فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی
سال دهم، شماره ۳۰، بهار ۱۳۹۹، صص۲۶۲-۲۱۵

Journal of Educational Measurement & Evaluation Studies
Vol.10, No. 30, Summer 2020

شناسایی رویکرد بهینه همترازسازی در آزمون‌های چندبعدی[1]

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

تاریخ دریافت مقاله: ۹۸/۰۱/۰۶
تاریخ پذیرش مقاله: ۹۹/۰۶/۰۲

چکیده

همترازسازی از مهم‌ترین مسائل در نظام سنجش تربیتی است و عدم برقراری مفروضات آن، اندازه‌گیری را با چالش‌های جدی مواجه می‌کند. در آزمون‌های چندبعدی، استفاده از روش‌های تک‌بعدی همترازسازی، سوگیری نتایج را به دنبال دارد. هدف از اجرای این پژوهش، شناسایی رویکرد بهینه همترازسازی در آزمون‌های چندبعدی بود. شش روش همترازسازی تک‌بعدی و چندبعدی مقایسه شدند. روش همصدک به لحاظ مقاوم بودن در برابر مفروضه تک‌بعدی 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 همترازسازی انتخاب شدند. ابعاد آزمون‌ها با استفاده از نرم‌افزارهای MPLUS و NOHARM مشخص شد. برای برآورد پارامترهای تک‌بعدی از بسته ltm و برای برآورد پارامترهای چندبعدی از بسته mirt در نرم‌افزار R استفاده شد. همترازسازی نمره مشاهده شده و نمره واقعی تک‌بعدی IRT با استفاده از نرم‌افزار PIE و همترازسازی همصدک با استفاده از نرم‌افزار R بسته equate انجام گرفت. از بسته‌های mirt، mvnorm و MASS در نرم‌افزار R نیز برای همترازسازی چندبعدی داده‌ها استفاده شد. نتایج نشان داد که در همترازسازی آزمون‌های چندبعدی بهینه‌ترین روش، همترازسازی چندبعدی کامل است و پس از آن روش نمره مشاهده شده چندبعدیِ تک‌بعدی شده، روش مناسبی است، اما استفاده از روش‌های تک‌بعدی نمره مشاهده شده و نمره واقعی در چنین شرایطی، کارایی مناسبی ندارد. بنابراین پیشنهاد می‌شود در همترازسازی آزمون‌هایی که ساختار چندبعدی دارند، از روش همترازسازی نمره مشاهده شده چندبعدی کامل استفاده شود.

واژگان کلیدی: همترازسازی چندبعدی کامل، همترازسازی نمره واقعی چندبعدیِ تک‌بعدی شده، همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده، همترازسازی همصدک، نظریه پرسشِ پاسخ چندبعدی

---

[1]. 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 واندازه‌گیری 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 واندازه‌گیری، 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[1]، ۲۰۰۰؛ گلزانز و وایبرگ[2]، 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[3]، ۲۰۱۰)، تعدیل دشواری فرم‌های مختلف آزمون و قابل مقایسه کردن نمره‌های آنها، یک ضرورت در اندازه‌گیری به شمار می‌آید (منگ[4]، ۲۰۱۲؛ ژانگ[5]، ۲۰۱۲؛ چن[6]، ۲۰۱٤؛ وایبرگ[7]، ۲۰۱۸).

با روی کار آمدن آزمون‌های با مقیاس بزرگ در دهه‌های ۱۹۳۰تا 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[8]) روش‌های جدیدی برای همتراز کردن داده‌ها را به وجود آورد (وون داویر، هالند و ثایر[9]، 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 Y همتراز می‌شوند. اما در عمل، نمره‌های فرم آزمون‌ها نه‌تنها 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[۱]،۲۰۲۰). همترازسازی همصدک رایج‌ترین نوع همترازسازی نظریه کلاسیک اندازه‌گیری است که در آن توزیع نمره‌ها در دو فرم همترازشده یکسان است (سنسیوری، وایبرگ و ماتیوکی[۲]، ۲۰۱۷). روش همترازسازی کرنل رویکرد منحصربه‌فردی به همترازسازی آزمون است و دارای پنج مرحله پیش‌هموارسازی، برآورد احتمال نمره‌ها، پیوسته‌سازی[۳]، همترازسازی و خطای استاندارد همترازسازی است (آکین آریکان[٤]، 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[٥] برقرار می‌شود (بروسمان[٦]، ۲۰۱۰) و ماهیت همترازسازی نمره مشاهده شده نظریه سؤال پاسخ همان همترازسازی همصدک است و الگوریتم توزیع دوجمله‌ای را در برمی‌گیرد (لرد و وینگرسکی[۷]، ۱۹۸٤: 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 این‌رو 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]، ۱۹۹٦). 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 (ICC)[1] 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، ۲۰۱٤).

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، جیائو و لیسیتز[2]، ۲۰۱۲؛ کاهرامان[3]، 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 چندبعدی است که وابستگی موضعی سؤال را در نظر می‌گیرند (چن، ۲۰۱٤). مدل‌های نظریه سؤال پاسخ چندبعدی[1](MIRT) برای مدل‌سازی روابط ابعاد چندگانه توانایی شخص و پارامترهای چندگانه سؤال‌ها به وجود آمده‌اند و احتمال پاسخ درست به سؤال‌های چندبعدی 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[2]، ۲۰۰۹).

مطالعه پیشینه نشان داد که مدل‌های چندبعدی نظریه سؤال پاسخ (MIRT) برای داده‌های شبیه‌سازی شده و داده‌های واقعی، دقت بیشتری نسبت به داده‌های تک‌بعدی دارند (لی و لیسیتز، ۲۰۰۰؛ بروسمان، ۲۰۱۰) و روایی کاربرد مدل‌های تک‌بعدی در همترازسازی، برازش مدل، برآورد پارامتر، نمره‌گذاری و سنجش انطباقی در موقعیتی که فضای چندبعدی معقول و منطقی است، مورد تردید واقع شده است (شو و وایز، ۲۰۱۵). با این وجود، ژانگ (۲۰۱۲) نشان داد که 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 و روش همترازسازی چندبعدی کامل بدترین عملکرد را بر اساس سوگیری و خطای همترازسازی به همراه داشت. سایر مطالعات نشان‌دهنده آن است که همترازسازی نمره واقعی چندبعدی ساختارساده SMT[3] در مقایسه با روش‌های معمول تک‌بعدی دقت همترازسازی بیشتری فراهم می‌کند و مفید بودن رویکرد چندبعدی در همترازسازی داده‌های چندبعدی نشان داده شده است (کیم[4]، ۲۰۱۸؛ پترسون[5]، ۲۰۱٤). در شرایط نقض مفروضه تک‌بعدی، روش همترازسازی نظریه سؤال پاسخ چندبعدی نسبت به روش تک‌بعدی بیشتر شبیه همترازسازی همصدک عمل می‌کند و روش‌های همترازسازی چندبعدی جایگزین مناسبی برای روش‌های تک‌بعدی هستند (بروسمان، ۲۰۱۰)، همچنین می‌توان از مدل‌های همترازسازی چندبعدی برای داده‌های تک‌بعدی

---

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[1] ، ۲۰۰۱). اما در پژوهش لی (۲۰۱۳) و لی، لی و برینان (۲۰۱٤) 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 چندبعدیِ تک‌بعدی شده[2]، روش همترازسازی نمره واقعی چندبعدیِ تک‌بعدی شده[3] و روشِ همترازسازی نمره مشاهده شده چندبعدی کامل[4] را مقایسه کرد. بنابراین استفاده از روش‌های چندبعدی همترازسازی در سطح جهان نو بوده و پس از بروسمان، پژوهشگران اندکی همچون لی[5] (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 چندبعدی 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 چندبعدی بود تا مشخص شود بهترین شیوه برای همترازسازی چنین داده‌هایی کدام است. بنابراین از میان تمام روش‌های موجود، ترکیبی از روش‌های تک‌بعدی و روش‌های نوین چندبعدی مدنظر قرار گرفت. در این پژوهش، روش همصدک به دلیل اینکه به روشنی مفروضات تک‌بعدی 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 چندبعدی

مدل‌های سنتی نظریه سؤال پاسخ تک‌بعدی 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 چندبعدی و پیچیده در عملکرد آزمون نقش دارند (مین و کیم[1]، ۲۰۰۳؛ آکرمان، گیرل و والکر[2]، 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 چندبعدی راه حل مناسبی را فراهم کرده است (ریکاس، ۲۰۰۹؛ واتزل و هل، ۲۰۱٤؛ کیم، ۲۰۱۸؛ کیم، لی و کولن[3]، ۲۰۲۰).

