

کاربرد مدل چندگزینه‌ای در تحلیل سؤال و مقایسه برآزش و آگاهی آن با مدل سه پارامتری (مورد: بخش زیست‌شناسی آزمون سراسری ۱۳۹۱)

بلال ایزانلو*
مسعود گرامی‌پور**
علی ترکاشوند***

چکیده

هدف اصلی پژوهش حاضر کاربرد مدل چندگزینه‌ای در تحلیل سؤال‌های چندگزینه‌ای و مقایسه برآزش و آگاهی آن با نتایج مدل سه‌پارامتری است. برای این منظور، از بین همه شرکت‌کنندگان آزمون سراسری ۹۲-۱۳۹۱ در گروه آزمایشی تجربی درس زیست‌شناسی (۵۰ سؤال)، نمونه‌ای ۵۰۰۰ نفری به صورت تصادفی انتخاب شد. پس از بررسی پیش‌فرض تک‌بعدی بودن با استفاده از نرم‌افزار NOHARM، سؤال‌های آزمون با برنامه MULTILOG تحلیل شد. مدل سه‌پارامتری که از نظر ساختاری در میان مدل‌های دو ارزشی، نزدیک‌ترین مدل به مدل چندگزینه‌ای است، به‌عنوان مدل دو ارزشی مناسب، برای مقایسه آگاهی برآزش آن با مدل چندگزینه‌ای انتخاب شد. نتایج این بررسی‌ها نشان‌دهنده کارایی این مدل در تحلیل گزینه‌های انحرافی سؤال‌های چندگزینه‌ای است. به‌علاوه مدل چندگزینه‌ای، توانایی آزمودنی‌های پایین‌تر از متوسط و مدل سه‌پارامتری، توانایی آزمودنی‌های بالاتر از متوسط را با دقت بیشتری برآورد می‌کنند. در مقایسه برآزش مدل سه‌پارامتری با مدل چندگزینه‌ای، شواهد از برتری مدل سه‌پارامتری حکایت دارد. بر اساس نتایج می‌توان از مدل چندگزینه‌ای در کنار سایر روش‌ها برای افزایش آگاهی به عملکرد گزینه‌های سؤال استفاده کرد.

واژگان کلیدی: روان‌سنجی، نظریه پاسخ سؤال، مدل چندگزینه‌ای، مدل سه‌پارامتری

تاریخ دریافت مقاله: ۱۳۹۰/۰۱/۲۹

تاریخ پذیرش مقاله: ۱۳۹۰/۰۶/۰۸

* استادیار دانشکده روان‌شناسی و علوم تربیتی دانشگاه خوارزمی (نویسنده مسئول، izan.b@khu.ac.ir)

** استادیار دانشکده روان‌شناسی و علوم تربیتی دانشگاه خوارزمی

*** کارشناس ارشد تحقیقات آموزشی دانشگاه خوارزمی

مقدمه

امروزه در حوزه اندازه‌گیری دو نظریه غالب وجود دارد؛ نظریه کلاسیک اندازه‌گیری، که می‌توان پیدایش آن را به اسپیرمن^۱ نسبت داد و نظریه پاسخ سؤال که منشأ آن به نظریه صفت پنهان^۲ بازمی‌گردد. مدل‌های نظریه پاسخ سؤال که گاهی در تقابل با نظریه کلاسیک اندازه‌گیری، نظریه جدید خوانده می‌شود، اطلاعات بیشتری را در مورد سؤال، فرد و آزمون فراهم می‌آورد (همبلتون^۳، ۱۹۸۹).

در تجزیه و تحلیل سؤال‌های آزمون‌های چندگزینه‌ای با استفاده از مدل‌های دوارزشی نظریه پاسخ سؤال، همانند نظریه کلاسیک، این محدودیت وجود دارد که در آنها نیز برای گزینه‌های انحرافی، هویت مستقلی قائل نمی‌شوند. در این مدل‌ها معمولاً برای پاسخ درست، ارزش یک و برای پاسخ به گزینه‌های نادرست، ارزش صفر در نظر گرفته می‌شود و پارامترهای سؤال، صرفاً برای گزینه درست برآورد می‌شوند. در صورتی که گزینه‌های انحرافی سؤال‌ها، حاوی اطلاعاتی در رابطه با سطح توانایی پاسخ‌دهندگان هستند و احتمال انتخاب هر یک از گزینه‌های نادرست از سوی آزمودنی‌ها در سطوح توانایی مختلف، متفاوت خواهد بود و نمی‌توان همه پاسخ‌های نادرست را هم‌ارزش دانست. در نظر نگرفتن اطلاعات بالقوه موجود در گزینه‌های انحرافی انتخاب شده توسط افراد، موجب از دست رفتن اطلاعات مفیدی خواهد شد که می‌تواند نقش مؤثری در افزایش دقت برآورد سطح توانایی آزمودنی‌ها ایفا کند. برخلاف مدل کلاسیک و مدل‌های دو ارزشی، به لحاظ نظری مدل‌های چندارزشی - نظریه پاسخ سؤال، از این قابلیت برخوردارند که می‌توانند در برآورد پارامترهای سؤال و آزمودنی‌ها در آزمون‌های حاوی سؤال‌های چندگزینه‌ای با تمرکز بر اطلاعات موجود در گزینه‌های نادرست موجب افزایش دقت برآورد پارامترهای سؤال و توانایی در مقایسه با مدل‌های دو ارزشی شوند. لذا استفاده از مدل چندارزشی مناسب در تجزیه و تحلیل سؤال‌های آزمون چندگزینه‌ای می‌تواند آگاهی‌دهنده‌تر از مدل‌های دوارزشی باشد. چون در صورت استفاده از مدل دوارزشی، همه پاسخ‌های نادرست در یک طبقه در نظر گرفته شده و اطلاعات موجود در آنها ارزیابی نمی‌شود اما مدل

1. Spearman

2. latent trait

3. Hambleton

چندارزشی از این مزیت برخوردار است که احتمال انتخاب هر گزینه را به‌عنوان تابعی از سطح توانایی برآورد می‌کند و پاسخ‌های نادرست را هم‌ارزش نمی‌داند. از آنجا که انتخاب هر گزینه تابعی از سطح توانایی است، به‌طور بالقوه می‌تواند اطلاعاتی نیز درباره سطح توانایی پاسخ‌دهنده دربرداشته باشد (لوین و دراسگو^۱، ۱۹۸۳؛ نینگ و همکاران^۲، ۲۰۱۵).

سؤال‌های مشتمل بر طبقات پاسخ چندگانه، علاوه بر حوزه نگرش، شخصیت و هوش، در حوزه آزمون‌های توانایی نیز به‌طور گسترده، به‌ویژه در قالب سؤال‌های چندگزینه‌ای، قابل استفاده‌اند. شاید ساده‌ترین و آشکارترین دلیل شکل‌گیری و توسعه مدل‌های چندارزشی نظریه پاسخ سؤال، وجود آزمون‌های حاوی سؤال‌های چندارزشی است که معمولاً در آزمون‌های شناختی و آموزشی، استفاده می‌شود. استفاده از چنین آزمون‌هایی در حوزه‌های مختلف سنجش به‌ویژه در زمینه سنجش متغیرهای شخصیتی و اجتماعی، امری غیرقابل اجتناب است. زیرا در چنین حوزه‌هایی، سنجش تفاوت‌های جزئی مورد توجه است که سؤال‌های مشتمل بر پاسخ‌های دو ارزشی، عموماً قادر به سنجیدن آنها نیستند. از این رو، در چنین مواردی استفاده از سؤال‌های چندارزشی که به دلیل برخورداری از بیش از دو طبقه پاسخ، می‌توانند طیف وسیع‌تری از پیوستار صفت را پوشش دهند، ضرورت می‌یابد. علاوه بر این به دلیل برخی نکات روان‌سنجی، جذابیت سؤال‌های چندارزشی در مقایسه با سؤال‌های دو ارزشی بیشتر است. یکی از نکات این است که سؤال‌های چندارزشی در مقایسه با سؤال‌های دو ارزشی، طیف وسیع‌تری از پیوستار توانایی را پوشش می‌دهند که این امر به‌واسطه قابلیت در برگرفتن طبقات پاسخ بیشتر در سؤال‌های چندارزشی در مقایسه با سؤال‌های دوارزشی است. به بیان دقیق‌تر می‌توان گفت، اگرچه همه انواع سؤال (دو و چندارزشی) می‌توانند سرتاسر پیوستار توانایی را دربرگیرند (از منفی بی-نهایت تا مثبت بی‌نهایت) اما باید توجه داشت که میزان آگاهی به دست آمده که با یک سؤال فراهم می‌شود در جایگاه مکانی سؤال (b_j) به نقطه اوج خود می‌رسد و هر چه از این نقطه به سوی دو سر پیوستار صفت پیش رویم، غالباً با شیب تندی، میزان آگاهی‌دهندگی کاهش می‌یابد. شگفت این‌که هرقدر یک سؤال در نقطه اوج

¹. Levine & Drasgow

². Ning et al

خود، آگاهی بیشتری را فراهم کند، برای طیف محدودتری از پیوستار توانایی، آگاهی - دهندگی مفیدی خواهد داشت. مزیت مدل‌های چندارزشی این است که با برخورداری از امتیاز تعداد طبقات بیشتر، از این قابلیت برخوردارند که آگاهی بیشتری را در طیف وسیع‌تری از پیوستار صفت فراهم کنند (استینی و نرینگ^۱، ۲۰۰۶). مسترز^۲ (۱۹۸۸) خاطرنشان ساخته که هدف کلی به‌کارگیری بیش از دو طبقه برای هر سؤال، تلاش در جهت دستیابی به آگاهی بیشتر درباره‌ی سطح صفت افرادی است که مورد اندازه‌گیری قرار می‌گیرند، به‌طوری که برآوردهای دقیق‌تری برای سطح صفت حاصل می‌شود. علاوه بر این وی نشان داد که اطلاعات تشخیصی جامع‌تری درباره‌ی پاسخ‌دهندگان و سؤال‌ها می‌تواند از سؤال‌های هر آزمون چندگزینه‌ای حاصل شود.

