

شناسایی رویکرد بهینه همتراسازی در آزمون‌های چندبعدي^۱

سمیه بهمن‌آبادی*
محمدرضا فلسفی‌نژاد**
علی دلاور***
نورعلی فرخی****
اصغر مینایی*****

چکیده

همتراسازی از مهم‌ترین مسائل در نظام سنجش تربیتی است و عدم برقراری مفروضات آن، اندازه‌گیری را با چالش‌های جدی مواجه می‌کند. در آزمون‌های چندبعدي، استفاده از روش‌های تک‌بعدي همتراسازی، سوگیری نتایج را به دنبال دارد. هدف از اجرای این پژوهش، شناسایی رویکرد بهینه همتراسازی در آزمون‌های چندبعدي بود. شش روش همتراسازی تک‌بعدي و چندبعدي مقایسه شدند. روش همصداک به لحاظ مقاوم بودن در برابر مفروضه تک‌بعدي بودن به‌عنوان ملاکی برای مقایسه سایر روش‌ها مدنظر قرار گرفت. جامعه آماری شامل همه داوطلبان گروه علوم ریاضی و فنی کنکور سراسری سال ۹۶ و ۹۷ بود. داده‌های آزمون ریاضی تعداد ۵۰۰۰ نفر از هر دو سال برای همتراسازی انتخاب شدند. ابعاد آزمون‌ها با استفاده از نرم‌افزارهای MPLUS و NOHARM مشخص شد. برای برآورد پارامترهای تک‌بعدي از بسته ltm و برای برآورد پارامترهای چندبعدي از بسته mirt در نرم‌افزار R استفاده شد. همتراسازی نمره مشاهده شده و نمره واقعی تک‌بعدي IRT با استفاده از نرم‌افزار PIE و همتراسازی همصداک با استفاده از نرم‌افزار R بسته equate انجام گرفت. از بسته‌های mvnorm، mirt و MASS در نرم‌افزار R نیز برای همتراسازی چندبعدي داده‌ها استفاده شد. نتایج نشان داد که در همتراسازی آزمون‌های چندبعدي بهینه‌ترین روش، همتراسازی چندبعدي کامل است و پس از آن روش نمره مشاهده شده چندبعدي تک‌بعدي شده، روش مناسبی است، اما استفاده از روش‌های تک‌بعدي نمره مشاهده شده و نمره واقعی در چنین شرایطی، کارایی مناسبی ندارد. بنابراین پیشنهاد می‌شود در همتراسازی آزمون‌هایی که ساختار چندبعدي دارند، از روش همتراسازی نمره مشاهده شده چندبعدي کامل استفاده شود.

واژگان کلیدی: همتراسازی چندبعدي کامل، همتراسازی نمره واقعی چندبعدي تک‌بعدي شده، همتراسازی نمره مشاهده شده چندبعدي تک‌بعدي شده، همتراسازی همصداک، نظریه پرسش پاسخ چندبعدي

^۱ این مقاله برگرفته از رساله دکتری رشته سنجش و اندازه‌گیری دانشگاه علامه طباطبائی است.

* دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران

** دانشیار گروه سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران (نویسنده مسئول)

falsafinejad@yahoo.co.uk

*** استاد ممتاز گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.

**** دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران

***** دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران

مقدمه

مقایسه‌پذیری اندازه‌گیری‌ها و آزمایش‌های انجام گرفته در شرایط مختلف با روش‌ها و پژوهشگران مختلف از شروط اساسی در همه علوم است و اندازه‌گیری‌های روانی و تربیتی نیز از این قاعده مستثنی نیست. مقایسه‌پذیری نمره‌های آزمون‌های آموزشی در تصمیمات مهمی همچون پذیرش دانشگاهی، دریافت بورسیه تحصیلی و تعیین شایستگی‌های افراد در مشاغل مختلف نقش اساسی دارد (دورانز و هالند^۱، ۲۰۰۰؛ گلزانز و وایبرگ^۲، ۲۰۱۷). در شرایطی که آزمون‌ها دارای فرم‌های چندگانه هستند و این فرم‌ها ویژگی‌های یکسان اما دشواری متفاوتی دارند، همترازسازی آزمون‌ها به دلیل کنترل کردن اثر متغیر مخدوش‌کننده فرم آزمون (وون داویر^۳، ۲۰۱۰)، تعدیل دشواری فرم‌های مختلف آزمون و قابل مقایسه کردن نمره‌های آنها، یک ضرورت در اندازه‌گیری به شمار می‌آید (منگ^۴، ۲۰۱۲؛ ژانگ^۵، ۲۰۱۲؛ چن^۶، ۲۰۱۴؛ وایبرگ^۷، ۲۰۱۸).

با روی کار آمدن آزمون‌های با مقیاس بزرگ در دهه‌های ۱۹۳۰ تا ۱۹۵۰، به جهت نیاز داشتن این آزمون‌ها به فرم‌های جدید برای تأمین امنیت آنها، همترازسازی به الزامی ضروری تبدیل شد و به‌عنوان معیار پاسخگویی سازمان‌ها برای برنامه‌های آزمون‌سازی مورد توجه قرار گرفت. بعدها مطرح شدن نظریه‌های مدرن روان‌سنجی و آماری همچون نظریه سؤال پاسخ (IRT^۸) روش‌های جدیدی برای همتراز کردن داده‌ها را به وجود آورد (وون داویر، هالند و ثایر^۹، ۲۰۰۳). تاکنون روش‌های متعددی هم در حوزه نظریه کلاسیک اندازه‌گیری و هم در حوزه نظریه سؤال پاسخ برای همترازسازی فرم‌های مختلف آزمون مورد توجه قرار گرفته است (منگ، ۲۰۱۲). در حوزه نظریه کلاسیک، در همترازسازی میانگین، میانگین نمره‌ها در دو فرم X و Y همتراز می‌شوند. اما در عمل، نمره‌های فرم آزمون‌ها نه تنها در میانگین که در انحراف

1. Dorans & Holland

2. González & Wiberg

3. Von Davier

4. Meng

5. Zhang

6. Chen

7. Wiberg

8. Item response theory

9. Von Davier, Holland & Thayer

استاندارد نیز متفاوت هستند، بنابراین روش همترازسازی خطی برای بهبود همترازسازی میانگین مورد توجه قرار گرفته است (ونگ، ژانگ و یو، ۲۰۲۰). همترازسازی همصدک رایج‌ترین نوع همترازسازی نظریه کلاسیک اندازه‌گیری است که در آن توزیع نمره‌ها در دو فرم همتراز شده یکسان است (سنسیوری، وایبرگ و ماتیوکی^۲، ۲۰۱۷). روش همترازسازی کرنل رویکرد منحصر به فردی به همترازسازی آزمون است و دارای پنج مرحله پیش‌هموارسازی، برآورد احتمال نمره‌ها، پیوسته‌سازی^۳، همترازسازی و خطای استاندارد همترازسازی است (آکین آریکان^۴، ۲۰۱۹). در نظریه سؤال پاسخ نیز همترازسازی نمره واقعی در رابطه بین نمره واقعی و برآورد توانایی در دو فرم آزمون از راه تابع منحنی ویژگی‌های آزمون^۵ برقرار می‌شود (بروسمان^۶، ۲۰۱۰) و ماهیت همترازسازی نمره مشاهده شده نظریه سؤال پاسخ همان همترازسازی همصدک است و الگوریتم توزیع دو جمله‌ای را در برمی‌گیرد (لرد و وینگرسکی^۷، ۱۹۸۴: به نقل از ژانگ، ۲۰۱۲).

موفقیت همترازسازی نظریه سؤال پاسخ به مفروضات آماری آن همچون تک‌بعدی بودن، استقلال موضعی و غیرسرعتی بودن آزمون بستگی دارد (ریکر^۸، ۲۰۰۷). اگر مفروضات مدل که مهم‌ترین آنها تک‌بعدی بودن صفت مکنون است، نقض شود، تبدیل نمره‌ها از فرمی به فرم دیگر مستقل از آزمودنی‌هایی که آزمون روی آنها اجرا شده است، نخواهد بود، از این رو ویژگی نامتغیر بودن آزمودنی‌ها در این شرایط نقض شده و تابع همترازسازی به دست آمده از آزمون‌ها برای زیر گروه‌های آزمودنی‌ها متفاوت خواهد بود (چامپلین^۹، ۱۹۹۶). از سوی دیگر، طبق نظر لرد^{۱۰} (۱۹۸۰) اگر فرم‌های آزمون بعد یکسانی را اندازه‌گیری نکنند، همترازسازی امکان‌پذیر نیست، بنابراین تک‌بعدی بودن در سطح نمره‌ها، پیش‌نیاز همترازسازی است. حال این پرسش

1. Wang, Zhang & You

2. Sansivieri, Wiberg & Matteucci

3. continuization

4. Akin Arikan

5. Test Characteristic Curve (TCC)

6. Brossman

7. Lord & Wingersky

8. Ricker

9. Champlain

10. Lord

پیش می‌آید که اگر مفروضه تک‌بعدی بودن نقض شود، چه اثری بر همترازسازی داده‌ها خواهد داشت و برای همترازسازی بهینه نمره‌ها در این شرایط، چه باید کرد؟ نقض تک‌بعدی بودن آزمون، مفروضه استقلال موضعی را به خطر انداخته و وجود وابستگی موضعی به چالش‌های جدی در همترازسازی منجر خواهد شد؛ وابستگی موضعی سؤال درون مجموعه آزمون بر دقت برآورد بیشینه درست‌نمایی اثر می‌گذارد و برآورد نادقیقی از پارامترهای سؤال ارائه خواهد داد. دقت نکردن در برآورد منحنی ویژگی‌های سؤال^۱ (ICC) و پارامترهای سؤال، فرایند تبدیل مقیاس را با خطا مواجه می‌کند و در نهایت بر برآورد منحنی ویژگی‌های آزمون (TCC) - که در همترازسازی نمره واقعی استفاده می‌شود- تأثیرگذار است (ریکر، ۲۰۰۷؛ چن، ۲۰۱۴).

با عنایت به اینکه سوگیری و خطا در نتایج ارزیابی‌های تربیتی و روان‌شناختی تهدیدی جدی به شمار می‌رود، مطلوب است که در همترازسازی اطمینان حاصل شود که داده‌ها تک‌بعدی هستند، اما این شرط در تمام موقعیت‌های آزمون، به‌ویژه هنگامی که با داده‌های واقعی سروکار داریم، امکان‌پذیر نیست و عوامل بسیاری تک‌بعدی بودن آزمون را نقض می‌کند، ضمن آنکه، بسیاری از آزمون‌های تربیتی و روانشناسی ماهیتاً چندبعدی هستند. در چنین شرایطی، چنانچه از روش‌های تک‌بعدی جهت همترازسازی داده‌ها استفاده شود، گزارش نادقیق و نادرستی به دست خواهد داد (لی، جیانو و لیسیتز^۲، ۲۰۱۲؛ کاهرامان^۳، ۲۰۱۳).

مدل‌های نظریه سؤال پاسخ تا حدودی در برابر نقض مفروضه تک‌بعدی مقاوم هستند و می‌توان اندک انحراف از این مفروضه را در آزمون‌ها با اغماض پذیرفت و روش‌های تک‌بعدی را برای آنها به کار برد، اما چنانچه این مفروضه به میزان زیادی نقض شود، استفاده از روش‌های تک‌بعدی به نتایج ضعیف و ناکارآمد منجر خواهد شد. بنابراین در شرایطی که تعداد عوامل تشکیل‌دهنده آزمون بیش از یک عامل است، باید از روش‌هایی استفاده شود که قابلیت مقابله با خطاهای ناشی از چنین اندازه‌هایی را داشته باشد. چندین روش برای مقابله با نقض تک‌بعدی بودن و وابستگی موضعی در مجموعه آزمون‌ها وجود دارد، یکی از این روش‌ها، استفاده از مدل‌های نظریه

1. Item characteristic curve

2. Li, Jiao & Lissitz

3. Kahraman

سؤال پاسخ چندبعدي است که وابستگی موضعی سؤال را در نظر می‌گیرند (چن، ۲۰۱۴). مدل‌های نظریه سؤال پاسخ چندبعدي^۱ (MIRT) برای مدل‌سازی روابط ابعاد چندگانه توانایی شخص و پارامترهای چندگانه سؤال‌ها به وجود آمده‌اند و احتمال پاسخ درست به سؤال‌های چندبعدي مبتنی بر ترکیب مهارت‌ها در توانایی‌های چندگانه است (ژانگ، ۲۰۱۲). به عبارت دیگر این مدل‌ها نشان‌دهنده تابع ریاضیاتی برای رابطه بردار ویژگی‌های چندگانه شخص و بردار ویژگی‌های سؤال‌ها است (ریکاس^۲، ۲۰۰۹).

مطالعه پیشینه نشان داد که مدل‌های چندبعدي نظریه سؤال پاسخ (MIRT) برای داده‌های شبیه‌سازی شده و داده‌های واقعی، دقت بیشتری نسبت به داده‌های تک‌بعدي دارند (لی و لیسیتز، ۲۰۰۰؛ بروسمان، ۲۰۱۰) و روایی کاربرد مدل‌های تک‌بعدي در همترازسازی، برآزش مدل، برآورد پارامتر، نمره‌گذاری و سنجش انطباقی در موقعیتی که فضای چندبعدي معقول و منطقی است، مورد تردید واقع شده است (شو و ایز، ۲۰۱۵). با این وجود، ژانگ (۲۰۱۲) نشان داد که روش همترازسازی نمره واقعی چندبعدي تک‌بعدي شده در مقایسه با سایر روش‌های همترازسازی در همه شرایط توزیع آزمودنی‌ها و همه روش‌های همترازسازی عملکرد بهتری داشته و روش همترازسازی چندبعدي کامل بدترین عملکرد را بر اساس سوگیری و خطای همترازسازی به همراه داشت. سایر مطالعات نشان‌دهنده آن است که همترازسازی نمره واقعی چندبعدي ساختار ساده SMT^۳ در مقایسه با روش‌های معمول تک‌بعدي دقت همترازسازی بیشتری فراهم می‌کند و مفید بودن رویکرد چندبعدي در همترازسازی داده‌های چندبعدي نشان داده شده است (کیم^۴، ۲۰۱۸؛ پترسون^۵، ۲۰۱۴). در شرایط نقض مفروضه تک‌بعدي، روش همترازسازی نظریه سؤال پاسخ چندبعدي نسبت به روش تک‌بعدي بیشتر شبیه همترازسازی همصدک عمل می‌کند و روش‌های همترازسازی چندبعدي جایگزین مناسبی برای روش‌های تک‌بعدي هستند (بروسمان، ۲۰۱۰)، همچنین می‌توان از مدل‌های همترازسازی چندبعدي برای داده‌های تک‌بعدي

1. Multidimensional Item response theory

2. Reckase

3. SS-MIRT true-score equating procedure

4. Kim

5. Peterson

نیز به‌خوبی استفاده کرد و خطای کمتری دارند (لی، ۱۹۹۷؛ بیگوین و هنسون^۱، ۲۰۰۱). اما در پژوهش لی (۲۰۱۳) و لی، لی و برینان (۲۰۱۴) حتی زمانی که آزمون‌های چندبعدی استفاده شد، روش‌های همترازسازی IRT تک‌بعدی نتایج دقیق‌تری ارائه داد، روش همترازسازی همصدک نیز نسبت به روش‌های IRT خطای بیشتری داشته و روش همترازسازی چندبعدی کامل زمانی که همبستگی بین ابعاد پایین بود، از دقت بیشتری نسبت به سایر روش‌ها برخوردار بود.

بررسی مطالعات انجام گرفته، نشان داد که با وجود اهمیت انکارناپذیر همترازسازی آزمون‌های چندبعدی و گستردگی قلمرو این موضوع، پژوهش‌های اندکی در این زمینه اجرا شده و پژوهشگران عمدتاً روش‌های همترازسازی نظریه‌های کلاسیک و نظریه IRT تک‌بعدی را مقایسه کرده‌اند (مقدم‌زاده، ۱۳۹۲؛ رضوانی‌فر، ۱۳۹۱؛ شاطریان محمدی، ۱۳۸۱؛ واشقانی فراهانی، ۱۳۸۰) و از سوی دیگر، جمع‌بندی مطالعات انجام گرفته در زمینه همترازسازی آزمون‌های چندبعدی، نشان‌دهنده این است که بیشتر این پژوهش‌ها با استفاده از داده‌های شبیه‌سازی اجرا شده و نتایج این پژوهش‌ها سمت و سوی واحدی را در این زمینه مشخص نکرده‌اند. بنابراین با توجه به ماهیت چندبعدی داده‌ها در روان‌شناسی و علوم تربیتی لازم است روش و رویکرد مناسب برای همترازسازی این ابعاد مشخص شود. همچنین همترازکردن فرم‌های مختلف آزمون‌های چندبعدی در مقایسه با آزمون‌های تک‌بعدی با پیچیدگی‌هایی مواجه خواهد بود که این پیچیدگی‌ها و مشکلات تاکنون به‌طور کامل بررسی نشده است. از سال ۲۰۱۰ بروسمان به این روش‌ها توجه دقیق و جزئی داشت. وی در رساله دکتری خود، سه روش همترازسازی چندبعدی شامل همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده^۲، روش همترازسازی نمره واقعی چندبعدی تک‌بعدی شده^۳ و روش همترازسازی نمره مشاهده شده چندبعدی کامل^۴ را مقایسه کرد. بنابراین استفاده از روش‌های چندبعدی همترازسازی در سطح جهان نو بوده و پس از بروسمان، پژوهشگران اندکی همچون لی^۵ (۲۰۱۳) و ژانگ

1. Béguin & Hanson

2. unidimensionalized MIRT observed score equating (UMIRTO)

3. unidimensionalized MIRT true score equating (UMIRTT)

4. Full MIRT observed score equating (FMIRTO)

5. Lee

(۲۰۱۲) آن را در داده‌های مختلف بررسی کرده‌اند. با توجه به نو بودن مدل‌های چندبعدي و لزوم اجرای پژوهش‌هایی با دغدغه روش‌شناختی، لازم است در آزمون‌های سرنوشت‌ساز در ایران، وضعیت این آزمون‌ها از لحاظ رعایت مفروضات و همچنین همترازسازی آنها با روش‌های معتبرتر مدنظر قرار گیرد. هدف از اجرای این پژوهش، همترازسازی آزمون ریاضی کنکور سراسری با رویکردهای نوین همترازسازی چندبعدي بود تا مشخص شود بهترین شیوه برای همترازسازی چنین داده‌هایی کدام است. بنابراین از میان تمام روش‌های موجود، ترکیبی از روش‌های تک‌بعدي و روش‌های نوین چندبعدي مدنظر قرار گرفت. در این پژوهش، روش همصداک به دلیل اینکه به روشنی مفروضات تک‌بعدي بودن را نقض نمی‌کند (بروسمان، ۲۰۱۰)، مبنای مقایسه سایر روش‌های همترازسازی IRT قرار گرفته است، بر این اساس، به منظور تعیین یک ملاک از این روش بهره‌برداری شده است.