مدل‌های نظریه سؤال پاسخ چندبعدی، یک بیان ریاضیاتی برای نشان دادن رابطه جایگاه افراد در فضای چندبعدی 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 گوایجاررن پانیشک[1]، 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:

$$P\left(U_{ij}=1\middle|\theta_{j,},a_i,d_i\right)=\frac{e^{a_i\dot{\theta}_j+d_i}}{1+e^{a_i\dot{\theta}_j+d_i}}$$

که U_{ij} اشاره به پاسخ شخص j به سؤال i، a بردار m×1 برای پارامترهای تمیز، Θ بردار m×1 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).

[1]. Suksuwan, Junpeng, Ngudgratoke & Guayjarernpanishk

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 مشاهده‌شده چندبعدی، گسترشی از همترازسازی نمره مشاهده‌شده تک‌بعدی است و با استفاده از سه فرایند به شرح زیر قابل محاسبه است.

برای انجام همترازسازی چندبعدی کامل، ابتدا توزیع شرطی نمره مشاهده‌شده (مانند $f(x\ |\theta_j)$) در هر سطح ترکیب توانایی (θ_j) با استفاده از یک فرمول بازگشتی[1] مانند الگوریتم لرد- وینگرسکی[2] 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:

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

$$= f_{g-1}(x\ |\theta_j)(1 - P_{jg}) + f_{g-1}(x - 1\ |\theta_j)\, P_{jg} \quad , 0 < x < g$$

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

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:

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

پس از این گام، n-1 تکرار برای محاسبه توزیع شرطی نمره مشاهده شده $f_g(x \mid \theta_j)$ 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 سنتی 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[1] (۱۹۹۹) و ژانگ[2] و ونگ (۱۹۹۸) گرفتند که مطالعات آنها مبتنی بر نظریه سؤال پاسخ چندبعدی وتوزیع 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. $\mathrm{P}(\theta)$ 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 استانداردی 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 چندبعدی 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 تک‌بعدی سؤال به دست می‌آید، برای برآورد تمیز تک‌بعدی از فرمول زیر استفاده می‌شود:

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

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

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

در فرمول فوق، $\mathrm{d_{iw}}$ پارامتر دشواری چندبعدی است ($\mathrm{d_{iw}} = (1 + \sigma_{\mathrm{iw}}^2)^{\frac{1}{2}}$).

در پایان این مراحل، پارامترهای سؤال تک‌بعدی با استفاده از پارامترهای سؤال چندبعدی درجه‌بندی شده و وزن‌های مرتبط با بهترین جهت اندازه‌گیری آزمون محاسبه می‌شوند. با در دست داشتن پارامترهای تک‌بعدی شده سؤال و توانایی می‌توان هر دو نوع همترازسازی نمره واقعی و نمره مشاهده شده را مانند آنچه در همترازسازی تک‌بعدی 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 ۱۳۹٦ و ۱۳۹۷ است که تعداد آنها در سال ۱۳۹٦ برابر با ۱٤۸٤۲۹ و در سال ۱۳۹۷ برابر با ۱٤٤٤۳۷ نفر بوده است. پاسخ تعداد ۵۰۰۰ نفر از داوطلبان به شیوه تصادفی ساده از بین تمامی داده‌های مربوط به آزمون ریاضی کنکور سراسری سال ۱۳۹٦ و ۱۳۹۷ انتخاب شدند. آزمون ریاضی کنکور سراسری در هر دو سال دارای ۵۵ سؤال بود. طرح گردآوری داده‌ها طرح گروه‌های تصادفی[1] 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[2] 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[3] باشد، و توزیع توانایی از توزیع نرمال استاندارد چندمتغیره با همبستگی صفر بین ابعاد پیروی کند، نیازی به روش‌های پیوندهی پیش از همترازسازی نیست (لی، لی و برینان، ۲۰۱٤). 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[4]) و هم در واحد اندازه‌گیری (دایلیشن[5]) 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[6]، 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. Orthognal
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[1]، ۲۰۱۲) به دلیل مناسب بودن این نرم‌افزار در تعیین ابعاد داده‌های چندبعدی و مفید بودن آن برای فقط تحلیل داده‌های دوارزشی (اسوتینا و لوی[2]، ۲۰۱۲) استفاده شد. همچنین برای اطمینان بیشتر در تعیین ابعاد آزمون از نرم‌افزار MPLUS8 (موشن و موشن[3]، ۲۰۱۷–۱۹۹۷) به لحاظ استفاده از روش WLSMV برای برآورد و مناسب بودن برای داده‌های دوارزشی استفاده شد (جاسپر[4]،۲۰۱۰؛ به نقل از ایزانلو، بازرگان، فرزاد، صادقی و کاوسی، ۱۳۹۱). تجزیه‌وتحلیل داده‌ها با استفاده از نرم‌افزار R (تیم توسعه R، ۲۰۱۹) و بسته‌های ltm (ریزوپولوس[5]، ۲۰۰۶)، mirt (چالمرز[6]، ۲۰۱۲)، MASS (ونبلز و ریپلی[7]، ۲۰۱۹) و mvtnorm (جنز و همکاران[8]، ۲۰۰۸) و نرم‌افزار PIE (هنسون و زنگ[9]، ۲۰۰۴) انجام گرفت.

برای همترازسازی داده‌ها با استفاده از روش همترازسازی تک‌بعدی نمره مشاهده شده و نمره واقعی پس از برآورد پارامترهای تک‌بعدی سؤال‌ها و توانایی در بسته ltm، پارامترها برای همترازسازی وارد نرم‌افزار PIE شدند. برای همترازسازی نمره مشاهده شده و نمره واقعی چندبعدی تک‌بعدی شده IRT، ابتدا پارامترهای چندبعدی 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 تتای چندبعدیِ تک‌بعدی شده، تمامی پارامترهای چندبعدیِ تک‌بعدی شده (a، b و

---

1. Fraser & McDonald
2. Svetina & Levy
3. Muthén & Muthén
4. Jasper
5. Rizopouos
6. Chalmers
7. Venables & Ripley
8. Genz et al
9. Hanson & Zeng

Θ) وارد نرم‌افزار PIE شدند. در نهایت، برای همترازسازی نمره مشاهده شده چندبعدی 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 چندبعدی تک‌بعدی شده IRT، همترازسازی نمره مشاهده شده چندبعدی تک‌بعدی شده و همترازسازی نمره مشاهده شده چندبعدی 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[1] (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 مد نظر 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.

[1]. Differences That Matter (DTM)

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

دواطلبان	میانگین	انحراف استاندارد	کمینه	بیشینه	کجی		کشیدگی	
					آماره	خطا	آماره	خطا
سال ۱۳۹۷	۷/۸۲	۷/۳۱	۱	۵۱	۱/۹۹	۰/۰۳	٤/٦٦	۰/۰٦
سال ۱۳۹٦	۷/٦٦	۷/٦۳	۱	۵۲	۱/۹۰	۰/۰۳	٤/۰۲	۰/۰٦

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 ۱۳۹٦ برابر با ۷/٦٦ و ۷/٦۳ 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 مجذورات باقی‌مانده‌ها برابر با ۰/۰۰۹۰ و ریشه دوم میانگین مجذورات باقی‌مانده‌ها (RMSR[1]) برابر با ۰/۰۰۲٦ 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[2] برابر با ۰/۹۹۱ و بیشتر از ۰/۹۵ است که نشان‌دهنده برازش مناسب مدل سه‌بعدی با داده‌ها است (هوپر، کاقلان و مولن[3]، 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 ۰/۰۰٤ 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	RMESA	TLI	CFI	df	X^2	X^2/df
یک‌بعدی	۰/۰۸۳	۰/۰۳۵	۰/۸٦	۰/۸۷	۱۲۷٤	۹۲٥٦/٦٤	۷/۲٦
دوبعدی	۰/۰٦۰	۰/۰۲٦	۰/۹۲	۰/۹۳	۱۲۲۳	٥۳٥٤/۲۰	٤/۳۷
سه‌بعدی	۰/۰۳۹	۰/۰۱۸	۰/۹٦	۰/۹۷	۱۱۷۳	۳۰٥۹/۳٦	۲/٦۰

