

## نظریه کلاسیک آزمون و محدودیتهای آن

دکتر حسین سپاسی\*

### چکیده

نظریه کلاسیک آزمون سالها است به عنوان یک نظریه منسجم مورد استفاده متخصصان آزمون‌سازی قرار گرفته است. اگرچه هنوز بسیاری از آزمونهای تربیتی و روانی براساس مفروضه‌های این نظریه ساخته می‌شوند، ولی در سالهای اخیر روان‌سنجان به این نتیجه رسیده‌اند که نظریه کلاسیک آزمون دیگر قادر به حل مسائل و مشکلاتی که بر سر راه ساختن این آزمونها و مقیاسها وجود دارد نیست. مثلاً، این نظریه تصویر روشنی درباره ارتباط بین سئوالات آزمون و خصیصه یا توانایی آزمودنی که به وسیله این سئوالات اندازه‌گیری می‌شود فراهم نمی‌آورد. متخصصان آزمون‌سازی تلاش کرده‌اند محدودیتهای این نظریه را روشن سازند و پایه‌های نظریه‌های جدیدی را در ساختن آزمونهای روانی و تربیتی بنا نهند. در این مقاله، ابتدا سعی شده است مختصراً به مفاهیم اساسی نظریه کلاسیک آزمون در ساختن آزمونها و مقیاسها اشاره گردد. آنگاه، مرور مختصری بر پایایی (Reliability) آزمون به عنوان یکی از مفاهیم زیربنایی نظریه کلاسیک خواهیم داشت، که ابتدا توسط اسپیرمن (Spearman) طی دو دهه اول قرن حاضر ارائه گردید. در پایان، مقاله به محدودیتهای این نظریه خواهد پرداخت و آنها را از دیدگاه متخصصان روان‌سنجی مورد نقد و بررسی کامل قرار می‌دهد. روان‌سنجان جدید جمله بر این باورند که چون آزمونها و مقیاسهای اندازه‌گیری باید برآوردهای واقعی از خصیصه‌های مورد اندازه‌گیری فراهم آورند، به نظریه‌های جدیدتری برای ساختن آزمونهای روانی و تربیتی نیاز داریم.

### مقدمه

نزدیک به یک قرن است که نظریه کلاسیک آزمون منشاء و مقصد بیشتر تصمیم‌گیریها در

\* عضو هیأت علمی گروه علوم تربیتی دانشگاه شهید چمران اهواز

زمینه اندازه‌گیری توانایی‌های روانی و تربیتی توسط محققان و پژوهشگران بوده است. روانشناسان، معلمان، مشاوران، مدیران، و دانشجویان علوم رفتاری علاقه‌مندند اطلاعاتی درباره این نظریه و سایر نظریه‌های آزمون کسب نمایند. آنان می‌خواهند بدانند نظریه‌های آزمون بر چه اصولی استوارند، برای رسیدن به این هدف از چه قواعد و ضوابطی استفاده کنند، و نکات مهم در خصیصه مورد اندازه‌گیری را چگونه تعبیر و تفسیر نمایند. به بیان دیگر، نظریه‌های آزمون چه هستند و در چه مواردی بحث می‌کنند. آیا روانشناسان در زمینه تحقیق، معلمان در زمینه تدریس و یا دانشجویان علوم رفتاری در تحصیل روزانه خود به نظریه‌های آزمون نیاز دارند؟ آیا معلمان نمی‌توانند پیشرفت دانش‌آموزان را با آزمون‌هایی که خود می‌سازند اندازه بگیرند؟ آیا این نظریه‌ها (نظریه‌های آزمون) ریشه در روانشناسی دارند یا این که خود رشته مستقلی را تشکیل می‌دهند؟

تا اواخر قرن حاضر به علت افزایش تقاضا شاهد انبوهی از آزمون‌ها خواهیم بود که امکان دارد هر کدام به نحوی در اندازه‌گیری توانایی‌های روانی و تربیتی مورد استفاده قرار گیرند و منشاء تصمیم‌گیری‌های مهمی واقع شوند. مهرنز و لهمان (Mehrens & Lehman, ۱۹۸۴) می‌گویند بسیاری از آزمون‌های موجود از اصول نظریه‌های آزمون پیروی نمی‌کنند و آزمون‌های پیشرفت تحصیلی معمولاً ضعیف تهیه می‌شوند. آنها اعتقاد دارند که این قبیل آزمون‌ها بندرت می‌توانند توانایی‌های آزمودنی‌ها را اندازه‌گیری کنند. با وجودی که آزمون‌ها تنها وسیله اندازه‌گیری خصوصیات روانی و تربیتی به‌شمار نمی‌روند، ولی به گفته ابل (Ebel, ۱۹۹۱) تا آن زمان که جایگزین بهتری برای آزمون‌ها پیدا نشده است باید به استفاده از آنها ادامه داد.