نظریه سؤال پاسخ چندبعدي

مدل‌های سنتی نظریه سؤال پاسخ تک‌بعدي مبتنی بر این مفروضه است که فقط یک بعد توانایی تعیین‌کننده عملکرد آزمودنی‌ها در آزمون است. با این حال، این مفروضه در موقعیت‌های عملی ممکن است نقض شود؛ چراکه مکانیزم و فرایند شناختی که ارزیابی‌شونده برای پاسخ به سؤال‌های آزمون استفاده می‌کند، ساده نیست و پژوهشگران روان‌شناسی و تعلیم‌وتربیت بر این باور هستند که صفات و توانایی‌های چندبعدي و پیچیده در عملکرد آزمون نقش دارند (مین و کیم، ۲۰۰۳؛ آکرمان، گیرل و والکر، ۲۰۰۳). همچنین آنها علاقه‌مند به توضیح فرایندها و بررسی توانایی‌های شناختی با مدل‌های نظریه سؤال پاسخ هستند که متغیرهای غیر قابل مشاهده چندگانه (مانند استعداد، توانایی، کارایی و ...) را اندازه بگیرد. امروزه استفاده از مدل‌های نظریه سؤال پاسخ چندبعدي راه حل مناسبی را فراهم کرده است (ریکاس، ۲۰۰۹؛ واتزل و هل، ۲۰۱۴؛ کیم، ۲۰۱۸؛ لی و کولن، ۲۰۲۰).

مدل‌های نظریه سؤال پاسخ چندبعدي، یک بیان ریاضیاتی برای نشان دادن رابطه جایگاه افراد در فضای چندبعدي و احتمال پاسخ آنها به سؤال‌های آزمون است (لی،

1. Min & Kim

2. Ackerman, Gierl & Walker

3. Kim, Lee & Kolen

۲۰۱۳؛ آکرمان، گیرل و والکر، ۲۰۰۳). به عبارت دیگر، این مدل‌ها یک مورد خاص از نظریه سؤال پاسخ هستند که تابع ریاضیاتی شامل پارامتر توانایی به‌عنوان محور ویژگی‌های چندگانه شخص - شامل مهارت و دانشی که شخص برای آزمون دارد - می‌شود و همچنین دربرگیرنده بردار ویژگی‌های سؤال است که دشواری و همچنین حساسیت سؤال‌های آزمون به تفاوت در ویژگی‌های شخص است را نشان می‌دهد (ریکاس، ۲۰۰۹).

نظریه سؤال پاسخ چندبعدی در فهم بهتر آنچه یک سؤال اندازه‌گیری می‌کند، سطح توانایی ارزیابی‌شونده در هر ویژگی، میزان دقت ارزیابی ترکیب‌های مختلف توانایی، فهم ساختار داده‌ها و بهبود طبقه‌بندی سؤال مفید است (ساسکوان، جانپنج، انگوگراتوکی و گویاجارن پانیشک، ۲۰۱۲). این نظریه مدل رشدیافته نظریه سؤال پاسخ تک‌بعدی و تحلیل عاملی است (ریکاس، ۲۰۰۹). مدل نظریه سؤال پاسخ چندبعدی برای مدل دو پارامتری لوجستیک به شرح زیر نشان داده می‌شود:

$$P(U_{ij} = 1 | \theta_j, a_i, d_i) = \frac{e^{a_i \theta_j + d_i}}{1 + e^{a_i \theta_j + d_i}}$$

که U_{ij} اشاره به پاسخ شخص j به سؤال i ، بردار a, d بردار $1 \times m$ برای پارامترهای تمیز، θ بردار $1 \times m$ پارامترهای توانایی است. در اینجا m به تعداد ابعاد در فضای مختصات اشاره دارد و d پارامتر عرض از مبدأ است. این مدل یک مدل چندبعدی نظریه سؤال پاسخ جبرانی است، زیرا بالا بودن توانایی در یک بعد می‌تواند پایین بودن توانایی در بعد دیگر را جبران کند (لی، ۲۰۱۳). در این فرمول پارامتر d ، پارامتر دشواری مانند آنچه در مدل‌های تک بعدی است، نبوده و مقدار منفی d تقسیم بر بردار پارامتر تمیز $(\frac{-d}{a_1})$ ، دشواری مرتبط با سؤال را در بعد مورد نظر می‌دهد، بنابراین دشواری (b) نظریه سؤال پاسخ تک‌بعدی در نظریه سؤال پاسخ چندبعدی با MDIFF نشان داده می‌شود که برابر با $MDIFF = \frac{-d}{a}$ است (ریکاس، ۲۰۰۹).

همترازسازی نمره مشاهده شده چندبعدی کامل

تاکنون مطالعات اندکی به بررسی همترازسازی تحت مدل‌های چندبعدی پرداخته‌اند (بروسمان، ۲۰۱۰؛ ژانگ، ۲۰۱۲؛ لی، ۲۰۱۳؛ لیم، ۲۰۱۶؛ لیم و لی، ۲۰۱۶ و کیم، ۲۰۱۸). نخستین کسی که در این زمینه تلاش کرد بروسمان (۲۰۱۰) بود که به همترازسازی چندبعدی آزمون اقدام کرد. همچنین لی و بروسمان (۲۰۱۲) همترازسازی چندبعدی نمره مشاهده شده در آزمون‌های با فرمت ترکیبی را ارائه کردند. همترازسازی نمره مشاهده شده چندبعدی، گسترشی از همترازسازی نمره مشاهده شده تک‌بعدی است و با استفاده از سه فرایند به شرح زیر قابل محاسبه است. برای انجام همترازسازی چندبعدی کامل، ابتدا توزیع شرطی نمره مشاهده شده (مانند $f(x|\theta_j)$) در هر سطح ترکیب توانایی (θ_j) با استفاده از یک فرمول بازگشتی^۱ مانند الگوریتم لرد-وینگرسکی^۲ به دست می‌آید (لی، ۲۰۱۳). در چارچوب نظریه سؤال پاسخ تک‌بعدی، توزیع شرطی نمره مشاهده شده (مانند $f(x|\theta)$) ابتدا در هر سطح توانایی (θ_j) ایجاد می‌شود. نظریه سؤال پاسخ چندبعدی توزیع شرطی را برای همه ترکیب‌های سطوح توانایی (مانند $f(x|\theta)$) به کار می‌برد که θ اشاره به سطح توانایی کلی دارد (کولن و ونگ، ۲۰۰۷؛ به نقل از بروسمان، ۲۰۱۰). در اینجا یک بردار سطح توانایی به جای یک سطح توانایی استفاده می‌شود (بروسمان، ۲۰۱۰) توزیع شرطی نمره‌ها به صورت زیر محاسبه می‌شود:

$$\begin{aligned} f_g(x|\theta_j) &= f_{g-1}(x|\theta_j)(1 - P_{jg}), x = 0 \\ &= f_{g-1}(x|\theta_j)(1 - P_{jg}) + f_{g-1}(x - 1|\theta_j)P_{jg}, 0 < x < g \\ &= f_{g-1}(x - 1|\theta_j)P_{jg}, x = g \end{aligned}$$

در معادله بالا $f_g(x|\theta_j)$ توزیع شرطی نمره مشاهده شده در اولین سؤال g است. $f_{g-1}(x|\theta_j)$ توزیع شرطی نمره مشاهده شده اولین سؤال $g-1$ است و P_{jg}

1. recursion

2. Lord-Wingersky

احتمال پاسخ درست برای سؤال g برای آزمودنی با بردار توانایی θ_j است. در عمل این الگوریتم به صورت زیر محاسبه می‌شود:

$$f_g(x | \theta_j) = (1 - P_{jg}) \quad , x = 0$$

$$= P_{jg} \quad , x = 1$$

پس از این گام، $n-1$ تکرار برای محاسبه توزیع شرطی نمره مشاهده شده $f_g(x | \theta_j)$ برای n سؤال لازم است.

در نظریه سؤال پاسخ تک‌بعدی این توزیع‌های شرطی در چگالی توانایی $(\Psi(\theta))$ ضرب شده و در همه سطوح توانایی با هم جمع می‌شود تا توزیع حاشیه‌ای نمره مشاهده شده برای هر فرم به دست آید. در نظریه سؤال پاسخ چندبعدی توزیع‌های شرطی در چگالی توانایی چند متغیری $(\Psi(\theta))$ ضرب شده و در همه فضای توانایی با هم جمع می‌شوند:

$$f(x) = \sum_1 \sum_2 \dots \sum_m f(x|\theta)\Psi(\theta)$$

یا

$$f(x) = \int_1 \int_2 \dots \int_m f(x|\theta)\Psi(\theta)d\theta$$

در این معادله m تعداد ابعاد است. بسته به اینکه توانایی دارای یک توزیع پیوسته یا توزیع گسسته باشد، یکی از دو فرمول بالا استفاده می‌شود. در نهایت زمانی که توزیع حاشیه‌ای نمره‌های مشاهده‌شده برای فرم X و Y تعیین شد، مشابه روش همترازسازی تک‌بعدی، روش همصدک سستی برای همترازسازی دو فرم استفاده می‌شود (لی، ۲۰۱۳؛ بروسمان، ۲۰۱۰).

همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده و همترازسازی نمره واقعی چندبعدی تک‌بعدی شده

بروسمان (۲۰۱۰) برای اجرای فرایند همترازسازی چندبعدی تک‌بعدی شده، پارامترهای سؤال تک‌بعدی و توزیع توانایی تک‌بعدی را از داده‌های چندبعدی برآورد

کرد. وی محاسبه پارامترهای تک‌بعدی را با تعیین بهترین جهت اندازه‌گیری آزمون برآورد کرد. منظور از بهترین جهت اندازه‌گیری در آزمون، جهتی است که تابع آگاهی نمره‌های چندبعدی را در تمامی جهت‌های ممکن بیشینه می‌کند. بروسمان (۲۰۱۰) و لی و بروسمان (۲۰۱۲) ایده بهترین جهت اندازه‌گیری را از مطالعات ژانگ (۱۹۹۶)، ژانگ و استوت^۱ (۱۹۹۹) و ژانگ^۲ و ونگ (۱۹۹۸) گرفتند که مطالعات آنها مبتنی بر نظریه سؤال پاسخ چندبعدی و توزیع نرمال چندمتغیری بود (بروسمان، ۲۰۱۰). ریکاس (۲۰۰۹) تابع آگاهی نمره‌های چندبعدی را به‌عنوان مجذور ریشه تابع آگاهی سؤال چندبعدی تعریف کرده است.

$$S_{\gamma}(\theta) = \sqrt{\frac{[\nabla_{\gamma} P(\theta)]^2}{P(\theta)Q(\theta)}}$$

در فرمول فوق، γ بردار زاویه‌ها با محور مختصات است که جهت به دست آمده از نقطه θ را مشخص می‌کند، ∇_{γ} مشتق جهت یا شیب در جهت γ است. $P(\theta)$ و $Q(\theta)$ نیز به ترتیب نشان‌دهنده احتمال پاسخ درست و احتمال پاسخ نادرست هستند.

بهترین جهت اندازه‌گیری با محاسبه میانگین تابع آگاهی نمره‌های چندبعدی در بین همه جهت‌ها به دست می‌آید و بردار استاندارد از وزن‌های متناسب با هر بعد به دست می‌دهد. با این توصیفات، بهترین جهت اندازه‌گیری داده‌ها، یک مقیاس تک‌بعدی است که با ترکیب خطی صفات ارائه می‌شود (لی، ۲۰۱۳).

همانند همترازسازی چندبعدی کامل برای همترازسازی چندبعدی تک‌بعدی شده سه گام مهم باید طی شود. در مرحله اول، باید پارامترهای چندبعدی سؤال محاسبه شود، سپس باید در هر دو فرم X و Y بهترین جهت اندازه‌گیری داده‌ها مشخص شود که این کار با تعیین پارامتر شیب چندبعدی سؤال‌ها (α) و با فرمول زیر انجام می‌گیرد:

1. Stout

2. Wang

$$W_k = \frac{\sum_{i=1}^N \alpha_{ik}}{\sqrt{\sum_{k=1}^m (\sum_{i=1}^n \alpha_{ik})^2}}$$

در این فرمول، W_k ضریب استاندارد در بردار ضرایب ترکیب خطی است، α_{ik} پارامتر تمیز چندبعدهی برآورد شده برای سؤال i در بعد k است، N تعداد کل سؤال‌های آزمون است و m تعداد کل ابعاد آزمون است. در مرحله سوم، پارامترهای تکبعدهی سؤال به دست می‌آید، برای برآورد تمیز تکبعدهی از فرمول زیر استفاده می‌شود:

$$\alpha_{iw} = (1 + \sigma_{iw}^2)^{\frac{1}{2}} \alpha_i^T \Sigma w$$

در فرمول فوق، Σ ماتریس کوواریانس توانایی چندبعدهی برآورد شده است، σ_{iw}^2 ماتریس کوواریانس شرطی $\alpha_i^T \theta$ ترکیب بردار مکنون کامل θ_w است ($\sigma_w^2 = (\alpha_i^T \Sigma \alpha_i - (\alpha_i^T \Sigma w)^2)$). همچنین جهت برآورد پارامتر دشواری تکبعدهی از فرمول زیر استفاده می‌شود:

$$b_{iw} = \frac{-d_{iw}}{\alpha_{iw}}$$

در فرمول فوق، d_{iw} پارامتر دشواری چندبعدهی است ($d_{iw} = (1 + \sigma_{iw}^2)^{\frac{1}{2}}$). در پایان این مراحل، پارامترهای سؤال تکبعدهی با استفاده از پارامترهای سؤال چندبعدهی درجه‌بندی شده و وزن‌های مرتبط با بهترین جهت اندازه‌گیری آزمون محاسبه می‌شوند. با در دست داشتن پارامترهای تکبعدهی شده سؤال و توانایی می‌توان هر دو نوع همترازسازی نمره واقعی و نمره مشاهده شده را مانند آنچه در همترازسازی تکبعدهی انجام می‌گیرد، برآورد کرد (لی، ۲۰۱۳).

روش پژوهش

با توجه به هدف و ماهیت مسئله پژوهش، روش‌های مختلف همترازسازی در داده‌های واقعی آزمون ریاضی گروه علوم ریاضی و فنی کنکور سراسری دو سال ۱۳۹۶ و ۱۳۹۷ بررسی و مقایسه شدند، بنابراین روش پژوهش توصیفی بود. جامعه

پژوهش شامل همه داوطلبان گروه علوم ریاضی و فنی کنکور سراسری سال‌های ۱۳۹۶ و ۱۳۹۷ است که تعداد آنها در سال ۱۳۹۶ برابر با ۱۴۸۴۲۹ و در سال ۱۳۹۷ برابر با ۱۴۴۴۳۷ نفر بوده است. پاسخ تعداد ۵۰۰۰ نفر از داوطلبان به شیوه تصادفی ساده از بین تمامی داده‌های مربوط به آزمون ریاضی کنکور سراسری سال ۱۳۹۶ و ۱۳۹۷ انتخاب شدند. آزمون ریاضی کنکور سراسری در هر دو سال دارای ۵۵ سؤال بود. طرح گردآوری داده‌ها طرح گروه‌های تصادفی^۱ بود. در طرح گروه‌های تصادفی، دو گروه از ارزیابی‌شوندگان مستقل از جامعه نمونه‌گیری می‌شوند و هر گروه یک فرم آزمون را پاسخ می‌دهد (گلزانز و وایبرگ، ۲۰۱۷).

در طرح گروه‌های تصادفی با مدل‌های تک‌بعدی نظریه سؤال پاسخ (UIRT) نیازی به فرایند پیونددهی^۲ و در مقیاس مشترک قراردادن نمرات نیست، زیرا دو گروه از جمعیت یکسانی استخراج شده‌اند، که این نیاز به تنظیم عرض از مبدأ و واحد مقیاس را از بین می‌برد (کیم، ۲۰۱۸). همچنین در همترازسازی نمره مشاهده شده و نمره واقعی چندبعدی نظریه سؤال پاسخ (MIRT) در طرح گروه‌های تصادفی، چنانچه هنگام درجه‌بندی برآوردهای سؤال و توانایی، راه‌حل‌ها متعامد^۳ باشد، و توزیع توانایی از توزیع نرمال استاندارد چندمتغیره با همبستگی صفر بین ابعاد پیروی کند، نیازی به روش‌های پیونددهی پیش از همترازسازی نیست (لی، لی و برینان، ۲۰۱۴). از آنجا که طرح مورد استفاده در این پژوهش طرح گروه‌های معادل است، بنابراین آزمودنی‌هایی که فرم X را گرفته‌اند با آزمودنی‌هایی که فرم Y را گرفته‌اند هم در مبدأ (ترنسلیشن^۴) و هم در واحد اندازه‌گیری (دایلشن^۵) در هر سطح تتا با هم معادل هستند. در این صورت فقط یک ماتریس چرخش برای نامعینی چرخش برای هموار کردن مقیاس لازم بود (تامپسون و همکاران^۶، ۱۹۹۷؛ به نقل از بروسمان، ۲۰۱۰). در این پژوهش، از چرخش متعامد در درجه‌بندی پارامترها استفاده شد.