جدول (۲) نشان می‌دهد که شاخص‌های برازش سه‌بعدی از مقادیر مناسبی برخوردار هستند، مقدار X^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 سه‌بعدی 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 ٤۱ 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	۰/۳٦٦	۰/۳۳۳	۲۹	۰/۵۹۰	۰/۷٦۸	۰/۲۷۷
۲	۰/۲۱۸	۰/۵۸۵	-۰/۱۹٤	۳۰	۰/۲۵۷	۰/۵۷۵	۰/۰۵٤
۳	۰/۷۲۲	۰/۱۲۳	۰/۳۵۵	۳۱	۰/۱۹۵	۰/٦۲۵	-۰/۰۷۲
٤	۰/٦۹۹	۰/٤٤۱	۰/۳۸٦	۳۲	۰/۵۰۹	۰/۵۲٦	۰/۳۸۰
۵	۰/٦۲٤	۰/۲۸۷	۰/۲۹۱	۳۳	۰/۳۱٤	۰/٤۸۰	۰/٤٦۰
٦	۰/٦۳۹	۰/٤٤۲	۰/۳۳۸	۳٤	۰/۷۰۱	۰/٤۱۰	۰/۳۸۲
۷	۰/۵۲۰	۰/٤٦۱	۰/۲۳٦	۳۵	۰/٦۹۹	۰/۳۰۸	۰/۳۹۵
۹	۰/٤٦۲	۰/٤۳٦	۰/۰۳۰	۳٦	۰/۳۲۵	۰/٤۳۷	۰/۳۲۵
۱۰	۰/۷۱۰	۰/۲۷۷	۰/۲۷٤	۳۷	۰/٦۵٤	۰/٤۸۷	۰/۲۱۸
۱۱	۰/۵٤۱	۰/٤۰۰	۰/۰٤۸	۳۸	۰/۷۱۱	۰/۵۳٦	۰/۳۱۷
۱۲	۰/۷۹۲	۰/۲۳۷	۰/۲۲۰	۳۹	۰/٦۱۲	۰/۳۱۰	۰/۳۹٤
۱٤	۰/٦۷۹	۰/٦۸۰	۰/۲۵۵	٤۰	۰/٦٦۳	۰/۵۱۳	۰/۳۹۰
۱۵	۰/۵۵۱	۰/۱٤۳	۰/۲۲۵	٤۲	۰/۱۵۳	۰/٤۱۵	-۰/۰۵۰
۱٦	۰/۸۳۵	۰/۱٦۰	۰/۲۷۵	٤۳	-۰/۰٤۱	۰/۰۱۳	۰/۵۵۸
۱۷	۰/۳۱٦	۰/٤۲۰	-۰/۰۳۳	٤٤	۰/٤۸۰	۰/۲۲۱	۰/٦۱۹
۱۸	۰/٦۷۰	۰/٤٦۸	۰/۲۵۰	٤۵	۰/۳۵۲	۰/۲۵٤	۰/۵۲۵
۱۹	۰/۷۷٦	۰/۳٤۰	۰/۲۱۲	٤٦	۰/٤۸۸	۰/۵٤۸	۰/٤۹۸
۲۰	۰/٦۳۱	۰/۳۰۹	۰/۰۹۱	٤۷	۰/۳۰۱	۰/۳۰۸	۰/۵۰۳
۲۱	۰/۲۵۹	۰/۳۰۹	۰/۰۸۱	٤۸	۰/۵۱۵	۰/٤٦٦	۰/٤۲۲

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، کاقلان 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 mplus استفاده شد. مقادیر برازش یک‌بعدی، دوبعدی و سه‌بعدی در جدول (٤) ارائه شده است.

جدول (٤) شاخص‌های برازش مدل یک، دو و سه‌بعدی با داده‌های آزمون ریاضی کنکور سال ۱۳۹٦

تعداد ابعاد	SRMR	RMESA	TLI	CFI	X^2	df	X^2/df
یک بعدی	۰/۰۷۲	۰/۰۳۰	۰/۹۱	۰/۹۲	٦۸٦۵/۸٦	۱۲۷٤	۵/۳۸
دو بعدی	۰/۰٤٦	۰/۰۲۱	۰/۹٦	۰/۹٦	۳۸۱۲/٦۲	۱۲۲۳	۳/۱۱
سه بعدی	۰/۰۳۸	۰/۰۱۵	۰/۹۷	۰/۹۸	۲۵۵۵/۳۸	۱۱۷۳	۲/۱۷

شاخص‌های برازش مدل در جدول (٤) نشان‌دهنده برازش مناسب داده‌های آزمون ریاضی سال ۱۳۹٦ 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 ۱۳۹٦ در جدول (۵) آمده است. سؤال‌های ۲۸، ٤۱ و ٤۷ به دلیل نداشتن بار عاملی مناسب از تحلیل حذف شدند.

جدول (۵) مقادیر بار عاملی مربوط به سه بعد آزمون ریاضی سال ۱۳۹٦

سؤال	عامل اول	عامل دوم	عامل سوم	سؤال	عامل اول	عامل دوم	عامل سوم
۱	۰/٤۹٦	۰/۵۹۲	۰/۱۹٤	۲۷	۰/٤۷۱	۰/٦۹۵	۰/۳۳٤
۲	۰/۵٦٦	۰/۳٦۱	۰/۳۰۲	۲۹	۰/۱۲۱	۰/۵۷٦	-۰/۰۱۱
۳	۰/٦۸٦	۰/۵۳۰	۰/۳۹۵	۳۰	۰/۵۱۱	۰/۷۱۲	۰/٤۲۳
٤	۰/۷۰۱	۰/۵۳٤	۰/۳۳٦	۳۱	۰/۱۲۹	۰/۳۵۲	۰/۱۵۳
۵	۰/٦۹۵	۰/٦۸۵	۰/۳٦٤	۳۲	۰/۲۷۳	۰/٦۱۸	۰/۲۷۳
٦	۰/۷۲۹	۰/٤۷۲	۰/۳٦۷	۳۳	۰/۵۹۹	۰/٦٤۸	۰/٤۹۲
۷	۰/۵۷۷	۰/٤٤۲	۰/۳۱۳	۳٤	۰/٤۰۲	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، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۸، 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 چندبعدی

مدل‌های چندبعدی دوپارامتری برای هر بعد، دارای یک پارامتر شیب (تشخیص) هستند که همانند پارامتر تشخیص در مدل‌های یک‌بعدی تفسیر می‌شود، با این تفاوت که پارامتر شیب در مدل یک‌بعدی در یک جهت (صفت مورد اندازه‌گیری) تمیز می‌دهد ولی در چندبعدی 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)

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

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

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

در جدول (٦) پارامترهای چندبعدی آزمون ریاضی در دو سال ۱۳۹٦ و ۱۳۹۷ ارائه شده است.

