

کاربرد مدل اندازه‌گیری راش در ارزشیابی ویژگی‌های اندازه‌گیری آزمون مهارت‌های دیداری- حرکتی (TVMS-R)

اصغر مینائی^۱

تاریخ وصول: ۹۳/۲/۲

تاریخ پذیرش: ۹۳/۹/۵

چکیده

ویژگی‌های اندازه‌گیری ابزارها، آزمون‌ها و مقیاس‌ها را می‌توان با استفاده مدل اندازه‌گیری راش بررسی و ارزشیابی کرد. هدف مطالعه حاضر معرفی ویژگی‌های اندازه‌گیری مدل راش و کاربرد آن در ارزشیابی کیفیت اندازه‌گیری نسخه تجدیدنظر شده آزمون مهارت‌های دیداری- حرکتی (TVMS-R) بود. داده‌های مورد استفاده در پژوهش مربوط به گروه نمونه پژوهش فرهنگ و مینائی (۱۳۸۳) است که در سال تحصیلی ۸۲-۱۳۸۱ در شهر تهران به اجرا درآمد و با استفاده از روش نمونه‌گیری چند مرحله‌ای تعداد ۱۲۸۱ دانش‌آموز (۶۵۵ پسر و ۶۲۶ دختر) با دامنه سنی ۳ تا ۱۴ سال در مهدهای کودک، مراکز پیش دبستانی و مدارس ابتدائی و راهنمائی دخترانه و پسرانه شهر تهران انتخاب و مورد آزمون قرار گرفتند. به طور کلی، یافته‌ها نشان داد که ۶۴ سؤال از ۱۴۲ سؤال آزمون TVMS-R با مدل راش برازش ندارند که در کنار نتایج حاصل از تحلیل مؤلفه‌های اصلی روی پس‌مانده‌ها و آزمون t بیانگر این است که آزمون مورد نظر بر خلاف ادعای مولف آن، تک بُعدی نیست و در مقایسه با سطح توانایی آزمودنی‌ها از دشواری پایین‌تری برخوردار است. در واقع، یافته‌ها نشان داد که آزمون TVMS-R برای اندازه‌گیری یکپارچگی دیداری- حرکتی افرادی که در دو انتهای بالا و پایین پیوستار صفت قرار دارند ابزار مناسبی نیست. تحلیل کارکرد افتراقی سوال نیز نشان داد که ۷ سوال در بین دو گروه دختر و پسر از کارکرد افتراقی (DIF) برخوردارند. بطور کلی با توجه به یافته‌های پژوهش حاضر می‌توان گفت که آزمون TVMS-R به شکل فعلی برای استفاده‌های بالینی چندان مناسب نیست و لازم است در ویرایش‌های بعدی که صورت می‌گیرد برخی اصلاحات اساسی روی آن صورت گیرد.

۱- استادیار دانشگاه علامه طباطبائی، as_minai@yahoo.com

واژگان کلیدی: مدل‌پردازی راش، مهارت‌های دیداری-حرکتی، ویژگی‌های اندازه‌گیری، ارزشیابی.

مقدمه

بسیاری از کودکان مبتلا به ناتوانایی‌های نورولوژیکی، تاخیرهای رشدی، مشکلات یادگیری، و مشکلات مهارت‌های حرکتی که به مراکز کاردرمانی ارجاع داده می‌شوند از نظر یکپارچگی دیداری-حرکتی نیز دارای مشکل هستند (دان، ۱۹۹۶، نقل از براون، آنس‌ورث و لیونز، ۲۰۰۹). یکپارچگی دیداری-حرکتی به میزان هماهنگی بین ادراک دیداری و حرکات ظریف دست اشاره دارد (بری، ۱۹۹۷، نقل از براون و همکاران، ۲۰۰۹) که نقش مهمی در جنبه‌های مختلف عملکرد تحصیلی مانند خواندن و دستخط دارد (گولداستند، کاسلو و پاروش، ۲۰۰۵، نقل از براون و همکاران، ۲۰۰۹).

