

« فراسوی مدیریت »

سال دوم - شماره ۷ - زمستان ۱۳۸۷

ص ص ۲۴۴ - ۲۱۷

## معرفی و مقایسه تحلیلی نظریه آزمون کلاسیک و نظریه سوال-پاسخ در روانسنجی

دکتر غلامرضا گل محمد نژاد<sup>۱</sup>

### چکیده

این مقاله به معرفی مختصر و مقایسه تحلیلی دو نظریه عمده روانسنجی، یعنی نظریه آزمون کلاسیک (CTT) و نظریه سؤال-پاسخ (IRT) همراه با محسنات و محدودیت‌های هر کدام از آنها پرداخته است.

هرچند نظریه آزمون کلاسیک در اوایل قرن بیستم پایه‌ریزی شد اما به دلیل داشتن محدودیت‌های متعدد، همواره متخصصان را بر آن می‌داشت تا جهت رفع این محدودیت‌ها تلاش کنند. در فرآیند پیدا کردن راه حل‌هایی برای از بین بردن نقاط ضعف نظریه کلاسیک، نظریه جدیدی به نام نظریه سؤال-پاسخ در حدود دهه پنجاه شروع به رشد کرد و علی‌رغم این که سالهای بعد رفته رفته تکمیل شد و در ظاهر محدودیت‌های نظریه کلاسیک را نداشت اما خود این نظریه هم هیچگاه فارغ از محدودیت نبوده و همواره ایراداتی به آن وارد شده است.

هم اکنون هر کدام از این نظریه‌ها در بین متخصصان تربیتی و روانشناسی و افراد مرتبط با اندازه‌گیری‌های روانی-تربیتی، کم و بیش طرفدارانی دارد، اما در

---

- عضو هیئت علمی دانشگاه آزاد اسلامی تبریز (rezaegitim@yahoo.com)

کشور ما هیچ کدام از این نظریه‌ها آن گونه که باید و شاید شناخته شده نیست. به همین دلیل مقاله حاضر سعی کرده هر چند مختصر، به معرفی نظریه‌های مذکور پرداخته و نقاط ضعف و قوت همچنین وجوه اشتراک و افتراق هر کدام از آنها را ارائه کند.

### واژگان کلیدی:

پارامتر تشخیص، پارامتر دشواری، پارامتر مجانب، تابع آگاهی، تابع ویژه سؤال، نظریه‌های روانسنجی، نظریه سؤال-پاسخ، نظریه آزمون کلاسیک، منحنی ویژه سؤال.

### مقدمه

در اواخر قرن نوزدهم، رواج روش علمی در مطالعات انسانی موجب جدایی بسیاری از شاخه‌های دانش از فلسفه و پیدایش علوم جدیدی نظیر روانشناسی، علوم اجتماعی، تربیتی و مانند آن شد. یکی از ارکان اصلی روش علمی، جمع آوری داده‌ها و اندازه‌های مربوط به صفات و ویژگی‌های افراد است. بنابراین همزمان با رواج روش علمی، ضرورت تولید فنون و روشهای کمی سازی و اندازه‌گیری صفات مختلف آدمی بیشتر از قبل آشکار گردید. از این دوره بود که کم کم روش‌های آماری در علوم رفتاری به کار برده شد و روشهای نوین اندازه‌گیری در این علوم توسط ریاضی‌دانان و آماردانان مطرح و پایه‌های حوزه جدیدی در روانشناسی به نام روانسنجی ریخته شد.

واژه روانسنجی<sup>۱</sup> به حوزه‌ای از علم روانشناسی اشاره دارد که در آن با استفاده از روش‌ها و فنون آماری-ریاضی، صفات روانی مورد سنجش قرار می‌گیرند.

<sup>۱</sup>Psychometrics

بنیانگذار روانسنجی را بی تردید می توان فرانسیس گالتون<sup>۱</sup> دانست. او با مطالعه در زمینه وراثت و مقایسه توزیع دو متغیر مرتبط و محاسبه همبستگی بین آنها، به توسعه و کاربرد روش های آماری در مطالعات زیستی و روانی کمک شایانی کرده است.

بعداز گالتون دانشمندانی همانند پیرسون<sup>۲</sup>، اسپیرمن<sup>۳</sup>، فیشر<sup>۴</sup>، ترستون<sup>۵</sup>، گاست<sup>۶</sup>، کندال<sup>۷</sup> و گاتمن<sup>۸</sup>، که اغلب از شاگردان و همکاران و پرورش یافته مکتب او بودند، به اندازه گیری تفاوت های فردی علاقه نشان دادند و به توسعه روش های آماری - ریاضی پرداختند و کاربرد آنها را در اندازه گیری های روانی و تربیتی گسترش دادند.

در دهه ۱۸۹۰ جیمز مک کین کتل<sup>۹</sup> که یکی دیگر از پیروان گالتون بود اولین بار اصطلاح آزمون روانی<sup>۱۰</sup> را به کار برد. وی به کمک دانشجوی خود، کلارک ویسلر<sup>۱۱</sup>، تعداد زیادی اندازه گیری روانشناختی انجام داده و همبستگی آنها را با نمرات مختلف مورد بررسی قرار داد که این همبستگی ها تقریباً همگی نزدیک به صفر بودند. البته افراد دیگری نظیر گیلبرت<sup>۱۲</sup> نیز تلاش مشابهی انجام دادند که باز هم نتایج این مطالعات مأیوس کننده بود (گالیکسن ۱۹۵۰). این نتایج سبب ناامیدی و عدم اعتماد نسبت به آزمون های روانی شد و کاربرد و توسعه روش های مربوط به آنرا دچار رکود ساخت.

<sup>1</sup>Francis Galton

<sup>2</sup>Karl Pearson

<sup>3</sup>Charles Spearman

<sup>4</sup>Ronald A. Fisher

<sup>5</sup>Louis L. Thurstone

<sup>6</sup>William S. Gossett

<sup>7</sup>Kendall

<sup>8</sup>Louis Guttman

<sup>9</sup>James Mc Keen Cattell

<sup>10</sup>Mental Test

<sup>11</sup>Clark Wissler

<sup>12</sup>Gilbert

با آغاز قرن بیستم تلاش علمی آلفرد بینه<sup>۱</sup> در توسعه یک مقیاس هوش و ابداع روشهای آماری-ریاضی توسط پیرسون و اسپیرمن در زمینه اندازه گیری های روانی، اعتماد علمی نسبت به استفاده از نظریه آزمون روانی را دوباره احیا نمود و از این زمان به بعد بود که توسعه و پیشرفت اندازه گیری استعداد و موفقیت هم به صورت نظری و هم به صورت عملی شتاب گرفت و کم کم مقیاس های گوناگون دیگری در زمینه ویژگیهای شخصیتی و سایر زمینه ها توسعه یافت و بطور گسترده در آموزش و پرورش و صنعت به کار برده شد. این روند در حال حاضر نیز با سرعت بیش از پیش و زمینه های گسترده تر ادامه دارد.

به طور کلی در تاریخ روانسنجی دو نظریه عمده وجود دارد: یکی نظریه آزمون کلاسیک<sup>۲</sup> (CTT) و دیگری نظریه سوال-پاسخ<sup>۳</sup> (IRT).

مقدمات نظریه آزمون کلاسیک در اوایل قرن بیستم توسط اسپیرمن پایه ریزی شد و به مرور زمان تکمیل گردید، اما علی رغم تکامل آن، بعضی از محدودیت های آن هیچ وقت از بین نرفت تا متخصصان را بر آن داشت در راستای غلبه بر این محدودیت ها نظریه جدیدی تدوین کنند که به نام نظریه سوال-پاسخ مطرح است. البته نظریه سوال-پاسخ نیز فارغ از محدودیت نیست، اما تا حدودی محدودیت های نظریه کلاسیک در آن اصلاح شده است.

