مقدمه ای بر مبانی رویکردهای جدید اندازه گیری در حوزه روان شناسی و علوم تربیتی

بلال ایزانلو[1]

مجتبی حبیبی عسگرآباد[2]

تاریخ وصول:۸۶/۱۲/۳ تاریخ پذیرش:۸۷/۶/۱۲

چکیده

امروزه در حوزه اندازه‌گیری دو نظریه غالب وجود دارد. نظریه کلاسیک اندازه‌گیری، که به وسیله اسپیرمن مطرح گردید و نظریه سوال پاسخ که سرمنشأ آن به نظریه خصیصه مکنون باز می‌گردد. در واقع، زمانی که نظریه خصیصه مکنون برای اندازه گیری متغیرهای دو ارزشی مورد استفاده قرار گرفت، به نظریه سوال پاسخ مشهور شد. اگرچه مبانی پیچیده ریاضی این نظریه، در مقایسه با نظریه کلاسیک، باعث شده که پژوهشگران تمایل کمتری برای کاربرد این نظریه داشته باشند، ولی مفروضه های زیربنایی آن منطقی تر از نظریه کلاسیک هستند. امروزه در چارچوب این نظریه، دو رویکرد مختلف به اندازه گیری وجود دارد. رویکرد اول که مدل های خانوداه راش از آن تبعیت می کنند، هدف اصلی اش اندازه گیری میزان خصیصه افراد است: در حالی که رویکرد دوم با هدف تحلیل سوال و برآورد پارامترهای آنها شکل گرفته است و از مدل هایی نظیر مدل ۲و۳ پارامتری برای این منظور استفاده می کند. با وجود تفاوت های موجود بین این دو رویکرد، کمتر منبعی به این تفاوت ها اشاره می کند و پژوهشگران نیز کمتر از این موضوع اطّلاع دارند. در مقاله حاضر به بررسی مبانی شکل گیری نظریه سوال پاسخ، مفروضه ها، مدل های مختلف و تفاوت بین آنها پرداخته می شود.

۱- دانشجوی دکتری سنجش و اندازه گیری، سازمان سنجش آموزش کشور

۲- دانشجوی دکتری تخصصی روانشناسی سلامت، دانشگاه تهران

واژگان کلیدی: نظریه کلاسیک، نظریه سوال پاسخ، خصیصه مکنون، مدل راش، مدل های یک، دو، سه پارامتری.

مقدمه

اولین مدل و نظریه منسجم در حوزه سنجش و اندازه‌گیری ویژگی های روانی به اسپیرمن[1]، آماردان انگلیسی برمی‌گردد. وی در خلال سال های ۱۹۰۴ تا ۱۹۱۳ با استفاده از مفهوم همبستگی، مطالعاتی در خصوص نمرات به دست آمده از آزمون ها انجام داد. ایده اصلی اسپیرمن این بود که "هر نمره مشاهده شده از دو جزء نمره واقعی[2] و نمره خطا[3] تشکیل شده است ". وی در تلاش برای توصیف این نمرات موفق شد اولین مدل آماری را در این خصوص فراهم کند (لین،۱۹۸۶؛ سپاسی، ۱۳۷۴،۱۳۷۵؛ کیامنش، ۱۳۷۸). ارائه اولین مدل منسجم آماری برای تحلیل داده‌های به دست آمده از اجرای آزمون‌ها که به نظریه کلاسیک آزمون[4] (CTT) معروف است، یکی از کارهای علمی اسپیرمن بوده است. علاوه بر این نظریه، وی اولین کسی است که مدل های تحلیل عامل[5] را برای بررسی عملکرد ذهنی افراد در آزمون های هوش به عنوان متغیرهای مکنون پیوسته[6]، به‌کار گرفت. هدف وی آن بود که با این روش متغیرهای مکنون زیر بنایی موثر در توانایی ذهنی را پیدا کند. این جنبه از کار اسپیرمن که همان تحلیل عامل سنتی است، بعد از وی بوسیله افرادی مثل لاولی[7]، ترستون[8] و ماکسول[9] ادامه یافت (بورسبوم، ملنبرگ، وانهیردن، ۲۰۰۳). در واقع اسپیرمن و این افراد با کارهایی که در این حوزه انجام دادند از پیش کسوتان روش موسوم به تحلیل خصیصه‌های مکنون هستند. این روش که بعدها به نظریه خصیصه مکنون[10] شهرت یافت، بعد از آنها بوسیله افرادی مثل یورسکوگ[11]، ویلی[12]، اسمیت[13]، برامبل[14] و سوربم[15]

1. Spearman
2. True score
3. Error score
4. Classical Test Theory
5. Factor analysis models
6. Continues latent variables
7. Lawley
8. Thurstone
9. Maxwell
10. Latent trait theory
11. Joreskog
12. Wiley
13. Schmidt
14. Brambel
15. Sorbom

در غالب کار بر روی تحلیل عامل تاییدی[1] (CFA) ادامه پیدا کرد. امروزه CFA یک تکنیک بسیار مشهور است که بخشی از آن بواسطه نرم افزارهایی مثل LISREL، برای متغیرهای پیوسته وTESTFACT، برای متغیرهای دوارزشی اجرا می‌گردد. اما همزمان با روش تحلیل عامل سنتی، ایده تحلیل متغیرهای مکنون که برای متغیرهای مکنون پیوسته مورد استفاده قرار می‌گرفت، بوسیله افرادی مثل گاتمن[۲]، لرد[۳]، راش[۴]، برن‌بام[۵] و موکن[۶] با موکن[۶] با متغیرهای مشاهده شده دو ارزشی[۷] مورد استفاده قرار گرفت. این مدل‌های اندازه‌گیری که عمدتاً در حوزه تعلیم و تربیت و آزمون‌های پیشرفت تحصیلی مورد استفاده قرار می‌گیرند، تحت عنوان نظریه سوال پاسخ[۸] معروف شدند. بعدها چهارچوب نظریه سوال پاسخ به وسیله افرادی مثل سامیجیما[۹]، بوک[۱۰]، تایسن[۱۱] و اشتنبرگ[۱۲] به متغیرهای مشاهده شده چند ارزشی[۱۳] گسترش یافت. در این میان افرادی مثل لازارسفلد[۱۴]، هنری[۱۵] و کودمن[۱۶]، نظریه خصیصه مکنون را با متغیرهای مکنون طبقه‌ای[۱۷] به کار بردند، که به تحلیل ساختار مکنون[۱۸] مشهور است (بورسبوم و همکاران، ۲۰۰۳). نظریه خصیصه مکنون وقتی برای تحلیل متغیرهای مختلف اسمی (دوارزشی، چندارزشی و مقوله ای)، رتبه ای، فاصله ای و نسبی به کار گرفته شد، نام های گوناگونی به خود گرفت. تحلیل خصیصه مکنون روشی است که برای اندازه گیری صفات غیر قابل مشاهده یا مکنونی که می‌توانند عملکرد مشاهده شده در آزمون را تبیین کنند، طراحی شده است. در مقابل نظریه سوال پاسخ بر روی عملکرد واقعی مشاهده شدهٔ در آزمون متمرکز است.

1. Confirmatory factor analysis
2. Guttman
3. Lord
4. Rasch
5. Birnbaum
6. Mokken
7. Dichotomous
8. Item response theory
9. Samejima
10. Bock
11. Thissen
12. Steinberg
13. Polytomous
14. lazarsfeld
15. Henry
16. Goodman
17. Categorical latent trait
18. Latent structure analysis

مقاله حاضر[1] با ارائه تبیینی از تحلیل خصیصه مکنون ونظریه سوال پاسخ به پژوهشگران پژوهشگران و خوانندگان حوزه تعلیم و تربیت کمک خواهد کرد، تا با مبانی منطقی و فلسفی این دو روش آشنا شده، درک مفهومی خود را از این روش ها گسترش دهند و در عین حال به تفاوت بین این دو نظریه نیز پی ببرند. ابتدا، مفروضه بنیادی این نظریه، یعنی تک بعدی[2] مورد بررسی قرار می گیرد. سپس، تفاوت موجود بین پاسخ مشاهده شدهٔ سوال سوال ها و خصیصه‌های مکنون زیربنایی تشریح می‌شود.

تک بعدی بودن مدل اندازه گیری

فرایند اندازه‌گیری ویژگی های روانشناختی مستلزم آن است که فرض کنیم ویژگی های آزمودنی‌های انسانی را می‌توان از یکدیگر مجزا و تفکیک نمود و هر بار یکی از این ویژگی ها را مورد اندازه‌گیری قرار داد، طوریکه هیچ یک از خصایص دیگر، خصیصه مورد اندازه‌گیری را تحت تاثیر قرار ندهند. از سوی دیگر در اندازه گیری ویژگی های آزمودنی های انسانی باید فرض کنیم که ابزارهای اندازه‌گیری را می‌توان طوری طراحی کرد که فقط به یک ویژگی منحصر به فرد انسان حساس باشند، بدون اینکه تحت تاثیر سایر خصوصیات انسان قرار بگیرند.

رابطه بین عملکرد مشاهده شده افراد در آزمون و خصیصه مکنون

تفاوت قائل شدن بین عملکرد قابل مشاهده شده آزمودنی ها در پاسخ دادن به سوالات آزمون با توانائی زیربنائی یا خصیصه مکنون مربوطه نکته بسیار مهمی است. در اینجا با یک مثال مفصل از یک موقعیت اندازه‌گیری در دنیای فیزیکی، تفاوت موجود را تشریح می‌کنیم. نمره‌های پنج دانش آموز در دو آزمون ۱۰ سوالی در جدول شماره۱ آمده است. در آزمون (الف) آزمودنی‌های A و B در مقایسه با آزمودنی C نمره‌های بالاتری گرفته‌اند. آزمودنی C هم در مقایسه با آزمودنی D نمره بالاتری گرفته است و آزمودنی E کمترین

۱- الگوی اصلی این مقاله از فصل "introduction to latent trait theory and item response theory" تالیف Jozef p. Rya اقتباس شده است.

