

شایسته‌گزینی در آزمون‌های استخدامی یک سازمان نظامی: کاربرد مدل‌های چندبعدی جبرانی و غیرجبرانی

نظریه سؤال پاسخ در بهینه‌سازی آزمون‌ها^۱

حسن جباری^۲؛ نورعلی فرخی^۳؛ عنایت‌الله زمان‌پور^۴؛ جلیل یونسی^۵

چکیده

نظریه‌های جدید اندازه‌گیری با تمرکز بر ویژگی‌های افراد و سوال‌ها بطور همزمان، افق جدیدی را برای بهینه‌سازی هرچه بیشتر آزمون‌ها در جهت شایسته‌گزینی بین داوطلبان، بخصوص متقاضیان ورود به سازمان‌های نظامی گشوده است. در این میان استفاده از مدل‌های جبرانی و غیرجبرانی چندبعدی نیل به اهداف مذکور را بیش از پیش میسر ساخته است. از این‌رو، پژوهش حاضر با هدف بهینه‌سازی سؤال‌های آزمون‌های استخدامی یک سازمان نظامی انجام شد. روش پژوهش، توصیفی از نوع علی-مقایسه‌ای و جامعه آماری، کلیه شرکت‌کنندگان در آزمون استخدامی سال ۱۴۰۲ آن سازمان (۱۲۰۰۰ نفر) بود که یک نمونه تصادفی ۲۵۰۰ نفری انتخاب گردید. ابزار پژوهش، آزمون‌های چهارگزینه‌ای ریاضی، زیست‌شناسی و ادبیات بود. یافته‌ها نشان داد که فرض تک‌بعدی بودن در هر سه آزمون رد شد. آزمون‌های ریاضی و ادبیات سه بعدی و آزمون زیست‌شناسی دو بعدی بود. هر سه آزمون دارای برازش بهتر با مدل جبرانی بود. در خصوص برازش سؤال‌های هر آزمون، مشخص شد که برخی از سؤال‌ها با مدل جبرانی، برخی با مدل غیرجبرانی و تعدادی هم با هر دو مدل برازش داشتند. تعدادی از سؤال‌ها هم با هیچ‌کدام از دو مدل برازش نداشتند. لزوم رعایت عدالت استخدامی در گزینش داوطلبان ورود به سازمان‌های نظامی ایجاب می‌کند که در طراحی سؤال‌های آزمون‌ها، از فنون قدرتمند نظریه سوال-پاسخ بهره گرفته شده و متناسب با حساسیت جایگاه‌های شغلی، از مدل‌های چندبعدی جبرانی یا غیرجبرانی در طراحی سؤال‌های آزمون‌ها استفاده گردد.

کلیدواژگان: نظریه سؤال پاسخ، مدل‌های چند بعدی، مدل‌های جبرانی و غیرجبرانی، آزمون‌های استخدامی

۱ - این مقاله برگرفته از رساله دکتری حسن جباری است.
۲ - دانشجوی دکتری سنجش و اندازه‌گیری دانشگاه علامه طباطبایی و
عضو هیات علمی دانشگاه امام حسین (ع)، نویسنده مسئول.
h_jabbari@atu.ac.ir

۳ - استاد دانشکده روانشناسی و علوم تربیتی دانشگاه علامه طباطبایی
۴ - استادیار دانشکده روانشناسی و علوم تربیتی دانشگاه علامه
طباطبایی
۵ - دانشیار دانشکده روانشناسی و علوم تربیتی دانشگاه علامه طباطبایی

مقدمه و بیان مساله

سازمان‌های نظامی از ارکان وجودی هر مملکتی به حساب می‌آیند که قوام و پایداری ملک به قوام و پایداری نیروهای مسلح آن وابسته است. کلام عمیق و زیبای امام علی علیه السلام که فرمود: ارتش به اذن خداوند، دژ مردم و زینت حاکمان و ارجمندی دین و راه‌های امنیت هستند (سید رضی، ۱۳۹۸)، نشان دهنده اهمیت نیروهای مسلح در هر کشوری است. این حد از اهمیت، اقتضا می‌کند تا در گزینش داوطلبان خدمت در این نیروها، نهایت دقت به عمل آید تا افرادی که شایستگی‌های لازم و شاخصه‌های نظامی‌گری را دارا هستند، برای طی کردن سیر خدمتی در این سازمان‌ها انتخاب شوند. یکی از این شایستگی‌های الزام آور در عصری فناوری، شایستگی علمی است که امروزه از طریق آزمون‌های مختلف، توسط سازمان سنجش آموزش کشور و نیز سایر مراکز دارای مجوز سنجیده می‌شود. با این وجود، تقریباً در همه آنها، با استفاده از رویکردهای مرسوم اندازه‌گیری مثل نظریه کلاسیک و مدل‌های نظریه سؤال پاسخ، اطلاعاتی کلی در قالب نمره و یا رتبه درصدی تهیه می‌شود و اطلاعات خاصی در مورد نقاط قوت و ضعف داوطلبان ارائه نمی‌شود تا به وسیله آن بتوان به یادگیرندگان در بررسی دلایل عدم موفقیت خود در آزمون و بهبود مهارت‌های شناختی خود کمک کرد.

یکی از آزمون‌های بسیار مهم که ارتباط مستقیم با آینده شغلی داوطلبان دارد، آزمون‌های استخدامی است. آزمون استخدامی عبارت است از اجرای آزمون‌های کتبی، شفاهی یا سایر آزمون‌ها به عنوان ابزاری برای تعیین شایستگی یا مطلوبیت داوطلب. با توجه به ماهیت آزمون‌های استخدامی و نوع رویکردی که در کشور ما نسبت به این آزمون‌ها وجود دارد، می‌توان آنها را جزء آزمون‌های سرنوشت‌ساز به شمار آورد. آزمون‌های سرنوشت‌ساز، ابزارهای مهمی برای ارزیابی پیشرفت فراگیران و نظارت بر کیفیت نظام‌های آموزشی هستند که این امر در مراکز آموزشی نظامی، اهمیت مضاعفی می‌یابد. در فرایند اجرای آزمون، فرض بر این است که اگر نمرات یک آزمون با عملکرد شغلی مرتبط باشد، از نظر اقتصادی برای سازمان مفید است که داوطلبان را بر اساس نمرات آن آزمون انتخاب کند. با توجه به ارتباط بین توانایی افراد و ویژگی‌های سؤال و احتمال پاسخ به سؤال، به نظر می‌رسد استفاده از نظریه سؤال-پاسخ، می‌تواند برای برآورد پارامترهای سؤال و فرد در آزمون‌های استخدامی، بخصوص آزمون‌های استخدامی سازمان‌های نظامی راهگشا باشد.

اهمیت و ضرورت انجام این تحقیق از دو جنبه حائز اهمیت است:

الف) از دیدگاه مساله محور: امروزه اهمیت آزمون‌های استخدامی بر کسی پوشیده نیست. وابستگی آینده شغلی داوطلبان ورود به مشاغل سازمان‌های مختلف کشور، بخصوص افرادی که داوطلب خدمت در مشاغل نظامی هستند؛ موجب سرنوشت‌ساز شدن آزمون‌های استخدامی شده است. لذا اصل بسیار مهم رعایت عدالت، شایسته‌گزینی و احتراز از وجود انواع مختلف سوگیری در سؤال‌ها که امروزه یکی از محورهای اصلی مورد بررسی در گستره نظریات جدید اندازه‌گیری را به خود اختصاص داده است؛ ایجاب می‌کند که در حوزه طراحی سؤال‌های آزمون‌های استخدامی هم؛ به سان سؤال‌های کنکور سراسری و حتی فراتر از آن، بررسی‌های جدی و عمیق صورت گیرد. لزوم طراحی سؤال‌ها و آزمون‌هایی که تا حد امکان بتواند عدالت و شایسته‌گزینی را در جنبه‌های مختلف

جنسیتی، نژاد، زبان و فرهنگ رعایت کند و منجر به استخدام افرادی شود که دارای همان ویژگی‌ها و مهارت‌های لازم برای مشاغل مربوطه باشند، ایجاب می‌کند که طراحی این سؤال‌ها با جدیدترین و به‌روزترین شیوه‌های موجود در حوزه سنجش و اندازه‌گیری انجام شود.