مدل پاسخ اسمی^۳ باک (۱۹۷۲) زیربنای مدل باک-سامی جیما^۴ و مدل چندگزینه‌ای^۵ (MCM) است. نتایج حاصل از برآزش مدل پاسخ اسمی^۶ (NRM) و مقایسه آنها با نتایج مدل دوپارامتری از نظر دقت اندازه‌گیری بر اساس پاسخ‌های ۵۷۰ دانش‌آموز به سؤال‌های آزمون ماتریس‌های پیش‌رونده ریون نشان داد که برای توانایی‌های کمتر از صفر در پیوستار توانایی میزان آگاهی‌دهندگی مدل پاسخ اسمی، ۱/۵ تا ۲ برابر مدل دو پارامتری است. اما در کل دامنه توانایی افزایش آگاهی ناچیزی مشاهده شد (تیسن، ۱۹۷۶). در کنار این موضوع در آزمون‌های چندگزینه‌ای پیشرفت تحصیلی نیز دقت برآورد توانایی مدل پاسخ اسمی در اکثر سطوح توانایی، به‌ویژه در مورد افراد با توانایی پایین، بسیار بیشتر از دقت برآورد مدل دو پارامتری است (معلمی اوره، ۱۳۸۷). به‌علاوه با به‌کارگیری مدل پاسخ اسمی (NRM)، به سؤال‌های کمتری نیاز است تا همان دقت برآورد مدل سه‌پارامتری حاصل شود (دی‌آیالا^۷، ۱۹۹۳).

از کاستی‌های مدل پاسخ اسمی آن است که بر اساس این مدل امکان تفکیک افرادی که گزینه‌ای را آگاهانه و آنهایی که گزینه‌ای را به‌صورت تصادفی انتخاب می‌کنند، وجود ندارد. به همین منظور، سامی جیما (۱۹۷۹) مدل اسمی باک (۱۹۹۷) را

^۱. Ostini & Nering

^۲. Masters

^۳. Nominal response model

^۴. Bock-samejima (BS) model

^۵. Multiple choice model

^۶. Nominal Response Model

^۷. De Ayala

طوری بسط داد که پارامتر حدس مجانب را نیز در برگیرد. این مدل که به مدل باک-سامی جیما نیز معروف است، میزان پارامتر حدس مجانب برای تمام گزینه‌های یک سؤال برابر با $1/m$ فرض می‌کند ($m \geq 2$ تعداد گزینه‌های سؤال است). دلیل اهمیت این موضوع، به‌خصوص توجه به نقش گزینه‌های انحرافی در سؤال‌های چندگزینه‌ای بوده است. زیرا نتایج پژوهش‌ها حاکی از آن است که بین انتخاب گزینه‌های نادرست در بسیاری از سؤال‌ها با سطح توانایی برآورد شده بر اساس دو آزمون سنجش استعداد، رابطه وجود دارد (دراسگو و همکاران^۱، ۱۹۹۵). با این حال بحث انتخاب گزینه‌ها در سایر آزمون‌ها نیز اهمیت دارد و فقط به آزمون‌های چندگزینه‌ای پیشرفت تحصیلی محدود نمی‌شود. اسمای‌لی‌گُف^۲ (۱۹۸۹) به مقایسه مدل باک-سامی جیما با مدل سه پارامتری در داده‌های یک آزمون چندگزینه‌ای تاریخ پرداخت. وی این آزمون را طوری ساخت که در گروهی از سؤال‌های آن، گزینه‌های انحرافی حاوی رابطه با صفت پنهان باشند و در گروهی دیگر از سؤال‌ها، گزینه‌های انحرافی آگاهی‌دهنده نباشند. بنابراین او کارکرد دو مدل را برای دو نوع سؤال بررسی کرد. نتایج نشان داد برازش مدل باک-سامی جیما برای هر دو نوع سؤال بهتر از مدل سه پارامتری بود. از نظر میزان آگاهی‌دهندگی نیز، وی نشان داد که دقت مدل باک-سامی جیما، به‌ویژه در مورد افراد ضعیف‌تر بیش از مدل سه پارامتری است.

نقص اصلی مدل باک-سامی جیما در آن است که برای تمام گزینه‌های یک سؤال حدس یکسانی در نظر می‌گیرد. تیسن و استینبرگ (۱۹۸۴) برای کنار آمدن با این مشکل، مدل باک-سامی جیما را (که با اعمال اصلاحاتی بر روی مدل اسمی باک (۱۹۷۲) توسط سامی جیما (۱۹۷۹) ارائه شده است) طوری بسط دادند که علاوه بر سایر پارامترها، پارامتر حدس تصادفی را نیز در برگیرد. این پژوهشگران با ارائه توصیفی از رابطه بین مدل و داده‌ها، برای برآورد پارامترهای مدل با روش بیشینه درست‌نمایی کناری طرحی ارائه دادند و نتایج این مدل و سایر مدل‌ها را که نسخه محدود شده این مدل هستند، با کمک آماره نسبت درست‌نمایی^۳ مقایسه کردند. توصیف و کاربرد مدل چندگزینه‌ای برای گزینه‌های انحرافی سؤال‌های چندگزینه‌ای

¹. Drasgow et al

². Smilaycoff

³. Likelihood Ratio Statistics

توسط تیسن، استینبرگ و فیتزپاتریک (۱۹۸۹) انجام شده است. آنها پاسخ‌های ۹۹۰ دانش‌آموز به ۴ سؤال ریاضی مربوط به مصرف کالری را که بخشی از یک آزمون ۵۰ سؤالی پیشرفت ریاضی بود با این مدل تحلیل کرده و به‌کارگیری این مدل برای تشخیص سؤال‌ها و گزینه‌های معیوب در طراحی و ساخت آزمون‌ها را مورد بحث قرار داده‌اند.

با وجود مزایای مختلف مدل‌های چندارزشی، به‌خصوص مدل چندگزینه‌ای در تحلیل سؤال، اینکه چرا از این مدل‌ها استفاده نمی‌شود، جای سؤال است. ای‌بد، اوی لی و پانسودا^۱ (۲۰۰۹) در مورد برخی از مشکلاتی که مانع به‌کارگیری مدل چندگزینه‌ای سامی‌جیما و مدل چندگزینه‌ای تیسن و استینبرگ می‌شود به بحث پرداخته و به چند مشکل در این خصوص اشاره کرده‌اند که عبارتند از: (۱) مشکلات مربوط به برآورد پارامترها به علت زیاد بودن پارامترها؛ (۲) مشکلات مربوط به شناسایی‌پذیری^۲ پارامترها در مدل تیسن و استینبرگ؛ (۳) برخورد این مدل‌ها با پاسخ‌های از دست رفته^۳. نتایج پژوهش‌ها در خصوص برازش مدل‌های چندارزشی تا اندازه‌ای ضد و نقیض است که در ادامه به برخی از مهم‌ترین آنها اشاره شده است. مطالعه دراسگو و همکاران^۴ (۱۹۹۵) برازش چهار مدل چندارزشی باک^۵ (۱۹۷۲)، چندگزینه‌ای سامی‌جیما^۶ (۱۹۷۹)، چندگزینه‌ای تیسن و استینبرگ (۱۹۸۴) و مدل ناپارامتریک لوین^۷ (۱۹۹۳) را روی سؤال‌های سه آزمون چندگزینه‌ای مهم (ASVAB، ACT، SAT) برای نمونه‌ای ۳۰۰۰ نفری بررسی کرد. نتایج نشان داد مدل پاسخ اسمی باک، برازش بسیار خوبی با داده‌های مورد نظر دارد و افزایش تعداد پارامترها در مدل سامی‌جیما و مدل چندگزینه‌ای تیسن و استینبرگ نسبت به مدل پاسخ اسمی باک، بهبودی در میزان برازش مدل‌ها با داده‌ها در پی ندارد. مطابق با نتایج پژوهش، مدل ناپارامتریک لوین نیز برای تمامی آزمون‌ها برازش خوبی نشان داد.

^۱. Abad, Olea and Ponsoda

^۲. identifiability

^۳. omitted responses

^۴. Drasgow et al

^۵. Bock

^۶. Samejima

^۷. Levine

در مقابل ای‌بد، اوی‌لی و پانسودا (۲۰۰۹) مدل چندگزینه‌ای سامی‌جیما، مدل چندگزینه‌ای تیسسن و استینبرگ، مدل پاسخ اسمی و یک مدل چندگزینه‌ای جدید (به نام مدل چندگزینه‌ای محدود شده سامی‌جیما^۱ که در مورد مکانیسم‌های فرض شده در خصوص داده‌های از دست رفته، شفاف‌تر است و از این مزیت برای تخمین پارامترها استفاده می‌کند) با نمونه‌ای ۳۲۲۴ نفری در یک آزمون نوشتاری زبان انگلیسی مورد برازش قرار دادند. نتایج شاخص برازش^۲ نشان داد که مدل پیشنهادی جدید مثل دو مدل چندگزینه‌ای قبلی بهتر از مدل پاسخ اسمی، با داده‌ها برازش دارند. از سوی دیگر به منظور تأیید این موضوع که ویژگی‌های برآورد مدل جدید واقعاً بهتر است، یک مطالعه شبیه‌سازی در خصوص بازیابی پارامتر نیز انجام شد. نتایج نشان داد که در مدل جدید نسبت به دو مدل چندگزینه‌ای قبلی، بازیابی پارامترها بهتر است.

