

آزمون تطبیق‌پذیر رایانه‌ای: شیوه‌ای موثر در ارزشیابی آموزشی

الهه امینی‌فر

استادیار آموزش ریاضی، دانشگاه تربیت دبیر شهید رجایی، تهران

محمد علی‌پور*

دانش‌آموخته کارشناسی ارشد آموزش ریاضی، دانشگاه تربیت دبیر شهید رجایی، تهران

تاریخ پذیرش نهایی: ۹۲/۷/۲۹

تاریخ دریافت مقاله: ۹۲/۵/۲

چکیده

در این مقاله، ابتدا به بررسی ویژگی‌های آزمون‌های معمولی طول - ثابت یا آزمون‌های کلاسیک پرداخته شده و نقاط ضعف این نوع آزمون‌ها مورد بررسی قرار می‌گیرد. سپس آزمون‌های تطبیق‌پذیر مورد بررسی قرار گرفته و برتری‌های این نوع آزمون‌ها نسبت به آزمون‌های معمولی بیان می‌شود. در ادامه نظریه‌های مختلف اندازه‌گیری به همراه شاخص‌های آماری دشواری و قدرت تشخیص سوال مورد بررسی قرار گرفته و برتری‌های نظریه پرسش پاسخ نسبت به نظریه کلاسیک آزمون بیان خواهد شد. همچنین مزیت‌های آزمون تطبیق‌پذیر نسبت به آزمون‌های معمولی طول - ثابت مورد بررسی قرار گرفته و محدودیت‌های این گونه آزمون‌ها نیز در نظر گرفته خواهد شد. مزیت‌هایی که استفاده از رایانه در امر آموزش ایجاد می‌کند آزمون تطبیق‌پذیر رایانه‌ای را به روشی موثر برای ارزشیابی آموزشی تبدیل کرده است. در واقع، هدف اصلی این مقاله پاسخ به این پرسش است: چرا باید از آزمون‌های تطبیق‌پذیر رایانه‌ای به جای آزمون‌های معمولی طول - ثابت استفاده کرد؟

واژه‌های کلیدی

آزمون معمولی طول - ثابت، آزمون تطبیق‌پذیر، آزمون تطبیق‌پذیر رایانه‌ای و نظریه‌ی پرسش پاسخ.

۱. مقدمه

پایایی درونی آزمون از طریق انتخاب سوالاتی با دشواری متوسط و قدرت تمیز بالا تاکید می‌شود (Weiss, 2011).

مشکل اصلی این نوع آزمون‌ها این است که نمرات آزمودنی‌هایی که توانایی آنها متناسب با دشواری آزمون نیست، با دقت کم اندازه‌گیری می‌شوند. به‌عنوان مثال، فرض کنید یک آزمون ریاضی بسیار دشوار از گروهی از دانشجویان به‌عمل آمده باشد. نمره‌ی دانشجویی که برحسب تصادف، توانایی کاملاً متناسب با این آزمون را دارد، با دقت بسیار بالا مشخص خواهد شد. اما یک دانشجوی ضعیف و یک دانشجوی بسیار ضعیف، هر دو در این امتحان صفر می‌گیرند و این آزمون نمی‌تواند تفاوت بین این دو دانشجو را آشکار کند.

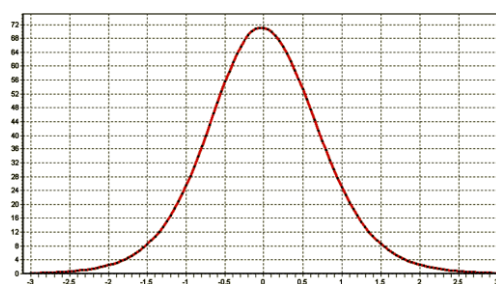
چون فرد آزمون‌گیر، قبل از اجرای آزمون، از توانایی واقعی آزمودنی‌ها آگاهی ندارد، هیچوقت نمی‌تواند مطمئن باشد که نمرات آزمودنی‌ها با چه دقتی اندازه‌گیری شده‌اند. یعنی در این نوع آزمون، باید اول توانایی واقعی آزمودنی‌ها مشخص باشد تا بتوان دقت نمره‌ی آنها را تعیین نمود. بدیهی است اگر توانایی واقعی آزمودنی‌ها مشخص باشد دیگر نیازی به آزمون‌گیری نیست!

