

# نظریه سؤال - پاسخ در آزمونهای روانی - تربیتی

## دکتر حسین سپاسی\*

### چکیده

در این مقاله سعی شده است تا نظریه جدید آزمون سازی موسوم به نظریه سؤال - پاسخ (Item Response Theory) مورد بحث و بررسی قرار گیرد. این نظریه عمدتاً به دلیل اشکالاتی به وجود آمده است که روانسجان جدید به نظریه کلاسیک آزمون، مطرح کرده‌اند. این نظریه نمره کی آزمون را برای تفسیر عملکرد آزمودنی مدنظر قرار نمی‌دهد. بلکه هر سؤال آزمون، به‌طور انفرادی مورد مطالعه قرار می‌گیرد. این نظریه از منحنی ویژگی سؤال و پدیده‌هایی چون دشواری و تغییر، برای برآورد احتمال پاسخ درست آزمودنی به هر سؤال آزمون استفاده می‌کند. به‌علاوه، این نظریه بر پایه دو فرض اساسی استثنای موضعی و تک‌بعدی بودن فضای خصیصه مکتون استوار است. و بر این اساس، احتمال پاسخ درست هر سؤال آزمون به عنوان تابعی از دشواری سؤال و توانایی آزمودنی توسط منحنی تراکمی برهان و اخیراً توسط منحنی منطقی که ساده‌تر است، بررسی می‌شود.

### مقدمه

از اوایل قرن جاری، الگوها و فرضهای نظریه کلاسیک آزمون (Classical Test Theory) در ساختن آزمونهای روانی-تربیتی مورد استفاده قرار گرفته است. اساس نظریه کلاسیک آزمون اولین بار در اوایل سالهای ۱۹۰۰ توسط اسپیرمن ارائه گردید (Spearman, ۱۹۱۲) و سپس با تلاش دانشمندان دیگری چون، گیلفورد (Guilford, ۱۹۵۲)، مگنوسون (Magnusson, ۱۹۶۷) بُرد و نوویک (Lord & Novick, ۱۹۶۸) به صورت آنچه امروزه مورد استفاده قرار می‌گیرد، گسترش یافت. با وجود اهمیتی که این نظریه تاکنون در

می شود و خطای اندازه‌گیری که خود تابعی از پایایی و واریانس آزمون به شمار می‌رود. برای تمام آزمودنیها یکسان فرض می‌گردد. چنانچه یک آزمودنی در گروههای متفاوتی قرار گیرد، با این که توانایی او در پاسخ دادن به سؤالات ثابت باقی می‌ماند، ولی نمره واقعی او با توجه به خطای اندازه‌گیری گروهی که در آن قرار می‌گیرد برآورد می‌شود. لذا، قبول فرض یکسان بودن خطا برای تمام آزمودنیها چندان آسان نیست. محدودیتهای فوق دشواری‌هایی بر سر راه سازندگان آزمون در ساختن و تفسیر نمره‌ها فراهم آورده است (سپاسی، ۱۳۷۴).

بنابراین، متخصصان روانسنجی جدید اساس و پایه نظریه دیگری برای ساختن و تفسیر نمره‌های آزمونهای روانی - تربیتی پیشنهاد می‌کنند. این نظریه جدید به نام نظریه خصیصه مکنون (Latent Trait Theory) یا نظریه سؤال - پاسخ (Item Response Theory)، یا اختصاراً به IRT مشهور است. خصوصیات این نظریه به عقیده همیلتون و همکاران (Hambelton et al., ۱۹۸۵) عبارتند از: (۱) پارامترهای سؤال (دشواری، تمیز) به گروه سررد آزمون بستگی ندارد، (۲)

ساختن آزمونهای روانی - تربیتی داشته است و از آن به عنوان یک نظریه منسجم در آزمون سازی نام برده می‌شود، متخصصان روانسنجی جدید معتقدند که مبانی این نظریه دیگر قادر نیست پارامترهایی را که آزمودنی به وسیله آنها به سؤالات مقیاس یا آزمون پاسخ می‌دهد توصیف کند. به بیان دیگر، به نظر این روانسنان، روشهایی که معمولاً براساس آنها آزمونهای روانی - تربیتی ساخته می‌شوند یا مورد ارزشیابی و استفاده قرار می‌گیرند، با محدودیتهایی مواجهند.

با توجه به حاصل این بررسیها می‌توان محدودیتهای این نظریه را به‌طور خلاصه به شرح زیر بیان کرد: (۱) پارامترهایی که به وسیله آنها سؤالات آزمون بررسی می‌شود (دشواری، تمیز) به گروه و ویژه‌ای از آزمودنیها وابسته است، (۲) نمره‌هایی که تواناییهای آزمودنیها را برآورده می‌کند به نوع آزمونی که مورد استفاده قرار گرفته است بستگی دارد. (۳) در این نظریه، همبستگی بین نمره‌های دو فرم همتای یک آزمون، پایایی تلقی می‌شود. اما در عمل حصول این نوع پایایی، اگر ممکن نباشد، بسیار دشوار است، (۴) ضریب پایایی برای کل آزمون محاسبه

مفاهیم اساسی و آنچه باید در رابطه با این نظریه دانسته شود بررسی می‌گردد. بحث درباره منحنی ویژگی سؤال همراه با ذکر مثال به دنبال خواهد آمد و سرانجام به آینده این نظریه از نگاه برخی متخصصان روانسنجی اشاره خواهد شد.

### پیشینه موضوع

پیشینه موضوع به زمانی برمی‌گردد که توانایی نظریه کلاسیک آزمون برای توصیف توانایی آزمودنیها در پاسخ به سؤالات آزمون مورد سؤال قرار گرفت. بنابراین، مبانی نظریه جدیدی برای ساختن آزمونهای روانی - تربیتی ایجاد گردید و روانسنان به صورت کنونی آن را گسترش دادند (Tucker, ۱۹۶۴). برخلاف نظریه کلاسیک آزمون که، نمره کل آزمون ملاکی برای بیان تواناییهای آزمودنی روی مقیاس اندازه‌گیری محسوب می‌شود، در نظریه جدید می‌توان توانایی هر آزمودنی را از روی منحنی احتمال پاسخ درستی که به هر سؤال آزمون به‌طور جداگانه می‌دهد برآورد کرد.

برخی از روان‌شناسان در گذشته، علیرغم اینکه در آثارشان از اصطلاح منحنی ویژگی

نمره‌هایی که توانایی آزمودنیها را نشان می‌دهد به نمره کل آزمون بستگی ندارد، (۳) الگوی ریاضی که اساس این نظریه بر آن استوار است به جای نمره کل آزمون، خصوصیات سؤالات را به طور انفرادی تفسیر می‌نماید، (۴) الگوی ریاضی این نظریه برای برآورد پایایی به در آزمون که دقیقاً با فرضهای آزمونهای همنا جور باشند نیازی ندارد و (۵) الگویی که این نظریه ارائه می‌دهد اندازه دقیقی برای برآورد توانایی آزمودنی فراهم می‌سازد.

نظریه سؤال - پاسخ توانسته است چهار چوب مفیدی برای حل مسائل گسترده اندازه‌گیری خصوصیات روانی - تربیتی فراهم آورد. تعداد روزافزونی از سازمانهای سنجش که آزمونهایی برای گزینش داوطلبان در مدارس، دانشگاهها و صنایع می‌سازند، از این نظریه برای ساختن آزمون، به منظور تعیین سوگیریهایی ضمنی که مکان دارد در سؤالات آزمون وجود داشته باشد و همچنین ساختن آزمونهای مختلف با فرمهای همتا از یک آزمون، و تفسیر دقیق نمره‌ها استفاده می‌کنند.