داده‌ها پس از دریافت از سازمان سنجش آموزش کشور پیرایش و آماده شدند. در مرحله اول در داده‌های دریافت شده از سازمان سنجش، مواردی که بیش از ۳۰ درصد

1. Random Group

2. linking

3. Orthogonal

4. translation

5. dialation

6. Thompson et al

داده گمشده داشتند، از تحلیل حذف شد و در نمونه‌ای که برای تحلیل انتخاب شدند، داده‌های گمشده به دلیل داشتن نمره منفی در آزمون سراسری به‌مثابه سؤال‌هایی در نظر گرفته شد که آزمودنی‌ها توان پاسخ‌دهی مناسب به آنها را نداشتند و از این‌رو، این سؤال‌ها با عدد صفر جایگزین شدند. برای تعیین ابعاد آزمون از نرم‌افزار NOHARM (فراسر و مک‌دونالد، ۲۰۱۲) به دلیل مناسب بودن این نرم‌افزار در تعیین ابعاد داده‌های چندبعدي و مفید بودن آن برای فقط تحلیل داده‌های دوارزشی (اسوتینا و لوی، ۲۰۱۲) استفاده شد. همچنین برای اطمینان بیشتر در تعیین ابعاد آزمون از نرم‌افزار MPLUS8 (موشن و موشن، ۲۰۱۷-۱۹۹۷) به لحاظ استفاده از روش WLSMV برای برآورد و مناسب بودن برای داده‌های دوارزشی استفاده شد (جاسپر، ۲۰۱۰؛ به نقل از ایزانلو، بازرگان، فرزاد، صادقی و کاوسی، ۱۳۹۱). تجزیه و تحلیل داده‌ها با استفاده از نرم‌افزار R (تیم توسعه R، ۲۰۱۹) و بسته‌های ltm (ریزوپولوس، ۲۰۰۶)، mirt (چالمرز، ۲۰۱۲)، MASS (ونبلز و ریپلی، ۲۰۱۹) و mvtnorm (جنز و همکاران، ۲۰۰۸) و نرم‌افزار PIE (هنسون و زنگ، ۲۰۰۴) انجام گرفت.

برای همترازسازی داده‌ها با استفاده از روش همترازسازی تک‌بعدي نمره مشاهده شده و نمره واقعی پس از برآورد پارامترهای تک‌بعدي سؤال‌ها و توانایی در بسته ltm، پارامترها برای همترازسازی وارد نرم‌افزار PIE شدند. برای همترازسازی نمره مشاهده شده و نمره واقعی چندبعدي تک‌بعدي شده IRT، ابتدا پارامترهای چندبعدي با استفاده از بسته mirt در R به دست آمدند، سپس بهترین جهت اندازه‌گیری داده‌ها (یعنی پارامتری که در ابعاد سه‌گانه هر سؤال قوی‌تر بود و در منحنی سطح سؤال جهت بهتری را نشان می‌داد)، در R مشخص شد و پس از به دست آوردن مقادیر تنای چندبعدي تک‌بعدي شده، تمامی پارامترهای چندبعدي تک‌بعدي شده (a، b و

1. Fraser & McDonald

2. Svetina & Levy

3. Muthén & Muthén

4. Jasper

5. Rizopoulos

6. Chalmers

7. Venables & Ripley

8. Genz et al

9. Hanson & Zeng

Θ) وارد نرم‌افزار PIE شدند. در نهایت، برای همترازسازی نمره مشاهده شده چندبعدي کامل، ابتدا توزیع حاشیه‌ای نمره‌های هر دو فرم با استفاده از بسته‌های MASS و mvtnorm به دست آمدند و سپس با استفاده از بسته equate همترازسازی همصدک انجام گرفت. در این پژوهش برای همترازسازی نمره واقعی چندبعدي تک‌بعدي شده IRT، همترازسازی نمره مشاهده شده چندبعدي تک‌بعدي شده و همترازسازی نمره مشاهده شده چندبعدي کامل، کدهای R طراحی شده توسط بروسمان (۲۰۱۰) اصلاح شده و از آنها استفاده شد. این کدها در پژوهش بروسمان (۲۰۱۰) در دسترس است.

برای ارزیابی عملکرد شش روش همترازسازی، از روش همصدک به‌عنوان معیار مقایسه استفاده شد. در این پژوهش ارزیابی عملکرد روش‌های همترازسازی و تعیین خطای آن به سبب استفاده از داده‌های واقعی و نامشخص بودن روابط همترازسازی جامعه (همترازسازی واقعی) دشوار بود، بنابراین ملاک تفاوت مهم^۱ (DTM) مدنظر قرار گرفت. مقدار DTM در پژوهش‌های مختلف تفاوت ۰/۵ با معیار بهینه در نظر گرفته است (کیم، ۲۰۱۸؛ بروسمان، ۲۰۱۰). روش همترازسازی همصدک به‌عنوان ملاک مقایسه (معیار بهینه) سایر روش‌ها در نظر گرفته شد و میانگین تفاوت‌های نمره‌های همتراز شده در روش‌های مختلف با مقدار DTM برابر با تفاوت ۰/۵ از روش همصدک برای قضاوت در خصوص روش‌ها مدنظر قرار گرفت؛ یعنی اگر تفاوت میانگین بین پنج روش همترازسازی با روش همصدک کمتر از ۰/۵ باشد، آن روش تفاوت اندکی با همصدک داشته و اگر این تفاوت بیش از ۰/۵ باشد به‌عنوان تفاوت زیاد و قابل توجه در نظر گرفته می‌شود.

یافته‌ها

شاخص‌های توصیفی نمره‌های درس ریاضی در دو گروه داوطلبان طی سال‌های ۱۳۹۶ و ۱۳۹۷ در جدول (۱) آمده است.

^۱. Differences That Matter (DTM)

جدول (۱) شاخص‌های توصیفی نمره‌های آزمودنی‌ها در دو فرم آزمون ریاضی

دوطلبان	میانگین	انحراف استاندارد	کمینه	بیشینه	کجی		کشیدگی	
					خطا	آماره	خطا	آماره
سال ۱۳۹۷	۷/۸۲	۷/۳۱	۱	۵۱	۱/۹۹	۰/۰۳	۴/۶۶	۰/۰۶
سال ۱۳۹۶	۷/۶۶	۷/۶۳	۱	۵۲	۱/۹۰	۰/۰۳	۴/۰۲	۰/۰۶

همان‌طور که در جدول (۱) دیده می‌شود، میانگین و انحراف استاندارد نمره‌های درس ریاضی در سال ۱۳۹۷ برابر با ۷/۸۲ و ۷/۳۱ و در سال ۱۳۹۶ برابر با ۷/۶۶ و ۷/۶۳ است. بنابراین بین نمره‌های دو سال تفاوت بسیار اندکی از لحاظ میانگین و انحراف استاندارد وجود دارد.

ابعاد آزمون‌ها

با توجه به هدف پژوهش که همترازسازی آزمون‌های چندبعدی است، ابتدا لازم بود مفروضه چندبعدی بودن آزمون تأیید شود. برای تعیین ابعاد آزمون از نرم‌افزار NOHARM (فراسر و مک‌دونالد، ۲۰۱۲) به دلیل مناسب بودن این نرم‌افزار در تعیین ابعاد داده‌های چندبعدی و مفید بودن آن برای فقط تحلیل داده‌های دوارزشی (اسوتینا و لوی، ۲۰۱۲) استفاده شد. نتایج مربوط به تعیین بعدیت در نرم‌افزار NOHARM نشان داد که با افزایش تعداد ابعاد میزان برازش مدل با داده‌ها مناسب‌تر است و فرض تک‌بعدی بودن داده‌ها رد می‌شود.

در نرم‌افزار NOHARM در مدل سه‌بعدی مقدار مجموع مجزورات باقی‌مانده‌ها برابر با ۰/۰۰۹۰ و ریشه دوم میانگین مجزورات باقی‌مانده‌ها ($RMSR^1$) برابر با ۰/۰۰۲۶ است که هر چه مقدار آن کوچک‌تر باشد، نشان‌دهنده برازش بهتر مدل است، همچنین می‌توان ملاک $RMSR$ را با چهار برابر ریشه دوم معکوس حجم نمونه مقایسه کرد که در این پژوهش مقدار $RMSR$ از ملاک ذکر شده کمتر است، شاخص نیکویی برازش تاناکا^۲ برابر با ۰/۹۹۱ و بیشتر از ۰/۹۵ است که نشان‌دهنده برازش مناسب مدل سه‌بعدی با داده‌ها است (هوپر، کافلان و مولن^۳، ۲۰۰۸). با

1. Root mean square of residuals

2. Tanaka index of goodness of fit

3. Hooper, Coughlan & Mullen

افزایش بعد چهارم میزان شاخص نیکویی برازش تاناکا به ۰/۹۹۵ افزایش می‌یابد از آنجا که افزایش ۰/۰۰۴ مقدار ناچیزی است و با توجه به تطبیق منطقی (با استفاده از نظر ۳ متخصص در حوزه ریاضی) عامل‌ها با سؤال‌های آزمون مدل سه‌عاملی پذیرفته شد.

همچنین برای اطمینان از مناسب بودن روش تعیین ابعاد، از روش تحلیل عاملی اکتشافی در نرم‌افزار **mplus** (موشن و موشن، ۱۹۹۷-۲۰۱۷) به لحاظ استفاده از روش **WLSMV** برای برآورد و مناسب بودن برای داده‌های دوارزشی استفاده شد (جاسپر، ۲۰۱۰؛ به نقل از ایزانلو و همکاران، ۱۳۹۱). شاخص‌های برازش در نرم‌افزار **Mplus** و بررسی منطقی سؤال‌های آزمون ریاضی نشان داد که می‌توان تعداد سه عامل را به‌عنوان عامل‌های این آزمون در نظر گرفت. شاخص‌های برازش برای مدل‌های یک، دو و سه‌عاملی در نرم‌افزار **Mplus** در جدول (۲) آمده است. همان‌طور که دیده می‌شود مدل سه‌عاملی برازش مناسب‌تری نسبت به مدل‌های یک و دوعاملی نشان داده است.

جدول (۲) شاخص‌های برازش مدل یک، دو و سه‌عاملی با داده‌های آزمون ریاضی سال ۱۳۹۷

تعداد ابعاد	SRMR	RMSEA	TLI	CFI	df	χ^2	χ^2/df
یک‌بعدي	۰/۰۸۳	۰/۰۳۵	۰/۸۶	۰/۸۷	۱۲۷۴	۹۲۵۶/۶۴	۷/۲۶
دو‌بعدي	۰/۰۶۰	۰/۰۲۶	۰/۹۲	۰/۹۳	۱۲۲۳	۵۳۵۴/۲۰	۴/۳۷
سه‌بعدي	۰/۰۳۹	۰/۰۱۸	۰/۹۶	۰/۹۷	۱۱۷۳	۳۰۵۹/۳۶	۲/۶۰

جدول (۲) نشان می‌دهد که شاخص‌های برازش سه‌بعدي از مقادیر مناسبی برخوردار هستند، مقدار χ^2/df از ۳ کمتر و مقادیر **CFI** و **TLI** بالای ۰/۹۰ هستند، همچنین مقدار **RMSE** از ۰/۰۸ کمتر است (کلاین، ۲۰۱۵). از این‌رو می‌توان گفت که شاخص‌های مدل سه‌بعدي از برازش مناسبی برخوردار هستند و آزمون کنکور ریاضی سال ۱۳۹۷ سه بعد را مورد سنجش قرار داده است.

در داده‌های مربوط به سال ۱۳۹۷ سه سؤال ۸، ۱۳ و ۴۱ به دلیل نداشتن بار عاملی مناسب حذف شدند. بارهای عاملی مربوط به سه بعد در جدول (۳) آمده است.

جدول (۳) مقادیر بار عاملی مربوط به سه بعد آزمون ریاضی سال ۱۳۹۷

سؤال	عامل اول	عامل دوم	عامل سوم	سؤال	عامل اول	عامل دوم	عامل سوم
۱	۰/۵۳۸	۰/۳۶۶	۰/۳۳۳	۲۹	۰/۵۹۰	۰/۷۶۸	۰/۲۷۷
۲	۰/۲۱۸	۰/۵۸۵	-۰/۱۹۴	۳۰	۰/۲۵۷	۰/۵۷۵	۰/۰۵۴
۳	۰/۷۲۲	۰/۱۲۳	۰/۳۵۵	۳۱	۰/۱۹۵	۰/۶۲۵	-۰/۰۷۲
۴	۰/۶۹۹	۰/۴۴۱	۰/۳۸۶	۳۲	۰/۵۰۹	۰/۵۲۶	۰/۳۸۰
۵	۰/۶۲۴	۰/۲۸۷	۰/۲۹۱	۳۳	۰/۳۱۴	۰/۴۸۰	۰/۴۶۰
۶	۰/۶۳۹	۰/۴۴۲	۰/۳۳۸	۳۴	۰/۷۰۱	۰/۴۱۰	۰/۳۸۲
۷	۰/۵۲۰	۰/۴۶۱	۰/۲۳۶	۳۵	۰/۶۹۹	۰/۳۰۸	۰/۳۹۵
۹	۰/۴۶۲	۰/۴۳۶	۰/۰۳۰	۳۶	۰/۳۲۵	۰/۴۳۷	۰/۳۲۵
۱۰	۰/۷۱۰	۰/۲۷۷	۰/۲۷۴	۳۷	۰/۶۵۴	۰/۴۸۷	۰/۲۱۸
۱۱	۰/۵۴۱	۰/۴۰۰	۰/۰۴۸	۳۸	۰/۷۱۱	۰/۵۳۶	۰/۳۱۷
۱۲	۰/۷۹۲	۰/۲۳۷	۰/۲۲۰	۳۹	۰/۶۱۲	۰/۳۱۰	۰/۳۹۴
۱۴	۰/۶۷۹	۰/۶۸۰	۰/۲۵۵	۴۰	۰/۶۶۳	۰/۵۱۳	۰/۳۹۰
۱۵	۰/۵۵۱	۰/۱۴۳	۰/۲۲۵	۴۲	۰/۱۵۳	۰/۴۱۵	-۰/۰۵۰
۱۶	۰/۸۳۵	۰/۱۶۰	۰/۲۷۵	۴۳	-۰/۰۴۱	۰/۰۱۳	۰/۵۵۸
۱۷	۰/۳۱۶	۰/۴۲۰	-۰/۰۳۳	۴۴	۰/۴۸۰	۰/۲۲۱	۰/۶۱۹
۱۸	۰/۶۷۰	۰/۴۶۸	۰/۲۵۰	۴۵	۰/۳۵۲	۰/۲۵۴	۰/۵۲۵
۱۹	۰/۷۷۶	۰/۳۴۰	۰/۲۱۲	۴۶	۰/۴۸۸	۰/۵۴۸	۰/۴۹۸
۲۰	۰/۶۳۱	۰/۳۰۹	۰/۰۹۱	۴۷	۰/۳۰۱	۰/۳۰۸	۰/۵۰۳
۲۱	۰/۲۵۹	۰/۳۰۹	۰/۰۸۱	۴۸	۰/۵۱۵	۰/۴۶۶	۰/۴۲۲

سؤال	عامل اول	عامل دوم	عامل سوم	سؤال	عامل اول	عامل دوم	عامل سوم
۲۲	۰/۵۶۰	۰/۶۴۰	۰/۱۲۷	۴۹	۰/۳۴۶	۰/۲۷۶	۰/۵۸۹
۲۳	۰/۶۸۱	۰/۵۵۲	۰/۱۸۱	۵۰	۰/۵۵۹	۰/۶۲۹	۰/۴۴۹
۲۴	۰/۸۲۱	۰/۲۹۶	۰/۲۲۰	۵۱	۰/۱۱۰	۰/۴۸۲	-۰/۱۱۴
۲۵	۰/۷۳۲	۰/۲۵۶	۰/۱۱۲	۵۲	۰/۵۶۶	۰/۲۱۲	۰/۵۰۳
۲۶	۰/۳۹۴	۰/۶۲۴	۰/۳۳۹	۵۳	۰/۷۲۰	۰/۴۵۸	۰/۵۶۱
۲۷	۰/۳۰۶	۰/۶۴۰	۰/۲۹۰	۵۴	۰/۲۴۸	۰/۴۹۹	۰/۲۹۱
۲۸	۰/۶۰۷	۰/۷۲۴	۰/۳۷۸	۵۵	۰/۵۷۴	۰/۴۰۲	۰/۷۵۷

نتایج جدول (۳) نشان می‌دهد که سؤال‌های ۱، ۳، ۴، ۵، ۶، ۷، ۹، ۱۰، ۱۱، ۱۲، ۱۴، ۱۵، ۱۶، ۱۸، ۱۹، ۲۰، ۲۱، ۲۲، ۲۳، ۲۴، ۲۵، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۵۲ و ۵۳ در عامل اول بار عاملی بیشتری دارند، این عامل با نام محاسبات جبری نامگذاری شد. سؤال‌های ۲، ۱۷، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۵۰، ۵۱ و ۴۲ در عامل دوم بار عاملی بیشتری نشان داده‌اند، عامل دوم با نام هندسه نامگذاری شد. همچنین سؤال‌های ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹، ۵۴ و ۵۵ در عامل سوم دارای بار عاملی بیشتری هستند، این عامل با نام احتمال نامگذاری شد. لازم به ذکر است که برخی از سؤال‌ها همچون سؤال ۲۱، ۳۶ و ۵۴ به دلیل داشتن بار عاملی در چند بعد، از لحاظ منطقی در بعد مورد نظر قرار گرفتند.