2. Unidimetionality

نمره را بدست آورده است و به نظر می‌رسد که در آزمون (الف) دارای کمترین توانایی است. نتایج بدست آمده در آزمون (ب) با نتایج به دست آمده از آزمون (الف) کاملاً منطبق نیست. در این آزمون، آزمودنی های A و B هر دو در مقایسه با آزمودنیC نمرات بالاتری گرفته‌اند و آزمودنی C در مقایسه با آزمودنی D نمره پائین‌تری به دست آورده است، اما آزمودنی E دوباره کمترین نمره را کسب کرده است.

به طور کلی آزمودنی ها در آزمون (الف) در مقایسه با آزمون (ب) نمرات بالاتری کسب کرده‌اند، بنابراین آزمون (الف) آسان‌تر از آزمون (ب) بوده است. اختلاف نمرات آزمودنی‌ها در دو آزمون در جدول ۱ آمده است، در مجموع متوسط اختلاف بین نمرات آزمون (الف) و (ب) سه نمره است. به عبارتی، نمرات آزمودنی ها در آزمون (ب) سه نمره پایین‌تر از نمراتشان در آزمون (الف) است. بنابراین می‌توان نمره مورد انتظار یک آزمودنی در آزمون (ب) را با استفاده از نمرات آنها در آزمون (الف) محاسبه کرد، به این صورت که نمره مورد انتظار یک آزمودنی در آزمون (ب) برابر است با نمره آن آزمودنی در آزمون (الف) منهای سه (جدول۱). تفاوت بین نمره مشاهد شده در آزمون ب و نمره مورد انتظار در این آزمون در آخرین ستون این جدول آمده است. مجموع این تفاوت ها برابر با صفر است، بنابراین متوسط خطا در نمرات پیش بینی شده آزمون (ب) که از طریق رابطه بین آزمون (الف) و (ب) محاسبه شده‌اند نیز برای همه آزمودنی ها برابر صفر است.

جدول شماره ۱: نمره های آزمودنی ها در آزمون های (الف) و (ب)

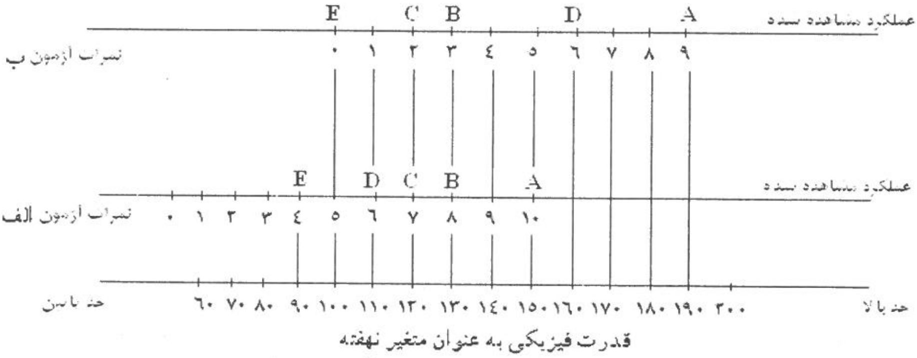
نمرات / آزمودنیها	آزمون(الف)	آزمون (ب)	تفاوت نمره‌های دو آزمون	نمره مورد انتظار در آزمون (ب)	تفاضل نمره مورد انتظار و نمره مشاهده شده در آزمون (ب)
A	۱۰	۹	۱	۷	۲
B	۸	۳	۵	۵	-۲
C	۷	۲	۵	۴	-۲
D	۶	۶	۰	۳	۳
E	۴	۰	۴	۱	-۱

اگر تفاوت های بین نمرات مشاهده شده و مورد انتظار آزمون (ب) را با نادیده گرفتن علامت های منفی و مثبت بررسی کنیم، نتایج متفاوتی مشاهده می‌شود. مجموع مقادیر قدر مطلق[1] تفاضل نمرات مشاهده شده و مورد انتظار در آزمون (ب) برابر با ۱۰ است (۱۰=۱+۳+۲+۲+۲)، که در این صورت متوسط اختلاف بین نمرات مشاهده شده و مورد انتظار برابربا ۲ می‌شود $\left(\frac{10}{5}=2\right)$. بنابراین در سطح فردی این روش منجر به یک خطای دو نمره‌ای در مقیاس ۱۰ نمره‌ای شده است که برابر با خطایی معادل ۲۰ درصد است.

سردرگمی بوجود آمده به خاطر اختلافات موجود در نمرات مشاهده شده این دو آزمون را می‌توان به وسیله خصیصه زیربنایی غیرقابل‌مشاهده آزمودنی‌ها، به شکل عمیق‌تری درک کرد. هر دو آزمون موجود در مثال قبل، برای اندازه‌گیری قدرت بدنی آزمودنی‌ها در بلند کردن وزنه‌ها طراحی شده بودند. ده سوال موجود در هر آزمون، ده وزن مختلفی هستند که از هر آزمودنی خواسته شده است، تا آنها را بلند کنند. اگر آزمودنی ها موفق به بلند کردن وزنه شوند، این کار به منزله جواب درست به سوال محسوب می‌شود. در حالی که عدم موفقیت در بلند کردن وزنه به منزله جواب غلط به سوال در نظر گرفته می‌شود. در اینجا خصیصه زیر بنایی و غیر قابل مشاهده آزمودنی ها که نمره مشاهده شده آنها را تبیین می کند، قدرت فیزیکی آزمودنی‌ها است. خصیصه‌های زیر بنایی یا غیر قابل مشاهده، تحت عنوان صفات مکنون مورد اشاره قرار می گیرند. توجه داشته باشید که یک خصیصه مکنون هیچ وقت به طور مستقیم قابل مشاهده نیست، بلکه این صفات تنها از طریق رفتارهای قابل مشاهده، نظیر پاسخ واقعی به یک سوال، استنتاج می‌شوند. رابطه بین عملکرد آزمودنی‌ها در آزمون (الف) و (ب) و خصیصه مکنون (قدرت فیزیکی) بیانگر عملکرد آزمودنی‌ها، در شکل ۱ نشان داده شده است. در این شکل قدرت فیزیکی یا همان خصیصه مکنون غیر قابل مشاهده، بوسیله یک خط افقی در قسمت پایین شکل نمایش داده شده است. نقاط موجود بر روی مقیاس به ترتیب از کم در سمت چپ

1. Absolute value

تا زیاد در سمت راست مرتب شده‌اند و خطوط نمایانگر عملکرد مشاهده شده در آزمون های (الف) و (ب) در بالای خصیصه مکنون به صورت افقی، نمایانگر دامنه خاص متغیر زیر بنایی مورد اندازه‌گیری در هر آزمون است. به علت ارتباط مستقیم این نمرات با خصیصه مکنون، می‌توان نمرات بدست آمده از آزمون های (الف) و (ب) را به صورت معنی‌دارتری تفسیر کرد. به طور مسلم، آزمون (الف) نسبت به آزمون (ب) آسان تر است، چرا که گسترهٔ آن دامنه پایین‌تری در خصیصه مکنون را به خود اختصاص داده است.



شکل شماره ۱: ارتباط بین عملکرد مشاهده شده در آزمونهای (الف) و (ب)

آزمودنی E در آزمون (الف) نمره‌ای برابر با ۴ گرفته است، عملکرد او نشان دهنده آن است که این آزمودنی توانایی بلند کردن ۹۰ کیلو را دارد. نمرهٔ صفر آزمودنی E در آزمون ب، در غیاب اطّلاعات تکمیلی دیگر، معنی و مفهوم واقعی خود را ندارد. نمره صفر این آزمودنی در آزمون (ب) نشان می دهد که قدرت آزمودنی در نقطه‌ای زیر حداقل توانایی لازم در آزمون (ب) (بلند کردن ۱۱۰ کیلو، یعنی سبک‌ترین وزن (آسان‌ترین سوال)) ، قرار دارد. بنابراین، بدون توجه به سایر اطلاعات، روش دیگری برای پیدا کردن محل آزمودنی با نمره‌ای معادل صفر در این آزمون وجود ندارد، چرا که جایگاه این آزمودنی، جایی زیر پایین‌ترین نقطه منعکس شده بوسیله آزمون (ب) تنزل پیدا کرده است. البته در عین حال، در انتهای دیگر پیوستار این آزمون، پیدا کردن جایگاه آزمودنی با نمره کامل در آزمون نیز غیر ممکن است. آزمودنی A در آزمون (الف) نمره ۱۰ گرفته است، این امر بدان معنی است که این آزمودنی می‌تواند حداقل ۱۵۰کیلو را بلند

کند، که این امر نشانگر آن است که بدون اطلاعات مکمل، هیچ روشی برای اندازه‌گیری توانایی این آزمودنی در خصیصه مکنون وجود ندارد. با استفاده از نتایج آزمون (ب) ما می‌توانیم بفهمیم که آزمودنی A توانایی بلندکردن ۱۹۰ کیلو را دارد، در حالی که نمی توانیم این توانایی را با بررسی نتایج آزمون (الف) برآورد کنیم.

عملکرد آزمودنی E به عنوان ضعیف‌ترین و عملکرد آزمودنی A به عنوان قوی ترین آزمودنی، بیانگر یک اصل بدیهی و مهم اندازه‌گیری است: غیر ممکن است بتوان توانایی مکنون آزمودنی‌هایی که در یک آزمون نمره کامل یا صفر گرفته‌اند را اندازه‌گیری کرد. عملکرد آزمودنی D در دو آزمون نشان دهنده ناهماهنگی برآوردهای توانایی مکنون‌اش در هر دو آزمون است. با توجه به این امر، می‌توان پرسید که آزمودنی D توانایی بلند کردن کدام یک از دو وزنه را دارد؟ ۱۱۰ کیلو یا ۱۶۰ کیلو؟ ناهماهنگی موجود در نمرات آزمودنی D مدرک روشنی است که نشان می‌دهد مفروضه تک بعدی بودن نقض شده است. نقض شدن این مفروضه می‌تواند از این موارد ناشی شده باشد: خود آزمودنی، آزمون یا موقعیت اجرای آزمون. به عنوان مثال، اگر آزمودنی D، شب قبل از آزمون (الف) بخوبی نخوابیده باشد، بنابراین عملکرد او در آزمون (الف) نمی‌تواند منعکس کننده توانایی واقعی او در آن زمینه باشد و یا ممکن است هنگام پاسخ گویی آزمودنی D به سوال یک در هر دو آزمون، وزنه‌ها به طور صحیح درجه بندی نشده باشند، در چنین موقعیت هایی ناهماهنگی در اندازه‌گیری توانایی از نقص در ابزار اندازه‌گیری ناشی شده است. در نهایت، این امکان وجود دارد که آزمودنی D در زمان انجام آزمون عجله کرده باشد. بنابراین، برآورد قدرت[1] این آزمودنی بوسیله سرعت[2] مختل شده است. در هر صورت تبیین عملکرد ناهماهنگ آزمودنی D در کنار تبیین نمره آزمون باید با استفاده از سایر منابع اطلاعاتی مکمل، استنتاج و تبیین شود.