جدول (٦) پارامتر سؤال‌های آزمون ریاضی سال ۹٦ و ۹۷ در مدل ۳ عاملی

سال ۱۳۹٦						
سؤال‌ها	MDIFF	MDISC	α_1	α_2	α_3	d
۱	۲/٤۰٤	۱/٤۷۳	۰/٥٥٦	−۱/۱۲۲	۰/۸٥۹	−۳/٥٤٤
۲	۱/۳۳۸	۱/۰۷۳	−۰/۰۱٤	۰/۰٥	۱/۰۹٥	−۱/٤۳۷
۳	۱/۲۲٤	۱/٦۳۷	−۰/۰٤۸	−۰/۲۷٦	۱/٤۲٤	−۲/۰۰٥
٤	۱/۱۷۰	۱/٦۹۹	۰/۲٥۸	−۰/٤۰۹	۱/٥۷٥	−۱/۹۸۹
٥	۱/٥۸٦	۲/۱۱۳	۰/۳۳٦	−۱/۱۱	۱/٤۷۲	−۳/۳٥۳
٦	۱/۲۱۰	۱/۷٥۹	۰/۱۱٥	−۰/۰۲۱	۱/۸۱۳	−۲/۱۲۹
۷	۱/٥۹۱	۱/۱۹۲	۰/۰۲٥	−۰/۱۸۱	۱/۰۹	−۱/۸۹۸
۸	۰/۸۹۲	۱/۸٤۳	−۰/۰۱۱	۰/۰٤۷	۱/۸٦٥	−۱/٦٤٦
۹	۱/۷۱۲	۱/۳۰۱	۰/۳۷۱	−۰/٤۲٥	۱/۲۱٤	−۲/۲۳
۱۰	۲/۲۸٥	۱/٥۰۸	۰/٤۷۳	−۱/۰٥٦	۰/۹۲٦	−۳/٤٤۷
۱۱	۰/۹۹٤	۱/۹٦٥	−۰/۲۷۹	۰/٦٤۹	۲/۱۳۳	−۱/۹٥٥
۱۲	۰/۹۲۱	۲/٥۷۷	−۰/۲۲٤	۰/٦۲	۲/۷۷۹	−۲/۳۷٥
۱۳	۰/۰۲۷	۰/۸۱۱	−۰/۳۹	۰/٥۸۷	۰/۷۹۲	−۰/۰۲۲
۱٤	۰/٥۸٥	۱/٦۹۷	−۰/۳۸۷	۰/۲۰٦	۱/٥٦٦	−۰/۹۹٤
۱٥	۱/۷۷٤	۱/٦۹۳	−۰/۱٤٥	−۰/۳۹٤	۱/۳۲٦	−۳/۰۰٥
۱٦	۱/۱٥٤	۱/۹۳۳	−۰/٤۲۸	۰/۲۰۲	۱/۷۷٤	−۲/۲۳۳
۱۷	۱/٦۸۷	۱/٥٤۰	−۰/۱۰٦	−۰/۲۷٦	۱/۲۸۷	−۲/٦
۱۸	۰/۸٤۱	۲/۸۳۳	−۰/۳٥۳	۰/٤۹	۲/۸۹۳	−۲/۳۸٥
۱۹	۱/٤٥۳	۲/۲۸٥	−۰/۳۹٥	۰/۰۲٥	۲/۰٤۳	−۳/۳۲۱
۲۰	۱/۰٥۹	۲/۳۳۱	−۰/۳۸۱	۰/٦۱٦	۲/٤۲۹	۲/٤۷۱
۲۱	۲/۰٦٥	۱/٦۷۸	۰/۰۳٥	−۰/۹٥٦	۰/۹٤۱	−۳/٤٦۷
۲۲	۲/٤٦۲	۱/٤۹۹	۰/۳۹٤	−۰/۹٤	۰/۹۹	−۳/٦۹۲
۲۳	۱/۸۱٥	۱/۸۸٤	۰/۱۳۳	−۰/۷۹٥	۱/۳۸۲	−۳/٤۲۱
۲٤	۱/٦۹٤	۱/۲۲۰	−۰/۱۰٥	−۰/۰۱٥	۱/۱٤٥	−۲/۰٦۸
۲٥	۲/۱۰۲	۱/٦٦۸	۰/٥٤۸	−۱/۹۳	۰/۰۰۳	−۳/٥۰۸
۲٦	۲/۷۰۳	۱/٦٦۸	۰/۷۰٦	−۲/۰۰۳	−۰/۰۳۹	−٤/٥۱۱

سال ۱۳۹۶						
سؤال‌ها	MDIFF	MDISC	α_1	α_2	α_3	d
۲۷	۱/۵۶۱	۱/۷۰۲	−۰/۰۹۸	−۱/۵۳۶	۰/۱۶۳	−۲/۶۵۸
۲۹	۲/۱۵۲	۱/۷۰۵	۰/۵۷۱	−۲/۲۲۹	−۰/۶۴۱	−۳/۶۷۱
۳۰	۲/۱۲۵	۱/۷۶۲	−۰/۳۴۴	−۱/۴۰۶	۰/۱۹۱	−۳/۷۴۷
۳۱	۲/۱۱۸	۰/۶۴۳	−۰/۱۸۱	−۰/۶۹۲	−۰/۳۴۳	−۱/۳۶۳
۳۲	۲/۴۶۶	۱/۵۶۶	−۰/۲۶۲	−۱/۶۹۹	−۰/۵۶۷	−۳/۸۶۴
۳۳	۱/۹۹۹	۱/۶۸۱	−۰/۵۱۷	−۰/۶۹۸	۰/۷۴۵	−۳/۳۶۲
۳۴	۲/۸۴۸	۱/۳۲۲	−۰/۳۲	−۱/۰۲۹	۰/۱۱۳	−۳/۷۶۶
۳۵	۳/۲۴۵	۱/۳۳۲	−۰/۰۶	−۱/۳۳۳	−۰/۰۶	−۴/۳۲۵
۳۶	۳/۱۱۵	۱/۴۰۰	−۰/۱۹۷	−۱/۱۸۳	۰/۱۳۸	−۴/۳۶۳
۳۷	۰/۹۲۱	۲/۰۵۷	−۲/۰۴۴	۰/۶۶۹	۰/۵۵۷	−۱/۸۹۶
۳۸	۱/۲۲۰	۱/۹۰۸	−۲/۰۶۸	۰/۲۶۹	−۰/۰۲۵	−۲/۳۲۹
۳۹	۱/۳۸۴	۱/۵۷۳	−۱/۰۴۴	۰/۲۷۱	۰/۴۷۶	−۲/۱۷۹
۴۰	۱/۷۲۱	۱/۷۲۳	−۱/۶۵۲	۰/۳۰۶	۰/۳۷۶	−۲/۹۶۸
۴۲	۲/۴۷۴	۱/۵۲۴	−۰/۱۸۵	−۱/۲۴۹	۰/۲۳۱	−۳/۷۷۲
۴۳	۲/۲۸۳	۱/۳۲۸	−۰/۳۷۵	−۰/۹۹۱	۰/۱۱۵	−۳/۰۳۳
۴۴	۱/۵۰۸	۱/۳۷۳	−۰/۶۴۲	−۰/۷۷۴	۰/۱۵۵	−۲/۰۷۲
۴۵	۱/۷۲۷	۱/۲۷۷	−۰/۸۶۸	−۰/۲۸۳	۰/۳۰۲	−۲/۲۰۷
۴۶	۲/۹۲۰	۱/۴۰۳	−۰/۲۷۳	−۱/۴۰۱	−۰/۲۹۹	−۴/۰۹۹
۴۸	۲/۸۴۸	۱/۱۰۹	−۰/۴۷۲	−۱/۰۰۹	−۰/۱۲۹	−۳/۱۶۲
۴۹	۱/۴۰۶	۰/۸۸۵	−۰/۸۲۸	−۰/۲۱۸	−۰/۱۴۵	−۱/۲۴۶
۵۰	۱/۰۱۰	۱/۱۹۳	−۱/۱۵۹	−۰/۱۵۹	−۰/۱۰۸	−۱/۲۰۷
۵۱	۲/۰۹۸	۱/۱۷۲	−۰/۷۳	−۰/۶۰۴	−۰/۰۲۷	−۲/۴۶
۵۲	۲/۹۰۳	۱/۳۴۹	−۰/۱۱۹	−۱/۴۸	−۰/۳۸	−۳/۹۱۷
۵۳	۳/۰۹۲	۱/۳۱۹	−۰/۱۸۹	−۱/۴۵۱	−۰/۴۸	−۴/۰۸۱
۵۴	۲/۳۶۱	۱/۵۱۷	−۰/۷۰۶	−۰/۹۱۹	۰/۰۹۵	−۳/۵۸۳
۵۵	۱/۸۳۲	۱/۲۶۵	−۰/۹۷۲	−۰/۱۵۵	۰/۲۷۸	−۲/۳۱۹

