسهیل نموده‌اند. از جمله کاربردهای بسیار اساسی نظریه‌های جدید اندازه‌گیری می‌توان به استفاده از آن در تهیه و تنظیم آزمون‌ها برای موارد خاص، سنجش انطباقی، شناسایی سوالاتی که نسبت به گروه‌های خاص سوگیری دارند، و همترازسازی نمرات آزمون‌ها اشاره کرد.

مدل‌های نظریه پرسش-پاسخ که گاهی در تقابل با نظریه کلاسیک اندازه‌گیری، نظریه جدید اندازه‌گیری خوانده می‌شود، اطلاعات بیشتری در مورد سوال، فرد و

آزمون فراهم می آورد. در واقع برخی از محققین این حوزه، نظریه کلاسیک را مورد خاصی از مدل عمومی تر IRT می دانند. با این همه IRT نه تنها صرفاً پالایش نظریه کلاسیک نیست، بلکه نگاهی نو و متفاوت نسبت به کل فرایند روان سنجی می باشد؛ مدلی که به لحاظ مفهومی پیچیده تر و به دلیل ماهیت ریاضی آن برای درک، به سطح عمیق تری از تفکر نیاز دارد. مدل کلاسیک برای کنترل خطا در تهیه آزمون بر تصادفی ساختن و استاندارد کردن تکیه دارد که در طرح تجربی شایع است. در حالی که مدل های IRT بر مدل های ریاضی تکیه داشته تا نمرات را از اثرات ویژگی های مزاحم (دشواری، تمیز و حدس) به لحاظ آماری تعدیل نماید. از تفاوت های دیگر مدل های پرسش-پاسخ از کلاسیک آن است که IRT بر نمره سوال تاکید دارد تا نمره آزمون و از روش های غیر خطی به جای روش های خطی استفاده می کند. همچنین IRT از روش های متفاوتی برای برآورد شاخص های تمیز، دشواری و حدس بهره می گیرد.

مدل های اولیه IRT ابتدا برای بررسی سوالات دو ارزشی آزمون های توانایی معرفی شدند. اما محققین خیلی زود متوجه شدند که می توان برای سایر صفات روانی و شخصیتی نیز از آن ها بهره برد. در دهه ۱۹۸۰ مدل هایی از IRT معرفی شدند که برای سوالاتی با بیش از دو طبقه پاسخ (سوالات چند ارزشی) نیز مناسب هستند. از دیر باز دروسی زیر عنوان روان سنجی، نظریه های جدید اندازه گیری، سنجش و اندازه گیری جزو برنامه آموزشی بسیاری از رشته های تحصیلی نظیر روانشناسی، علوم تربیتی و رشته های مربوط به آن بوده است. حتی دوره کارشناسی ارشد رشته سنجش و اندازه گیری از سال تحصیلی ۱۳۷۰-۱۳۶۹ در دانشگاه علامه طباطبائی و از سال تحصیلی ۱۳۷۲-۱۳۷۱ در دانشگاه آزاد اسلامی واحد تهران مرکز دایر بوده و دوره دکتری این رشته از سال ۱۳۸۶ در دانشگاه علامه طباطبائی دایر شده، با این همه، فقدان منابع کافی از مشکلات جدی علاقمندان به این حوزه از دانش در ایران می باشد.

قبلاً کتاب هایی در زمینه سنجش و اندازه گیری به زبان فارسی ترجمه و مورد استقبال فراوان قرار گرفته است. این کتاب ها نسبت به سنجش و اندازه گیری دیدی