در ۵ تا ۶ درصد از کودکان سن مدرسه، مشارکت در تکالیف و فعالیت‌های روزانه مدرسه، خانه و جامعه، تحت تاثیر مشکل در مهارت‌های حرکتی فرار می‌گیرد (مک‌هال و سرماک، ۱۹۹۲، نقل از براون و همکاران، ۲۰۰۹). تشخیص نامناسب و نادرست مشکلات یکپارچگی دیداری-حرکتی بر زندگی روزمره و رشد حرفه‌ای کودکان سن مدرسه تاثیر منفی می‌گذارد (همیلتون، ۲۰۰۱، نقل از براون و آنس‌ورث، ۲۰۰۹). از اینرو برای کمک به متخصصان حوزه آموزش و سلامت که بتوانند نیازهای این دسته از مددجویان را به شکل مناسب برآورده سازند، لازم است تا ابزارهایی مانند آزمون مهارت‌های دیداری-حرکتی^۲ (TVMS-R) از ویژگی‌های مطلوب اندازه‌گیری برخوردار باشند. بنابراین، ارزشیابی ویژگی‌های روان‌سنجی این آزمون به منظور اطمینان از اینکه نتایج حاصل، معتبر، روا و دقیق است، به ویژه زمانی که نمره‌های آزمون برای اهداف تشخیصی به کار می‌روند، از اهمیت

1. Brown, Unsworth & Lyons
2. Test of Visual Motor Skills- Revised

خاصی برخوردار است. در طول ساخت، رواسازی و ارزشیابی آزمون‌ها می‌توان از مدل اندازه‌گیری راش برای مطالعه ویژگی‌های روان‌سنجی آزمون‌ها و بررسی این مسأله که سؤال‌های تشکیل دهنده آزمون تا چه اندازه با ویژگی‌های خاصی که برای مدل راش بیان می‌شود برازش دارند استفاده نمود. هدف این مقاله، معرفی مدل راش و کاربرد آن در ارزشیابی کیفیت روان‌سنجی آزمون TVMS-R است. به بیان دقیق‌تر در این مقاله تلاش می‌شود تا با کاربرد اصول مدل اندازه‌گیری راش بررسی شود که ۱- آیا تمام سوال‌های آزمون TVMS-R سازه یکپارچگی دیداری- حرکتی را اندازه می‌گیرند (تک بُعدی بودن)؟، ۲- آیا سطح دشواری سؤال‌ها متناسب با سطح توانایی آزمودنی‌ها است؟ و ۳- آیا سوال‌های آزمون TVMS-R در ارتباط با جنسیت از ویژگی تغییرناپذیری^۱ (فقدان DIF) برخوردارند؟

مروری بر مدل اندازه‌گیری راش

تمدن زمانی ظهور پیدا کرد که بشر توانست نظام‌های پولی و استانداردهای وزن و اندازه را ابداع و از مبادله پایاپای اجناس و خدمات دست بردارد (شوماکر و اسمیت^۲، ۲۰۰۷). وزن، قد، سن و جرم دارای واحدهای استاندارد و متریک هستند که امکان انجام مبادله‌ها و مقایسه‌هایی را که لازمه تمدن امروزی است فراهم می‌آورند. با این حال بسیاری از اندازه‌گیری‌های سنتی روانی و تربیتی، فاقد این ویژگی‌های مطلوب هستند. میشل^۳ (۱۹۹۷) بر این اعتقاد است که اندازه‌گیری‌های روان‌سنجی فاقد ویژگی‌های اندازه‌گیری‌های واقعاً علمی هستند. به نظر وی برای ایجاد یک علم کمی، دانشمند بایستی در دو حوزه علمی و ابزاری به تحقیق و فعالیت پردازد. در حوزه علمی باید ساختار کمی صفت یا ویژگی مورد مطالعه را به اثبات برساند و در حوزه ابزاری، شیوه‌هایی برای اندازه‌گیری صفتی که ساختار کمی آن به اثبات رسیده است

1. Invariance
2. Schumacker & Smith
3. Michell

ابداع نماید. به عبارت دیگر دانشمند قبل از ابداع روشی برای اندازه‌گیری صفت باید ثابت کند که صفت مورد نظر دارای ساختار کمی و جمع‌پذیر است.