به‌عبارت دقیق‌تر، هر آزمون معمولی طول - ثابت، دارای یک «قله» است که نمرات نزدیک به آن قله با دقت بالا اندازه‌گیری می‌شوند اما هر چه از آن قله دورتر شویم، دقت اندازه‌گیری نمرات کاهش می‌یابد. شکل ۱ نشان‌دهنده‌ی میزان اطلاعات حاصل شده از آزمون و شکل ۲ نشان‌دهنده‌ی میزان خطای استاندارد نمرات حاصل از آزمون برای یک آزمون معمولی طول - ثابت فرضی می‌باشند.

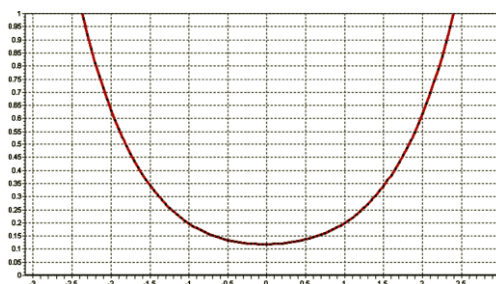
امروزه در اکثر مراکز آموزشی و دانشگاهی ایران، از آزمون‌های معمولی طول - ثابت برای اندازه‌گیری میزان توانایی یادگیرندگان استفاده می‌شود. در این نوع آزمون‌ها، یک مجموعه‌ی ثابت از سوالات به تمام آزمودنی‌ها ارائه شده و در پایان نمرات آزمودنی‌ها با هم مقایسه می‌گردد. ایراد اصلی این روش این است که نمرات آزمودنی‌هایی که توانایی آنها نزدیک به سطح دشواری آزمون است با دقت بالایی اندازه‌گیری می‌شوند؛ اما هرچه توانایی آزمودنی از سطح دشواری سوالات دورتر باشد، دقت محاسبه‌ی نمره‌ی وی نیز کاهش خواهد یافت (Weiss, 2011). آزمون تطبیق‌پذیر روشی است که با ارائه‌ی سوالات متناسب با توانایی آزمودنی، میزان خطای استاندارد اندازه‌گیری را کاهش می‌دهد، اما اجرای آزمون تطبیق‌پذیر به روش دستی، بسیار وقت‌گیر و دشوار است. به همین دلیل، امروزه برای اجرای آزمون‌های تطبیق‌پذیر، از رایانه استفاده می‌گردد. آزمون تطبیق‌پذیر رایانه‌ای به دلیل سرعت بالای محاسبه‌ی نمرات و ارائه‌ی سوالات شخصی برای آزمودنی‌ها، یکی از جدیدترین و موثرترین روش‌ها برای تخمین توانایی آزمودنی‌ها است.

۲. آزمون‌های معمولی طول - ثابت

آزمون‌های معمولی طول - ثابت، که با نام آزمون‌های کلاسیک نیز شناخته می‌شوند، نخستین بار در آزمون ورودی ارتش آمریکا برای جنگ جهانی اول در سال ۱۹۱۸ به‌طور گسترده به‌کار گرفته شدند (DuBois, 1970). در این نوع آزمون‌ها، معمولاً بر بالا بردن



شکل ۱. میزان اطلاعات به‌دست آمده در هر سطح از توانایی



شکل ۲. میزان خطای استاندارد اندازه‌گیری در هر سطح از توانایی

آزمون تطبیق‌پذیر پویا است، یعنی می‌تواند همزمان با اجرای آزمون دشواری سوالات را با سطح توانایی آزمودنی تنظیم کند.