تخصصی داشته و در ارایه و شفاف سازی مبانی آماری و ریاضی تکنیک های اندازه گیری صفات روانی، اعم از توانایی و شخصیت، از موفقیت نسبی هم برخوردار بوده اند. نقطه مشترک اکثر آن ها این است که با تاکید بر نظریه کلاسیک اندازه گیری نگاشته شده و به صورت کلی و گذرا به نظریات جدید پرداخته اند. لذا، دید و افق روشنی را در مورد نظریه های جدید اندازه گیری پیش روی دانشجویان قرار نمی دهند. دلیل این امر نیز روشن است؛ متون به فارسی ترجمه شده به منابع انگلیسی مربوط می شوند که در زمان انتشار آن ها، نظریه های جدید اندازه حتی در امریکا و کشورهای اروپایی دوره های اولیه رشد و تکامل خود را می گذرانید. از این رو است که نیاز به وجود منابعی جامع، ساده و قابل فهم در زمینه نظریه های جدید اندازه گیری کاملاً احساس می شود. برای مرتفع کردن چنین کمبودی کتاب "مبانی نظریه پرسش-پاسخ" نوشته رونالد همبلتون و دیگران جهت ترجمه انتخاب گردید.

علاوه بر سادگی به موارد زیر می توان به عنوان دلایل دیگر انتخاب این کتاب برای ترجمه اشاره نمود. اول آن که، کتاب تقریباً شالوده مطالب اصلی نظریه پرسش-پاسخ را در بر دارد و می تواند برای درسی اختصاصی در مورد نظریه های جدید اندازه گیری مورد استفاده مدرسین، دانشجویان و محققین قرار گیرد. دوم آن که، کتاب بدون به ابتذال کشیدن مطالب از اثبات های ریاضی غیر ضرور که متن را پیچیده نموده و غالباً دانشجویان و افراد تازه وارد در این حوزه را با مشکل مواجه کرده و سدی در برابر یادگیری آن ها می باشد، بدور است. این ویژگی باعث می شود که افراد و دانشجویان با کمترین اطلاعات فنی و ریاضی نیز بتوانند کلیات نظریه پرسش-پاسخ را بفهمند و آن را در مورد مسایل عملی سنجش به کار بگیرند. سوم آن که، این کتاب به وسیله نویسندگانی به زیور طبع آراسته شده که از معروف ترین نویسندگان در حوزه IRT بوده و علاوه بر تالیفات و تحقیقات فراوان در مورد سنجش، در مواردی صاحب سبک و نظر می باشند و در واقع آشنایی با آثار آن ها به نوعی نوشیدن آب از سرچشمه می باشد. مهم تر آن که این کتاب منبع اصلی درس نظریه های جدید اندازه گیری در بسیاری از دانشگاه های معتبر دنیا می باشد. همچنین، تجربه اینجانب در تدریس محتوی کتاب و پیروی از ساختار آن در

کلاس‌های درس نظریه‌های جدید اندازه‌گیری برای دانشجویان کارشناسی ارشد رشته سنجش و اندازه‌گیری دانشگاه علامه طباطبایی در سال‌های گذشته نشان داد که دانشجویان به خوبی با مطالب ارتباط برقرار کرده و آن‌را فرا می‌گیرند.

امید است ترجمه این کتاب بتواند به دانشجویان و پژوهشگران حوزه سنجش و اندازه‌گیری کمک نموده و آن‌ها را به یادگیرندگانی مستقل در این زمینه بدل نماید. در ترجمه اصطلاحات و واژگان اختصاصی این اثر نهایت تلاش به عمل آمد، تا از واژه‌ها و کلماتی استفاده شود که هم اکنون در ادبیات سنجش و روان‌سنجی متداول می‌باشد. در مواردی که مترجم ناگزیر از جعل اصطلاح جدید بوده، نهایت سعی به عمل آمده تا واژه انتخابی به خوبی بیانگر مقصود و منظور مطلب و نویسندگان اثر باشد. به طور مثال، همکار عزیز و فرهیخته جناب آقای دکتر علی اکبر سیف پیشنهاد نمودند که در مقابل Item واژه ماده به کار گرفته شود. اگر چه با نظر ایشان مبنی بر این که آیتم در زبان انگلیسی الزاماً به معنی سوال نیست موافق هستم، به علت مصطلح بودن، از کلمه پرسش در برابر کلمه انگلیسی آیتم استفاده شد. از ایشان و سایر همکارانی که در موارد مشابه تذکرات ارزنده‌ای را گوشزد نمودند، نهایت تشکر و قدردانی را دارم.

