

Comparison of Validity, Accuracy, and Reliability of the Scores obtained from Thurstonian Item Response Theory in forced-choice scales

Ali Tayarani Rad	Ph.D. Student in Assessment and Measurement, Allameh Tabataba'i University, Tehran, Iran
Mohammadreza Falsafinejad *	Associate Professor, Assessment and Measurement Dept., Allameh Tabataba'i University, Tehran, Iran
Noorali Farokhi	Associate Professor, Assessment and Measurement Dept., Allameh Tabataba'i University, Tehran, Iran
Ahmad Borjali	Professor, Clinical Psychology Dept., Allameh Tabataba'i University, Tehran, Iran

Abstract

The main purpose of this study was to identify the optimum approach to analysis of epistemic data in measuring personality traits. For this purpose, MPLUS software and lavan software package simulate different formats of response scales, and two traditional scoring approaches and Thurston's IRT method were compared in some of the most important psychometric outcomes. Overall the findings of the study indicated that the Thurstonian IRT approach performs better than the traditional scoring approach in accurately estimating actual scores, criterion validity, locating a person on the attribute continuum. Also, the research findings showed that the average correlation between estimated scores and actual scores for all three correlation coefficients (0, 0.25 and 0.50) in Thurstonian Item Response Theory is higher than the scores obtained from the classical method; True for all three correlation coefficients, in Thurstonian Item Response Theory more than classical method scores and mean experimental reliability scores for correlation coefficients 0.25 and 0.5 with classical method are higher than scores obtained by Thurstonian IRT method, but for coefficient The mean correlation of the mean scores of the Turestonian IRT method is higher than the classical scores. The findings also showed that the difference between the two tests, both Classical and Trestonium IRT, was significant for both the 48 and 96 questions, and the correlation between estimated scores and actual scores in the Thursetonium IRT method was more than the classical method.

Keywords: Ipsative scores, Forced-choice Scale, Traditional Scoring, Thurstonian Item Response Theory

* Corresponding Author: falsafinejad@yahoo.com

How to Cite: TAYARANI RAD, A., falsafinejad, M. R., farrokhi, N., & borjali, A. (2022). Comparison of Validity, Accuracy, and Reliability of the Scores obtained from Thurstonian Item Response Theory in forced-choice scales. *Quarterly of Educational Measurement*, 13(49), 7-29. doi: 10.22054/jem.2021.45918.1954

روایی و پایایی نمرات بدست آمده از مدل نظریه سؤال پاسخ ترستونی در آزمون‌های گزینه-بایست

علی طیرانی راد

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

محمد رضا فلسفی

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

نژاد *

نورعلی فرخی

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

احمد برجعلی

دانشیار گروه روان‌شناسی بالینی، دانشگاه علامه طباطبائی، تهران، ایران.

چکیده

هدف اصلی این پژوهش شناسایی رویکرد بهینه تجزیه و تحلیل داده‌های ایستو در سنجش ویژگی‌های شخصیتی بود. شاخص‌های روایی ملاکی، پایایی تجربی و پایایی واقعی نمرات، با توجه به تأثیر همبستگی بین عامل‌ها در دو رویکرد نمره‌گذاری به روش کلاسیک و روش IRT ترستونی مورد مقایسه قرار گرفت. جامعه آماری پژوهش حاضر عبارت از کلیه موقعیت‌هایی است که در آن پرسشنامه‌های شخصیت جهت ارزیابی و گزینش افراد مورد استفاده قرار گرفت. بدین منظور با استفاده از نرم افزار MPLUS به شبیه سازی فرمت‌های مختلف مقیاس‌های گزینه-بایست پرداخته شد و دو رویکرد نمره‌گذاری به روش سنتی و روش IRT ترستونی مورد مقایسه قرار گرفتند. در کل یافته‌های پژوهش گویای این بود که رویکرد IRT ترستونی نسبت به رویکرد کلاسیک نمره‌گذاری، در دقت برآورد نمرات واقعی، پایایی تجربی و روایی ملاکی، بهتر عمل می‌کند. همچنین، یافته‌های تحقیق نشان دادند که میانگین همبستگی بین نمرات برآورد شده و نمرات واقعی برای هر سه ضریب همبستگی (صفر، ۰/۲۵ و ۰/۵۰)، در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است؛ میانگین نمرات پایایی واقعی برای هر سه ضریب همبستگی، در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است و میانگین نمرات پایایی تجربی برای ضرایب همبستگی ۰/۲۵ و ۰/۵ با روش کلاسیک بیشتر از نمرات حاصل از روش IRT ترستونی است، ولی برای ضریب همبستگی صفر میانگین نمرات روش IRT ترستونی بیشتر از نمرات کلاسیک است.

کلیدواژه‌ها: مقیاس پاسخ بایست، نمرات ایستو، نمره‌گذاری سنتی، IRT ترستونی

* نویسنده مسئول: falsafinejad@yahoo.com

مقدمه

از نظر تاریخی، سازه‌های غیرشناختی نظیر شخصیت (Schmidt & Hunter, 1998)، هوش هیجانی (VanRooy & Viswesvaran, 2004)، و مهارت‌های اجتماعی (Semadar et al., 2006) غالباً توسط مقیاس‌های دارای طیف لیکرتی مورد سنجش قرار می‌گیرند. مقیاس‌های نوع لیکرتی مجموعه‌ای تک گزاره‌ای بوده و نیاز به این دارد که پاسخ دهندگان سطح توافق خود را با هر یک از جملات یا گزاره‌ها برای مثال، در فرمت ۱ (کاملاً مخالفم) تا ۵ (کاملاً موافقم) مشخص سازد.

این روش شناسی به علت آمادگی آن به انواع مختلفی از سوگیریهای پاسخ مورد نقد قرار گرفته است. به ویژه، سبک پاسخ دهی وانمود خوب موجب متورم شدن میانگین‌های مقیاس و همبستگی‌های بالا می‌شود، و این امر می‌تواند موجب کاهش روایی و کاربرد بهینه ابزارهای بکار رفته برای تصمیمات حساس و سرنوشت ساز شود. مقیاس‌های از نوع لیکرت همچنین آمادگی برای خطاهای درجه بندی کننده (نظیر هالو^۱) و سوگیریهای وابسته به فرهنگ دارند که احتمالاً موجب متورم شدن همبستگی‌های بین ابعادی (Borman et al., 2001) و تضعیف روابط با پیامدها در بافت‌های پژوهشی بین فرهنگی شود (He & vandeVijver, 2013).

برای کاهش این سوگیریهای پاسخ و خطاهای درجه بندی، مقیاس‌های گزینه-بایست چند بعدی به عنوان جایگزینی برای مقیاس‌های لیکرتی در سنجش‌های غیر شناختی شده است (Stark et al., 2012). مقیاس‌های گزینه-بایست چند بعدی جملات ارائه شده در بلوک‌های دو، سه، و چهار تایی در درون بلوک‌ها هستند، جملات ارائه کننده سازه‌های متفاوتی هستند که احتمالاً با مطلوبیت اجتماعی پیوند خورده‌اند. کار یا وظیفه پاسخ دهنده انتخاب از بین این جملات در هر بلوکی است که "بیشترین شباهت با فرد دارد" یا درجه بندی جملات در هر بلوک از "بیشترین شباهت" تا "کمترین شباهت" است.

در نظریه، پیوند دادن مطلوبیت اجتماعی و پاسخ‌های افراطی موجب می‌شود که تشخیص بهترین پاسخ‌ها بسیار سخت شود، و با اجبار پاسخ دهندگان به انتخاب از بین جایگزین‌ها، به جای نشان دادن سطح توافق خود با جملات، سوگیری پاسخ‌ها و خطای درجه بندی کننده کاهش یابد (Brown & Maydeu-Olivares, 2014).