از آنجایی که توسعه روانسنجی در کشور ما هنوز در مراحل اولیه قرار دارد، در ارتباط با نظریه های سوال-پاسخ به جز چند ترجمه مختصر و گزارش برخی کارهای عملی و پایان نامه های کارشناسی ارشد که بر اساس نظریه های سوال-پاسخ انجام گرفته منابع فارسی معدود و اندکی برای معرفی و شناسایی این نظریه وجود دارد. در مورد نظریه کلاسیک هم علی رغم این که در دانشگاهها تدریس می شود و

<sup>1</sup>- Alfred Binet

<sup>2</sup>- Classical Test Theory

<sup>3</sup>- Item Response Theory

کارهای علمی و عملی زیادی هم بر اساس آن صورت گرفته، اما آنچنان که باید و شاید هنوز با این نظریه آشنایی کافی وجود ندارد.

این مقاله سعی دارد هر دو نظریه را هرچند به اختصار، معرفی نموده و نقاط اشتراک و افتراق آنها، همچنین نقاط ضعف و قوت هر کدام را همراه با مقایسه تحلیلی ارائه نماید تا زمینه برای شناسایی بهتر نظریه‌های مذکور فراهم گردد.

### نظریه آزمون کلاسیک (CTT)

نظریه آزمون کلاسیک یا سنتی، ابتدای قرن بیستم توسط پیشگامان روانسنجی - که شاخص‌ترین آنها اسپیرمن بود - پایه‌ریزی شد؛ اما بعدها فرم نهایی آن توسط نوویک<sup>۱</sup> تکمیل گردید (واندر لیندن و دیگران ۱۹۹۶). اساس این نظریه مبتنی بر معادله‌ای است که در آن نمره مشاهده شده (به دست آمده) به طور فرضی از یک نمره حقیقی و یک نمره خطا تشکیل شده است. در این مدل نمره حقیقی هر فرد ثابت است و در اندازه‌گیری‌های مکرر تغییری نخواهد کرد. آنچه متغیر خواهد بود خطاهای اندازه‌گیری است که یک متغیر تصادفی فرض می‌شود.

$$X_o = X_t + X_e \quad \text{این معادله اساسی بدین صورت است:}$$

در این معادله،  $X_o$  نمره مشاهده شده،  $X_t$  نمره حقیقی و  $X_e$  نمره خطاست (نمره خطا می‌تواند منفی و یا مثبت باشد).

در نظریه کلاسیک اعتبار اندازه‌گیری زمانی کامل خواهد بود که میزان نمره خطا صفر بوده و نمره مشاهده شده برابر با نمره حقیقی باشد. به عبارت دیگر هر قدر اندازه خطا کمتر باشد اعتبار نیز بالا خواهد بود. در این نظریه، بیشتر از خطاهای نامنظم بحث شده که بر خلاف عنوانشان از توزیع معینی برخوردار بوده و اعتبار را تحت تاثیر قرار می‌دهند؛ برعکس، خطاهای منظم که همگی منفی یا همگی مثبت هستند و روایی اندازه‌گیری را تحت تاثیر قرار می‌دهند، در این نظریه

<sup>۱</sup> Melvin R. Novick

کمتر مورد توجه واقع شده‌اند. خطاهایی که در این مقاله مد نظر بوده، خطاهای نامنظم (تصادفی) است.

معادله‌ای که در بالا به آن اشاره شد و اساس نظریه کلاسیک را تشکیل می‌دهد، خود بر مفروضه‌هایی استوار است که به نام مفروضه‌های زیر بنایی نظریه کلاسیک شناخته می‌شود. این مفروضه‌ها به صورت زیر است:

۱- اندازه مورد انتظار خطا صفر است.

اگر اندازه‌گیری بی‌نهایت بار تکرار شود و یا در مورد تعداد بی‌نهایت فرد اجرا شود میانگین خطاها صفر است.

۲- همبستگی بین اندازه‌های حقیقی و خطا صفر است.

۳- همبستگی بین خطاهای مربوط به دفعات متعدد اندازه‌گیری صفر است.

۴- در تکرار مجدد اندازه‌گیری، همبستگی بین اندازه‌های حقیقی در اندازه‌گیری اول و اندازه‌های خطا در اندازه‌گیری بار دوم و بالعکس صفر است (گالیکسن<sup>۱</sup> ۱۹۵۰، لرد<sup>۲</sup> و نوویک<sup>۳</sup> ۱۹۶۸، ثرندایک<sup>۳</sup> ۱۹۸۲).

شاخصهای مربوط به سوالات و کل تست در نظریه کلاسیک

در نظریه کلاسیک، شاخصهای تست همانند دشواری تست، میانگین، واریانس، اعتبار و روایی را می‌توان با استفاده از شاخصهای سوالات منتخب برآورد و تنظیم کرد. میانگین تست را می‌توان با انتخاب سوالاتی با درجه دشواری خاص به صورت دلخواه تنظیم نمود یا می‌توان از راه انتخاب سوالاتی با قدرت تشخیص و واریانس بالا، اعتبار و روایی تست را بالا برد. همچنین از راه افزایش تعداد سوال با نسبتی معین می‌توان اعتبار تست را افزایش داد.

<sup>1</sup>- Harold Gulliksen

<sup>2</sup>- Frederic M. Lord

<sup>3</sup>- Robert L. Thorndike

### درجه دشواری سوال

درجه دشواری سوال عبارت است از نسبت کسانی که به سوال جواب درست داده‌اند به کل افراد. درجه دشواری سوال از طریق زیر محاسبه می‌شود:

$$P_j = \frac{n_r}{N}$$

در این فرمول  $P_j$  درجه دشواری سوال  $j$ ،  $n_r$  تعداد کسانی که به سوال  $j$  جواب درست داده‌اند و  $N$  نیز تعداد کل افراد است. همچنین درجه دشواری سوال را می‌توان بر اساس انتخاب ۲۷٪ کل افراد از بالا به عنوان گروه قوی و ۲۷٪ از پایین به عنوان گروه ضعیف نیز محاسبه نمود که فرمول آن به صورت زیر است:

$$P_j = \frac{N_u + N_l}{U + L}$$

در فرمول بالا نیز  $N_u$  تعداد کسانی که در گروه قوی و  $N_l$ ، تعداد کسانی که در گروه ضعیف به سوال  $j$  جواب درست داده‌اند و  $U$  و  $L$  به ترتیب تعداد افراد گروه قوی و ضعیف را شامل می‌شود.

درجه دشواری سوال ارزش‌های بین صفر و یک را می‌تواند اختیار کند. اگر مقدار آن ۰/۵ یا نزدیک به آن باشد حاکی از درجه دشواری متوسط و هر قدر به یک نزدیک شود نشان از آسان بودن آن و هر اندازه به صفر نزدیک شود نشان‌دهنده دشواری سوال است.

### قدرت تشخیص سوال<sup>۱</sup>

قدرت تشخیص سوال از مفاهیمی است که هم در نظریه کلاسیک و هم در نظریه سوال-پاسخ از جمله مفاهیم محوری و اساسی به شمار می‌آید و بیان‌کننده قدرت سوال در تفکیک و تمییز افراد افراد قوی و ضعیف است. به عبارت دیگر هر قدر پاسخ افراد قوی و ضعیف به آن سوال، متفاوت باشد بیان‌کننده قدرت

<sup>۱</sup>Discrimination

تشخیص بالای آن سوال است. بالابودن قدرت تشخیص سوال خود نشان‌دهنده هماهنگی و همبستگی بالای آن سوال با کل تست است.