در رابطه با انتخاب مدل چندارزشی مناسب برای تحلیل باید گفت اگرچه برای سؤال‌های چندگزینه‌ای می‌توان از طیف گسترده‌ای از مدل‌های چندارزشی استفاده کرد اما از آنجا که آزمون‌های پیشرفت تحصیلی، نظیر آزمون سراسری ورود به دانشگاه معمولاً طوری طراحی نمی‌شوند که بتوان گزینه‌های پاسخ را برحسب درجه حصول به پاسخ صحیح و یا گذراندن مراحل از حل مسئله مرتب کرد (مانند آنچه در مدل پاسخ مدرج^۲ (GRM) و یا مدل امتیاز تفکیکی^۳ (PCM) مورد نظر است) بنابراین باید از مدلی استفاده شود که برای هر نوع پاسخ اسمی مناسب باشد. مدل چندگزینه‌ای که تیسسن و استینبرگ^۴ (۱۹۸۴) مطرح کردند از انواع مدل‌های چندارزشی نظریه پاسخ سؤال است که شرط فوق را دارد و در آن از تمامی اطلاعات موجود در پاسخ‌های آزمودنی‌ها و نه فقط پاسخ‌های صحیح برای بیشینه کردن دقت برآورد توانایی استفاده می‌شود. فرض کنید سؤال z دارای m طبقه پاسخ است. در مدل چندگزینه‌ای احتمال آنکه فردی با توانایی θ به طبقه k ام ($k = 1, 2, \dots, m$) سؤال z پاسخ دهد برابر است با:

¹. Restricted Samejima Multiple-Choice Model

². Graded Response Model

³. Partial Credit Model

⁴. Thissen and Steinberg

$$p_j(x = k | \theta; \underline{\alpha}, \underline{c}, \underline{\phi}) = \frac{\exp(c_{jk} + \alpha_{jk}\theta) + d_{jk} \exp(c_{j0} + \alpha_{j0}\theta)}{\sum_{k=1}^{m_j} \exp(c_{jk} + \alpha_{jk}\theta)} \quad (1)$$

که در آن θ پارامتر توانایی، a_{jk} پارامتر شیب، c_{jk} پارامتر عرض از مبدأ و d_{jk} پارامتر طبقه پاسخ پنهان "نمی‌دانم" است و نسبت پاسخ‌دهنده‌هایی را نشان می‌دهد که "نمی‌دانند" ولی طبقه k سؤال z را انتخاب می‌کنند. توجه شود که بردارهای $\underline{\alpha}$ ، \underline{c} و \underline{d} دارای طولی برابر با تعداد گزینه‌ها، m هستند. به علاوه طبقه دارای کد صفر همان طبقه "نمی‌دانم" است (جزئیات بیشتر در تیسن و استینبرگ (۱۹۸۴) و تیسن، استینبرگ و فیتزپاتریک (۱۹۸۹) ارائه شده است). در این مدل، هم گزینه درست و هم گزینه‌های انحرافی، تجزیه و تحلیل می‌شوند. مدل چندگزینه‌ای تیسن و استینبرگ، خلاصه‌ای تصویری و پارامتری از عملکرد هر گزینه در سؤال‌های چندگزینه‌ای را فراهم می‌کند که این خلاصه‌ها رابطه بین احتمال انتخاب هر گزینه با توانایی اندازه‌گیری شده را توصیف می‌کنند. به طور کلی، کاربرد گسترده تحلیل سؤال بر اساس مدل چندگزینه‌ای به فهم بیشتر عملکردها و رفتار گزینه‌های انحرافی منجر می‌شود. لذا دانش اینکه مدل چندگزینه‌ای چگونه می‌تواند در تحلیل سؤال، استفاده شود برای طراحان سؤال و تحلیل سؤال‌های هرگونه آزمون چندگزینه‌ای، می‌تواند مفید باشد (تیسن، استینبرگ و فیتزپاتریک، ۱۹۸۹). با توجه به مطالب گفته شده، پژوهش حاضر با هدف تعیین ویژگی‌های روان‌سنجی سؤال‌های بخش زیست‌شناسی در آزمون سراسری سال ۱۳۹۱ بر اساس مدل چندگزینه‌ای انجام گرفته است. به علاوه، آگاهی کل آزمون زیست‌شناسی در مدل چندگزینه‌ای با آگاهی مدل سه‌پارامتری و برازش مدل چندگزینه‌ای با برازش مدل سه‌پارامتری، مقایسه می‌شود.

جامعه و نمونه

جامعه آماری این پژوهش شامل همه داوطلبان ورود به دانشگاه‌های کشور در گروه آزمایشی علوم تجربی سال ۱۳۹۱ است. در این سال تعداد افراد رشته علوم تجربی ۴۴۹۹۵۵ نفر شامل ۳۰۴۶۰۸ نفر زن و ۱۴۵۳۴۷ نفر مرد بود. با توجه به اینکه در هر یک از گروه‌های آزمایشی، سؤال‌ها در چهار دفترچه طراحی و توزیع می‌شود،

¹. don't know

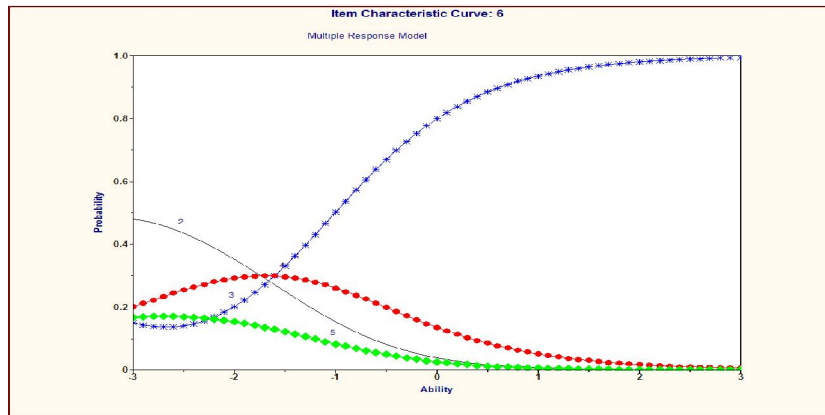
در این پژوهش فقط داده‌های مربوط به داوطلبانی تحلیل و بررسی شد که به دفترچه A پاسخ داده بودند. شایان ذکر است که تفاوت دفترچه‌ها فقط در ترتیب گزینه‌ها است. برای نمونه‌گیری با استفاده از نرم‌افزار SPSS نمونه تصادفی به تعداد ۵۰۰۰ نفر انتخاب شد. داده‌ها شامل پاسخ‌های داوطلبان به سؤال‌های چهارگزینه‌ای آزمون ۵۰ سؤالی بخش زیست‌شناسی بود. برای تحلیل داده‌ها از نرم‌افزارهای MULTILOG و jMetrik، NOHARM استفاده شد. به این صورت که پس از بررسی پیش‌فرض تک‌بعدی بودن با NOHARM به منظور برآورد پارامترهای سؤال و توانایی برای مدل چندگزینه‌ای و سه پارامتری از نرم‌افزار MULTILOG استفاده شد. برای ترسیم نمودارهای داده‌های مشاهده شده نیز از jMetrik استفاده شد.

یافته‌ها

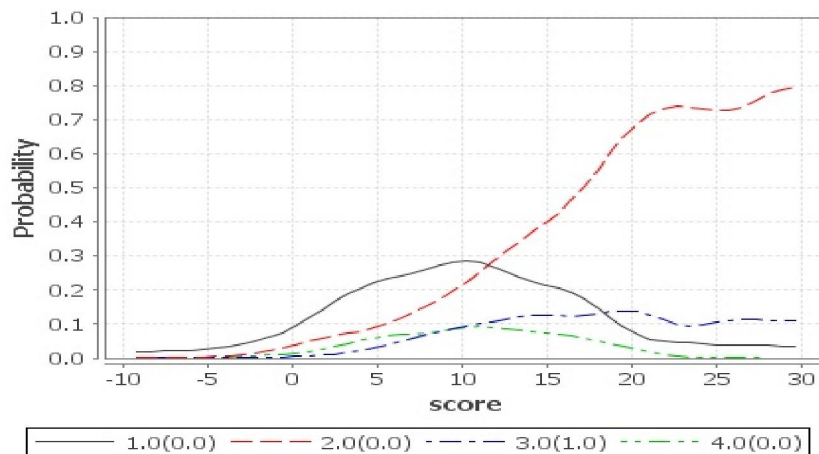
پیش از تحلیل پاسخ‌ها با استفاده از مدل چندگزینه‌ای IRT، پیش‌فرض تک‌بعدی بودن با استفاده از NOHARM بررسی شد. شاخص ریشه دوم میانگین مجذور مانده‌ها^۱ (RMSR) ۰/۰۳۷ است که به برازش خوب مدل تک‌بعدی با داده‌ها اشاره دارد. شاخص خوبی برازندگی تاناکا^۲، در این پژوهش ۰/۹۸۷ است که برازش قابل قبولی را نشان می‌دهد (ایزانلو و همکاران، ۱۳۹۳).