املاي نانوشته غلط ندارد. لذا، پیشاپیش از کاستی‌های احتمالی موجود در اثر پوزش خواسته و شخصاً مسئولیت آن‌ها را به عهده می‌گیرم. از کلیه همکاران، صاحب‌نظران و دانشجویانی که زحمت ارایه بازخورد را به هر نحو مقتضی به خود داده و مرا از راهنمایی‌های خود بهره‌مند ساختند نهایت سپاسگزاری را دارم. در پایان، از اعضای محترم گروه سنجش و اندازه‌گیری دانشگاه علامه طباطبایی به‌ویژه، جناب آقای دکتر علی دلاور به خاطر تشویق‌ها و دلگرمی‌ها و همچنین از انتشارات دانشگاه علامه طباطبایی به دلیل فراهم آوردن امکان چاپ این اثر، نهایت قدردانی را دارم.

محمد رضا فلسفی نژاد

پیشگفتار ویرایشگر مجموعه (سری)

در دهه گذشته، همچنان که کاربرد نظریه کلاسیک اندازه گیری عرصه را به استفاده از نظریه پرسش - پاسخ واگذار می کرد، ما شاهد انقلابی در اندازه گیری روانی و تربیتی بوده ایم. امروزه نظریه پرسش - پاسخ به وسیله بزرگترین شرکت های آزمون سازی در ایالات متحده آمریکا و اروپا برای طراحی، اجرا، مقیاس سازی و مدرج سازی آزمون ها، تهیه آزمون و بانک های سوال، بررسی سوگیری سوال و سایر شیوه های متداول در فرایند تهیه آزمون به کار می رود. محققین اندازه گیری، سیستم مدارس عمومی، نیروهای نظامی، و تعدادی از شاخه های غیرنظامی دولت فدرال، نظریه پرسش - پاسخ را با دقت و اشتیاق فراوان پذیرفته و بکار برده اند.

این کتاب مقدمه ای ساده ولی محکم در مورد مفاهیم پایه نظریه پرسش - پاسخ فراهم می آورد و با توصیف های دقیق قابل فهم، کاربرد این نظریه در زمینه مسائل مربوط به تهیه آزمون، شناسایی سوالاتی که به طور بالقوه تورش دارند، معادل سازی آزمون و سنجش انطباقی کامپیوتری دنبال می کند. خلاصه ای در مورد جهت گیری های جدید تحقیق و توسعه نظریه پرسش - پاسخ، کتاب را کامل می کند.

همپتون، سوامیناتان و راجرز با استفاده از مباحث به دقت تنظیم شده، مفاهیم آشنای نظریه کلاسیک و آمار پایه را با هم در آمیخته و با به کار گیری مثال های عددی تفصیلی، به طور قدم به قدم نظریه پرسش - پاسخ و کاربردش را گسترش داده اند. کتاب به وسیله جداولی که حاوی نتایج تحلیل های واقعی IRT است و شکل هایی که اثرپذیری نتایج IRT از مسائلی چون مدل هایی با تعداد متفاوت پارامتر

سوال، آزمودنی‌ها در سطوح مختلف توانایی و درجات متنوع برازش میان داده‌ها و مدل‌های IRT را ترسیم می‌کند، به خوبی تشریح می‌شود. اگرچه معادلات لازم برای محاسبه آماره‌های IRT در کتاب ارائه شده ولی استنتاجات ریاضی آن‌ها حذف شده است. به رغم آن‌که این اثر یک "کتاب آشپزی"¹ در مورد روش‌های IRT نیست، خواننده می‌تواند بحث‌های دقیقی را در ارتباط با روش‌های جایگزین برآورد پارامترها در IRT بیابد. بحث‌های که به برآورد بیشینه درست‌نمایی، برآورد بیشینه درست‌نمایی حاشیه‌ای، برآورد بیزی و غیره مربوط می‌شود. توانایی در درک مباحث زیربنایی حساب برای درک منشاء این روش‌ها و وجوه تمایزشان ضروری نیست. همبلتون و دیگران به هدف سری روش‌های اندازه‌گیری در علوم انسانی، که همانا قابل دسترس کردن موضوعات، روش‌ها و مفاهیم پیچیده اندازه‌گیری برای خواننده‌ای که کمترین سابقه ریاضی را دارد و مشتاق یادگیری و به کارگیری روش‌هایی که در محدوده منجش علوم انسانی است، وفادار بوده‌اند. این کتاب مفاهیم و کاربردهای نیرومند اندازه‌گیری را به شکلی معرفی می‌کند که می‌تواند به وسیله هزاران تن از کسانی که از این پس IRT چیزی جز رازی زیبا برایشان نیست به درستی فهمیده و به کاربرده شود.