۴. نظریه‌های اندازه‌گیری

قبل از شروع بحث در مورد برتری‌های آزمون تطبیق‌پذیر نسبت به آزمون‌های کلاسیک، مرور ویژگی‌های دو نظریه‌ی اصلی اندازه‌گیری مفید است:

۴-۱. نظریه‌ی کلاسیک آزمون

نظریه‌ی کلاسیک آزمون (CTT) روشی بود که برای رفع مشکلات آزمون‌های معمولی، مورد بررسی قرار گرفت (Weiss, 2011). در نظریه‌ی کلاسیک آزمون، نمره‌ی کسب شده توسط هر آزمودنی (X_O) از دو جزء توانایی واقعی وی (X_T) و خطای اندازه‌گیری (X_E)، تشکیل می‌شود که به صورت زیر با هم مرتبط می‌شوند (امبرتسون و رایس^۱ ۱۳۸۸):

$$X_O = X_T + X_E \quad ۱$$

در نظریه‌ی کلاسیک آزمون، برای هر سوال، دو شاخص آماری محاسبه می‌گردد:

۱. دشواری سوال: نسبتی از آزمودنی‌ها که در حل سوال شکست خورده‌اند.

۲. قدرت تشخیص سوال: همبستگی بین نحوه‌ی پاسخ‌دهی آزمودنی‌ها به آن سوال و نمره‌ی کل آزمون.

پس از محاسبه‌ی دو شاخص فوق برای تمامی سوالات، سوالاتی که دشواری متوسط و قدرت تشخیص بالایی دارند برای آزمون انتخاب خواهند شد (Weiss, 2011).

نظریه‌ی کلاسیک آزمون دو مشکل عمده داشت: یکی اینکه خطای اندازه‌گیری سوالات آزمون را برای تمام آزمودنی‌ها یکسان فرض می‌کرد (امبرتسون و رایس، ۱۳۸۸) و در نتیجه نمرات آزمودنی‌هایی که توانایی آنها متناسب با دشواری آزمون نبود با خطای بالایی محاسبه می‌گردید. دومین مشکل به نحوه‌ی محاسبه‌ی شاخص‌های دشواری و قدرت تشخیص مربوط می‌شد. در این روش ابتدا توسط یک پیش‌آزمون و با تعداد محدودی آزمودنی، شاخص‌های دشواری و قدرت تشخیص هر سوال را مشخص کرده و سپس از این مقادیر برای اندازه‌گیری توانایی افراد جامعه استفاده می‌کنند (وایز، ۲۰۱۱؛ امبرتسون و رایس، ۲۰۰۰). بدیهی است که اگر نمونه‌ی اولیه، معرف کل جامعه نباشد این شاخص‌ها نمی‌توانند توانایی افراد جامعه را به درستی محاسبه کنند.

۳. آزمون تطبیق‌پذیر

اصول زیربنایی آزمون تطبیق‌پذیر توسط آلفرد بینه در سال ۱۹۰۵ تدوین و اجرا شد و بعدها در آزمون هوش استنفورد - بینه به کار گرفته شد (Weiss, 2011). هسته‌ی مرکزی این نوع آزمون‌ها، بسیار ساده و بدیهی است. ابتدا یک توانایی اولیه را برای آزمودنی (معمولاً با حدس زدن) مشخص می‌کنیم، سپس براساس این توانایی اولیه، سوالی با دشواری مناسب به آزمودنی ارائه می‌دهیم. اگر آزمودنی به این سوال درست پاسخ داد، توانایی تخمینی وی بالاتر رفته و سوال بعدی، دشوارتر انتخاب می‌شود؛ اگر هم آزمودنی نتوانست به سوال ابتدایی پاسخ صحیح بدهد توانایی تخمینی وی کاهش یافته و سوال بعدی وی، سوال آسان‌تری خواهد بود (Linacre, 2000).

در آزمون‌های هوشمند، از آزمون تطبیق‌پذیر استنفورد - بینه به عنوان «استاندارد طلایی» یاد می‌شود. آزمون بینه تمامی ویژگی‌های آزمون‌های تطبیق‌پذیر امروزی را داشت. یک آزمون تطبیق‌پذیر پنج ویژگی دارد که آن را از آزمون‌های عادی متمایز می‌سازد (Weiss, 2011).