^۱. Root mean square of residual

^۲. Tanaka



شکل (۱) خطوط اثر گزینه‌های سؤال ۶ در مدل چندگزینه‌ای



نمودار (الف) نسبت انتخاب گزینه‌های سؤال ۶ در پیوستار نمره خام (کلید ۳)

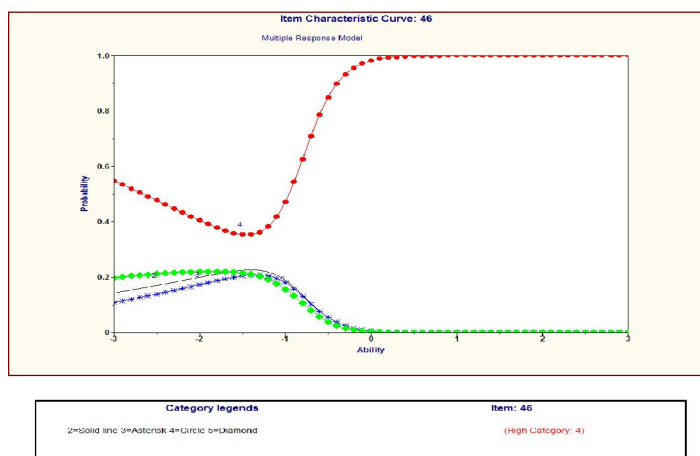
سؤال	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵	۳۶	۳۷	۳۸	۳۹				
پارامتر	d_k	Γ_{pbi_s}	$P+$	a_k	C_k	d_k	Γ_{pbi_s}	$P+$	a_k	C_k	d_k	Γ_{pbi_s}	$P+$	a_k	C_k	d_k	Γ_{pbi_s}	$P+$	
(1)DK	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
(2)DK	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
۵	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
۴	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
۳	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
۲	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
(1)DK	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰
(2)DK	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰	۳۰۰

از این‌رو، شواهد کافی برای رد پیش‌فرض تک‌بعدی بودن آزمون مذکور وجود ندارد. پارامترهای مربوط به تحلیل داده‌ها با استفاده از مدل چندگزینه‌ای در جدول (۱) ارائه شده است. این مدل برای تمام گزینه‌های هر سؤال^۱، سه پارامتر شیب (a_k)، حدس (d_k) و عرض از مبدأ (c_k) ارائه می‌کند. لازم به ذکر است که پارامترهای مربوط به گزینه نمی‌دانم (DK=Don't know)، افرادی که به سؤال، پاسخ نداده و آن را رها کرده‌اند نیز در این جدول ارائه شده است. در زیر پارامترهای هر سؤال، آماره‌های همبستگی دو رشته‌ای نقطه‌ای (r_{pt-bis}) و نسبت افرادی (P+) که هر گزینه را انتخاب کرده‌اند، ارائه شده است. تحلیل سؤال‌های آزمون زیست‌شناسی با استفاده از مدل چندگزینه‌ای نشان داد در مجموع چهار دسته سؤال در این آزمون وجود دارد. اولین دسته سؤال‌های ۶، ۲۶، ۳۷ و ۴۷ هستند که سؤال‌های نامناسبی محسوب می‌شوند. زیرا گزینه کلید در این سؤال‌ها، شیب نزولی دارد؛ به این معنی که با افزایش سطح توانایی داوطلبان احتمال انتخاب پاسخ درست کاهش می‌یابد.

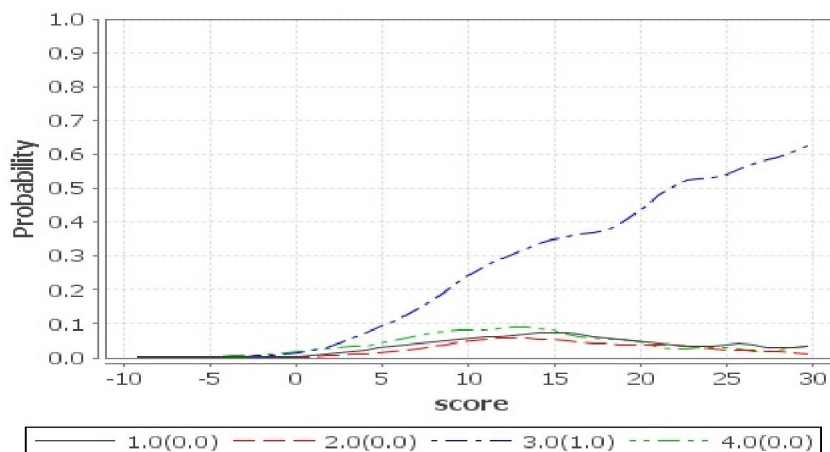
برای نمونه، به پارامترهای سؤال ۶ آزمون زیست‌شناسی (جدول ۱) و توابع پاسخ مربوط به هریک از گزینه‌های این سؤال در شکل (۱) توجه کنید (نتایج را با شواهد حاصل از داده‌های مشاهده شده هر چهار گزینه سؤال در نمودار الف مقایسه کنید). مقادیر پارامتر شیب در جدول (۱) نشان می‌دهد که گزینه انحرافی ۳ دارای بیشینه شیب مثبت است و همان‌طور که از نمودار مربوط به این گزینه انحرافی در شکل (۱) پیداست با افزایش سطح توانایی احتمال انتخاب این گزینه انحرافی به سرعت افزایش می‌یابد. این بدان معناست که بیشتر افرادی که توانایی بالایی دارند به جای انتخاب گزینه ۴ (پاسخ درست) جذب این گزینه انحرافی شده‌اند که نشان‌دهنده عملکرد معیوب گزینه انحرافی ۳ است. مقدار پارامتر c_k برای این گزینه نشان می‌دهد که بین گزینه‌های سؤال، گزینه انحرافی ۳ بیشترین میزان محبوبیت را بین داوطلبان داشته است و از بین تمام افرادی که به این سؤال پاسخ داده‌اند ۰/۲۵ درصد گزینه ۳ را انتخاب کرده‌اند. پارامتر حدس بالای گزینه انحرافی ۳ نشان می‌دهد که ۸۸ درصد از داوطلبانی که توانایی پایینی داشته‌اند و قادر به تشخیص پاسخ درست نبوده‌اند، گزینه

^۱ توجه کنید که برای تحلیل داده‌ها با مدل چندگزینه‌ای داده‌ها به جای ۰، ۱، ۲، ۳، ۴ داده‌ها به صورت (DK=۱، ۲، ۳، ۴ و ۵) کدگذاری شده‌اند. یعنی تمام گزینه‌ها، از جمله گزینه کلید، به اضافه یک شده‌اند. در مقایسه نتایج نمودارهای مربوط به داده‌های مشاهده شده با نمودارهای مدل به این نکته توجه شود.

۳ را انتخاب کرده‌اند. مقدار عددی ضریب همبستگی دو رشته‌ای نقطه‌ای (r_{pt-bis}) مربوط به گزینه انحرافی ۳ در جدول (۱) برابر $0/44$ است که نشان‌دهنده رابطه مثبت بین نمره این گزینه انحرافی با نمره کل آزمون است. همان‌طور که از خط اثر مربوط به گزینه کلید (گزینه ۴) در شکل (۱) پیداست، از سطوح پایین توانایی تا نزدیک توانایی $1/5-$ با افزایش توانایی احتمال پاسخ صحیح نیز افزایش می‌یابد و در نزدیکی توانایی $1-$ این احتمال به بیشترین مقدار خود، $0/3$ می‌رسد. اما از این سطح توانایی به بعد با افزایش توانایی احتمال پاسخ درست کاهش می‌یابد. گزینه انحرافی ۲ بین داوطلبانی که توانایی پایینی دارند از جذابیت بالایی برخوردار است و تقریباً 20 درصد از افرادی که توانایی پایینی دارند، گزینه انحرافی ۲ را انتخاب کرده‌اند. خط اثر این گزینه انحرافی نشان می‌دهد که با افزایش توانایی احتمال انتخاب گزینه ۲ کاهش می‌یابد.



شکل (۲) خطوط اثر گزینه‌های سؤال ۴۶ در مدل چندگزینه‌ای



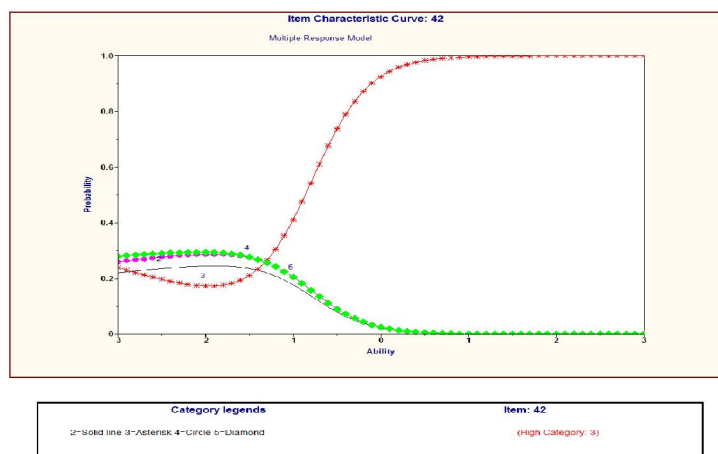
نمودار (ب) نسبت انتخاب گزینه‌های سؤال ۴۶ در پیوستار نمره خام (کلید ۳)

خط اثر مربوط به گزینه انحرافی ۵ در شکل فوق نشان می‌دهد که با افزایش توانایی، احتمال انتخاب این گزینه انحرافی کاهش می‌یابد. همچنین گزینه ۵ دارای کمترین مقدار پارامتر عرض از مبدأ در جدول بالا است ($C_5 = -0.09$) و بیانگر آن است که تنها ۵ درصد از پاسخ‌دهندگان به سؤال ۶، جذب گزینه انحرافی ۵ شده‌اند که از نظر توانایی در سطح پایینی قرار داشته‌اند. با توجه به نتایج فوق می‌توان گفت که سؤال ۶ عملکرد معیوبی داشته است.