در این مقاله ابتدا، سعی خواهد شد تا مفاهیم نظریه کلاسیک آزمون، اولین نظریه منسجم در آزمون‌سازی، به‌طور مختصر بیان گردد و آنچه را که لازم است خواننده در این باره بداند توضیح داده شود. دیر زمانی طولانی است که اصول اندازه‌گیری آزمون‌های روانی و تربیتی بر اساس نظریه کلاسیک آزمون استوار است، ولی با گذشت زمان و پیشرفت مبانی علمی نظریه‌ها و توسعه تکنولوژی، اکنون اصول این نظریه آزمون مورد انتقاد روان‌سنجان جدید قرار گرفته است. در پایان مقاله سعی خواهد شد به محدودیت‌های این نظریه که چشم‌انداز روشنی برای

ایجاد نظریه‌های دیگر در آزمون‌سازی فراهم آورده‌اند پرداخته شود.

### نظریه کلاسیک آزمون

به واسطه پیشرفت چشمگیری که در زمینه مطالعه رفتار توسط دانشمندان علوم رفتاری به عمل آمده، نظریه‌های متفاوتی در زمینه آزمون و آزمون‌سازی ایجاد شده است که دانستن آنها برای معلمان و دانشجویان علوم تربیتی و روانشناسی از اهمیت فراوانی برخوردار است. معمولاً، متخصصان روان‌سنجی در بیان نظریه‌های خود از مدل ریاضی برای ساختن شکل آزمون و تفسیر نمره‌های آن استفاده می‌کنند. یک مدل ریاضی شامل مجموعه فرضیهایی درباره داده‌هاست که به وسیله آن روابط ویژه بین سازه‌های قابل مشاهده و سازه‌های غیرقابل مشاهده مدل را توصیف می‌کند (فلدت Feldt, ۱۹۷۵). شایان ذکر است که همه مدل‌های ریاضی نمی‌توانند اساسی علمی برای اندازه‌گیری تمام جنبه‌های مختلف توانایی فراهم آورند. ولی، در عین حال وجود چنین مدل‌هایی ما را قادر می‌سازد که رابطه بین نمره‌های مشاهده شده و رفتار مورد نظر را به شیوه علمی مورد بررسی قرار دهیم و از آن برای پیش‌بینی و تخمین بعضی ویژگی‌های آزمودنی‌ها استفاده کنیم. این نظریه‌ها معمولاً چهارچوبی کلی برای بررسی فرایند تهیه آزمون‌ها فراهم می‌آورند. نظریه کلاسیک آزمون نیز در همین راستا به وجود آمده است.

سابقه نظریه کلاسیک آزمون به کار اسپیرمن، (Spearman) روانشناس انگلیسی در اوائل سالهای ۱۹۰۰ میلادی برمی‌گردد. وی در خلال سالهای ۱۹۰۴ و ۱۹۱۳ با استفاده از مفهوم همبستگی مطالعاتی انجام داد و از لحاظ ریاضی نشان داد که نمره‌های آزمون اندازه‌های گمراه‌کننده‌ای از خصوصیات آزمودنی‌ها فراهم می‌کنند. کوشش‌های اسپیرمن (۱۹۱۳) برای توصیف نمره‌های خطا و مقادیر نمره‌های واقعی موجب گردید تا وی اساس مدل نظریه کلاسیک آزمون را به وجود آورد. بعداً پژوهشگران دیگری چون گیلفورد (Guilford, ۱۹۳۰)، گولیکسن (Guliksen, ۱۹۴۵)، مگنوسون (Magnuson, ۱۹۶۷) و لرد و نویک، (Lord & Novick, ۱۹۶۸)، به تجدید نظر در مدل اسپیرمن پرداختند و آن را به صورتی که امروز مورد استفاده قرار می‌گیرد درآوردند.