ب) از دیدگاه روش‌محور: آنچه که در یک نگاه روش‌محورانه به مساله در پژوهش حاضر، حائز اهمیت است اینکه درک پیچیدگی‌ها در نحوه تعامل افراد با سؤال‌ها می‌تواند راهگشای بسیاری از مشکلات سنجش عادلانه و شایسته‌گزینی باشد، زیرا در هنگام پاسخ دادن به سؤال‌های آزمون‌های مختلف، تفاوت قابل ملاحظه‌ای هم در تعداد فعالیت شناختی و هم در نوع آن در افراد مختلف وجود دارد و درک پیچیدگی‌های موجود در ورای داده‌های تجربی می‌تواند بر کاربرد بسیاری از مدل‌های آماری و روان‌سنجی در حوزه آموزش تأثیر داشته باشد. لذا مد نظر قرار دادن ساختار شناختی در طراحی سؤال‌های آزمون‌ها و تعیین ابعاد شناختی درگیر در پاسخ‌گویی به آن‌ها دارای اهمیت بسیار زیادی است تا از این رهگذر، بتوان مناسب‌ترین افراد را برای استخدام در سازمان انتخاب کرد. با توجه به اینکه اولاً عمده بررسی‌های انجام شده در حوزه مقایسه مدل‌های جبرانی و غیرجبرانی نظریه سؤال پاسخ با داده‌های واقعی حاصل از آزمون‌ها نبوده و با داده‌های شبیه‌سازی شده انجام شده است و ثانیاً، تاکنون در حوزه بهینه‌سازی آزمون‌های استخدامی، بخصوص در سازمان‌های نظامی؛ هیچ پژوهشی با استفاده از مدل‌های دوگانه ذکرشده انجام نشده است، لازم است یافته‌های این حوزه با استفاده از داده‌های واقعی و در آزمون‌های استخدامی یک سازمان نظامی نیز مقایسه شود تا میزان معتبر بودن مقادیر انتخابی برای پارامترها در شبیه‌سازی جهت بهینه‌سازی آزمون‌ها مشخص شود. بنابراین در پژوهش حاضر سعی بر آن است که مقایسه بین دو مدل با استفاده از داده‌های واقعی آزمون‌های استخدامی یک سازمان نظامی انجام شود. لذا این نیاز احساس شد که میزان برآورد پارامترهای سؤال و فرد در این آزمون‌ها در قالب مدل‌های چندبُعدی جبرانی^۱ و غیرجبرانی^۲ سنجیده شود. بی‌شک، با شناخت مدل بهینه، می‌توان با سنجش تطابق سؤال‌های آزمون‌های مختلف با این مدل، سؤال‌هایی را برای ذخیره در بانک سؤال انتخاب کرد که بهترین کارایی را در گزینش داوطلبان متقاضی ورود به سازمان داشته باشند تا از این رهگذر بتوان به شایسته‌گزینی داوطلبان ورودی همت گماشت. از این‌رو، هدف‌های پژوهش حاضر عبارتند از:

- تعیین میزان کارایی هر یک از مدل‌های جبرانی و غیرجبرانی نظریه سؤال پاسخ در بهینه‌سازی سؤال‌های آزمون‌های استخدامی یک سازمان نظامی،
- تعیین جبرانی یا غیرجبرانی بودن هر یک از سؤال‌های آزمون‌ها.
- و سؤال‌های پژوهش عبارتند از:
- در برآورد پارامترهای سؤال و فرد در آزمون‌های استخدامی سازمان نظامی، کدام یک از مدل‌های جبرانی یا غیرجبرانی، برازش مناسب‌تری با داده‌ها جهت انتخاب مناسب‌ترین داوطلبان برای ورود به سازمان را نشان می‌دهد؟
- کدام سؤال‌ها با مدل جبرانی و کدام سؤال‌ها با مدل غیرجبرانی برازش بهتری دارند؟

پیشینه پژوهش

پیشینه داخلی: تنها پیشینه داخلی که در حوزه کاربرد مدل‌های جبرانی و غیرجبرانی در برآورد پارامترهای سوال و فرد در آزمون‌ها انجام شده، پژوهش ایزانلو (۱۳۹۳) با عنوان «مقایسه کارایی مدل‌های جبرانی و غیرجبرانی چند بُعدی نظریه سؤال پاسخ در برآورد پارامترهای سؤال و فرد: مورد آزمون‌های سراسری ورود به دانشگاه» است. نتایج پژوهش وی نشان داد که سؤال‌های موجود در آزمون‌های تحلیل شده دارای ترکیبی از ابعاد جبرانی و غیرجبرانی هستند؛ هرچند که در کل، اکثر سؤال‌ها با مدل جبرانی برازش بهتری داشتند. همچنین سؤال‌های دارای ابعاد غیرجبرانی در آزمون‌هایی که فقط به خواندن وابسته بودند (مثل آزمون ادبیات)، نیز وجود داشت. در اکثر موارد، مدل غیرجبرانی، دومین توانایی آزمون را دقیق‌تر از مدل جبرانی برآورد کرد.

پیشینه‌های خارجی: وانگ و نایدیک^۱ (۲۰۱۵) در پژوهشی با عنوان «مقایسه دو الگوریتم برای مدرج^۲ کردن مدل چند بُعدی غیرجبرانی محدود سؤال پاسخ» به این نتیجه رسیدند که دسته غیرجبرانی مدل‌های چند بُعدی نظریه سؤال پاسخ^۳ (MIRT) اغلب، فرآیندهای شناختی زیربنایی یک سری از سؤال‌های آزمون را بهتر از دسته جبرانی مدل‌های چند بُعدی نظریه سؤال پاسخ نشان می‌دهد. دی‌مارس^۴ (۲۰۱۵) در پژوهشی با عنوان «مدل‌های نظریه سؤال پاسخ چند بُعدی جبرانی جزئی: دو فرم مدل جایگزین»، به این یافته رسیدند که هر دو مدل مورد استفاده برای شبیه‌سازی داده‌ها و مدل جبرانی به طور کلی بهترین برازش را با سوال‌های آزمون داشتند. همچنین بدون توجه به مدلی که برای شبیه سازی داده‌ها استفاده می‌شود، پارامترهای توانایی به طور مشابه توسط همه مدل‌ها بازیابی شدند. سو^۵ و وانگ (۲۰۱۹) در پژوهشی با عنوان «سنجش انطباقی چند بُعدی کامپیوتری با استفاده از مدل‌های نظریه سؤال پاسخ غیرجبرانی» به این نتیجه رسیدند که هرچه همبستگی بین صفات نهفته بیشتر باشد، دقت اندازه‌گیری و پایایی آزمون بیشتر است. همچنین فارغ از همبستگی بین صفات نهفته، میزان پایایی آزمون در ابعاد بسیار مشابه بود. نتیجه نهایی آن بود که به طور متوسط، دشواری سؤال‌ها در سطح پایین‌تری نسبت به توانایی‌های آزمون شوندگان بود.

تامانو و موکیه‌اشی (۲۰۲۳) در پژوهشی با عنوان «مدل نظریه سؤال پاسخ چند بُعدی غیر جبرانی پویا با استفاده از تقریب متغیر» به این نتیجه رسیدند که مطالعات شبیه‌سازی موید این است که روش پیشنهادی آنها قادر به بازتولید دقیق مهارت‌های نهفته است. همچنین مدل غیر جبرانی پویای پیشنهاد شده قادر به استنباط ردیابی مهارت عملی است و تمایزها را در ردیابی مهارت بین مدل‌های غیرجبرانی و جبرانی روشن می‌کند.

آیانواله^۶ و همکاران (۲۰۲۴) در تحقیقی با عنوان «مدرج‌سازی چند بُعدی نظریه سوال-پاسخ ساختار

۱ - Wang and Nydick

۲ - calibration

۳ - Multidimensional item response theory

۴ - DeMars

۵ - Hsu

۶ - Ayanwale

پاسخ دوارزشی با استفاده از زبان R برای محاسبات آماری» به این نتیجه رسیدند که از طریق تکنیک‌های تأییدی و اکتشافی، شواهد قوی مبنی بر پشتیبانی از چند بعدی بودن آزمون پیشرفت فیزیک وجود دارد. همچنین آنها به سه بعد متمایز رسیدند: پردازش شناختی، توانایی خواندن، و مهارت‌های حل مسئله.

مبانی نظری و ادبیات تحقیق

تعریف مفاهیم و اصطلاحات

نظریه سؤال-پاسخ^۱: حوزه‌ای از علم روان‌سنجی است که بر اندازه‌گیری و ارزیابی ویژگی‌های نهفته روان‌شناختی یا آموزشی، یعنی ویژگی‌های غیرقابل مشاهده مانند توانایی، هوش یا سطوح شایستگی تمرکز دارد. در این نظریه که خانواده‌ای از مدل‌های ریاضی را در بر می‌گیرد، تلاش می‌شود با بهره‌گیری از نظریه احتمالات، بازنمایی روشن و دقیقی از تعامل بین فرد و سؤال‌ها انجام شود تا امکان پیش‌بینی عملکرد فرد در موقعیت‌های مختلف فراهم گردد. همبلتن، سوامیناتان و راجرز^۲ (ترجمه فلسفی‌نژاد، ۱۳۸۹) نظریه سؤال-پاسخ را بر دو مفروضه اساسی استوار می‌دانند: الف) عملکرد آزمودنی روی یک سؤال می‌تواند به وسیله مجموعه‌ای از عوامل که صفات، صفات مکنون یا توانایی نامیده می‌شوند، پیش‌بینی شود. ب) رابطه میان عملکرد آزمودنی در سؤال و مجموعه‌ای از صفات زیربنایی به وسیله یک تابع افزایشی یکنواخت که تابع ویژه سؤال^۳ یا خم ویژه سؤال^۴ (ICC) خوانده می‌شود، قابل توصیف است. این تابع مشخص می‌سازد که هر چه سطح صفت بالاتر رود، احتمال پاسخ صحیح به سؤال نیز افزایش می‌یابد.