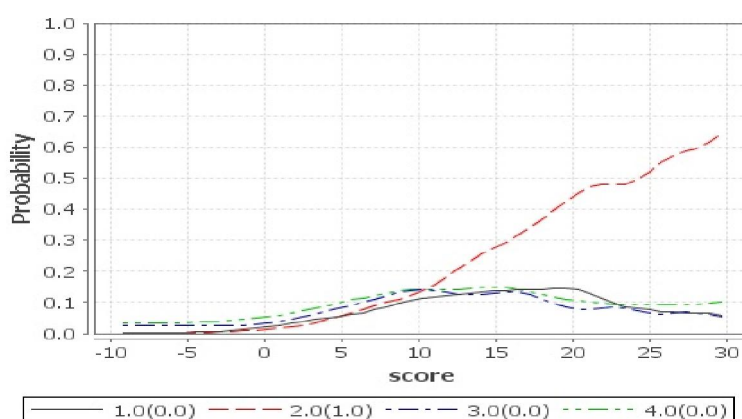
دسته دوم سؤال‌های ۲، ۴، ۲۲، ۲۹، ۳۳ و ۴۶ هستند که اینها نیز سؤال‌های نامناسبی هستند؛ زیرا گزینه کلید آنها بر اساس توانایی دارای شیب صعودی یک‌نواخت و گزینه‌های انحرافی نیز دارای شیب نزولی یک‌نواخت نیست. در مجموع، پیوستار توانایی حتی در پایین‌ترین سطوح توانایی نیز احتمال انتخاب گزینه کلید از تمامی گزینه‌های انحرافی در تمامی سؤال‌های ذکر شده بیشتر است که این امر می‌تواند به علت ساده بودن سؤال یا ناکارآمدی و جذابیت پایین گزینه‌های انحرافی باشد. به عنوان مثال، به پارامترهای سؤال ۴۶ آزمون زیست‌شناسی در جدول (۱) و توابع پاسخ مربوط به هریک از گزینه‌های آن در شکل (۲) توجه کنید (این نتایج را با نتایج نمودار ب مقایسه کنید).

مقادیر پارامتر شیب نشان می‌دهد که گزینه کلید (گزینه ۴) دارای بیشترین شیب مثبت در بین گزینه‌های سؤال ۴۶ است. همچنین گزینه کلید، بیشترین مقدار عرض از مبدأ

را به خودش اختصاص داده که نشان می‌دهد گزینه ۴ دارای بیشترین محبوبیت در بین داوطلبان بوده است، به طوری که از بین ۳۶ درصد از افرادی که به سؤال پاسخ داده‌اند نزدیک ۲۳ درصد به درستی گزینه کلید را انتخاب کرده‌اند. همان‌طور که شکل (۲) نشان می‌دهد احتمال پاسخ صحیح در سطوح پایین توانایی بسیار بالاست که این می‌تواند یا به علت آسانی بیش از حد سؤال یا به علت پارامتر حدس بالای گزینه کلید ($c_4=0/91$) باشد؛ زیرا این پارامتر حدس نشان می‌دهد که ۹۱ درصد از پاسخ‌دهندگان که قادر به تشخیص گزینه درست سؤال نبوده‌اند، توانسته‌اند به درستی گزینه کلید را حدس بزنند. مقادیر پارامترهای شیب و عرض از مبدأ گزینه‌های انحرافی ۲ و ۳ بسیار به هم نزدیک‌اند و خطوط اثر مربوط به هر دو گزینه نیز مشابه است. حتی در برخی از نقاط پیوستار توانایی کاملاً بر هم منطبق می‌شوند و هر دو گزینه انحرافی در نزدیکی سطح توانایی ۱- دارای بیشترین جذابیت هستند. لذا می‌توان این دو گزینه انحرافی را معادل یکدیگر دانست. به عبارت دیگر، هر دو گزینه در جذب داوطلبان با سطح توانایی پایین عملکرد مشابهی داشته‌اند. خط اثر گزینه انحرافی ۵ نشان می‌دهد که در فاصله توانایی $[1/5- و 3-]$ احتمال انتخاب گزینه انحرافی ۵ تقریباً ثابت است. اما از سطح توانایی $1/5-$ به بعد با افزایش توانایی احتمال انتخاب این گزینه انحرافی به تدریج کاهش می‌یابد. در مجموع می‌توان نتیجه گرفت که گزینه‌های انحرافی سؤال ۴۶ از جذابیت کافی برای جذب داوطلبان با توانایی پایین برخوردار نبوده‌اند (این نتیجه توسط ضرایب همبستگی تقریباً صفر نیز تأیید می‌شود) و بیشتر داوطلبان با توانایی پایین و متوسط به درستی گزینه کلید را حدس زده‌اند. به علاوه بر اساس توانایی گزینه کلید دارای شیب صعودی یکنواخت و گزینه‌های انحرافی نیز دارای شیب نزولی یکنواخت نیستند.



شکل (۳) خطوط اثر گزینه‌های سؤال ۴۲ در مدل چندگزینه‌ای



نمودار (ج) نسبت انتخاب گزینه‌های سؤال ۴۲ در پیوستار نمره خام (کلید ۲)

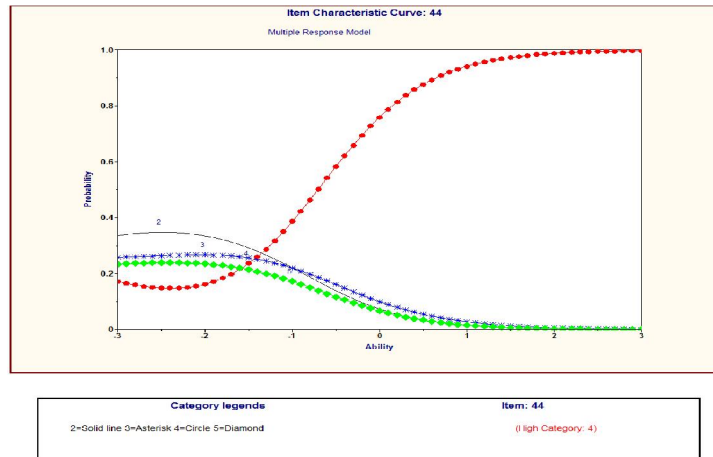
دسته سوم، سؤال‌های ۱۵، ۱۷، ۱۹ و ۴۲ که دارای سه گزینه انحرافی معادل و سؤال‌های ۱۱، ۱۲، ۱۸، ۲۱، ۲۸، ۳۶ و ۴۱ که دارای دو گزینه انحرافی معادل هستند. این وضعیت سبب می‌شود این سؤال‌ها بیشتر به صورت سؤال‌های دوگزینه‌ای عمل کنند تا سؤال‌های چهارگزینه‌ای. برای نمونه به پارامترهای سؤال ۴۲ در جدول (۱) و شکل (۳) توجه کنید (این نتایج را با نمودار ج مقایسه کنید).

همان‌طور که انتظار می‌رود گزینه کلید این سؤال (گزینه ۳) دارای بیشترین شیب مثبت در گزینه‌های سؤال ۴۲ است. همان‌گونه که در شکل (۳) مشاهده می‌شود خط

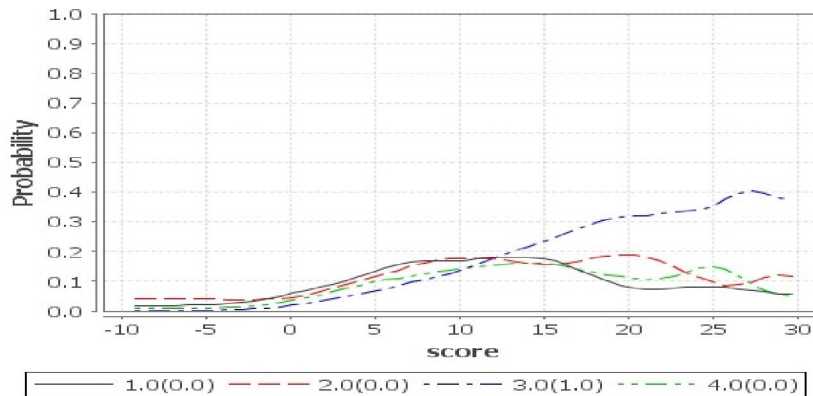
اثر گزینه کلید دارای رابطه‌ای مثبت با سطح توانایی θ است. مقادیر پارامتر عرض از مبدأ در جدول نشان می‌دهد که گزینه ۳ دارای بیشترین عرض از مبدأ است و این نشان‌دهنده محبوبیت بالای گزینه کلید در آزمودنی‌ها است. به طوری که نزدیک به ۱۸ درصد از کل پاسخ‌دهندگان به سؤال، به درستی گزینه کلید را انتخاب کرده‌اند. پارامتر حدس گزینه کلید نشان می‌دهد که ۸۰ درصد از داوطلبانی که به سؤال ۴۲ پاسخ داده‌اند و قادر به تشخیص گزینه درست نبوده‌اند، توانسته‌اند به طور صحیح، گزینه کلید را حدس بزنند. در این سؤال، پارامتر شیب گزینه‌های انحرافی ۴ و ۵ دقیقاً برابر $(-1/49)$ و پارامتر شیب گزینه انحرافی ۲ نیز تقریباً با آنها برابر است $(a_3 = -0/48)$. همچنین خطوط اثر مربوط به این گزینه‌های انحرافی مشابه هم است. همان‌طور که در مورد گزینه‌های انحرافی ۴ و ۵ مشاهده می‌شود، خطوط اثر مربوط به این دو گزینه تقریباً بر هم منطبق است. از این رو، می‌توان گفت که هر سه گزینه انحرافی معادل یکدیگرند؛ یعنی هر سه گزینه در جذب داوطلبان با توانایی پایین، عملکرد مشابهی داشته‌اند. مقادیر پارامتر عرض از مبدأ برای گزینه‌های انحرافی نشان می‌دهد که بعد از گزینه کلید به ترتیب گزینه‌های انحرافی ۵ و ۴ و ۲ دارای بیشترین محبوبیت یا فراوانی انتخاب بین داوطلبان هستند. پارامتر حدس گزینه‌های انحرافی ۲ و ۴ تقریباً با هم برابر است که نشان می‌دهد هر دو گزینه از شانس برابری برای انتخاب تصادفی برخوردارند. اما در مورد گزینه انحرافی ۵ مشاهده می‌شود که ۱۳ درصد از داوطلبانی که به سؤال پاسخ داده‌اند و قادر به تشخیص گزینه درست نبوده‌اند، به طور تصادفی گزینه ۵ را انتخاب کرده‌اند. بر اساس آماره‌های همبستگی دو رشته‌ای نقطه‌ای (r_{pt-bis}) گزینه‌های انحرافی دارای همبستگی تقریباً صفر با نمره کل هستند. توجه کنیم که در حالت ایده‌آل منفی بودن این همبستگی‌ها بهتر است. بنابراین، با توجه به اینکه هر سه گزینه انحرافی سؤال ۴۲ معادل یکدیگرند، می‌توان گفت که سؤال ۴۲ بیشتر به صورت سؤالی دوگزینه‌ای عمل کرده است تا سؤال چهارگزینه‌ای و گزینه‌های انحرافی آن نیاز به اصلاح و جایگزینی با گزینه‌های انحرافی بهتری دارند.