در نظریه کلاسیک، همبستگی سوال با کل تست شاخص هماهنگی و وجوه اشتراک آن سوال با تست (سایر سوالات) و در نتیجه شایستگی و مناسب بودن آن سوال است.

در نظریه کلاسیک، قدرت تشخیص سوال از طریق فرمول زیر به دست می‌آید:

$$D_j = P_u - P_l$$

در این فرمول،  $D_j$  قدرت تشخیص سوال  $j$ ،  $P_u$  نسبت کسانی که در گروه قوی (بالا) و  $P_l$  نسبت کسانی که در گروه ضعیف (پایین) به سوال مورد نظر پاسخ درست داده‌اند. فرمول قدرت تشخیص سوال به شکل زیر نیز نوشته می‌شود:

$$D_j = \frac{N_u - N_l}{U/L}$$

نمادهای فرمول مذکور همانند نمادهای گفته شده در فرمول درجه دشواری است.

### محدودیت‌های نظریه کلاسیک

با وجود این که نظریه کلاسیک سالهاست در مورد توسعه تست‌ها و اجرا، تحلیل و تفسیر نتایج آنها مورد استفاده قرار می‌گیرد دارای برخی از محدودیت‌هاست. از محدودیت‌های نظریه کلاسیک می‌توان به موارد زیر اشاره کرد:

۱- در این نظریه، شاخص‌های مربوط به تست و سوالات، وابسته به گروه نمونه است. سطوح توانایی گروه نمونه و توزیع آن، شاخص‌های سوالات مانند قدرت تشخیص، درجه دشواری، انحراف استاندارد، واریانس و میانگین تست را تحت تاثیر شدید قرار می‌دهد. به همین دلیل تعمیم نتایج به سایر گروه‌ها و جوامع با محدودیت روبرو می‌شود.



- ۲- آشکار شدن سطح توانایی افراد به تست و سوالات وابسته است. یک فرد ممکن است در دو تست که صفت واحدی را اندازه گیری می کنند اما به لحاظ سطح دشواری متفاوت هستند نمرات متفاوتی به دست آورد.
- ۳- تستی که بر اساس نظریه های کلاسیک توسعه یافته است بیشتر، افراد با سطح توانایی متوسط را مورد خطاب قرار می دهد. بدین معنی که اعتبار تست در مورد افراد گروه بالا و گروه پایین توانایی نسبت به گروه متوسط کم است.
- ۴- خطای استاندارد اندازه گیری در مورد همه افراد یکسان فرض می شود. این فرض موجب تصمیم گیری اشتباه در مورد افراد می شود زیرا خطای استاندارد بر اساس سطح توانایی می تواند متفاوت باشد.
- ۵- اعتبار تست به گروه نمونه وابسته است. در تعریف مفهوم اعتبار، اتفاق نظر وجود ندارد و روش های اعتباریابی نیز متفاوت است. با توجه به روش اعتباریابی، میزان اعتبار محاسبه شده نیز به صورت های مختلف قابل توجیه و تفسیر است
- ۶- درجه دشواری سوالات در نظریه کلاسیک برای همه سطوح توانایی یکسان فرض می شود و این در حالی است که آن می تواند با توجه به سطوح توانایی متفاوت باشد.
- ۷- در نظریه کلاسیک در صورت تحقق برخی شرایط از درجه دشواری سوال بعنوان احتمال کلی پاسخ درست افراد به یک سوال معین سخن گفته می شود که در این مورد نیز باید گفت چون احتمال پاسخ درست به سوال براساس سطح توانایی می تواند متفاوت باشد پس درجه دشواری نمی تواند به طور دقیق احتمال پاسخ صحیح به یک سوال معین در بین افراد با توانایی متفاوت باشد.
- ۸- در نظریه کلاسیک در صورتی که سطح توانایی افراد نزدیک هم باشد تست ها و سوالات آن کارایی لازم در تشخیص این سطوح از یکدیگر را نخواهند داشت.
- ۹- مفروضه های زیادی در نظریه کلاسیک وجود داشته و احتمال عدم تحقق همه این مفروضه ها در هر زمان وجود دارد.

علاوه بر محدودیت‌های مذکور، مواردی مانند تشخیص سوالات سوگیر و هم‌ترازسازی تست‌ها نیز آن‌چنان که باید و شاید از دقت بالایی برخوردار نیست.

### محسنات نظریه کلاسیک

با وجود موارد بالا که به عنوان محدودیت‌های نظریه کلاسیک شناخته می‌شود، این نظریه دارای محسناتی نیز هست.

همبلتون<sup>۱</sup> و جونز<sup>۲</sup> (۱۹۹۳) از جمله محسنات بزرگ نظریه کلاسیک را، قابلیت استفاده آن در مورد تست‌های مختلف می‌دانند. از محسنات دیگر آن می‌توان به تحلیل آسان نتایج تست‌ها و برآورد شاخصهای مربوط به تست‌ها و سوالات اشاره کرد که سبب شده نسبت به سایر نظریه‌ها مورد استفاده بیشتری پیدا کند.

نظریه کلاسیک در پیدایش نظریه‌های دیگر نیز نقش مهمی ایفا نموده است که پیدایش و توسعه نظریه سوال-پاسخ از آن جمله است. جهت از بین بردن محدودیت‌های نظریه کلاسیک تلاش‌هایی صورت گرفته که منجر به رشد نظریه‌های سوال پاسخ شده است.

### نظریه سوال - پاسخ (IRT)

اساس نظریه سوال-پاسخ، نظریه کلاسیک است. همان‌گونه که گفته شد تلاش علمای سنجش و اندازه‌گیری در راستای غلبه بر نارسایی‌ها و محدودیت‌های نظریه کلاسیک، منجر به توسعه نظریه سوال-پاسخ شده است.

سنگ بنای نظریه سوال-پاسخ در دهه پنجاه طی مقاله‌ای توسط لرد (۱۹۵۳) و چند سال بعد وسیله بیرن‌بام<sup>۳</sup> (۱۹۵۷، ۱۹۵۸) در قالب پژوهش و گزارش‌های فنی که در نیروی هوایی امریکا انجام داده بود گذاشته شد. از علمای دیگری که در

<sup>۱</sup>- Hambleton

<sup>۲</sup>- Jones

<sup>۳</sup>- Birnbaum

توسعه و رشد نظریه سوال-پاسخ نقش داشته‌اند می‌توان به جورج راش<sup>۱</sup> (۱۹۶۰) اشاره نمود. جورج راش یک ریاضی‌دان دانمارکی است که سالیان زیادی در مشاوره و تدریس آمار مشغول به کار بود. او خانواده‌ای از مدل‌های نظریه سوال-پاسخ برای بهبود سنجش خواندن و تهیه تست‌هایی برای استفاده در ارتش دانمارک به وجود آورد (امبرتسون و رایز<sup>۲</sup>، ۲۰۰۰).