توانایی: منظور از توانایی که در متون نظریه سؤال-پاسخ، با حرف θ نمایش داده می‌شود، نمره محاسبه شده فرد در آزمون بر اساس مدل‌های نظریه سؤال-پاسخ است که معرف جایگاه فرد در پیوستار صفت است. (وندربلیندن و همبلتن^۵، ۱۹۹۷)

بُعد^۶: ریکاس^۷ (۲۰۰۹) بُعد را معادل تعداد توانایی ذهنی لازم برای پاسخ‌گویی به سؤال‌ها تعریف کرده است. به عبارت دیگر، اگر دادن پاسخ به سؤال‌های یک آزمون، فقط به یک توانایی شناختی وابسته باشد، آن آزمون آن را تک‌بعدی و اگر به چند توانایی شناختی وابسته باشد، آن را چندبعدی می‌نامند.

پارامتر^۸: از دیدگاه گلس و هاپکینز^۹ (۱۹۹۵)، پارامتر یک ویژگی عددی نامشخص از تابع توزیع یک متغیر در جامعه مد نظر است که بنا به علل مختلف، غیرقابل مشاهده یا محاسبه است و عملاً با استفاده از داده‌های یک نمونه برآورد می‌شود.

۱ - item response theory

۲ - Hamblen, Soaminatan & Rojers

۳ - Item Characteristic Function

۴ - Item Characteristic Curve (ICC)

۵ - Vander Linden and Hapkins

۶ - dimension

۷ - Reckase

۸ - Parameter

۹ - Glass and Hopkins

مدل‌های تک‌بعدی^۱: مدل‌های سؤال-پاسخی که در آن‌ها فرض می‌شود یک توانایی غالب برای تبیین و یا به حساب آوردن عملکرد پاسخ‌گویان کافی است، مدل‌های تک‌بعدی خوانده می‌شوند (همبلتن، سوامیناتان و راجرز؛ ترجمه فلسفی‌نژاد، ۱۳۸۹).

مدل‌های چندبعدی^۲: دسته‌ای از مدل‌های مطرح در نظریه سؤال-پاسخ هستند که نقض شدن فرض تک‌بعدی بودن را عامل اصلی عدم برازش داده‌ها با مدل‌های تک‌بعدی نظریه سؤال-پاسخ دانسته و سعی می‌کنند پیچیده بودن داده‌ها را با در نظر گرفتن سایر ابعاد شناختی لازم به‌طور دقیق و درست بازنمایی کنند (ریکاس، ۲۰۰۹) این مدل‌ها فرض می‌کنند بیش از یک توانایی برای تعیین عملکرد پاسخ‌گویان در آزمون ضروری است (همبلتن، سوامیناتان و راجرز؛ ترجمه فلسفی‌نژاد، ۱۳۸۹).

مدل چندبعدی جبرانی^۳: یک مدل چندبعدی که معتقد به جبران کاستی در یک صفت به وسیله سایر صفات درگیر است. لذا فردی که در یک توانایی یا صفت نقص دارد، این کاستی را به وسیله صفت یا صفات دیگر جبران کرده و به سؤال مورد نظر پاسخ درست می‌دهد. (انسلی و فرسیت^۴، ۱۹۸۵)

مدل چندبعدی غیرجبرانی^۵: یکی از مدل‌های چندبعدی است که جبران نقص و کاستی در یک صفت توسط صفات دیگر را ناممکن می‌داند. لذا در این نوع مدل‌ها، فردی که دارای نقص یا کاستی در یکی از صفات است، با استفاده از سایر صفات و توان‌مندی‌های ذهنی خود نمی‌تواند به سؤال-پاسخ درست بدهد. (سیمپسن^۶، ۱۹۷۸)

تعاریف عملیاتی

در این پژوهش، منظور از آزمون‌های ریاضی و آمار مقدماتی، معارف اسلامی، زبان انگلیسی و هوش و توانمندی‌های عمومی؛ پاسخ‌های داوطلبان به سؤال‌های چند گزینه‌ای آزمون‌های ذکر شده است که توسط سازمان سنجش برای متقاضیان شغل در مراکز دولتی در سال ۱۴۰۰ اجرا شده است.

مبانی نظری

اولین استفاده از آزمون‌های استاندارد استخدامی به سال ۶۰۵ بعد از میلاد در چین برمی‌گردد، زمانی از یک آزمون استاندارد سراسری به نام امتحان امپراطوری^۷ برای گزینش داوطلبان استفاده می‌شد. (کلافکورن^۸، ۲۰۱۷). به تدریج، با پیشرفت مباحث سنجش و اندازه‌گیری و ورود به حوزه آزمون‌سازی و الزام رعایت مباحث حقوقی در حوزه آزمون‌سازی، سازندگان و طراحان آزمون‌ها را در نشر بی‌مهابای آن‌ها محتاط‌تر کرد و باعث شد که توجه

۱ - Unidimensional Models

۲ - Multidimensional Models

۳ - compensational multidimensional models

۴ - Ansley & Forsyth

۵ - non-compensational multidimensional models

۶ - Sympton

۷ - imperial examination

۸ - Cloefkorn

عمیق‌تری به مبحث صفات و ابعاد آزمون صورت گیرد و تلاش شود تا فرایند سنجش برای گروه‌های مختلف جنسی و نژادی؛ مسیری عادلانه را طی کند و عاری از هر گونه سوگیری باشد (کوهن^۱، ۲۰۱۵).

یکی از تحولات اخیر در حوزه سنجش و اندازه‌گیری، ظهور دیدگاهی جدید در نظریه صفت مکنون^۲ است که عمدتاً به نظریه سؤال-پاسخ^۳ معروف است. نظریه سؤال-پاسخ، نظریه‌ای بر پایه مدل است و در پاسخ به نارسایی‌های نظریه کلاسیک اندازه‌گیری^۴ توسعه داده شده است. در حالی که در نظریه کلاسیک اندازه‌گیری، تمامی سؤال‌های آزمون دارای وزن یکسان هستند، در مدل‌های نظریه سؤال پاسخ، سؤال‌ها برای برآورد سطح صفت، وزن‌های برابری ندارند. لذا برآورد صفت دقیقاً بستگی به این دارد که به کدام سؤال‌ها پاسخ درست و به کدام سؤال‌ها پاسخ غلط داده شده است (امبرتسون و رایس^۵، ۲۰۰۰). در واقع، نظریه سؤال-پاسخ مجموعی از مدل‌هاست که یکی از راه‌های برقراری ارتباط بین متغیرهای پنهان و نشانگرهای^۶ آنها را تعریف و از خصوصیات پنهان افراد و سؤال‌ها به عنوان پیش‌بینی کننده پاسخ‌های مشاهده شده استفاده می‌کند (آیالا^۷، ۲۰۲۲). لذا در این نظریه، پارامترهای سؤال و فرد نقش متغیر پیش‌بین و احتمال پاسخ صحیح فرد به سؤال، نقش متغیر ملاک را دارد. نکته قابل توجه این است که برخلاف نظریه کلاسیک آزمون که آزمون‌محور است و نمره فرد را با جمع زدن مجموع نمرات سؤال‌های آزمون محاسبه می‌کند، نظریه سؤال پاسخ، سؤال محور است و پاسخ‌های افراد را به تک‌تک سؤال‌ها مدل‌سازی می‌کند. در این نظریه، پاسخ فرد به یک سؤال آزمون به عنوان تابعی از ویژگی فرد و ویژگی سؤال در نظر گرفته می‌شود. در نظریه سؤال-پاسخ، سطح توانایی یک فرد از روی پاسخ‌های وی به سؤال‌ها برآورد می‌شود. یک مدل نظریه سؤال پاسخ مشخص می‌کند که چگونه سطح توانایی فرد و ویژگی‌های سؤال با پاسخ‌های وی به سؤال‌ها مرتبط هستند (امبرتسون و رایس، ۲۰۰۰). لذا در این راستا لازم است میزان برآورد این پارامترها در قالب مدل‌های چندبُعدی نظریه سؤال پاسخ^۸ سنجیده شود. مدل‌های چند بُعدی که حالت خاصی از مدل‌های نظریه سؤال پاسخ هستند، بر این فرض بنا شده‌اند که تابع ریاضی مدل از یک طرف متشکل از پارامترهای یک بردار ویژگی‌های چندگانه فرد است که توصیف کننده دانش و مهارت‌های وی می‌باشد و از طرف دیگر، بردار ویژگی‌های سؤال که توصیف کننده دشواری سؤال آزمون و میزان حساسیت آن برای تمایز قائل شدن بین ویژگی‌های شخص است (ریکاس، ۲۰۰۹). دلیل انتخاب مدل‌های چندبُعدی به جای مدل‌های تک بُعدی^۹ این است که اگرچه تک بُعدی بودن، تمایل آرمانی و البته تحقق ناپذیر محققین حوزه سنجش در خصوص آزمون‌هاست ولی شواهدی وجود دارد که نشان می‌دهد مدل‌سازی داده‌های آزمون واقعی توسط مدل‌های تک بُعدی محل تردید است (اسپری^{۱۰} و همکاران، ۱۹۹۰). زیرا بعید به نظر می‌رسد تنها یک توانایی ذهنی یا شناختی بتواند به تنهایی تعیین کننده احتمال پاسخ فرد به یک یا چند سؤال باشد. اصولاً ساختار حاکم بر داده‌های حاصل از