در این مقاله ابتدا تاریخچه تحول نظریه سؤال - پاسخ مرور می‌شود. پس از آن

سؤالات را به عنوان تابعی از خصیصه مکنون که اساس عملکرد را روی سؤالات آزمون تشکیل می‌دهد بر روی محور مختصات نمودار ساخت. علاوه بر تاکر (Lazarsfeld, ۱۹۶۸) از پژوهشهایی که در حوزه اندازه‌گیری انجام داد، از اصطلاح خصیصه مکنون نام می‌برد.

باین که مبانی این نظریه در مقایسه با نظریه قدیم برای بیان پارامترهای آزمون پیچیده‌تر است ولی در واقع می‌توان آنها را شکل گسترش یافته نظریه کلاسیک آزمون دانست. لولی (۱۹۴۳، ۱۹۲۴) سعی کرد پارامترهای نظریه سؤال - پاسخ را به نظریه کلاسیک آزمون ربط دهد. وی چندین روش برای برآورد پارامترهای مورد نیاز پیشنهاد کرد. کار و بر این اساس بود که همبستگی درونی بین سؤالات را مساری فرض می‌کرد و حدس را عامل مهمی در پاسخ دادن به سؤالات آزمون به‌شمار نمی‌آورد. افزون بر آن، ریچاردسون (۱۹۳۶) و تاکر (۱۹۶۴) روانسنجانی بودند که روابط بین پارامترهای انگوی کلاسیک آزمون و پارامترهای منحنی ویژگی سؤال را مطرح و تدوین کردند.

چنانچه بخواهیم از روانسنجانی که در سالهای گذشته بیشترین سهم را در اشاعه

سؤال ذکری به میان نیامده، اما از بررسی پژوهشهای آنها چنین استنباط می‌شود که آنها با چنین منحنی کار کرده‌اند. از جمله این روان‌شناسان می‌توان از بینه و سیمون (Bient & Simon, ۱۹۱۶) نام برد. آنها عملکرد کودکان را بر اساس تکالیف شناختی که به آنها می‌دادند بر روی نموداری برآورد می‌کردند. به بیان دیگر، بینه و سیمون می‌خواستند بدانند چرا برخی کودکان از انجام برخی تکالیف شناختی عاجزند. برای این که عملکرد آنها را با سایر کودکان مقایسه کنند از نمودار استفاده می‌کردند. امروزه به این نمودار "منحنی ویژگی سؤال" گفته می‌شود که عامل اصلی در نظریه سؤال - پاسخ به‌شمار می‌رود. وی بحث درباره آنچه را که امروزه به آن نظریه سؤال - پاسخ می‌گوییم در واقع می‌توان در کارهای ریچاردسون (Richardson, ۱۹۳۶، لولی (Lowely, ۱۹۴۳، ۱۹۴۴)، و تاکر (Tucker, ۱۹۶۴) ردیابی کرد. از تاکر می‌توان به عنوان روانسنجی نام برد که نخستین بار اصطلاح "منحنی ویژگی سؤال" (Item Characteristics Curve) را به کار برد. وی احتمال پاسخ درست آزمودنیها به

سؤال را، سهولت بخشید و منحنیهای منطقی Logistic Curves را جایگزین منحنیهای تراکمی نرمال یا منحنی نرمال اوجسایو (Ogive Normal Curve) که توسط لرد و همکارانش (۱۹۸۰) ارائه شده بود کرد و امکان استفاده از این الگوها را برای روانسنجان آسان کرد.

علی‌رغم توجه خاصی که در ابتدا توسط متخصصان روان‌سنجی به گسترش این نظریه معطوف شده بود، عواملی موجب گردید تا در خلال سالهای ۱۹۶۰ تا ۱۹۷۶ رشد این نظریه کند گردد. به عنوان مثال، دشواری مفاهیم ریاضی که برای توجیه مبانی این نظریه به کار برده شده بود درک آن را برای علاقمندان با مشکل مواجه ساخته بود. افزون بر این، هنوز نرم‌افزارهای کامپیوتری که به وسیله آنها بتوان به سهولت به تحلیل داده‌ها بر اساس نظریه سؤال - پاسخ پرداخت ساخته نشده بود. در نهایت این شک وجود داشت که نتوان از مبانی این نظریه در ساختن آزمونهای روانی - تربیتی استفاده کرد.

پس از سالها متعاقب آن متخصصان روانسنجی تلاش کردند تا نمره‌گذاری سؤالات آزمون را یکنواخت کنند، محاسبه

نظریه سؤال - پاسخ داشته‌اند نام ببریم، بدون شک نام لرد (۱۹۵۲، ۱۹۵۳ الف، ۱۹۵۲ ب، ۱۹۶۸) جایگاه ویژه‌ای دارد. بیکر (Baker, ۱۹۹۲). لرد علاوه بر مقالات مهمی که انتشار داد، می‌توان به رساله‌ای که تحت عنوان تک‌نگاشت روانسنجی (Psychometric Monograph) در سال ۱۹۵۱ به رشته تحریر درآورد اشاره کرد. وی اولین کسی است که الگوی نظریه سؤال - پاسخ و روشهای مربوطه را برای برآورد پارامترهای سؤال به وجود آورد و به طور موفقیت‌آمیزی این الگو را به داده‌های آزمونهای پیشرفت و توانایی بسط داد. سامیجیما (Samegima, ۱۹۶۹، ۱۹۷۲) گسترشهایی در نظریه سؤال - پاسخ به وجود آورده است که از آن جمله می‌توان گسترش داده‌های دو - وجهی Dichotomous به چند - وجهی Poly-choctmous، الگوهای یک - بعدی Unidimensional به چند - بعدی Poly-dimentional را ذکر کرد.

برین بام (Birnbaum, ۱۹۶۹) کارهای پژوهشی اساسی در زمینه نظریه سؤال - پاسخ انجام داده که در تحول این نظریه بسیار چشمگیر بوده است. وی با ایجاد روشهای آماری، محاسبه پارامترهای منحنی ویژگی

- پارامتری توجه خود را به گسترش الگوهای دو و سه - پارامتری معطوف ساختند و مبنای ریاضی و محاسباتی نیز برای این الگوها فراهم آوردند. در سالهای اخیر می توان به کارهای برجسته همبلتون و سوامی ناتان (Hambelton & Swaminathan, ۱۹۸۵) هیلین، دراسگر و پارسونز (Hulin, Drasgow & Parsons, ۱۹۸۳)، رایت و استون (Wright & Stone, ۱۹۷۹) و بیکر (Baker, ۱۹۹۲) اشاره کرد. علاوه بر طرح و ارائه الگوهای جدید برای توصیف نظریه سؤال - پاسخ، برنامه های کامپیوتری دقیق و پیشرفته برای بررسی پارامترها و سرعت بخشیدن به تحلیل نتایج آزمون به طور گسترده ای افزایش پیدا کرده است.