هماهنگی نمرات آزمودنی های C و B در هر دو آزمون، حاکی از برآورد توانایی مکنون آنها است. نمره بدست آمده در هر دو آزمون نشان می‌دهد که آزمودنی C توانائی بلند کردن ۱۲۰ کیلو را دارد و آزمودنی B می‌تواند ۱۳۰ کیلو را بلند کند. واضح است که

1. Strength
2. Speed

این برآوردهای توانایی مکنون، مستقل از آزمون مورد استفاده برای استنباط این توانایی می‌باشد. نمرات این آزمودنی‌ها در آزمون (الف) از نمرات آنها در آزمون (ب) به اندازهٔ پنج نمره متفاوت است. شکل ۱ نشان می‌دهد که دامنه توانایی‌های ارائه شده بوسیله هر دو آزمون بر روی پیوستار خصیصه مکنون، به اندازه پنج نمره در همه نقاط پیوستار خصیصه مکنون متفاوت است. بنابراین آزمون (الف) به اندازهٔ پنج نمره از آزمون (ب) آسانتر است، نه سه نمره‌ای که قبلا براساس نمرات مشاهده شده در جدول ۱ به دست آمده است.

در مثال قبلی، عملکرد آزمودنی‌ها در آزمون بر اساس پاسخ های مشاهده شده سوال ها، بیانگر نوع داده‌هایی است که در نظریه سوال پاسخ بررسی می‌شوند. نظریه سوال پاسخ تلاش می‌کند تا از نظر آماری توصیف دقیقی از چنین داده‌هایی ارائه کند. بازنمایی عملکرد مشاهده شده در خصیصه مکنون غیر قابل مشاهده، همانند تفسیر عملکرد مشاهده شده آزمودنی‌ها بر حسب خصیصه مکنون، انواع فعالیت های بیانگر تحلیل خصیصه مکنون را نشان می‌دهد. با استناد به این مثال می‌توان نتیجه گیری کرد که: اول، تک بعدی بودن به معنای واقعی خود برای اندازه‌گیری انواع متغیرهای علوم طبیعی ضروری است. دوم، همه موقعیت های مربوط به مشاهده عملکرد آزمودنی‌ها، شرایط تک بعدی بودن را به معنای واقعی برآورده نمی‌کنند. شرایط و موقعیت مشاهده عملکرد آزمودنی‌ها تا جایی که امکان دارد باید به دقت کنترل شود. به عنوان مثال، در تلاش برای اندازه‌گیری قدرت آزمودنی‌ها، بهتر آن است که هر تلاش آزمودنی در جواب به هر یک از سوال ها به صورت متوالی باشد، تا بدین طریق تأثیر عوامل خارجی را به حداقل برسانیم. سوم، داده‌های جمع آوری شده مربوط به عملکرد باید به دقت بررسی شوند، تا وجود مفروضه بنیادی تک بعدی بودن مورد بررسی قرار گیرد. به عبارتی، مفروضه تک بعدی بودن در واقع یک مفروضه نیست، بلکه فرضیه‌ای است که اعتبار آن می‌تواند مورد قضاوت قرار گیرد. چهارم، به طور مستقیم نمی‌توان، خصیصه مکنون تک بعدی را در خصوص آزمودنی‌هایی که در یک آزمون نمره کامل یا صفر گرفته‌اند، مورد سنجش قرار داد. پنجم، آزمودنی‌هایی که نمرهٔ مشاهده شدهٔ آنها اطلاعات ناهماهنگی در خصوص خصیصه مکنون ارائه می‌دهند (مثل آزمودنی D)، به معنی واقعی مورد اندازه‌گیری قرار نگرفته‌اند.

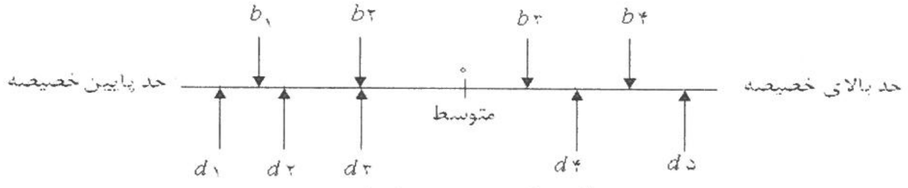
عملکرد چنین آزمودنی‌هایی را می‌توان به منظور بررسی و تبیین این قبیل ناهماهنگی‌ها مورد بررسی بیشتری قرار داد. اما داده‌های مربوط به عملکرد این آزمودنی‌ها نمی‌تواند به اندازه‌گیری در یک خصیصه مکنون تک بعدی واقعی منجر شود. چنین آزمودنی‌هایی باید از فرایند تحلیل کنار گذاشته شوند. ششم، توانایی‌های افراد در خصیصه مکنون را می‌توان مستقل از آزمون معینی برآورد نمود، نظیر آزمودنیهای C و B. هفتم، حتی اگر فرضیه تک‌بعدی بودن به معنای واقعی رد شود، برای بیرون کشیدن اطلاعات آماری دربارهٔ آزمودنی‌ها، روشهای خاصی را می‌توان مورد استفاده قرار داد. این روش ها را می‌توان برای آزمودنی‌هایی با نمرهٔ کامل یا صفر و حتی برای دانش‌آموزانی که عملکرد آنها در آزمون با خصیصه مکنون واقعی ناسازگار است، مورد استفاده قرار داد. اما باید بدانید که کاربرد این روشها، فرضیه تک بعدی بودن به معنی دقیق و واقعی را نقض می‌کند، بنابراین نتایج چنین روش هایی توانایی معلمان و پژوهشگران را برای ایجاد ارتباط مستقیم بین عملکرد مشاهده شده در آزمون با خصیصه مکنون تک‌بعدی واقعی را محدود می‌کند.

اندازه‌گیری خصیصه مکنون

اولین گام در فرایند اندازه‌گیری متغیر مورد نظر تعریف مفهومی آن است. متغیرهای مورد علاقه پژوهشگران حوزه تعلیم و تربیت را می‌توان به وسیله آموزش تحت تأثیر قرار داد تا جایگاه[1] دانش‌آموزان در خصیصه مکنون مربوطه ارتقاء یابد. از آنجا که از نظر آموزشی اندازه‌گیری دقیق متغیرهای مربوطه برای تشخیص، جایابی درست دانش‌آموزان، ایجاد و اصلاح برنامه‌های درسی، تعیین اهداف و استانداردهای آموزشی و ارزشیابی آموزشی بسیار حیاتی است، بنابراین متغیرهایی از این نوع و اندازه‌گیری آنها از عمده علایق پژوهشی معلمان و پژوهشگران می‌باشند. در محیط های آموزشی متغیرهایی که به یادگیری در مدرسه ارتباط دارند، از اهمیت عمده‌ای برخوردارند. فرض کنید که گروهی از معلمان به اندازه‌گیری توانایی دانش‌آموزان در مهارت های ریاضی پایه ششم یا توانایی مهارتهای زبانی پایه سوم و یا هر خصیصه دیگر دانش آموزان که آموزش بر روی آن تاثیر دارد،

1. Location

علاقمند باشند. تحت مفروضه بنیادی تک بعدی بودن، متغیر آموزشی را می‌توان به عنوان یک خصیصه مکنون در نظر گرفت که هم دانش‌آموزان و هم سوالات یا فعالیت های موجود در تست را می توان بر روی پیوستار آن خصیصه به ترتیب جایگاه نشان داد. یک موقعیت فرضی در شکل شماره۲ نشان داده شده است. در این شکل خط افقی نشانگر خصیصه مکنون می‌باشد و مقدار توانایی که یک آزمودنی در خصیصه مکنون دارد در بالای این خط با علامت b نمایش داده شده است. به طور کلی، توانایی j امین آزمودنی با علامت b_j نشان داده شده است، مثلا b_1 و b_2 به ترتیب توانایی اولین و دومین آزمودنی در مقیاس خصیصه مکنون می‌باشد. دشواری سوال ها در زیر این خط با علامت d نمایش داده شده است، در کل دشواری iامین سوال با علامت d_i نشان داده شده است، به طوریکه d_1 و d_2 به ترتیب دشواری اولین و دومین سوال است.



شکل شماره ۲: خصیصه نهفته فرضی

آزمودنی‌هایی با توانایی کم و سوال های آسان در سمت چپ این شکل و مقادیر بالاتر توانایی و سوال های دشوار در سمت راست شکل قرار دارند. با انتخاب صفر به عنوان مبدأ، آزمودنی‌ها و سوال هایی که جایگاه آنها در زیر مقدار متوسط خصیصه مکنون (نقطه صفر) است در دامنه منفی قرار می گیرند و آزمودنی‌ها و سوال ها بالاتر از مقدار متوسط خصیصه مکنون، در دامنه مثبت واقع می شوند.