۱. این آزمون بر پایه یک بانک سوال که پارامترهای روان‌سنجی سوالات آن محاسبه شده‌اند اجرا می‌گردد.

۲. این آزمون می‌تواند با استفاده از اطلاعات موجود درباره آزمودنی، نقطه‌ی شروع وی را در آزمون تعیین نماید، به عبارت دیگر لازم نیست تمامی آزمودنی‌ها از یک سوال مشخص شروع به آزمون دادن کنند.

۳. سوالات در همان لحظه‌ای که پاسخ داده شوند، نمره‌دهی می‌شوند و نمره‌ی کل آزمون می‌تواند به صورت جداگانه از مجموعه سوالات هر آزمودنی به دست آید.

۴. برای انتخاب سوالات بعدی آزمودنی، از نحوه‌ی پاسخ‌دهی وی به سوالات قبلی استفاده می‌شود.

۵. آزمون هنگامی به پایان می‌رسد که آزمودنی بتواند شرط معینی را تأمین نماید، یعنی لازم نیست حتماً تمام آزمودنی‌ها به تعداد یکسانی از سوالات پاسخ دهند.

چهار ویژگی پایانی آزمون تطبیق‌پذیر نشان می‌دهد که این آزمون یک آزمون شخصی شده است. در این نوع آزمون، همه‌ی آزمودنی‌ها لزوماً از یک سوال مشخص آزمون را شروع نمی‌کنند. هر آزمودنی می‌تواند مجموعه‌ی خاصی از سوالات را دریافت کند و لازم نیست همه‌ی آزمودنی‌ها تعداد یکسانی از سوالات را دریافت نمایند. یک

۲-۴. نظریه‌ی پرسش پاسخ

نظریه‌ی پرسش پاسخ^۲ (IRT)، نظریه‌ای است که برای حل مشکلات نظریه‌ی کلاسیک آزمون مورد توجه قرار گرفت (Mislevy et al., 2000). نظریه‌ی پرسش پاسخ دارای دو ویژگی اساسی است (همبلتون و همکاران، ۱۳۸۹): اول اینکه از روی توانایی آزمودنی، می‌توان احتمال پاسخگویی وی به هر یک از سوالات آزمون را مشخص کرد. دوم اینکه مقدار این احتمال، توسط تابع مشخصه‌ی سوال قابل محاسبه است. تابع مشخصه‌ی سوال با معادله‌ی زیر تعریف می‌شود:

$$P(\theta) = \frac{1}{1 + e^{a(b-\theta)}} \quad ۲$$

در این روش، در هر مقدار از θ ، خطای مشخصی تعریف می‌شود که مقدار آن برابر است با:

$$SE(\theta) = \sqrt{\frac{1}{\sum_{i=1}^m a_i^2 P_i(1-P_i)}} \quad ۳$$

در فرمول‌های فوق، m تعداد پاسخ‌های ثبت شده، a قدرت تشخیص، b پارامتر دشواری و P تابع مشخصه‌ی سوالات است. در نتیجه، نظریه‌ی پرسش پاسخ برای هر آزمودنی، خطای جداگانه‌ای را محاسبه می‌کند (بیکر^۳ ۱۳۸۱) و از این رو می‌توان هنگام مقایسه‌ی نمره‌ی دو آزمودنی، خطای محاسبه‌ی نمرات آنها را نیز در نظر گرفت.

۵. علت استفاده از آزمون تطبیق‌پذیر به جای آزمون کلاسیک

برتری‌های آزمون تطبیق‌پذیر نسبت به آزمون کلاسیک را می‌توان به ترتیب زیر بیان کرد:

۱. نمرات حاصل از آزمون تطبیق‌پذیر، خطای کمتری دارند. در هر یک از نظریه‌های اندازه‌گیری، هنگامی که توانایی آزمودنی با دشواری سوال متناسب نباشد، دو نوع خطا بروز می‌کند: الف. خطای استاندارد اندازه‌گیری افزایش می‌یابد (Partchev, 2004; Weiss, 2011). ب. خطاهای ناشی از بی‌حوصلگی آزمودنی هنگام حل سوالات بسیار راحت یا ناامیدی وی هنگام حل سوالات بسیار مشکل، افزایش می‌یابد (Linacre, 2000). در آزمون تطبیق‌پذیر، به دلیل انتخاب سوالاتی با دشواری متناسب با توانایی آزمودنی، بروز چنین خطاهایی به حداقل می‌رسد.