میشل (۱۹۹۷) معتقد است در روان‌شناسی بطور عمد و آگاهانه از انجام تکلیف علمی خودداری می‌شود. یعنی بدون اینکه تلاشی برای اثبات ساختار کمی صفت مورد اندازه‌گیری به عمل آید و صرفاً با فرض اینکه صفت مورد مطالعه دارای ساختار کمی است، ابزاری برای اندازه‌گیری آن تهیه می‌شود. در همین ارتباط می‌توان گفته معروف ادوارد لی. ثرندایک مبنی بر اینکه «هر چیزی که وجود داشته باشد دارای کمیت است و هر چیزی که دارای کمیت باشد قابل اندازه‌گیری است» را به عنوان شاهی بر این ادعای میشل ارائه داد. بدین ترتیب مشاهده می‌کنیم که زیربنای منطقی اندازه‌گیری‌های روان‌سنجی مورد انتقاد است.

کاملاً واضح است انتقاداتی که به اندازه‌گیری‌های روانی وارد است و تفاوتی که بین این نوع اندازه‌گیری‌ها و اندازه‌گیری‌های علوم طبیعی وجود دارد به ماهیت متغیرهای مورد مطالعه در روان‌شناسی بر می‌گردد. به عنوان مثال، به هیچ عنوان معلوم نیست که برای اندازه‌گیری هوش از چه واحدهای اندازه‌گیری می‌توان استفاده کرد و معنای صفر واقعی در این زمینه چیست. این موضوع در مورد اکثر مفاهیم و متغیرهای روان‌شناختی صادق است. برای مقابله با این مشکل دو راه حل از سوی متخصصان روان‌سنجی ارائه شده است. راه حل اول را استیونس (۱۹۴۶) و راه حل دوم را لوسی و توکی (۱۹۶۴) ارائه داده‌اند (نقل از کلاین، ۲۰۰۰).

راه حل استیونس علی‌رغم اینکه به طور گسترده از سوی متخصصان روان‌سنجی پذیرفته شده و در زمینه‌های عملی نیز موفقیت‌های زیادی به دنبال داشته است، لکن نمی‌تواند پایه و اساس اندازه‌گیری‌های واقعاً علمی در روان‌شناسی قرار گیرد. زیرا تعریف عملیاتی او از اندازه‌گیری تلویحا وجود واقعیت بیرونی و دنیای خارجی را که لازمه فعالیت علمی است نفی

می‌کند (کلاین^۱، ۱۹۹۸). روان‌سنجی کلاسیک و آزمون‌هایی که با استفاده از روش‌های نظریه کلاسیک آزمون^۲ (CTT) ساخته می‌شود بر تعریف استیونس از اندازه‌گیری استوار است. راه حل لوسی و توکی به اندازه‌گیری مشترک جمع‌پذیر^۳ معروف است. آنها معتقدند زمانی که جمع‌پذیری (جمع ریاضی) یا تسلسل فیزیکی متغیر مورد اندازه‌گیری ناممکن باشد، و یا عبارت دیگر زمانی که اندازه‌گیری متغیر در سطح رتبه‌ای باشد، از این روش می‌توان برای اثبات ساختار کمی و جمع‌پذیر آن استفاده کرد.