برای تعیین تعداد عامل‌های آزمون ریاضی کنکور سراسری در سال ۱۳۹۶ نیز از روش تحلیل عاملی اکتشافی در نرم‌افزار NOHARM استفاده شد. نتایج نشان داد که مقدار مجموع مجذورات باقی‌مانده‌ها برابر با ۰/۰۰۷۰ و ریشه دوم میانگین مجذورات باقی‌مانده‌ها (RMSR) برابر با ۰/۰۰۲۳ است که چون مقدار آن از ۰/۰۸ کمتر است و شاخص نیکویی برازش تاناکا برابر با ۰/۹۹۳ و بیشتر از ۰/۹۵ است، نشان‌دهنده برازش مناسب مدل سه بعدی با داده‌ها است (هوپر، کافلان و مولن، ۲۰۰۸؛ کلاین، ۲۰۱۱). با افزودن بعد چهارم به مدل، میزان شاخص نیکویی برازش تاناکا به ۰/۹۹۶ افزایش می‌یابد که افزایش ۰/۰۰۳ مقدار ناچیزی است و با توجه به تطبیق منطقی عامل‌ها با سؤال‌های آزمون مدل سه‌عاملی پذیرفته شد.

به منظور اطمینان بیشتر برای تعیین ابعاد آزمون از روش تحلیل عاملی اکتشافی با plus استفاده شد. مقادیر برازش یک‌بعدی، دوبعدی و سه‌بعدی در جدول (۴) ارائه شده است.

جدول (۴) شاخص‌های برازش مدل یک، دو و سه‌بعدی با داده‌های آزمون ریاضی کنکور سال ۱۳۹۶

تعداد ابعاد	SRMR	RMSEA	TLI	CFI	X^2	df	X^2/df
یک بعدی	۰/۰۷۲	۰/۰۳۰	۰/۹۱	۰/۹۲	۶۸۶۵/۸۶	۱۲۷۴	۵/۳۸
دو بعدی	۰/۰۴۶	۰/۰۲۱	۰/۹۶	۰/۹۶	۳۸۱۲/۶۲	۱۲۲۳	۳/۱۱
سه بعدی	۰/۰۳۸	۰/۰۱۵	۰/۹۷	۰/۹۸	۲۵۵۵/۳۸	۱۱۷۳	۲/۱۷

شاخص‌های برازش مدل در جدول (۴) نشان‌دهنده برازش مناسب داده‌های آزمون ریاضی سال ۱۳۹۶ با مدل سه‌بعدی است. همانطور که دیده می‌شود، شاخص‌های برازش به‌ویژه نسبت X^2/df در مدل‌های یک و دوبعدی بالاتر از ۳ است. بارهای عاملی مربوط به آزمون ریاضی در سال ۱۳۹۶ در جدول (۵) آمده است. سؤال‌های ۲۸، ۴۱ و ۴۷ به دلیل نداشتن بار عاملی مناسب از تحلیل حذف شدند.

جدول (۵) مقادیر بار عاملی مربوط به سه بعد آزمون ریاضی سال ۱۳۹۶

سؤال	عامل اول	عامل دوم	عامل سوم	سؤال	عامل اول	عامل دوم	عامل سوم
۱	۰/۴۹۶	۰/۵۹۲	۰/۱۹۴	۲۷	۰/۴۷۱	۰/۶۹۵	۰/۳۳۴
۲	۰/۵۶۶	۰/۳۶۱	۰/۳۰۲	۲۹	۰/۱۲۱	۰/۵۷۶	-۰/۰۱۱
۳	۰/۶۸۶	۰/۵۳۰	۰/۳۹۵	۳۰	۰/۵۱۱	۰/۷۱۲	۰/۴۲۳
۴	۰/۷۰۱	۰/۵۳۴	۰/۳۳۶	۳۱	۰/۱۲۹	۰/۳۵۲	۰/۱۵۳
۵	۰/۶۹۵	۰/۶۸۵	۰/۳۶۴	۳۲	۰/۲۷۳	۰/۶۱۸	۰/۲۷۳
۶	۰/۷۲۹	۰/۴۷۲	۰/۳۶۷	۳۳	۰/۵۹۹	۰/۶۴۸	۰/۴۹۲
۷	۰/۵۷۷	۰/۴۴۲	۰/۳۱۳	۳۴	۰/۴۰۲	۰/۵۹۷	۰/۳۵۹

سؤال	عامل اول	عامل دوم	عامل سوم	سؤال	عامل اول	عامل دوم	عامل سوم
۸	۰/۷۴۹	۰/۴۷۶	۰/۴۰۲	۳۵	۰/۲۹۳	۰/۵۵۹	۰/۲۴۳
۹	۰/۵۶۹	۰/۴۸۰	۰/۲۲۵	۳۶	۰/۳۷۳	۰/۶۰۳	۰/۳۱۴
۱۰	۰/۵۲۱	۰/۶۰۴	۰/۲۳۰	۳۷	۰/۵۳۴	۰/۴۱۸	۰/۷۶۰
۱۱	۰/۷۵۱	۰/۴۰۸	۰/۴۵۶	۳۸	۰/۴۳۳	۰/۴۵۳	۰/۷۲۷
۱۲	۰/۸۳۳	۰/۴۷۷	۰/۴۸۳	۳۹	۰/۵۴۵	۰/۵۶۳	۰/۵۸۳
۱۳	۰/۴۲۲	۰/۱۴۱	۰/۳۲۷	۴۰	۰/۴۹۵	۰/۴۷۱	۰/۶۷۸
۱۴	۰/۷۲۳	۰/۴۵۱	۰/۴۷۲	۴۲	۰/۴۶۲	۰/۶۸۱	۰/۳۵۴
۱۵	۰/۶۷۳	۰/۶۰۸	۰/۴۳۵	۴۳	۰/۴۲۳	۰/۵۷۳	۰/۳۸۱
۱۶	۰/۷۴۹	۰/۵۰۷	۰/۵۱۰	۴۴	۰/۴۷۹	۰/۶۱۲	۰/۴۷۳
۱۷	۰/۶۵۴	۰/۵۶۱	۰/۴۰۷	۴۵	۰/۴۷۰	۰/۵۰۴	۰/۵۱۲
۱۸	۰/۸۵۹	۰/۵۱۴	۰/۵۳۰	۴۶	۰/۲۷۲	۰/۵۵۱	۰/۲۸۰
۱۹	۰/۷۹۲	۰/۵۸۳	۰/۵۴۱	۴۸	۰/۳۱۲	۰/۵۲۱	۰/۲۷۹
۲۰	۰/۸۰۶	۰/۴۶۷	۰/۵۰۶	۴۹	۰/۲۹۱	۰/۳۸۷	۰/۴۵۵
۲۱	۰/۶۰۶	۰/۶۶۲	۰/۳۸۳	۵۰	۰/۳۶۲	۰/۴۴۱	۰/۵۶۰
۲۲	۰/۵۲۲	۰/۵۸۷	۰/۲۴۷	۵۱	۰/۳۸۵	۰/۵۳۲	۰/۴۴۹
۲۳	۰/۶۷۵	۰/۶۶۱	۰/۳۹۴	۵۲	۰/۲۳۷	۰/۵۳۹	۰/۲۲۰
۲۴	۰/۵۸۷	۰/۴۳۸	۰/۳۵۶	۵۳	۰/۲۳۲	۰/۵۶۹	۰/۲۱۳
۲۵	۰/۳۲۶	۰/۶۳۹	۰/۱۱۴	۵۴	۰/۵۰۶	۰/۶۵۱	۰/۴۶۶
۲۶	۰/۲۷۸	۰/۶۰۱	۰/۰۵۶	۵۵	۰/۵۲۶	۰/۵۱۰	۰/۴۴۹

همان‌طور که در جدول (۵) دیده می‌شود، از لحاظ آماری و منطقی در آزمون ریاضی سال ۱۳۹۶، سؤال‌های ۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰، ۲۱، ۲۲، ۲۳ و ۲۴ در عامل اول بار عاملی بیشتری نشان دادند، این عامل با نام محاسبات جبری نامگذاری شد. در عامل دوم نیز سؤال‌های ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹، ۵۰، ۵۱، ۵۲، ۵۳، ۵۴ و ۵۵ بار عاملی بیشتری نشان دادند و این عامل با نام هندسه و احتمال نام گرفت. سؤال‌های ۳۷، ۳۸، ۳۹، ۴۰، ۴۹ و ۵۰ در عامل سوم بار عاملی بیشتری داشتند که این

عامل جبر خطی نام گرفت. لازم به ذکر است که سؤال‌های ۱، ۱۰ و ۲۱ به دلیل داشتن بار عاملی بالا در چند بعد، به لحاظ منطقی در ابعاد مناسب جای گرفتند.

ویژگی‌های روان‌سنجی سؤال‌های آزمون چندبعدهی

مدل‌های چندبعدهی دوپارامتری برای هر بعد، دارای یک پارامتر شیب (تشخیص) هستند که همانند پارامتر تشخیص در مدل‌های یک‌بعدهی تفسیر می‌شود، با این تفاوت که پارامتر شیب در مدل یک‌بعدهی در یک جهت (صفت مورد اندازه‌گیری) تمیز می‌دهد ولی در چندبعدهی نشانگر چگونگی تمیز در هر بعد است (بروسمان، ۲۰۱۰؛ لی، ۲۰۱۳). در این پژوهش با توجه به اینکه آزمون دارای سه بعد است، سه ضریب تمیز یعنی α_1 ، α_2 و α_3 گزارش شده است.

شاخص‌های $MDIFF_i$ اشاره به فاصله بین مبدأ و عمیق‌ترین نقطه در سطح پاسخ سؤال دارد و $MDISC_i$ اشاره به ضریب تمیز سؤال i دارد. این دو شاخص به صورت زیر تعریف می‌شوند (لی، ۲۰۱۳)

$$MDISC_i = \left(\sum_{k=1}^m a_{ik}^2 \right)^{1/2}$$

$$MDIFF_i = \frac{-d_i}{MDISC_i}$$

در تعریف فوق، پارامتر d مقدار عرض از مبدأ است.

در جدول (۶) پارامترهای چندبعدهی آزمون ریاضی در دو سال ۱۳۹۶ و ۱۳۹۷ ارائه شده است.

جدول (۶) پارامتر سؤال‌های آزمون ریاضی سال ۹۶ و ۹۷ در مدل ۳ عاملی

سال ۱۳۹۶						
d	α_3	α_2	α_1	MDISC	MDIFF	سؤال‌ها
-۳/۵۴۴	۰/۸۵۹	-۱/۱۲۲	۰/۵۵۶	۱/۴۷۳	۲/۴۰۴	۱
-۱/۴۳۷	۱/۰۹۵	۰/۰۵	-۰/۰۱۴	۱/۰۷۳	۱/۳۳۸	۲
-۲/۰۰۵	۱/۴۲۴	-۰/۲۷۶	-۰/۰۴۸	۱/۶۳۷	۱/۲۲۴	۳
-۱/۹۸۹	۱/۵۷۵	-۰/۴۰۹	۰/۲۵۸	۱/۶۹۹	۱/۱۷۰	۴
-۳/۳۵۳	۱/۴۷۲	-۱/۱۱	۰/۳۳۶	۲/۱۱۳	۱/۵۸۶	۵
-۲/۱۲۹	۱/۸۱۳	-۰/۰۲۱	۰/۱۱۵	۱/۷۵۹	۱/۲۱۰	۶
-۱/۸۹۸	۱/۰۹	-۰/۱۸۱	۰/۰۲۵	۱/۱۹۲	۱/۵۹۱	۷
-۱/۶۴۶	۱/۸۶۵	۰/۰۴۷	-۰/۰۱۱	۱/۸۴۳	۰/۸۹۲	۸
-۲/۲۳	۱/۲۱۴	-۰/۴۲۵	۰/۳۷۱	۱/۳۰۱	۱/۷۱۲	۹
-۳/۴۴۷	۰/۹۲۶	-۱/۰۵۶	۰/۴۷۳	۱/۵۰۸	۲/۲۸۵	۱۰
-۱/۹۵۵	۲/۱۳۳	۰/۶۴۹	-۰/۲۷۹	۱/۹۶۵	۰/۹۹۴	۱۱
-۲/۳۷۵	۲/۷۷۹	۰/۶۲	-۰/۲۲۴	۲/۵۷۷	۰/۹۲۱	۱۲
-۰/۰۲۲	۰/۷۹۲	۰/۵۸۷	-۰/۳۹	۰/۸۱۱	۰/۰۲۷	۱۳
-۰/۹۹۴	۱/۵۶۶	۰/۲۰۶	-۰/۳۸۷	۱/۶۹۷	۰/۵۸۵	۱۴
-۳/۰۰۵	۱/۳۲۶	-۰/۳۹۴	-۰/۱۴۵	۱/۶۹۳	۱/۷۷۴	۱۵
-۲/۲۳۳	۱/۷۷۴	۰/۲۰۲	-۰/۴۲۸	۱/۹۳۳	۱/۱۵۴	۱۶
-۲/۶	۱/۲۸۷	-۰/۲۷۶	-۰/۱۰۶	۱/۵۴۰	۱/۶۸۷	۱۷
-۲/۳۸۵	۲/۸۹۳	۰/۴۹	-۰/۳۵۳	۲/۸۳۳	۰/۸۴۱	۱۸
-۳/۳۲۱	۲/۰۴۳	۰/۰۲۵	-۰/۳۹۵	۲/۲۸۵	۱/۴۵۳	۱۹
۲/۴۷۱	۲/۴۲۹	۰/۶۱۶	-۰/۳۸۱	۲/۳۳۱	۱/۰۵۹	۲۰
-۳/۴۶۷	۰/۹۴۱	-۰/۹۵۶	۰/۰۳۵	۱/۶۷۸	۲/۰۶۵	۲۱
-۳/۶۹۲	۰/۹۹	-۰/۹۴	۰/۳۹۴	۱/۴۹۹	۲/۴۶۲	۲۲
-۳/۴۲۱	۱/۳۸۲	-۰/۷۹۵	۰/۱۳۳	۱/۸۸۴	۱/۸۱۵	۲۳
-۲/۰۶۸	۱/۱۴۵	-۰/۰۱۵	-۰/۱۰۵	۱/۲۲۰	۱/۶۹۴	۲۴
-۳/۵۰۸	۰/۰۰۳	-۱/۹۳	۰/۵۴۸	۱/۶۶۸	۲/۱۰۲	۲۵
-۴/۵۱۱	-۰/۰۳۹	-۲/۰۰۳	۰/۷۰۶	۱/۶۶۸	۲/۷۰۳	۲۶

سال ۱۳۹۶						
d	α_3	α_2	α_1	MDISC	MDIFF	سؤال‌ها
-۲/۶۵۸	۰/۱۶۳	-۱/۵۳۶	-۰/۰۹۸	۱/۷۰۲	۱/۵۶۱	۲۷
-۳/۶۷۱	-۰/۶۴۱	-۲/۲۲۹	۰/۵۷۱	۱/۷۰۵	۲/۱۵۲	۲۹
-۳/۷۴۷	۰/۱۹۱	-۱/۴۰۶	-۰/۳۴۴	۱/۷۶۲	۲/۱۲۵	۳۰
-۱/۳۶۳	-۰/۳۴۳	-۰/۶۹۲	-۰/۱۸۱	۰/۶۴۳	۲/۱۱۸	۳۱
-۳/۸۶۴	-۰/۵۶۷	-۱/۶۹۹	-۰/۲۶۲	۱/۵۶۶	۲/۴۶۶	۳۲
-۳/۳۶۲	۰/۷۴۵	-۰/۶۹۸	-۰/۵۱۷	۱/۶۸۱	۱/۹۹۹	۳۳
-۳/۷۶۶	۰/۱۱۳	-۱/۰۲۹	-۰/۳۲	۱/۳۲۲	۲/۸۴۸	۳۴
-۴/۳۲۵	-۰/۰۶	-۱/۳۳۳	-۰/۰۶	۱/۳۳۲	۳/۲۴۵	۳۵
-۴/۳۶۳	۰/۱۳۸	-۱/۱۸۳	-۰/۱۹۷	۱/۴۰۰	۳/۱۱۵	۳۶
-۱/۸۹۶	۰/۵۵۷	۰/۶۶۹	-۲/۰۴۴	۲/۰۵۷	۰/۹۲۱	۳۷
-۲/۳۲۹	-۰/۰۲۵	۰/۲۶۹	-۲/۰۶۸	۱/۹۰۸	۱/۲۲۰	۳۸
-۲/۱۷۹	۰/۴۷۶	۰/۲۷۱	-۱/۰۴۴	۱/۵۷۳	۱/۳۸۴	۳۹
-۲/۹۶۸	۰/۳۷۶	۰/۳۰۶	-۱/۶۵۲	۱/۷۲۳	۱/۷۲۱	۴۰
-۳/۷۷۲	۰/۲۳۱	-۱/۲۴۹	-۰/۱۸۵	۱/۵۲۴	۲/۴۷۴	۴۲
-۳/۰۳۳	۰/۱۱۵	-۰/۹۹۱	-۰/۳۷۵	۱/۳۲۸	۲/۲۸۳	۴۳
-۲/۰۷۲	۰/۱۵۵	-۰/۷۷۴	-۰/۶۴۲	۱/۳۷۳	۱/۵۰۸	۴۴
-۲/۲۰۷	۰/۳۰۲	-۰/۲۸۳	-۰/۸۶۸	۱/۲۷۷	۱/۷۲۷	۴۵
-۴/۰۹۹	-۰/۲۹۹	-۱/۴۰۱	-۰/۲۷۳	۱/۴۰۳	۲/۹۲۰	۴۶
-۳/۱۶۲	-۰/۱۲۹	-۱/۰۰۹	-۰/۲۷۲	۱/۱۰۹	۲/۸۴۸	۴۸
-۱/۲۴۶	-۰/۱۴۵	-۰/۲۱۸	-۰/۸۲۸	۰/۸۸۵	۱/۴۰۶	۴۹
-۱/۲۰۷	-۰/۱۰۸	-۰/۱۵۹	-۱/۱۵۹	۱/۱۹۳	۱/۰۱۰	۵۰
-۲/۴۶	-۰/۰۲۷	-۰/۶۰۴	-۰/۷۳	۱/۱۷۲	۲/۰۹۸	۵۱
-۳/۹۱۷	-۰/۳۸	-۱/۴۸	-۰/۱۱۹	۱/۳۴۹	۲/۹۰۳	۵۲
-۴/۰۸۱	-۰/۴۸	-۱/۴۵۱	-۰/۱۸۹	۱/۳۱۹	۳/۰۹۲	۵۳
-۳/۵۸۳	۰/۰۹۵	-۰/۹۱۹	-۰/۷۰۶	۱/۵۱۷	۲/۳۶۱	۵۴
-۲/۳۱۹	۰/۲۷۸	-۰/۱۵۵	-۰/۹۷۲	۱/۲۶۵	۱/۸۳۲	۵۵