علی‌رغم نقاط قوت پرسش‌نامه گزینه-بایست در مقابله با وانمودسازی، چالش‌های خاصی در نمره‌گذاری و تحلیل آنها وجود دارد. نمره‌گذاری مبتنی بر رویکرد کلاسیک برای چنین قالب‌هایی منجر به بدست آمدن نمرات ipsativenesscores می‌شود که در این حالت نمرات کل بدست آمده برای تمام افراد مقدار ثابتی است. پرسشنامه‌های گزینه-بایست تنها یکی از چندین شیوه‌ای هستند که منجر به ایجاد داده‌های ایپسیتیو می‌شوند. اگرچه نمرات ایپسیتیو برای سنجش‌های نمرات درون فردی قابل تفسیر است، ولی زمانی که برای مقایسه‌های بین فردی استفاده شود مشکل ساز می‌شوند. این نوع از داده‌ها دارای ویژگی‌های خاصی هستند که روایی سازه و تفسیر نمرات و سایر ویژگی‌های روانسنجی را به مخاطره می‌اندازد (Dunlap & Cornwell, 1994؛ Meade, 2004). استفاده از روش‌های تحلیل عاملی معمول برای داده‌های ایپسیتیو منجر به استخراج عوامل دوقطبی می‌شود که مقیاس‌هایی متضاد از دو عامل کاملاً متضاد را شامل می‌شود (Dunlap & Cornwell, 1994؛ Baron, 1996). همچنین داده‌های ایپسیتیو، همسانی درونی ابزار را تحریف می‌کند، اما اینکه در چه جهتی و به چه مقداری ظاهر می‌شود، تحت شرایط مختلف متفاوت خواهد بود. به طور کلی آلفای کرونباخ برای پایایی قالب گزینه-بایست نامناسب است، به این دلیل که داده‌های ایپسیتیو برخی مفروضات مهم که آماره آلفا بر آن متکی است را نقض می‌کند. مشکل اول این است که اساسی‌ترین مفروضه، استقلال واریانس خطا، در داده‌های ایپسیتیو نقض می‌شود. به علاوه پاسخ به ماده‌ها در یک بلوک با توجه به صفات مکنون، مستقل نیست- در عوض پاسخ داده شده به یک ماده به پاسخ‌های داده شده به دیگر ماده‌های بلوک بستگی دارد (Meade, 2004).

پیشرفت‌های اخیر در روش‌های برآورد، توسعه سریع مدل‌های سوال-پاسخ برای داده‌های مقایسه‌ای، شامل مدل IRT ترستون (Brown & Maydeu-Olivares, 2014)، مدل رجحان‌های زوجی چند-تک بعدی (Stark et al., 2012) را میسر ساخته است. مدل‌یابی مناسب سوال-پاسخ، برآورد نمرات شخص به نحوی که به طور مستقیم برای مقایسه‌های بین فردی قابل تفسیر باشد را ممکن می‌سازد.

مدل IRT ترستونی یکی از روش‌های مورد استفاده برای داده‌های ایپسیتیو، است. بر اساس روش قضاوت‌های مقایسه‌ای ترستون می‌توان، مدلی مبتنی بر نظریه سؤال پاسخ معرفی کرد که قادر به مدل‌یابی پاسخ‌های داده شده به تمامی انواع پرسشنامه‌های گزینه-بایست باشد

(Baron & Meade, 2010). با استفاده از مدل براون و میدو در مدل یابی تصمیم‌گیری ترجیحی^۱ برای پرسشنامه‌های گزینه-بایست، نمراتی مربوط به ویژگی‌های اندازه‌گیری شده بدست می‌دهد که فارغ از مشکلات مربوط به داده‌های ایپستيو است. مدل سوال-پاسخ ترستون، مدل سوال-پاسخ چند بعدی با برخی ویژگی‌های خاص است که می‌تواند به طور مستقیم و با استفاده از نرم افزارهای مدل یابی برآورد شود.

مطالعه Baron (2010) یکی از مطالعات اخیر صورت گرفته در زمینه کارایی این روش در برآورد پارامترهای فرد و ویژگی‌های روانسنجی آزمون است. در این مطالعه مقایسه‌ای بین روش کلاسیک و IRT تورستونی در برآورد نمرات صورت نگرفته است، و برتری این روش نسبت روش کلاسیک به طور خاص مورد بررسی قرار نگرفته است. یکی از مشکلات روش نمره‌گذاری کلاسیک برای داده‌های ایپستيو بیش برآورد ضرایب روایی است که کارایی نمرات برآورد شده به روش ترستونی در اصلاح این نقیصه مورد بررسی قرار نگرفته است. با توجه به اینکه رویکرد مناسب تجزیه و تحلیل داده‌های ایپستيو تحت شرایط گوناگون و با در نظر گرفتن تعداد عامل‌های مختلف، تغییر در طول آزمون، فرمت و محتوی سؤالات، مشخص نیست و شواهد پژوهشی در این موارد ناکافی و مبهم است، بنابراین انجام پژوهش‌هایی جهت بررسی دقت و روایی این روش نمره‌گذاری ضرورت دارد. این پژوهش به عنوان یک پژوهش پیشگام در جهت بررسی روایی و پایایی نمرات آزمون‌های گزینه-بایست در روش نمره‌گذاری IRT ترستونی با روش سنتی نمره‌گذاری انجام می‌شود.

روش

این پژوهش از حیث هدف از نوع پژوهش‌های کاربردی است و با توجه به ماهیت مساله و کاربرد روش شبیه‌سازی و دستکاری شرایط مختلف مورد بررسی، در زمره تحقیقات تجربی است. نمرات واقعی برآورد شده و خطاهای استاندارد، روایی و پایایی آزمون، متغیرهای وابسته پژوهش را تشکیل می‌دهند که تأثیر متغیرهای مستقل شامل تعداد عامل‌ها، طول آزمون، فرمت و همچنین محتوی سؤالات بر روی آنها به روش آزمایشی و با استفاده از داده‌های شبیه‌سازی شده مورد بررسی قرار می‌گیرد.

جامعه آماری پژوهش حاضر عبارت است از کلیه موقعیت‌هایی که در آن پرسشنامه‌های شخصیت جهت ارزیابی و گزینش افراد مورد استفاده قرار می‌گیرد. از بین موقعیت‌های

گوناگون تأثیر مربوط به تعداد عامل‌های مختلف، تغییر در طول آزمون، فرمت متفاوت سؤالات و همچنین محتوی سؤالات مورد بررسی قرار می‌گیرد. داده‌های مورد استفاده در پژوهش حاضر از طریق شبیه‌سازی تولید می‌شود. مطالعات شبیه‌سازی ابزار بسیار ارزشمندی محسوب می‌شوند که به دلیل سادگی و مفید بودن در بسیاری از حوزه‌های تحقیق به منظور بررسی نظریه‌ها و پژوهش مورد استفاده قرار می‌گیرند. روش‌های شبیه‌سازی به پژوهشگر اجازه می‌دهد که به طور تجربی ویژگی خاصی را با استفاده از نمونه‌های تصادفی استخراج شده از جامعه معلوم مورد بررسی قرار دهد (ایوانس، 2010). به منظور شبیه‌سازی داده‌ها، شرایط متفاوتی شامل تعداد عامل‌های مختلف، طول آزمون، فرمت سؤالات و همچنین محتوی سؤالات مورد بررسی قرار می‌گیرد. داده‌های پرسشنامه گزینه-بایست با شرایط مختلف آزمایشی و تکرارهای مختلف توسط نرم افزار Mplus شبیه‌سازی خواهد شد.

در این پژوهش دو روش تحلیل کلاسیک و مبتنی بر نظریه سؤال پاسخ استفاده خواهد شد. در تجزیه و تحلیل داده‌ها به روش کلاسیک از همبستگی پیرسون برای برآورد ضریب روایی و برای نمره گذاری از روش نمره گذاری مرسوم استفاده می‌شود که در آن به ماده‌ای که به عنوان اولویت نخست انتخاب شده بیشترین نمره و به ماده‌ای که به عنوان پایین‌ترین اولویت انتخاب شده کمترین نمره اختصاص می‌یابد. از روش IRT ترستونی در تجزیه و تحلیل مبتنی بر مدل سوال-پاسخ بهره گرفته خواهد شد. در این روش از همبستگی پیرسون برای برآورد ضریب روایی و روش‌های تحلیل بیز در برآورد نمره استفاده خواهد شد. بعلاوه از تحلیل واریانس چندمتغیری برای مقایسه نمرات کلاسیک و نمرات مبتنی بر مدل IRT ترستونی استفاده شد.

یافته‌ها

فرضیه اول: رویکرد IRT ترستونی در برآورد نمرات واقعی آزمون‌های ایستتو نسبت به رویکردهای سنتی از دقت بیشتری برخوردار است.