برخی از نظریه‌پردازان اندازه‌گیری از جمله میشل (۱۹۹۷) و رایت^۴ (۱۹۸۵) معتقدند اندازه‌گیری مشترک جمع‌پذیر، واقعاً علمی است و انتقاداتی را که به لحاظ منطقی بر اندازه‌گیری‌های روانی وارد است برطرف می‌سازد. در اندازه‌گیری مشترک جمع‌پذیر، نمره فرد در متغیر سوم، تابع دو متغیر مستقل از یکدیگر است. در مدل راش احتمال پاسخ درست به سوال (متغیر سوم)، تابع دشواری سوال و جایگاه فرد در صفت مکنون است. از آنجائیکه در مدل راش، دو پارامتر دشواری سوال و توانایی فرد مستقل از یکدیگرند، بنابراین از نظر بسیاری از متخصصان روان‌سنجی از جمله امبرتسون و ریس^۵ (۲۰۰۰)، آندریش^۶ (۱۹۸۸)، بارت^۷ (۲۰۰۳)، باند و فاکس^۸ (۲۰۰۱)، بورسبوم و ملنبرگ^۹ (۲۰۰۴)، رایت (۱۹۹۹)، فیشر^{۱۰} (۱۹۹۵)، کاراباتسوس^{۱۱} (۲۰۰۱)، پرلین، رایت و وینر^{۱۲} (۱۹۷۹) و گرین^{۱۳} (۱۹۸۶) این مدل را می‌توان

-
1. Kline
 2. Classical Test Theory
 3. Additive Conjoint Measurement
 4. Wright
 5. Embretson & Riecke
 6. Andrich
 7. Barrett
 8. Bond & Fox
 9. Borsboom & Mellenbergh
 10. Fischer
 11. Karabatsos
 12. Perline, Wright & Wainer
 13. Green

مصادق اندازه‌گیری مشترک جمع‌پذیر یا نسخه احتمالاتی آن دانست. این اعتقاد و ادعای این افراد از جملاتی که از آنها در زیر نقل شده است کاملاً آشکار است:

امبرتسون و ریس (۲۰۰۱) اظهار می‌دارند «...صرفاً مدل راش است که شرایط اندازه‌گیری مشترک جمع‌پذیر را برآورده می‌سازد و از اینرو در اغلب مواردی که ویژگی‌های مقیاسی اندازه‌گیری، بسیار مهم تلقی می‌شود از این مدل استفاده می‌شود (صص ۱۵۰-۱۴۹)». گرین (۱۹۸۶) بیان می‌کند «از آنجائی که داده‌های روانی و تربیتی از اعتبار کامل برخوردار نیستند لذا فقدان یک نظریه خطا، سودمندی مدل‌های اندازه‌گیری جمع‌پذیر را محدود کرده است. یک استثناء جالب در این زمینه، کاربرد گسترده مدل‌های راش است (ص ۱۴۱)». پرلین و همکاران (۱۹۷۹) معتقدند «ظاهراً بسیاری از روان‌شناسان آگاه نیستند که مدل راش، تحقق عملی اندازه‌گیری مشترک جمع‌پذیر است (ص ۲۳۷)».

در CTT پایه و اساس تفسیر عملکرد آزمودنیها و انجام مقایسه‌ها نمره‌های خام است. متأسفانه این نمره‌ها و نمره‌های حاصل از تبدیل خطی آنها ویژگی‌های اساسی و مطلوب، مانند واحدهای اندازه‌گیری یکسان و عینیت خاص^۱ را ندارند و در نتیجه امکان انجام مقایسه‌های با معنا و تغییرناپذیر براساس این نمره‌ها وجود ندارد. خوشبختانه، مدل اندازه‌گیری راش این ضعفها و کمبودهای CTT را برطرف می‌کند (بریگز و ویلسون^۲، ۲۰۰۳؛ اسمیت، ۲۰۰۲). مدل راش که توسط ریاضی‌دان دانمارکی بنام جورج راش در ۱۹۶۰ مطرح گردید، و در گذشته عمدتاً در اندازه‌گیری‌های آموزشی به کار می‌رفت، امروزه در بسیاری از حوزه‌ها و شاخه‌های علم، از جمله در روانشناسی، پزشکی، بازاریابی، بهداشت و سلامت، پرستاری و... کاربرد دارد. در ادامه برخی از ویژگی‌های مدل اندازه‌گیری راش مورد بحث قرار می‌گیرد.