سال ۱۳۹۷						
سؤال‌ها	MDIFF	MDISC	α_1	α_2	α_3	d
۱	۱/۸۲۰	۱/۱۲۹	−۰/٦۹٦	۰/۱۸۷	۰/٤٦۸	−۲/۰۵۷
۲	۲/۵۳۸	۱/۹۲۷	۰/۱۲٤	۲/۱۵۵	−۰/٦۳۱	−٤/۸۹۳
۳	۰/۵٤۹	۱/۹٦۵	−۲/۰۳۲	−۰/۷۸۳	۰/٤٤۱	−۱/۰۸۱
٤	۱/۲۸۷	۱/٦۸٤	−۱/۱۹	۰/۲٤۸	۰/۵٤۱	−۲/۱۷
۵	۱/۰۷۲	۱/۲۸۸	−۱/۱٤۳	−۰/۰۵۲	۰/۲۹۵	−۱/۳۸۲
٦	۱/٦۰۲	۱/٤۳٦	−۰/۹۳۱	۰/۳٤۷	۰/٤٤۲	−۲/۳۰۲
۷	۲/۱۳۸	۱/۱٦۰	−۰/٦۰۲	۰/۵٦	۰/۲٤۱	−۲/٤۸۳
۹	۱/۲۱٦	۱/٦۸٦	−۱/٦۹	−۰/۱۸۸	۰/۱٦۹	−۲/۰۵۲
۱۰	۲/۳٦۲	۱/٤۰۳	−۱/۱٦۳	۰/۵۹۷	−۰/۳٦٤	−۳/۳۱٦
۱۱	۰/۸۹۹	۲/۳۱٤	−۲/۵۵٦	−۰/٤٤۲	−۰/۰۸۳	−۲/۰۸۲
۱۲	۲/۱۷۹	۲/۲۲۹	−۱/۱۲۷	۱/٤۰٦	۰/۰۲۲	−٤/۸٦
۱٤	۰/۸۸۲	۱/۰۵۷	−۱/۱۱۳	−۰/۳۵۸	۰/۱۷۳	−۰/۹۳۳
۱۵	۰/٦۵٤	۲/۹۰٦	−۳/۲۵۸	−۰/۹۷۷	۰/۰۹۸	−۱/۹۰۱
۱٦	۲/۹۹٦	۱/۱۸۱	−۰/٤۸۸	۰/۹۹۹	−۰/۳۷۳	−۳/۵٤۱
۱۷	۲/۲۱۰	۱/۷۰۷	−۱/۳۸۸	۰/٤۷٦	۰/۰۳۳	−۳/۷۷٤
۱۸	۱/۵۱٦	۲/۲۹۱	−۲/۳۵۸	۰/۰۷	−۰/۲۲٤	−۳/٤۷٤
۱۹	۱/٦۷٦	۱/۵۰۳	−۱/۵۵۱	۰/۱۵۱	−۰/۳۰۲	−۲/۵۱۹
۲۰	۲/۳۸۲	۱/۳۱۹	−۱/۲۲٦	۰/۳٤۹	−۰/۲۷	−۳/۱٤٤
۲۱	۲/٤۸٦	۱/۹۰۷	−۰/۷۸۹	۱/٤٦٤	−۰/۱۹۱	−٤/۷٤۲
۲۲	۲/۳۸۳	۲/۰٦۲	−۱/۵۸۳	۰/۹۰۹	−۰/۲۹٤	−٤/۹۱٦
۲۳	۱/۱۱٦	۲/٤۹۲	−۲/٦۷	−۰/۲۳۹	−۰/۱۲٦	−۲/۷۸۳
۲٤	۱/٤۹۷	۲/۰۵۲	−۲/۲۹۹	−۰/۱٦۷	−۰/۳۹٤	−۳/۰۷۳
۲۵	۱/۸۸۳	۱/٤۸۳	۰/۳٤۵	۱/۱٦٦	۰/۸۰۱	−۲/۷۹٤
۲٦	۱/۸۲۹	۱/۵۳۹	۰/٦۹۵	۱/۳۷۸	۰/۷۹۹	−۲/۸۱۷
۲۷	۲/۰٦۹	۱/۹۹۰	−۰/۱٦۸	۱/٤۲٦	۰/۷۷۲	−٤/۱۱۸
۲۸	۲/۰۲۲	۲/۳۰٦	−۰/۲٤۲	۱/۹٦	۰/٤۰۳	−٤/٦٦٦
۲۹	۲/۸۹۳	۱/۳۹۸	۰/٤۰٤	۱/۵۲۵	۰/۱۱۸	−٤/۰٤۵

سال ۱۳۹۷						
سؤال‌ها	MDIFF	MDISC	α_1	α_2	α_3	d
۳۰	۲/۲۹۳	۱/۵۹۳	۰/۵۵٤	۱/۸۵٦	−۰/۱٤۷	−۳/٦۵٤
۳۱	۱/۷۸۳	۱/۲۹۸	−۰/۲۰۳	۰/٦۷۳	۰/۷۱	−۲/۳۱٦
۳۲	۱/۷۷۰	۱/۲٦٦	۰/٤٦	۰/۵۸۵	۱/۱٤۱	−۲/۲٤۲
۳۳	۱/٦۳۹	۱/٦٦۳	−۱/۲٦۹	۰/۱۸۱	۰/٤٦٦	−۲/۷۲۸
۳٤	۱/۱٤٦	۱/۵۸۸	−۱/۳۳۳	−۰/۱۷۵	۰/۵۵۳	−۱/۸۲۲
۳۵	۱/۷٦۹	۱/۸۷۱	−۱/٤۷٦	۰/٤۳۵	۰/۲۲۱	−۳/۳۱۱
۳٦	۲/۵۱۲	۱/۸٤٦	−۱/٤٤۸	۰/۷۰۹	−۰/۱٤۷	−٤/٦۳۹
۳۷	۱/۸۷۱	۱/۸۸۷	−۱/۲۷۳	۰/٦۹۳	۰/۲۵	−۳/۵۳۲
۳۸	۲/۸۷۱	۰/٦۲٦	۰/۰۲۷	۰/۰۸۲	۰/٦۰۲	−۱/۷۹۸
۳۹	۱/۵۷۱	۱/۳۲۰	−۰/۹۷۷	−۰/۱۲۱	۰/٦۰۷	−۲/۰۷۵
٤۰	۲/۲۱۹	۱/٦۵۳	−۱/۰۱۱	۰/٤٦۷	۰/۵۱۹	−۳/٦۷۱
٤۲	٤/۰۱۳	۰/۹۰۸	۰/۱٦۹	۱/۰۲۲	−۰/۱۱۳	−۳/٦٤٦
٤۳	−۰/۵۹٦	۱/۳۲٦	۰/۷۹۹	−۰/۵۸۱	۱/٤۸۳	۰/۷۹۱
٤٤	۰/۵۸۸	۱/۵۵۰	۰/٤۵٤	−۰/۵۱۳	۱/٤٦٤	−۰/۹۱۳
٤۵	۱/۲۸۰	۱/۱۳۵	−۰/۰۷۱	−۰/۲۷	۱/۱۹	−۱/٤۵٤
٤٦	۲/۰٦۳	۱/٤۵۸	۰/۰۹۲	۰/۵۵٦	۱/۱۹۲	−۳/۰۱
٤۷	۱/۵٦٦	۱/۰٤۵	۰/۱۷۳	۰/۰۰۹	۱/۱۱۸	−۱/٦۳۸
٤۸	۱/۹۷۱	۱/۱٦۱	−۰/۳۱۵	۰/۳۲	۰/۷۷	−۲/۲۹۱
٤۹	۱/۲۱۹	۱/۲۹۹	۰/۱۰۳	−۰/۲۳٤	۱/٤۲۳	−۱/۵۸٦
۵۰	۲/۱۸۱	۱/٦۷۸	−۰/۰۸۵	۰/۹۳۲	۰/۹۹۷	−۳/٦٦۲
۵۱	۳/۲۳۷	۱/۳۹٤	۰/٤۳۵	۱/٦۳۷	−۰/۲۵٤	−٤/۵۱۳
۵۲	۱/۲۳۳	۱/۳٤۸	−۰/۸۵۳	−۰/۵۱۷	۰/۹٤٤	−۱/٦٦۳
۵۳	۱/٤۳۹	۱/۹۰۹	−۱/۰۵۵	۰/۰۲۳	۱/۱۳٤	−۲/۷٤۸
۵٤	۲/٤٤۵	۱/۰۳۳	۰/٤۸۸	۰/۷۷	۰/۷۲۲	−۲/۵۲٦
۵۵	۱/۱۵۷	۲/۱۷۸	−۱/۳٤۷	−۰/۲۵٤	۱/۳۱۷	−۲/۵۲۲

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[1] (SS)، ساختار تقریباً ساده[2] (APSS) و ساختار پیچیده[3] (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، ۲۰۱٦؛ کیم، ۲۰۱۸).