مفروضه بنیادی تک بعدی بودن سه فرضیه آزمون‌پذیر را فراهم می کند: اول، تنها عامل تاثیرگذار بر موقعیت آزمودنی ها در خصیصه مکنون، توانایی مورد اندازه‌گیری آنها است نه سایر ویژگی های آنها. دوم، تنها دشواری سوال ها بر موقعیت سوال ها در مقیاس خصیصه مکنون تاثیر می گذارد نه سایر ویژگی های سوال ها. سوم، خصیصه‌ای که

آزمودنی ها بر اساس توانایی و سوال ها بر اساس دشواری بر روی آن مرتب شده‌اند، مشابه بوده و مقیاس هر دو مورد بر روی توزیع طبیعی است. بنابراین برآوردهایی که از توانایی آزمودنی ها و دشواری سوالات صورت می گیرد بر روی این توزیع قرار می گیرد که به لحاظ نظری دامنه‌ای از منفی بی‌نهایت تا مثبت بی‌نهایت دارد (لرد ۱۹۸۰؛ امبرستون و رایز، ۲۰۰۰). تحت این شرایط فرضی، بحث کردن در مورد احتمال این که یک آزمودنی با مقدار توانایی معین به سوالی با مقدار دشواری مشخص جواب صحیح دهد، به لحاظ نظری و عملی بسیار آگاهی دهنده و مفید خواهد بود. منطقی است که هر چه جایگاه فرد در پیوستار خصیصه بالاتر باشد (یعنی به سمت مثبت پیوستار حرکت کند) احتمال پاسخ درست به سوال افزایش می یابد و بالعکس. به عنوان مثال، زمانی که یک آزمودنی با توانایی b_2 به سوالی با دشواری d_2 پاسخ دهد، محتمل‌ترین نتیجه این است به سوال مورد نظر جواب درست بدهد. چرا که بر اساس شکل۲ توانایی این آزمودنی بیشتر از دشواری سوال است. اما زمانی که یک آزمودنی با توانایی b_2 به یک سوال با دشواری d_4 پاسخ دهد، محتمل‌ترین نتیجه این است که به آن سوال جواب غلط بدهد، چرا که توانایی آزمودنی پایین‌تر از دشواری سوال است. سرانجام این که بر اساس شکل۲ می‌توان گفت اگر یک آزمودنی با توانائی b_2 به سوالی با دشواری d_3 جواب دهد، محتمل‌ترین نتیجه نامشخص است، چرا که $b_2 = d_3$ است. در چنین شرایطی این آزمودنی برخی اوقات به سوال مورد نظر جواب درست و برخی اوقات نیز به آن جواب غلط خواهد داد، به عبارتی منطقی‌ترین احتمال در چنین مواردی ۰/۵ است. با توجه به این مطلب، اگر مفروضه بنیادی تک بعدی بودن برقرار باشد، می‌توان سه گزاره زیر را دربارهٔ احتمال پاسخ صحیح به مقادیر گوناگون $\left(b_j - d_i\right)$ مطرح کرد:

اگر $b_j < d_i$ آنگاه $P<۰/۵$

اگر $b_j = d_i$ آنگاه $P<۰/۵$

اگر $b_j > d_i$ آنگاه $P<۰/۵$

شکل ریاضی رابطه بین توانایی آزمودنی و دشواری سوال بر روی پیوستار خصیصه مکنون را متخصصان اندازه‌گیری از زمان راش به صورتی که در پی می‌آید بیان کرده‌اند

(لرد و نویک، ۱۹۶۸). همان طور که می دانیم دامنه احتمال از صفر تا یک است، در حالی که به لحاظ نظری دامنه $(b_j - d_i)$ می‌تواند از منفی بی‌نهایت $(-\infty)$تا مثبت بی‌نهایت $(+\infty)$ است. مقادیر منفی بزرگ در رابطه $(b_j - d_i)$ زمانی رخ می‌دهند که آزمودنی‌ها با توانایی بسیار پایین به سوال های بسیار دشوار پاسخ می‌دهند و مقادیر مثبت بزرگ در این رابطه زمانی به دست می‌آیند که آزمودنی هایی با توانایی بسیار بالا به سوال های بسیار آسان پاسخ دهند. بنابراین امکان تفسیر مستقیم عبارت $(b_j - d_i)$ به صورت احتمال پاسخ صحیح وجود ندارد. آنچه که برای انجام این کار، مورد نیاز است یک تابع ریاضی از عبارت $(b_j - d_i)$ می‌باشد، تا امکان تفسیر مستقیم این عبارت به صورت احتمال پاسخ صحیح، به وجود آید. بررسی منابع مختلف نشان می دهد که سه روش برای تبدیل این رابطه به صورت احتمال وجود دارد که دو روش آن در زیر توصیف می شوند و روش سوم در قسمت مدل دو پارامتری ارائه می شود.

روش اول

اولین بار راش ریاضی دان دانمارکی با استناد به اصل امساک[1] علمی، تابعی به نام تابع لوجستیک ساده[2] را برای این منظور انتخاب کرد.

(معادله ۱)

$$F(b_j - d_i) = \frac{A^{(b_j - d_i)}}{1 + A^{(b_j - d_i)}}$$

در تابع لوجستیک ساده $(b_j - d_i)$ نما یا توانی است که مبنای آن A می‌باشد. مقدار مبنای A کاملاً اختیاری است. صرف نظر از این که چه مقداری در مبنا به جای A قرار گیرد، دامنه این تابع همیشه بین صفر و یک است. در عمل معمولاً مقدار لگاریتم طبیعی نپر یعنی ۲/۷۱۸۳ یکی از پر استفاده‌ترین مقادیری است که به عنوان مبنا استفاده می‌شود. به عنوان مثال، اگر پاسخ یک آزمودنی به هر یک از سوال های یک آزمون را با نماد X

1. Parsimony

2. Simple logistic function

مشخص کنیم، پس متغیر تصادفی X می‌تواند یکی از دو مقدار (۱) و (۰) را به خود اختصاص دهد. اگر آزمودنی به سوال مورد نظر پاسخ درست بدهد این متغیر ارزش (۱) و در صورتی که پاسخ غلط بدهد این متغیر ارزش (۰) می گیرد. از سوی دیگر، استفاده از علامت های b_i و d_j که در معادله شماره۱ به ترتیب به توانایی و دشواری برآورد شده از نمونه اشاره دارند بسیار رایج است. اما برای نشان دادن این پارامترها در جامعه، باید θ_i (توانایی) را جایگزین b_i و β_j (دشواری) را جایگزین d_j نمود. به عبارتی توانایی و دشواری واقعی آزمودنی ها و سوال های را هرگز نمی توان مستقیماً مشاهده کرد، بلکه در واقع توانایی واقعی آزمودنی ها (θ_i) و دشواری واقعی سوال های (β_j) با استفاده از مقادیر به دست آمده از نمونه برآورد می شوند. در اینجا b_i و d_j مقادیر برآورد شده پارامترهای خصیصه مکنون هستند. با استفاده از این مفهوم، احتمال پاسخ آزمودنی ها به سوال هایی که به صورت دو مقوله‌ای نمره گذاری می‌شوند به صورت زیر بیان می‌شود.

(معادله۲)

$$P(\mathrm{X}=1 \mid \theta,\beta)=\frac{e^{\mathrm{X}(\theta-\beta)}}{1+e^{(\theta-\beta)}}$$

برای نشان دادن چگونگی عملکرد این معادله فرض کنید که مقادیر ۲= b و۱= d برای یک سوال فرضی در دست باشند:

احتمال برای $\mathrm{X}=1$ برابراست با:

$$P(X=۱ \mid b=۲,d=۱)=\frac{e^{۱(۲-۱)}}{۱+e^{(۲-۱)}}=\frac{۲/۷۱۸}{۱+۲/۷۱۸}=./۷۴$$

و احتمال برای $X=۰$ برابراست با:

$$P(X=0 \mid b=2,d=1)=\frac{e^{0(2-1)}}{1+e^{(2-1)}}=\frac{1}{1+2/718}=0/26$$

مدل ارائه شده در معادله شماره ۲، در منابع تخصصی تحت دو عنوان مدل تک پارامتری[1] و مدل راش معروف است.

روش دوم

روش دیگری نیز برای رسیدن به یک الگوی احتمالی در خصوص رابطه $(\theta_j - \beta_i)$ مورد استفاده قرار گرفته است. ارائه این روش ها نیازمند برخی مفاهیم آماری است. شانس[2] وقوع یک حادثه برابر است با:

(معادله ۳)

$$odds = \frac{P}{1-P} = \frac{P}{q}$$

در ارتباط با موضوع حاضر p برابر با احتمال پاسخ صحیح به سوال و q احتمال عدم پاسخ صحیح به سوال است، که با توجه به این مطلب می توان شانس موفقیت در یک سوال را با توجه به رابطه $(\theta_j - \beta_i)$ به صورت زیر تعریف کرد.

(معادله ۴)

$$odds = \frac{P}{1-P} = (\theta_j - \beta_i)$$

شانس به عنوان شاخصی از احتمال وقوع یک حادثه با مشکلی همراه است. به این معنی که اگر حادثه ای دارای احتمال زیادی باشد، شانس وقوع آن می تواند مقادیر بی نهایت بزرگ و اگر احتمال آن خیلی کم باشد شانس تنها می تواند کسری بین صفر تا یک باشد. به عبارتی دامنه این شاخص بین صفر تا مثبت بی نهایت است (ناگفته نماند که راش نیز دامنه θ و β را به اعداد بین $[o, \infty]$ محدود کرد). ناقرینگی[3] موجود در نسبت شانس را می توان با گرفتن لگاریتم طبیعی از شانس برطرف کرد. به این ترتیب مقیاس جدیدی با عنوان لگاریتم شانس یا لوجیت[4]، بدست می آید، که براساس آن دامنه p بین صفر و یک قرار می گیرد. در نتیجه فرم کلی معادله قبلی به صورت زیر تبدیل می شود:

1. One parameter model
2. Odds
3. Asymmetry
4. logit

(معادله ۵)

$$\ln\frac{P}{1-P}=\ln(\theta_j-\beta_i)\rightarrow\ln\frac{P}{1-P}=\ln\theta_j-\ln\beta_i$$

می توان با مساوی قرار دادن $\ln\theta_j=\theta_j$ و $\ln\beta_i=\beta_i$ رابطه بالا را به صورت زیر نوشت:

(معادله ۶)

$$\ln\frac{P}{1-P}=(\theta_j-\beta_i)\Rightarrow\frac{P}{1-P}=e^{(\theta_j-\beta_i)}\Rightarrow P=e^{(\theta_j-\beta_i)}-Pe^{(\theta_j-\beta_i)}\Rightarrow P+Pe^{(\theta_j-\beta_i)}=e^{(\theta_j-\beta_i)}$$

$$P\left(1+e^{(\theta_j-\beta_i)}\right)=e^{(\theta_j-\beta_i)}\Rightarrow P=\frac{e^{(\theta_j-\beta_i)}}{1+e^{(\theta_j-\beta_i)}}\Rightarrow P=\left[\frac{1+e^{(\theta_j-\beta_i)}}{e^{(\theta_j-\beta_i)}}\right]^{-1}\Rightarrow P=\left[\frac{1}{e^{(\theta-\beta)}}+\frac{e^{(\theta_j-\beta_i)}}{e^{(\theta_j-\beta_i)}}\right]^{-1}$$

$$P=\left[e^{-(\theta_j-\beta_i)}+1\right]^{-1}\Rightarrow P=\frac{1}{1+e^{-(\theta_j-\beta_i)}}\Rightarrow P=\left[1+\exp-(\theta_j-\beta_i)\right]^{-1}$$

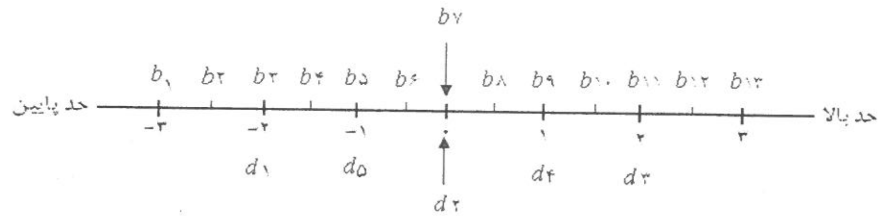
این رابطه همان مدل یک پارامتری لوجستیک است که در منابع مختلف نظریه سوال پاسخ ذکر شده است. با توجه به این رابطه، در نهایت می توان مدل های دو و سه پارامتری را نیز از طریق آن استخراج کرد. وجه تسمیه این مدل به مدل تک پارامتری آن است که شکل ریاضی این تابع فقط یکی از پارامترهای سوال یعنی دشواری سوال را الگوسازی می‌کند. از طرفی دیگر، وجه تسمیه این مدل به مدل راش، به خاطر اذعان نمودن به کارهای پیشگامانه جرج راش است. البته، در ادامه خواهیم دید که بین مدل راش و یک پارامتری تفاوت هایی وجود دارد. راش(۱۹۶۰) اولین تبیین کامل از منطق و کارکرد این مدل را ارائه کرد. مدل راش (معادله شماره۲) برای موقعیت هایی با پاسخ های دو مقوله ای صفر و یک کاربرد دارد، بنابراین به مدل ترسم شده در معادله شماره۲ مدل دو ارزشی[1] نیز می گویند.

شکل شماره۳ و جدول شماره ۲ چگونگی کارکرد مدل دو ارزشی راش را با استفاده از توانایی۱۳ آزمودنی برای سه سوال در خصیصه مکنون را نشان می‌دهد. در شکل شماره۳ دامنه توانایی از ۳- تا ۳+ با فواصل ۰/۵ درجه بندی شده و دشواری سوال ها برابر

1. dichotomous model

با $d_3 = ۲$ ، $d_2 = ۰$ ، $d_1 = -۲$ می باشد. پاسخ های مورد انتظار آزمودنی‌ها با استفاده از ۱۳ توانایی مختلف در هر سه سوال به صورت احتمال در جدول شماره ۲ ارائه شده اند.

یکی از ویژگی های بسیار مهم مدل راش این است که یک مدل احتمالی[1] است نه جبری[2]. به عنوان مثال با استفاده از مدل راش در جدول شماره۲، احتمال این که یک دانش آموز با توانائی۱- به دومین سوال پاسخ صحیح بدهد برابر $P(X=۱ \mid b_5 ، d_2) = ۰/۲۶۹$ است. بنابراین، اگر این آزمودنی به سوال شماره۲ پاسخ دهد، محتمل‌ترین نتیجه آن است که به سوال پاسخ غلط بدهد، این مطلب به زبان احتمالات به این معنی است که اگر صد نفر از آزمودنی‌های مشابه این آزمودنی، به این سوال پاسخ دهند، انتظار داریم که تقریباً ۲۷ آزمودنی به این سوال پاسخ درست بدهند. این نتیجه بیانگر این واقعیت است که مفروضه بنیادی تک بعدی بودن حالت ساده شده واقعیت پیچیده دنیای خارجی است.



شکل شماره ۲- توزیع توانایی آزمودنی ها و دشواری سوال ها

1. probabilistic

2. deterministic

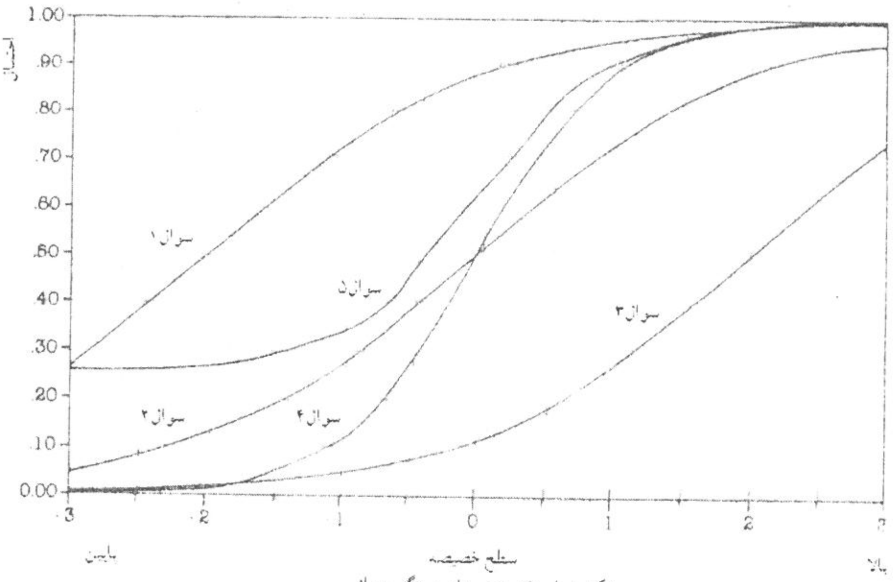
جدول شماره۲: احتمال پاسخ صحیح برای آزمودنی هایی با توانایی های مختلف در ۳ سوال

توانایی(تتا)	دشواری					
	سوال ۱		سوال ۲		سوال ۳	
	بتا = ۲-		بتا=۰		بتا =۲	
	(b-d)	احتمال	(b-d)	احتمال	(b-d)	احتمال
۳-	۱/۰-	۰/۲۶۹	۳/۰-	۰/۰۴۷	۵/۰-	۰/۰۰۷
۲/۵-	۰/۵-	۰/۳۷۸	۲/۵-	۰/۰۷۶	۴/۵-	۰/۰۱۱
۲-	۰/۰	۰/۵۰۰	۲-	۰/۱۱۹	۴/۰-	۰/۰۱۸
۱/۵-	۰/۵۰	۰/۶۲۳	۱/۵-	۰/۱۸۲	۳/۵-	۰/۰۲۹
۱-	۱/۰	۰/۷۳۱	۱-	۰/۲۶۹	۳/۰-	۰/۰۴۷
۰/۵-	۱/۵	۰/۸۱۸	۰/۵-	۰/۳۷۸	۲/۵-	۰/۰۷۶
۰	۲/۰	۰/۸۸۱	۰	۰/۵۰۰	۲/۰-	۰/۱۱۹
۰/۵	۲/۵	۰/۹۲۴	۰/۵۰	۰/۶۲۳	۱/۵-	۰/۱۸۲
۱	۳/۰	۰/۹۵۳	۱/۰	۰/۷۳۱	۱/۰-	۰/۲۶۹
۱/۵	۳/۵	۰/۹۷۱	۱/۵	۰/۸۱۸	۰/۵-	۰/۳۷۸
۲	۴/۰	۰/۹۸۲	۲/۰	۰/۸۸۱	۰	۰/۵۰۰
۲/۵	۴/۵	۰/۹۸۹	۲/۵	۰/۹۲۴	۰/۵	۰/۶۲۳
۳	۵/۰	۰/۹۹۳	۳/۰	۰/۹۵۳	۱/۰	۰/۷۳۱

منحنی ویژگی سوال

احتمال های ارائه شده برای سوالات (شکل شماره۳) را می‌توان به صورت بصری با منحنی ویژگی سوال[1] ترسیم کرد. در این نمودارها، محور افقی خصیصه مکنون و محور عمودی احتمال پاسخ صحیح را نشان می دهد. در شکل شماره۴ پنج منحنی ویژگی سوال (ICC) نشان داده شده‌اند. ابتدا به ICC سه سوال ۱، ۲ و۳ از شکل شماره۳ توجه کنید. برای پی بردن به چگونگی کارکرد منحنی‌های ویژگی سوال توانایی معادل با خصیصه مکنون ۱/۵- را در محور افقی پیدا کنید و یک خط عمودی از این نقطه به طرف بالا بکشید تا ICC مربوط به سوال۱ را قطع کند، سپس از محل تقاطع یک خط افقی به طرف محور احتمالات در سمت چپ رسم کنید. چنانکه خواهید دید این خط نقطه P=۰/۶۲۳ را بر روی محور عمودی قطع می‌کند. این مطلب به این معنا است که احتمال اینکه یک آزمودنی با توانایی ۱/۵- به سوال اول جواب درست بدهد برابر با ۰/۶۲۳ خواهد بود. به این ترتیب، احتمال این که آزمودنی‌ها در هر موقعیتی بر روی مقیاس خصیصه مکنون به هر یک از سوالات جواب درست بدهند به این صورت قابل تبیین خواهد بود.