نظریه سوال-پاسخ مبتنی بر نظریه خصیصه مکنون است. خصیصه مکنون به طور نظری یک صفت تک بعدی بنیادی مفروض است که توسط تست مورد اندازه‌گیری قرار می‌گیرد. این خصیصه می‌تواند یک صفت ساده یا ترکیبی از صفات مختلف باشد. همین طور می‌تواند تک بعدی یا چند بعدی باشد. ناگفته نماند که مرکب بودن خصیصه مکنون مغایر با مفروضه تک بعدی بودن آن نیست؛ چرا که در نظریه سوال-پاسخ آن (خصیصه مرکب) به عنوان یک سازه واحد که تعیین کننده پاسخ فرد به سوال است، شناخته می‌شود. هر چند مدل‌های ابتدایی بر مفروضه تک بعدی بودن استوار هستند، اما مدتی است مدل‌هایی بر اساس چند بعدی نیز توسعه یافته‌اند.

در نظریه خصیصه مکنون، موضوع مورد بحث رابطه بین سطوح مختلف این صفت (خصیصه مکنون) و سوال است. به عبارت دیگر این نظریه، احتمال پاسخ درست یک فرد به یک سوال را به سطح صفت بنیادی که فرد در آن قرار دارد به وسیله معادله‌ای ارتباط می‌دهد. این معادله بدین صورت است:

$$P_j = f(\theta_1, \theta_2, \theta_3, \dots, \theta_k)$$

در این معادله  $P_j$  احتمال موفقیت در سوال  $j$ ام و  $\theta$  سطح صفت بنیادی (توانایی) که دارای  $k$  بعد بوده و وسیله سوال، مورد اندازه‌گیری قرار می‌گیرد. در واقع احتمال موفقیت در سوال تابعی از جایگاه فرد در پیوستار صفت بنیادی

<sup>۱</sup>- Georg Rasch

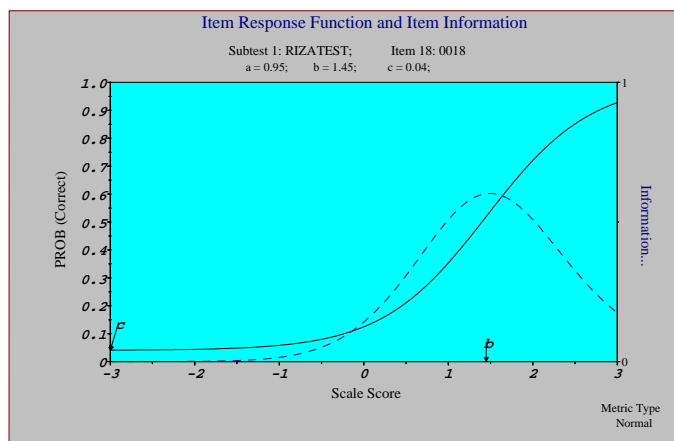
<sup>۲</sup>- Susan E. Embretson & Steven P. Reise

(توانایی) است. به همین دلیل بررسی رابطه بین سوال و صفت بنیادی از اهمیت زیادی برخوردار است.

در واقع، نظریه سوال-پاسخ، توسعه و به کارگیری مدل‌هایی بر اساس نظریه خصیصه مکنون، برای سوالات مختلفی است که توانایی را مورد اندازه‌گیری قرار می‌دهند. در این نظریه، توانایی اعم از هوش، استعداد، پیشرفت و مانند آن، به عنوان خصیصه مکنون یا همان صفت بنیادی شناخته می‌شوند.

### منحنی ویژه سوال<sup>۱</sup> (ICC)

در نظریه سوال-پاسخ هر سوالی دارای منحنی ویژه است که توزیع آن سوال در سطوح مختلف توانایی را نشان می‌دهد و تحت عنوان منحنی ویژه سوال (ICC) شناخته می‌شود. به طور نظری این منحنی که حالتی تک‌نوا داشته و افزایشی است، ارتباط بین احتمال پاسخ درست به سوال و سطوح مختلف توانایی را ترسیم می‌کند. البته امکان دارد در عمل، منحنی ویژه سوال تک‌نوا و یا افزایشی نباشد. به عنوان مثال در زیر نمونه‌ای از یک منحنی ویژه سوال (ICC) همراه با پارامترهای مربوط به آن نمایش داده شده است:



<sup>1</sup>- Item Characteristic Curve (ICC)

پارامترهای سوال نقش تعیین کننده در نحوه شکل گیری و پیدایش منحنی سوال ایفا می کنند. این پارامترها عبارتند از پارامتر دشواری<sup>۱</sup> ( $b_j$ )، پارامتر تشخیص<sup>۲</sup> ( $a_j$ ) و مقدار مجانب<sup>۳</sup> ( $c_j$ ) که احتمال پاسخ درست به سوال را در پایین سطح توانایی نشان می دهد و گاهی اوقات تحت نام پارامتر شانس معرفی می شود.

پارامتر دشواری سوال ( $b_j$ ) در واقع نشان دهنده سطح دشواری سوال بوده و عبارت است از سطحی از توانایی ( $\theta$ ) که در آن احتمال پاسخ به سوال برابر با  $\frac{1+c_j}{2}$  باشد. در صورتی که مقدار مجانب ( $c_j$ ) یا همان احتمال پاسخ درست شانسی صفر باشد در واقع پارامتر دشواری سوال ( $b_j$ ) عبارت خواهد بود از نقطه ای از سطح توانایی ( $\theta$ ) که احتمال پاسخ درست برای افرادی که در آن نقطه هستند ۵۰ درصد باشد. در مثال سوال بالا مقدار  $b$  برای آن سوال برابر با  $1/45$  است. این مقدار نشان می دهد کسانی که توانایی آنها به اندازه  $1/45$  است، یعنی در پیوستار ( $\theta$ ) در نقطه  $\theta = 1.45$  قرار دارند، با توجه به اینکه مقدار  $c$  برای این سوال مساوی  $0.04$  است، احتمال پاسخ درست در سوال مذکور برای آنها، برابر با  $\frac{1+0.04}{2} = 0.52$  خواهد بود.

پارامتر تشخیص ( $a_j$ ) که پارامتر شیب نیز گفته می شود به مفهوم قدرت سوال در تفکیک افراد در سطوح مختلف توانایی است و نشان می دهد که با افزایش سطح توانایی احتمال پاسخ درست با چه سرعتی افزایش می یابد. هنگامی که قدرت تشخیص سوال بالا باشد منحنی ویژه سوال بیشتر به حالت عمودی و هنگامی که پایین باشد به حالت افقی نزدیک خواهد شد. معادله مربوط به پارامتر تشخیص سوال به صورت زیر است:

<sup>1</sup>- Difficulty parameter

<sup>2</sup>- Discrimination parameter

<sup>3</sup>- Asymptote

$$a_j = \frac{r_{bisj}}{\sqrt{1 - r_{bisj}^2}}$$

در این معادله  $r_{bisj}$  ضریب همبستگی دو رشته‌ای سوال با صفت بنیادی بوده که دارای توزیع نرمال است. مقدار  $a$  در سوال بالا در کل برابر با ۰/۹۵ بوده و همان گونه که مشاهده می‌شود شیب منحنی در سطوح بالای توانایی مناسب است. بدین معنی که این سوال، افراد با سطح توانایی بالای ۰/۰۰ ( $\theta > 0.00$ ) را بهتر از افراد با سطح توانایی پایین ( $\theta < 0.00$ ) تفکیک می‌کند. این سوال در سطوح بالا آگاهی بیشتری نسبت به سطوح پایین داده و به همین جهت برای سطوح بالا مناسب‌تر از سطوح پایین خواهد بود.