سال ۱۳۹۷						
d	α_3	α_2	α_1	MDISC	MDIFF	سؤال‌ها
-۲/۰۵۷	۰/۴۶۸	۰/۱۸۷	-۰/۶۹۶	۱/۱۲۹	۱/۸۲۰	۱
-۴/۸۹۳	-۰/۶۳۱	۲/۱۵۵	۰/۱۲۴	۱/۹۲۷	۲/۵۳۸	۲
-۱/۰۸۱	۰/۴۴۱	-۰/۷۸۳	-۲/۰۳۲	۱/۹۶۵	۰/۵۴۹	۳
-۲/۱۷	۰/۵۴۱	۰/۲۴۸	-۱/۱۹	۱/۶۸۴	۱/۲۸۷	۴
-۱/۳۸۲	۰/۲۹۵	-۰/۰۵۲	-۱/۱۴۳	۱/۲۸۸	۱/۰۷۲	۵
-۲/۳۰۲	۰/۴۴۲	۰/۳۴۷	-۰/۹۳۱	۱/۴۳۶	۱/۶۰۲	۶
-۲/۴۸۳	۰/۲۴۱	۰/۵۶	-۰/۶۰۲	۱/۱۶۰	۲/۱۳۸	۷
-۲/۰۵۲	۰/۱۶۹	-۰/۱۸۸	-۱/۶۹	۱/۶۸۶	۱/۲۱۶	۹
-۳/۳۱۶	-۰/۳۶۴	۰/۵۹۷	-۱/۱۶۳	۱/۴۰۳	۲/۳۶۲	۱۰
-۲/۰۸۲	-۰/۰۸۳	-۰/۴۴۲	-۲/۵۵۶	۲/۳۱۴	۰/۸۹۹	۱۱
-۴/۸۶	۰/۰۲۲	۱/۴۰۶	-۱/۱۲۷	۲/۲۲۹	۲/۱۷۹	۱۲
-۰/۹۳۳	۰/۱۷۳	-۰/۳۵۸	-۱/۱۱۳	۱/۰۵۷	۰/۸۸۲	۱۴
-۱/۹۰۱	۰/۰۹۸	-۰/۹۷۷	-۳/۲۵۸	۲/۹۰۶	۰/۶۵۴	۱۵
-۳/۵۴۱	-۰/۳۷۳	۰/۹۹۹	-۰/۴۸۸	۱/۱۸۱	۲/۹۹۶	۱۶
-۳/۷۷۴	۰/۰۳۳	۰/۴۷۶	-۱/۳۸۸	۱/۷۰۷	۲/۲۱۰	۱۷
-۳/۴۷۴	-۰/۲۲۴	۰/۰۷	-۲/۳۵۸	۲/۲۹۱	۱/۵۱۶	۱۸
-۲/۵۱۹	-۰/۳۰۲	۰/۱۵۱	-۱/۵۵۱	۱/۵۰۳	۱/۶۷۶	۱۹
-۳/۱۴۴	-۰/۲۷	۰/۳۴۹	-۱/۲۲۶	۱/۳۱۹	۲/۳۸۲	۲۰
-۴/۷۴۲	-۰/۱۹۱	۱/۴۶۴	-۰/۷۸۹	۱/۹۰۷	۲/۴۸۶	۲۱
-۴/۹۱۶	-۰/۲۹۴	۰/۹۰۹	-۱/۵۸۳	۲/۰۶۲	۲/۳۸۳	۲۲
-۲/۷۸۳	-۰/۱۲۶	-۰/۲۳۹	-۲/۶۷	۲/۴۹۲	۱/۱۱۶	۲۳
-۳/۰۷۳	-۰/۳۹۴	-۰/۱۶۷	-۲/۲۹۹	۲/۰۵۲	۱/۴۹۷	۲۴
-۲/۷۹۴	۰/۸۰۱	۱/۱۶۶	۰/۳۴۵	۱/۴۸۳	۱/۸۸۳	۲۵
-۲/۸۱۷	۰/۷۹۹	۱/۳۷۸	۰/۶۹۵	۱/۵۳۹	۱/۸۲۹	۲۶
-۴/۱۱۸	۰/۷۷۲	۱/۴۲۶	-۰/۱۶۸	۱/۹۹۰	۲/۰۶۹	۲۷
-۴/۶۶۶	۰/۴۰۳	۱/۹۶	-۰/۲۴۲	۲/۳۰۶	۲/۰۲۲	۲۸
-۴/۰۴۵	۰/۱۱۸	۱/۵۲۵	۰/۴۰۴	۱/۳۹۸	۲/۸۹۳	۲۹

سال ۱۳۹۷						
d	α_3	α_2	α_1	MDISC	MDIFF	سؤال‌ها
-۳/۶۵۴	-۰/۱۴۷	۱/۸۵۶	۰/۵۵۴	۱/۵۹۳	۲/۲۹۳	۳۰
-۲/۳۱۶	۰/۷۱	۰/۶۷۳	-۰/۲۰۳	۱/۲۹۸	۱/۷۸۳	۳۱
-۲/۲۴۲	۱/۱۴۱	۰/۵۸۵	۰/۴۶	۱/۲۶۶	۱/۷۷۰	۳۲
-۲/۷۲۸	۰/۴۶۶	۰/۱۸۱	-۱/۲۶۹	۱/۶۶۳	۱/۶۳۹	۳۳
-۱/۸۲۲	۰/۵۵۳	-۰/۱۷۵	-۱/۳۳۳	۱/۵۸۸	۱/۱۴۶	۳۴
-۳/۳۱۱	۰/۲۲۱	۰/۴۳۵	-۱/۴۷۶	۱/۸۷۱	۱/۷۶۹	۳۵
-۴/۶۳۹	-۰/۱۴۷	۰/۷۰۹	-۱/۴۴۸	۱/۸۴۶	۲/۵۱۲	۳۶
-۳/۵۳۲	۰/۲۵	۰/۶۹۳	-۱/۲۷۳	۱/۸۸۷	۱/۸۷۱	۳۷
-۱/۷۹۸	۰/۶۰۲	۰/۰۸۲	۰/۰۲۷	۰/۶۲۶	۲/۸۷۱	۳۸
-۲/۰۷۵	۰/۶۰۷	-۰/۱۲۱	-۰/۹۷۷	۱/۳۲۰	۱/۵۷۱	۳۹
-۳/۶۷۱	۰/۵۱۹	۰/۴۶۷	-۱/۰۱۱	۱/۶۵۳	۲/۲۱۹	۴۰
-۳/۶۴۶	-۰/۱۱۳	۱/۰۲۲	۰/۱۶۹	۰/۹۰۸	۴/۰۱۳	۴۲
۰/۷۹۱	۱/۴۸۳	-۰/۵۸۱	۰/۷۹۹	۱/۳۲۶	-۰/۵۹۶	۴۳
-۰/۹۱۳	۱/۴۶۴	-۰/۵۱۳	۰/۴۵۴	۱/۵۵۰	۰/۵۸۸	۴۴
-۱/۴۵۴	۱/۱۹	-۰/۲۷	-۰/۰۷۱	۱/۱۳۵	۱/۲۸۰	۴۵
-۳/۰۱	۱/۱۹۲	۰/۵۵۶	۰/۰۹۲	۱/۴۵۸	۲/۰۶۳	۴۶
-۱/۶۳۸	۱/۱۱۸	۰/۰۰۹	۰/۱۷۳	۱/۰۴۵	۱/۵۶۶	۴۷
-۲/۲۹۱	۰/۷۷	۰/۳۲	-۰/۳۱۵	۱/۱۶۱	۱/۹۷۱	۴۸
-۱/۵۸۶	۱/۴۲۳	-۰/۲۳۴	۰/۱۰۳	۱/۲۹۹	۱/۲۱۹	۴۹
-۳/۶۶۲	۰/۹۹۷	۰/۹۳۲	-۰/۰۸۵	۱/۶۷۸	۲/۱۸۱	۵۰
-۴/۵۱۳	-۰/۲۵۴	۱/۶۳۷	۰/۴۳۵	۱/۳۹۴	۳/۲۳۷	۵۱
-۱/۶۶۳	۰/۹۴۴	-۰/۵۱۷	-۰/۸۵۳	۱/۳۴۸	۱/۲۳۳	۵۲
-۲/۷۴۸	۱/۱۳۴	۰/۰۲۳	-۱/۰۵۵	۱/۹۰۹	۱/۴۳۹	۵۳
-۲/۵۲۶	۰/۷۲۲	۰/۷۷	۰/۴۸۸	۱/۰۳۳	۲/۴۴۵	۵۴
-۲/۵۲۲	۱/۳۱۷	-۰/۲۵۴	-۱/۳۴۷	۲/۱۷۸	۱/۱۵۷	۵۵

در سؤال‌های سال ۱۳۹۷، سؤال‌های ۱۶، ۲۹، ۴۲ و ۵۱ با داشتن بالاترین ضریب MDIFF دشوارترین سؤال‌ها بودند. در سؤال‌های کنکور ریاضی سال ۱۳۹۶ نیز سؤال‌های ۳۵، ۳۶، ۵۲ و ۵۳ دارای بالاترین ضریب MDIFF و جزء دشوارترین سؤال‌ها بودند.

در زمینه ضریب تمیز چندبعدی (MDISC) نیز سؤال ۱۵ در آزمون ریاضی سال ۱۳۹۷ و سؤال‌های ۱۲ و ۱۸ در سال ۱۳۹۶ جزء سؤال‌های با ضریب تمیز چندبعدی بالا هستند.

تفسیر پارامترهای سؤال در نظریه سؤال پاسخ چندبعدی آسان نیست. دشواری و تمیز یک سؤال نظریه سؤال پاسخ چندبعدی به ترکیب خاص صفات مکنون بستگی دارد (لی، ۲۰۰۷). در بررسی پیشینه سه نوع ساختار سؤال برای نظریه سؤال پاسخ چندبعدی شامل ساختار ساده^۱ (SS)، ساختار تقریباً ساده^۲ (APSS) و ساختار پیچیده^۳ (CS) وجود دارد (ژانگ، ۲۰۱۲). در حوزه نظریه سؤال پاسخ چندبعدی، یک ساختار عاملی ساده اشاره به این دارد که سؤال‌های آزمون فقط یک بعد را مورد سنجش قرار دهند که به آن چندبعدیتی بین سؤالی می‌گویند. مانند زمانی که مثلاً در یک آزمون ۲۰ سؤالی ۱۰ سؤال اول، یک بعد آزمون و ۱۰ سؤال دوم، بعد دوم آزمون را مورد سنجش قرار دهند. ساختار ساده یک ساختار ایدئال آزمون است و در عمل کمتر اتفاق می‌افتد، یعنی هر سؤال فقط یک بعد را مورد سنجش قرار دهد (ژانگ، ۲۰۱۲). در ساختار تقریباً ساده (APSS) هر سؤال یک بار عاملی بالاتر در یک بعد و بار عاملی بسیار کمی در ابعاد دیگر دارد. به عبارت دیگر برای یک خوشه از سؤال‌ها، یک ضریب تمیز بالا در بعد یکسان و تمیز پایین در سایر ابعاد دارد. به طور کلی سؤال‌ها در ساختار تقریباً ساده یک ضریب تمیز بالا و چند ضریب تمیز غیر صفر اما پایین دارند. اما ساختار پیچیده اشاره به این دارد که برخی سؤال‌های آزمون دو یا چند صفت مکنون را به طور هم‌زمان مورد سنجش قرار دهند که به آن چندبعدیتی درون سؤالی می‌گویند. برای مثال، هنگامی که در یک آزمون ۲۰ سؤالی، ۱۰ سؤال اول به طور کامل بعد اول را مورد سنجش قرار دهند اما ۱۰ سؤال دوم، ترکیبی از بعد اول و دوم را مورد سنجش قرار دهند، هر سؤال بار عاملی مشابهی در یک بعد و ابعاد دیگر

1. simple structure

2. approximate simple structure

3. complex structure

دارد. این سؤال‌ها در واقعیت و در عمل بیشتر از سؤال‌های ساختار تقریباً ساده دیده می‌شوند (کاهرامان، ۲۰۱۳؛ ژانگ، ۲۰۱۲؛ فریمان، ۲۰۱۶؛ کیم، ۲۰۱۸). همان‌طور که در جدول (۶) دیده می‌شود، برخی از سؤال‌های آزمون ریاضی مانند سؤال‌های ۵۳، ۵۴، ۵۵ و ۹۷ دارای بار عاملی و ضریب تمیز تقریباً یکسان در دو بعد هستند، بنابراین چنین سؤال‌هایی از ساختار پیچیده برخوردار بودند. سؤال‌هایی همچون سؤال ۱، ۳، ۴ و ۵ یک شیب بالاتر در یک بعد نسبت به ابعاد دیگر و شیب بسیار کمی در ابعاد دیگر دارند، از این‌رو، این سؤال‌ها از ساختار تقریباً ساده برخوردار هستند و با توجه به ضریب تمیز مشابه سؤال‌ها در همه سؤال‌های آزمون می‌توان گفت در سؤال‌های ریاضی، سؤالی با ساختار کاملاً ساده وجود ندارد.

نتایج همترازی نمره‌های ریاضی در روش‌های مورد مطالعه

نمره‌های همتراز شده در شش روش مختلف همترازی چندبعدي و تک‌بعدي در جدول (۷) ارائه شده است.

جدول (۷) نمره‌های همتراز شده در شش روش همترازسازی

همترازسازی همصدک	روش نمره مشاهده شده IRT چندبعدي کامل	روش نمره مشاهده شده IRT چندبعدي تکبعدي شده	روش نمره واقعي IRT چندبعدي تکبعدي شده	روش نمره مشاهده شده IRT تکبعدي ^۲	روش نمره واقعي IRT تکبعدي ^۱	نمره‌های خام
۰/۵۴۰۳	-۰/۱۸۱۲	۰/۶۶۳۹	۰	۰/۶۲۶۴	۰	۰
۱/۴۴۹۲	۰/۷۲۰۶	۱/۴۳۳۴	۱/۴۷۰۴	۱/۴۴۱۴	۱/۴۷۶۵	۱
۲/۵۶۷۰	۱/۶۹۰۶	۲/۳۳۳۱	۲/۲۹۹۳	۲/۴۴۴۵	۲/۴۸۹۳	۲
۳/۵۵۵۶	۲/۶۵۲۸	۳/۲۵۵۰	۳/۲۰۸۳	۳/۴۶۲۲	۳/۵۰۱۸	۳
۴/۴۹۲۹	۳/۶۵۳۶	۴/۱۹۱۸	۴/۱۴۹۶	۴/۴۶۵۵	۴/۴۸۷۹	۴
۵/۴۱۲۷	۴/۶۹۱۷	۵/۱۴۵۲	۵/۱۱۰۱	۵/۴۴۹۱	۵/۴۴۵۴	۵
۶/۳۳۲۳	۵/۷۶۲۳	۶/۱۱۳۵	۶/۰۸۴۹	۶/۴۱۷۱	۶/۳۷۹۳	۶
۷/۳۲۷۰	۶/۸۵۹۰	۷/۰۹۴۱	۷/۰۷۱۱	۷/۳۵۹۰	۷/۲۹۶۸	۷
۸/۱۱۱۳	۷/۹۷۰۳	۸/۰۸۵۰	۸/۰۶۶۹	۸/۲۴۵۰	۸/۲۰۴۶	۸
۸/۹۴۶۷	۹/۰۸۱۶	۹/۰۸۴۵	۹/۰۷۰۷	۹/۰۸۵۴	۹/۱۰۸۵	۹
۹/۹۳۹۵	۱۰/۱۸۰۳	۱۰/۰۹۰۹	۱۰/۰۸۰۹	۹/۹۶۰۸	۱۰/۰۱۲۸	۱۰
۱۰/۹۴۷۳	۱۱/۲۵۶۳	۱۱/۱۰۲۹	۱۱/۰۹۶۱	۱۰/۸۹۴۴	۱۰/۹۲۰۰	۱۱
۱۱/۹۰۸۱	۱۲/۳۰۳۸	۱۲/۱۱۹۱	۱۲/۱۱۴۹	۱۱/۸۵۸۶	۱۱/۸۳۱۸	۱۲
۱۲/۷۴۴۰	۱۳/۳۲۰۵	۱۳/۱۳۸۴	۱۳/۱۳۶۴	۱۲/۸۲۶۳	۱۲/۷۴۸۶	۱۳
۱۳/۵۷۸۷	۱۴/۳۰۷۸	۱۴/۱۵۹۸	۱۴/۱۵۹۵	۱۳/۷۸۹۸	۱۳/۶۷۰۴	۱۴
۱۴/۳۷۰۳	۱۵/۲۶۸۹	۱۵/۱۸۲۶	۱۵/۱۸۳۶	۱۴/۷۳۰۲	۱۴/۵۹۷۰	۱۵
۱۵/۴۳۴۴	۱۶/۲۰۸۵	۱۶/۲۰۶۳	۱۶/۲۰۸۲	۱۵/۶۲۶۳	۱۵/۵۲۸۱	۱۶
۱۶/۴۱۶۶	۱۷/۱۳۱۶	۱۷/۲۳۰۵	۱۷/۲۳۳۰	۱۶/۴۹۵۸	۱۶/۶۴۳۴	۱۷
۱۷/۵۵۳۵	۱۸/۰۴۲۵	۱۸/۲۵۴۹	۱۸/۲۵۷۹	۱۷/۳۸۲۸	۱۷/۴۰۲۸	۱۸
۱۸/۶۴۷۰	۱۸/۹۴۵۰	۱۹/۲۷۹۳	۱۹/۲۸۲۹	۱۸/۳۰۴۱	۱۸/۳۴۶۵	۱۹