1. Item Characteristic Curves (ICC)



شکل شماره ۲: منحنی های ویژگی سوال

مدل دو پارامتری

دو منحنی ویژگی سوال در شکل شماره۴، نمایانگر سوالات چهارم و پنجم موجود در شکل شماره۳ هستند. همان طور که می‌بینید بین ICC_2 و ICC_4 تفاوت عمده‌ای وجود دارد، در واقع ICC_4 در مقایسه با ICC_2 شیب بیشتری دارد. مشخص است که وقتی توانائی از ۰/۵- به ۰/۵ تغییر می کند احتمال این که آزمودنی‌ها به سوال ۴ به طور صحیح پاسخ دهند از ۰/۱۸ به ۰/۷۲ افزایش می یابد. احتمال پاسخ صحیح آزمودنی هایی با توانایی ۰/۵- به سوال ۴، بسیار کم است. یعنی سوال مورد بررسی برای آزمودنی هایی که در این سطح از توانایی می‌باشند، بسیار دشوار است. برای آزمودنی هایی با توانایی ۰/۵+ احتمال پاسخ درست به این سوال بسیار زیاد است، یعنی سوال مورد بررسی برای آزمودنی هایی با این سطح از توانایی بسیار آسان است. بنابراین سوال ۴ در مقایسه با سوال ۲ می‌تواند بین آزمودنی‌هایی که در این دو سطح از توانایی قرار دارند، تمایز بیشتری ایجاد کند. وجود سوالاتی از این نوع که دارای قدرت تشخیص متفاوتی با سایر سوالات می‌باشند، اغلب در داده های سوال پاسخ واقعی دیده می‌شوند. اینگونه موارد باعث شده‌اند که بسیاری از روان

سنج ها دومین پارامتر سوال را که قدرت تشخیص(آلفا) نام دارد، در تحلیل داده‌ها به کار برند، معادله ریاضی این مدل از این قرار است:

معادله(۷)

$$P\left(X=1\middle|\alpha_i,\theta_j,\beta_i\right)=\frac{e^{\alpha_i(\theta_j-\beta_i)}}{1+e^{\alpha_i(\theta_j-\beta_i)}}$$

با توجه به اثبات معادله مدل یک پارامتری می‌توان این معادله را نیز به راحتی اثبات کرد. در این معادله، α_i بیانگر قدرت تشخیص هر سوال است. سایر نمادهای به کار رفته در این معادله همان نمادهای تعریف شده در مدل تک پارامتری هستند. اکنون وقت آن رسیده که سومین روش برای رسیدن به یک الگوی احتمالی در خصوص رابطه $(\theta_j-\beta_i)$ توصیف شود.

روش سوم

روش دیگری که برای رسیدن به یک الگوی احتمالی در خصوص رابطه $(\theta_j-\beta_i)$ مورد استفاده می گیرد تابع تراکمی توزیع نرمال[1] است. در واقع پژوهش های اولیه در خصوص نظریه سوال پاسخ با مدل اجایو نرمال انجام گرفت و ما در اینجا نشان می دهیم که چگونه با توجه با استفاده از تابع اجایو نرمال یک الگوی احتمالی در خصوص عبارت $(\theta_j-\beta_i)$ ارائه می شود (بیکر و کیم، ۲۰۰۴). اکثر دانشجویان رشته های علوم رفتاری با فرمول تابع تراکمی توزیع نرمال و عناصر آن آشنایی دارند، در این جا این تابع به عنوان معادله یک منحنی مورد استفاده قرار گرفته نه به عنوان تابع توزیع تراکمی که در نظریه آمار مورد استفاده قرار می گیرد:

(معادله۸)

$$p_i(\theta)=p(\mu_i,\sigma_i,\beta_i)=\Phi(z_i)\int_{z_i=-(\theta-\mu_i)\sigma_i}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz,$$

1. Cumulative normal distribution function

آنچه در این تابع تعیین کننده است عبارت $-Z_i = -(\theta - \mu_i)/\sigma_i = 1/\sigma_i(\theta - \mu_i)$ است. در این فرمول، μ_i میانگین، σ_i انحراف استاندارد، θ توانایی و z انحراف نرمال است. در عبارت $(\theta_j - \beta_i)$، β_i برابر با نقطه ای در پیوستار توانایی است که احتمال پاسخ صحیح به آن ۰/۵ است و متناظر با میانه منحنی ویژگی سوال است. در تابع اجایو نرمال نقطه ای که احتمال پاسخ صحیح به آن برابر ۰/۵ است میانگین است. تحت این شرایط می‌توان $\beta_i = \mu_i$ قرار داد. در تابع اجایو نرمال σ_i برابر با پراکندگی توزیع نرمال است. در یک توزیع نرمال هرچه σ_i افزایش یابد شیب منحنی در نقطه β_i کاهش می یابد و برعکس. به علاوه قدرت تشخیص سوال(α_i) برابر با شیب منحنی ویژگی سوال در نقطه β_i است. بنابراین، می توان گفت بین شیب منحنی ویژگی سوال و انحراف استاندارد اجایو نرمال رابطه معکوس وجود دارد. در این صورت می توان رابطه $\alpha_i = 1/\sigma_i$ را در نظر گرفت. در این صورت می توان گفت هرچه قدر σ_i افزایش یابد α_i کاهش خواهد یافت و بر عکس. بر این اساس می توان رابطه زیر را در نظر گرفت:

(معادله ۹)

$$Z_i = (\theta - \mu_i)/\sigma_i = 1/\sigma_i(\theta - \mu_i) \Rightarrow \alpha_i(\theta - \beta_i)$$

بر این اساس می توان گفت پارامترهای β و σ در منحنی ویژگی سوال به ترتیب توابع یک به یک از پارامترهای μ و σ در تابع اجایو نرمال هستند. بنابراین می توان احتمال پاسخ گویی به یک سوال را از طریق مدل اجایو نرمال نیز برآورد کرد، که نتایج آن به مقادیر به دست آمده از مدل لوجستیک نزدیک است. ذکر این نکته نیز لازم است که در این جا σ_i پارامتر اجایو نرمال است و رابطه ای با σ_θ (انحراف استاندارد نمره های افراد) در پیوستار توانایی آزمون دهندگان ندارد. معمولا مقیاس نمره های حاصل را به توزیعی با میانگین صفر و انحراف استاندار یک تبدیل می کنند. اکنون به مدل دو پارامتری برگشته و برخی از ویژگی های آن را ذکر می کنیم.

بکارگیری مدل دوپارامتری دو مشکل منطقی بوجود می‌آورد: اول، پارامتر تشخیص برای مدل سازی یک ویژگی مکنون بنیادی سوال های آزمون به کار نمی‌رود. بلکه این

شاخص به این دلیل در این مدل گنجانده شده که توصیف مفصل تری از پاسخ‌های مشاهده شده سوال‌های آزمون فراهم کند.دوم، گنجاندن پارامتر تشخیص، تفسیر دشواری سوال های را مبهم می کند. به عنوان مثال، احتمال این که آزمودنی‌هایی با توانایی ۰/۵- سوال۲ را به طور صحیح جواب دهند بیشتر از سوال۴ است. بنابراین احتمال پاسخ صحیح به سوال۲ درمقایسه با سوال۴ بیشتر است، از این رو سوال۲ برای چنین افرادی باید آسان‌تر باشد. در مورد سوال هایی با قدرت تشخیص بسیار متغیر این امکان وجود ندارد که سوال ها را بر اساس دشواری در سرتاسر دامنه توانایی آزمودنی‌ها رتبه‌بندی کنیم، علت این امر آن است که رتبه‌بندی نسبی سوال ها وابسته به توانایی آزمودنی‌ها است.

مدل سه پارامتری[۱]

ICC_5 در شکل شماره۴ نشان دهنده نوع دیگری از سوال های می‌باشد. کرانه پایین‌تر ICC_5 (در سمت چپ) با نقطه $\rho = ۰$ فاصله دارد، طوریکه این کرانه منحنی، در نقطه $\rho = ۰/۲۵$ با محور x موازی می‌شود و به صورت یکنواخت در می آید. منحنی ویژگی این نوع سوال ها، نشان دهندهٔ سوال هایی است که آزمودنی‌هایی با توانایی پایین با استفاده از حدس زدن به آنها جواب صحیح می‌دهند. به طور معمول، اگر در یک سوال چهار گزینه‌ای آزمودنی‌هایی با توانایی کم به طور تصادفی پاسخ صحیح را حدس بزنند، تقریباً ۲۵ درصد آزمودنی‌ها جواب صحیح را حدس می‌زنند. به خاطر پاسخ های حدسی، روانسنج ها از مدل سه پارامتری برای تحلیل داده‌های آزمون استفاده می کنند. معادله ریاضی این مدل عبارت است از:

معادله (۱۰)

$$P\left(X=۱ \mid C_i,\alpha_i,\theta_j,\beta_i\right)=C_i+(۱-C_i)\frac{e^{\alpha_i(\theta_j-\beta_i)}}{۱+e^{\alpha_i(\theta_j-\beta_i)}}$$

۱- توجه داشته باشید که اطلاق عنوان مدل سه پارامتری لوجستیک به این مدل اشتباه است، زیرا این مدل جزو خانواده توابع لوجستیک نیست، بلکه یک مدل ریاضی خاص است که با مدل های لجیستیک متفاوت است (بیکر و کیم، ۲۰۰۴).