جدول ۱. همبستگی بین نمرات برآورد شده و نمرات واقعی، پایایی واقعی و پایایی تجربی در دو روش IRT ترستونی و کلاسیک

کلاسیک		IRT ترستونی		روش تحلیل
انحراف معیار	میانگین	انحراف معیار	میانگین	
۰,۱۵۷	۰,۶۳۷	۰,۲۳۵	۰,۷۳۸	همبستگی بین نمرات برآورد شده و واقعی
۰,۱۸۷	۰,۴۹۶	۰,۱۴۵	۰,۶۵۹	پایایی واقعی
۰,۰۹۱	۰,۸۰۱	۰,۲۱۱	۰,۷۳۳	پایایی تجربی

در جدول ۱ آمار توصیفی مربوط به میانگین و انحراف معیار برآوردهای مربوط به IRT همبستگی بین نمرات برآورد شده و واقعی، پایایی واقعی و تجربی حاصل از دو روش ترستونی و کلاسیک ارائه شده است. همان طور که ملاحظه می‌گردد میانگین همبستگی بین ترستونی بیشتر IRT نمرات برآورد شده و نمرات واقعی و همچنین پایایی واقعی در روش از نمرات حاصل از روش کلاسیک است. در مقابل میانگین پایایی تجربی برآورده شده به ترستونی است. IRT روش کلاسیک بالاتر از روش

جدول ۲. آزمون t وابسته جهت مقایسه همبستگی بین نمرات برآورد شده و نمرات واقعی، پایایی واقعی و تجربی در دو روش IRT ترستونی و کلاسیک

متغیر	آماره t	سطح معنی داری	تفاوت میانگین
همبستگی بین نمرات برآورد شده و واقعی	۴۰/۰۰۹	۰/۰۱	۰/۱۰۱
پایایی واقعی	۷۴/۱۵۸	۰/۰۱	۰/۱۶۴
پایایی تجربی	-۲۷/۱۴۰	۰/۰۱	۰/۰۶۸

با توجه به نتایج مندرج در جدول ۲ تفاوت بین همبستگی بین نمرات برآورد شده و ترستونی و IRT نمرات واقعی، پایایی واقعی و پایایی تجربی برآورد شده در روش کلاسیک در سطح آلفای ۰/۰۱ معنی دار است. بر اساس نتایج بدست آمده همبستگی بین ترستونی نسبت IRT نمرات برآورد شده و نمرات واقعی و همچنین پایایی واقعی در روش به روش کلاسیک بالاتر برآورد شده است. در مقابل میانگین پایایی تجربی در روش ترستونی برآورده شده است. IRT کلاسیک بالاتر از روش

فرضیه دوم: رویکرد IRT ترستونی در برآورد روایی ملاکی آزمون‌های ایستو نسبت به رویکردهای سنتی از دقت بیشتری برخوردار است.

جدول ۳. اطلاعات توصیفی روایی ملاکی در دو روش IRT ترستونی و کلاسیک

روش تحلیل	IRT ترستونی		کلاسیک	
	میانگین	انحراف معیار	میانگین	انحراف معیار
روایی ملاکی	۰,۲۵۱	۰,۱۲۵	۰,۰۲۸	۰,۲۷۰

در جدول ۳ آمار توصیفی مربوط به میانگین و انحراف معیار روایی ملاکی برآورد شده ترستونی و کلاسیک ارائه شده است. همان طور که ملاحظه IRT با استفاده از دو روش ترستونی بیشتر از نمرات IRT می‌گردد میانگین ضرایب روایی ملاکی برآورد شده در روش حاصل از روش کلاسیک است.

جدول ۴. آزمون t وابسته جهت مقایسه روایی ملاکی در دو روش IRT ترستونی و کلاسیک

متغیر	آماره t	سطح معنی داری	تفاوت میانگین
روایی ملاکی	۴۹/۶۶۳	۰/۰۱	۰/۲۲۳

با توجه به نتایج مندرج در جدول ۴ تفاوت بین روایی ملاکی برآورد شده در روش ترستونی و کلاسیک در سطح آلفای ۰/۰۱ معنی دار است. بر اساس نتایج بدست آمده IRT ترستونی نسبت به روش کلاسیک بالاتر برآورد شده است. IRT روایی ملاکی در روش **فرضیه سوم:** کارایی رویکردهای سنتی و IRT در تعیین جایگاه شخص روی پیوستار خصیصه، تحت تأثیر همبستگی بین عوامل مندرج در ابزار قرار دارد.

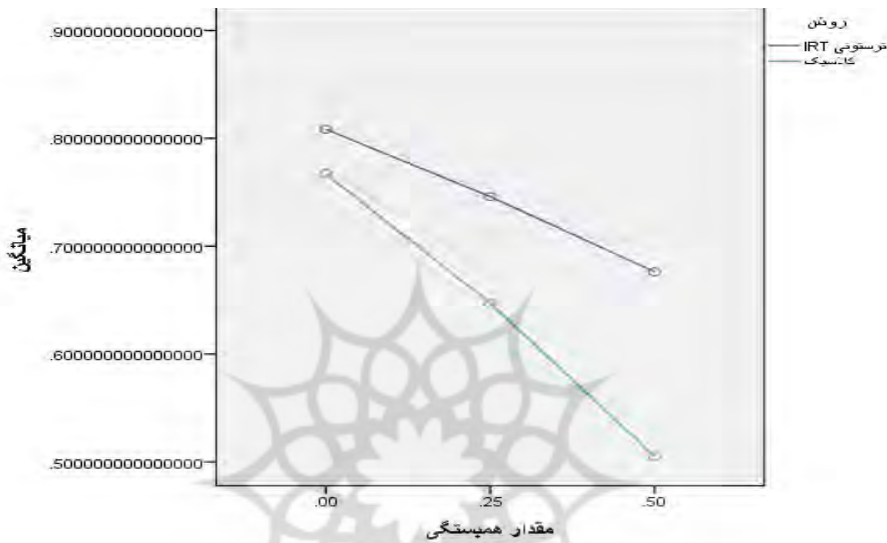
جدول ۵. همبستگی بین نمرات برآورد شده و نمرات واقعی در دو روش IRT ترستونی و کلاسیک به تفکیک مقدار همبستگی بین عامل‌ها

روش تحلیل	IRT ترستونی		کلاسیک	
	میانگین	انحراف معیار	میانگین	انحراف معیار
مقدار همبستگی	۰,۸۰۸	۰,۰۰۶	۰,۷۶۷	۰,۰۰۳
	۰,۷۴۶	۰,۰۰۶	۰,۶۴۷	۰,۰۰۳
	۰,۶۷۶	۰,۰۰۷	۰,۵۰۵	۰,۰۰۳

در جدول ۵ آمار توصیفی مربوط به همبستگی بین نمرات برآورد شده و نمرات واقعی ترستونی و کلاسیک به تفکیک برای مقادیر مختلف همبستگی بین IRT در دو روش عامل‌ها نشان داده شده است. همان طور که ملاحظه می‌گردد میانگین همبستگی بین نمرات

ترستونی بیشتر IRT برآورد شده و نمرات واقعی برای هر سه ضریب همبستگی، در روش از نمرات حاصل از روش کلاسیک است.

نمودار ۱. میانگین همبستگی بین نمرات برآورد شده و نمرات واقعی با توجه به مقادیر همبستگی به تفکیک دو روش IRT ترستونی و کلاسیک



در نمودار ۱ میانگین همبستگی بین نمرات برآورد شده و نمرات واقعی برای ضرایب ترستونی و کلاسیک نشان داده شده است. IRT مختلف همبستگی، به تفکیک دو روش

جدول ۶. نتایج آزمون چند متغیری برای مقایسه همبستگی بین نمرات برآورد شده و نمرات واقعی در دو روش IRT ترستونی و کلاسیک

اثر	مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
تکرار	اثر پیلایی	۲۰۱۶,۶۴۰,۳۶۷	۱	۳۴۷۴	۰,۰۰۱	۰,۳۶۷
	لامبدای ویلکز	۲۰۱۶,۶۴۰,۶۳۳	۱	۳۴۷۴	۰,۰۰۱	۰,۳۶۷
	اثر هتلینگ	۲۰۱۶,۶۴۰,۵۸۰	۱	۳۴۷۴	۰,۰۰۱	۰,۳۶۷
	بزرگ‌ترین ریشه روی	۲۰۱۶,۶۴۰,۵۸۰	۱	۳۴۷۴	۰,۰۰۱	۰,۳۶۷
تکرار * گروه	اثر پیلایی	۲۶۴,۷۷۱,۰۱۳۲	۲	۳۴۷۴	۰,۰۰۱	۰,۱۳۲
	لامبدای ویلکز	۲۶۴,۷۷۱,۰۸۶۸	۲	۳۴۷۴	۰,۰۰۱	۰,۱۳۲
	اثر هتلینگ	۲۶۴,۷۷۱,۰۱۵۲	۲	۳۴۷۴	۰,۰۰۱	۰,۱۳۲
	بزرگ‌ترین ریشه روی	۲۶۴,۷۷۱,۰۱۵۲	۲	۳۴۷۴	۰,۰۰۱	۰,۱۳۲