نماد  $c_j$  همان مجانب پایین‌تر منحنی ویژه سوال است که احتمال موفقیت در پاسخ‌گویی به سوال در پایین‌ترین سطح توانایی را نشان می‌دهد. به عبارت دیگر این پارامتر شاخص احتمال موفقیت تصادفی در سوال است. مقدار  $c_j$  از داده‌های مربوط به امتحان شوندگانی که در پایین‌ترین سطح توانایی قرار دارند به دست می‌آید و از آنجایی که این افراد سوالات تست را با استفاده از شانس و حدس پاسخ می‌دهند، این پارامتر با نام پارامتر شانس و حدس نیز شناخته می‌شود. مقدار  $c_j$  در سوالات تشریحی بطور طبیعی برابر با صفر است و در سوالات گزیده-پاسخ نیز هرچه تعداد گزینه‌ها زیاد باشد به صفر گرایش پیدا خواهد کرد؛ چرا که احتمال پاسخ درست شانسی در سوالاتی که دارای گزینه‌های زیادی باشد پایین می‌آید.

همان‌طور که در منحنی ویژه سوال بالا نیز مشاهده می‌شود، مقدار مجانب برابر ۰/۰۴ است ( $c = 0.04$ ) که نشان می‌دهد احتمال پاسخ درست به این سوال در پایین‌ترین سطوح توانایی نزدیک به صفر است؛ یعنی حتی با استفاده از شانس هم نمی‌توان به این سوال پاسخ درست داد.

مقدار مجانب ( $c_j$ ) با پارامتر دشواری ( $b_j$ ) معمولاً یک ارتباط معکوس دارد. بدین ترتیب که هر اندازه مقدار  $b_j$  بالا باشد مقدار  $c_j$  پائین خواهد بود. به عبارت دیگر هر مقدار سوال دشوارتر باشد احتمال موفقیت تصادفی یا همان مقدار مجانب نیز پایین خواهد بود.

### مدل‌های تک بعدی نظریه سوال-پاسخ

به غیر از تک بعدی بودن خصیصه مکنون، نوع داده‌های به دست آمده از سوالات نیز در تنوع مدل‌های نظریه سوال-پاسخ دخیل‌اند. داده‌های حاصل از سوالات به چهار دسته داده‌های دو ارزشی<sup>۱</sup>، چند ارزشی<sup>۲</sup>، داده‌های پیوسته<sup>۳</sup> و طبقه‌ای<sup>۴</sup> (داده‌های طبقه‌ای هم‌زمان گسسته<sup>۵</sup> نیز محسوب می‌شوند) تقسیم می‌گردند. نظریه سوال-پاسخ در ابتدا شامل داده‌های دو ارزشی (صفر و یک) و تست‌های توانایی و استعداد می‌شد. برای سوالاتی که به صورت دو ارزشی نمره‌گذاری می‌شوند مدل‌های یک، دو و سه پارامتری مبتنی بر دو مدل اجایو نرمال<sup>۶</sup> و تابع منطقی<sup>۷</sup> وجود دارد که با توجه به تعداد پارامترهای سوال تعریف می‌شوند و تابع مربوط به هر کدام از مدل‌ها نیز تعداد معینی از پارامترهای سوال که قبلاً تعریف شده بود (پارامترهای  $a, b, c$ ) را شامل می‌شود.

چنان که در بالا نیز اشاره شد در نظریه سوال-پاسخ، تست‌هایی که تک بعدی هستند و پاسخ‌های آن به صورت دو ارزشی (صفر و یک) نمره‌گذاری شده باشند، می‌توانند بر اساس تعداد پارامتر در یکی از مدل‌های یک، دو و یا سه پارامتری که هر کدام از آنها نیز به دو بخش لجستیک (منطقی) و او جایو نرمال تقسیم می‌شوند

<sup>1</sup>Dichotomous

<sup>2</sup>Multichotomous

<sup>3</sup>Continuous

<sup>4</sup>Categorical

<sup>5</sup>Discrete

<sup>6</sup>Normal Ogive

<sup>7</sup>Logistic Function

مورد بررسی و تحلیل قرار گیرند. به دلیل گستردگی مطالب، در این مقاله تنها به معرفی مختصر سه مدل از مدل‌های منطقی (لجستیک) پرداخته شده است:

#### ۱- مدل تک پارامتری:

این مدل ساده‌ترین مدل در نظریه سوال-پاسخ بوده که از طرف راش ارائه شده است و مورد خاصی از مدل دو و سه پارامتری بوده که در آن فرض بر این است که قدرت تشخیص همه سوالات با هم برابر و احتمال پاسخ شانس نیز حداقل است. در مدل تک پارامتری، به غیر از سطوح توانایی ( $\theta$ )، پارامتر دشواری ( $b_j$ ) تنها متغیری است که در تابع منطقی سوال نقش دارد. این تابع به صورت زیر است (همبلتون و سووامیناتان، ۱۹۸۵):

$$P_j(\theta) = \frac{e^{D(\theta-b_j)}}{1 + e^{D(\theta-b_j)}}$$

در این معادله،  $P_j(\theta)$  احتمال پاسخ صحیح فردی با سطح توانایی معین ( $\theta$ ) که به طور تصادفی انتخاب شده، به سوال  $j$ ام است. مقادیر مربوط به سطوح توانایی یا  $\theta$ ، بطور نظری از  $-\infty$  تا  $+\infty$  می‌تواند متغیر باشد اما در عمل ارزش‌های آن معمولاً بین ۳- تا ۳+ مد نظر قرار می‌گیرد.  $b_j$  پارامتر دشواری سوال  $j$ ام و برابر با نقطه‌ای در سطوح توانایی ( $\theta$ ) بوده که احتمال پاسخ درست افراد واقع در آن نقطه به آن سوال ۵۰ درصد است (البته با فرض  $c_j = 0$ ) و بالاخره  $D$  و  $e$ ، مقادیر ثابت تابع و به ترتیب برابر با  $1/7$  و  $2/7183$  هستند.

محدودیت عمده مدل تک پارامتری، فرض برابری قدرت تشخیص همه سوالات است که در موارد عملی شواهد محکمی وجود دارد که به جز زمانی که سوالات با آگاهی قبلی انتخاب می‌شوند امکان‌پذیر نیست. امتیاز این مدل نیز سهولت محاسبات نسبت به مدل‌های دو و سه پارامتری است.



## ۲- مدل دو پارامتری:

در مقابل مدل تک پارامتری، که تنها بر اساس پارامتر دشواری ( $b_j$ ) استوار است، مدل دو پارامتری که از طرف بیرن بام (۱۹۵۷) پیشنهاد شده، علاوه بر پارامتر دشواری، پارامتر تشخیص ( $a_j$ ) را نیز در بر می گیرد. تابع منطقی سوال در این مدل به این شکل است (همبلتون و سووامیناتان، ۱۹۸۵):

$$P_j(\theta) = \frac{e^{Da_j(\theta-b_j)}}{1+e^{Da_j(\theta-b_j)}}$$

در این تابع،  $a_j$  همان پارامتر تشخیص سوال بوده و بقیه نمادهای این تابع همانند مدل تک پارامتری است.

به طور نظری مقدار  $a_j$  می تواند از  $-\infty$  تا  $+\infty$  متغیر باشد، اما در عمل و هنگام محاسبه، اکثر مقادیر بین  $-2/8$  تا  $+2/8$  را اختیار می کند. اگر مقدار  $a_j$  بالا باشد نشان دهنده واریانس بالای سوال و به نوعی نشان دهنده اعتبار بالای سوال است و در صورتی که به صفر متمایل شود نشان می دهد سوال قدرتی در تشخیص سطوح مختلف افراد ندارد (در این حالت شکل منحنی ویژه سوال به صورت افقی خواهد بود) و در صورتی که منفی باشد نشان از اندازه گیری خصیصه ای متضاد و یا به غیر از خصیصه مورد اندازه گیری کل تست است (در این حالت نیز شکل منحنی ویژه سوال به صورت نزولی خواهد بود). بنابر این در شرایطی که  $a_j \leq 0$  باشد سوال با مدل همخوانی نخواهد داشت. یکی از مفروضه های این مدل، صفر یا حداقل بودن احتمال پاسخ شانس به سوال است یعنی  $c_j = 0$ . در صورتی که  $a_j < 0$  باشد مفروضه مذکور ( $c_j = 0$ ) نیز خدشه دار خواهد شد. بیکر<sup>۱</sup> (۲۰۰۱) توصیه می کند سوالاتی که قدرت تشخیص آنها زیر  $0/8$  ( $a_j < 0.8$ ) است نباید در فرم نهایی تست گنجانده شده و مورد استفاده قرار گیرند.