مفاهیم اساسی نظریه سؤال - پاسخ آنچه که باید در رابطه با این نظریه آزمون دانسته شود پاسخ به سؤالاتی است که در زیر ارائه می شوند:

۱. خصیصه مکنون چیست؟

۲. چه عواملی منحنی ویژگی سؤال را تشکیل می دهند؟

۳. فرض استقلال موضعی چه معنی

توانایی آزمودنیها را روی مقیاس اندازه گیری سهولت بخشند و کاربردی برای نظریه جدید نشان دهند تا بدین وسیله توجه علاقمندان را به سوی نظریه سؤال - پاسخ جلب نمایند. از جمله عواملی که به جلب علاقه عموم نسبت به نظریه جدید کمک کرد می توان به کارهای برجسته راش (Rasch, ۱۹۶۰) در دانمارک اشاره کرد. الگویی که راش برای توجیه مبانی نظریه سؤال - پاسخ ارائه کرده بود نه تنها در اروپا مورد استقبال قرار گرفت، بلکه بین متخصصان روانسنجی در ایالات متحده نیز طرفداران بسیاری پیدا کرد. از جمله می توان به رایت (Wright, ۱۹۷۹) اشاره کرد که با تشکیل کارگاههای آموزشی به مدت ۱۲ سال به اشاعه الگوی راش و سهولت اجرای آن در ساختن آزمونهای روانی - تربیتی پرداخت.

عامل دیگری که به ایجاد علاقه نسبت به نظریه سؤال - پاسخ کمک کرد انتشار کتاب نظریه آماری نمره های آزمون روانی (Statistical Theory Of Mental Test Scores)

بود که توسط لرد و نوویک در اواخر دهه

۱۹۶۰ انتشار یافت. در دهه ۱۹۸۰

متخصصات روانسنجی علاوه بر الگوی یک

کیفی یک مفهوم یاسازه است و نه یک اندازه فیزیکی، نمی توان آن را مثل قد و وزن به طور مستقیم مشاهده و اندازه گرفت. هدف اصلی اندازه گیریهای روانی - تربیتی این است که مشخص سازد فرد چقدر از خصیصه مکنون مورد نظر را دارد. چون بیشتر پژوهشها در زمینه پیشرفت به متغیرهایی چون توانایی خواندن، نوشتن، حساب کردن اشاره می کنند، اصطلاح کلی توانایی (ability)، در نظریه سؤال - پاسخ برای اطلاق به چنین خصیصه ای به کار برده شده است (بیکر، ۱۹۸۶).

همانطور که از متر برای اندازه گیری فاصله استفاده می کنیم، اگر بخواهیم میزان خصیصه مکنونی را که فرد دارا است اندازه بگیریم، به مقیاسی برای اندازه گیری آن نیاز است. تعریف مقیاس اندازه گیری، اعداد مورد استفاده در آن و تعیین مقدار خصیصه هایی که اعداد نمایشگر آنها هستند، کار بسیار دشواری است. بنابراین، نظریه سؤال - پاسخ از یک مقیاس توانایی فرضی استفاده می کند. صرف نظر از این که بخواهیم چه نوع توانایی را اندازه بگیریم، سه طور قراردادی می توان مقیاسی ساخت که وسط آن صفر و هر واحد

دارد و چگونه به بُعدی بودن آزمون مربوط می شود؟

### خصیصه مکنون و منحنی ویژگی سؤال

در بسیاری از مطالعات روانی - تربیتی صرفاً اندازه گیری یک متغیر مورد نظر است. مثلاً، می توان از هوش به عنوان یک متغیر نام برد که فراد در این ویژگی با هم تفاوت دارند. بنابراین افراد را می توان با صفت باهوش یا دارای هوش متوسط توصیف کرد و مطمئن بود که شنونده به آسانی متوجه پیام می شود. همچنین می توان درباره تواناییهای تحصیلی و عواملی از قبیل گرفتن نمره خوب یادگیری آسانتر مطلب درسی، توانایی برقرار کردن ارتباط بین اطلاعات به دست آمد، و استفاده مطلوب و مؤثر از وقت مطالعه بحث کرد و یا اینکه در زمینه های تحصیلی می توان از توانایی خواندن یا توانایی در ریاضیات یا علوم سخن گفت. هر یک از این موارد را متخصصان روانسنجی، خصیصه غیر قابل مشاهده یا خصیصه مکنون (latent trait) می گویند. در حالی که این متغیر را می توان به آسانی توصیف کرد و افراد آگاه می توانند به ذکر مشخصه های آن بپردازند، چون متغیر

می‌باشد. ولی در نظریه سؤال - پاسخ علاقه و توجه اصلی به جای این که به نمره خام آزمون معطوف شود به این است که آیا هر آزمودنی به هر سؤال به طور جداگانه پاسخ درست داده است یا خیر. به این دلیل است که مفاهیم اصلی در نظریه سؤال - پاسخ بر روش تحلیل هر یک از سؤالات استوار است و به نمره خام آزمون که از جمع نمره پاسخهای درست به سؤال به دست می‌آید کاری ندارد.

در عمل استفاده از سؤالات تشریحی در آزمون بسیار دشوار است، به ویژه اگر بخواهیم آنها را به شیوه باثباتی نمره‌گذاری کنیم یا مشکل مواجه خواهیم شد. در نتیجه، نظریه سؤال - پاسخ از سؤالات آزمون چند-گزینه‌ای استفاده می‌کند. فرض قابل قبولی که می‌توانیم اختیار کنیم این است که هر آزمودنی که به سؤال آزمونی پاسخ می‌دهد از مقداری توانایی برای پاسخ دادن به آن برخوردار است. از این رو، می‌توان فرض کرد هر آزمودنی برای پاسخ به هر سؤال یک ارزش عددی یک نمره‌ای دارد که موفقیت او را در امتداد محور مقیاس اندازه‌گیری توانایی نشان می‌دهد این نمره توانایی در نظریه سؤال - پاسخ با حرف

اندازه‌گیری آن برابر یک باشد. واحدهای اندازه‌گیری این مقیاس می‌تواند از دامنه بی‌نهایت منفی شروع و تا بی‌نهایت مثبت ادامه یابد. از این مقیاس می‌توان برای اندازه‌گیری توانایی مورد نظر استفاده کرد.

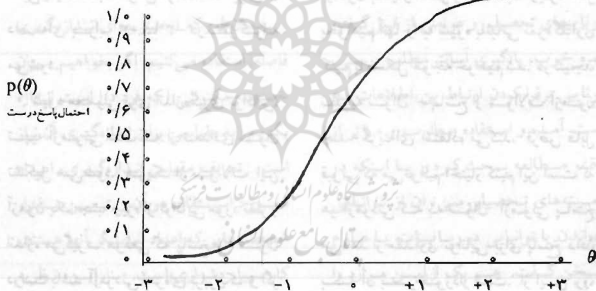
در این جا فرض بر این است که مانند اکثر اندازه‌گیرهای فیزیکی از این مقیاس نیز می‌توان برای میزان توانایی یک فرد در خصیصه مورد نظر، و یا برای مقایسه توانایی چند نفر استفاده کنیم. علی‌رغم این که دامنه توانایی را می‌توان از بی‌نهایت منفی تا بی‌نهایت مثبت فرض کرد، عملاً برای آن دامنه‌ای برابر  $+3$  تا  $-3$  در نظر گرفته می‌شود.