در جدول فوق نتایج آزمون‌های چند متغیری جهت مقایسه همبستگی بین نمرات برآورد ترستونی و کلاسیک، ارائه شده است. براساس IRT شده و نمرات واقعی در دو روش اطلاعات مندرج در جدول آماره‌های چند متغیری مربوط به عامل تکرار (تفاوت بین دو (که نشان P ترستونی و کلاسیک) در سطح آلفای ۰/۰۱ معنی دار هستند ($IRT < 0/01$) روش ترستونی با نمرات حاصل از روش IRT دهنده وجود تفاوت بین نمرات حاصل از روش کلاسیک است. آماره‌های مربوط به اثر تعاملی بین گروه‌ها و تکرار نیز معنی دار می‌باشد. که ترستونی و کلاسیک با توجه به مقدار ضرایب IRT نشان می‌دهد میزان اختلاف بین دو روش همبستگی بین عامل‌ها متفاوت بوده است.

جدول ۷. نتایج آزمون چند متغیری برای مقایسه همبستگی بین نمرات برآورد شده و نمرات واقعی به تفکیک مقادیر همبستگی بین عامل‌ها

مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
صفر لامبدای ویلکز ۰,۹۷۰	۱۰۷,۶۸۲	۱	۳۴۷۴	۰,۰۰۱	۰,۰۳۰
۰/۲۵ لامبدای ویلکز ۰,۸۵۱	۶۰۸,۲۶۸	۱	۳۴۷۴	۰,۰۰۱	۰,۱۴۹
۰/۵ لامبدای ویلکز ۰,۶۵۷	۱۸۱۱,۲۸۶	۱	۳۴۷۴	۰,۰۰۱	۰,۳۴۳

در جدول فوق نتایج آزمون‌های چند متغیری جهت مقایسه همبستگی بین نمرات برآورد ترستونی و کلاسیک به تفکیک برای ضرایب IRT شده و نمرات واقعی در دو روش همبستگی صفر، ۰/۲۵ و ۰/۵ ارائه شده است. بر اساس مندرجات جدول تفاوت بین دو روش ترستونی و کلاسیک برای هر سه ضریب همبستگی معنی دار بوده و میزان همبستگی IRT ترستونی بیشتر از روش کلاسیک IRT بین نمرات برآورد شده و نمرات واقعی در روش است. بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی ۰/۵ است.

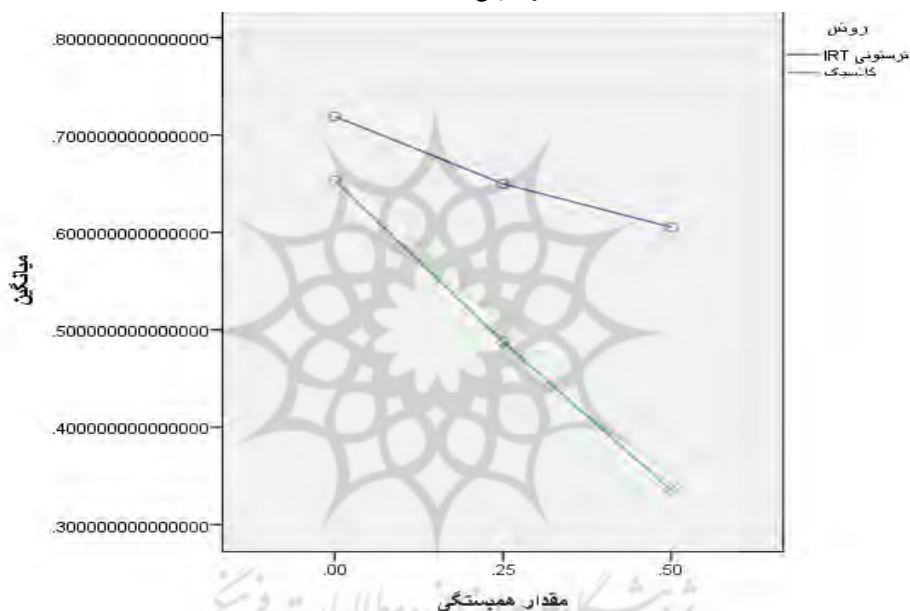
جدول ۸. میانگین و انحراف استاندارد نمرات پایایی واقعی در دو روش IRT ترستونی و کلاسیک به تفکیک مقدار همبستگی بین عامل‌ها

روش تحلیل مقدار همبستگی	IRT ترستونی		کلاسیک	
	میانگین	انحراف معیار	میانگین	انحراف معیار
صفر	۰,۷۱۹	۰,۰۰۴	۰,۶۵۳	۰,۰۰۴
۰/۲۵	۰,۶۵۰	۰,۰۰۴	۰,۴۸۸	۰,۰۰۴
۰/۵	۰,۶۰۵	۰,۰۰۴	۰,۳۳۵	۰,۰۰۴

در جدول ۸ آمار توصیفی مربوط به میانگین و انحراف معیار نمرات پایایی واقعی در دو ترستونی و کلاسیک به تفکیک برای مقادیر مختلف همبستگی بین عامل‌ها IRT روش نشان داده شده است. همان طور که ملاحظه می‌گردد میانگین نمرات پایایی واقعی برای هر ترستونی بیشتر از نمرات حاصل از روش کلاسیک IRT سه ضریب همبستگی، در روش است.

نمودار ۲. میانگین نمرات پایایی واقعی با توجه به مقادیر همبستگی به تفکیک دو روش IRT

ترستونی و کلاسیک



در نمودار ۲ میانگین نمرات پایایی واقعی برای ضرایب مختلف همبستگی بین عامل‌ها، ترستونی و کلاسیک نشان داده شده است. IRT به تفکیک دو روش

جدول ۹. نتایج آزمون چند متغیری برای مقایسه پایایی واقعی در دو روش IRT ترستونی و کلاسیک

اثر	مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
اثر پیلایی	۹۴۶۹,۰۷۰,۷۳۲	۱	۳۴۷۴	۰,۰۰۱	۰,۷۳۲	
لامبدای ویلکز	۹۴۶۹,۰۷۰,۲۶۸	۱	۳۴۷۴	۰,۰۰۱	۰,۷۳۲	تکرار
اثر هتلینگ	۹۴۶۹,۰۷۲,۷۲۶	۱	۳۴۷۴	۰,۰۰۱	۰,۷۳۲	
بزرگ‌ترین ریشه روی	۹۴۶۹,۰۷۲,۷۲۶	۱	۳۴۷۴	۰,۰۰۱	۰,۷۳۲	

اثر	مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
اثر پیلایی	۱۱۹۸,۹۸۰,۴۰۸	۲	۳۴۷۴	۰,۰۰۱	۰,۴۰۸	
لامبدای ویلکز	۱۱۹۸,۹۸۰,۵۹۲	۲	۳۴۷۴	۰,۰۰۱	۰,۴۰۸	
اثر هتلینگ	۱۱۹۸,۹۸۰,۶۹۰	۲	۳۴۷۴	۰,۰۰۱	۰,۴۰۸	
بزرگ‌ترین ریشه روی	۱۱۹۸,۹۸۰,۶۹۰	۲	۳۴۷۴	۰,۰۰۱	۰,۴۰۸	

در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت میانگین نمرات پایایی ترستونی و کلاسیک، ارائه شده است. براساس اطلاعات مندرج در IRT واقعی در دو روش ترستونی و IRT جدول آماره‌های چند متغیری مربوط به عامل تکرار (تفاوت بین دو روش) که نشان دهنده وجود تفاوت P کلاسیک) در سطح آلفای ۰/۰۱ معنی دار هستند (۰/۰۱ < ترستونی با پایایی واقعی حاصل از روش کلاسیک IRT بین پایایی واقعی حاصل از روش است. آماره‌های مربوط به اثر تعاملی بین گروه‌ها و تکرار نیز معنی دار می‌باشد. که نشان تورستونی و کلاسیک با توجه به مقدار ضرایب IRT می‌دهد میزان اختلاف بین دو روش همبستگی متفاوت بوده است.