مدل اسپیرمن بر این اساس است که هر نمره مشاهده شده آزمون را می‌توان ترکیبی از دو عامل یکی نمره واقعی و دیگری نمره خطا دانست که به لحاظ ریاضی به صورت فرمول زیر بیان می‌گردد:

$$X=T+E$$

در این فرمول،  $X$  نمره مشاهده شده آزمون،  $T$  نمره واقعی، و  $E$  نمره خطای تصادفی نام دارند. با استفاده از این مدل ریاضی، می‌توان نتایج معین دیگری درباره نمره واقعی ( $T$ ) و نمره مشاهده شده ( $X$ ) به دست آورد. به عنوان مثال، چنانچه در فرمول فوق نمره مشاهده شده و نمره خطا در دست باشند می‌توانیم نمره واقعی مربوط به خصیصه مورد نظر را محاسبه کنیم. به بیان دیگر، چنانچه نمره واقعی و نمره مشاهده شده را داشته باشیم، مقدار  $E$ ، خطای اندازه‌گیری را می‌توان از فرمول زیر به دست آورد.

$$E=X-T$$

بر اساس مدلی که نظریه کلاسیک برای توصیف روابط بین نمره‌ها و خصوصیات مورد اندازه‌گیری بیان می‌کند فرضهای دیگری نیز به شرح زیر بیان شده‌اند:

- ۱- نمره‌های خطا تصادفی هستند و میانگینی برابر صفر دارند، ۲- همبستگی بین نمره‌های واقعی و نمره‌های خطا برای جامعه‌ای از آزمودنیها برابر صفر است، ۳- موقعی که آزمودنیها به دو آزمون متفاوت پاسخ می‌دهند و چنانچه دو نمره هر آزمودنی به‌طور تصادفی از دو توزیع مستقل نمره‌های مشاهده شده ممکن به دست آمده باشند، همبستگی بین نمره‌های خطای دو آزمون صفر است، و ۴- بین نمره‌های واقعی و نمره‌های مشاهده شده و نمره‌های خطا رابطه خطی وجود دارد (آناستازی Anastasi، ۱۹۸۸).

پس از این که نمره‌های مشاهده شده آزمودنیها محاسبه شد و خطای معیار اندازه‌گیری معلوم گردید، چنانچه بخواهیم نمره واقعی آزمودنیها را برآورد کنیم، و یا به بیان دیگر بخواهیم رابطه بین سئوالات آزمون و خصیصه مورد بررسی را اندازه بگیریم، نظریه کلاسیک آزمون از فرمول آماری برای بسط و توصیف این مفاهیم استفاده می‌کند. مفهوم پایایی (Reliability) در نظریه کلاسیک آزمون مفهوم بسیار مهمی به شمار می‌رود، به همین جهت در زیر سعی خواهد شد تا

به اختصار به شرح آن پرداخته شود.

## پایایی

مطرح کردن نمره واقعی و نمره مشاهده شده گویای آن است که علی‌رغم میل باطنی معلم که به نمره واقعی دانش‌آموزان علاقه دارد، در عمل آنچه از میزان آموخته‌های آنان در اختیار دارد صرفاً نمره‌های مشاهده شده است. ولی سئوالی که می‌توان مطرح کرد این است که آیا نمره‌های مشاهده شده تا چه اندازه به نمره‌های واقعی نزدیکند؟ شاخصی که می‌تواند این ارتباط را بررسی کند ضریب همبستگی نام دارد. آن شاخص همبستگی که درجه ارتباط بین نمره‌های واقعی و نمره‌های مشاهده شده را تعیین می‌کند به شاخص پایایی (Reliability Index) مشهور است.

فرمولی که برای محاسبه شاخص پایایی ساخته شده است، براساس مدل ریاضی رابطه بین نمره‌های واقعی، نمره‌های مشاهده شده و نمره‌های خطا که در بالا به آنها اشاره شد استوار است. براساس این مدل ریاضی، شاخص پایایی را می‌توان به‌عنوان نسبت انحراف معیار نمره‌های واقعی به انحراف معیار نمره‌های مشاهده شده به شرح زیر بیان کرد.

$$\rho_{XT} = \frac{\sigma_T}{\sigma_X}$$

در فرمول فوق  $\rho_{XT}$  به‌عنوان شاخص پایایی یا رابطه بین نمره‌های واقعی و نمره‌های مشاهده شده،  $\sigma_T$  انحراف معیار نمره‌های واقعی، و  $\sigma_X$  انحراف معیار نمره‌های مشاهده شده تعریف شده‌اند. توجه کنید که فرمول فوق یک مفهوم نظری است و تنها همبستگی بین تمام نمره‌های واقعی و نمره‌های مشاهده شده ممکن را که از آزمون کردن همه اعضای جامعه بزرگی از آزمودنی‌ها به دست می‌آید توصیف می‌کند.