-
1. specific objectivity
 2. Briggs & Willson

ویژگی‌ها و مزایای مدل راش

مدل راش نه تنها از لحاظ نظری بلکه از نقطه نظر عملی و کاربردی نیز دارای امتیازات قابل توجهی است. این مدل با فراهم آوردن یک روش مناسب برای تبدیل غیرخطی داده‌های خام رتبه‌ای به اندازه‌های فاصله‌ای بر یکی از نگرانی‌ها و دغدغه‌های رایج در ساخت ابزارهای اندازه‌گیری‌های روانی و تربیتی، یعنی دستیابی به عینیت خاص یا تغییرناپذیری مقایسه‌ها غلبه می‌کند (هاگ کوئیست، بروس و گوستاوسون^۱، ۲۰۰۹). علی‌رغم اینکه مفاهیم عینیت و تغییرناپذیری در اندازه‌گیری‌ها و لزوم اندازه‌گیری‌های اساسی^۲ در روان‌شناسی توسط افرادی چون ترستون (۱۹۲۷)، ثرندایک (۱۹۲۶) و لاونیجر (۱۹۴۶) مطرح شده بود (هاگ کوئیست و همکاران، ۲۰۰۹)، با این حال در سراسر تاریخچه روان‌سنجی، سنت و تعریف استیونس از اندازه‌گیری غالب بوده است و تنها در دهه ۱۹۶۰ بود که راش مفهوم عینیت و تغییرناپذیری را در قالب یک مدل آماری مطرح ساخت و روشی برای آزمون آنها ابداع نمود.

مدل راش در واقع خانواده‌ای از مدل‌ها را شامل می‌شود که از آن جمله می‌توان به مدل لوجستیک ساده^۳ (SLM؛ راش، ۱۹۶۰)، مدل مقیاس درجه‌بندی^۴ (RSM؛ آندریش، ۱۹۷۸) و مدل امتیازدهی پاره‌ای (PCM؛ ماسترز، ۱۹۸۲) اشاره نمود. مدل راش اگرچه از ویژگی‌های مطلوب و جذابی برخوردار است لکن به هیچ عنوان وسیله‌ای برای درمان و حل همه مشکلات اندازه‌گیری نیست. این مدل را می‌توان برای تحلیل انواع مختلف داده‌های حاصل از آزمون‌ها استفاده کرد اما وسیله‌ای برای اصلاح و بهبود اندازه‌گیری‌های بد و بی کیفیت نیست.

1. Hagquist, Bruce & Gustavsson
2. fundamental measurement
3. simple logistic model
4. partial credit model

۱- مقایسه‌های تغییرناپذیر

مدل راش یک مدل اندازه‌گیری تک بُعدی است که ملاک تغییرناپذیری را که از خصوصیات مهم اندازه‌گیری‌های اساسی استل دارا است (آندریش، ۱۹۸۸). در واقع می‌توان گفت که مهمترین ویژگی مدل‌های خانواده راش، ملاک تغییرناپذیری است (آندریش، ۲۰۰۴). منظور از تغییرناپذیری این است که ابزار اندازه‌گیری برای تمام افراد به نحو یکسان عمل نماید. معنای تلویحی این مطلب آن است که عملکرد وسیله اندازه‌گیری در میان گروه‌های مختلف افراد، به عنوان مثال در بین دو گروه دختران و پسران، تغییرناپذیر و یکسان باشد. برای مثال، زمانی که از ترازو برای وزن کردن دختران و پسران استفاده می‌شود، کاملاً مسلم فرض می‌شود که ترازو برای افراد هر دو جنس به نحو یکسان عمل می‌کند. به همین نحو، لازمه انجام مقایسه‌های با معنا و روا در بین افراد گروه‌های مختلف این است که کارکرد سوال‌های آزمون در میان افراد گروه‌های مختلف، تغییرناپذیر باشد. از این مطالب چنین بر می‌آید که اندازه‌گیری نبایستی تحت تاثیر هیچ عامل دیگری به غیر از صفت مورد نظر قرار گیرد (تنت و کوناگان، ۲۰۰۷).