جدول ۱۰. نتایج آزمون چند متغیری برای مقایسه پایایی واقعی به تفکیک مقادیر همبستگی

مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
صفر لامبدای ویلکز	۵۰۲,۵۷۴	۱	۳۴۷۴	۰,۰۰۱	۰,۱۲۶
۰/۲۵ لامبدای ویلکز	۲۹۹۷,۰۷	۱	۳۴۷۴	۰,۰۰۱	۰,۴۶۳
۰/۵ لامبدای ویلکز	۸۲۷۸,۹۹	۱	۳۴۷۴	۰,۰۰۱	۰,۷۰۴

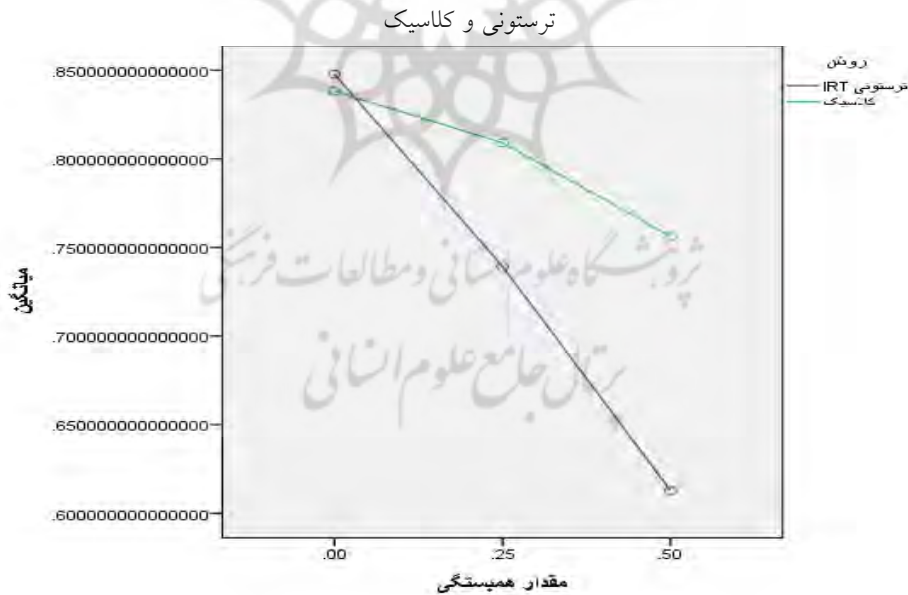
در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت بین میانگین نمرات ترستونی و کلاسیک به تفکیک برای ضرایب همبستگی IRT پایایی واقعی در دو روش IRT صفر، ۰/۲۵ و ۰/۵ ارائه شده است. بر اساس مندرجات جدول تفاوت بین دو روش ترستونی و کلاسیک برای هر سه ضریب همبستگی معنی دار بوده و میزان پایایی واقعی در ترستونی بیشتر از روش کلاسیک است. بیشترین تفاوت بین دو روش نیز برای IRT روش ضریب همبستگی ۰/۵ است.

جدول ۱۱. میانگین و انحراف استاندارد نمرات پایایی تجربی در دو روش IRT ترستونی و کلاسیک به تفکیک مقدار همبستگی بین عامل‌ها

روش تحلیل مقدار همبستگی	IRT ترستونی		کلاسیک	
	میانگین	انحراف معیار	میانگین	انحراف معیار
صفر	۰,۸۴۸	۰,۰۰۵	۰,۸۳۸	۰,۰۰۲
۰/۲۵	۰,۷۳۹	۰,۰۰۵	۰,۸۰۹	۰,۰۰۲
۰/۵	۰,۶۱۳	۰,۰۰۵	۰,۷۵۶	۰,۰۰۲

در جدول ۱۱ آمار توصیفی مربوط به میانگین و انحراف معیار نمرات پایایی تجربی در ترستونی و کلاسیک به تفکیک برای مقادیر مختلف همبستگی بین عامل‌ها IRT دو روش نشان داده شده است. همان طور که ملاحظه می‌گردد میانگین نمرات پایایی تجربی برای IRT ضرایب همبستگی ۰/۲۵ و ۰/۵ با روش کلاسیک بیشتر از نمرات حاصل از روش ترستونی بیشتر IRT ترستونی است. ولی برای ضریب همبستگی صفر میانگین نمرات روش از نمرات کلاسیک است.

نمودار ۳. میانگین نمرات پایایی تجربی با توجه به مقادیر همبستگی به تفکیک دو روش IRT



در نمودار ۳ میانگین نمرات پایایی تجربی برای ضرایب مختلف همبستگی، به تفکیک ترستونی و کلاسیک نشان داده شده است. IRT دو روش

جدول ۱۲. نتایج آزمون چند متغیری برای مقایسه پایایی تجربی در دو روش IRT ترستونی و کلاسیک

اثر	مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
اثر پیلایی	۹۰۰,۵۱۱۰,۲۰۴	۱	۳۵۱۸	۰,۰۰۱	۰,۲۰۴	
لامبدای ویلکز	۹۰۰,۵۱۱۰,۷۹۶	۱	۳۵۱۸	۰,۰۰۱	۰,۲۰۴	تکرار
اثر هتلینگ	۹۰۰,۵۱۱۰,۲۵۶	۱	۳۵۱۸	۰,۰۰۱	۰,۲۰۴	
بزرگ‌ترین ریشه روی	۹۰۰,۵۱۱۰,۲۵۶	۱	۳۵۱۸	۰,۰۰۱	۰,۲۰۴	
اثر پیلایی	۳۸۲,۹۴۳۰,۱۷۹	۲	۳۵۱۸	۰,۰۰۱	۰,۱۷۹	
لامبدای ویلکز	۳۸۲,۹۴۳۰,۸۲۱	۲	۳۵۱۸	۰,۰۰۱	۰,۱۷۹	تکرار * گروه
اثر هتلینگ	۳۸۲,۹۴۳۰,۲۱۸	۲	۳۵۱۸	۰,۰۰۱	۰,۱۷۹	
بزرگ‌ترین ریشه روی	۳۸۲,۹۴۳۰,۲۱۸	۲	۳۵۱۸	۰,۰۰۱	۰,۱۷۹	

در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت میانگین نمرات پایایی ترستونی و کلاسیک، ارائه شده است. براساس اطلاعات مندرج IRT تجربی در دو روش ترستونی IRT در جدول آماره‌های چند متغیری مربوط به عامل تکرار (تفاوت بین دو روش) که نشان دهنده وجود تفاوت P و کلاسیک) در سطح آلفای ۰/۰۱ معنی‌دار هستند ($< 0/01$) ترستونی با نمرات پایایی تجربی حاصل از IRT بین نمرات پایایی تجربی حاصل از روش کلاسیک است. آماره‌های مربوط به اثر تعاملی بین گروه‌ها و تکرار نیز معنی‌دار ترستونی و کلاسیک با توجه به IRT می‌باشد. که نشان می‌دهد میزان اختلاف بین دو روش مقدار ضرایب همبستگی متفاوت بوده است.

جدول ۱۳. نتایج آزمون چند متغیری برای مقایسه پایایی تجربی به تفکیک مقادیر همبستگی

مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
۰,۹۹۸	۶,۱۰۸	۱	۳۵۱۸	۰,۰۱۴	۰,۰۰۲
۰,۹۱۸	۳۱۶,۲۵۰	۱	۳۵۱۸	۰,۰۰۱	۰,۰۸۲
۰,۷۲۴	۱۳۳۹,۹۷	۱	۳۵۱۸	۰,۰۰۱	۰,۲۷۶

در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت بین میانگین نمرات ترستونی و کلاسیک به تفکیک برای ضرایب همبستگی IRT پایایی تجربی در دو روش IRT صفر، ۰/۲۵ و ۰/۵ ارائه شده است. بر اساس مندرجات جدول تفاوت بین دو روش