۲. طول آزمون‌های تطبیق‌پذیر معمولاً کوتاه‌تر از طول آزمون‌های معمولی است. یک آزمون تطبیق‌پذیر، می‌تواند بدون کاهش دقت اندازه‌گیری، طول آزمون را تا نصف کاهش دهد (Weiss & Kingsbury, 1984; Linacre, 2000; Thompson & Weiss, 2011).

۳. آزمون به صورت شخصی اجرا می‌شود، بنابراین هر آزمودنی با تمام کردن سوالات یک مبحث، بدون اینکه مجبور باشد منتظر بقیه بماند، می‌تواند سوالات مبحث بعدی را دریافت کند.

۴. امنیت اجرای آزمون افزایش و احتمال تقلب کاهش می‌یابد.

۵. آزمون تطبیق‌پذیر با کاهش تعداد سوال مورد نیاز، از خسته شدن آزمودنی جلوگیری کرده و مانع از بروز اشتباهات ناشی از خستگی می‌شود.

۶. آزمون تطبیق‌پذیر در هر سطحی از توانایی قادر است تخمین دقیقی را ارائه دهد در حالی که آزمون‌های معمولی فقط در نقطه‌ی میانگین خود از دقت بالایی برخوردارند (Rudner, 1998).

۶. محدودیت‌های استفاده از آزمون تطبیق‌پذیر

اینها فهرستی از مشکلات آزمون‌های تطبیق‌پذیر هستند که باید مورد بررسی قرار گیرند. در پایان هر مشکل دفاعیه‌ی آقای لیناکر (۲۰۰۰) آورده شده است:

۱. نمره‌های آزمون تطبیق‌پذیر با نمره‌های آزمون مداد کاغذی همخوانی ندارند (Fair Test, 1992).

چیزی که در آزمون مهم است «نمره»ی آزمودنی نیست بلکه «دقت» تخمین توانایی وی است. آزمون تطبیق‌پذیر با حذف سوالات بسیار آسان و بسیار مشکل، توانایی آزمودنی را با دقت بالاتری تخمین می‌زند.

۲. آزمون‌های تطبیق‌پذیر در هر مبحث یا مهارتی قابل استفاده نیستند. این آزمون‌ها بر پایه‌ی نظریه‌ی IRT بنا شده‌اند که این نظریه در هر حوزه‌ای قابل به‌کارگیری نیست (همان).

این اشکال درست است. البته چنین محدودیتی در مورد آزمون‌های مداد کاغذی نیز وجود دارد.

۳. درجه‌بندی‌های رایج در آزمون‌های مداد کاغذی برای استفاده در آزمون‌های تطبیق‌پذیر مناسب نیستند (Rudner, 1998).

همان‌طور که رایت و داگلاس (۱۹۷۵) و سایر مطالعات نشان داده‌اند، هیچ مقیاس درجه‌بندی‌ای در ذات خود بهتر از بقیه نیست. در نتیجه لزومی ندارد تا درجه‌بندی آزمون‌های تطبیق‌پذیر، منطبق بر آزمون‌های مداد کاغذی باشد. درواقع، به نظر می‌رسد که درجه‌بندی

نشان داده شد، زمانی که توانایی آزمودنی در نقطه‌ای دور از قله‌ی آزمون قرار می‌گیرد، دقت اندازه‌گیری آزمون بسیار کاهش یافته و خطای اندازه‌گیری آن بسیار زیاد می‌شود. نتیجه‌ی چنین خطاهای بزرگی، کاهش کاربردی بودن این نمرات برای استفاده در تحلیل‌های آماری می‌باشد. انحراف استانداردها و واریانس‌های مملو از اشتباه، قدرت آزمون t را برای تشخیص تفاوت میانگین‌ها کاهش داده و به خطادار شدن تمامی آزمون‌های همبستگی منجر می‌شوند (Weiss, 2011).