معنای دیگر تغییرناپذیری این است که مقایسه بین توانایی دو نفر، بایستی مستقل از نمونه سوال‌های مورد استفاده در آزمون و همچنین مستقل از سایر افراد موجود در تحلیل باشد. به همین نحو، مقایسه بین پارامتر دشواری دو سوال نیز بایستی مستقل از نمونه افراد مورد استفاده و سایر سوال‌های موجود در تحلیل باشد. یعنی باید نسبت بین دشواری هر جفت از سوال‌ها در طول پیوستار صفت (توانایی) ثابت باقی بماند که برای مقیاس‌پردازی در سطح فاصله‌ای لازم و ضروری است (تنت و کوناگان، ۲۰۰۷).

۲- ساختار پاسخ در مدل راش

ساختار پاسخ مورد نیاز برای مدل راش، شکل احتمالی ساختار پاسخ گاتمن است. بدین معنا که برای یک شخص، احتمال پاسخ درست به سوال آسان‌تر بایستی بالاتر از احتمال پاسخ درست

به سوال دشوارتر باشد و برعکس (آندریش، ۱۹۸۸). از این مطلب چنین بر می‌آید که با مقایسه افرادی با توانایی‌های مختلف انتظار می‌رود شخصی که از توانایی بالاتری برخوردار است به تمام سوال‌هایی که شخص ضعیف‌تر، پاسخ درست داده است پاسخ درست بدهد و علاوه بر این به یک یا چند سوال دشوارتر نیز پاسخ درست بدهد. به بیان دقیق‌تر، در مدل یا ساختار گاتمن اگر شخص A و شخص B به ترتیب به ۴ و ۲ سوال پاسخ درست داده باشند، در این صورت شخص A بایستی علاوه بر آنکه به دو سوالی که شخص B پاسخ درست داده است پاسخ درست بدهد باید به دو سوال دیگری نیز که قاعدتا دو سوال دشوارتر هستند نیز پاسخ درست بدهد. الگوی کامل ساختار گاتمن برای ۶ سوال فرضی که براساس دشواری مرتب شده‌اند در جدول ۱ ارائه شده است.

جدول ۱. الگوی ساختار گاتمن

نمره کل	سوال					
	۱	۲	۳	۴	۵	۶
۰	۰	۰	۰	۰	۰	۰
۱	۰	۰	۰	۰	۰	۱
۲	۱	۱	۰	۰	۰	۰
۳	۱	۱	۱	۰	۰	۰
۴	۱	۱	۱	۱	۰	۰
۵	۱	۱	۱	۱	۱	۰
۶	۱	۱	۱	۱	۱	۱

با بررسی جدول فوق می‌توان نتیجه گرفت که در ساختار گاتمن، به عنوان مثال فردی که نمره خام ۳ گرفته است، بایستی به همان دو سوالی که شخصی با نمره ۲، پاسخ درست داده است پاسخ درست بدهد و علاوه بر این باید به سوال ۳ نیز که دشوارتر از آن دو سوال است پاسخ درست بدهد.

مدل گاتمن یک مدل قطعی گرایانه است و لذا در دنیای واقعی امکان مشاهده چنین الگویی از پاسخ‌ها وجود ندارد. در مجموعه داده‌های واقعی، الگوی پاسخ‌ها بطور دقیق و کامل با ساختار گاتمن یکسان نیست. به عبارت دیگر از آنجائی که پاسخ‌ها قطعی گرایانه نبوده بلکه احتمالی هستند، لذا وجود پراش تصادفی در پاسخ‌ها تلویحاً به این معناست که الگوی پاسخ‌های دو نفر که نمره کل یکسان دارند به هیچ عنوان ضرورتی ندارد که دقیقاً یکسان باشد.