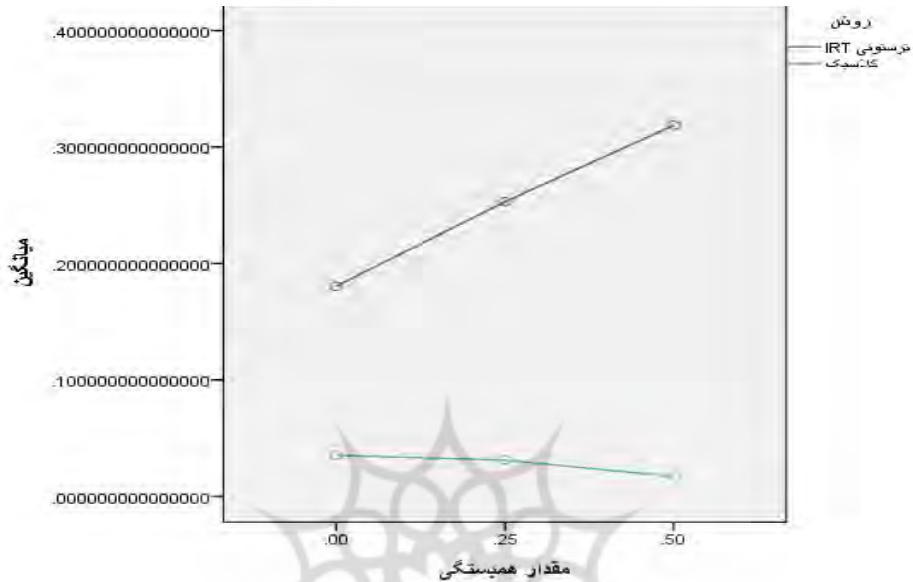
ترستونی و کلاسیک برای هر سه ضریب همبستگی معنی دار بوده و میزان پایایی تجربی برای کلاسیک بیشتر و برای ضریب همبستگی صفر در ضرایب همبستگی ۰/۲۵ و ۰/۵ در روش ترستونی بیشتر است. بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی IRT روش ۰/۵ است.

جدول ۱۴. میانگین و انحراف استاندارد نمرات روایی ملاکی در دو روش IRT ترستونی و کلاسیک به تفکیک مقدار همبستگی بین عامل‌ها

کلاسیک		IRT ترستونی		روش تحلیل
انحراف معیار	میانگین	انحراف معیار	میانگین	مقدار همبستگی
۰,۰۰۸	۰,۰۳۵	۰,۰۰۳	۰,۱۸۰	صفر
۰,۰۰۸	۰,۰۳۱	۰,۰۰۳	۰,۲۵۳	۰/۲۵
۰,۰۰۸	۰,۰۱۷	۰,۰۰۳	۰,۳۱۹	۰/۵

در جدول ۱۴ آمار توصیفی مربوط به میانگین و انحراف معیار نمرات روایی ملاکی در ترستونی و کلاسیک به تفکیک برای مقادیر مختلف همبستگی نشان داده IRT دو روش شده است. همان طور که ملاحظه می گردد میانگین نمرات روایی ملاکی برای هر سه ضریب تورستونی بیشتر از نمرات حاصل از روش کلاسیک است. IRT همبستگی، در روش

نمودار ۴. میانگین نمرات روایی ملاکی با توجه به مقادیر همبستگی به تفکیک دو روش IRT ترستونی و کلاسیک



در نمودار ۴ میانگین نمرات روایی ملاکی برای ضرایب مختلف همبستگی، به تفکیک ترستونی و کلاسیک نشان داده شده است. IRT دو روش

جدول ۱۵. نتایج آزمون چند متغیری برای مقایسه روایی ملاکی در دو روش IRT ترستونی و کلاسیک

اثر	مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
اثر پیلایی	۰,۴۲۷	۲۶۱۸,۹۸	۱	۳۵۲۰	۰,۰۰۱	۰,۴۲۷
لامبدای ویلکز	۰,۵۷۳	۲۶۱۸,۹۸	۱	۳۵۲۰	۰,۰۰۱	۰,۴۲۷
اثر هتلینگ	۰,۷۴۴	۲۶۱۸,۹۸	۱	۳۵۲۰	۰,۰۰۱	۰,۴۲۷
بزرگ‌ترین ریشه روی	۰,۷۴۴	۲۶۱۸,۹۸	۱	۳۵۲۰	۰,۰۰۱	۰,۴۲۷
اثر پیلایی	۰,۰۵۸	۱۰۷,۷۰۱	۲	۳۵۲۰	۰,۰۰۱	۰,۰۵۸
لامبدای ویلکز	۰,۹۴۲	۱۰۷,۷۰۱	۲	۳۵۲۰	۰,۰۰۱	۰,۰۵۸
اثر هتلینگ	۰,۰۶۱	۱۰۷,۷۰۱	۲	۳۵۲۰	۰,۰۰۱	۰,۰۵۸
بزرگ‌ترین ریشه روی	۰,۰۶۱	۱۰۷,۷۰۱	۲	۳۵۲۰	۰,۰۰۱	۰,۰۵۸

در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت میانگین نمرات روایی ترستونی و کلاسیک، ارائه شده است. براساس اطلاعات مندرج IRT ملاکی در دو روش ترستونی IRT در جدول آماره‌های چند متغیری مربوط به عامل تکرار (تفاوت بین دو روش

(که نشان دهنده وجود تفاوت P و کلاسیک) در سطح آلفای ۰/۰۱ معنی دار هستند (۰/۰۱ < ترستونی با روایی ملاکی حاصل از روش کلاسیک IRT بین روایی ملاکی حاصل از روش است. آماره‌های مربوط به اثر تعاملی بین گروه‌ها و تکرار نیز معنی دار می‌باشد. که نشان ترستونی و کلاسیک با توجه به مقدار ضرایب IRT می‌دهد میزان اختلاف بین دو روش همبستگی متفاوت بوده است.

جدول ۱۶. نتایج آزمون چند متغیری برای مقایسه روایی ملاکی به تفکیک مقادیر همبستگی

مقادیر	F	درجه آزادی اثر	درجه آزادی خطا	سطح معناداری	اندازه اثر
صفر لامبدای ویلکز ۰,۹۰۵	۳۷۱,۲۴۴	۱	۳۵۲۰	۰,۰۰۱	۰,۰۹۵
۰/۲۵ لامبدای ویلکز ۰,۸۰۳	۸۶۴,۲۷۱	۱	۳۵۲۰	۰,۰۰۱	۰,۱۹۷
۰/۵ لامبدای ویلکز ۰,۶۸۸	۱۵۹۵,۷۶	۱	۳۵۲۰	۰,۰۰۱	۰,۳۱۲

در جدول فوق نتایج آزمون‌های چند متغیری جهت بررسی تفاوت بین میانگین نمرات ترستونی و کلاسیک به تفکیک برای ضرایب همبستگی IRT روایی ملاکی در دو روش IRT صفر، ۰/۲۵ و ۰/۵ ارائه شده است. بر اساس مندرجات جدول تفاوت بین دو روش ترستونی و کلاسیک برای هر سه ضریب همبستگی معنی دار بوده و میزان ضریب روایی ترستونی بیشتر از روش کلاسیک است. بیشترین تفاوت بین دو روش IRT ملاکی در روش نیز برای ضریب همبستگی ۰/۵ است.

بحث و نتیجه گیری

هدف اصلی این پژوهش شناسایی رویکرد بهینه تجزیه و تحلیل داده‌های ایستو در سنجش ویژگی‌های شخصیتی بود. بدین منظور با استفاده از نرم افزار MPLUS و بسته نرم افزاری IRT به شبیه سازی فرمت‌های مختلف مقیاس‌های پاسخ بایست پرداخته شد و دو رویکرد نمره گذاری به روش سنتی و روش IRT تورستونی در برخی از مهمترین پیامدهای روانسنجی مورد مقایسه قرار گرفتند که در ادامه خلاصه یافته‌های حاصل پرداخته می‌شود. یافته‌های تحقیق نشان داد که میانگین همبستگی بین نمرات برآورد شده و نمرات واقعی و همچنین پایایی واقعی در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است. در مقابل میانگین پایایی تجربی برآورده شده به روش کلاسیک بالاتر از روش IRT ترستونی است. همچنین، بر اساس نتایج بدست آمده همبستگی بین نمرات برآورد شده و