دسته چهارم، سؤال‌های ۲۵، ۳۵ و ۴۴ هستند. در این سؤال‌ها گزینه‌های انحرافی، نسبت به بقیه سؤال‌ها، تقریباً عملکرد مناسبی داشته و کارآمد بوده‌اند. برای مثال، به پارامترهای سؤال ۴۴ در جدول (۱) و توابع پاسخ مربوط به هریک از گزینه‌های این سؤال در شکل (۴) توجه کنید (با نتایج نمودار د مقایسه کنید).

مقادیر پارامتر شیب در جدول (۱) نشان می‌دهد که گزینه کلید (گزینه ۴) دارای بیشترین شیب مثبت در گزینه‌های سؤال ۴۴ است. خط اثر مربوط به این گزینه نیز دارای رابطه مثبت قوی با سطح توانایی است، به طوری که احتمال پاسخ صحیح در توانایی پایین بسیار کم است و به تدریج با افزایش توانایی احتمال آن به سرعت افزایش می‌یابد و بیشتر داوطلبانی که سطح توانایی بالای حد متوسط (صفر) دارند به احتمال زیاد گزینه کلید را انتخاب خواهند کرد.



شکل (۴) خطوط اثر گزینه‌های سؤال ۴۴ در مدل چندگزینه‌ای

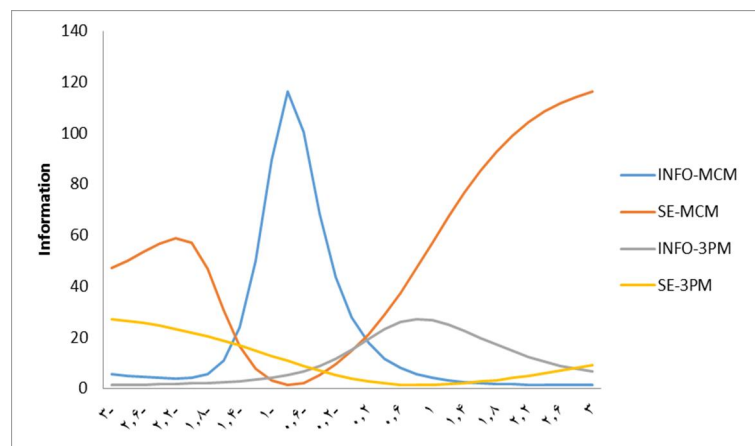


نمودار (د) نسبت انتخاب گزینه‌های سؤال ۴۴ در پیوستار نمره خام (کلید ۳)

به‌علاوه گزینه کلید دارای بیشترین مقدار پارامتر عرض از مبدأ است که نشان می‌دهد این گزینه دارای بیشترین محبوبیت در بین داوطلبان بوده است. به‌طوری که از بین ۴۴ درصد از افرادی که به سؤال ۴۴ پاسخ داده‌اند نزدیک به ۱۵ درصد گزینه کلید را انتخاب کرده‌اند. پارامتر حدس گزینه ۴ نشان می‌دهد که ۴۳ درصد از پاسخ‌دهندگان به سؤال که قادر به تشخیص گزینه درست نیستند، موفق شده‌اند به‌درستی گزینه کلید را حدس بزنند. خط اثر گزینه انحرافی ۲ دارای رابطه‌ای منفی با سطح توانایی θ است و با افزایش توانایی احتمال انتخاب این گزینه انحرافی کاهش می‌یابد. این گزینه انحرافی در حدود ۱۳ درصد از پاسخ‌دهندگان به سؤال را به خود جذب کرده است که بیشتر داوطلبانی هستند که سطح توانایی آنها کمتر از ۱- است. با افزایش سطح توانایی در فاصله [۱- و ۳-] احتمال انتخاب گزینه انحرافی ۳ ثابت است و تغییر محسوسی نمی‌کند. اما از سطح توانایی ۱- به بعد با افزایش توانایی احتمال انتخاب این گزینه انحرافی به‌تدریج کاهش می‌یابد. از بین همه پاسخ‌دهندگان به سؤال، نزدیک به ۱۴ درصد گزینه انحرافی ۳ را انتخاب کرده‌اند که بعد از گزینه کلید، بیشترین محبوبیت را در بین داوطلبان داشته است. پارامتر حدس این گزینه نشان می‌دهد که ۳۱ درصد از داوطلبانی که قادر به تشخیص گزینه کلید نبوده‌اند به‌صورت تصادفی گزینه درست را انتخاب کرده‌اند. گزینه انحرافی ۵ دارای پایین‌ترین مقدار پارامتر عرض از مبدأ است که نشان می‌دهد این گزینه انحرافی دارای کمترین محبوبیت در بین داوطلبان بوده است و تنها ۱۱ درصد از کل پاسخ‌دهندگان به سؤال را به خود اختصاص داده است. خط اثر این گزینه انحرافی نشان می‌دهد که با افزایش سطح توانایی، احتمال انتخاب این گزینه انحرافی نیز به‌تدریج کاهش می‌یابد. بر اساس آماره‌های همبستگی دو رشته‌ای نقطه‌ای، گزینه‌های انحرافی دارای همبستگی تقریباً صفر با نمره کل هستند. توجه کنیم که در حالت ایده‌آل منفی بودن این همبستگی‌ها بهتر است. با توجه به نتایج فوق می‌توان گفت که هر سه گزینه انحرافی سؤال ۴۴ در جذب داوطلبان با توانایی پایین به‌خوبی عمل کرده‌اند.

به‌منظور مقایسه میزان آگاهی مدل چندگزینه‌ای با مدل سه پارامتری، به‌عنوان مدل دو ارزشی مناسب برای مقایسه آگاهی کل در آزمون زیست‌شناسی انتخاب شد. در شکل (۵) تابع آگاهی کل آزمون بر اساس مدل چندگزینه‌ای با تابع آگاهی کل آزمون در مدل سه پارامتری ارائه شده است.

شکل (۵) نشان می‌دهد که بیشینه آگاهی آزمون بر اساس مدل چندگزینه‌ای در نزدیکی توانایی $0/8-$ قرارداد و مقدار آن تقریباً برابر 120 است. در این مدل توانایی در دامنه $0 \leq \theta \leq 1.5-$ با دقت بالایی برآورد می‌شود. اما هرچه در سطح توانایی از نقطه بیشینه فاصله بگیریم و به دو انتهای مقیاس نزدیک‌تر شویم مقدار آگاهی آزمون به شکل معناداری کاهش می‌یابد. نمودار تابع آگاهی آزمون در مدل سه پارامتری در شکل (۴) نشان می‌دهد که آزمون در دامنه $0 < \theta \leq 3$ توانایی را تا حدی با دقت برآورد می‌کند و بیشینه آگاهی که آزمون در این دامنه فراهم می‌کند در سطح توانایی $1/5$ قرار دارد که مقدار آن تقریباً 25 است.