همان‌طور که در جدول (٦) دیده می‌شود، برخی از سؤال‌های آزمون ریاضی مانند سؤال‌های ۵۲،۵۳، ٥٤ و ۵۵ در سال 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، ٤ 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 همترازشده در شش روش مختلف همترازسازی چندبعدی و تک‌بعدی در جدول (۷) ارائه شده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول (۷) نمره‌های همتراز شده در شش روش همترازسازی

نمره‌های خام	روش نمره واقعی IRT تک‌بعدی[1]	روش نمره مشاهده شده IRT تک‌بعدی[2]	روش نمره واقعی IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی کامل	همترازسازی همصدک
۰	۰	۰/۶۲۶۴	۰	۰/۶۶۳۹	−۰/۱۸۱۲	۰/۵۴۰۳
۱	۱/۴۷۶۵	۱/۴۴۱۴	۱/۴۷۰۴	۱/۴۳۳۴	۰/۷۲۰۶	۱/۴۴۹۲
۲	۲/۴۸۹۳	۲/۴۴۴۵	۲/۲۹۹۳	۲/۳۳۳۱	۱/۶۹۰۶	۲/۵۶۷۰
۳	۳/۵۰۱۸	۳/۴۶۲۲	۳/۲۰۸۳	۳/۲۵۵۰	۲/۶۵۲۸	۳/۵۵۵۶
۴	۴/۴۸۷۹	۴/۴۶۵۵	۴/۱۴۹۶	۴/۱۹۱۸	۳/۶۵۳۶	۴/۴۹۲۹
۵	۵/۴۴۵۴	۵/۴۴۹۱	۵/۱۱۰۱	۵/۱۴۵۲	۴/۶۹۱۷	۵/۴۱۲۷
۶	۶/۳۷۹۳	۶/۴۱۷۱	۶/۰۸۴۹	۶/۱۱۳۵	۵/۷۶۲۳	۶/۳۳۲۳
۷	۷/۲۹۶۸	۷/۳۵۹۰	۷/۰۷۱۱	۷/۰۹۴۱	۶/۸۵۹۰	۷/۳۲۷۰
۸	۸/۲۰۴۶	۸/۲۴۵۰	۸/۰۶۶۹	۸/۰۸۵۰	۷/۹۷۰۳	۸/۱۱۱۳
۹	۹/۱۰۸۵	۹/۰۸۵۴	۹/۰۷۰۷	۹/۰۸۴۵	۹/۰۸۱۶	۸/۹۴۶۷
۱۰	۱۰/۰۱۲۸	۹/۹۶۰۸	۱۰/۰۸۰۹	۱۰/۰۹۰۹	۱۰/۱۸۰۳	۹/۹۳۹۵
۱۱	۱۰/۹۲۰۰	۱۰/۸۹۴۴	۱۱/۰۹۶۱	۱۱/۱۰۲۹	۱۱/۲۵۶۳	۱۰/۹۴۷۳
۱۲	۱۱/۸۳۱۸	۱۱/۸۵۸۶	۱۲/۱۱۴۹	۱۲/۱۱۹۱	۱۲/۳۰۳۸	۱۱/۹۰۸۱
۱۳	۱۲/۷۴۸۶	۱۲/۸۲۶۳	۱۳/۱۳۶۴	۱۳/۱۳۸۴	۱۳/۳۲۰۵	۱۲/۷۴۴۰
۱۴	۱۳/۶۷۰۴	۱۳/۷۸۹۸	۱۴/۱۵۹۵	۱۴/۱۵۹۸	۱۴/۳۰۷۸	۱۳/۵۷۸۷
۱۵	۱۴/۵۹۷۰	۱۴/۷۳۰۲	۱۵/۱۸۳۶	۱۵/۱۸۲۶	۱۵/۲۶۸۹	۱۴/۳۷۰۳
۱۶	۱۵/۵۲۸۱	۱۵/۶۲۶۳	۱۶/۲۰۸۲	۱۶/۲۰۶۳	۱۶/۲۰۸۵	۱۵/۴۳۴۴
۱۷	۱۶/۶۴۳۴	۱۶/۴۹۵۸	۱۷/۲۳۳۰	۱۷/۲۳۰۵	۱۷/۱۳۱۶	۱۶/۴۱۶۶
۱۸	۱۷/۴۰۲۸	۱۷/۳۸۲۸	۱۸/۲۵۷۹	۱۸/۲۵۴۹	۱۸/۰۴۲۵	۱۷/۵۵۳۵
۱۹	۱۸/۳۴۶۵	۱۸/۳۰۴۱	۱۹/۲۸۲۹	۱۹/۲۷۹۳	۱۸/۹۴۵۰	۱۸/۶۴۷۰

[1]. unidimensional true score equating (UIRTT)
[2]. Unidimensional observed score equating (UIRTO)

نمره‌های خام	روش نمره واقعی IRT تک‌بعدی[1]	روش نمره مشاهده شده IRT تک‌بعدی[2]	روش نمره واقعی IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی کامل	همترازسازی همصدک
۲۰	۱۹/۲۹٤۷	۱۹/۲٦٤۳	۲۰/۳۰۷۹	۲۰/۳۰۳۷	۱۹/۸٤۲٤	۱۹/٦٦٦٦
۲۱	۲۰/۲٤۷۸	۲۰/۲٤٤۱	۲۱/۳۳۲۹	۲۱/۳۲۸۰	۲۰/۷۳۸۳	۲۰/٤۷۲۲
۲۲	۲۱/۲۰٦٥	۲۱/۲۳۱۷	۲۲/۳٥۷۹	۲۲/۳٥۲۱	۲۱/٦۳۷۳	۲۱/۲۷۲۷
۲۳	۲۲/۱۷۱۳	۲۲/۲۲۲۳	۲۳/۳۸۲۸	۲۳/۳۷٥۹	۲۲/٥٤٤۱	۲۲/۲۰۳۱
۲٤	۲۳/۱٤۳۰	۲۳/۲۱۲۳	۲٤/٤۰۷٥	۲٤/۳۹۹۱	۲۳/٤۷۳۱	۲۳/۲٥۰۰
۲٥	۲٤/۱۲۲٥	۲٤/۱۹٦۸	۲٥/٤۳۱٦	۲٥/٤۲۱٥	۲٤/٤۲۹۳	۲٤/۱۲۹٦
۲٦	۲٥/۱۱۰٦	۲٥/۱۷۰۷	۲٦/٤٥٤٦	۲٦/٤٤۲۳	۲٥/٤۰۷۳	۲٥/۲۲۰۰
۲۷	۲٦/۱۰۸۳	۲٦/۱٤۱۸	۲۷/٤۷٥٦	۲۷/٤٦۰٦	۲٦/٤۰۷۹	۲٦/۳٦٥۳
۲۸	۲۷/۱۱٦٥	۲۷/۱۳۰۲	۲۸/٤۹۳۳	۲۸/٤۷٥۱	۲۷/٤۳۰۰	۲۷/٥۷۸۹
۲۹	۲۸/۱۳٦۰	۲۸/۱۳۹۰	۲۹/٥۰٥۷	۲۹/٤۸۳۸	۲۸/٤۷۰۸	۲۸/٥۸۸۲
۳۰	۲۹/۱٦۷۰	۲۹/۱٥۹۲	۳۰/٥۱۰۳	۳۰/۳۸٤۳	۲۹/٥۳٦٤	۲۹/٤٤۱۱
۳۱	۳۰/۲۰۹۳	۳۰/۱۸٥۳	۳۱/٥۰۳٦	۳۱/٤۷۳۷	۳۰/٦۲٦۲	۳۰/٥۰۰۰
۳۲	۳۱/۲٦۳۲	۳۱/۲۱۷٤	۳۲/٤۸۱۸	۳۲/٤٤۸٥	۳۱/۷۲۳۰	۳۱/۳٤۳۷
۳۳	۳۲/۳۲٥٦	۳۲/۲٦۰۸	۳۳/٤٤۰٤	۳۳/٤۰٥٥	۳۲/۸۲۲٤	۳۳/۳۱۲٥
۳٤	۳۳/۳۹٤۰	۳۳/۳۲٦۷	۳٤/۳۷٤۷	۳٤/۳٤۱٥	۳۳/۹۲۳۲	۳٤/٤۰۹۰
۳٥	۳٤/٤٦٤٥	۳٤/٤۱۹۱	۳٥/۲۸۰۳	۳٥/۲٥۲٥	۳٥/۰۲٤٥	۳٥/۲۰۰۰
۳٦	۳٥/٥۳۱٥	۳٥/٥۱۰۸	۳٦/۱٥۳۰	۳٦/۱۳۰۹	۳٦/۱۲۸٦	۳٦/۱۲٥۰
۳۷	۳٦/٥۸۸٥	۳٦/٥۷۱۳	۳٦/۹۸۹۸	۳٦/۹٦۸٦	۳۷/۲۳۷۹	۳۷/۱٦٦٦
۳۸	۳۷/٦۲۷۹	۳۷/٦۰۳٦	۳۷/۷۸۸٦	۳۷/۷٦۷۳	۳۸/۳٥۱۲	۳۷/۷۷۷۷
۳۹	۳۸/٦٤۱۳	۳۸/٦۱۷۲	۳۸/٥٤۸۸	۳۸/٥٤۰۹	۳۹/٤٦٥۲	۳۸/۲۲۲۲
٤۰	۳۹/٦۱۹۸	۳۹/٦۱۰۳	۳۹/۲۷۰۸	۳۹/۳۱۹۳	٤۰/٦۲٥۰	۳۹
٤۱	٤۰/٥٥٥۰	٤۰/٥۷٤۳	۳۹/۹٥٦٦	٤۰/۱۲۳۳	٤۱/۷۹۳٤	٤۰/۷٥۰۰