می توان اثبات کرد که در این معادله اثر حدس از روی احتمال پاسخ صحیح (p) حذف شده است. برای این کار ابتدا عبارت $a_i(q_j - b_i)$ را مساوی z_i قرار داده و سپس اثبات مورد نظر را انجام می دهیم.

$$P=C+\frac{(1-C)e^{z_i}}{1+e^{z_i}} \Rightarrow p=\frac{c+ce^{z_i}+e^{z_i}-ce^{z_i}}{1+e^{z_i}} \Rightarrow p=\frac{c+e^{z_i}}{1+e^{z_i}} \Rightarrow p+pe^{z_i}=c+e^{z_i}$$

$$p-c=e^{z_i}-pe^{z_i} \Rightarrow p-c=e^{z_i}(1-p) \Rightarrow \frac{p-c}{1-p}=e^{z_i} \Rightarrow \ln\frac{p-c}{1-p}=z_i \Rightarrow \ln\frac{p-c}{1-p}=\alpha_i(\theta_j-\beta_i)$$

در این رابطه C_i احتمال پاسخ صحیح به سوال با استفاده از حدس و $(1-C_i)$ احتمال حدس زدن غلط جواب سوال است. سایر علائم به کار رفته همان علائم مدل دو پارامتری است. مدل سه پارامتری علاوه بر مشکلات مربوط به مدل دو پارامتری، که ناشی از اعمال پارامتر تشخیص برای سوال ها بود، چند مشکل خاص خودش را نیز دارد. اول، بر طبق مدل سه پارامتری، آزمودنی ها می‌توانند با استفاده از یکی از دو راهبرد عامل حدس و توانایی فردی به یک سوال جواب صحیح بدهند. بنابراین مشکل است که ببینیم چگونه مدلی که دو راهبرد برای درست جواب دادن به یک سوال ارائه می کند، می‌تواند برای اندازه‌گیری توانایی آزمودنی‌ها در یک خصیصه مکنون تک بعدی مورد استفاده قرار گیرد. دوم، با اعمال حدس در مدل سه پارامتری این طور به نظر می‌رسد که عامل حدس را به عنوان یک ویژگی سوال وارد تحلیل می‌کنیم، اما توجه داشته باشید که به احتمال زیاد، حدس یک ویژگی مربوط به آزمودنی‌ها باشد نه سوال ها. البته غیر محتمل است که حدس زدن هم ویژگی سوال ها باشد و هم ویژگی آزمودنی ها. چرا که سوال ها کمی باعث برانگیختن حدس برای همه آزمودنی ها می‌شوند و آزمودنی های کمی در واقع برای همه سوال ها از حدس استفاده می کنند.

رویکردهای اندازه‌گیری

باید تاکنون متوجه تفاوت بین مدل راش و مدلهای دو و سه پارامتری شده باشید. در واقع نگرش و منطق زیربنایی این مدل ها به اندازه گیری با هم متفاوت است. همان طور که تیسن و اورلند(۲۰۰۱) مطرح کرده اند، امروزه مدل سازی در اندازه گیری با نظریه سوال پاسخ از دو رویکرد تبعیت می کند. رویکرد اول با هدف تحلیل سوال، مناسب ترین مدل را برای برآورد پارامترهای سوال و خصیصه مورد نظر به کار می گیرد. بر اساس این رویکرد مدلی که برای تحلیل استفاده می‌شود باید ویژگی های مربوط به داده‌های مشاهده شده را با دقت کافی منعکس کند، طوریکه عملکرد یک سوال در پارامترهای آن خلاصه شود. فلسفه این رویکرد آن است که فرض می کند سوال ها همان طور که اجرا شده‌اند، اندازه‌گیری می کنند، نه آن طور که باید عمل می‌کردند. به عبارتی این رویکرد هیچ پیش فرضی در مورد سوال ها ندارد و معتقد است که نظریه و مدل های اندازه‌گیری باید داده‌های به دست آمده را تبیین کنند. دومین رویکرد برای مدل سازی در اندازه گیری با نظریه سوال پاسخ این است که داده‌های سوال پاسخ باید با ویژگی های خاص اندازه‌گیری که به وسیله مدل تعریف شده‌اند، هماهنگ باشند. اگر سوال یا فردی با ویژگی‌های اندازه‌گیری مدل برازش نداشته باشد، کنار گذاشته می‌شوند. این دیدگاه که از رویکرد مدلهای راش پیروی می‌کند، در مواردی که داده‌ها با مدل برازش داشته باشند، یک تفسیر ساده نیز برای تحلیل سوال ها و نمره گذاری آزمون ارائه می کند. این رویکرد مدل سازی در نظریه سوال پاسخ معتقد است که اندازه‌گیری بهینه از نظر ریاضی تعریف شده است، بنابراین گروهی از مدل های نظریه سوال پاسخ با هدف انجام این نوع اندازه‌گیری مطرح می‌کند.

دو رویکرد توصیف شده در بالا باعث به وجود آمدن دوگانگی در بین روانسنج ها شده اند. گروهی که معتقدند یک الگوی خوب اندازه‌گیری باید رفتار مکنون در الگوهای پاسخ به دست آمده از یک بررسی را توصیف کند، با بررسی مدل های مختلف و دقت برازش آنها با داده ها، مناسب ترین مدل نظریه سوال پاسخ را برای برازش با داده‌ها انتخاب می‌کنند. به عبارتی این گروه معتقدند که انتخاب مدل های نظریه سوال پاسخ وابسته به الگوهای مشاهده شده پاسخ ها یا داده‌ها است. مدلهای یک پارامتری، دو پارامتری، سه

پارامتری و سایر مدل های موجود در نظریه سوال پاسخ که جزء خانواده مدل های راش نیستند (جدول شماره ۳) بر اساس این دیدگاه شکل گرفته اند (تیسن و استینبرگ، ۱۹۸۶).

جدول شماره ۳: انواع مدل های متداول در نظریه سوال پاسخ

مدل*	نوع پاسخ	ویژگی های مدل
مدل اولیه راش* و مدل تک پارامتری لوجستیک	دو ارزشی	قدرت تشخیص همه سوال ها یکسان ولی دشواری سوال ها تغییر می کند
مدل دو پارامتری لوجستیک	دو ارزشی	پارامتر تشخیص و دشواری در سوال های مختلف متفاوت است
مدل سه پارامتری	دو ارزشی	پارامتر حدس را نیز شامل می‌شود
مدل درجه بندی شده[1]	چند ارزشی	پاسخهای مرتب شده[2] و پارامتر تشخیص در سوال های مختلف تغییر می کند
مدل اسمی[3]	چند ارزشی	هیچ ترتیب از پیش تعیین شده‌ای وجود ندارد و تشخبص سوال های مختلف تغییر می کند
مدل اعتبار تفکیکی*[4]	چند ارزشی	پارامتر تشخیص بین سوال های مختلف مساوی است
مدل مقیاس درجه بندی*[5] بندی*[5]	چند ارزشی	پارامتر تشخیص و دشواری در سوال های مختلف یکسان است
* نشان دهنده مدلهای متعلق به خانواده راش		

گروه دیگر با پیروی از رویکرد راش، معتقدند که تنها مدل های مناسب برای اندازه گیری های روانی و رفتاری، مدل های خانواده راش هستند. چونکه ویژگی های قوی ریاضی نظیر عینیت مشخص[6] (یعنی برآورد جداگانه و مستقل پارامترهای افراد و سوال ها) و کفایت ساده نمره کل[7] (یعنی عدم نیاز به اطلاع داشتن از الگوی پاسخ) را دارا هستند. مدل راش دارای چندین مزیت است: زمانی که تعداد پاسخ دهندها کم است یا نمونه‌های

1. graded model
2. ordered response
3. nominal model
4. credit partial model
5. rating scale model
6. Specific objectivity
7. Summed score simple sufficiency

فوق‌العاده نامعرف مورد استفاده قرار می گیرند و یا توزیع جامعه در خصیصه زیربنایی مورد نظر به شدت دارای کجی باشد، در هر سه مورد این مدل از توانائی کافی برای برآوردهای پایدارتری از پارامترهای افراد و ویژگی های سوال ها برخوردار است.

به نظر امبرستون و رایز (۲۰۰۰) زمانی که در تعریف متغیر زیربنایی، همه سوال ها وزن یکسانی دارند، یعنی همه سوالات اهمیت یکسانی دارند- و زمانی که ویژگی های قوی مدل اندازه‌گیری - یعنی عینیت مشخص، کفایت ساده نمره کل- مورد نظر است، باید از مدل های خانواده راش استفاده کرد. اما اگر برازش یک مدل نظریه سوال پاسخ با داده‌های موجود مورد نظر باشد، یا برآوردهای دقیق‌تری از پارامترها برای شما مطلوب است، باید از مدل های پیچیده تر نظیر مدل دو پارامتری یا مدل مدرج شده استفاده کرد.

تفاوت مدل ساده راش و مدل یک پارامتری

مدل اولیه راش و مدل یک پارامتری، هر دو ویژگی های مشابهی دارند و از نظر ریاضی نیز یکسانند. ولی این مدل ها تفاوت هایی نیز با هم دارند که کمتر مورد توجه قرار می گیرند. اولین تفاوت این دو مدل در زیربنای منطقی آنها است. راش از بین دو رویکرد مطرح شده در نظریه سوال پاسخ طرفدار رویکرد دوم بود و مدلش نیز بر اساس همین رویکرد شکل گرفته است. در حالی که مدل تک پارامتری از رویکرد اول نشأت گرفته است که هدف آن تحلیل الگوهای پاسخ به سوال ها می‌باشد. در رویکرد دوم، اگر سوال یا فردی با مدل برازش نداشته باشد، حذف می‌شود، روش راش و طرفداران وی نیز همین بوده است. در حالی که، رویکرد اول به جای کنار گذاشتن سوال یا آزمودنی، مدل را عوض می کند تا با یافتن مناسب ترین مدل دقیق‌ترین پارامترها را برآورد کند. دومین تفاوت این دو مدل آن است که راش مدل خود را به این قاعده محدود کرده که مجموع پارامترهای دشواری برای همه سوالات مقیاس برابر صفر است ($\sum b=0$) و بر اساس همین محدودیت، مقیاس پارامتر θ را مشخص می کند. با توجه به این محدودیت توزیع θ در جامعه نامشخص است، مقدار میانگین این توزیع متناسب با متوسط دشواری سوال های و مقدار واریانس آن متناسب با شیب خطوط ممتد، یعنی یک است. مدل راش فرض می کند که همه سوالات بر