1. unidimensional true score equating (UIRTT)

2. Unidimensional observed score equating (UIRTO)

همترازسازی همصدک	روش نمره مشاهده شده IRT چندبعدي کامل	روش نمره مشاهده شده IRT چندبعدي تکبعدي شده	روش نمره واقعي IRT چندبعدي تکبعدي شده	روش نمره مشاهده شده IRT تکبعدي ^۲	روش نمره واقعي IRT تکبعدي ^۱	نمره‌های خام
۱۹/۶۶۶۶	۱۹/۸۴۲۴	۲۰/۳۰۳۷	۲۰/۳۰۷۹	۱۹/۲۶۴۳	۱۹/۲۹۴۷	۲۰
۲۰/۴۷۲۲	۲۰/۷۳۸۳	۲۱/۳۲۸۰	۲۱/۳۳۲۹	۲۰/۲۴۴۱	۲۰/۲۴۷۸	۲۱
۲۱/۲۷۲۷	۲۱/۶۳۷۳	۲۲/۳۵۲۱	۲۲/۳۵۷۹	۲۱/۲۳۱۷	۲۱/۲۰۶۵	۲۲
۲۲/۲۰۳۱	۲۲/۵۴۴۱	۲۳/۳۷۵۹	۲۳/۳۸۲۸	۲۲/۲۲۲۳	۲۲/۱۷۱۳	۲۳
۲۳/۲۵۰۰	۲۳/۴۷۳۱	۲۴/۳۹۹۱	۲۴/۴۰۷۵	۲۳/۲۱۲۳	۲۳/۱۴۳۰	۲۴
۲۴/۱۲۹۶	۲۴/۴۲۹۳	۲۵/۴۲۱۵	۲۵/۴۳۱۶	۲۴/۱۹۶۸	۲۴/۱۲۲۵	۲۵
۲۵/۲۲۰۰	۲۵/۴۰۷۳	۲۶/۴۴۲۳	۲۶/۴۵۴۶	۲۵/۱۷۰۷	۲۵/۱۱۰۶	۲۶
۲۶/۳۶۵۳	۲۶/۴۰۷۹	۲۷/۴۶۰۶	۲۷/۴۷۵۶	۲۶/۱۴۱۸	۲۶/۱۰۸۳	۲۷
۲۷/۵۷۸۹	۲۷/۴۳۰۰	۲۸/۴۷۵۱	۲۸/۴۹۳۳	۲۷/۱۳۰۲	۲۷/۱۱۶۵	۲۸
۲۸/۵۸۸۲	۲۸/۴۷۰۸	۲۹/۴۸۳۸	۲۹/۵۰۵۷	۲۸/۱۳۹۰	۲۸/۱۳۶۰	۲۹
۲۹/۴۴۱۱	۲۹/۵۳۶۴	۳۰/۳۸۴۳	۳۰/۵۱۰۳	۲۹/۱۵۹۲	۲۹/۱۶۷۰	۳۰
۳۰/۵۰۰۰	۳۰/۶۲۶۲	۳۱/۴۷۳۷	۳۱/۵۰۳۶	۳۰/۱۸۵۳	۳۰/۲۰۹۳	۳۱
۳۱/۳۴۳۷	۳۱/۷۲۳۰	۳۲/۴۴۸۵	۳۲/۴۸۱۸	۳۱/۲۱۷۴	۳۱/۲۶۳۲	۳۲
۳۳/۳۱۲۵	۳۲/۸۲۲۴	۳۳/۴۰۵۵	۳۳/۴۴۰۴	۳۲/۲۶۰۸	۳۲/۳۲۵۶	۳۳
۳۴/۴۰۹۰	۳۳/۹۲۳۲	۳۴/۳۴۱۵	۳۴/۳۷۴۷	۳۳/۳۲۶۷	۳۳/۳۹۴۰	۳۴
۳۵/۲۰۰۰	۳۵/۰۲۴۵	۳۵/۲۵۲۵	۳۵/۲۸۰۳	۳۴/۴۱۹۱	۳۴/۴۶۴۵	۳۵
۳۶/۱۲۵۰	۳۶/۱۲۸۶	۳۶/۱۳۰۹	۳۶/۱۵۳۰	۳۵/۵۱۰۸	۳۵/۵۳۱۵	۳۶
۳۷/۱۶۶۶	۳۷/۲۳۷۹	۳۶/۹۶۸۶	۳۶/۹۸۹۸	۳۶/۵۷۱۳	۳۶/۵۸۸۵	۳۷
۳۷/۷۷۷۷	۳۸/۳۵۱۲	۳۷/۷۶۷۳	۳۷/۷۸۸۶	۳۷/۶۰۳۶	۳۷/۶۲۷۹	۳۸
۳۸/۲۲۲۲	۳۹/۴۶۵۲	۳۸/۵۴۰۹	۳۸/۵۴۸۸	۳۸/۶۱۷۲	۳۸/۶۴۱۳	۳۹
۳۹	۴۰/۶۲۵۰	۳۹/۳۱۹۳	۳۹/۲۷۰۸	۳۹/۶۱۰۳	۳۹/۶۱۹۸	۴۰
۴۰/۷۵۰۰	۴۱/۷۹۳۴	۴۰/۱۲۳۳	۳۹/۹۵۶۶	۴۰/۵۷۴۳	۴۰/۵۵۵۰	۴۱

همترازسازی همصدک	روش نمره مشاهده شده IRT چندبعدي کامل	روش نمره مشاهده شده IRT چندبعدي تکبعدي شده	روش نمره واقعي IRT چندبعدي تکبعدي شده	روش نمره مشاهده شده IRT تکبعدي ^۲	روش نمره واقعي IRT تکبعدي ^۱	نمره‌های خام
۴۳/۹۰۰۰	۴۲/۹۴۹۹	۴۰/۹۵۱۶	۴۰/۶۰۹۸	۴۱/۵۰۰۶	۴۱/۴۳۹۳	۴۲
۴۴/۵۰۰۰	۴۴/۱۰۵۲	۴۱/۷۸۹۶	۴۱/۲۳۵۲	۴۲/۴۱۳۸	۴۲/۲۶۷۴	۴۳
۴۵	۴۵/۲۴۹۹	۴۲/۶۱۸۲	۴۱/۸۳۹۴	۴۳/۲۹۷۳	۴۳/۰۳۷۴	۴۴
۴۷	۴۶/۳۵۷۱	۴۳/۴۶۳۹	۴۲/۴۳۱۰	۴۴/۱۴۷۴	۴۳/۷۵۱۹	۴۵
۴۸	۴۷/۳۷۵۰	۴۴/۳۸۴۶	۴۳/۰۲۰۷	۴۴/۹۵۱۶	۴۴/۴۱۹۳	۴۶
۴۸	۴۸/۲۵۰۰	۴۵/۳۲۰۳	۴۳/۶۲۳۷	۴۵/۶۸۳۹	۴۵/۰۵۴۷	۴۷
۴۸	۴۹	۴۶/۲۶۸۷	۴۴/۲۶۲۴	۴۶/۴۳۴۲	۴۵/۶۸۲۶	۴۸
۴۹	۵۲/۵۰	۴۷/۲۳۱۸	۴۴/۹۷۵۴	۴۷/۳۱۵۵	۴۶/۳۴۲۴	۴۹
۵۲/۵۰	۵۲/۵۰	۴۸/۲۱۷۷	۴۵/۸۴۳۶	۴۸/۲۱۴۱	۴۷/۱۰۹۶	۵۰
۵۲/۵۰	۵۲/۵۰	۴۹/۲۴۴۸	۴۷/۰۹۸۲	۴۹/۱۳۶۷	۴۸/۱۸۵۷	۵۱
۵۲/۵۰	۵۲/۵۰	۵۰/۳۴۲۱	۵۲	۵۰/۱۵۳۲	۵۲	۵۲

همانطور که در جدول (۷) دیده می‌شود دو روش همترازسازی تکبعدي نمره مشاهده‌شده و نمره واقعي در همترازسازی نمره‌ها مشابه هم عمل کرده‌اند. همچنین دو روش همترازسازی چندبعدي تکبعدي شده نمره واقعي و نمره مشاهده شده نیز مشابه هم عمل کرده‌اند. بنابراین بین روش‌های همترازسازی نمره واقعي و نمره مشاهده شده تفاوت چندانی وجود ندارد. بلکه تفاوت موجود در روش‌های مختلف همترازسازی به ساختار و تکبعدي و چندبعدي بودن روش‌ها برمی‌گردد.

روش همترازسازی چندبعدي کامل نمره مشاهده‌شده تقریباً مشابه روش همترازسازی همصدک عمل کرده است. در نمره‌های پایین شباهت زیادی بین شش

روش همترازسازی در نمره‌های همتراز شده وجود دارد، اما در نمره‌های بالا میزان شباهت بین روش‌ها کمتر می‌شود.

مقایسه روش‌های همترازسازی چندبعدی و تک‌بعدی با روش همصدک در جدول (۸) تفاوت بین روش‌های چندبعدی و تک‌بعدی همترازسازی نظریه سؤال پاسخ و کلاسیک با روش همترازسازی همصدک آمده است.

جدول (۸) تفاوت پنج روش همترازسازی تک‌بعدی و چندبعدی با روش همترازسازی همصدک

روش چندبعدی کامل نمره مشاهده شده IRT	روش نمره مشاهده شده چندبعدی تک‌بعدی شده IRT	روش نمره واقعی چندبعدی تک‌بعدی شده IRT	روش تک‌بعدی نمره مشاهده شده IRT	روش تک‌بعدی نمره واقعی IRT	نمره‌های خام
-۰/۷۲۱۶	۰/۱۲۳۵	-۰/۵۴۰۴	۰/۰۸۶۰	-۰/۵۴۰۴	۰
-۰/۷۲۸۶	-۰/۰۱۵۸	۰/۰۲۱۲	-۰/۰۰۷۸	۰/۰۲۷۳	۱
-۰/۸۷۶۴	-۰/۲۳۳۹	-۰/۲۶۷۷	-۰/۱۲۲۵	-۰/۰۷۷۷	۲
-۰/۹۰۲۸	-۰/۳۰۰۷	-۰/۳۴۷۴	-۰/۰۹۳۵	-۰/۰۵۳۹	۳
-۰/۸۳۹۳	-۰/۳۰۱۱	-۰/۳۴۳۳	-۰/۲۷۴	-۰/۰۰۵۰	۴
۰/۷۲۱	-۰/۲۶۷۶	-۰/۳۰۲۷	۰/۰۳۶۳	۰/۰۳۲۶	۵
۰/۵۶۹۹	-۰/۲۱۸۸	-۰/۲۴۷۴	۰/۰۸۴۸	۰/۰۴۷۰	۶
-۰/۴۶۸	-۰/۲۳۳۹	۰/۲۵۵۹	۰/۰۳۲۰	-۰/۰۳۰۲	۷
-۰/۱۴۱	-۰/۰۲۶۳	-۰/۰۴۴۴	۰/۱۳۳۷	۰/۰۹۳۳	۸
۰/۱۳۴۹	۰/۱۳۷۸	۰/۱۲۴۰	۰/۱۳۸۷	۰/۱۶۱۸	۹
۰/۲۴۰۷	۰/۱۵۱۳	۰/۱۴۱۳	۰/۰۲۱۲	۰/۰۷۳۲	۱۰
۰/۳۰۹	۰/۱۵۵۵	۰/۱۴۸۷	-۰/۰۵۳۰	-۰/۰۲۷۴	۱۱
۰/۳۹۵۷	۰/۲۱۰۹	۰/۲۰۶۷	-۰/۰۵۱۴	-۰/۰۷۶۴	۱۲
۰/۵۷۶۵	۰/۳۹۴۳	۰/۳۹۲۳	۰/۰۸۲۲	۰/۰۰۴۵	۱۳
۰/۷۲۹۱	۰/۵۸۱۱	۰/۵۸۰۸	۰/۲۱۱۱	۰/۰۹۱۷	۱۴
۰/۸۹۸۶	۰/۸۱۲۲	۰/۸۱۳۲	۰/۳۵۹۸	۰/۲۲۶۶	۱۵

روش چندبعدی کامل نمره مشاهده شده IRT	روش نمره مشاهده شده چندبعدی تک بعدی شده IRT	روش نمره واقعی چندبعدی تک بعدی شده IRT	روش تک بعدی نمره مشاهده شده IRT	روش تک بعدی نمره واقعی IRT	نمره‌های خام
۰/۷۷۴۲	۰/۷۷۱۹	۰/۷۷۳۸	۰/۱۹۱۹	۰/۰۹۳۷	۱۶
۰/۷۱۴۹	۰/۸۱۳۸	۰/۸۱۶۳	۰/۰۷۹۱	۰/۰۴۶۷	۱۷
۰/۴۸۸۹	۰/۷۰۱۳	۰/۷۰۴۳	-۰/۱۷۰۸	-۰/۱۵۰۸	۱۸
۰/۲۹۸	۰/۶۳۲۲	۰/۶۳۵۸	-۰/۳۴۳۰	-۰/۳۰۰۶	۱۹
۰/۱۷۵۸	۰/۶۳۷۰	۰/۶۴۱۲	-۰/۴۰۲۴	-۰/۳۷۲۰	۲۰
۰/۲۶۶۲	۰/۸۵۵۸	۰/۸۶۰۷	-۰/۲۲۸۱	-۰/۲۲۴۴	۲۱
۰/۳۶۴۶	۱/۰۷۹۴	۱/۰۸۵۲	-۰/۰۴۱۰	-۰/۰۶۶۲	۲۲
۰/۳۴۱	۱/۱۷۲۸	۱/۱۷۹۷	۰/۰۱۹۲	-۰/۰۳۱۸	۲۳
۰/۲۲۳۱	۱/۱۴۹۱	۱/۱۵۷۵	-۰/۰۳۷۷	-۰/۱۰۷۰	۲۴
۰/۲۹۹۷	۱/۲۹۱۹	۱/۳۰۲۰	۰/۰۶۷۲	-۰/۰۰۷۱	۲۵
۰/۱۸۷۴	۱/۲۲۲۳	۱/۲۳۴۶	-۰/۰۴۹۳	-۰/۱۰۹۴	۲۶
۰/۰۴۲۶	۱/۰۹۵۲	۱/۱۱۰۲	-۰/۲۲۳۶	-۰/۲۵۷۱	۲۷
-۰/۱۴۸۹	۰/۸۹۶۲	۰/۹۱۴۴	-۰/۴۴۸۷	-۰/۴۶۲۴	۲۸
-۰/۱۱۷۳	۰/۸۹۵۶	۰/۹۱۷۵	-۰/۴۴۹۲	-۰/۴۵۲۲	۲۹
۰/۰۹۵۳	۱/۰۴۳۱	۱/۰۶۹۱	-۰/۲۸۲۰	-۰/۲۷۴۱	۳۰
۰/۱۲۶۳	۰/۹۷۳۷	۱/۰۰۳۶	-۰/۳۱۴۷	-۰/۲۹۰۲	۳۱
۰/۳۷۹۳	۱/۱۰۴۸	۱/۱۳۸۱	-۰/۱۲۶۳	-۰/۰۸۰۵	۳۲
-۰/۴۹۰	۰/۰۹۳۰	۰/۱۲۷۹	-۱/۰۵۱۷	-۰/۹۸۶۳	۳۳
-۰/۴۸۵۸	-۰/۰۶۷۶	-۰/۰۳۴۴	-۱/۰۸۲۴	-۱/۰۱۵۱	۳۴
-۰/۱۷۵۵	۰/۰۵۲۵	۰/۸۰۳	-۰/۷۸۰۹	-۰/۷۳۵۵	۳۵