اما، در عمل این مفهوم ارزش چندانی ندارد، چه نمره‌های واقعی مستقیماً قابل مشاهده نیستند و امکان این که تمام نمره‌های مشاهده شده جامعه بزرگی از آزمودنی‌ها را به دست آوریم نیز وجود ندارد.

بنابر این، به جای آن می‌توان گروهی از آزمودنیها را در دو موقعیت متفاوت و با آزمونی یکسان یا با دو فرم هم‌تا از یک آزمون مورد آزمایش قرار می‌داد. چنانچه دو آزمون شرایط لازم آزمونهای هم‌تا را دارا باشند، این امکان وجود دارد که رابطه‌ای ریاضی بین  $\rho_{XT}$  (ضریب همبستگی بین نمره‌های واقعی و مشاهده شده)، و  $\alpha_{XX}$  (ضریب همبستگی بین نمره‌های مشاهده شده دو آزمون هم‌تا)، پیدا کنیم. برحسب نظریه کلاسیک آزمون زمانی می‌توانیم دو آزمون را هم‌تا بنامیم که ۱- هر آزمودنی در هر دو فرم آزمون نمره یکسانی بیاورد، و ۲- واریانس خطاهای دو فرم آزمون مساوی باشند. به بیان دیگر، دو آزمون دارای میانگین و واریانسهای یکسانی هستند. علاوه بر این هاپکینز (Hopkins, ۱۹۹۰) می‌گوید می‌توان چنین فرض کرد که محتوای دو آزمون هم‌تا با همدیگر جور باشند.

به نظر می‌رسد خوانندگان این مقاله در حال حاضر اطلاعاتی درباره جزئیات نظریه کلاسیک آزمون داشته باشند و از اصول نظری آن در تدریس، تحقیق، و نوشته‌های خود به‌طور گسترده استفاده می‌کنند. افراد دیگری هم که علاقه‌مندند اطلاعات دقیقتری درباره این نظریه کلاسیک کسب کنند می‌توانند به اکثر کتابهای منبعش و اندازه‌گیری و روانسنجی مراجعه نمایند. بنابر این به همین مرور مختصر درباره این نظریه اکتفا می‌شود. آنچه در این مقاله بیشتر مورد توجه قرار خواهد گرفت پرداختن به محدودیت‌های این نظریه منسجم است که در حال حاضر اساس ساخت و اعتباریابی بسیاری از آزمونهای روانی و تربیتی را در کشورمان تشکیل می‌دهد. در باقی مانده این مقاله به بررسی محدودیت‌های نظریه کلاسیک آزمون می‌پردازیم.

### محدودیت‌های نظریه کلاسیک آزمون

امروزه تعداد آزمونهای پیشرفت، استعداد، و شخصیت که بر اساس مفروضه‌های نظریه کلاسیک ساخته شده‌اند بسیارند. ولی، با وجود سابقه طولانی شهرت نظریه کلاسیک، در سالهای اخیر این نظریه مورد انتقاد متخصصان روانسنجی قرار گرفته است. از بررسی مقالات پژوهشگران می‌توان مدارک زیادی گردآوری کرد که به محدودیت‌های آزمونها و مقیاسهایی که بر اساس نظریه کلاسیک آزمون ساخته شده‌اند اشاره می‌کنند. برای مثال، برخی از روان‌سنجان

چون لرد (۱۹۸۰) و رامبرگ (Ramberg, ۱۹۹۲) از نظریه کلاسیک به عنوان یک نظریه قدیمی یاد می‌کنند و معتقدند که مفروضه‌های آن قادر به حل مسائل و مشکلات آزمون‌سازی نیستند. بنابراین، نظریه‌های دیگری برای توصیف نمره‌های آزمون یا مقیاس اندازه‌گیری (به عنوان مثال، نمره واقعی و نمره خطا) بر اساس مفروضه‌های جدیدتری ارائه داده‌اند. برای این که خواننده از این محدودیتها آگاهی پیدا کند و دریابد که چرا نظریه کلاسیک آزمون، که زمانی از آن به عنوان نظریه منسجم یاد می‌شد، دیگر قادر به حل مسائل و مشکلات در آزمون‌سازی نیست، در زیر به تشریح این محدودیتها می‌پردازیم.