<sup>۱</sup>Frank B. Baker

**۳- مدل سه پارامتری:**

هنگامی که احتمال پاسخ شانس به سوال وجود داشته باشد ( $c_j > 0$ ) مدل یک و دو پارامتری مناسب نخواهد بود. در این شرایط مدل سه پارامتری که در آن پارامتر شانس نیز به تابع سوال اضافه شده مناسب خواهد بود. مدل سه پارامتری با علاوه کردن پارامتر شانس ( $c_j$ ) به مدل دو پارامتری به دست می آید. تابع منطقی سوال در مدل سه پارامتری به صورت زیر است (همبلتون و سووایناتان، ۱۹۸۵):

$$P_j(\theta) = c_j + (1 - c_j) \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}}$$

در این تابع همه نمادها همانند مدل دو پارامتری است و  $c_j$  نیز احتمال یا سطح پاسخ شانس به سوال یا کمترین مقدار مجانب در منحنی ویژه سوال بوده و نشان دهنده احتمال پاسخ درست به سوال در بین افرادی که در پایین ترین سطح توانایی قرار دارند است.

در مورد سوالاتی که احتمال پاسخ درست شانس آنها بالاست مدل سه پارامتری مناسب تر از مدل دو پارامتری خواهد بود، اما در مورد سوالات باز پاسخ، تست های عملی یا سوالاتی که احتمال پاسخ شانس آنها پایین باشد (نزدیک صفر باشد) از مدل دو پارامتری به جای سه پارامتری می توان استفاده کرد.

**مفروضه های نظریه سوال- پاسخ**

کاربرد مدل های نظریه سوال- پاسخ نیازمند تحقق برخی مفروضه ها است. هر چند در عمل این مفروضه ها به سختی تحقق پیدا می کند، اما از آنجایی که این مفروضه ها اساس نظریه مذکور است، تحقق آنها (یا حداقل تحقق تقریبی این مفروضه ها) شرط استفاده از مدل های مربوط به این نظریه محسوب می شود. این مفروضه ها به این صورت است:

**الف - تک بعدی بودن :**

یکی از مفروضه‌های نظریه سوال-پاسخ این است که صفت مورد اندازه‌گیری وسیله تست یک صفت تک بعدی است یا سوالات تست به طور مشترک تنها یک بعد از یک خصیصه مکنون را اندازه‌گیری میکنند.

تحقق تک بعدی بودن تست امر بسیار دشواری است. در شرایطی هم که این امر تحقق پیدا می‌کند و تست یک خصیصه تک بعدی را مورد اندازه‌گیری قرار می‌دهد شناسایی و تعریف این بعد مورد اندازه‌گیری دشوار خواهد بود (کراکر<sup>۱</sup> و الجینا<sup>۲</sup> ۱۹۸۶).

**ب- استقلال موضعی:**

یکی دیگر از مفروضه‌های نظریه سوال-پاسخ، مفروضه استقلال موضعی سوالات است. استقلال موضعی در واقع نتیجه تک بعدی بودن تست است. به همین خاطر مفروضه تک بعدی بودن و استقلال موضعی از طرف متخصصان سنجش و اندازه‌گیری تحت عنوان واحدی مورد بررسی قرار می‌گیرد (لرد ۱۹۸۰: همبلتون و سووامیناتان ۱۹۸۵: الجینا و کراکر ۱۹۸۶).

استقلال موضعی در واقع بیان‌کننده این است که پاسخ به یک سوال تأثیری در چگونگی پاسخ به سوال دیگر نداشته باشد. بعبارت دیگر پاسخ به یک سوال تعیین‌کننده پاسخ فرد به سوال دیگر نباشد. در غیر این صورت تابع احتمال سوال در نظریه‌های سوال-پاسخ توجیه‌پذیر و اساساً درست نخواهد بود. متخصصان استقلال موضعی سوال را از زاویه‌ای دیگر نیز توضیح داده‌اند. از این زاویه، استقلال موضعی سوال در یک سطح از سطوح توانایی به میزان احتمال یکسان پاسخ‌گویی به هر کدام از سوالات تست در بین افراد آن سطح توانایی اشاره دارد. این مفهوم از استقلال موضعی بیشتر با قاعده تک بعدی بودن تست همپوشی دارد.

<sup>1</sup>- Crocker

<sup>2</sup>- Algina

به فرض این که تست دو بعدی باشد، در بین افرادی که در یک سطح خاصی از توانایی ( $\theta$ ) هستند ممکن است گروهی در بعد اول و گروهی دیگر در بعد دوم توانا تر باشند. حال ممکن است احتمال پاسخ‌گویی در میان افراد گروه اول نسبت به سوال  $j$  که بیشتر با بعد اول ارتباط دارد بیشتر از احتمال پاسخ‌گویی گروه دوم باشد که نسبت به گروه اول در بعد اول ضعیف‌تر هستند. بدین ترتیب است که گاهی اوقات متخصصان مفهوم تک بعدی بودن تست و استقلال موضعی سوال را زیر یک چتر واحد بررسی می‌کنند.

### ج- نرمال بودن توزیع داده‌ها:

نظریه سوال-پاسخ بر نرمال بودن توزیع اندازه‌ها استوار است، بنابراین یکی دیگر از مفروضه‌های این مدل، نرمال بودن توزیع صفت مورد اندازه‌گیری و اندازه‌های به دست آمده است.

### تابع آگاهی سوال<sup>۱</sup>

یکی از مفاهیمی که در نظریه سوال-پاسخ در مورد دقت سوال در تشخیص و برآوردهای مربوط به سطوح توانایی، بطور فراوان مورد استفاده قرار می‌گیرد مفهوم تابع آگاهی سوال است.

تابع آگاهی سوال، میزان دقت، آگاهی و اطلاعات سوال درباره سطح خاصی از توانایی را نشان می‌دهد که بیشینه مقدار آن در نقطه‌ای خواهد بود که معادل  $b_j$  باشد. به طور نظری تابع آگاهی سوال با قدرت تشخیص سوال رابطه مستقیم دارد. بدین معنی که در هر نقطه از پیوستار توانایی میزان قدرت تشخیص سوال بالا باشد تابع آگاهی در آن بالا بوده و هر اندازه قدرت تشخیص پایین باشد تابع آگاهی نیز پایین خواهد بود. در کنار منحنی ویژه سوال بالا همان طور که دیده می‌شود یک منحنی دیگر به صورت نقطه چین وجود دارد. در واقع این خم میزان آگاهی سوال در هر نقطه از پیوستار توانایی را نشان می‌دهد. به طوری که ملاحظه می‌شود بیشینه

<sup>1</sup>Item Information Function

مقدار آن در نقطه‌ای معادل  $1/45$  بوده که معادل با پارامتر دشواری سوال ( $b = 1/45$ ) است.