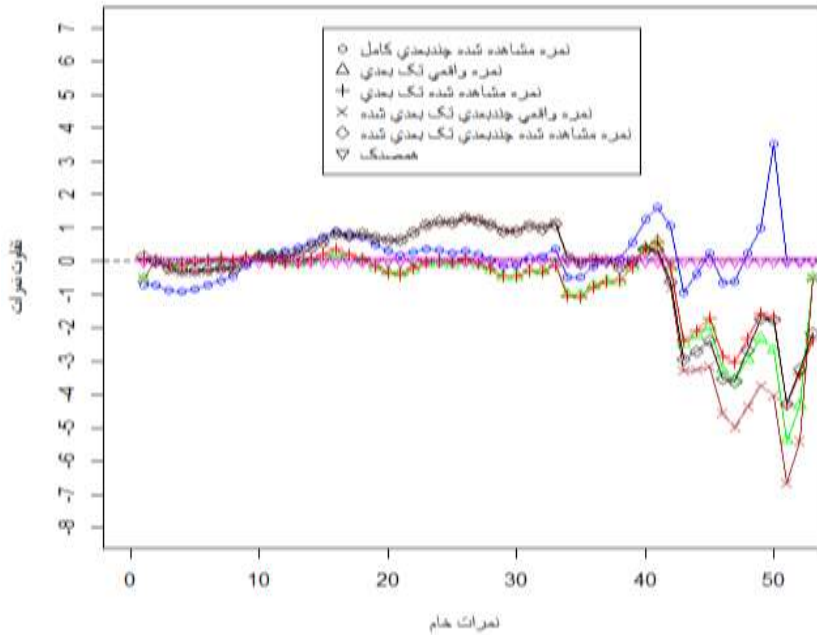
نمره‌های خام	روش تک‌بعدی نمره واقعی IRT	روش تک‌بعدی نمره مشاهده شده IRT	روش تک‌بعدی نمره مشاهده شده IRT	روش واقعی تک‌بعدی نمره مشاهده شده IRT	روش مشاهده شده تک‌بعدی نمره مشاهده شده IRT
۳۶	۰/۵۹۳۵	-۰/۶۱۴۲	۰/۲۸۰	۰/۰۳۷	۰/۰۵۹
۳۷	-۰/۵۷۸۲	-۰/۵۹۵۴	-۰/۱۷۶۹	۰/۰۷۱۳	۰/۱۹۸۱
۳۸	-۰/۱۴۹۹	-۰/۱۷۴۲	۰/۰۱۰۸	۰/۵۷۳۴	-۰/۰۱۰۵
۳۹	۰/۴۱۹۱	۰/۳۹۵۰	۰/۳۲۶۶	۱/۲۴۳۱	۰/۳۱۸۷
۴۰	۰/۶۱۹۸	۰/۶۱۰۳	۰/۲۷۰۸	۱/۶۲۵	۰/۳۱۹۳
۴۱	-۰/۱۹۵۰	-۰/۱۷۵۷	-۰/۷۹۳۴	۱/۰۴۳۵	-۰/۶۲۶۷
۴۲	-۲/۴۶۰۷	-۲/۳۹۹۴	-۳/۲۹۰۲	-۰/۹۵۰۰	-۲/۹۴۸۴
۴۳	-۲/۲۳۲۶	-۲/۰۸۶۲	-۳/۲۶۴۸	-۰/۳۹۴۷	-۲/۷۱۰۴
۴۴	-۱/۹۶۲۶	۱/۷۰۲۷	-۳/۱۶۰۶	۰/۲۴۹۹	-۲/۳۸۱۸
۴۵	-۳/۲۴۸۱	-۲/۸۵۲۶	-۴/۵۶۹۰	-۰/۶۴۲۹	-۳/۵۳۶۱
۴۶	-۳/۵۸۰۷	-۳/۰۴۸۴	-۴/۹۷۹۳	-۰/۶۲۵	-۳/۶۱۵۴
۴۷	-۲/۹۴۵۳	-۲/۳۱۶۱	-۴/۳۷۶۳	۰/۲۵۰	-۲/۶۷۹۷
۴۸	-۲/۳۱۷۴	-۱/۵۶۵۸	-۳/۷۳۷۶	۱/۰۰۰	-۱/۷۳۱۳
۴۹	-۲/۶۵۷۶	-۱/۶۸۴۵	-۴/۰۲۴۶	۳/۵۰۰	-۱/۷۶۸۲
۵۰	-۵/۳۹۰۴	-۴/۲۸۵۹	-۶/۶۵۶۴	۰	-۴/۲۸۲۳
۵۱	-۴/۳۱۴۳	-۳/۳۶۳۳	-۵/۴۰۱۸	۰	-۳/۲۵۵۲
۵۲	-۰/۵۰۰۰	-۲/۳۴۶۸	-۰/۵۰۰۰	۰	-۲/۱۵۷۹
میانگین تفاوت‌ها	-۰/۷۱۵۵	-۰/۶۲۳۵	-۰/۵۱۸۸	۰/۱۴۳۸	-۰/۲۶۱۷

همانطور که در جدول (۸) دیده می‌شود کمترین میزان تفاوت بین روش همصدک ناهموار با روش‌های چندبعدی و تک‌بعدی همترازسازی مربوط به روش همترازسازی نمره مشاهده شده چندبعدی کامل و بیشترین تفاوت مربوط به روش همترازسازی نمره واقعی تک‌بعدی است.

همچنین برای بررسی تفاوت بین روش‌ها از ملاک تفاوت مهم (DTM) استفاده شد. نتایج نشان می‌دهد که در روش همترازسازی نمره مشاهده شده چندبعدی کامل و روش همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده، میانگین تفاوت نمره‌های همتراز شده از همترازسازی همصدک کمتر از ملاک $0/5$ است، اما میانگین تفاوت‌ها در روش همترازسازی نمره مشاهده شده و نمره واقعی تک‌بعدی و همترازسازی نمره واقعی چندبعدی تک‌بعدی شده بیشتر از $0/5$ است.

همان‌طور که دیده می‌شود، نزدیک‌ترین روش به روش همصدک، روش نمره مشاهده شده چندبعدی کامل، پس از آن روش همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده و روش همترازسازی نمره واقعی چندبعدی تک‌بعدی شده قرار می‌گیرند. دو روش همترازسازی نمره مشاهده شده و نمره واقعی تک‌بعدی نیز بیشترین تفاوت را با روش همترازسازی همصدک داشتند.

نمودار (۱) تفاوت بین پنج روش همترازسازی تک‌بعدی و چندبعدی را با روش همصدک ناهموار به صورت دیداری نشان می‌دهد.



نمودار (۱) مقایسه نتایج همترازسازی پنج روش همترازسازی IRT تک‌بعدی و چندبعدی با روش همصدک ناهموار

نمودار (۱) نشان‌دهنده تفاوت روش‌های مختلف همترازسازی با همترازسازی همصدک است. همان‌طور که دیده می‌شود در نمره‌های پایین، نمره‌های همتراز شده در پنج روش مورد نظر تقریباً شبیه به هم هستند و تفاوت بین روش‌ها در نمره‌های پایین اندک است. در نمره‌های وسط بین نمره‌های ۲۲ تا ۳۲ تفاوت روش‌های همترازسازی نمره واقعی چندبعدی تک‌بعدی شده و روش همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده با روش همصدک نسبت به سه روش دیگر بیشتر است و هر چه نمره‌ها بیشتر می‌شود میزان تفاوت بین روش‌ها و به‌خصوص تفاوت بین روش‌های همترازسازی نظریه سؤال پاسخ تک‌بعدی با روش همصدک بیشتر می‌شود. در نمودار (۱) دیده می‌شود که نزدیک‌ترین روش به همصدک، روش همترازسازی نمره مشاهده چندبعدی کامل و دورترین روش در نمره‌های بالا، روش همترازسازی نمره واقعی چندبعدی تک‌بعدی شده است. همچنین در همترازسازی، روش‌های نمره

واقعی و نمره مشاهده شده تک‌بعدي شبیه به هم عمل کرده و روش‌های همترازسازی نمره واقعی و نمره مشاهده شده چندبعدي تک‌بعدي شده مشابه هم عمل کرده‌اند.

بحث و نتیجه‌گیری

هدف از اجرای پژوهش حاضر، شناسایی رویکرد بهینه به منظور همترازسازی آزمون‌های چندبعدي بود. به این منظور، شش روش همترازسازی شامل روش‌های همترازسازی نمره مشاهده شده تک‌بعدي (UIRTO)، همترازسازی نمره واقعی تک‌بعدي (UIRTT)، همترازسازی نمره مشاهده شده چندبعدي تک‌بعدي شده (UMIRTT)، همترازسازی نمره واقعی چندبعدي تک‌بعدي شده (UMIRTO)، همترازسازی نمره مشاهده شده چندبعدي کامل (FMIRT) و روش همترازسازی همصداک برای همترازسازی داده‌های آزمون چندبعدي ریاضی مقایسه شد. با بررسی پیشینه مشخص شد که نقض مفروضه تک‌بعدي بودن آزمون، اثر سوء بر نتایج همترازسازی دارد و در صورت نقض کامل این مفروضه نمی‌توان به روش‌های همترازسازی تک‌بعدي اعتماد کرد. از این‌رو، مدل‌های نظریه سؤال پاسخ چندبعدي، اخیراً مدنظر پژوهشگران قرار گرفته و در زمینه همترازسازی و پیونددهی چندبعدي تلاش‌هایی از سوی پژوهشگران صورت گرفته است (بروسمان، ۲۰۱۰؛ ژانگ، ۲۰۱۲؛ لی، ۲۰۱۳؛ لی و برینان، ۲۰۱۴). اما مطالعات انجام گرفته، نتیجه مشخص و همسویی را در زمینه همترازسازی آزمون‌های چندبعدي نشان نداده‌اند که با اطمینان بتوان از نتایج آنها استفاده کرد. بنابراین، لزوم اجرای پژوهشی که بتواند خلأ موجود در پیشینه روش‌های پیشرفته و همترازسازی چندبعدي را پر کند، احساس شد.

نتایج به دست آمده نشان داد که در نمره‌های پایین بین شش روش همترازسازی تفاوت چندانی از لحاظ نمره‌های همتراز شده وجود ندارد. در نمره‌های وسط بین نمره‌های ۲۲ تا ۳۲ یعنی نمره‌های میانی تفاوت روش‌های همترازسازی نمره مشاهده شده و نمره واقعی چندبعدي تک‌بعدي شده با روش همترازسازی همصداک بیشتر از سایر روش‌ها است. در نمره‌های بالا نیز بین روش‌های همترازسازی تک‌بعدي و چندبعدي نظریه سؤال پاسخ با روش همصداک تفاوت بیشتری وجود دارد. این یافته با پژوهش هان، کولن و پوهلمان^۱ (۱۹۹۷) همسو است. آنها در پژوهشی که تفاوت و

^۱. Han, Kolen & Pohlmann

شباهت‌های همترازسازی نمره مشاهده شده و نمره واقعی نظریه سؤال پاسخ با همترازسازی همصدک در آزمون‌های ورودی ریاضی و زبان دانشگاه‌های آمریکا را بررسی کردند، نشان دادند که همترازسازی نمره واقعی نظریه سؤال پاسخ نتایج پایاتری نسبت به دو روش دیگر داشت. همچنین همترازسازی نمره مشاهده شده نظریه سؤال پاسخ نتایج پایاتری نسبت به همترازسازی همصدک نشان داد. دستیابی به چنین یافته‌ای در پژوهش حاضر به جهت تفاوتی که بین دو نظریه سؤال پاسخ و نظریه کلاسیک از لحاظ خطاها وجود دارد، قابل انتظار است. افزون بر این، تفاوت اندک روش‌های همترازسازی در نمره‌های پایین همترازسازی و تفاوت زیاد روش‌های همترازسازی در نمره‌های بالا ممکن است مربوط به فراوانی اندک و نزدیک به صفر نمره‌ها در طیف بالا باشد (کیم، ۲۰۱۸). به این دلیل که فراوانی تعداد افرادی که در کنکور سراسری نمره‌های پایین کسب کرده‌اند بسیار بیشتر از افرادی است که نمره‌های بالا کسب کرده‌اند و در نمره‌های بالا فراوانی افراد به شدت کم بوده است، چراکه؛ با توجه به دشواری آزمون ریاضی کنکور سراسری تعداد انگشت‌شماری موفق می‌شوند که نمره‌های بسیار بالا کسب کنند. یکی از عوامل اثرگذار بر همترازسازی نمره‌ها، حجم نمونه است که این اثر در پژوهش‌های مختلف به تأیید رسیده است (کولن و برینان، ۲۰۰۴؛ لی، ۱۹۹۷؛ سیمون، ۲۰۰۸؛ کائو، ۲۰۰۸). به نظر می‌رسد در شرایطی که حجم نمونه بالا است، به دلیل کم شدن خطای نمونه‌گیری، استفاده از روش‌های چندبعدي و تک‌بعدي نتایج یکسانی داشته باشد، اما برای اطمینان از این موضوع باید پژوهش‌هایی در خصوص اثرات حجم نمونه بر روش‌های مختلف همترازسازی اجرا شود. در پژوهش کیم (۲۰۱۸) نیز همه روش‌های همترازسازی به جز در نمره‌های بالا و پایین نتایج مشابهی داشتند و در نمره‌های نزدیک ۵۰ همترازسازی تک‌بعدي نظریه سؤال پاسخ خطای کمتری نسبت به روش‌های همترازسازی چندبعدي داشتند. اما در سایر طیف نمره‌ها روش همترازسازی چندبعدي بهتر از تک‌بعدي عمل کرد.

در همترازسازی، روش‌های نمره واقعی و نمره مشاهده شده تک‌بعدي شبیه به هم عمل کرده و روش‌های همترازسازی نمره واقعی و نمره مشاهده چندبعدي تک‌بعدي شده مشابه هم عمل کرده‌اند. این یافته، نشان‌دهنده تأثیر نقض مفروضه تک‌بعدي

بودن بر همترازسازی داده‌ها است و با پژوهش لی (۲۰۱۳) همسو است. لی در پژوهش خود نشان داد که روش‌های مختلف همترازسازی در دو دسته قرار می‌گیرند، روش همترازسازی چندبعدي شامل روش‌های همترازسازی نمره مشاهده‌شده چندبعدي کامل و روش‌های نمره مشاهده شده و نمره واقعي چندبعدي تک‌بعدي شده در یک گروه مشابه هم عمل کردند و روش همترازسازی نمره مشاهده شده و نمره واقعي تک‌بعدي نیز در یک گروه قرار گرفتند و مشابه هم عمل کردند. در پژوهش بروسمان (۲۰۱۰) و لی و بروسمان (۲۰۱۲) هم دو روش همترازسازی تک‌بعدي شبیه به هم و سه روش همترازسازی چندبعدي مشابه هم عمل کردند. این یافته به دلیل خطای بیشتر روش‌های تک‌بعدي برای همترازسازی داده‌های چندبعدي قابل توجیه است. به این دلیل که؛ با توجه به نقض مفروضه تک‌بعدي، خطای منظم استفاده از روش‌های تک‌بعدي افزایش یافته و تفاوت بین روش‌های تک‌بعدي و چندبعدي مورد انتظار است. در همین خصوص، پژوهش چوی (۲۰۱۹) نشان داد که در داده‌های چندبعدي عملکرد روش‌های همترازسازی چندبعدي نسبت به روش‌های تک‌بعدي بهتر است و روش‌های همترازسازی چندبعدي کامل و روش همترازسازی چندبعدي با ساختار ساده نتایج دقیق‌تری نسبت به روش‌های تک‌بعدي فراهم کردند. همچنین، این یافته در تأیید پژوهش اندریو (۲۰۱۱) است که نشان داد روش‌های همترازسازی تک‌بعدي نظریه سؤال پاسخ برای داده‌های تک‌بعدي بهتر عمل کرد و روش همترازسازی نمره واقعي تک‌بعدي بهترین روش در بین روش‌های تک‌بعدي بود. صرف‌نظر از میزان همبستگی بین ابعاد، هر چه میزان چندبعدي بودن آزمون افزایش یافت روش همترازسازی نمره مشاهده شده چندبعدي کامل روش مناسب‌تری برای همترازی مرتبه اول و مرتبه دوم بود.

نزدیک بودن نمره‌های همتراز شده نمره واقعي و نمره مشاهده شده تک‌بعدي و روش‌های نمره واقعي و نمره مشاهده شده چندبعدي تک‌بعدي شده نشان‌دهنده این است که بین نمره واقعي و نمره مشاهده شده تفاوت چندانی از لحاظ نتایج همترازسازی وجود ندارد، بلکه عامل اثرگذار بر تفاوت نتایج همترازسازی ساختار و ابعاد آزمون است. این یافته همسو با پژوهش چن (۲۰۱۴)، لی (۲۰۱۳) و بروسمان (۲۰۱۰) است که در پژوهش آنها بین روش‌های همترازسازی نمره واقعي و نمره مشاهده شده تفاوت چندانی وجود نداشت. بنابراین می‌توان گفت عامل تأثیرگذار بر

تفاوت عملکرد همترازسازی در روش‌های مختلف به ساختار متفاوت آزمون و نه به فرایند همترازسازی مربوط می‌شود.