شیوه معمولی برای اندازه‌گیری توانایی، تهیه آزمونی است که از تعدادی سؤال تشکیل می‌شود. هر یک از سؤالات این آزمون یک جنبه ویژه از توانایی مورد نظر را اندازه می‌گیرد. موقعی که پاسخ به سؤال درست باشد آزمودنی نمره‌ای برابر یک و اگر پاسخ غلط باشد نمره‌ای برابر صفر می‌گیرد. به عبارتی، سؤال را به صورت دو ارزشی نمره‌گذاری می‌کنیم. در ارتباط با نظریه کلاسیک آزمون، نمره خام آزمودنی، مجموع پاسخهای صحیح او به سؤالات آزمون



شکل ۱ نشان داده شده است در می آید. احتمال پایین ترین سطوح توانایی نزدیک به صفر است و در بالاترین سطح توانایی احتمال پاسخ درست به یک می رسد. این منحنی S شکل رابطه بین احتمال پاسخ درست به سؤال را روی مقیاس توانایی نشان می دهد. در نظریه سؤال - پاسخ به این شکل منحنی ویژگی سؤال (Item Characteristic) گفته می شود. هر سؤال مندرج در آزمون منحنی ویژگی خود را دارد.

یروانی تا  $\theta$  (Theta) نشان داده می شود. در هر سطح توانایی احتمال معینی وجود دارد که آزمودنی با آن احتمال می تواند به سؤال پاسخ درست بدهد. این احتمال را می توانیم به صورت  $P(\theta)$  نشان دهیم. در مورد یک آزمون معمولی، این احتمال برای آزمودنیهایی که توانایی کمی دارند پایین و برای آزمودنیهایی که توانایی زیادی دارند بالا خواهد بود. چنانچه  $P(\theta)$  به عنوان تابعی از توانایی روی نمودار رسم شود نتیجه به صورت منحنی S بگونه ای که در



شکل ۱. منحنی ویژگی سؤال

اقتباس از (بیکر Baker، ۱۹۸۵، ص. ۳۴۱)

مفاهیم دیگر مربوط به این نظریه به این منحنی بستگی دارد. بیکر (۱۹۸۵) عقیده

منحنی ویژگی سؤال استخوانبندی نظریه سؤال - پاسخ را تشکیل می دهد و تمام

بالای توانایی است. با استفاده از این دو توضیح، می توان شکل کلی منحنی ویژگی سؤال را بیان کرد. از این دو پارامتر همچنین می توان برای بحث درباره ویژگی های فنی هر سؤال استفاده کرد. بیکر (۱۹۸۵) مشدار می دهد که باید متوجه بود این دو پارامتر مطلبی درباره این که خود سؤال واقعا قدرت اندازه گیری جنبه های مختلف توانایی مورد نظر را دارد یا نه بیان نمی کند این سوالی است که به اعتبار آزمون بستگی دارد. به عقیده وی این دو پارامتر به سادگی شک منحنی ویژگی سؤال را بیان می کنند. علاوه بر منحنی ویژگی سؤال، فرضهای اساسی دیگری در مورد نظریه سؤال - پاسخ بیان شده است. همبلتون و کوک (Hambelton & Cook, ۱۹۷۷) علاوه بر منحنی ویژگی سؤال به دو فرض دیگر یعنی، بُعدی بودن فضای مکنون (Dimensional Latent space) و استقلال موضعی (local independence) اشاره می کنند. به بیان دیگر، موفقیت هر آزمودنی را می توان از نمرههایی که او در فضای مکنون به خود اختصاص می دهد برآورد کرد. نمره های خصیصه مکنون به عنوان محور مختصات برای تعیین محل قرار گرفتن واقعی آزمودنی در فضای مکنون به کار برده می شود. بُعدی بودن فضای

دارد که منحنی ویژگی سؤال دو ویژگی فنی دارد که منحنی با آنها توصیف می شود. اولین ویژگی، پارامتر دشواری سؤال (Item Difficulty) نام دارد. در نظریه سؤال - پاسخ، دشواری سؤال محلی را روی مقیاس اندازه گیری توصیف می کند که با توانایی آزمودنی متناظر است. به عنوان مثال، یک سؤال آسان در ارتباط با آزمودنیها دارای توانایی کم تمایز قابل می شود در حالی که یک سؤال دشوار در ارتباط با آزمودنیهای دارای توانایی بالا، عمل می کند. بنابراین، دشواری سؤال شاخصی است که محل عملکرد آزمودنیها را روی مقیاس توانایی تعیین می کند. دومین ویژگی، پارامتر تمیز سؤال (Item discrimination) نام دارد که نشان می دهد چگونه سؤال می تواند آزمودنیهای با توانایی مختلف را (بالا، متوسط و پایین) از یکدیگر تفکیک کند. این پارامتر اساسا به شیب منحنی ویژگی سؤال مربوط می شود و بیشتر در قسمت وسط منحنی اثر می گذارد. بنابراین، هر چه شیب منحنی بیشتر باشد سؤال از قدرت تمیز بیشتری برخوردار است. برعکس هر چه منحنی مسطح تر باشد، سؤال قدرت تمیز کمتری دارد؛ احتمال پاسخ درست در سطوح پایین توانایی تقریبا مثل احتمال پاسخ درست در سطوح

مکنون به تعداد عواملی که روی سؤال آزمون تأثیر می‌گذارند بستگی دارد. معمولاً می‌توان چنین فرض کرد که فضای مکنون یک بعدی (Unidimensional) است یعنی فرض بر این است که سؤالات آزمون ز لحاظ این که فقط یک توانایی یا یک خصیصه را اندازه‌گیری می‌کنند همانند. به عقیده لرد (Lord, ۱۹۶۸) ماهیت فرض یک بعدی بودن مجموعه سؤالات آزمون در مورد اکثر آزمونها صحت ندارد، ولی اضافه می‌کند که امکان دارد یک بعدی بودن در بعضی شرایط، تقریب خوب و قابل قبولی به وجود آورد. هیولین، دواسگو و پارسونز (۱۹۸۳) در این که آزمونهای روانی‌ای که دقیقاً یک بعدی باشند وجود ندارد، با لرد (۱۹۸۰) هم عقیده هستند. احتمال بسیار وجود دارد که در آنها بعد جزئی بر روی پاسخ سؤالات اثر بگذارند. آنان اظهار می‌دارند که چون نمی‌توان با بررسی سؤالات، مشکل یک بعدی بودن آزمون را برطرف کرد و از طرفی فرض یک بعدی بودن در تک‌مل نظریه سؤال - پاسخ اساسی است، یک بعدی بودن هر آزمون باید قبل از به کار بستن این نظریه مورد بررسی قرار گیرد.

همبستگی و تناسب (Trub & Hambelton, ۱۹۷۱) مناسب

بودن فرض یک بعدی بودن برای مجموعه سؤالات آزمونهای روانی را از طریق تحلیل عاملی مورد مطالعه قرار داده‌اند. از نتایج مطالعات آنها چنین استنباط می‌شود که موقعی که سؤالات آزمونی بیش از یک توانایی واحد را اندازه می‌گیرند این امکان وجود دارد که براساس نتایج به دست آمده از تحلیل عوامل بتوانیم سؤالات را در گروههای متجانستری گروهیم آوریم و آن گاه فرض خصیصه مکنون را در باره مجموعه‌ای از سؤالات ممکن تحلیل کنیم. کروکر و الگینا (Croker & Algina, ۱۹۸۶) یک بعدی بودن را با همبستگی بین سؤالات آزمون مشخص می‌کنند. به عقیده آنها، زمانی فرض یک بعدی بودن حاصل می‌شود که بتوان از لحاظ آماری نشان داد که همبستگی معنی‌داری بین جامعه‌های متفاوتی از سؤالات وجود دارد. یعنی این که ابتدا تنها به اندازه‌گیری یک خصیصه مکنون می‌پردازند. سپس الگوی خصیصه مکنون را در مورد هر مجموعه‌ای از سؤالات ممکن به کار می‌برند.