¹ ریچارد ام. جایگر

² دانشگاه کارولینا شمالی در گرینز بورو

-
1. Cook book
 2. Jaeger
 3. University of North Carolina at Greensboro

پیشگفتار مولف

مدل‌های معروف (یا کلاسیک) و شیوه‌های اندازه‌گیری ساخت آزمون‌های روانی و تربیتی و تفسیر نمرات آزمون، در دوره‌ای طولانی به متخصصین آزمون به خوبی خدمت کرده است. مروری بر کاتالوگ آزمون‌ها و نسخه‌های جدید سالنامه اندازه‌گیری‌هایی روانی^۱ و نقدهای آزمون^۲ نشان می‌دهد که آزمون‌های پیشرفت تحصیلی، استعداد و شخصیت بسیاری با به کارگیری این مدل‌ها و رویکردهای کلاسیک ساخته شده است. با این همه، روش‌هایی که بر اساس آن‌ها معمولاً آزمون‌های روانی و تربیتی ساخته، ارزشیابی و به کار می‌روند نارسائی‌های مستند شده زیادی دارند (برای مثال همبلتون^۳ ۱۹۸۹ را نگاه کنید). این نارسائی‌ها شامل موارد زیر می‌شوند: الف) از شاخص‌های سوالاتی استفاده می‌کنند که مقدارشان به گروه خاصی آزمودنی‌ها که مرجع به حساب می‌آیند وابسته است و ب) توانایی برآورده شده آزمون‌ها به سوالات خاصی که در آزمون گنجانیده شده است، وابسته می‌باشد.

متخصصین روان‌سنجی سیستم اندازه‌گیری جدیدی را نظریه پرسش - پاسخ^۴، IRT، برای مقابله با این نارسائی‌ها و سایر نارسائی‌هایی که در اندازه‌گیری معمول وجود دارد، ابداع کرده‌اند. در دهه ۱۹۸۰ نظریه پرسش - پاسخ یکی از موضوعات غالب در میان متخصصین اندازه‌گیری بود. در این کتاب، مدل‌های مشهور IRT زیادی توصیف خواهد شد. به دلیل آن که نظریه پرسش - پاسخ چارچوب

-
1. Manual Measurement year book
 2. Test critiques
 3. Hambleton
 4. Item Response Theory

مفیدی را برای حل مسائل متنوع و گسترده اندازه گیری فراهم می آورد، بسیاری از ناشران آزمون‌ها، دپارتمان‌های کشوری و استانی تعلیم و تربیت، آژانس‌های اعتباری، مدارس ناحیه‌ای، نیروهای ارتش و صنعت از این نظریه برای ساخت آزمون‌ها، شناسایی سوالاتی که به طور بالقوه سوگیری^۱ دارند، معادل‌سازی نمرات آزمون‌های مختلف یا فرم‌های مختلف یک آزمون و گزارش نمرات آزمون کمک می گیرند. نظریه پرسش - پاسخ کاربردهای نوید بخش دیگری نیز دارد. در فصول بعد، چندین مورد از این کاربردها به تفصیل مورد بحث قرار می گیرند.