### توسعه تست<sup>۱</sup>

یکی دیگر از موضوعاتی که در روانسنجی کاربردی وجود دارد توسعه تست و انتخاب سوال مناسب است. ممکن است روانشناسان، متخصصان علوم تربیتی و حوزه‌های دیگر علمی برای هدفهای مختلف و متنوع خود درصدد تهیه تست یا بانکی از سوالات باشند که برای اهداف کاربردی آنان مناسب‌ترین بوده و بالاترین کارایی را داشته باشد. به همین منظور لازم خواهد بود بعد از تهیه مجموعه مقدماتی از سوالات و تشکیل فرم اولیه‌ای از تست، آن بر روی نمونه گسترده‌ای از افراد اجرا شود. نکته حائز اهمیت است که بعد از اجرای فرم مقدماتی تست وجود خواهد داشت تصمیم در مورد حذف یا گنجاندن سوال در فرم نهایی خواهد بود. درحالی که در نظریه کلاسیک تصمیم در مورد حذف یا گنجاندن سوال در فرم نهایی، بستگی به هماهنگی آن سوال با کل تست دارد در نظریه سوال-پاسخ، تصمیم در مورد بالا و مناسب بودن سوال (برازندگی سوال)، بستگی به هماهنگی آن با مدل مربوطه دارد.

### امتیازات (محسنات) نظریه سوال-پاسخ

در شرایطی که مفروضه‌های زیر بنایی مدل‌های نظریه سوال-پاسخ تحقق پیدا کند این نظریه‌ها نسبت به نظریه‌های کلاسیک از امتیازاتی برخوردارند. از آنجایی که مدل‌های نظریه سوال-پاسخ بر پایه پارامترهای سوال استوار است و به خاطر آن که پارامترها به جامعه مربوط می‌شوند (از نمونه مستقل هستند)، این ویژگی امتیازی برای این نظریه محسوب می‌شود. در حالی که در نظریه‌های کلاسیک مشخصه‌های سوال همانند درجه دشواری (p) به گروه نمونه وابسته

<sup>1</sup> Test Development

است. مثلاً اگر گروه نمونه را افراد با توانایی بالا تشکیل می‌دهند درجه دشواری سوال، بالا خواهد بود و اگر گروه نمونه از افراد با توانایی پایین تشکیل شده باشد درجه دشواری، پایین خواهد بود. همچنین، در صورتی که گروه نمونه از واریانس پایین برخوردار باشد قدرت تشخیص سوال و اعتبار تست پایین بوده و در صورتی که گروه نمونه دارای واریانس بالا باشد قدرت تشخیص سوال و اعتبار تست بالا خواهد بود. علاوه بر آن در نظریه‌های کلاسیک سطح توانایی افراد ممکن است از تستی به تست دیگر تغییر یابد. به عبارتی دیگر سطح توانایی افراد وابسته به ویژگی‌های تستی است که روی آنها اجرا می‌شود؛ اما در نظریه‌های سوال-پاسخ، نه پارامترهای سوال وابسته به گروه است و نه سطح توانایی افراد وابسته به تستی است که بر روی آنها اجرا می‌شود. همین طور تابع آگاهی سوالات و تست که معادل مفهوم اعتبار در نظریه‌های کلاسیک است در نظریه سوال-پاسخ به گروه نمونه وابسته نیست.

### محدودیت‌های نظریه سوال-پاسخ

نظریه سوال-پاسخ علی‌رغم محسناتی که دارد دارای محدودیت‌هایی نیز هست که مهم‌ترین آنها، تحقق مفروضه‌های زیر بنایی است. به خاطر همین محدودیت، استفاده از این نظریه زیاد گسترده نیست. یکی دیگر از محدودیت‌های این نظریه، پیچیدگی توابع مربوط به خصوصیات و ویژگی‌های سوال و برآورد پارامترها و سطوح توانایی است. به همین علت در صورت وجود پارامترهای زیاد یا چند ارزشی بودن سوال، استفاده از برنامه‌های کامپیوتری غیر قابل گریز است. این محدودیت‌ها یا مواردی مانند آن موجب شده تا این نظریه در مقایسه با نظریه کلاسیک کاربرد کمتری داشته باشد.

### مقایسه دو نظریه کلاسیک و سوال-پاسخ به لحاظ توسعه تست

با وجود پیشرفته بودن مدل سوال-پاسخ نسبت به مدل کلاسیک، اکثر روانشناسان و افراد مرتبط با تعلیم و تربیت برای توسعه تست‌ها بیشتر از مدل‌های کلاسیک استفاده کرده و می‌کنند. در این مدل‌ها، به اندازه تفاوت در ترسیم نیمرخ روانی افراد بررسی و توسعه تست‌ها نیز متفاوت است. اما هر دو مدل بر جواب‌های درست و اشتباه افراد به سوالات استوار است.

با این که در حدود نیم قرن از مطرح شدن نظریه سوال-پاسخ می‌گذرد استفاده از آن بعد از بوجود آمدن کامپیوتر و توسعه برنامه‌های کامپیوتری شروع شده است.

در مقایسه نظریه سوال-پاسخ و کلاسیک می‌توان به این موارد اشاره کرد:

۱- فرض نرمال بودن صفت مورد اندازه‌گیری و خطای اندازه‌گیری در هر دو نظریه مشترک است. در صورت تحقق این فرض پیش‌بینی درجه دشواری سوال از طریق قدرت تشخیص سوال و بالعکس امکان‌پذیر خواهد شد (لرد و نوویک ۱۹۶۸؛ کراکر والجینا ۱۹۸۶).

۲- در نظریه‌های کلاسیک شاخص‌های مربوط به سوالات به گروهی که تست روی آنها اجرا شده بستگی دارد (از گروهی به گروه دیگر ممکن است این شاخص‌ها متفاوت باشد) اما در نظریه سوال-پاسخ این ادعا وجود دارد که پارامترهای سوال از گروه و حتی جامعه مستقل است. به عبارت دیگر این پارامترها در همه جوامع یکسان تلقی می‌شود.

۳- در نظریه کلاسیک سطح توانایی و استعداد افراد (نمراتی که کسب می‌نمایند) به سطح دشواری تست و سوالات آن بستگی دارد، اما در نظریه سوال-پاسخ می‌توان سطح توانایی افراد را مستقل از سطح دشواری تست و سوالات آن برآورد کرد (لرد و نوویک ۱۹۶۸).

۴- در نظریه‌های کلاسیک، میزان اعتبار تست برای همه افراد به صورت یکسان در نظر گرفته می‌شود. در حالی که در نظریه سوال-پاسخ برای هر کدام از سطوح توانایی از طریق محاسبه میزان تابع آگاهی هر سوال در آن سطح از توانایی و محاسبه مجموع توابع آگاهی سوالات که تابع آگاهی تست در آن سطح از توانایی محسوب می‌شود به عنوان اعتبار تست در آن سطح مد نظر قرار می‌گیرد. هر قدر تابع آگاهی تست بالا باشد می‌توان گفت اعتبار تست در آن سطح از توانایی بالاست.

در نظریه‌های سوال-پاسخ برای هر کدام از افراد نیز با توجه به این که چه ترکیبی از سوالات تست را پاسخ درست یا غلط داده‌اند، میزان اعتبار و خطای اندازه‌گیری جداگانه محاسبه می‌گردد. مثلاً ممکن است در مورد دو نفر که به تعداد مساوی پاسخ درست دارند اما ترکیب مختلفی از سوالات را پاسخ داده‌اند، اعتبار و خطای اندازه‌گیری متفاوتی محاسبه شود.