در ارتباط با دومین فرض در مورد نظریه سؤال - پاسخ، یعنی استقلال موضعی، می‌توان گفت که براساس این فرض پاسخهایی که آزمودنی به سؤالات آزمون می‌دهد به لحاظ آماری از یکدیگر مستقل

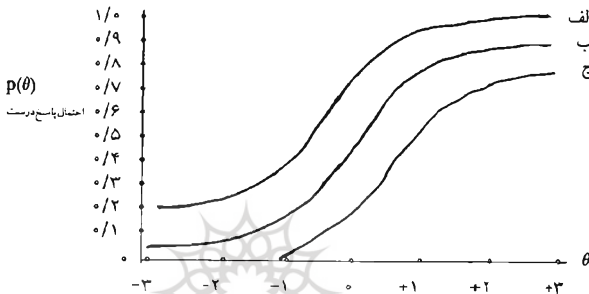
آن سؤال ویژگی استقلال‌ی موضعی را ندارد و مناسب‌الگویی خصیصه مکنون نیست. به نظر کروکر و الگینا (۱۹۸۶)، استقلال موضعی فرض محدودی را بیان می‌کند و بنابراین، امکان دارد نتوان آن را در مورد داده‌های آزمونهای روانی به کار برد. لرد (۱۹۸۰)، بین فرض یک بُعدی و فرض استقلال موضعی رابطه نزدیکی نشان می‌دهد و معتقد است که مناسب بودن فرض استقلال را می‌توان بر روی هر مجموعه‌ای از داده‌ها با استفاده از روش تحلیل عوامل آزمایش کرد.

پارامتر دشواری سؤال به عنوان شاخص محل قرار گرفتن آزمودنی روی مقیاس توانایی، به بررسی بیشتری نیاز دارد. بنابراین در زیر آن را دقیق‌تر مورد بحث قرار می‌دهیم. در شکل ۲، سه منحنی ویژگی سؤال در روی نمودار ارائه شده است. تمام آنها سطح تمیز یکسانی دارند ولی در رابطه با دشواری با یکدیگر متفاوتند. منحنی بالا سمت چپ (الف) معرف سؤال آسان است برای این که احتمال پاسخ درست برای آزمودنیها با توانایی کم زیاد است و برای آزمودنیهای با توانایی بالا به یک می‌رسد. منحنی وسط (ب) سؤال را با دشواری متوسط نشان می‌دهد برای اینکه احتمال پاسخ درست در پایین‌ترین سطوح توانایی کم، حدود ۵۰/۵۰

هستند. برای این که این فرض درست باشد، عملکرد آزمودنی در یک سؤال نباید بر عملکردش در دیگر سؤالات آزمون به گونه‌ای بمر یا بدتر تأثیر بگذارد. به عنوان مثال، محتوای یک سؤال نباید علایم و نشانه‌هایی برای پاسخ به سؤالات دیگر فراهم آورد. در واقع آنچه که هر یک از سؤالات آزمون اندازه می‌گیرند توانایی واحدی است و هیچ توانایی دیگری در سؤالات آزمون مشترک نیست. برای اینکه این فرض تحقق یابد می‌توان احتمال هر الگوی پاسخ آزمودنی (یکها و صفرها) را به وسیله حاصل ضرب نمره‌های حاصل از هر سؤال برای هر آزمودنی به دست آورد. این فرض چنانچه تمام سؤالات آزمون یک توانایی مشخص را اندازه بگیرد صادق است (هملتون و تراب، ۱۹۷۱). مثلاً، اگر احتمال این که آزمودنی به سؤال اول پاسخ درست دهد، مساوی  $0/70$  و همین احتمال برای سؤال دوم  $0/80$  باشد، چنانچه این دو سؤال دارای استقلال موضعی باشند، احتمال این که آزمودنی به هر دو سؤال پاسخ درست بدهد مساوی  $0/56 = 0/70 \times 0/80$  است. آلن و یسن (Allen & Yen, ۱۳۷۴) اظهار می‌دارند، در صورتی که پاسخ دادن به یک سؤال تحت تأثیر پاسخ دادن به سؤالی دیگر قرار گیرد،

بالاتر توانایی می‌رسد، افزایش پیدا می‌کند. حتی در بالاترین سطح توانایی یعنی +۳، احتمال پاسخ درست برای دشوارترین سؤال فقط ۰/۸۰ می‌باشد.

در وسط مقیاس توانایی، و در بالاترین سطح توانایی نزدیک به یک است. منحنی پایین سمت راست (ج) مبین سؤال دشوار است. احتمال پاسخ درست برای مقیاس توانایی پایین و تنها موقعی که به سطوح



شکل ۲. منحنی خصوصیات سه سؤال با درجه دشواری متفاوت

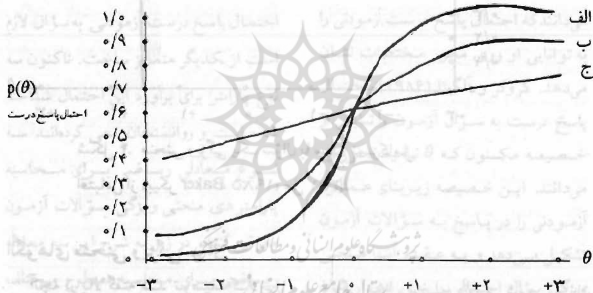
(اقتباس از بیکر Baker, ۱۹۸۵, ص ۳۴۹)

کمتر از ۰/۵۰ می‌باشد، و در فاصله کوتاهی در سمت راست، احتمال از ۰/۵۰ بسیار بیشتر است. منحنی (ب) سؤالی را با قدرت تمیز متوسط نشان می‌دهد. شیب این منحنی از منحنی قبلی بسیار کمتر است و همانطور که بر سطح توانایی افزوده می‌شود، احتمال پاسخ درست کمتر از منحنی قبلی تغییر می‌کند. اما، احتمال پاسخ درست برای کمترین توانایی نزدیک به

مفهوم تمیز در شکل ۳ نشان داده شده است. این شکل سه منحنی ویژگی سؤال را که همه سطح دشواری یکسانی دارند ولی در قدرت تمیز متفاوتند نشان می‌دهد. منحنی (الف) از سطح تمیز بالایی برخوردار است برای این که در وسط، یعنی جایی که احتمال پاسخ درست به سرعت تغییر می‌یابد شیب پیدا می‌کند. در فاصله کوتاهی در سمت چپ منحنی، احتمال پاسخ درست بسیار

افزایش پیدا می‌کند. همانطور که قبلاً گفتیم با وجودی که شکلها دامنه توانایی را بین +۳ و -۳ نشان می‌دهند، ولی از لحاظ تئوری دامنه این توانایی از بی‌نهایت مثبت تا بی‌نهایت منفی ادامه دارد. بنابراین، تمام منحنیهای ویژگی سؤال که در اینجا به کار گرفته شده‌اند نسبت به احتمال صفر در یک دامنه و بر احتمال یک در دنباله دیگر مجانب می‌شوند.