۱ - Cohen

۲ - latent trait

۳ - Item Response Theory (IRT)

۴ - classical Test Theory

۵ - Embretson and Reise

۶ - manifestations

۷ - Ayala

۸ - multidimensional item response theory (MIRT)

۹ - unidimensional Models

۱۰ - Spray

آزمون‌ها، یک ساختار ساده نیست بلکه یک ساختار پیچیده است، به این معنی که سؤال‌ها اغلب بیش از یک توانایی ذهنی را اندازه‌گیری می‌کنند. در این راستا، اندیشه اساسی ریکاس (۲۰۰۹) این است که مدل‌های تک‌بعدی قادر به نشان دادن پیچیدگی تعامل بین آزمودنی‌ها و سؤال‌های آزمون نیستند. همچنان که به زعم شنگ و وی‌کل^۱ (۲۰۰۷)، مدل‌های چندبعدی در مقایسه با مدل‌های تک‌بعدی، می‌توانند بازنمایی^۲ بهتری از داده‌های حاصل از اجرای آزمون‌ها داشته باشند. اهمیت این مطلب آنجا مضاعف می‌شود که بدانیم اگر ساختار داده‌ها چندبعدی باشد، تأثیرات نامطلوبی بر برون‌داد نرم‌افزارهایی که بر اساس تک‌بعدی بودن طراحی شده‌اند، خواهد داشت (نانداکمار^۳، ۱۹۹۱).

از طرف دیگر، در حوزه مدل‌های چندبعدی نظریه سؤال پاسخ، که متناسب با این بحث است؛ دو رویکرد عمده و کلی در خصوص نحوه تعامل توانایی‌های ذهنی در موقع پاسخ دادن به سؤال‌های یک آزمون وجود دارد: مدل‌های جبرانی^۴ و مدل‌های غیرجبرانی^۵. مدل‌های جبرانی، بر این اندیشه اساسی استوار هستند که ماهیت صفات درگیر فرد هنگام تعامل با یک سؤال یا مجموعی از سؤال‌ها جمع‌پذیر^۶ است به‌طوری که می‌توان کاستی در یک توانایی را از طریق سایر توانایی‌ها جبران کرد. اما در مدل‌های غیرجبرانی، وضع به گونه دیگری است و ابعاد ذهنی، مستقل در نظر گرفته می‌شوند. به عبارت دیگر، در این مدل‌ها، سایر توانایی‌ها قادر به جبران کاستی در یک توانایی مفروض نیستند (اکرم، گیرل و والکر^۷، ۲۰۰۳). حال، صرف‌نظر از این مطلب که داده‌های واقعی دارای ساختاری چندبعدی هستند، نتایج به دست آمده از پژوهش‌هایی که به جای داده‌های واقعی، از داده‌های شبیه‌سازی شده استفاده کرده‌اند نیز نشان می‌دهد اینکه ابعاد دارای ماهیت جبرانی یا غیرجبرانی باشند، بر آورد پارامترهای مدل‌های تک‌بعدی را تحت تأثیر قرار می‌دهد. همچنین نتایج بررسی‌های وی، انسلی و فرسیت^۸ (۱۹۸۸) نشان داده است که اگر فرایند تحلیل عاملی ماتریس همبستگی چهارخانه‌ای را در خصوص داده‌های دو ارزشی که دارای ابعاد جبرانی و غیرجبرانی هستند، انجام دهیم؛ به ساختار عاملی متفاوتی خواهیم رسید.

روش‌شناسی پژوهش

پژوهش حاضر از لحاظ روش و ماهیت جمع‌آوری داده‌ها، توصیفی از نوع علی مقایسه‌ای و از جهت هدف، کاربردی است. زیرا از نتایج آن می‌توان جهت شناخت نقاط قوت و ضعف سؤال‌ها از یک سو و افراد پاسخ‌دهنده به آن سؤال‌ها و نیز شناخت مدل بهینه (جبرانی یا غیر جبرانی) در طراحی سؤال‌های آزمون‌های استخدامی و در نهایت، انتخاب افرادی شایسته برای خدمت در سازمان استفاده کرد. جامعه آماری این پژوهش، کلیه داوطلبانی بودند که در آزمون استخدامی سال ۱۴۰۲ این سازمان شرکت کرده بودند. آمار اعلامی سازمان در خصوص تعداد شرکت‌کنندگان، ۱۲۰۰۰ نفر بود. در خصوص اندازه نمونه، شواهدی که در پیشینه وجود دارد همگی بر نمونه‌های حداقل ۱۰۰۰ نفر توافق دارند (ین^۹ و کیم^{۱۰}، به نقل از ایزانلو، ۱۳۹۳)، با توجه به موارد فوق و همچنین رسیدن به اطمینان کافی در خصوص کفایت نمونه، اندازه نمونه ۲۵۰۰ نفری

۱ - Sheng & Wikle

۲ - representation

۳ - NandaKumar

۴ - multidimensional compensatory models

۵ - multidimensional non-compensatory models

۶ - additive

۷ - Ackerman, Gierl and Walker

۸ - Way, Ansley and Forsyth

۹ - Yen

۱۰ - Kim

برای پژوهش حاضر در نظر گرفته شد که با استفاده از فرایند نمونه‌گیری تصادفی ساده انتخاب گردید. ابزار پژوهش، سؤال‌های چهارگزینه‌ای دروس ریاضی از گروه آزمایشی ریاضی فیزیک (۳۵ سؤال)، ادبیات از گروه علوم انسانی (۲۰ سؤال) و زیست‌شناسی از گروه علوم تجربی (۲۵ سؤال) در آزمون استخدامی سازمان مربوطه بود که داوطلب می‌بایست در زمان مشخص شده به آن‌ها پاسخ می‌داد و روش نمره‌گذاری سؤال‌های آزمون نیز به صورت صفر و یک (دو ارزشی) بود. در خصوص روایی ابزار، سؤال‌های چندگزینه‌ای با توجه به اینکه توسط اساتید مجرب و منتخب دروس، طراحی می‌شوند؛ دارای روایی صوری و محتوایی لازم هستند. لذا در پژوهش‌های با ماهیت روانسنجی (مثل پژوهش حاضر) به آنها پرداخته نمی‌شود، اما سایر ویژگی‌های روانسنجی سؤال‌ها مثل دشواری و قدرت تشخیص و نیز تحلیل عاملی غیرخطی که با روایی سازه ابزار در ارتباط است، در بخش یافته‌ها مورد بررسی قرار می‌گیرد. پایایی سؤال‌های آزمون‌های مورد بررسی در قالب آزمون آلفای کرونباخ عبارت بود از: ادبیات ۰/۷۷، ریاضی ۰/۸۱ و زیست‌شناسی ۰/۷۱ که نشان‌دهنده پایایی مناسب برای ابزارهای ذکر شده است. در نهایت، داده‌های پژوهش، بنا به مورد با استفاده از بسته‌های نرم‌افزاری R و همچنین نرم‌افزار آماری SPSS نسخه ۲۱ مورد تجزیه و تحلیل آماری قرار گرفت.

یافته‌های پژوهش و تحلیل

جدول ۱: شاخص‌های توصیفی نمرات داوطلبان

تعداد سؤال	میانگین	انحراف معیار	حداقل	حداکثر	کجی	کشیدگی
۲۰	۶/۳۶	۳/۸۰	۰	۲۰	۰/۸۷	۰/۶۶
۳۵	۵/۲۷	۴/۳۷	۰	۲۹	۱/۱۲	۱/۲۹
۲۵	۲/۸۲	۲/۷۴	۰	۱۹	۲/۶۲	۳/۴۶

نتایج جدول ۱ نشان می‌دهد میانگین و انحراف معیار به ترتیب برای ادبیات فارسی ۶/۳۶ و ۳/۸۰، ریاضی ۵/۲۷ و ۴/۳۷ و زیست‌شناسی ۲/۸۲ و ۲/۷۴ است (برای هر پاسخ درست یک نمره مثبت در نظر گرفته شده و پاسخ‌های اشتباه هم نمره منفی ندارند). توزیع نمرات هر سه آزمون دارای کجی و کشیدگی مثبت که بیشترین کجی و کشیدگی مربوط به آزمون زیست‌شناسی است. در ادامه، برای هر آزمون، ابتدا تک‌بعدی بودن با استفاده از نرم‌افزار NOHARM آزمون شد. تک‌بعدی بودن ادبیات فارسی ($X^2=5853/79$, $df=169$, $P<0/001$)، ریاضی ($X^2=20407/25$, $df=559$, $P<0/001$) و زیست‌شناسی ($X^2=1083/55$, $df=274$, $P<0/001$) رد شد. لذا هر سه آزمون چندبعدی هستند و بیش از یک توانمندی ذهنی برای پاسخ‌گویی به سؤال‌های آنها لازم است. برای تعیین تعداد عوامل هر آزمون از تحلیل موازی بر روی ماتریس استفاده شد. برای آزمون ادبیات فارسی سه عامل، ریاضی سه عامل و زیست‌شناسی دو عامل شناسایی شدند. برای هر آزمون براساس تعداد عوامل شناسایی شده به روش تحلیل موازی، مدل‌های جبرانی و غیرجبرانی دو و سه پارامتری اجرا شدند و با توجه به شاخص‌های برازش مدل، مناسب انتخاب شد.