در بررسی تفاوت روش‌های همترازسازی چندبعدی و تک‌بعدی با روش همصدک، نتایج نشان داد که نزدیک‌ترین روش به همصدک، روش همترازسازی نمره مشاهده شده چندبعدی کامل و دورترین روش به همصدک روش همترازسازی نمره واقعی تک‌بعدی است. همچنین نتایج نشان داد که در روش همترازسازی نمره مشاهده شده چندبعدی کامل و روش همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده، میانگین تفاوت نمره‌های همتراز شده از همترازسازی همصدک کمتر از ملاک ۰/۵ است، اما میانگین تفاوت‌ها در روش همترازسازی نمره مشاهده شده و نمره واقعی تک‌بعدی و همترازسازی نمره واقعی چندبعدی تک‌بعدی شده بیشتر از ۰/۵ است. از این رو، روش همترازسازی چند بعدی کامل نسبت به سایر روش‌ها شباهت بیشتری با روش همصدک دارد. این یافته همسو با پژوهش بروسمان (۲۰۱۰) است که در آن تفاوت روش‌های چندبعدی با همصدک کمتر از ۰/۵ و تفاوت روش‌های تک‌بعدی با همصدک بیشتر از ۰/۵ بود. بنابراین، می‌توان گفت با توجه به هدف پژوهش که مقایسه پنج روش همترازسازی تک‌بعدی و چندبعدی نظریه سؤال پاسخ با روش کلاسیک همصدک بود، روش همترازسازی چندبعدی کامل نسبت به روش‌های تک‌بعدی و سایر روش‌های چندبعدی شباهت بیشتری به همصدک داشت. و در شرایطی که داده‌ها چندبعدی است، اگرچه می‌توان از سه روش همترازسازی چندبعدی به لحاظ شباهت نتایج همترازسازی به‌ویژه در نمره‌های پایین استفاده کرد، اما با توجه به شباهت بیشتر روش همترازسازی نمره مشاهده شده چند بعدی کامل، این روش نسبت به سایر روش‌ها از ارجحیت بیشتری در همترازسازی داده‌های چندبعدی برخوردار است. پژوهشگرانی همچون واتزل و هل^۱ (۲۰۱۴) و شو و وایز^۲ (۲۰۱۵) نیز نشان دادند که تحلیل چندبعدی دارای مزایایی از جمله تفسیر آسان صفات پنهان، برآورد مستقیم پارامترهای شخص و سؤال به‌طور هم‌زمان، دقت بالاتر در برآورد پارامترهای شخص و سؤال، برآورد مستقیم همبستگی بین ابعاد و دقت بیشتر در ارائه مفروضات نظری است.

1. Wetzel & Hell

2. Seo & Weiss

نتیجه حائز اهمیت در این پژوهش این بود که در نمره‌های بالا، روش همترازسازی نمره واقعی چندبعدی تک‌بعدی شده بیشترین تفاوت را با روش همصدک داشت. این یافته با پژوهش کیم (۲۰۱۸) که نشان داد روش همترازسازی نمره واقعی چندبعدی ساختار ساده SMT در مقایسه با روش‌های معمول تک‌بعدی دقت همترازسازی بیشتری فراهم کرد و نتایج مفید بودن رویکرد چندبعدی را در همترازسازی داده‌های چندبعدی نشان داد، ناهمسو است، همچنین پژوهش ژانگ (۲۰۱۲) با این یافته ناهمسو است، چراکه وی نشان داد روش همترازسازی نمره واقعی چندبعدی تک‌بعدی شده در مقایسه با سایر روش‌های همترازسازی در همه شرایط توزیع آزمودنی‌ها و همه روش‌های همترازسازی عملکرد بهتری داشته است و روش همترازسازی چند بعدی کامل در شرایط پیونددهی چندبعدی با روش NOP بدترین عملکرد را بر اساس سوگیری و خطای همترازسازی به همراه داشت. به نظر می‌رسد ناهمسویی نتایج این دو پژوهش، ناشی از عوامل مختلفی مانند روش‌های مختلف گردآوری داده‌ها، نوع متفاوت داده‌ها و استفاده از روش‌های پیونددهی پیش از همترازسازی در دو پژوهش باشد، چراکه در پژوهش ژانگ (۲۰۱۲) طرح گردآوری داده‌ها طرح گروه‌های لنگر بود و در پژوهش حاضر از طرح گروه‌های تصادفی داده‌ها استفاده شده است. در پژوهش ژانگ (۲۰۱۲) پیش از همترازسازی به دلیل استفاده از طرح لنگر، استفاده از روش‌های مختلف پیونددهی چندبعدی ضروری بود، اما در پژوهش حاضر به لحاظ استفاده از طرح گروه‌های تصادفی نیازی به پیونددهی نبود. افزون بر این، در پژوهش حاضر، داده‌های مورد مطالعه داده‌های واقعی بوده، حال آنکه در پژوهش ژانگ (۲۰۱۲) از داده‌های شبیه‌سازی در شرایط مختلف استفاده شده بود. بنابراین با توجه به نتایج پژوهش می‌توان گفت در شرایطی که طرح گردآوری داده‌ها، طرح گروه‌های تصادفی باشد و در شرایط داده‌های واقعی، بهینه‌ترین روش برای همترازسازی داده‌های چندبعدی، استفاده از روش همترازسازی نمره مشاهده شده چندبعدی کامل است و برای قضاوت در خصوص سایر شرایط و طرح‌ها باید کارایی این روش‌ها در داده‌ها و طرح‌های مختلف همترازسازی در آزمون‌های اجرا شده در کشور بررسی شود.

با توجه به یافته‌های پژوهش، اکنون با قطعیت بیشتری می‌توان ادعا کرد که استفاده از روش‌های چندبعدی همترازسازی در شرایطی که داده‌ها چندبعدی هستند، نتایج متفاوت‌تری نسبت به روش‌های معمول تک‌بعدی به دست می‌دهند. بنابراین، بعدیت

عامل مهمی در همترازسازی داده‌ها است و در همترازسازی داده‌ها باید ساختار داده‌ها به‌عنوان یک عامل تأثیرگذار مد نظر قرار گیرد. از این رو، پیشنهاد می‌شود در همترازسازی داده‌های آزمون‌هایی با مقیاس بزرگ که سرنوشت‌ساز بوده و با تحصیل و آینده شغلی افراد مرتبط است، ابتدا وضعیت داده‌ها از لحاظ تعداد ابعاد مورد سنجش بررسی شود. پس از تأیید شدن چندبعدی بودن داده‌ها، در درجه اول بهتر است از روش همترازسازی نمره مشاهده شده چندبعدی کامل استفاده شود و پس از آن استفاده از روش همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده مناسب‌ترین روش است. اما با توجه به تفاوت معنی‌دار روش‌های همترازسازی تک‌بعدی با روش بهینه همترازسازی، استفاده از روش‌های همترازسازی تک‌بعدی نظریه سؤال پاسخ در شرایطی که داده‌ها چندبعدی باشد، پیشنهاد نمی‌شود.

یکی از محدودیت‌های پژوهش، نبود معیار قابل مقایسه دقیق و متقن برای مقایسه نتایج بود. در این پژوهش طبق مطالعات بروسمان (۲۰۱۰)، لی و بروسمان (۲۰۱۲) و چوی (۲۰۱۹) که به این نتیجه رسیدند روش همصدک به اعتبار مقاوم بودن در برابر نقض مفروضه تک‌بعدی بودن، می‌تواند ملاک مقایسه قرار گیرد، این ملاک برای قضاوت و مقایسه نتایج استفاده شد. لی (۲۰۱۳) در پژوهش خود، روش همترازسازی نمره مشاهده شده چندبعدی کامل را به‌عنوان معیار مقایسه در نظر گرفت، اما چون این روش در ایران برای نخستین بار و در پژوهش حاضر مورد بررسی قرار گرفته شد، نمی‌توانست به‌عنوان ملاک مقایسه مدنظر قرار گیرد. از این رو، ابتدا باید اعتبار این روش در داده‌های موجود بررسی می‌شد. با توجه به یافته‌های این پژوهش، پیشنهاد می‌شود پژوهشگران بعدی با مدنظر قرار دادن ملاک روشی همچون روش همترازسازی چندبعدی کامل، وضعیت سایر روش‌های همترازسازی را بررسی کنند. همچنین پیشنهاد می‌شود پژوهشگران آینده به همترازسازی آزمون‌های چندبعدی با استفاده از روش‌های چندبعدی نمره واقعی اقدام کنند. یک رویکرد برای همترازسازی داده‌های چندبعدی توسط کیم (۲۰۱۸) بررسی شده است که این روش، روش همترازسازی نمره واقعی در آزمون‌های چندبعدی با ساختار ساده است، بهتر است این روش در کشور ما نیز روی داده‌های چندبعدی بررسی شود.

با توجه به اینکه امکان بررسی اثر عوامل مختلف بر همترازسازی تک‌بعدی و چندبعدی در این پژوهش نبود، پیشنهاد می‌شود در همترازسازی داده‌ها، اثر عواملی همچون ساختار ساده یا پیچیده بودن داده‌ها مد نظر قرار گیرد. همچنین لازم است با

توجه به عملکرد متفاوت روش‌های مختلف همترازسازی در نمره‌های بالا، از آنجا که یکی از دلایل احتمالی آن حجم نمونه اندک در نمره‌های بالا است، پژوهشگران بعدی اثر حجم نمونه بر روش‌های مختلف همترازسازی را بررسی کنند.

منابع

- ایزانلو، بلال؛ بازرگان، عباس؛ فرزاد، ولی‌الله؛ صادقی، ناهید؛ کاوسی، امیر (۱۳۹۳). تفکیک ابعاد متعامد از خوشه‌های سؤال بر اساس هشت روش تعیین بعد در داده‌های دوارزشی: مورد آزمون ریاضی رشته ریاضی فیزیک کنکور ۹۲-۹۱. فصلنامه اندازه‌گیری تربیتی، ۱۸(۵)، ۲۰۷-۲۴۰.
- رضوانی‌فر، شیرین (۱۳۹۱). همترازسازی نمرات دروس ریاضی و فیزیک رشته علوم تجربی آزمون کنکور سراسری سال‌های ۱۳۸۸ و ۱۳۸۹ براساس نظریه‌های کلاسیک و جدید اندازه‌گیری. پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی.
- شاطریان محمدی، فاطمه (۱۳۸۲). مقایسه سه روش همترازسازی همصداک هموار نشده نمره مشاهده شده *IRT* و نمره واقعی *IRT* در طرح گروه‌های نامعادل با سؤالات لنگر. پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی.
- لرد، فردریک، ام (۱۹۸۰). کاربردهای نظریه سؤال-پاسخ؛ ترجمه علی دلاور و جلیل یونسی (۱۳۹۱). تهران: انتشارات رشد.
- مقدم‌زاده، علی (۱۳۹۲). روش بهینه همترازسازی با توجه به ویژگی‌های بومی آزمونهای ملی ایران: مورد مطالعه آزمون تولیمو و آزمون‌های جامع کنکورهای آزمایشی سازمان سنجش آموزش کشور. رساله دکتری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی.
- واشقانی فراهانی، مریم (۱۳۸۰). کاربرد روش همترازسازی همصداک در معادل‌سازی نمرات آزمون‌های ورودی دانشگاه‌ها (کنکور ورودی سال ۱۳۸۷). پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues & Practice*, 22(3), 37-51.
- Akin Arikan, C. (2019). A Comparison of Kernel Equating Methods Based on Neat Design. *Eurasian Journal of Educational Research*, 82, 27-44.
- Béguin, A. A., & Hanson, B. A. (2001). Effect of noncompensatory multidimensionality on separate and concurrent estimation in IRT

- observed score equating. *annual meeting of the National Council on Measurement in Education, Seattle, WA.*
- Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory* [Doctoral dissertation, University of Iowa].
- Brossman, B. G., & Lee, W. C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement, 37*(6), 460-481.
- Cao, Y. (2008). *Mixed-format test equating: Effects of test dimensionality and common-item sets.* [Doctoral dissertation, University of Maryland].
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29.
- Champlain, A. F. (1996). The Effect of Multidimensionality on IRT True-Score Equating for Subgroups of Examinees. *Journal of Educational Measurement, 33*(2), 181-201.
- Chen, J. (2014). *Model selection for IRT equating of Testlet-based tests in the random groups design.* [Doctoral dissertation, university of Iowa].
- Choi, J. (2019). *Comparison of MIRT observed score equating methods under the common-item nonequivalent groups design.* [Doctoral dissertation, The University of Iowa].
- Dorans, N. J. & Holland, P. W. (2000). Population invariance and equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement, 37*, 281-306
- Fraser, C., & McDonald, R. P. (2012). NOHARM 4: A Windows program for fitting both unidimensional.
- Freeman, L. (2016). Assessing model-data fit for compensatory and non-compensatory multidimensional item response models using Vuong and Clarke statistics. [Doctoral Dissertation, University of Wisconsin-Milwaukee].
- Genz, F., Bretz, T., Hothorn, T., Miwa, X., Mi, F., Leisch, & F. Scheipl (2008). *Mvtnorm: Multivariate Normal and T Distribution.* URL <http://CRAN.R-project.org>. R package version 0.9-0.
- González, J., & Wiberg, M. (2017). *Applying test equating methods.* New York: Springer. doi, 10, 978-3.
- Han, T., Kolen, M., & Pohlmann, J. (1997). A comparison among IRT true-and observed-score equatings and traditional

- equipercentile equating. *Applied Measurement in Education*, 10(2), 105-121.
- Hanson, B., & Zeng, L. (2004). *PIE: A computer program for IRT equating*. (Windows Console Version, Revised by Z. Cui, May 20, 2004) [Manual]. Unpublished
- Hanson, B., & Zeng, L. (2004). *PIE: A computer program for IRT equating*. Iowa City, IA, US: CASMA.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Jasper, F. (2010). Applied dimensionality and test structure assessment with the START-M mathematics test. *International Journal of Educational & Psychological Assessment*, 6(1), 104-125.
- Kahraman, N. (2013). Unidimensional interpretations for multidimensional test items. *Journal of Educational Measurement*, 50(2), 227-246.
- Kim, S. Y. (2018). *Simple structure MIRT equating for multidimensional tests theory*. [Doctoral dissertation, University of Iowa].
- Kim, S. Y., Lee, W. C., & Kolen, M. J. (2020). Simple-Structure Multidimensional Item Response Theory Equating for Multidimensional Tests. *Educational & Psychological Measurement*, 80(1), 91-125.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York: Springer-Verlag.
- Lee, E., Lee, W.C., Brennan, R. L. (2014). *Equating Multidimensional Tests under a Random Groups Design: A Comparison of Various Equating Procedures*, Center for Advanced Studies in Measurement and Assessment, *CASMA Research Report*, 40, <http://education.uiowa.edu/sites/education.uiowa.edu/files/documents/centers/casma/publications/casma-research-report-40.pdf>
- Lee, E. (2013). *Equating multidimensional tests under a random groups design: a comparison of various equating procedures*. [Doctoral dissertation, The University of Iowa].
- Lee, S. H. (2007). *Multidimensional item response theory: A SAS MDIRT macro and empirical study of PIAT math test*. [Doctoral dissertation, University of Oklahoma].

- Li, Y. H. (1997). *An evaluation of multidimensional IRT equating methods by assessing the accuracy of transforming parameters onto a target test metric*. [Doctoral dissertation, University of Maryland].
- Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement*, 24(2), 115-138.
- Li, Y., Jiao, H., & Lissitz, R. W. (2012). Applying multidimensional item response theory models in validating test dimensionality: An example of K-12 large-scale science assessment. *Journal of Applied Testing Technology*, 13(2), 1-27.
- Lim, E. (2016). *Subscore equating with the random groups design*. [Doctoral dissertation, University of Iowa].
- Lim, E.; Lee, W. C. (2016). *Subscore Equating and Reporting*. Center for Advanced Studies in Measurement and Assessment, *CASMA Research Report*.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*, 1st Ed. Lawrence Erlbaum Associates.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, 8, 453-461 .
- Meng, Y. (2012). *Comparison of Kernel Equating and Item Response Theory Equating Methods*. [Doctoral dissertation, University of Massachusetts Amherst].
- Min, K. S., & Kim, J. P. (2003). A Comparison of Two Linking Methods for Multidimensional IRT Scale Transformations. *ACT*.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests*. [Doctoral dissertation, University of Iowa].
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*. New York, NY: Springer.
- Ricker, K. L. (2007). *The Consequence of Multidimensionality IRT Equating Outcomes Using a Common-Items Nonequivalent Groups Design*. [Doctoral dissertation, university of Alberta].
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25.

- Sansivieri, V., Wiberg, M., & Matteucci, M. (2017). A review of test equating methods with a special focus on IRT-based approaches. *Statistica*, 77(4), 329-352.
- Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational & Psychological Measurement*, 75(6), 954-978.
- Simon, M. K. (2008). *Comparison of concurrent and separate multidimensional IRT linking of item parameters*. [Doctoral Dissertation, University of Minnesota].
- Spence, P. D. (1996). *The effect of multidimensionality on unidimensional equating with item response theory*. [Doctoral dissertation, University of Florida].
- Suksuwan, S., Junpeng, P., Ngudgratoke, S., & Guayjarernpanishk, P. (2012). The Effect of Proportion Common Item's with Mixed Format Test on Multidimensional Item Response Theory Linking. *Procedia-Social & Behavioral Sciences*, 69, 1505-1511.
- Svetina, D., & Levy, R. (2012). An overview of software for conducting dimensionality assessment in multidimensional models. *Applied Psychological Measurement*, 36(8), 659-669.
- Venables, W. N. and Ripley, B. D. (2019). *Modern Applied Statistics with S-PLUS*. Third Edition.
- Von Davier, A. (Ed.). (2010). *Statistical models for test equating, scaling, and linking*. Springer Science & Business Media.
- Von Davier, A. A., Holland, P. W., & Thayer, D. T. (2003). *The kernel method of test equating*. Springer Science & Business Media.
- Wang, S., Zhang, M., & You, S. (2020). A Comparison of IRT Observed Score Kernel Equating and Several Equating Methods. *Frontiers in Psychology*, 11, 308.
- Wetzel, E., & Hell, B. (2014). Multidimensional Item Response Theory models in vocational interest measurement: An illustration using the AIST-R. *Journal of Psychoeducational Assessment*, 32(4), 342-355.
- Wiberg, M. (2018). equateIRT Package in R. *Measurement: Interdisciplinary Research and Perspectives*, 16(3), 195-202.
- Zhang, O. (2012). *Observed score and true score equating for multidimensional item response theory under nonequivalent group anchor test design*. [Doctoral Dissertation, University of Florida].