۱. اولین محدودیت به یکی از مفاهیم زیربنایی نظریه کلاسیک آزمون یعنی برآورد پایایی که بر اساس فرمهای همتای آزمون استوار است اشاره می‌کند. در عمل، اندازه‌های همتایی که به وسیله آنها بتوان به اندازه‌گیری پیشرفت پرداخت مشکل یافت می‌شوند چرا که افراد در اجرای نوبت دوم به هیچ وجه عیناً همان افراد در اجرای نوبت اول آزمون نیستند. آنان مسائلی را فراموش می‌کنند، مهارتهای جدیدی را فرا می‌گیرند، و یا ممکن است سطح انگیزش و هیجان آنها تغییر کند و جزء اینها (هیولین و دراسگو Huline & Drasgow, ۱۹۸۰). به عقیده این پژوهشگران چون نظریه کلاسیک قویاً بر مفهوم فرمهای همتای آزمون استوار است، تعجبی ندارد که کاربرد آن امروزه با مشکلاتی مواجه گردد.

۲. به نظر وایس (Weiss, ۱۹۸۱)، محدودیت دیگر نظریه کلاسیک این است که فقط از نمونه‌ای از آزمودنیها برای تخمین پایایی آزمون استفاده می‌کند. وی اشاره می‌کند که تخمین پایایی در چنین حالتی تابعی از مجموعه سئوالات آزمون و نمونه افرادی خواهد بود که آزمون را در مورد آنها اجرا کرده‌ایم. به گمان وی استنباط منطقی غلط این است چنانچه یک آزمودنی در دو نمونه متفاوت از آزمودنیها مورد اندازه‌گیری قرار گیرد امکان دارد دو خطای اندازه‌گیری متفاوت به دست آید و تخمین نمره واقعی او بر اساس این که در کدام نمونه مورد اندازه‌گیری قرار بگیرد متفاوت خواهد بود. همبلتون (Hambelton, ۱۹۸۴) هم نظر مشابهی ابراز می‌دارد. وی می‌گوید این یک محدودیت نظریه کلاسیک آزمون به‌شمار می‌رود و تا زمانی که تخمین پایایی بر اساس نمونه‌ای از آزمودنیها استوار باشد به‌طور مقتضی نمی‌توان آن را در

تفسیر نمره‌ها یا در تخمین دقت اندازه‌گیریها مورد استفاده قرار داد.

۳. گاهی امکان دارد بخواهیم خصوصیات اندازه‌گیری آزمون یا یک مقیاس را در مورد گروه‌های دیگری که ممکن است از اعضاء گروه اصلی بزرگتر و یا ناهماهنگتر باشند به کار ببریم. تفاوتی که ممکن است معانی سئوالات برای گروه‌های مختلف داشته باشد و هم چنین فقدان ربط داشتن تک تک سئوالات برای ارزیابی خصیصه‌ها و سازه‌ها در برخی از جامعه‌های فرعی، این امکان را به وجود می‌آورد که به اندازه‌گیریهای غیرهم‌ارز در مورد گروه‌های مختلف منجر گردد. این کار، تفاوت‌گذاری آزمون یا مقیاس را در مورد این گروه‌های فرعی بیشتر ساختگی تا واقعی نشان می‌دهد. هیولین و دراسگو (۱۹۹۳) معتقدند که نظریه کلاسیک آزمون راه‌حل مناسبی برای این مسأله ارائه نمی‌دهد. آنها اظهار می‌دارند مواقعی که خصوصیات یک ابزار اندازه‌گیری باید در مورد گروه‌های دیگر غیر از گروه موردنظر مورد مقایسه قرار گردد، سازنده آزمون باید تعیین کند که آزمون، مثلاً یک آزمون روانی یا هوشی، اندازه‌گیری سوداری از تواناییهای گروه‌های دیگر، مثلاً گروه‌های اقلیت، فراهم نمی‌آورد. در نظریه کلاسیک آزمون تفاوت در شاخصهای استاندارد مثلاً، دشواری سؤال (Item difficulty)، نسبت پاسخهای درست) نمی‌تواند برای ارزیابی این سوگیریها به کار برده شود. گاهی اوقات شاخص تمیز سؤال (Item Discrimination Index)، تفاوت نسبت پاسخهای درست گروه‌های ملاک به سئوالات آزمون یا مقیاس) نیز قادر به از میان برداشتن این سوگیری نمی‌باشد. به عنوان مثال، گاهی مشاهده می‌شود که شاخص تمیز نسبت پاسخهای درست گروه‌های ملاک (بالا و پایین) به هر سؤال آزمون را معکوس ارزیابی می‌کند. همبلتون، ۱۹۸۴ می‌گوید این شاخصها از پیدا کردن سوگیری تفاوت درون - گروهی در خصوصیتی که به وسیله آزمون اندازه‌گیری می‌شود عاجزند.