آزمون تطبیق‌پذیر رایانه‌ای برای حل مشکلات بالا راه‌حل مناسبی است. به دلیل پویا بودن آزمون‌های CAT، که هر آزمون را منطبق بر توانایی یادگیرنده تنظیم می‌کنند، این آزمون‌ها کارآمد و بهینه هستند. کارآمد بودن آزمون CAT از آن جهت است که این آزمون برای هر آزمودنی آزمون قله‌دار مناسب وی را ارائه می‌دهد؛ یعنی با تشخیص سریع سطح توانایی آزمودنی، زیرمجموعه‌ای از سوالات بانک سوال را که مناسب این آزمودنی است، وی ارائه می‌دهد. چنین زیرمجموعه‌ای از سوالات، مجموعه‌ای است که آزمودنی بتواند ۵۰٪ از سوالات آن را صحیح پاسخ دهد. همچنین چون آزمون‌های CAT از روش پیشینه‌ی درست‌نمایی برای انتخاب سوالات استفاده می‌کنند، بصورت بهینه عمل می‌کنند یعنی با استفاده از حداقل سوالات ممکن می‌توانند توانایی آزمودنی را با حداقل خطای استاندارد اندازه‌گیری نمایند. همانطور که گفته شد، آزمون‌های CAT برای تمام آزمودنی‌ها و در تمام سطوح توانایی به‌شکلی مناسب عمل می‌کنند.

ساختن یک آزمون CAT پیچیده‌تر از ساختن یک آزمون عادی است. این نوع آزمون نیازمند بانک سوالاتی گسترده است که سوالات آن با استفاده از روش IRT مدرج شده باشند همچنین به دلیل نیاز به پیش شرط‌هایی برای استفاده از بانک سوال، آنها باید از نرم‌افزارهایی نظیر CATSim (Weiss & Guyer 2010) استفاده نمایند. تامپسون و وایز (۲۰۱۱) مراحل مورد نیاز برای ایجاد یک آزمون CAT را بیان کرده‌اند. اما علی‌رغم پیچیده‌تر بودن، این نوع آزمون‌ها بسیار دقیق‌تر و کم‌خطاتر از آزمون‌های معمولی بوده و نتایج آنها برای اندازه‌گیری توانایی آزمودنی، قابل اتکاتر است.

آزمون تطبیق‌پذیر، به دلیل وجود نمونه‌های متناسب‌تر، بهتر از بقیه باشد.

۴. وقتی هر آزمودنی سوالات متفاوتی را نسبت به سایر آزمودنی‌ها دریافت می‌کند، چگونه می‌توان ادعا کرد عدالت در این آزمون رعایت شده است؟ (همان)

به همین دلیل تا زمانی که توانایی آزمودنی با دقت بالا تخمین زده نشود، آزمون متوقف نمی‌شود.

۷. آزمون تطبیق‌پذیر رایانه‌ای

اجرای آزمون تطبیق‌پذیر به شیوه‌ی دستی به دلیل هزینه‌ی بالا و کمبود آزمون‌گیرندگان مجرب، به صورت محدود پی‌گیری شد. اما با ورود رایانه‌ها به عرصه‌ی آموزش، مشکلات اجرای آزمون تطبیق‌پذیر مرتفع گردید (Weiss, 2011). آزمون‌های تطبیق‌پذیر رایانه‌ای (CAT) با ترکیب کردن دقت بالای نظریه‌ی پرسش پاسخ با توانایی محاسباتی رایانه‌ها، امکان نمره‌دهی فوری و تطبیق‌پذیری آنی آزمون را پس از هر سوال فراهم می‌آورند.