چرا اکنون کتابی در مورد IRT منتشر کنیم؟ گرایش به دانستن در مورد این نظریه جدید اندازه گیری و به کارگیری آن جهانی است و نیاز به منابع کاربردی آموزشی نیز وجود دارد. بنابراین هدف این کتاب، فراهم سازی مقدمه ای کاربردی و جامع در حوزه نظریه پرسش - پاسخ است. محدودیت های شیوه کلاسیک اندازه گیری مورد اشاره قرار می گیرد تا توجیهی برای یک نظریه جایگزین فراهم نمایند. مبانی نظریه پرسش - پاسخ شامل مدل ها، مفروضات، ویژگی ها و همچنین برآورد پارامترها، روش هایی برای سنجش برازش داده ها با مدل، مقیاس های گزارش دهی جایگزین و کارایی و آگاهی سوال و آزمون بخش های اصلی این کتاب را تشکیل می دهند. چندین کاربرد مهم IRT در فصولی بعدی توصیف می شوند. هر جا ممکن بوده، روابط میان نظریه پرسش - پاسخ و نظریه کلاسیک مورد توجه قرار گرفته تا وضوح مطالب را بهبود بخشد.

از آنجایی که این کتاب برای افراد تازه کار در IRT که مهارت های متوسط آماری دارند تدوین شده، رویکرد ما بر مفاهیم پایه نظریه پرسش - پاسخ تاکید می کند و از بحث در مورد استنتاجات ریاضی و جنبه های پیچیده آماری نظریه اجتناب می کند. منابعی برای پیگیری این جنبه های مهم ارائه می گردند. هر جا ممکن باشد، مثال ها و توضیحات ارائه می گردند. برای توانمند کردن کاربران به کسب تجربه در مورد مدل ها و شیوه های IRT، در پایان هر فصل تمرین ها و پاسخ های

کامل آن‌ها ارایه شده است. سرانجام برخی از برنامه‌های معروف کامپیوتری به همراه بحثی راجع به نقاط قوت و ضعفشان معرفی می‌شوند. آگاهی از برنامه‌های کامپیوتری بکارگیری موفقیت‌آمیز مدل‌های IRT را تسهیل می‌کند.

به طور خلاصه، نظریه پرسنش - پاسخ عبارت از خانواده‌ای از مدل‌های ریاضی می‌باشند که برای طراحی، ساخت و ارزشیابی آزمون‌های روانی و تربیتی مفید می‌باشند. انتظار می‌رود همانطور که تحقیقات بعدی پیش می‌روند، مشکلات تکنیکی موجود در بکارگیری مدل‌ها حل شوند. به علاوه، انتظار می‌رود که در سال‌های آتی، مدل‌های جدیدتر و کاربردی‌تر IRT توسعه یابند و IRT را قادر به ارائه راه حل‌های بهتری برای مسائل مهم اندازه‌گیری نمایند. امیدواریم که این کتاب برای متخصصین اندازه‌گیری، که به کارگیری IRT در کارشان علاقه‌مندند، سودمند باشد.

ما سپاسگزار چندین همکار، دانشجویان قبلی و دانشجویان فعلی که به طور جامع پیش‌نویس قبلی کتاب را مرور کردند، هستیم: لوید باند¹، دانشگاه کارولینای شمالی در گریتزبورو²؛ لیندا ال. کوک³ و دانیل ایگنور⁴، آژانس آزمون‌های تربیتی⁵؛ وندی ین⁶ و آن فیتزپاتریک⁷؛ سی تی بی / مک میلان / مک گروهیل⁸؛ و راسل دبلیو جونز⁹، دانشگاه ماساچوست در امهرست¹⁰. غالباً نقطه نظرات آن‌ها ما را مجبور به شفاف ساختن بحث‌ها و موضع‌مان در مورد موضوعات تکنیکی متفاوتی می‌کرد. به دلیل تجربه و بینش بررسی کنندگان، کتاب قابل فهم تر و از نظر تکنیکی صحیح تر شد.

-
1. Lloyd Bond
 2. University of North Carolina at Greensboro
 3. Linda L. Cook
 4. Daniel Eignor
 5. Educational Testing Service
 6. Wendy Yen
 7. Anne Fitzpatrick
 8. CTB/Macmillan/McGraw-Hill
 9. Russell W. Jones
 10. University of Massachusetts at Amherst