نمره‌های خام	روش نمره واقعی IRT تک‌بعدی[1]	روش نمره مشاهده شده IRT تک‌بعدی[2]	روش نمره واقعی IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی تک‌بعدی شده	روش نمره مشاهده شده IRT چندبعدی کامل	همترازسازی همصدک
۴۲	۴۱/۴۳۹۳	۴۱/۵۰۰۶	۴۰/۶۰۹۸	۴۰/۹۵۱۶	۴۲/۹۴۹۹	۴۳/۹۰۰۰
۴۳	۴۲/۲۶۷۴	۴۲/۴۱۳۸	۴۱/۲۳۵۲	۴۱/۷۸۹۶	۴۴/۱۰۵۲	۴۴/۵۰۰۰
۴۴	۴۳/۰۳۷۴	۴۳/۲۹۷۳	۴۱/۸۳۹۴	۴۲/۶۱۸۲	۴۵/۲۴۹۹	۴۵
۴۵	۴۳/۷۵۱۹	۴۴/۱۴۷۴	۴۲/۴۳۱۰	۴۳/۴۶۳۹	۴۶/۳۵۷۱	۴۷
۴۶	۴۴/۴۱۹۳	۴۴/۹۵۱۶	۴۳/۰۲۰۷	۴۴/۳۸۴۶	۴۷/۳۷۵۰	۴۸
۴۷	۴۵/۰۵۴۷	۴۵/۶۸۳۹	۴۳/۶۲۳۷	۴۵/۳۲۰۳	۴۸/۲۵۰۰	۴۸
۴۸	۴۵/۶۸۲۶	۴۶/۴۳۴۲	۴۴/۲۶۲۴	۴۶/۲۶۸۷	۴۹	۴۸
۴۹	۴۶/۳۴۲۴	۴۷/۳۱۵۵	۴۴/۹۷۵۴	۴۷/۲۳۱۸	۵۲/۵۰	۴۹
۵۰	۴۷/۱۰۹۶	۴۸/۲۱۴۱	۴۵/۸۴۳۶	۴۸/۲۱۷۷	۵۲/۵۰	۵۲/۵۰
۵۱	۴۸/۱۸۵۷	۴۹/۱۳۶۷	۴۷/۰۹۸۲	۴۹/۲۴۴۸	۵۲/۵۰	۵۲/۵۰
۵۲	۵۲	۵۰/۱۵۳۲	۵۲	۵۰/۳۴۲۱	۵۲/۵۰	۵۲/۵۰

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

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

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

روش همترازسازی در نمره‌های همترازشده 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

نمره‌های خام	روش تک‌بعدی نمره واقعی IRT	روش تک‌بعدی نمره مشاهده شده IRT	روش نمره واقعی چندبعدی تک‌بعدی شده IRT	روش نمره مشاهده شده چندبعدی تک‌بعدی شده IRT	روش چندبعدی کامل نمره مشاهده شده IRT
۰	-۰/۵٤۰٤	۰/۰۸٦۰	-۰/۵٤۰٤	۰/۱۲۳۵	-۰/۷۲۱٦
۱	۰/۰۲۷۳	-۰/۰۰۷۸	۰/۰۲۱۲	-۰/۰۱۵۸	-۰/۷۲۸٦
۲	-۰/۰۷۷۷	-۰/۱۲۲۵	-۰/۲٦۷۷	-۰/۲۳۳۹	-۰/۸۷٦٤
۳	-۰/۰۵۳۹	-۰/۰۹۳۵	-۰/۳٤۷٤	-۰/۳۰۰۷	-۰/۹۰۲۸
٤	-۰/۰۰۵۰	-۰/۲۷٤	-۰/۳٤۳۳	-۰/۳۰۱۱	-۰/۸۳۹۳
۵	۰/۰۳۲٦	۰/۰۳٦۳	-۰/۳۰۲۷	-۰/۲٦۷٦	۰/۷۲۱
٦	۰/۰٤۷۰	۰/۰۸٤۸	-۰/۲٤۷٤	-۰/۲۱۸۸	۰/۵٦۹۹
۷	-۰/۰۳۰۲	۰/۰۳۲۰	۰/۲۵۵۹	-۰/۲۳۲۹	-۰/٤٦۸
۸	۰/۰۹۳۳	۰/۱۳۳۷	-۰/۰٤٤٤	-۰/۰۲٦۳	-۰/۱٤۱
۹	۰/۱٦۱۸	۰/۱۳۸۷	۰/۱۲٤۰	۰/۱۳۷۸	۰/۱۳٤۹
۱۰	۰/۰۷۳۲	۰/۰۲۱۲	۰/۱٤۱۳	۰/۱۵۱۳	۰/۲٤۰۷
۱۱	-۰/۰۲۷٤	-۰/۰۵۳۰	۰/۱٤۸۷	۰/۱۵۵۵	۰/۳۰۹
۱۲	-۰/۰۷٦٤	-۰/۰۵۱٤	۰/۲۰٦۷	۰/۲۱۰۹	۰/۳۹۵۷
۱۳	۰/۰۰٤۵	۰/۰۸۲۲	۰/۳۹۲۳	۰/۳۹٤۳	۰/۵۷٦۵
۱٤	۰/۰۹۱۷	۰/۲۱۱۱	۰/۵۸۰۸	۰/۵۸۱۱	۰/۷۲۹۱
۱۵	۰/۲۲٦٦	۰/۳۵۹۸	۰/۸۱۳۲	۰/۸۱۲۲	۰/۸۹۸٦

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

نمره‌های خام	روش تک‌بعدی نمره واقعی IRT	روش تک‌بعدی نمره مشاهده شده IRT	روش نمره واقعی چندبعدی تک‌بعدی شده IRT	روش نمره مشاهده شده چندبعدی تک‌بعدی شده IRT	روش چندبعدی کامل نمره مشاهده شده IRT
۳٦	۰/۵۹۳۵	-۰/٦۱٤۲	۰/۰۲۸۰	۰/۰۰۵۹	۰/۰۰۳۷
۳۷	-۰/۵۷۸۲	-۰/۵۹۵٤	-۰/۱۷٦۹	۰/۱۹۸۱	۰/۰۷۱۳
۳۸	-۰/۱٤۹۹	-۰/۱۷٤۲	۰/۰۱۰۸	-۰/۰۱۰۵	۰/۵۷۳٤
۳۹	۰/٤۱۹۱	۰/۳۹۵۰	۰/۳۲٦٦	۰/۳۱۸۷	۱/۲٤۳۱
٤۰	۰/٦۱۹۸	۰/٦۱۰۳	۰/۲۷۰۸	۰/۳۱۹۳	۱/٦۲۵
٤۱	-۰/۱۹۵۰	-۰/۱۷۵۷	-۰/۷۹۳٤	-۰/٦۲٦۷	۱/۰٤۳۵
٤۲	-۲/٤٦۰۷	-۲/۳۹۹٤	-۳/۲۹۰۲	-۲/۹٤۸٤	-۰/۹۵۰۰
٤۳	-۲/۲۳۲٦	-۲/۰۸٦۲	-۳/۲٦٤۸	-۲/۷۱۰٤	-۰/۳۹٤۷
٤٤	-۱/۹٦۲٦	۱/۷۰۲۷	-۳/۱٦۰٦	-۲/۳۸۱۸	۰/۲٤۹۹
٤۵	-۳/۲٤۸۱	-۲/۸۵۲٦	-٤/۵٦۹۰	-۳/۵۳٦۱	-۰/٦٤۲۹
٤٦	-۳/۵۸۰۷	-۳/۰٤۸٤	-٤/۹۷۹۳	-۳/٦۱۵٤	-۰/٦۲۵
٤۷	-۲/۹٤۵۳	-۲/۳۱٦۱	-٤/۳۷٦۳	-۲/٦۷۹۷	۰/۲۵۰
٤۸	-۲/۳۱۷٤	-۱/۵٦۵۸	-۳/۷۳۷٦	-۱/۷۳۱۳	۱/۰۰۰
٤۹	-۲/٦۵۷٦	-۱/٦۸٤۵	-٤/۰۲٤٦	-۱/۷٦۸۲	۳/۵۰۰
۵۰	-۵/۳۹۰٤	-٤/۲۸۵۹	-٦/٦۵٦٤	-٤/۲۸۲۳	۰
۵۱	-٤/۳۱٤۳	-۳/۳٦۳۳	-۵/٤۰۱۸	-۳/۲۵۵۲	۰
۵۲	-۰/۵۰۰۰	-۲/۳٤٦۸	-۰/۵۰۰۰	-۲/۱۵۷۹	۰
میانگین تفاوت‌ها	-۰/۷۱۵۵	-۰/٦۲۳۵	-۰/۵۱۸۸	-۰/۲٦۱۷	۰/۱٤۳۸