۴. بیکر (Baker, ۱۹۸۵) محدودیت دیگر این نظریه را عدم مشخص بودن چگونگی روپروشدن آزمودنی با سئوالات آزمون ذکر می‌کند. در توضیح این محدودیت لازم به یادآوری است که باید رابطه‌ای بین پاسخهای سؤال و خصوصیتی که به وسیله آزمون یا مقیاس اندازه‌گیری می‌شود وجود داشته باشد. به بیان دیگر، می‌توان احتمال پاسخ مثبت به سؤال را



تابعی از خصوصیت یا توانایی مورد ارزیابی به وسیله آزمون یا مقیاس فرض کرد. اساس فرض فوق این است که افرادی که از خصوصیت مورد اندازه‌گیری به وسیله آزمون یا مقیاس از مقادیر بیشتری برخوردارند، به عنوان مثال، افرادی که رضایت شغلی بالاتری دارند یا دانش‌آموزانی که پیشرفت ریاضیات عالیتری نشان می‌دهند، قاعده‌تاً باید احتمال پاسخ مثبت آنان در مقایسه با افرادی که مقادیر کمتری از این خصوصیات را دارند بیشتر باشد. موقعی که سعی می‌کنیم سئوالات آزمون یا مقیاسی را با سطح توانایی آزمودنیها جور کنیم دانستن احتمال پاسخ مثبت آزمودنی به هر سؤال یا مقیاس بسیار بااهمیت است. مثلاً، چنانچه سازنده آزمون بخواهد از نمره‌های آزمون یا مقیاس، خصوصیات پاسخ آزمودنیها را در یک جامعه یا جامعه‌های بزرگتر برآورد کند و یا این که بخواهد آزمونی را با خصوصیات ویژه‌ای برای جامعه خاصی از آزمودنیها طرح‌ریزی کند، آنگاه داشتن اطلاعات درباره احتمال پاسخ دادن مثبت به هر یک از سئوالات آزمون یا مقیاس برای وی بسیار با اهمیت به شمار می‌رود.

هیولین و دراسگو (۱۹۸۲) می‌گویند نمره‌های واقعی و خطا و یا سایر مفروضه‌هایی که اساس نظریه کلاسیک آزمون را تشکیل می‌دهند معمولاً به صورت نمره کل آزمون تعریف می‌شوند. بنابر این، واحد تحلیل در نظریه کلاسیک کل آزمون محسوب می‌گردد. آنان می‌افزایند چنانچه بین احتمال پاسخ به سئوالات خصیصه‌ای که توسط آزمون یا مقیاس اندازه‌گیری می‌شود ارتباط باشد، نمره‌های واقعی و خطای آزمودنیها را بهتر می‌توان برآورد کرد و این‌کاری است که نظریه کلاسیک آزمون قادر به انجام آن نیست.

۵. یکی دیگر از محدودیتهای نظریه کلاسیک آزمون به شیوه و تکنیک طرح و ساختن آزمون یا مقیاس برمی‌گردد، یعنی این‌که مقایسه توانایی آزمودنیها معمولاً به وسیله یک آزمون یا فرمهای همتای آن محدود می‌گردد. یک اشکال این شیوه این است چون بسیاری از آزمونها پیشرفت و استعداد نوعاً برای دانش‌آموزان با توانایی درحد متوسط ساخته می‌شوند، تخمینهای بسیار دقیقی از توانایی دانش‌آموزان قوی یا ضعیف فراهم نمی‌آورند. به عقیده لرد، (۱۹۸۰) موقعی که دشواری آزمون با سطح توانایی تقریبی هر آزمودنی جور نباشد، شاهد تغییر در پایایی آزمون خواهیم بود. وی اشاره می‌کند چنانچه سئوالات آزمون را طوری بسازیم که با

سطوح توانایی آزمودنیها جور باشند، می‌توانیم از طول آزمون بدون این که کاهشی در پایایی آن فراهم گردد بکاهیم. ولی، اگر آزمونی درست کنیم که سئوالات آن از لحاظ سطح دشواری متفاوت باشند، در این حالت کار مقایسه بین آزمودنیها با مشکل مواجه خواهد شد. لرد برای توجیه این مطالب می‌گوید دو دانش‌آموز را که در دو آزمون با دشواری متفاوت عملکردی معادل ۵۰ درصد به دست آورده‌اند نمی‌توان دارای توانایی یکسانی به حساب آوردیم. و یا دانش‌آموزی را که در یک آزمون آسان به ۶۰ درصد سئوالات پاسخ درست داده است نمی‌توان به لحاظ توانایی بالاتر یا پایین‌تر از دانش‌آموزی دانست که در یک آزمون مشکل به ۴۰ درصد سئوالات پاسخ درست داده است. کار مقایسه آزمودنیهای را که به نمونه‌هایی از سئوالات آزمون با دشواری متفاوت پاسخ داده‌اند به آسانی نمی‌توان به وسیله مدل نظریه کلاسیک آزمون مورد بررسی قرار داد (رامبرگ، ۱۹۹۲).