نتایج آزمون ادبیات

جدول ۲: شاخص‌های برازش مدل‌های جبرانی و غیرجبرانی آزمون ادبیات فارسی

مدل	AIC	SABIC	BIC	logLik	X ²	df	P
جبرانی ۲ پارامتری	۱۰۱۵۴۰/۱	۱۰۱۶۷۲/۳	۱۰۱۷۹۹/۴	-۵۰۷۳۰/۱	-	-	-
جبرانی ۳ پارامتری	۹۹۳۱۴/۰۴	۹۹۵۷۵/۱۱	۹۹۸۲۶/۱۴	-۴۹۵۷۸	۲۳۰۴/۰۸	۳۹	۰/۰۰۱
غیرجبرانی ۲ پارامتری	۹۹۴۳۵/۵۵	۹۹۸۲۸/۸۲	۱۰۰۲۰۷	-۴۹۵۹۸/۸	-	-	-
غیرجبرانی ۳ پارامتری	۹۹۳۲۷/۹۴	۹۹۷۸۷/۳	۱۰۰۲۲۹	-۴۹۵۲۵	۱۴۷/۶۱	۲۰	۰/۰۰۱

نتایج جدول ۲ نشان می‌دهد در آزمون ادبیات فارسی مدل جبرانی سه پارامتری در مقایسه با مدل جبرانی دو پارامتری ($X^2=230.4/0.8$, $df=20$, $P<0.001$)، و مدل غیرجبرانی سه پارامتری در مقایسه با مدل غیرجبرانی دو پارامتری ($X^2=147/61$, $df=20$, $P<0.001$) برازش بهتری دارند. بنابراین برای آزمون ادبیات فارسی مدل سه پارامتری جبرانی و سه پارامتری غیرجبرانی انتخاب شدند.

در ادامه تحلیل، مدل جبرانی در مقایسه با مدل غیرجبرانی بر اساس شاخص‌های AIC، BIC و SABIC برازش بهتری با داده‌های آزمون ادبیات فارسی داشت. برازش سؤال‌ها با مدل با آماره $S-X^2$ بررسی شد. برای کنترل خطای نوع اول، سطوح معناداری به روش هولم اصلاح شدند. برازش سؤال‌های ادبیات فارسی با مدل‌های جبرانی و غیرجبرانی در جدول ۳ نشان داده شده است.

جدول ۳. برازش سؤال‌های ادبیات فارسی با مدل‌های سه پارامتری جبرانی و غیرجبرانی

سوال	مدل جبرانی		مدل غیرجبرانی	
	S_X2	سطح معناداری	S_X2	سطح معناداری
۱	۳۵/۹۱۳	۰/۰۱۷	۴۸/۶۱۱	۰/۰۰۱
۲	۲۶/۶۸۷	۰/۲۱۱	۲۲/۹۵۰	۰/۲۵۳
۳	۱۵/۴۱۳	۱/۰۰۰	۱۲/۹۵۷	۰/۸۸۸
۴	۲۷/۶۲۰	۰/۲۰۴	۳۱/۵۸۸	۰/۰۲۶
۵	۲۳/۳۶۶	۰/۲۰۴	۲۲/۳۱۱	۰/۱۵۰
۶	۱۵/۰۷۲	۱/۰۰۰	۲۳/۶۱۵	۰/۱۵۰
۷	۴۷/۸۶۶	۰/۰۰۱	۲۴/۰۸۲	۰/۱۴۹
۸	۲۶/۱۴۶	۰/۲۰۴	۲۸/۵۹۶	۰/۰۳۹
۹	۳۶/۹۳۲	۰/۰۰۴	۴۳/۳۱۳	۰/۰۰۱
۱۰	۴۶/۸۵۵	۰/۰۰۱	۱۹/۲۹۸	۰/۲۵۳
۱۱	۲۴/۵۰۶	۰/۳۵۸	۳۵/۰۰۴	۰/۰۰۴
۱۲	۱۳/۴۹۹	۱/۰۰۰	۱۶/۰۴۶	۰/۴۹۲
۱۳	۱۱/۳۶۵	۱/۰۰۰	۱۶/۱۱۷	۰/۷۴۴
۱۴	۱۰/۱۵۳	۱/۰۰۰	۱۱/۲۵۵	۰/۸۸۸

۰/۰۰۱	۴۱/۰۶۵	۰/۰۰۱	۵۷/۱۲۰	۱۵
۰/۲۵۳	۲۱/۵۳۶	۰/۰۲۹	۳۲/۶۲	۱۶
۰/۲۵۳	۱۹/۹۳۶	۰/۴۷۶	۲۱/۲۶۱	۱۷
۰/۱۴۰	۲۴/۷۲۶	۰/۰۴۱	۳۱/۴۳۳	۱۸
۰/۸۸۸	۱۲/۷۲۱	۱/۰۰۰	۱۱/۱۹۵	۱۹
۰/۱۴۰	۲۳/۱۱۷	۰/۳۵۸	۲۱/۶۶۹	۲۰

نتایج جدول ۳ نشان می‌دهد از ۲۰ سوال ادبیات فارسی ۷ سوال با مدل جبرانی و ۶ سوال با مدل غیرجبرانی برازش نداشتند (معیار برازش، بزرگتر بودن سطح معناداری از مقدار ۰,۰۵ است). سؤال‌های شماره ۱، ۹ و ۱۵ با هیچ‌کدام از دو مدل برازش نداشتند. سؤال‌های شماره ۷، ۱۰، ۱۶ و ۱۸ با مدل جبرانی و سؤال‌های ۴، ۸ و ۱۱ با مدل غیرجبرانی برازش نداشتند.

همان‌گونه که مشاهده می‌شود، برخی از سوال‌ها با هیچ‌کدام از دو مدل جبرانی و غیرجبرانی برازش ندارند. عدم برازش سؤال‌های مذکور با هر دو مدل می‌تواند به دلیل دشواری زیاد و در نتیجه، حدس زدن بیشتر پاسخ صحیح از سوی افراد باشد. به‌خصوص توجه به این نکته نیز لازم است که وجود حدس در سؤال‌های یک آزمون، می‌تواند نشانه وجود ردپای اطلاعات نامربوط در داده‌ها باشد. لذا وجود عواملی همچون حدس‌پذیری سؤال‌ها، دشواری بیش از حد به همراه محدودیت زمانی، پرسیدن موارد خاص، تکراری بودن و وجود گزینه‌های معیوب در سؤال‌ها را از عوامل احتمالی برازش ضعیف یا عدم برازش یک سؤال با هیچ یک از دو مدل جبرانی و غیرجبرانی می‌توان به حساب آورد. بخصوص اینکه مدل غیرجبرانی به ابعاد تک‌تک سؤال‌ها حساستر از مدل جبرانی است. مطلب دیگر در حوزه چنین سؤال‌هایی که با هیچ‌کدام از دو مدل برازش مناسبی نشان نمی‌دهند این است که به نظر می‌رسد برای پاسخ‌گویی به آنها به دانش محتوایی زیادی نیاز است و یادآوری دانش محتوایی زیاد باعث می‌شود افزایش احتمال پاسخ‌گویی به سؤال‌ها فقط در بخش خاصی از سطوح توانایی رخ دهد نه در همه آن. به عبارت دیگر، در این حوزه، توانایی‌ها بیشتر از آنکه ماهیت پیوسته‌ای داشته باشند، دارای ماهیت طبقه‌ای هستند. لذا طبیعی است که برازش داده‌ها در جمیع ابعاد با هیچ‌کدام از دو مدل محقق نشود. آنچه که ماحصل این مباحث است اینکه سوال‌هایی که با هیچ‌یک از دو مدل برازش نداشته باشند، سوال‌هایی هستند که به هر دلیلی نتوانسته‌اند آنچه را که مدنظر طراح سوال است بسنجند که این امر، روایی آزمون را که از پایه‌های بنیادی در هر ابزار سنجش است، دچار مخاطره می‌کند. لذا اینگونه سوال‌ها بایست از مجموع سوال‌ها کنار گذاشته شوند تا خللی در امر شایسته‌گزینی در آزمون‌ها ایجاد نشود. از طرف دیگر، نتایج نشان داد که تعدادی از سوال‌ها با مدل غیرجبرانی برازش داشتند. محتوای نوشتاری این سؤال‌ها نشان می‌دهد که در تمامی آنها عبارت «به ترتیب» ذکر شده است. ذکر چنین عبارت‌های در تنه سؤال‌ها، بر پیچیدگی آنها می‌افزاید و روند پاسخ را با دشواری بیشتری مواجه می‌کند (دشواری یکی از ویژگی‌های روانسنجی و مهم سؤال‌های آزمون است). این موضوع حاکی از این حقیقت است که دشواری سؤال‌ها سهم مهمی در منعکس کردن ابعاد شناختی نهفته در داده‌ها دارد و تا زمانی که داده‌ها حاوی چنین اطلاعاتی باشند، شاخص‌های برازش از مدل غیرجبرانی حمایت خواهند کرد. ویژگی دیگر این سؤال‌ها این