۳- نقش نمره کل و نمره‌های خام

برای دستیابی به اندازه‌گیری، لازم است تا پارامترهای سوال و شخص مستقل از یکدیگر برآورد شوند. این موضوع در واقع، عملیاتی‌سازی ملاک تغییرناپذیری است. استقلال پارامترهای سوال و شخص به این معناست که پارامتر سوال‌ها بایستی مستقل از توزیع توانایی افراد و پارامتر افراد نیز مستقل از سوال‌ها باشد. بنابراین لازمه برآورد پارامتر سوال‌ها و اشخاص وجود یک آماره مکفی است که امکان حذف پارامتر مورد نظر را فراهم آورد. اگر داده‌ها با مدل راش برازش داشته باشد مجموع نمره‌های خام فرد در سوال‌ها (نمره خام کل آزمون) برای برآورد پارامتر شخص آماره مکفی محسوب می‌شود. این به آن معناست که پارامتر سوال‌ها را می‌توان از طریق مشروط کردن بر نمره‌های خام افراد برآورد کرد. به همین شکل، مجموع نمره‌های خام سوال که در واقع برابر با تعداد افرادی است که به سوال پاسخ درست داده‌اند برای برآورد پارامتر دشواری سوال، آماره مکفی به حساب می‌آید. آماره مکفی یک ابزار منحصر به فردی است که این امکان را فراهم می‌آورد که در فرایند برآورد، امکان جداسازی پارامترهای سوال و شخص را از یکدیگر جدا نمود که برای اندازه‌های تغییرناپذیر لازم و ضروری است فراهم می‌آورد با این وجود، چنانچه بین پاسخ‌ها و ساختار گاتمن همخوانی وجود نداشته باشد نمره کل را نمی‌توان به عنوان آماره مکفی در نظر گرفت (آندریش، ۱۹۸۵).

از آنجائی که تبدیل نمره‌های خام به برآوردهای پارامتر افراد (یعنی توانایی) یک تبدیل تکنوا است لذا این احتمال وجود دارد که بین آنها همبستگی بالایی مشاهده شود (آندریش،

۱۹۸۲). از این رو، ممکن است افراد وسوسه شوند تا با نمره‌های خام، به خصوص به خاطر جذابیت و سهولت تفسیر آنها، به صورت اندازه‌های خطی برخورد نمایند. با این حال، نمره‌های خام، هر چند که نقش تعیین‌کننده‌ای در برآورد پارامترها دارند لکن حتی زمانی که داده‌ها با مدل راش نیز برازش دارند به خودی خود و فی‌نفسه اندازه‌های خطی نیستند. در واقع رابطه بین نمره‌های خام و برآوردهای پارامترهای اشخاص در طول پیوستار، متغیر است که این موضوع، خود را از طریق تبدیل غیرخطی یا لوجستیک نمره‌های خام به برآوردهای پارامترهای اشخاص نشان می‌دهد. معنای تلویحی این مطلب آن است که تفاوت‌های یکسان بین دو نمره خام در قسمت‌های مختلف پیوستار ممکن است معانی متفاوتی داشته باشد.

۴- برازش مدل - فقدان تغییرناپذیری

از آنجائیکه ویژگی تغییرناپذیری، بخش لاینفک مدل راش است لذا از آن می‌توان برای واری و کنترل داده‌ها استفاده نمود. انطباق بین داده‌ها و مدل، تلویحاً به آن معنا است که از ابزار مورد نظر می‌توان برای انجام مقایسه‌های تغییرناپذیر استفاده نمود (آندریش، ۱۹۸۸). از شرط تغییرناپذیری چنین بر می‌آید که جایگاه نسبی سوال‌ها، یعنی نسبت بین ارزش‌های مکانی تمام جفت سوال‌ها، در سراسر پیوستار صفت مکنون بایستی یکسان باشد (آندریش، ۱۹۸۸). از مدل راش می‌توان برای بررسی تغییرناپذیری در طول پیوستار صفت مکنون و همچنین در میان گروه‌های مختلف استفاده نمود. فقدان تغییرناپذیری در میان گروه‌های معین، به عنوان مثال دو گروه مردان و زنان، معمولاً کارکرد افتراقی سوال^۱ (DIF) نامیده می‌شود و به این معنا است که افرادی از دو گروه مردان و زنان که از توانایی یکسان برخوردارند، در یک سوال خاص بطور متفاوتی عمل می‌کنند (هاگ کوئیست و آندریش، ۲۰۰۴).