صفر است و برای آزمودنیهای با بالاترین سطح توانایی نزدیک به یک می‌رسد. سومین منحنی یعنی منحنی (ج) سؤالی را با قدرت تمیز پایین نشان می‌دهد. منحنی دارای شیب بسیار کمی است و احتمال پاسخ درست در دامنه تواناییهای نشان داده شده به آستگی تغییر می‌کند. حتی در سطوح پایین توانایی، احتمال پاسخ درست به‌طور قابل قبولی بزرگ و فقط موفقی که به سطوح توانایی بالا می‌رسد مقدار کمی



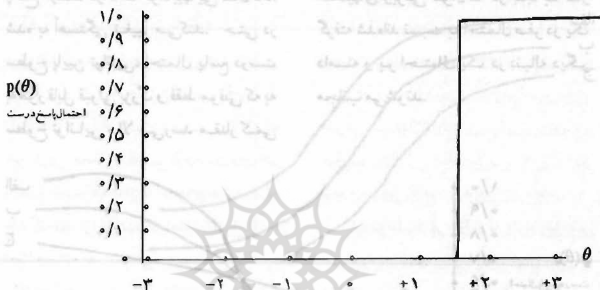
شکل ۳. مقایسه سه منحنی خصوصیت سؤال با قدرت تمیز متفاوت

اقتباس از (بیکر Baker، ۱۹۸۵، ص ۳۵۰)

شکل ۴ منحنی چنین سؤال را نشان می‌دهد. در قسمت چپ خط عمودی جایی که  $\theta = 1/5$  است، احتمال پاسخ درست صفر و در سمت راست خط احتمال پاسخ درست

حالت خاصی وجود دارد که هر سؤال دارای سطح تمیز کامل می‌باشد. منحنی ویژگی چنین سؤالی تشکیل خط عمودی در نقطه‌ای در امتداد مقیاس توانایی می‌دهد.

یا پایین  $1/5$  قرار دارد ایده آل است. ولی، چنین سؤالی بین آزمودنیها با تواناییهای بالای  $1/5$  و آزمودنیهای پایین تر از  $1/5$  تمایزی فائل نمی شود (بیکر، ۱۹۹۲).



شکل ۴. منحنی ویژگی یک سؤال با سطح تمیز کامل

اقتباس از (بیکر Baker، ۱۹۸۵، ص ۳۵۱)

### الگوهای منحنی ویژگی سؤال

آنچه در بالا گفته شد توصیف کلامی ویژگیهای منحنی سؤال بود. با وجودی که احتمال دارد خواننده تصویر نسبتاً روشنی از زمینه بحث منحنیهای ویژگی سؤال به دست آورده باشد، ولی تا زمانی که این گفته‌ها به لحاظ نظریه‌ای تأیید نشوند از دقت لازم برخوردار نخواهند بود. بنابراین، در زیر مفاهیم ریاضی زیربنای الگوهای منحنی

ویژگی سؤال توضیح داده می‌شود. به بیان دیگر، در هر کاربرد نظریه سؤال - پاسخ سازنده آزمون باید شکل معادل ریاضی را برای منحنی ویژگی سؤال معین کند تا برآوردهایی از پارامترهایی که برای توصیف منحنی لازم است به دست دهد. هر الگو از چند پارمتر که مقادیر عددی آنها ویژگی سؤال خاصی را تعریف می‌کند درست شده

نظر مستقل است. بدیهی است، احتمال پاسخ درست برای یک آزمودنی به این که چند آزمودنی دیگر در نقطه یکسانی در امتداد پیوستار توانایی (با در نقطه متفاوتی) قرار گرفته‌اند بستگی نخواهد داشت. بنابر این، شکل منحنی ویژگی سؤال به توزیع توانایی در جامعه آزمودنی‌ها بستگی ندارد. لذا در بعضی مواقع شکل منحنی برای نمونه‌های متفاوتی از آزمودنی‌های آن جامعه، بدون توجه به اینکه نمونه آزمودنی‌های چگونه انتخاب شده است بدون تغییر باقی خواهد ماند (ص ۴۶).

الگوهای نظریه سؤال - پاسخ را می‌توان با توجه به تعداد پارامترهایی که برای برآورد احتمال پاسخ درست آزمودنی به سؤال لازم است از یکدیگر متمایز ساخت. تاکنون سه نوع پارامتر برای برآورد این احتمال شناخته شده است و روانسجان سعی کرده‌اند سه نوع معادلی ریاضی برای محاسبه پارامترهای منحنی ویژگی سؤالات آزمون معرفی کنند. در زیر به شرح این سه معادله ریاضی برای محاسبه سطوح مختلف توانایی می‌پردازیم.

### تابع منطقی Logistic Function

در ارتباط با نظریه سؤال - پاسخ، الگوی ریاضی استاندارد که برای منحنی ویژگی سؤال به کار می‌یریم شکل تراکمی به خود

است. اگر بخواهیم یک نظریه اندازه‌گیری را به گونه‌ای تکامل بخشیم که بخوبی قابل تعریف باشد، به این قبیل الگوما نیاز خواهیم داشت. به علاوه، ارائه الگوها و پارامترهای آنها وسیله‌ای برای انتقال اطلاعات درباره ویژگیهای فنی سؤال فراهم می‌آورند (بیکر ۱۹۸۶).

اکثر متخصصان روانسنجی تعریف مشابهی برای منحنی ویژگی سؤال بیان می‌کنند. همبلتون و سوامی ناتان (۱۹۸۵)، منحنی ویژگی سؤال را تابعی ریاضی می‌دانند که احتمال پاسخ درست آزمودنی را به توانایی او روی محور مختصات نشان می‌دهد. کروکر و آگینا (۱۹۸۶)، احتمال پاسخ درست به سؤال آزمون را تابعی از خصیصه مکنون که  $\theta$  نامیده می‌شود می‌دانند. این خصیصه زیربنای عملکرد آزمودنی را در پاسخ به سؤالات آزمون تشکیل می‌دهد و به عقیده آنها منحنی ویژگی سؤال احتمالی را روی نمودار نشان می‌دهد که با آن احتمال آزمودنی می‌تواند به سؤالات آزمون پاسخ درست دهد. علاوه بر تعاریف فوق، همبلتون و کوک (۱۹۷۷)، به اهمیت نقش منحنی ویژگی سؤال در این نظریه اشاره می‌کنند. آنها می‌گویند:

احتمال پاسخ درست به یک سؤال از توزیع توانایی آزمودنی در جامعه‌ای از آزمودنی‌های مورد



سؤال را در امتداد پیوستار کنترل می‌کند، یعنی احتمال پاسخ مثبت آزمودنی‌هایی که توانایی آنها برابر  $b$  برآورد می‌شود مساوی  $0/50$  است و بنابراین این  $b$  نقطه‌ای در امتداد پیوستار توانایی است که در آن نقطه آزمودنیها  $50\%$  شانس دارند که به سؤال پاسخ مثبت دهند. دامنه نظری مقادیر این پارامتر  $+\infty < b < -\infty$  ولی مقادیری که در محاسبات عملی به دست می‌آیند در دامنه بین  $-2$  تا  $+3$  قرار می‌گیرد.

به واسطه این که منحنی ویژگی سؤال نیکل  $S$  دارد، شیب منحنی به عنوان تابعی از سطح توانایی تغییر می‌کند و موقعی به حداکثر مقدار خود می‌رسد که توانایی ساری دشواری سؤال می‌شود. بنابراین، پارامتر نیز شیب کلی منحنی ویژگی سؤال را معرفی می‌نماید. پارامتر تمیز به لحاظ ریاضی نسبت شیب منحنی ویژگی سؤال در جایی که  $\theta$  با  $b$  برابر است تعریف می‌شود. دامنه نظری مقادیر این پارامتر از منهای بی‌نهایت تا به اضافه بی‌نهایت ادامه دارد ولی دامنه‌ای که معمولاً در محاسبات به دست می‌آوریم معمولاً از  $-2/80$  تا  $+2/80$  گسترده است.