هیولین و دراسگو (۱۹۸۳) نیز به شیوه تهیه، تجدیدنظر، و هم‌ارزسازی در نظریه کلاسیک آزمون اشاره می‌کنند. آنان اظهار می‌دارند که این نظریه قادر نیست برآوردهای مستقیمی از مقدار اطلاعات موجود در نمره‌های مشاهده شده درباره توانایی آزمودنیها در نقاط ویژه‌ای در امتداد این پیوستار فراهم آورد. به گفته ایشان آماره‌هایی که برای نظریه کلاسیک آزمون محاسبه می‌شوند (مثلاً، خطای معیار اندازه‌گیری و ضریب آلفا) دقت آزمون یا مقیاس را به صورت گروهی نشان می‌دهد و هیچ مطلبی برای گفتن درباره دقت آزمون یا مقیاس در نقاط ویژه‌ای روی پیوستار خصیصه مکتون (مثلاً، نمره برش برای تصمیم‌گیری در مورد استخدام با رضایت شغلی بالا) ندارد.

۶. لوین (Levin, ۱۹۷۸) به محدودیت دیگر نظریه کلاسیک آزمون که واریانس خطای آزمون را برای تمام آزمودنیها یکسان فرض می‌کند اشاره می‌نماید. وی می‌گوید معمولاً می‌توان افرادی را مشاهده کرد که مثلاً به سئوالات آزمونی با ثبات‌تر و دقیق‌تر از دیگران پاسخ می‌دهند. به بیان دیگر، میزان ثبات و دقت آزمودنیها در اجرای امتحانات یا تکالیفی که به آنها واگذار می‌شود نسبت به توانایی آنها متفاوت است. بنابر این، عملکرد آزمودنیها با سطح توانایی عالی در پاسخ به چندین فرم آزمونهای هم‌تا امکان دارد از عملکرد آزمودنیها با سطح توانایی

متوسط یا ثبات تر باشد.

۷. محدودیت دیگر نظریه کلاسیک آزمون به عدم توانایی آزمون یا مقیاس در توصیف برخی مسائل که آزمودنیها در پاسخ دادن به سئوالات آزمون یا مقیاس مواجهه می‌شوند مربوط می‌گردد. به عنوان مثال، پاسخ‌دهندگان امکان دارد پرسشنامه نگرشی را بدون خواندن برخی سئوالات پاسخ دهند، یا دانش‌آموز ضعیفی ممکن است با رونویسی از روی ورقه دانش‌آموز قوی بغل دستی نمره بالایی بگیرد، یا این که یک آزمودنی امکان دارد بخشی از پاسخنامه را جا بگذارد و بنابراین نمره پایینی در آزمون کسب نماید. سئوالی که متخصصان آزمون‌سازی در این زمینه می‌پرسند این است که چگونه می‌توان این الگوهای نامناسب پاسخ آزمودنیها را پیدا کرد؟ هیولین و دراسگو (۱۹۸۳) می‌گویند پاسخ این سئوال در آزمون‌سازی از اهمیت ویژه‌ای برخوردار است و از آنجا که نظریه کلاسیک آزمون کل آزمون را به عنوان واحد تحلیل به شمار می‌آورد از عهده پاسخ دادن به آن بر نمی‌آید. آنان توجه را به نظریه‌های دیگر آزمون‌سازی معطوف می‌سازند که برای شناسایی الگوهای غیرمعمول پاسخ به سئوالات آزمون روش جدیدی در اختیار سازنده آزمون قرار می‌دهد. این روان‌سنجان معتقدند که اغلب می‌توان پاسخهای غیرمعمول را به سهولت شناسایی کرد، چه این که این قبیل پاسخها در مقایسه با الگوهای پاسخ معمول که بسیار باثباتند الگوهای بی‌ثباتی فراهم می‌آورند.