است که چند سوال در یک سوال جا داده شده و تنه سوال یا گزینه‌ها را طولانی کرده است که این امر هم به نوبه خود بر دشواری سؤال‌ها می‌افزاید. بیشتر این سؤال‌ها از نظر محتوایی به درک مطلب و قرابت معنایی مربوط هستند. علاوه بر این، دادن پاسخ به این سؤال‌ها نیازمند زمان بیشتری است، چرا که داوطلب بایست یک بیت یا عبارت را با چندین بیت یا عبارت مقایسه کرده و پاسخ دهد. سؤال‌های ۴، ۸ و ۱۱ نیز با مدل جبرانی برازش داشتند. بررسی شکل نوشتاری این سؤال‌ها نیز نشان می‌دهد در هر سه سوال، عبارت «به‌جز» ذکر شده است. به‌نظر می‌رسد وجه متمایز این سؤال‌ها از بقیه این باشد که در این سؤال‌ها به دنبال گزینه غلط هستیم در صورتی در بقیه سؤال‌ها، گزینه درست مطلوب است. در نهایت، سؤال‌های ۲، ۳، ۵، ۶، ۱۲، ۱۳، ۱۴، ۱۷، ۱۹ و ۲۰ نیز با هر دو مدل جبرانی و غیرجبرانی برازش داشتند.

نتایج آزمون ریاضی

جدول ۴: شاخص‌های برازش مدل‌های جبرانی و غیرجبرانی آزمون ریاضی

مدل	AIC	SABIC	BIC	logLik	X2	df	P
جبرانی ۲ پارامتری	۵۵۰۴۹/۱۲	۵۵۴۱۶/۸۶	۵۵۸۵۸/۵	-۲۷۳۸۵/۶	-	-	-
جبرانی ۳ پارامتری	۵۵۰۰۷/۰۶	۵۵۴۶۷/۴	۵۶۰۲۰/۲۴	-۲۷۳۲۹/۵	۱۱۲/۰۶	۳۵	۰/۰۰۱
غیرجبرانی ۲ پارامتری	۵۵۰۹۷/۸۲	۵۵۵۳۹/۶۶	۵۶۳۱۴/۷۹	-۲۷۳۳۹/۹	-	-	-
غیرجبرانی ۳ پارامتری	۵۷۵۹۵/۱	۵۸۱۱۰/۹۴	۵۹۰۱۵/۸۷	-۲۸۵۵۳/۶	-۲۴۲۷/۲۸	۳۵	۱/۰۰۰

نتایج جدول ۴ نشان می‌دهد در آزمون ریاضی، مدل جبرانی سه پارامتری در مقایسه با مدل جبرانی دو پارامتری ($X^2=112/06$, $df=35$, $P<0/001$) برازش بهتری داشت. لذا مدل جبرانی سه پارامتری برای سؤال‌های آزمون ریاضی انتخاب شد. اما مدل غیرجبرانی سه پارامتری در مقایسه با مدل غیرجبرانی دو پارامتری ($-2427/28$ ، $X^2=35$, $df=35$, $P>0/050$) منجر به بهبود برازش مدل با داده‌ها نشد و مدل دو پارامتری غیرجبرانی انتخاب شد. مدل جبرانی در مقایسه با مدل غیرجبرانی بر اساس شاخص‌های AIC، BIC و SABIC برازش بهتری با داده‌های آزمون ریاضی داشت.

جدول ۵: برازش سؤال‌های ریاضی با مدل‌های سه پارامتری جبرانی و دو پارامتری غیرجبرانی

سوال	مدل جبرانی		مدل غیرجبرانی	
	S_X2	سطح معناداری	S_X2	سطح معناداری
۱	۴۵/۹۵۲	۰/۰۰۶	۲۶/۱۸۶	۰/۷۵۵
۲	۲۷/۴۶۶	۰/۹۲۸	۲۴/۸۱۶	۰/۷۵۵
۳	۳۲/۱۰۷	۰/۲۱۳	۳۵/۹۵۱	۰/۰۳۲
۴	۸/۵۶۶	۱/۰۰۰	۹/۴۰۳	۱/۰۰۰
۵	۳۷/۲۲۱	۰/۰۵۹	۳۵/۴۰۲	۰/۰۹۴
۶	۳۳/۰۱۴	۰/۰۷۵	۳۸/۳۵۵	۰/۰۰۸

۷	۱۷/۰۳۸	۱/۰۰۰	۲۶/۸۱	۰/۶۰۶
۸	۲۴/۴۱۳	۰/۹۳۴	۲۳/۴۹۳	۰/۷۵۵
۹	۱۹/۸۶۱	۱/۰۰۰	۳۷/۰۲۵	۰/۰۳۶
۱۰	۲۴/۰۳۰	۰/۹۶۹	۲۳/۶۶۱	۰/۷۵۵
۱۱	۳۱/۹۱۴	۰/۲۱۵	۲۴/۱۶۴	۰/۷۴۴
۱۲	۳۰/۴۱۱	۰/۳۲۰	۲۷/۵۱۱	۰/۵۴۶
۱۳	۱۲/۵۶۲	۱/۰۰۰	۱۴/۶۴۳	۱/۰۰۰
۱۴	۲۱/۹۵۰	۱/۰۰۰	۱۶/۳۵۷	۱/۰۰۰
۱۵	۲۷/۴۳۹	۰/۹۲۸	۲۵/۶۷۷	۰/۷۵۵
۱۶	۳۲/۷۹۰	۰/۱۱۵	۲۷/۲۷۲	۰/۱۶۹
۱۷	۴۵/۴۲۸	۰/۰۰۷	۳۴/۱۰۸	۰/۱۳۱
۱۸	۳۶/۴۷۳	۰/۰۹۵	۲۵/۳۷۳	۰/۷۴۴
۱۹	۳۹/۷۱۵	۰/۰۴۴	۳۹/۷۴۹	۰/۰۲۶
۲۰	۲۵/۱۶۱	۰/۹۶۹	۲۷/۹۷۵	۰/۳۳۰
۲۱	۲۳/۸۹۵	۱/۰۰۰	۱۹/۷۱۸	۱/۰۰۰
۲۲	۴۴/۰۷۹	۰/۰۰۶	۲۴/۴۶۰	۰/۵۷۰
۲۳	۱۸/۶۹۲	۱/۰۰۰	۱۸/۳۳۹	۱/۰۰۰
۲۴	۲۲/۹۵۴	۱/۰۰۰	۲۴/۳۷۸	۰/۷۴۲
۲۵	۲۰/۵۹۱	۱/۰۰۰	۲۱/۶۴۱	۰/۷۵۵
۲۶	۶۱/۶۷۳	۰/۰۰۱	۶۳/۶۰۱	۰/۰۰۱
۲۷	۲۹/۶۷۹	۰/۵۴۶	۲۸/۰۵۰	۰/۶۰۶
۲۸	۳۶/۸۳۶	۰/۰۶۲	۳۱/۴۸۵	۰/۱۲۳
۲۹	۳۶/۶۷۳	۰/۰۹۳	۲۱/۵۹۸	۰/۷۸۵
۳۰	۱۵/۰۸۵	۱/۰۰۰	۳۲/۴۶۱	۰/۰۵۷
۳۱	۱۲/۲۳۲	۱/۰۰۰	۱۱/۰۵۴	۱/۰۰۰
۳۲	۱۶/۳۰۷	۱/۰۰۰	۱۲/۴۱۷	۱/۰۰۰
۳۳	۲۲/۸۱۹	۱/۰۰۰	۲۱/۸۵۷	۰/۸۹۳
۳۴	۳۳/۹۵۱	۰/۰۶۱	۵۴/۵۶۰	۰/۰۰۱
۳۵	۳۳/۷۴۶	۰/۰۶۲	۳۸/۳۴۵	۰/۰۰۸

نتایج جدول ۵ نشان می‌دهد از ۳۵ سوال ریاضی، ۵ سوال با مدل جبرانی و ۷ سوال با مدل غیرجبرانی برازش نداشتند. سؤال‌های شماره ۱۹ و ۲۶ با هیچ‌کدام از دو مدل برازش نداشتند. سؤال‌های شماره ۱، ۱۷، ۲۲ با مدل جبرانی سه پارامتری و سؤال‌های ۳، ۶، ۹، ۳۴ و ۳۵ با مدل غیرجبرانی دو پارامتری برازش نداشتند.