نمرات واقعی و همچنین پایایی واقعی در روش IRT ترستونی نسبت به روش کلاسیک بالاتر برآورد شد. در مقابل میانگین پایایی تجربی در روش کلاسیک بالاتر از روش IRT ترستونی برآورده شد. همچنین یافته‌های تحقیق نشان داد که میانگین ضرایب روایی ملاکی برآورد شده در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است. از طرف دیگر، یافته‌های تحقیق نشان داد که میانگین همبستگی بین نمرات برآورد شده و نمرات واقعی برای هر سه ضریب همبستگی (صفر، ۰/۲۵ و ۰/۵۰)، در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است. نتایج آزمون‌های چند متغیری جهت مقایسه همبستگی بین نمرات برآورد شده و نمرات واقعی در دو روش IRT ترستونی و کلاسیک، نشان داد که میزان اختلاف بین دو روش IRT ترستونی و کلاسیک با توجه به مقدار ضرایب همبستگی بین عامل‌ها متفاوت بوده است. نتایج آزمون‌های چند متغیری جهت مقایسه همبستگی بین نمرات برآورد شده و نمرات واقعی در دو روش IRT ترستونی و کلاسیک به تفکیک برای ضرایب همبستگی صفر، ۰/۲۵ و ۰/۵ نشان داد که میزان همبستگی بین نمرات برآورد شده و نمرات واقعی در روش IRT ترستونی بیشتر از روش کلاسیک است و بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی ۰/۵ است. همچنین یافته‌های تحقیق نشان داد میانگین نمرات پایایی واقعی برای هر سه ضریب همبستگی، در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است. بطوری که بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی ۰/۵ بود. از طرف دیگر، یافته‌های پژوهش نشان داد که میانگین نمرات پایایی تجربی برای ضرایب همبستگی ۰/۲۵ و ۰/۵ با روش کلاسیک بیشتر از نمرات حاصل از روش IRT ترستونی است. ولی برای ضریب همبستگی صفر میانگین نمرات روش IRT ترستونی بیشتر از نمرات کلاسیک است. یافته‌های تحقیق نشان داد که تفاوت بین دو روش IRT ترستونی و کلاسیک برای هر سه ضریب همبستگی معنی‌دار بوده و میزان پایایی تجربی برای ضرایب همبستگی ۰/۲۵ و ۰/۵ در روش کلاسیک بیشتر و برای ضریب همبستگی صفر در روش IRT ترستونی بیشتر است. بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی ۰/۵ بود. همچنین یافته‌های تحقیق نشان داد که میانگین نمرات روایی ملاکی برای هر سه ضریب همبستگی، در روش IRT ترستونی بیشتر از نمرات حاصل از روش کلاسیک است. بطوریکه، بیشترین تفاوت بین دو روش نیز برای ضریب همبستگی ۰/۵ بود.

در زمینه برتری روش نمره گذاری IRT ترستونی نسبت به روش سنتی یافته‌های این پژوهش با یافته‌های پژوهش‌های Maydeu-Olivares and Brown، 2010؛ وانگ و همکاران 2016؛ وانگ و همکاران (2017) همسو است. بطوریکه اکثر پژوهش‌های انجام شده در این حوزه نشانگر این هستند که مدل IRT ترستونی دارای توانایی حصول برآوردهای صفت مکنون دقیق‌تری است. برای مثال Maydeu-Olivares (2011) تعدادی مطالعه شبیه سازی به منظور سنجش بهبود پارامترهای سؤال و نمرات مکنون تحت شرایط مختلف انجام دادند. آن‌ها به این نتیجه رسیدند که هنگامی که ملزومات طرح پارامترهایی نظیر تعداد صفات مورد سنجش، اندازه بلوک، تعداد سؤالات در هر صفت، همبستگی بین صفات، جهت نمره گذاری سؤالات را در نظر می‌گیرد نتایج برآوردهای صفت مکنون دقیق بدست می‌آید.

در تبیین این یافته باید گفت مدل IRT ترستونی چارچوبی برای برآورد پارامترهای سؤال و بدست آوردن نمرات فردی برای هر پرسشنامه‌ای که از فرمت رتبه بندی یا مقایسه‌ای تشکیل شده فراهم می‌کند. بنابراین، بعد از کد گذاری داده‌های پاسخ-بایست با استفاده متغیرهای پیامدی دو جمله‌ای، مدل می‌تواند در چارچوب آشنای SEM برای برآورد و نمره گذاری توسط نرم افزار MPLUS برقرار شود. پرسشنامه‌ها هر تعداد از صفات را اندازه گیری می‌کند و با استفاده از رتبه بندی بلوک‌های دارای اندازه‌های مختلف می‌تواند مدل سازی شود. Maydeu-Olivares and Brown (2011) مطالعات شبیه سازی گسترده‌ای را انجام دادند که نتایج آن نشان گر این است که پارامترها و نمرات فردی پارامترها می‌تواند به طور خیلی دقیق در طرح‌های گزینه-بایست با استفاده از رتبه بندی کامل مورد برآورد قرار بگیرد. در تبیین نتایج حاصله برای فرضیه‌های پژوهش می‌توان گفت که اولاً، نمرات IRT گزینه-بایست اجازه تغییر پذیری در مکان‌های نیمرخ را می‌دهند و بنابراین به طور مستقیم برای مقایسه بین افراد قابل تفسیر هستند. نمره گذاری IRT تفاوت‌های غیر تکراری با نمرات ایپسیتیو در رتبه بندی پاسخ دهندگان را بدست می‌آورد. دوباره ترتیب بندی به طور سیستماتیک بوده و نیمرخ کل را (به جای نمرات فردی) به نیمرخ هنجاری مربوطه مرتبط می‌کند. ثانیاً، نمرات IRT گزینه=بایست نسبت به نمرات ایپسیتیو دقت اندازه گیری بهتری را فراهم می‌کنند. این نتیجه، شگفت‌انگیز نیست، چرا که روش شناسی نظریه سؤال پاسخ بر اساس مدل استوار است و اکثر اطلاعات در هر تصمیم‌گیری را اتخاذ می‌کند. بر

عکس، نمره گذاری ایپسیتیو مدلی را فرض می‌کند که یک پیشینی غلط است، و منجر به تخطی از مفروضه‌های نظریه سنتی می‌شود. علاوه بر این، مدل IRT میانگین‌های برآورد خطاهای استاندارد شرطی را برای هر یک از ترکیبات فردی نمرات فراهم می‌کند.

یافته‌های پژوهش فعلی نشان داد که روایی معیار نمره‌گذاری مبتنی بر IRT ترستونی نسبت به روش سنتی نمره‌گذاری نتایج بهتری دارد. یافته‌های این پژوهش با نتایج یافته‌های والتون و همکاران (2019)؛ آنگویانو کاراسکو و همکاران (2015) همسو است. برای مثال، در یکی از مطالعات که توسط آنگویانو کاراسکو و همکاران (2015) انجام شد، محققان شواهد روایی برای مقیاس هوش هیجانی که با روش گزینه-بایست ساخته شده بود را مورد بررسی قرار دادند و به این نتیجه رسیدند که نمرات حاصل شده آر روش IRT ترستونی دارای شواهد روایی معیار و همگرایی بزرگتری نسبت به رویکرد ایپسیتیو کلاسیک هستند. بر عکس، یافته‌های حاصل از پژوهشی که بر روی مقیاس خود شیفتگی توسط وزل و همکاران (2016) انجام شد نشان داد که نمرات حاصل از دو رویکرد دارای همبستگی‌های مشابهی با متغیرهای مرتبط با معیار هستند. در تبیین این یافته می‌توان گفت که نمره‌گذاری IRT محدودیت‌های نمره‌گذاری ایپسیتیو را که مجبور می‌کند همه همبستگی‌ها با یک معیار خارجی با مجموع صفر شود را از بین می‌برد، بنابراین هر ضریب روایی جعلی را حذف می‌کند. در همین زمان، به خاطر اینکه نمره‌گذاری IRT بر اساس پاسخ‌های گزینه-بایست است، این نوع نمره‌گذاری روایی آزمون را با کاهش سوگیری‌های پاسخ افزایش می‌دهد. همچنین، ساختار زیربنایی (روایی سازه) نمرات گزینه-بایست IRT محور تفاوتی با نمرات هنجاری ندارد. زمانی که نمره‌گذاری IRT اعمال می‌شود، این ساختار به راحتی پیکربندی می‌شود. متوسط همبستگی درونی مقیاس منفی نمی‌شود، و پرسشنامه‌های گزینه-بایست را می‌توان با فنون تحلیل داده‌های استاندارد که کاربران داده‌های هنجاری استفاده می‌کنند تحلیل عاملی کرد.