می‌گیرد. این الگوی ریاضی استاندارد گروهی از منحنیها را با شکل عمومی منحنیهای ویژگی سؤال، که قبلاً شرح آنها داده شد، معین می‌کند. این تابع منطقی ابتدا در سال ۱۸۴۴ به دست آمد و تاکنون به‌طور گسترده‌ای در علوم زیستی برای توضیح رشد گیاهان و حیوانات از تولد تا بلوغ به کار برده شده است (کروکر و آلگینا، ۱۹۸۶). اولین بار در اواخر سالهای ۱۹۵۰ از این الگو به واسطه سادگی آن برای بررسی منحنی ویژگی سؤال در آزمونه‌های روانی-تربیتی نیز استفاده شد (بیکر، ۱۹۸۵).

معادله ریاضی برای الگوی دو-پارامتری به

شرح زیر است: 
$$p(\theta) = \frac{1}{1+e^{-L}} = \frac{1}{1+e^{-a(\theta-b)}}$$
 در این فرمول:

$e$  مقدار ثابت  
 $b$  پارامتر دشواری  
 $a$  پارامتر تمیز  
 $L = a(\theta - b)$  انحراف لگاریتمی  
 $\theta$  سطح توانایی

این معادله گروهی از منحنیها را به وجود می‌آورد که اعضایش، پارامترهای دشواری و تمیز، به وسیله مقادیر عددی ویژه مشخص می‌شود. پارامتر دشواری که با  $b$  نشان داده می‌شود شکل منحنی ویژگی

## الگوی منطقی یک - پارامتری یا الگوی راش **Rasch Model**

الگوی دیگری در اواسط سالهای ۱۹۶۰ به وسیله راش ریاضیدان دانمارکی تدوین شد. ویژگی الگوی راش این است که پارامتر تمیز در الگوی دو - پارامتر در مقدار  $a=1$  برای تمام سوالات ثابت است و این فقط پارامتر دشواری است که می تواند مقادیر متفاوتی به خود بگیرد بنا بر این. به الگوی راش الگوی یک - پارامتری (One Parameter Logistic Function) می گویند (راش، ۱۹۶۰). معادله الگوی

راش به شرح زیر است 
$$p(\theta) = \frac{1}{1 + e^{-1(\theta-b)}}$$
 در این فرمول:

$$e \quad 2/718$$

b پارامتر دشواری

$\theta$  سطح توانایی

هستند. دقت کنید که پارامتر تمیز (a) در معادله وجود دارد، ولی چون مقدار آن همیشه مساوی یک است، معمولاً آن را در معادله نشان نمی دهند:

## الگوی سه - پارامتری

### Three-Parameter Model

واقعیتی که در اجرای آزمون وجود دارد

این است که امکان دارد آزمودنیها با حدس زدن پاسخ درست سوالی را پیدا کنند. بنا بر این، در محاسبه احتمال پاسخ درست جزء کوچکی ناشی از حدس زدن مداخله دارد. هیچ یک از دو الگویی که قبلاً به آنها اشاره شد پدیده حدس زدن را مورد ملاحظه قرار نمی دهند. برین بام (۱۹۶۹) الگوی دو - پارامتری را تغییر داد تا پارامتر دیگری به نام حدس زدن را که در احتمال پاسخ درست مداخله می کند معرفی نماید. الگویی که برین بام ایجاد کرد به الگوی سه - پارامتری مشهور است (برین بام، ۱۹۶۹).

معادله الگوی سه پارامتری به شرح زیر

است:

$$p(\theta) = c + (1-c) \frac{1}{1 + e^{-a(\theta-b)}}$$

در این فرمول:

b پارامتر دشواری

a پارامتر تمیز

c پارامتر حدس زدن

$\theta$  سطح توانایی

هستند. پارامتر c احتمال رسیدن به پاسخ درست فقط به وسیله حدس زدن است و به لحاظ تعریفی که در فرمول به کار رفته است مقدار آن به عنوان تابعی از سطح توانایی تغییر نمی کند. از این رو، آزمودنیها در پایین ترین و بالاترین سطح توانایی احتمال

می‌دهد کدامند؟ آیا این نظریه جدید می‌تواند در آینده کاربرد عملی در ساختن آزمونهای روانی-تربیتی داشته باشد و برای حل مسائل و مشکلاتی که سازنده آزمون در ساختن و تفسیر نمره‌ها با آنها مواجه است پاسخی منطقی فراهم آورد؟

نورشته‌ها و مقاله‌های غربی در زمینه‌های نظریه سؤال-پاسخ (IRT) چشم‌انداز بسیار روشنی پیش روی سازندگان آزمون قرار داده است. متخصصان روان‌سنجی به محدودیتهای نظریه کلاسیک آزمون آگاهی دارند و معتقدند که نظریه جدید در آینده می‌تواند برای مشکلاتی که آزمون‌سازی در گذشته با آنها مواجه بوده است و تاکنون لاینحل باقی مانده است پاسخی منطقی فراهم آورد. آنان همچنین بر این باورند که الگوهای این نظریه در آینده ابزارهای مطمئنی در طراحی و ساختن آزمونها و متیاسها در اختیار سازندگان آزمون قرار می‌دهد. هم‌اکنون تعداد کتب و مجله‌های غربی که به نشر مطالب درباره این نظریه و چشم‌اندازهای آن در آینده اختصاص یافته رو به افزایش است. برخی از مجله‌های علمی- پژوهشی مطالب کامل بعضی از شماره‌های خود را به انعکاس نظر

ساوای برای رسیدن به پاسخ درست به وسیله حدس زدن را دارند. پارامتر  $c$  دامنه نظری برابر  $0 < c < 1$  دارد، ولی در عمل مقادیر بالای  $0/35$  قابل قبول نیستند و از این‌رو دامنه آن  $0 < c < 0/25$  می‌باشد.

تأثیر جانبی به کاربردن پارامتر حدس زدن ( $c$ ) این است که تعریف پارامتر دشواری تغییر می‌کند. باتوجه به دو الگوی قبلی،  $b$  روی مقیاس توانایی نقطه‌ای بود که احتمال پاسخ درست در آن نقطه برابر  $0/50$  است. ولی در این الگو، حد پایین منحنی ویژگی سؤال بستگی به میزان شانس که آزمودنی در پاسخ دادن به سؤال دارد، تغییر می‌کند.

### بحث و نتیجه‌گیری

این روزها بحث درباره نظریه جدید آزمون، نظریه سؤال-پاسخ بین متخصصان روان‌سنجی، سازندگان آزمون و دانشجویان در کشورمان در حال گسترش است. آنها می‌خواهند بدانند ماهیت این نظریه چیست و در چه مواردی صحبت می‌کند؟ با نظریه موجود، یعنی نظریه کلاسیک آزمون، چه تفاوتی دارد؟ الگوها و شیوه‌هایی که برای ساختن آزمونهای روانی-تربیتی ارائه

تفکیک کرد، و سرانجام از بررسی خصوصیات آماری آنها در ساختن و تفسیر نمره‌های آزمون یا مقیاس استفاده کرد.