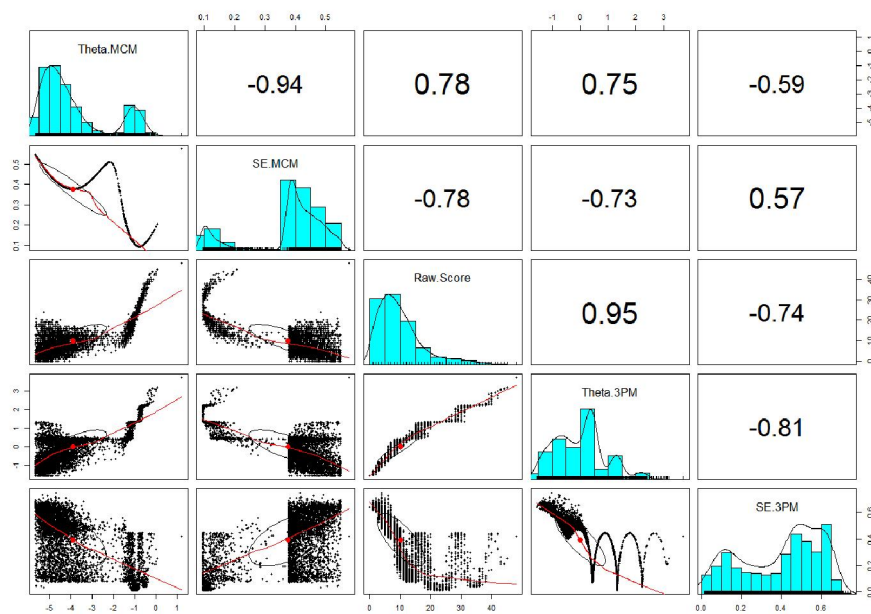


شکل (۵) تابع آگاهی کل آزمون زیست‌شناسی در مدل سه پارامتری و مدل چندگزینه‌ای همراه با خطای استاندارد آنها

برای مقایسه برازش دو مدل با داده‌ها، از آماره منفی دو لگاریتم درست‌نمایی^۱ استفاده شد. تفاوت آماره مدل چندگزینه‌ای ($614227/9$) از آماره مربوط به مدل سه پارامتری ($124581/3$) برابر $289666/6 = 614227/9 - 124581/3$ است. این تفاوت که دارای توزیع کای دو است، با توجه به درجه آزادی حاصل از تفاوت درجات آزادی مدل چندگزینه‌ای (550) از مدل سه پارامتری (150) که برابر 400 است، به‌طور معناداری بزرگ‌تر از کای دو بحرانی در سطح $0/01$ است. پس می‌توان نتیجه گرفت که مدل سه پارامتری نسبت به مدل چندگزینه‌ای برازش بهتری با مجموعه داده‌ها دارد.

^۱. -2Log-Likelihood

اگرچه تعداد پارامترهای مدل سه‌پارامتری و چندگزینه‌ای برای هر سؤال به ترتیب ۳ و ۱۱ است، ولی نتایج حاصل از آماره لگاریتم درست‌نمایی از برازش بهتر مدل سه‌پارامتری با داده‌ها حکایت دارد. شاخص‌های AIC و BIC که از آماره لگاریتم درست‌نمایی به دست می‌آیند، برای مدل چندگزینه‌ای به ترتیب $415327/9$ و $418912/4$ و برای مدل سه‌پارامتری به ترتیب $125681/3$ و $129265/8$ به دست آمد که حاکی از برازش بهتر مدل سه‌پارامتری دارد. بر اساس روش موجود در MULTILOG مدل سه‌پارامتری در مدل چندگزینه‌ای آشیانه گرفته است، به همین دلیل از رویکرد فوق برای مقایسه برازش دو مدل می‌توان استفاده کرد. در صورتی که به جای داده‌های فردی، الگوهای مشاهده شده در نمونه ۵۰۰۰ نفری تحلیل شوند باز نتایج حاکی از برازش بهتر مدل سه‌پارامتری بود.



شکل (۶) نمودار مقایسه توانایی و خطای استاندارد برآورد شده مدل چندگزینه‌ای با توانایی و خطای استاندارد مدل سه‌پارامتری و نمره خام

بر اساس توصیه تیسن و استینبرگ (۱۹۸۹) در حال حاضر برآورد توانایی با این مدل چندگزینه‌ای توصیه نمی‌شود که دلیل آن می‌تواند به خطوط اثر غیریکنواخت^۱ گزینه‌های کلید برخی از سؤال‌ها مربوط باشد. پراکندگی نامتعارف نمره‌های توانایی حاصل از مدل چندگزینه‌ای در ارتباط با نمره خام در شکل (۶) مشهود است. توانایی برآورد شده آزمودنی‌ها در مدل چندگزینه‌ای با نمره خام دارای همبستگی بالایی است در حالی که رابطه بین آنها رابطه‌ای غیرخطی است، اما رابطه توانایی برآورد شده آزمودنی‌ها در مدل سه‌پارامتری با نمره خام آزمودنی‌ها رابطه‌ای خطی است و میزان همبستگی آن نسبت به مدل چندگزینه‌ای بیشتر است. بررسی هیستوگرام توانایی هر دو مدل نشان می‌دهد که پراکندگی سطوح توانایی در مدل چندگزینه‌ای نسبت به مدل سه‌پارامتری بیشتر بوده و توزیعی دونمایی را نشان می‌دهد. از این رو اتکا به توانایی برآورد شده با مدل چندگزینه‌ای نیازمند بررسی بیشتر است.

بحث و نتیجه‌گیری

تحلیل سؤال‌های آزمون زیست‌شناسی با استفاده از مدل چندگزینه‌ای نشان داد که سؤال‌های ۶، ۲۶، ۳۷ و ۴۷ سؤال‌های نامناسبی هستند. زیرا گزینه کلید در سؤال‌های فوق دارای شیب نزولی است و این به آن معناست که با افزایش سطح توانایی داوطلبان احتمال انتخاب پاسخ درست کاهش می‌یابد و داوطلبانی که سطح توانایی پایینی دارند نسبت به داوطلبانی که از نظر توانایی در سطح بالایی قرار دارند در انتخاب گزینه کلید، عملکرد بهتری داشته‌اند؛ از این رو لازم است طراحان سؤال این سؤال‌ها را بازبینی کرده و دلیل این موضوع را مشخص کنند تا در آینده مرتکب چنین اشتباهات احتمالی نشوند. برای مثال به سؤال ۶ از این دسته در زیر توجه کنید. صرف نظر از هیچ دانش زیست‌شناسی، فقط با توجه به شکل ظاهری سؤال می‌توان گفت، سؤال از نظر اصول روان‌سنجی، نادرست طراحی شده است، به طوری که بخشی زیادی از تنه سؤال نامشخص است و فرد باید اطلاعات زیادی را یادآوری کند. آیا افراد با روش‌های مهندسی معکوس به سؤال پاسخ داده‌اند؟ انتخاب بیش از حد گزینه ۲ توسط افراد به دلیل وجود چند کلمه و نماد انگلیسی و طول کوتاه آن تا حدی این فرضیه را تأیید می‌کند.

¹. Nonmonotonic trace lines

در همه سلول‌های یوکاریوتی،.....

- (۱) پوشش هسته در پروفاز ناپدید و در تلوفاز دوباره ظاهر می‌شود.
 - (۲) همانندسازی DNA قبل از آغاز پروفاز I انجام می‌شود.
 - (۳) در شروع تقسیم سلول، رشته‌های دوک به کروموزوم‌ها اتصال می‌یابند.
 - (۴) بلافاصله پس از تقسیم هسته، غشای سلول به درون فرو رفتگی پیدا می‌کند.
- همچنین سؤال‌های ۲، ۴، ۲۲، ۲۹، ۳۳ و ۴۶ نیز سؤال‌های نامناسبی هستند، زیرا بر اساس توانایی گزینه کلید دارای شیب صعودی یکنواخت و گزینه‌های انحرافی نیز دارای شیب نزولی یکنواخت نیستند. در مجموع، پیوستار توانایی حتی در پایین‌ترین سطوح توانایی نیز احتمال انتخاب گزینه کلید از تمامی گزینه‌های انحرافی در تمامی سؤال‌های ذکر شده بیشتر است که این می‌تواند به علت ساده بودن سؤال یا ناکارآمدی و جذابیت پایین گزینه‌های انحرافی باشد. به‌عنوان نماینده این دسته به سؤال ۴۶ در زیر توجه کنید. همان مشکل سؤال قبلی در اینجا نیز مشهود است. بالا بودن پارامتر حدس گزینه کلید و طولانی‌تر بودن طول گزینه کلید نسبت بقیه فرضیه قبلی را در اینجا نیز تقویت می‌کند.
- در همه گیاهان دو ساله،.....

- (۱) شیر خام فقط توسط یک نوع آوند چوبی هدایت می‌شود.
 - (۲) تشکیل بافت‌های حاصل از رشد پسین غیرممکن است.
 - (۳) مریستم‌های نخستین در نوک ساقه و نزدیک به نوک ریشه تشکیل می‌شوند.
 - (۴) ساقه تنها محل ذخیره مواد غذایی برای تشکیل محور گل است.
- سؤال‌های ۱۵، ۱۷، ۱۹ و ۴۲ دارای سه گزینه انحرافی معادل و سؤال‌های ۱۱، ۱۲، ۱۸، ۲۱، ۲۸، ۳۶ و ۴۱ دارای دو گزینه انحرافی معادل است که سبب می‌شود این سؤال‌ها بیشتر به‌صورت سؤال دوگزینه‌ای عمل کنند تا سؤال چهارگزینه‌ای. به‌عنوان نماینده این دسته به سؤال ۴۲ در زیر توجه کنید. همان‌طور که مشخص است در تنه سؤال زیر، کلمه منفی مشخص نشده است. به‌علاوه بالا بودن پارامتر حدس گزینه کلید حاکی از ناکارآمدی گزینه‌های انحرافی سؤال است.
- در چرخه زندگی کلامیدوموناس،..... ممکن نیست.