در حوزه سؤال‌های آزمون ریاضی، نتایج نشان داد که این آزمون هم چندبعدی بوده و بیش از یک توانمندی ذهنی برای پاسخ‌گویی به سؤال‌های آن لازم است. سایر نتایج نیز نشان داد که برای آزمون، سه بعد شناسایی شد. علاوه بر آن، مدل جبرانی در مقایسه با مدل غیرجبرانی، برازش بهتری با داده‌های آزمون ریاضی داشت. در حوزه سؤال‌های آزمون ریاضی، نتایج نشان داد از ۳۵ سوال آزمون، سؤال‌های ۱۹ و ۲۶ با هیچ‌کدام از دو مدل برازش نداشتند. این سؤال‌ها دارای دشواری زیادی بوده و به هندسه و احتمالات مربوط هستند. سؤال‌های ۱، ۶، ۱۷ و ۲۲ با مدل

غیرجبرانی برازش داشت. بررسی این سؤال‌ها نشان می‌دهد که چندین توانایی برای پاسخ‌گویی به آنها لازم است و هیچ‌یک از این توانمندی‌ها قادر به جایگزین شدن توسط توانمندی دیگر نیستند. سؤال‌های ۳، ۹، ۳۴ و ۳۵ نیز با مدل جبرانی برازش داشتند. این سؤال‌ها به حوزه جبر و مثلثات تعلق داشتند و ویژگی مشترکشان این بود در صورت عدم بهره‌مندی داوطلب از یک توانمندی خاص، با استفاده از توانمندی‌های جایگزین می‌توان تا حدی به آنها پاسخ داد. در نهایت، اکثر سؤال‌های آزمون ریاضی یعنی سؤال‌های ۲، ۴، ۵، ۷، ۸، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۲۰، ۲۱، ۲۳، ۲۴، ۲۵، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲ و ۳۳ با هر دو مدل جبرانی و غیرجبرانی برازش داشتند.

نتایج آزمون زیست‌شناسی

جدول ۶: شاخص‌های برازش مدل‌های جبرانی و غیرجبرانی آزمون زیست‌شناسی

مدل	AIC	SABIC	BIC	logLik	X ²	df	P
جبرانی ۲ پارامتری	۳۹۷۷۱/۴۷	۳۹۹۶۹/۳۴	۴۰۲۰۴/۴۶	-۱۹۸۱۱/۷	-	-	-
جبرانی ۳ پارامتری	۳۹۷۹۶/۷۴	۴۰۰۶۱/۴۷	۴۰۳۷۶/۰۲	-۱۹۷۹۹/۴	۲۴/۷۳	۲۵	۰/۴۸۰
غیرجبرانی ۲ پارامتری	۳۹۷۱۴/۸۸	۳۹۹۷۹/۶۱	۴۰۲۹۴/۱۶	-۱۹۷۵۸/۴	-	-	-
غیرجبرانی ۳ پارامتری	۴۵۴۸۷/۹۷	۴۵۸۱۹/۵۴	۴۶۲۱۳/۵۳	-۲۲۶۲۰	-۵۷۲۳/۰۹	۲۵	۱/۰۰۰

نتایج جدول ۶ نشان می‌دهد در آزمون زیست‌شناسی، مدل جبرانی سه پارامتری در مقایسه با مدل جبرانی دو پارامتری ($X^2=24/73$, $df=25$, $P>0/050$)، و مدل غیرجبرانی سه پارامتری در مقایسه با مدل غیرجبرانی دو پارامتری ($X^2=5723/09$, $df=25$, $P>0/050$) به بهبود معنادار برازش مدل با داده‌ها منجر نشد. بنابراین برای زیست‌شناسی مدل دو پارامتری جبرانی و دو پارامتری غیرجبرانی انتخاب شدند. مدل جبرانی در مقایسه با مدل غیرجبرانی بر اساس شاخص‌های AIC، BIC و SABIC برازش بهتری با داده‌های آزمون زیست‌شناسی داشت.

جدول ۷: برازش سؤال‌های زیست‌شناسی با مدل‌های دو پارامتری جبرانی و دو پارامتری غیرجبرانی

سوال	مدل جبرانی		مدل غیرجبرانی	
	S_X2	سطح معناداری	S_X2	سطح معناداری
۱	۱۷/۷۳۹	۱/۰۰۰	۱۲/۶۰۶	۱/۰۰۰
۲	۲۸/۰۷۳	۰/۱۳۰	۱۹/۶۵۱	۱/۰۰۰
۳	۲۱/۸۸۲	۱/۰۰۰	۱۲/۸۹۸	۱/۰۰۰
۴	۲۲/۴۰۶	۱/۰۰۰	۱۷/۷۸۶	۱/۰۰۰
۵	۱۵/۹۵۸	۱/۰۰۰	۱۵/۰۴۷	۱/۰۰۰
۶	۲۴/۸۱۰	۰/۵۳۸	۱۶/۷۲۹	۱/۰۰۰
۷	۹/۳۲۶	۱/۰۰۰	۵/۴۰۰	۱/۰۰۰
۸	۱۶/۳۹۶	۱/۰۰۰	۱۵/۲۲۱	۱/۰۰۰
۹	۷/۸۸۹	۱/۰۰۰	۶/۳۵۰	۱/۰۰۰

۱۰	۳۰/۲۸۸	۰/۰۶۳	۱۸/۸۸۳	۱/۰۰۰
۱۱	۱۵/۶۶۴	۱/۰۰۰	۱۴/۷۷۶	۱/۰۰۰
۱۲	۱۹/۱۲۳	۱/۰۰۰	۱۹/۳۸۳	۱/۰۰۰
۱۳	۱۴/۲۰۹	۱/۰۰۰	۱۳/۸۱۶	۱/۰۰۰
۱۴	۲۴/۷۲۶	۰/۳۷۲	۲۸/۰۶۳	۰/۰۷۹
۱۵	۷/۵۸۴	۱/۰۰۰	۹/۰۳۰	۱/۰۰۰
۱۶	۱۰/۶۹۶	۱/۰۰۰	۱۱/۳۶۵	۱/۰۰۰
۱۷	۱۹/۷۹۲	۱/۰۰۰	۱۶/۸۶۳	۱/۰۰۰
۱۸	۲۱/۲۸۵	۰/۹۲۷	۲۳/۶۶۹	۰/۳۴۱
۱۹	۱۳/۰۳۱	۱/۰۰۰	۱۲/۹۲۴	۱/۰۰۰
۲۰	۷/۶۵۴	۱/۰۰۰	۹/۴۱۴	۱/۰۰۰
۲۱	۱۰/۶۶۲	۱/۰۰۰	۹/۴۲۷	۱/۰۰۰
۲۲	۱۸/۷۰۵	۱/۰۰۰	۱۷/۰۱۶	۱/۰۰۰
۲۳	۱۱/۵۶۵	۱/۰۰۰	۱۲/۲۰۹	۱/۰۰۰
۲۴	۲۱/۶۲۶	۰/۸۸۰	۱۸/۱۷۶	۱/۰۰۰
۲۵	۸/۲۰۵	۱/۰۰۰	۸/۳۴۱	۱/۰۰۰

نتایج جدول ۷ حاکی از این است که همه ۲۵ سوال آزمون زیست‌شناسی با هر دو مدل دو پارامتری جبرانی و دوپارامتری غیرجبرانی برازش دارند.

در حوزه سؤال‌های آزمون زیست‌شناسی، نتایج نشان داد که این آزمون چندبعدی بوده و بیش از یک توانمندی ذهنی برای پاسخ‌گویی به سؤال‌های آن لازم است. همچنین بر خلافت دو آزمون ادبیات و ریاضی که هر کدام دارای سه بعد بودند، برای این آزمون، دو بعد شناسایی شد. علاوه بر آن، مانند دو آزمون قبل، مدل جبرانی در مقایسه با مدل غیرجبرانی، برازش بهتری با داده‌های آزمون زیست‌شناسی داشت. در حوزه سؤال‌های آزمون، نتایج نشان داد تمامی ۲۵ سوال آزمون با هر دو مدل برازش داشتند. این یافته می‌تواند ناشی از دشواری یکنواخت سؤال‌ها در طول آزمون باشد. به نظر می‌رسد تا زمانی که پاسخ دادن به سوالی وابستگی تقریباً یکسانی به دو توانایی داشته باشد، هر دو مدل قادر به ارایه تبیین یکسانی از داده‌ها هستند. این یافته نشان می‌دهد سؤال‌های مفهومی با هر دو مدل برازش قابل قبولی دارند.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف تعیین میزان کارایی هر یک از مدل‌های جبرانی و غیرجبرانی نظریه سؤال پاسخ در بهینه‌سازی سؤال‌های آزمون‌های استخدامی یک سازمان نظامی و تعیین جبرانی یا غیرجبرانی بودن هر یک از سوال‌های آزمون‌ها انجام شد. سؤال‌های پژوهش این بود که در برآورد پارامترهای سؤال و فرد در آزمون‌های استخدامی سازمان نظامی، کدامیک از مدل‌های جبرانی یا غیرجبرانی، برازش مناسب‌تری با داده‌ها جهت انتخاب مناسب‌ترین داوطلبان برای ورود به سازمان را نشان می‌دهد و کدام سؤال‌ها با مدل جبرانی و کدام سؤال‌ها با مدل غیرجبرانی برازش بهتری دارند؟