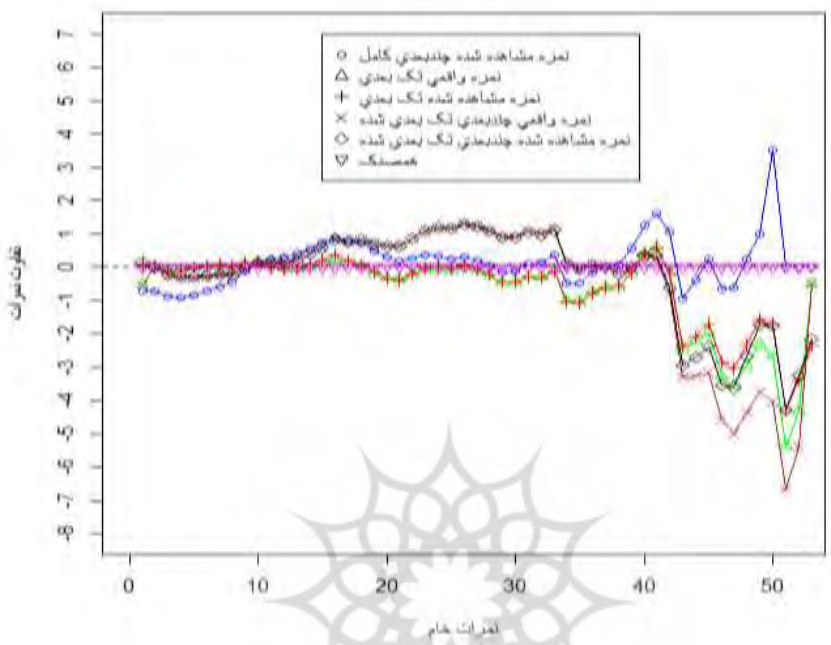
همانطور که در جدول (۸) دیده می‌شود کمترین میزان تفاوت بین روش همصدک ناهموار با روش‌های چندبعدی و تک‌بعدیِ همترازسازی مربوط به روش همترازسازی نمره مشاهده شده چندبعدی کامل و بیشترین تفاوت مربوط به روش همترازسازی نمره واقعی تک‌بعدی است.

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 همترازشده 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.

همان‌طور که دیده می‌شود، نزدیک‌ترین روش به روش همصدک، روش نمره مشاهده شده چندبعدی کامل، پس از آن روشِ 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.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



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 همترازشده 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 (۱دیده 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 تک‌بعدی شبیه به هم عمل کرده و روش‌های همترازسازی نمره واقعی و نمره مشاهده شده چندبعدیِ تک‌بعدی شده مشابه هم عمل کرده‌اند.

بحث و نتیجه‌گیری

هدف از اجرای پژوهش حاضر، شناسایی رویکرد بهینه به منظور همترازسازی آزمون‌های چندبعدی بود. به این منظور، شش روش همترازسازی شامل روش‌های همترازسازی نمره مشاهده شده تک‌بعدی (UIRTO)، همترازسازی نمره واقعی تک‌بعدی (UIRTT)، همترازسازی نمره مشاهده شده چندبعدیِ تک‌بعدی شده (UMIRTO)، همترازسازی نمره واقعی چندبعدیِ تک‌بعدی شده (UMIRTT)، همترازسازی نمره مشاهده شده چندبعدی کامل (FMIRT) و روش همترازسازی همصدک برای همترازسازی داده‌های آزمون چندبعدی ریاضی مقایسه شد. با بررسی پیشینه مشخص شد که نقض مفروضه تک‌بعدی بودن آزمون، اثر سوء بر نتایج همترازسازی دارد و در صورت نقض کامل این مفروضه نمی‌توان به روش‌های همترازسازی تک‌بعدی اعتماد کرد. از این‌رو، مدل‌های نظریه سؤال پاسخ چندبعدی، اخیراً مدنظر پژوهشگران قرار گرفته و در زمینه همترازسازی و پیونددهی چندبعدی تلاش‌هایی از سوی پژوهشگران صورت گرفته است (بروسمان، ۲۰۱۰؛ ژانگ، ۲۰۱۲؛ لی، ۲۰۱۳؛ لی، لی و برینان، ۲۰۱٤). اما مطالعات انجام گرفته، نتیجه مشخص و همسویی را در زمینه همترازسازی آزمون‌های چندبعدی 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 چندبعدی را پر کند، احساس شد.

نتایج به دست آمده نشان داد که در نمره‌های پایین بین شش روش همترازسازی تفاوت چندانی از لحاظ نمره‌های همترازشده وجود ندارد. در نمره‌های وسط بین نمره‌های ۲۲ تا ۳۲ یعنی نمره‌های میانی تفاوت روش‌های همترازسازی نمره مشاهده شده و نمره واقعی چندبعدیِ تک‌بعدی شده با روش همترازسازی همصدک بیشتر از سایر روش‌ها است. در نمره‌های بالا نیز بین روش‌های همترازسازی تک‌بعدی و چندبعدی نظریه سؤال پاسخ با روش همصدک تفاوت بیشتری وجود دارد. این یافته با پژوهش هان، کولن و پوهلمان[1] (۱۹۹۷) همسو است. آنها در پژوهشی که تفاوت و

[1]. 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، ۲۰۰٤؛ لی،۱۹۹۷؛ سیمون، ۲۰۰۸؛ کائو[1]، ۲۰۰۸). به نظر می‌رسد در شرایطی که حجم نمونه بالا است، به دلیل کم شدن خطای نمونه‌گیری، استفاده از روش‌های چندبعدی و تک‌بعدی 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 تک‌بعدی نظریه سؤال پاسخ خطای کمتری نسبت به روش‌های همترازسازی چندبعدی داشتند. اما در سایر طیف نمره‌ها روش همترازسازی چندبعدی بهتر از تک‌بعدی عمل کرد.

در همترازسازی، روش‌های نمره واقعی و نمره مشاهده شده تک‌بعدی شبیه به هم عمل کرده و روش‌های همترازسازی نمره واقعی و نمره مشاهده چندبعدی تک‌بعدی شده مشابه هم عمل کرده‌اند. این یافته، نشان‌دهنده تأثیر نقض مفروضهٔ تک‌بعدی

---

[1]. Cao

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

نزدیک بودن نمره‌های همترازشده نمره واقعی و نمره مشاهده شده تک‌بعدی و روش‌های نمره واقعی و نمره مشاهده شده چندبعدی تک‌بعدی شده نشان‌دهنده این است که بین نمره واقعی 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 (۲۰۱٤)، لی (۲۰۱۳) و بروسمان (۲۰۱۰) است که در پژوهش آنها بین روش‌های همترازسازی نمره واقعی 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 چندبعدیِ 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 همترازشده 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 این‌رو، 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 تک بعدی 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[1] (۲۰۱٤) و شو و وایز[2] (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 چندبعدیِ 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 این‌رو، 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 روشنی 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 چند بعدی 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.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

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 همصدک هموار نشده نمره مشاهده شده IRT و نمره واقعی IRT در طرح گروه های 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 ویژگی های 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 همصدک در معادل‌سازی نمرات آزمون‌های ورودی دانشگاه‌ها (کنکور ورودی سال ۱۳۸۷)*. پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبایی.

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 Lowa].

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 Lowa].

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 Lowa].

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 Lowa].

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].