- | | |
|----------------------------------|------------------------------|
| (۱) تولید هاگ با تقسیم میتوز | (۲) تولید گامت با تقسیم میوز |
| (۳) تولید زئوسپور در شرایط مساعد | (۴) وجود تنوع در گامت‌های |
- ملحق شده

تنها در سه سؤال ۲۵، ۳۵ و ۴۴ تمامی گزینه‌های انحرافی، عملکرد تقریباً مناسبی داشته‌اند. به‌عنوان نماینده این دسته به محتوای سؤال ۴۴ در زیر توجه کنید. متأسفانه این سؤال نیز از نظر شکل ظاهری مشکلات دو سؤال ۶ و ۴۶ را داراست. با این تفاوت که گزینه‌های انحرافی آن برای افراد ضعیف، جذاب‌تر از گزینه کلید بوده است. البته بازهم طول گزینه کلید از بقیه کمتر است. همه آرکی باکتری‌هایی که.....

- ۱) تولید کننده‌اند، در آب‌های شیرین زندگی می‌کنند.
 - ۲) هالوفیل‌اند، در خاک‌های شور زندگی می‌کنند.
 - ۳) گاز متان تولید می‌کنند، دیواره سلولی دارند.
 - ۴) ترموفیل‌اند، دمای بین ۸۰ تا ۱۲۰ درجه را به خوبی تحمل می‌کنند.
- همان‌طور که نتایج تحلیل با مدل چندگزینه‌ای نشان داد، در بیشتر سؤال‌ها یک یا دو گزینه انحرافی معیوب وجود داشت. باید توجه داشت که مهم‌ترین بخش هر آزمون چندگزینه‌ای، نوشتن گزینه‌های انحرافی است و کارآمدی و مؤثر بودن این گزینه‌ها به کارآمدی سؤال کمک زیادی می‌کند. بنابر این با توجه به نتایج حاصل از تحلیل، لازم است که در طراحی گزینه‌های انحرافی دقت بیشتری صورت گیرد تا کیفیت سؤال‌ها افزایش یابد. البته در کنار این موضوع باید به انتشار سؤال‌ها پس از اجرای آزمون بنا به دلایل نه‌چندان منطقی هم اشاره کرد. روشن است که این روند در طول زمان باعث محدودیت طراحان در طراحی سؤال و گزینه‌های انحرافی مناسب می‌شود. به‌طوری که پس از هر آزمون عرصه بر آنها تنگ‌تر می‌شود. از این رو، روی آوردن طراحان سؤال به روش‌های نامتعارف طراحی سؤال برای حفظ جدید بودن سؤال‌ها و گزینه‌ها چندان عجیب نیست.

یافته‌های پژوهش نشان داد که در کل آزمون، مدل چندگزینه‌ای در دامنه $-1.5 \leq \theta \leq 0$ توانایی را با دقت نسبتاً بالایی برآورد و مقدار آگاهی برآورد شده در $\theta = -1$ بیشتر از سطوح دیگر است (لازم به ذکر است که حداکثر آگاهی آزمون تقریباً برابر ۱۲۰ است)؛ بنابراین بر اساس مدل چندگزینه‌ای، آزمون مذکور برای آزمودنی‌هایی با این سطح توانایی بیشترین کارایی را دارد. اما برای توانایی‌های بزرگ‌تر از صفر دقت برآورد مدل سه‌پارامتری نسبت به مدل چندگزینه‌ای بیشتر است. هرچند بیشینه آگاهی فراهم شده توسط مدل سه‌پارامتری در قیاس با مدل چندگزینه‌ای بسیار پایین است. به‌علاوه، نتایج تحلیل حاکی از برتری دقت برآورد

توانایی مدل چندگزینه‌ای، نسبت به مدل سه‌پارامتری، برای آزمودنی‌ها با توانایی کمتر از صفر است. با توجه به شکل (۵) می‌توان نتیجه گرفت که برای توانایی‌های کمتر از صفر، آگاهی پاسخ‌های نادرست سبب افزایش آگاهی کل آزمون می‌شود. بنابراین می‌توان گفت که برای توانایی‌های زیر صفر در کل آزمون، مدل چندگزینه‌ای به‌طور قابل ملاحظه‌ای توانایی را با دقت بیشتری نسبت به مدل سه‌پارامتری برآورد می‌کند. اما برای توانایی‌های بزرگ‌تر از صفر مدل سه‌پارامتری میزان آگاهی نسبتاً بیشتری فراهم می‌کند. یافته‌های پژوهش در خصوص برازش مدل نشان می‌دهد که داده‌های آزمون زیست‌شناسی با مدل سه‌پارامتری نسبت به مدل چندگزینه‌ای برازش بهتری دارند. دلیل این امر می‌تواند زیادی بیش از حد پارامترهای مدل چندگزینه‌ای باشد. در کنار این موضوع کمبود شاخص‌های برازش برای بررسی برازش مدل چندگزینه‌ای است، چراکه استفاده از شاخص برازش کای دو برای تعیین برازش مدل با داده‌ها با محدودیت‌هایی همراه است که از مهم‌ترین آنها می‌توان به حجم نمونه و روش‌های برآورد پارامترها اشاره کرد، زیرا هم اندازه نمونه و هم روش برآورد پارامترهای سؤال و آزمودنی شاخص برازش کای دو را تحت تأثیر قرار می‌دهند (یوتان، ۲۰۰۷). همان‌طور که ای‌بد و همکاران (۲۰۰۹) بیان کرده‌اند مدل چندگزینه‌ای، اولین مدل جامع برای آنالیز کامل سؤال‌های چندگزینه‌ای است که آنالیز سؤال‌ها را به‌صورت خیلی خوب انجام می‌دهد اما روش‌های برازش این مدل، کمی پیچیده است. تیسن و استینبرگ (۱۹۸۹) توصیه می‌کنند که از مدل چندگزینه‌ای برای تحلیل آزمون‌های مشتمل بر سؤال‌های چندگزینه‌ای استفاده شود؛ زیرا این مدل عملکرد تمامی گزینه‌های انحرافی را به‌طور کامل بررسی می‌کند.

¹. Yutong

منابع

- ایزانلو، بلال؛ بازرگان، عباس؛ فرزاد، ولی‌اله؛ صادقی، ناهید و کاوسی، امیر (۱۳۹۳). تفکیک ابعاد متعامد از خوشه‌های سؤال بر اساس هشت روش تعیین بعد در داده های دوارزشی: مورد، آزمون ریاضی رشته ریاضی فیزیک کنکور ۹۲-۹۱. *اندازه گیری تربیتی*، ۵ (۱۸)، ۲۰۷-۲۴۰.
- معلمی اوره، مهرناز (۱۳۸۷). مقایسه دقت برآورد توانایی در سؤالات چندگزینه‌ای با به‌کارگیری مدل‌های سؤال پاسخ دو و چندارزشی. پایان‌نامه کارشناسی ارشد رشته سنجش و اندازه‌گیری دانشگاه علامه طباطبائی.
- Abad, F. J.; Olea, J., & Ponsoda, V. (2009). The Multiple-Choice Model Some Solutions for Estimation of Parameters in the Presence of Omitted Responses. *Applied Psychological Measurement*, 33 (3), 200-221.
- Bock, R. D. (1972). Estimating item parameters and latent ability when response are scored in two or more nominal categories. *Psychometrika*, 37 (1), 29- 51.
- Bock, R. D. (1997). *The nominal categories model*. In *Handbook of modern item response theory* (pp. 33-49). New York: Springer.
- De Ayala, R. J. (1993). An introduction to polytomous Item Response Theory Models. *Measurement and Evaluation in counseling Development*, 25 (4).
- Drasgow, F.; Levine, M. V.; Tsien, S.; Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19 (2), 143-166.
- Hambleton, R. k. (1989). Principles and selected applications of item Response Theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). (pp. 147-200).
- Levine, M. V. (1993). *Orthogonal functions and the niteness of continuous item response theories*. Unpublished manuscript.
- Levine, M. V., & Drasgow, F. (1983). The relation between incorrect option choice and estimated ability. *Educational and Psychological Measurement*, 43 (3), 675-685.
- Masters, G. N. (1988). An analysis of partial credit scoring. *Applied Measurement in Education*, 1, 279-297.

- Ning, R.; Waters, A. E.; Studer, C., & Baraniuk, R. G. (2015). SPRITE: A Response Model for Multiple Choice Testing. *ArXiv preprint arXiv*: 1501.02844.
- Ostini, R. & Nering, L. (2006). *Polytomous Item Response Theory Models*. Sage Publications.
- Samejima, F. (1979). *A New Family of Models for the Multiple-Choice Item* (No. RR-79-4). Tennessee Univ Knoxville Dept of Psychology.
- Smilaycoff, M. P. (1989). *An application of the Bock Samejima model for multiple category scoring to test items in which distractors contains information related to latent ability*. Dissertation Abstracts International, (12-A), 4100, (University Microfilms No. 9112783).
- Thissen, D. & Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika*, 49 (4) 567-577.
- Thissen, D. & Steinberg, L. & Fitzpatrick, A. R. (1989). Multiple-Choice Models; the Distractors are also part of Item. *Journal of Educational Measurement*. 26 (2), 161-176.
- Thissen, D. (1976). Information in wrong responses to the Raven Progressive Matrices. *Journal of Educational Measurement*, 13 (3), 201-214.
- Yutong, Yin (2007). *Using Beaton fit indices to assess goodness-of-fit of IRT models*. PHD thesis.

استناد به این مقاله:

ایزائلو، بلال؛ گرامی‌پور، مسعود و ترکاشوند، علی (۱۳۹۵). کاربرد مدل چندگزینه‌ای در تحلیل سؤال و مقایسه برآزش و آگاهی آن با مدل سه‌پارامتری (مورد: بخش زیست‌شناسی آزمون سراسری). *فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۶ (۱۴)، ۱۱ - ۳۹.