در پایان با این که از ابتدای کار نظریه سؤال - پاسخ در سالهای ۱۹۴۰ بیش از پنجاه سال می‌گذرد و مبانی نظری آن در تمام سالهای ۱۹۸۰ و اوائل سالهای ۱۹۹۰ با شتاب بیشتری گسترش یافته است، این نظریه در بین متخصصان روانسنجی و سازندگان آزمون در کشورمان متوزنا شناخته است. علیرغم این که کتاب و مقالات فراوانی در زمینه مفاهیم اساسی این نظریه و کاربردهای آن در غرب انتشار یافته، تاکنون کتاب یا مقاله‌ای که به‌طور کامل به معرفی این نظریه بپردازد در کشورمان منتشر نشده است. چنین به نظر می‌رسد در آینده باید کتابها با مقاله‌های زیادی در این زمینه به رشته تحریر درآید، مطالعات کاربردی بیشتری، صورت پذیرد، برنامه‌های کامپیوتری مفیدی تهیه شود تا خواننده بتواند با مبانی این نظریه آشنایی پیدا کند و از الگوها و شیوه‌های آن برای ساختن و تفسیر نمره‌های آزمون یا مقیاس استفاده کند. تحریر این مقاله در این راستا انجام پذیرفته است. در این مقاله سعی شد اساس

منخصصان روانسنجی برای بحث و بررسی در مورد مفاهیم اساسی این نظریه و آینده آن اختصاص داده‌اند. به عنوان مثال، می‌توان از تلاشهای لرد (۱۹۸۰) و همپتون (۱۹۸۵) در جهت توسعه این نظریه و نقشی که در آینده می‌تواند در آزمون‌سازی داشته باشد نام برد. لرد (۱۹۸۰) یک بُعدی بودن فضای خصیصه مکنون را قبول نمی‌کند و معتقد است که این فرض مسأله بُعدی بودن فضای خصیصه مکنون را که تاکنون لاینحل مانده است حل نمی‌کند. وی معتقد است که روانشناسان و متخصصان علوم تربیتی می‌خواهند از نمره‌های آزمون اطلاعاتی بیش از رتبه‌بندی آزمودنیها و یا دسترسی به سطحی از استاندارد کسب نمایند.

وی اظهار می‌دارد که الگوهای نظریه سزای - پاسخ باید در آینده بتواند اطلاعات تشخیصی بیشتری در مورد توانایی آزمودنیها و خصوصیات سوالات آزمون فراهم آورد. آنچه از گفته‌های این روانسازان استنباط می‌شود این است که چگونه می‌توان الگوهایی ارائه داد که توسط آنها بهتر بشود عملکرد آزمودنیها را به توانایشان ارتباط داد، و پارامترهایی که برای تکمیل تکالیف شناختی یا عاطفی لازم است از هم

این نظریه و ویژگیهای آن با زبانی ساده توضیح داده شود و چشم‌اندازهایی که این نظریه در آینده امکان دارد در ساختن و تفسیر نمره‌های آزمودنیهای روانی - تربیتی داشته باشد، پیش روی خواننده علاقمند قرار دهد. بدون شک پرداختن به تمام

جوانب این نظریه آزمون از عهده یک مقاله خارج است. ولی، پژوهشگران باید تلاش کنند بیشتر به تبیین ابعاد گوناگون این نظریه پردازند و کاربرد عملی آن را در ساختن و تفسیر نمره‌های آزمون با ذکر مثال برای علاقمندان روشن سازند.

## منابع

### فارسی

- آلن. مری جی وین، وندی ام (۱۳۷۴). مقدمه‌ای بر نظریه‌های اندازه‌گیری (روانشنجی)، ترجمه علی دلاور. تهران سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت).  
 پاشا شریفی، حسن (۱۳۷۳). اصول روان‌سنجی در روان‌آزمایی. تهران، انتشارات رشد.  
 توراندایک، آل (۱۳۶۹). روان‌سنجی کاربردی. ترجمه حیدرعلی هومن. تهران، مؤسسه انتشارات دانشگاه تهران.  
 سپاسی، حسین (۱۳۷۴). نظریه کلاسیک آزمون و محدودیت‌های آن. مجله علوم تربیتی و روانشناسی دانشگاه شهید چمران، شماره‌های اول و دوم ص ۱۲۳-۱۳۲.

### خارجی

- Baker. B. F. (1985). *The basics of item response theory*. New Hampshire Heinemann Education Book.
- Baker. B. F. (1992). *Item-response theory*. New hampshire, Heinemann Education book.
- Biggs, J. B, & Collis. K. F. (1982), *Evaluating the quality of learning*. New York, Academic Press.
- Binet, A. L. & Simon, T. H. (1916). *The development of intelligence in young children*. New Jersey, Vineland.
- Birnbaum, A. (1969). A staistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, 6:258-76.

- Crocker, L. & Algina. j. (1986), *Introduction to classical and modern test theory*, New York. Holt, Rinehart & Winston. New York Press.
- Guilford, G. P. (1954). *Psychometric Methods*, (sec. Ed.) New Jersey, McGraw-Hill.
- Guilford, G. P. (1965). *Fundamental statistics in psychology and education*, New York. McGraw-Hill Book Company.
- Guliksen, H. (1950). *Theory of mathematic assessment and evaluation*. New York, State University of New York Press.
- Hambelton, K. R. & Swaminathan. H (1985). *Item response theory*, principles and application. Boston. Nighff Publications.
- Hambelton, K. R. Cook, I.I. (1977). Latent trait models and their use in the analysis of education test data. *Journal of Educational Measurement*, 14:75-96.
- Hambelton, K. R. & Traub, K. e. (1971). Information curves and efficiency of three logistic models. *British Journal of Mathematics and Psychology*, 24: 273-281.
- Hulin, C. L. & Drasgow, F. & Parsons, C. K. (1983) *Applications of Psychological measurements*. Illinois. Dow Jones Irwin, Homewood.
- Lazarsfield, P. F. (1968). *The logical and mathematic foundation of latent Structure analysis*. Princeton: Princeton University Press.
- Lord. F. M. (1952). A theory of test scores. *Psychometric monograph* No. 7.
- Lord. F. M. (1953). Relation of test score to the trait underlying test. *Educational and Psychological Measurement*, 13: 577-584.
- Lord. F. M. (1953). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18:57-75.
- Lord. F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*, Massachussets. Addison-Wesley.
- Lord. F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey. Lawrence Erlbaum Associations Publishers.
- Lowely, D. N. (1943). Problems connected with item section and test construction. *Royal Society of Edinburgh* 6, 273-283.
- Lowely, D. N. (1944). The factor analysis of multiple item tests. *Royal Society of Edinburgh* 69: 74-82.
- Magnusson. D. (1967). *Test theory*. Boston, Addison-Wesley.
- Rasch. G. (1960). *Probabilistic models for some intelligence and attainment test*. Denmark. Danish Institute for Educational Research.
- Richardson. M. W. (1936). The relationship between difficulty, and

- the differential Validity of a test. *Psychometrika*, 33:46-52.
- Samejima, F. (1969). Estimaion of latent ability using a response pattern of graded scores. *Psychometric Monograph* No. 17.
- Samejima, F. (1972), A General model for free-response data. *Psychometric Monograph*, No. 18.
- Spearman, C. (1913). Correlation of sum and difference. *British Journal of Psychology*, 5:417-426.
- Spearman, C(1904). General intelligence objectively determind and measured. *American Journal of Psychology*. 15:201-293.
- Thorndike, E. L. (1904). *An intruduction to the theory of mental and social meaurments.* New York. Teachers College, Columbia University.
- Tucker. L. R. (1964). Maximum Validity of a test with equivalent itens. *Psychometrika*, 11:1-13.
- Wright, B. D. Stone, H. H. (1979), *Best test desgin*, Chicago MESA.



ژوئیه ۱۳۹۸  
سال هفتم  
شماره ۱  
مجله علوم تربیتی و روانشناسی  
دانشگاه شهید چمران اهواز