همان‌گونه که نتایج تحقیق نشان داد، سؤال‌های آزمون‌های سه‌گانه مورد بررسی، دارای ترکیبی از ابعاد جبرانی و غیرجبرانی هستند. علاوه بر آن، به نظر می‌رسد سؤال‌هایی که ناوابسته به بافت کتاب و مستلزم ترکیب کردن مباحث چندین حوزه محتوایی مختلف هستند و شکل طراحی آنها با قالب‌های رایج طراحی سوال متفاوت است یا نیازمند استدلال هستند، دارای ماهیت غیرجبرانی در ابعاد خود هستند. متقابلاً، سؤال‌هایی که با سطح توانایی داوطلبان متناسب هستند یا دارای قابلیت حدس‌پذیری بیشتری هستند، ابعادشان جبرانی است. توجه به این موارد برای طراحان سوال، بخصوص کسانی که در حوزه طراحی سؤال‌های آزمون‌های استخدامی نهادها و ارگان‌های مختلف مشغول به فعالیت هستند، حایز کمال اهمیت است. این اهمیت هم در حوزه طراحی سؤال‌های عمومی و هم در حوزه سؤال‌های تخصصی مطرح است. نتایج پژوهش حاضر با وضوح تمام نشان داد که لزوماً این‌گونه نیست که همه یا بیشتر سؤال‌های یک آزمون با مدل جبرانی یا غیرجبرانی برازش بهتری داشته باشد. بلکه همان‌گونه که ایزانلو (۱۳۹۳) نیز اذعان دارد، این نوع طراحی سؤال، میزان وابستگی آن به فرایندهای شناختی بالاتر، طراحی سؤال بر اساس ترکیب مباحث چندین حوزه محتوایی و طراحی سؤال بدون وابستگی به متن کتب درسی، همگی قادر است تا ماهیت ابعاد درگیر در پاسخ‌گویی به سؤال‌ها را به سمت مدل غیرجبرانی سوق دهد. ضمن اینکه در دروسی که صرفاً به خواندن یا یادآوری متکی هستند، نیز امکان طراحی سؤال‌های دارای ابعاد غیرجبرانی وجود دارد. این یافته با نتایج تحقیق‌های ایزانلو (۱۳۹۳) و بُلت و لال (۲۰۰۳) همسو است. مطلب دیگر اینکه هرچند هر سه آزمون با مدل جبرانی برازش مناسب‌تری نشان دادند، اما هر آزمون ترکیبی از سوال‌هایی بود که به‌طور عمده با هر دو مدل جبرانی و غیرجبرانی برازش داشت. این یافته تلویحات مهمی به دنبال دارد، از جمله اینکه لزومی ندارد سوال‌های یک آزمون فقط جبرانی یا فقط غیرجبرانی باشند زیرا جبرانی و غیرجبرانی بودن دارای یک وضعیت گسسته و صفر و صدی نیست و جبرانی یا غیرجبرانی بودن یک پیوستار است که در یک انتهای آن حالت جبرانی بودن قرار دارد و هر چه به سمت دیگر پیوستار حرکت نماییم، از جبرانی بودن سوال کاسته شده و ویژگی غیرجبرانی بودن آن افزایش می‌یابد. این یافته با نتایج تحقیق اکرمی و ترنر (۲۰۰۳)، وانگ و نایدیک (۲۰۱۵)، دی‌مارس (۲۰۱۵) و سو و وانگ (۲۰۱۹) همسو است.

پیشنهادهای کاربردی: با توجه به یافته‌های این پژوهش، پیشنهاد می‌شود:

۱. سوال‌هایی که به هیچ یک از دو مدل جبرانی و غیرجبرانی برازش مناسبی نشان نمی‌دهند، از مجموعه سوال‌های بانک سوال حذف شوند.

۲. در چینش سؤال‌ها، کل محتواهای یک آزمون در نیمه اول آزمون گنجانده شود و سؤال‌های نیمه دوم، تکرار همان محتواها باشد تا تاثیر محدودیت زمانی به کمترین حد خود برسد و عامل سرعت کمترین خلل را در پاسخ‌دهی داوطلبان ایجاد کرده و توانمندی‌های داوطلبان بهتر به منصه ظهور برسد زیرا با انجام اینکار، اثر مخرب محدودیت زمانی که به ایجاد تورش در میزان توانمندی‌های داوطلبان منجر می‌شود، از روی نمرات برداشته شده و ضعف‌های نهفته در پاسخ‌های داوطلبان برای متولیان امر روشن‌تر می‌شود.

۳. تجارب آزمون‌های مختلف نشان داده است که یکنواختی طراحی سؤال‌ها در قالب‌های کلیشه‌ای و تکراری مثل سؤال‌های چهار گزینه‌ای رایج، به مرور باعث تبحر داوطلبان در پاسخ‌گویی به سؤال‌ها شده، بدون اینکه بر اساس توانمندی‌ها یا

مطالعات پیشین خود پاسخ داده باشند که ایجاد یک واریانس خطای قابل توجه می‌کند. لذا لازم است طراحان سوال خود را از قالب‌های رایج جدا کرده و در طراحی خود، خلاقیت و ابتکار به خرج دهند.

۳. شناخت و سنجش منابع واریانس موجود در داده‌ها امر بسیار مهم و قابل توجهی برای مراکز طراحی سوال است. لازم است تا حد امکان سؤال‌ها به نحوی طراحی شوند تا با شناخت سهم هر یک از ویژگی‌های مربوط یا نامربوط در واریانس داده‌ها، گام‌های مهمی در راستای پالایش داده‌ها از منابع واریانس نامربوط به سؤال‌ها و محتوای آموزشی برداشته شده و سهم توانمندی‌های دانشی و شناختی داوطلب در پاسخ دادن به سؤال‌ها به حداکثر مقدار خود برسد.

منابع

- ایزانلو، بلال (۱۳۹۳). مقایسه کارایی مدل‌های جبرانی و غیرجبرانی چندبعدی نظریه سؤال پاسخ در برآورد پارامترهای سوال و فرد: مورد آزمون‌های سراسری ورود به دانشگاه. رساله دکتری دانشگاه تهران.
- رضی، سید مرتضی (۱۳۹۸). نهج البلاغه. ترجمه حسین انصاریان. آیین دانش.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational measurement: issues and practice*, 22(3), 37-51.
- Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 37-48.
- Babcock, B.G. (2011). Estimating a non-compensatory IRT model using Metropolis within Gibbs sampling. *Applied Psychological Measurement*, 35, 317-329.
- Bielinski, J., & Davison, M. L. (2001). A Sex Difference by Item Difficulty Interaction in Multiple-Choice Mathematics Items Administered to National Probability Samples. *Journal of Educational Measurement*, 38(1), 51-77.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, 27(6), 395-414.
- Cloefkorn, Sh. (2017). The History of Skills Testing: <https://www.eskill.com/>
- Cohen, B. (2015). Hiring the best talents, Whitepaper-Article. <https://painlesshire.com/history-pre-employment-testing/>
- De Ayala, R. J. (2022). *The theory and practice of item response theory*. Guilford Publications.

- DeMars, C. E. (2016). Partially compensatory multidimensional item response theory models: Two alternate model forms. *Educational and Psychological Measurement*, 76(2), 231-257.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for Psychologists*. Psychology Press.
- Freeman, L. L. (2016). *Assessing model-data fit for compensatory and non-compensatory multidimensional item response models using Vuong and Clarke statistics* (Doctoral dissertation, The University of Wisconsin-Milwaukee).
- Fu, Y., Strachan, T., Ip, E. H., Willse, J. T., Chen, S. H., & Ackerman, T. (2020). The recovery of correlation between latent abilities using compensatory and noncompensatory multidimensional IRT models. *International Journal of Testing*, 20(2), 169-186.
- Hsu, C. L., & Wang, W. C. (2019). Multidimensional computerized adaptive testing using non-compensatory item response theory models. *Applied psychological measurement*, 43(6), 464-480.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Luecht, R. M., & Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement*, 16, 279-293.
- Nandakumar, R. (1991). Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement*, 28(2), 99-117.
- Ozdemir, B., & Gelbal, S. (2022). Measuring language ability of students with compensatory multidimensional CAT: A post-hoc simulation study. *Education and Information Technologies*, 1-22.
- Reckase, M. D. (2009). *Multidimensional item response theory*: Springer Verlag.
- Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20(4), 355-371.
- Seong, T. J., & Subkoviak, M. J. (1987). A Comparative Study of Recently Proposed Item Bias Detection Methods.
- Sheng, Y., & Wikle, C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67(6), 899-919.
- Spray, J. A., Davey, T. C., Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1990). Comparison of two logistic multidimensional item response theory models: DTIC Document.

- Tamano, H., Mochihashi, D. (2023). Dynamical Non-compensatory Multidimensional IRT Model Using Variational Approximation. *Psychometrika* 88, 487–526 .
<https://doi.org/10.1007/s11336-023-09903-y>
- Wang, C., & Nydick, S. W. (2015). Comparing two algorithms for calibrating the restricted non-compensatory multidimensional IRT model. *Applied Psychological Measurement*, 39(2), 119-134.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-252.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4), 479–494.