روی عامل مورد نظر وزن یکسانی دارند، بنابراین پارامتر تشخیص را برای همه سوال ها برابر با یک در نظر می گیرد و معتقد است که عامل حدس نیز پاسخ افراد را تحت تاثیر قرار نمی‌دهد. چون شکل توزیع θ در جامعه این مدل نامشخص است، پس این توزیع می‌تواند هر شکلی داشته باشد و توزیع نمرات مشاهده شده، شکل این توزیع را مشخص می کند. در مقابل این ویژگی ها، مدل یک پارامتری فقط مستلزم آن است که شیب برای همه سوال ها یکسان باشد، بنابراین، شیب در این مدل مقادیری متناسب با واحد انحراف استاندارد متغیر مکنون می‌گیرد، توزیع جامعه برای متغیر زیر بنایی در مدل یک پارامتری (همانند مدلهای دو پارامتری و سه پارامتری) معمولا دارای میانگین صفر و واریانس یک در نظر گرفته می‌شود. پارامتر دشواری این مدل نیز متناسب با مقدار متوسط میزان خصیصه مورد اندازه گیری- یعنی صفر- در نظر گرفته می شود (تیسن و اورلند، ۲۰۰۱). بنابراین، این مدل فرض می کند که متغیر مکنونθ دارای توزیع نرمال است، نه پاسخ های مقوله‌ای سوال[1] (تیسن و استینبرگ، ۱۹۸۸). با وجود تفاوت های ذکر شده در بالا، نتایج این دو مدل از نظر ریاضی یکسان است. به طوری که در اکثر مواقع مدل یک پارامتری و مدل راش را به جای هم به کار می‌برند. اگر تفاوت های ذکر شده را کنار بگذاریم همیشه برازش این دو مدل با داده های جمع آوری شده به دشواری حاصل می‌شود. این امر، به این خاطر است که مفروضات قوی تری را در مقایسه با مدل های دو و سه پارامتری در نظر می‌گیرد. در این دو مدل معتقدند که تنها عوامل تاثیر گذار در پاسخ گویی افراد به سوال ها (متغیر وابسته)، توانایی افراد و دشواری سوال ها (دو متغیر مستقل) می‌باشند. از نظر این دو مدل، عامل حدس نقشی در پاسخ دادن افراد ندارد و قدرت تشخیص همه سوالات نیز با هم یکسان است. به همین خاطر است که در عمل همیشه مدل های دوپارامتری و سه پارامتری بیشتر از مدل یک پارامتری راش با داده ها برازش پیدا می کنند(آلن و ین، ۱۹۷۹؛ رایز، اینثوورث، هاویلند، ۲۰۰۵).

1. Categorical item response

بحث و نتیجه گیری

موضوع اصلی مقاله حاضر این بود که نظریه‌های خصیصه مکنون و سوال پاسخ روش های متفاوتی هستند. نظریه خصیصه مکنون با اندازه‌گیری ویژگی های مکنون یا زیربنایی افراد سر و کار دارد، اما نظریه سوال پاسخ برای توصیف و خلاصه کردن عملکرد مشاهده شده به صورت آماری به کار می‌رود. نگرش اکثر روانسنج ها به تحلیل خصیصه مکنون و نظریه سوال پاسخ یک رویکرد ریاضی است، که این جهت یابی می‌تواند تفاوت بین این دو نظریه را بپوشاند. در مدل سه پارامتری (معادله ۴) اگر $C = 0$ باشد در این صورت مدل سه‌پارامتری از نظر جبری به مدل دو پارامتری تقلیل می‌یابد و اگر در مدل دو پارامتری (معادله ۳) $\alpha = 1$ باشد در این صورت مدل دو پارامتری به تک پارامتری تقلیل می‌یابد. بررسی هر سه مدل در غالب شکل ریاضی‌شان ما را به این نتیجه می‌رساند که مدل تک پارامتری مورد خاصی از مدل دو پارامتری که دارای $\alpha = 1$ است و مدل دو پارامتری حالت خاصی از مدل سه پارامتری است که در آن $C = 0$ است. شباهت مبنای ریاضی هر سه مدل، یک مزیت اساسی مدل تک پارامتری راش را از نظر خوانندگان مخفی نگه می‌دارد. به عبارتی، فرم خاص مدل تک پارامتری راش این توانایی فوق العاده را دارا است که می‌توان مفروضه بنیادی تک بعدی بودن را بررسی نمود. اما زمانی که دومین وسومین پارامتر سوال به وسیله مدل‌های دو و سه پارامتری وارد صحنه می‌شود، توان بررسی این مفروضه از دست می‌رود.

از منظر ریاضی کاربردی، تفاوت موجود بین نظریه‌های خصیصه مکنون و سوال پاسخ به آسانی قابل مشاهده نیست. در مطالعات زیادی نتایج داده‌های تحلیل شده با مدل تک پارامتری را با نتایج مدل های دو و سه پارامتری مقایسه کرده‌اند، که نتایج اغلب این مطالعات ضد و نقیض و مبهم بوده است. اگر ناهماهنگی[1] بین پاسخ های مشاهده شده و پاسخ های پیش بینی شده با استفاده از مدل تک پارامتری، تحت عنوان خطا در مفهوم آماری در نظر گرفته شوند تبیین ابهام موجود در نتایج مطالعات مقایسه‌ای آسان خواهد

1. Inconsistence

شد. اگر مفروضه معروف تصادفی بودن توزیع چنین خطاهایی را بپذیریم و اگر نمونه بزرگی از آزمودنی‌ها وجود داشته باشند و آزمون ها نیز به اندازه کافی طولانی باشند، می‌توان انتظار داشت که این گونه خطاهای تصادفی کاملا همدیگر را خنثی کنند. وقتی که خطاها همدیگر را خنثی کنند نتیجه این خواهد بود که توانایی مکنون آزمودنی‌ها و دشواری سوال ها، داده‌های مشاهده‌شده را تبیین کنند. بنابراین، زمانی که مفروضه خطاهای تصادفی معقول باشد، مقایسه تجربی مدل تک پارامتری با سایر مدلها نتایج مشابهی بدست می‌دهد. اما اگر مفروضه خطای تصادفی برقرار نباشد، نتایج چنین مقایسه‌هایی می‌تواند متناقض باشد. نتیجه‌گیری نهایی این است که نظریه‌های خصیصه مکنون و سوال پاسخ بر اساس مفروضه‌های فلسفی متفاوتی قرار دارند و اهداف متفاوتی را دنبال می‌کنند. زمانی که این دو روش برای تحلیل داده‌های آزمون به کار روند، سوال های متفاوتی را مدنظر قرار می‌دهند. بنابراین، توجه به این نکته مهم است که هر سوال جواب خاص خودش را دارد. بر عهده پژوهشگران است که تعیین کنند قصد پاسخ گویی به چه نوع سوال هایی را دارند تا بتوانند مسائل کاربردی مربوط به اندازه گیری شان را حل کنند.

منابع

آلن،مری جی؛ ین، وندی ام.(۱۹۷۹). *مقدمه ای برنظریه های اندازه گیری*، ترجمه علی دلاور(۱۳۷٤). تهران: سمت.

سپاسی، حسین (۱۳۷۵). نظریه سوال‌پاسخ در آزمونهای روانی تربیتی، *مجله علوم تربیتی و روانشناسی دانشگاه شهید چمران اهواز*، شماره ۳و٤.

سپاسی، حسین(۱۳۷٤). نظریه کلاسیک آزمون و محدودیتهای آن، *مجله علوم تربیتی و روانشناسی دانشگاه شهید چمران اهواز*: دوره سوم، شماره ۱و۲،ص ١٤٦-۱۳۲.

کیامنش، علیرضا(۱۳۷۸). نظریه تعمیم پذیری در اندازه گیری آموزشی. *مجله علوم تربیتی دانشگاه تهران*: دوره ١٦، شماره۱و۲، ص ۲۵-٤٨.

Baker F. & Seock-ho. kim (٢٠٠٤). Item response theory parameter estimation techniques, second edition, revised and expanded .new York. Basel. Marcel Dekker,inc

Borsboom, Denn; Mellenbergh J. Gideon and Van Heerden , Jaap (٢٠٠٣). The Theoretical Status of Latent Variables. Psychological Review. , Vol. ١١٠, No. ٢, ٢٠٣–٢١٩

Courville, Troy G.(٢٠٠٤). An Emprical comporison of Item Response theory and classical Test Theory Item/Person statistics. Dissertation of Doctoral Philosophy, Texas A&M University.

Embereston, Susan E.; Reise, Steven P. (٢٠٠٠). Item Response Theory for psychologists. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey. London.

Lord, F.M. & Norick, M.R. (١٩٦٨). Statistical theories of Mental test Scores. Reading MA: Addison – Wesley.

Lord, F.M. (١٩٨٠). Applications of item response theory to practical testing problems. New Jersey: Lawrence Erlbaum Associates, Inc

Rasch G. (١٩٦٠). Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press (Reprinted ١٩٨٠).

Reise, Steven P.; Ainsworth, Andrew T.; Haviland, Mark G.(٢٠٠٥). Item Response Theory: Fundamentals, Applications, and promise in Psychological Research. American Psychological Society, Vol.١٤, NO, ٢.

Ryan, J. Garcia – Quintana, R. and Hamm, D.W. (١٩٨٠). Testing the fit of subjects to a latent Trait Model. Paper Presented at the Annual Meeting of the National Council on Measurement in Education. Boston.

Ryan, Joseph P. (١٩٨٣). Introduction to latent Trait Analysis and Item Respons Theory. In Hathaway, W.E. (Ed). Testing in the schools: New Directions for Testing and Measurment. No.١٩ (pp.٤٩-٦٥). San Francisco, CA: Jossey-Bass.

Thissen, d. & Orlando. m. (۲۰۰۱). Chapter ۳-item response theory for items scored in two categories. In d. tissen & h. wainer (eds.), test scoring. Hillsdale, NJ: Erlbaum.
Thissen, D. & Steinberg, L. (۱۹۸۸). Data analysis using item response theory. Psychological Bulletin, ۱۰٤, ۳۸٥-۳۹٥.
Thissen, D., & Steinberg, /L. (۱۹۸٦). A Toxonomy of Item Response Models. Psychomtrika, No. ٥۱, Vol. ٤, pp. ٥٦۷-٥۷۷.