به عنوان مثال، چنانچه فرض کنیم یک آزمودنی باهوش در پاسخ دادن به تمام سئوالات باهوش باقی خواهد ماند (با پاسخ دادن صحیح به اکثر سئوالات)، چنانچه او بخشی از سئوالات را جا بگذارد، بدون شک بررسی پاسخ او به سئوالات الگویی غیرمعمول را نشان می‌دهد، یعنی الگویی باثبات (دادن پاسخ درست) در بخشهایی از آزمون و الگویی بی‌ثبات (دادن پاسخ غلط) در بخش دیگری از آزمون دارد. هیولین و دراسگو (۱۹۸۳)، معتقدند که نظریه کلاسیک آزمون توان شناسایی این مسائل را ندارد. در مقابل، نظریه‌های جدید آزمون روشهای مطمئن‌تری برای پیدا کردن چنین الگوهای غیرمعمول آزمودنیها ارائه می‌دهند.

به واسطه این محدودیتها، نظریه کلاسیک و شیوه‌های مربوط به آن نتوانسته است راه حل رضایت‌بخشی برای مسائلی که آزمون با آن مواجه است پیدا کند. بنابراین، پژوهشگران و

متخصصان روان‌سنجی سعی کرده‌اند نظریه‌های مناسب‌تری برای ساختن آزمونهای روانی و تربیتی تدوین کنند. در حال حاضر، مشهورترین مدلها برای تخمین پاسخ آزمودنی به مجموعه‌ای از سؤالات آزمون، حول «نظریه خصیصه مکنون» (Latent Trait Theory) فراهم آمده است. نام دیگر این نظریه که امروزه شهرت بسیار یافته «نظریه پاسخ به سؤال» (Item Response Theory) است که تشریح آن به مقاله دیگری نیازمند است.

منابع

منابع انگلیسی

- Anastasi, A. (1988). *Psychology testing* (6th ed.). New York: McGraw-Hill
- Baker, B. F. (1985). *The basics of item response theory*. Portsmouth, New Hampshire: Heineman.
- Bloom, S. B. (1978). *Toward a theory of testing, measurement, evaluation, and assessment*. Cseir Report # 10 UCB.
- Cronbach, L. J. (1969). Review of the theory of Achievement Test Issues. *Psychometrika*, 35: 509 - 511.
- Ebel, L. R. & Frisbie. A. D. (1991). *Essentials of educational measurement*. Engelwood Cliffs, New York: Prentice Hall.
- Feldet, L. S. (1975). Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika*, 40: 557-561.
- Guilford, J. P. (1930). *Psychometric methods* (2nd ed.) New York: McGraw-Hill
- Guliksen, H. (1945). *Theory of Mental Test*. New York: John Wiley.
- Hambleton, K. R & Novick, M. R. (1973). Toward an integration of theory and methods for criterion testing and measurement. *Journal of Educational Measurement*, 11: 159-170.
- Hambelton, K. R. (1984). *Item Response Theory*. Boston: Kluwer Nijhof Publishing

- Hoyt, E. J. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6: 153-160.
- Hulin. C. L, & Drasgow, D. E. (1982) Recovery of two and three parametrics logistic item characteristic courses. *Applied Psychology Measurement*, 6: 153-160.
- Hulin, C. L. & Drasgow, D. E (1983). Item Response Theory. Homewood, Ill: Dow Jonse Boston.
- Levin. J. R. (1978). Correcting planning on experiment in the company of measurement errors. *Applied Psycholog Measurement*. 2: 382-85.
- Lord, M. F. (1980). Application of Item Response Theory to practical testing problems. New Jersey: Larrence Association Publisher.
- Lord, M. F & Novick, M. R. (1968). Theories of Mental Test Scores. Reading, Massachusset: Addison-Wesley.
- Mehrens, A. W & Kaminiski, D. C. (1989). Methods of improving standardised test scores. *Educational Measurement, Issues & Practic*, 8: 14-22.
- Mehrens, A. W & Lehman, J. C. (1984). *Measurement and evaluation in education and psychology*. New York: Rienhart & Winston, Inc.
- Magnuson, D. (1967). *Test Theory*. Boston: Addison - Wesley
- Romberg, A. T. (1992). Mathematic Assessment and evaluation. New York: State University of New York.
- Spearman. C. (1907). Demonstration of formula for true

measurement correlation. *American Psychology*, 18: 161-196.

Spearman, C. (1913). Correlation of sum and differences. *British Journal of Psychology*, 5: 17-42.

Stevens, S. S. (1946). *On the theory of scales of measurement*. New York: John Wiley.

Torgerson, W. S. (1958). Theory of scores of measurement. *Science*, 103: 677-680.

Weiss, J. D. (1981). Test theory and methods. *Annual Review of Psychology*, 32: 629-658.