یافته‌های پژوهش فعلی نشان داد که پایایی حقیقی نمره‌گذاری IRT ترستونی نسبت به روش سنتی بالاتر است، نتایج یافته‌های این پژوهش با نتایج یافته‌های Maydeu-Olivares and Brown (2011) و Merk و همکاران (2017) همسو است. برای مثال، Merk و همکاران (2017) پژوهشی با عنوان "پرسشنامه سیستم‌های ارزشی انگیزشی (MVSQ): تحلیل روانسنجی با استفاده از مدل‌های IRT ترستونی گزینه-بایست انجام دادند. نمونه

پژوهش متشکل از ۱۲۱۷ دانشجوی و متخصصی بودند. مقایسه نمرات مقیاسی هنجاری (IRT) و نمرات ایپستيو دلالت بر این داشت که نمرات IRT به طور زیادی از محدودیت‌های ناشی از ایپستيو بودن آزاد است و بنابراین اجازه مقایسه بین فردی نمرات مقیاس را می‌دهد. پایایی تجربی با استفاده از رویکرد شبیه سازی وابسته به نمونه برآورد شد که نتایج دال بر برآورد خوب و قابل قبول آن بود. علاوه بر این، مطالعات روایی شواهدی از هر دو روایی سازه و وابسته به معیار ارائه داد. نمرات مقیاسی و ارتباطات نمرات با سن و جنسیت عمدتاً در راستای انتظارات تجربی و نظری بودند و نتایج تحلیل چند صفتی - چند روشی از روایی سازه همگرا و واگرا حمایت کرد.

در زمینه تأثیر طول آزمون بر برآورد پارامترهای آزمون و تأثیر آن بر پیامدهای آزمون نظیر پایایی، روایی و... یافته‌های پژوهش فعلی با یافته‌های پژوهش‌های انجام شده در این حوزه از پژوهش همسویی دارد (پرلر، ۲۰۰۷؛ کروین و همکاران، ۲۰۱۴) همسو است. در تبیین این یافته می‌توان گفت که برای آزمون‌های کوتاه‌تر، نتایج بدست آمده مبهم بوده و به نظر می‌رسد که نظریه کلاسیک نتایج مشابه با نتایج IRT بدست می‌آورد. و در آزمون‌های کوتاه‌تر استفاده از نظریه کلاسیک موجب می‌شود که تغییرات در نمرات نسبت به نظریه IRT تغییرات را بهتر آشکار می‌کند، اما باید توجه داشت که نرخ آشکار سازی تغییرات باید با در نظر گرفتن نرخ‌های خطای نوع اول تجربی در شرایطی که هیچ تغییری وجود ندارد مورد تفسیر قرار گیرد. برای آزمون‌های کوتاه‌تر، برای آزمون‌های همگن از نظر دشواری سؤال، نرخ‌های خطای نوع اول تجربی در کل برای آزمون‌های نمره گذاری شده با روش سنتی نسبت به روش IRT ترستونی بالاتر است، و در وسط مقیاس تنها نرخ‌های خطای نوع اول بالاتر از سطح آلفا است. بنابراین، برای آزمون‌های کوتاه نظریه سؤال پاسخ پیشنهاد می‌کند که تغییر فردی نسبت به نظریه کلاسیک کمتر اتفاق می‌افتد.

مدل IRT ترستونی معرفی شده در این پژوهش، فرآیند تصمیم‌گیری برای پاسخ به پرسشنامه‌های شخصیت گزینه-بایست، که صفات چندگانه را اندازه می‌گیرند، توصیف می‌کند. این مدل می‌تواند برای هر ابزار گزینه-بایست، که از آیتم‌های متناسب با مدل پاسخ غالب تشکیل شده، و با هر تعداد صفت مورد اندازه‌گیری و برای بلوک‌های با اندازه‌های مختلف (زوجی، سه تایی، چهارتایی و ...) مورد استفاده قرار گیرد. این باعث می‌شود که این مدل به شکل گسترده‌ای برای بسیاری از پرسشنامه‌های گزینه-بایست مانند پرسشنامه

شخصیت شغلی (SHL, 2006)، پرسشنامه سبک ارتباط مشتری (SHL, 1997)، و بررسی ارزش‌های بین فردی (Gordon, 1976) قابل استفاده باشد و همچنین برای طراحی پرسشنامه در آینده نیز مفید خواهد بود.

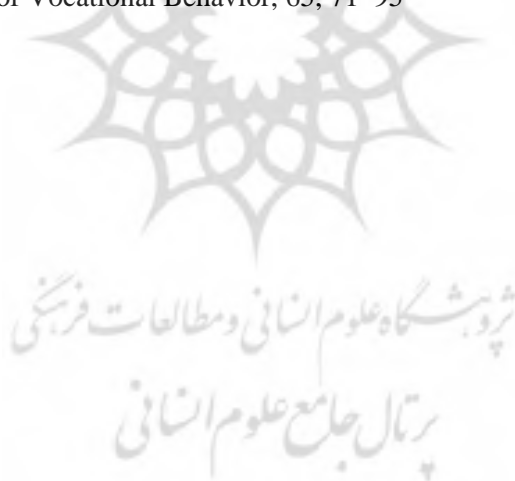
مدل IRT ترستونی را می‌توان برای برآورد و نمره‌گذاری بوسیله یک نرم‌افزار عمومی، در چارچوب یک SEM آشنا، قرار داد (در این مقاله از MPLUS استفاده شد). این مدل همچنین ابزاری برای برآورد پایایی پرسشنامه‌های گزینه-بایست، که با استفاده از روش‌های کلاسیک مشکل ساز بوده است، را فراهم می‌کند (Baron, 1996؛ Tenopyr, 1998).

رویکرد IRT پیشنهاد شده امکان استفاده از قالب گزینه بایست، که موجب کاهش سوگیری پاسخ می‌شود، را فراهم می‌کند، در حالی که مزایای تکنیک‌های تحلیل داده برای پرسشنامه‌های تک-محرك را نیز شامل می‌شود. البته قالب گزینه-بایست نمی‌تواند ایرادات موجود در ساختار آزمون را اصلاح کند، و گاهی ممکن است آنها را بیشتر آشکار سازد. همانطور که نشان داده شد، ساخت یک پرسشنامه گزینه-بایست در مقایسه با پرسشنامه تک-محرك نیازمند توجه و در نظر گرفتن عوامل بیشتری است. مشروط بر اینکه این فاکتورها به دقت مورد توجه قرار گیرند و دقت کافی در ترکیب عبارات برای بلوک‌های گزینه-بایست صورت گیرد، این قالب می‌تواند مزایای قابل توجهی داشته باشد. با از بین بردن ویژگی‌های مختص داده‌های ایستتو، امید می‌رود موانع نظری بر سر راه استفاده از قالب گزینه-بایست از میان برداشته شود.

References

- Baron, H., 1996. Strength and limitations of ipsative measurement. *J. Occup. Organ. Psych.* 69, 49-56.
- Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology*, 86, 965-973.
- Brown, A., & Maydeu-Olivares, A. (2014). Modeling forced-choice response formats. In P. Irwing, T. Booth, & D. Hughes (Eds.). *The Wiley handbook of psychometric testing*. John Wiley & Sons.
- Dunlap, W.P., & Cornwell, J.M. (1994). Factor analysis of ipsative measures. *Journal Multivariate Behavioral Research*. 29,115-126.
- Gordon, L.V. 1976. "Survey of interpersonal values. Revised manual". Chicago, IL: Science
- He, J., & van de Vijver, F. J. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, 55, 794-800.

- Maydeu-Olivares, A., & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research*, 45(6), 935-974.
- Meade, A. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organisational Psychology*, 77, 531-552.
- [Merk J.](#), [Schlotz W.](#), [Falter T.](#) (2017). The Motivational Value Systems Questionnaire (MVSQ): Psychometric Analysis Using a Forced Choice Thurstonian IRT Model. *Front Psychol*. 8(1), 1-14.
- Semadar, A., Robins, G., & Ferris, G. R. (2006). Comparing the validity of multiple social effectiveness constructs in the prediction of managerial job performance. *Journal of Organizational Behavior*, 27, 443-461.
- SHL. (2006). OPQ32 technical manual. Surrey, UK. Author.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods*, 15(3), 463-487.
- Tenopyr, M.L. (1988). Artfactual reliability of forced-choice scales. *Journal of Applied Psychology*, 78, 749-751.
- Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, 65, 71-95



استناد به این مقاله: طیرانی راد، علی، فلسفی نژاد، محمدرضا، فرخی، نورعلی و برجعلی، احمد. (۱۴۰۱). روایی و پایایی نمرات بدست آمده از مدل نظریه سوال پاسخ ترستونی در آزمون های گزینه-بایست. *فصلنامه اندازه گیری تربیتی*، ۱۳(۴۹)، ۷-۲۹. doi: 10.22054/jem.2021.45918.1954



Educational Measurement is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.