۵- در نظریه‌های کلاسیک اعتبار تست از گروهی به گروه دیگر و از جامعه‌ای به جامعه دیگر ممکن است متفاوت باشد، اما در نظریه سوال-پاسخ ادعا می‌شود تابع آگاهی هر سوال و در نتیجه کل تست، از جامعه‌ای به جامعه دیگر یکسان است.

۶- همانند اعتبار، خطای استاندارد اندازه‌گیری نیز در نظریه‌های کلاسیک برای همه افراد یکسان محاسبه می‌شود، اما در نظریه سوال-پاسخ با توجه به تابع آگاهی هر سطح از توانایی، برای آن سطح خطای استاندارد اندازه‌گیری متفاوت از سطوح دیگر محاسبه می‌شود. البته همان گونه که در بالا نیز گفته شد در نظریه سوال-پاسخ برای هر فرد نیز با توجه به این که چه ترکیبی از سوالات را پاسخ درست یا غلط داده، جدا از افراد دیگر اعتبار و خطای استاندارد اندازه‌گیری محاسبه می‌شود.



۷- در مورد بررسی تورش و تابع سوگیری سوال و کل تست نسبت به یک گروه خاص<sup>۱</sup> به نظر می‌رسد نظریه سوال-پاسخ از امتیاز بیشتری برخوردار باشد. اما تحقیقات نشان داده است نتایج به دست آمده در مورد بررسی مسأله تورش و تابع سوگیری سوال و کل تست نسبت به یک گروه خاص، در نظریه سوال-پاسخ و کلاسیک تقریباً مشابه است.

۸- در انتخاب سوالات مناسب برای قرار دادن در فرم نهایی تست در نظریه کلاسیک، قدرت تشخیص، واریانس سوال و همبستگی سوال با کل تست به عنوان پایه و اساس انتخاب آن سوال مد نظر قرار گرفته می‌شود؛ در حالی که در نظریه سوال-پاسخ ارزش پارامترهای سوال، تابع آگاهی سوال و متناسب بودن سوال با مدل به عنوان ملاک انتخاب آن سوال شناخته می‌شوند. علاوه بر آن در نظریه سوال-پاسخ ممکن است فرم نهایی تست جهت سنجش سطح خاصی تهیه گردد، بنابراین تابع آگاهی سوال در آن سطح مورد اهمیت و توجه خواهد بود.

۹- در نظریه سوال-پاسخ مشخصات سوال بیشتر از نظریه کلاسیک در مرکز توجه قرار دارد، در حالی که در نظریه کلاسیک به مشخصات تست اهمیت بیشتری داده می‌شود.

۱۰- در نظریه کلاسیک نمره حقیقی و در نظریه سوال-پاسخ،  $(\theta)$  یا همان سطح توانایی، مشخص و قابل مشاهده نیست، بلکه از نمرات مشاهده شده و عملکرد افراد تخمین زده می‌شود. در هر دو نظریه فرض وجود رابطه خطی بین متغیرهای مشاهده شده و غیر قابل مشاهده یک فرض زیربنایی و اساسی محسوب می‌شود. این فرض پایه و اساس همه مراحل رشد و توسعه تست در هر دو نظریه است.

<sup>1</sup>Differential Item Function (DIF)

### خلاصه بحث

از آنجا که هر اندازه مفروضه‌های یک نظریه کمتر باشد اعتبار آن نظریه بالاست و از آنجایی که مفروضه‌های نظریه کلاسیک از نظریه سوال-پاسخ بیشتر است، بنابر این اغلب متخصصان، نظریه سوال-پاسخ را معتبرتر از نظریه کلاسیک می‌دانند و کاربرد آن را بر نظریه کلاسیک ترجیح می‌دهند.

البته لازم به یادآوری است که مفروضه‌های مربوط به نظریه کلاسیک تنها در مورد اندازه مشاهده شده، اندازه حقیقی و اندازه خطا خلاصه می‌شود  $(X_o = X_t + X_e)$  و خود نظریه سوال-پاسخ ادامه همین نظریه کلاسیک است. در نتیجه، مفروضه‌های مذکور در نظریه کلاسیک، در قالب خصیصه مکنون در نظریه سوال-پاسخ نیز وجود دارد. و آخر این که به دلیل وجود مفروضه‌های مربوط به پیاده کردن نظریه سوال-پاسخ و پیچیدگی توابع و فرمولهای آن، عملاً استفاده از آن با محدودیت روبرو بوده و کمتر استفاده می‌شود. بنابر این در انتقاداتی که از نظریه کلاسیک می‌شود خود نظریه سوال-پاسخ نیز سهم است.

## منابع:

- Baker, F. (2001). The Basic of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F., & Kim, Seok-Ho. (2004). Item Response Theory Parameter Estimation Techniques. New York: Marcel Dekker, Inc.
- Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. Research Bulletin 81-20. Princeton, NJ: Educational Testing Service.
- Birnbaum, A. (1957). Efficient design and use of tests of a mental ability for various decision-making problems. Series Report No. 58-16. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas: January.
- Birnbaum, A. (1958a). Further considerations of efficiency in tests of a mental ability. Technical Report No. 17. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas.
- Birnbaum, A. (1958b). On the estimation of mental ability. Series Report No. 15. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas: January.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores. Reading, Mass: Addison-Wesley.
- Crocker, L., & Algina, J. (1986). Introduction to classical & modern test theory. Holt, Rinehart and Winston, Inc. University of Florida.
- Embretson, S. E.; & Riese, S. P. (2000). Item response theory for psychologists. New jersey: Lawrence Erlbaum associates, Mahwah.
- Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.
- Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. Applied Psychological Measurement, 10, 287-302.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 147-200). New York: American Council on Education, Macmillan.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to

- Test Development. Educational measurement: Issues and Practices, 12,3,38-47.
- Hambleton, R. K., & Swaminathan, H. (1985). Item Response Theory: Principles and Application. Kluwer-Nijhoff Publishing, Boston.
  - Hambleton, R. K., & Swaminathan, H. & Roger, J. H. (1991). Fundamentals of Item Response Theory. Sage Publishing, Boston.
  - Harvey, R. J., & Hammer, A. L. (1999). Item Response Theory ; MYERS-Briggs Type Indicator. Counseling Psychologist, 27,3,353-374.
  - Henrysson, S. (1971). Gathering, analyzing and using data on test items. In R. L. Thorndike (Ed.). Educational Measurement (2nd Ed.). Washington, D.C.: American Council on Education.
  - Lord, F. M. (1952). A theory of test scores. Psychometric monograph, No. 7.
  - Lord, F. M. (1953). The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 13, 517-548.
  - Lord, F. M. (1980). Application of item response theory to practical testing problems. Hillsdale, New jersey: Lawrence Erlbaum Associates.
  - Lord, F. M.; & Novick, M. R. (1968). Statistical theories of Mental test scores. Reading, Mass: Addison-Wesley.
  - Ostini, Remo.; & Nering, Michael L. (2006). Polytomous Item Response Thoery Models. Sage. Inc.
  - Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
  - Rasch, G. (1966a). An item analysis which takes individual differences into account. British journal of Mathematical and Statistical Psychology, 19, 49-57.
  - Rasch, G. (1966b). An individualistic approach to item analysis. In P. Lazarsfeld, & N. V. Henry (Eds.), Readings in Mathematical social science. Chicago: Science Research Associations, 89-107.
  - Thorndike, R. L. (1982). Applied Psychometrics. Houghton Mifflin Company, Boston.
  - Van der Linden, W. J., & Hambleton, R. K. (1996). Handbook of modern item response theory. New York: Springer.