علاوه بر مزیت‌های ذکر شده برای آزمون‌های تطبیق‌پذیر، آزمون‌های تطبیق‌پذیر رایانه‌ای در موارد زیر نیز بر آزمون‌های کلاسیک برتری دارند:

۱. طراحی، اجرا و گزارش‌گیری آزمون‌های تطبیق‌پذیر به سرعت به وسیله‌ی رایانه انجام می‌پذیرد (Linacre, 2000)؛
۲. اجرای آزمون تطبیق‌پذیر با رایانه، برای آزمون‌ها لذت‌بخش است (Rudner, 1998)؛

۳. در آزمون‌های تطبیق‌پذیر رایانه‌ای، چون رایانه مسوول انتخاب سوالات و نمره‌دهی به آنها است، اشتباهات انسانی از بین می‌روند (همان).

البته اجرای آزمون‌های تطبیق‌پذیر رایانه‌ای به دلیل نیاز به تجهیزات سخت‌افزاری گران‌قیمت، با دشواری‌های خاصی روبه‌روست (Fair Test, 1992). البته با پیشرفت فناوری و مجهز شدن مراکز آموزشی به تجهیزات جدید، این مشکلات روز به روز کمتر می‌گردند.

۸. نتیجه‌گیری

آزمون‌های معمولی قله‌دار که در قرن گذشته بیشترین کاربرد را داشته‌اند فقط و فقط هنگامی کارایی مناسبی دارند که توانایی فرد آزمون‌شونده دقیقاً برابر با نقطه‌ی قله‌ی آزمون باشد. اما واضح است که هدف اصلی اندازه‌گیری، یافتن توانایی آزمودنی است و قبل از شروع آزمون نمی‌توان توانایی وی را ملاک قرار داد. همان‌گونه که

8. Partchev, I. (2004), "A visual guide to item response theor", Friedrich-Schiller-Universit at Jena.
9. Rudner, L. (1998), "An on-line, interactive, computer adaptive testing mini-tutorial", ERIC clearing house on assessment and evaluation.
10. Thompson, N.A. & Weiss D.J. (2011), "A framework for the development of computerized adaptive tests", Practical assessment, research & evaluation 16, pp. 1-9.
11. Weiss, D.J. (2011), "Better data from better measurements using computerized adaptive testing", journal of methods and measurement in the social sciences 2, pp. 1-27.
12. Weiss, D.J. & Guyer, R. (2010), "Manual for CATSIM: comprehensive simulation of computerized adaptive testing", St. Paul MN: Assessment systems corporation.
13. Weiss, D.J. & Kingsbury, G.G. (1984), "Application of computerized adaptive testing to educational problems", Journal of educational measurement 21, pp. 361-375.
14. Wright, B.D. & Douglas, G. (1975), "Best test design and self-tailored testing", MESA memorandum 19, Department of education, Univ. of Chicago.

1. Embretson & Reise
2. Item Response Theory
3. Baker

منابع

۱. امبرتسون، سوزان ای. و رایس، استیون بی. (۱۳۸۸)، «نظریه‌های جدید روان‌سنجی برای روان‌شناسان (IRT)»، ترجمه حسن پاشاشریفی، ولی‌الله فرزاد، مجتبی حبیبی عسگرآباد و بلال ایزانلو، تهران، انتشارات رشد.
۲. بیکر، فرانک بی. (۱۳۸۱)، «پایه‌های اساسی تئوری سؤال پاسخ»، ترجمه حیدرعلی هومن و علی عسگری، تهران، نشر پارسا.
۳. همبلتون، رونالد ک.؛ سوامیناتان، اچ. و راجرز، اچ. جین (۱۳۸۹)، «مبانی نظریه پرسش-پاسخ»، ترجمه محمدرضا فلسفی نژاد، تهران، انتشارات دانشگاه علامه طباطبائی.
4. DuBois, P.H. (1970), "A history of psychological testing", Boston: Allyn and Bacon.
5. Fairtest (1992), "Computerized testing: more questions than answers", Cambridge MA: The national center for fair and open testing.
6. Linacre, J.M. (2000), "Computer-Adaptive Testing: A methodology whose time has come", MESA memorandum 69.
7. Mislevy, R.J.; Wilson, M R.; Ercikan, K. & Chudowsky, N. (2002), "Psychometric principles in student assessment", International handbook of educational evaluation, the Netherlands: Kluwer academic press .