مدل راش نه بر مفروضه‌های آماری درباره داده‌ها بلکه بر الزامات و مقتضیات اندازه‌گیری استوار است (آندریش، ۱۹۸۸). ویژگی جذاب مدل راش، یعنی همان تغییرناپذیری، صرفاً زمانی صادق است که داده‌ها با مدل راش برازش داشته و الزامات مدل راش را برآورده سازند. بنابراین کاملاً واضح است که آزمون و بررسی برازش داده‌ها با مدل از اهمیت خاصی برخوردار است. با این حال، باید به خاطر داشت که برای بررسی رابطه داده‌ها و مدل، صرفاً از یک آزمون نمی‌توان استفاده کرد. این بدان معناست که برای بررسی برازش داده‌ها با مدل بایستی از چندین آزمون استفاده کرد. برای آزمون انطباق داده‌ها با مدل راش می‌توان از آماره‌های رسمی مانند آماره‌های X^2 و پس‌ماندهای استاندارد شده، استفاده نمود. علاوه بر این، روشهای نموداری، مانند منحنی ویژگی سوال^۱ (ICC) نیز ابزاری سودمند برای بررسی سوال‌ها هستند (هاگ کوئیست و آندریش، ۲۰۰۴).

مسأله برازش در مدل راش نیز همانند سایر موضوعات و زمینه‌های آماری، یک امر نسبی است، یعنی به توان آزمون آماری بستگی دارد. از این رو، هر چقدر حجم نمونه بزرگتر باشد احتمال شناسایی یا تشخیص انحرافات از مدل راش بالاتر خواهد بود. اگر حجم نمونه به اندازه کافی بزرگ باشد حتی تفاوت‌های ناچیز بین مقادیر مشاهده شده و مقادیر مورد انتظار ممکن است به عنوان عدم برازش مشخص شوند (آندریش، ۱۹۸۸). نکته قابل توجه در ارزیابی برازش سوال‌ها در مدل راش این است که آماره‌های برازش که در این مدل به کار می‌روند، انحرافات از یک وضعیت ایده‌آل را می‌سنجند. در رویکرد راش، مدل، مستقل از داده‌ها ساخته می‌شود و در نتیجه، دیدگاه سنتی در خصوص رابطه بین داده‌ها و مدل، دستخوش تغییر و دگرگونی می‌شود (آندریش، ۲۰۰۴). در دیدگاه سنتی اگر بین داده‌ها و مدل، برازش وجود نداشته باشد مدل دیگری انتخاب و به داده‌ها برازش داده می‌شود. به عبارت دیگر، در این دیدگاه بر مدل‌پردازی داده‌ها تأکید می‌شود. اما در دیدگاه راش اگر بین داده‌ها و مدل همخوانی وجود

نداشته باشد به جای تغییر مدل و انتخاب یک مدل جدید تلاش می‌شود تا با استفاده از اطلاعات تشخیصی که از تحلیل به دست می‌آید، منابع و دلایل عدم برازش داده‌ها با مدل شناسایی و برطرف گردد. به عبارت دقیق‌تر، در دیدگاه راش اعتقاد بر این است که این داده‌ها هستند که به مدل برازش داده می‌شوند و نه برعکس، و لذا اگر بین این دو، همخوانی وجود نداشته باشد منابع و دلایل این عدم برازش در داده‌ها جستجو می‌شود.

۵- چند بُعدی بودن و وابستگی پاسخ

یکی از الزامات مهم مدل راش، استقلال موضعی است. استقلال موضعی به این معنا است که همبستگی بین سوال‌ها بایستی بطور کامل به وسیله صفت مکنون تبیین شود. آن بخش از همبستگی‌های بین سوال‌ها که به وسیله صفت مکنون (یعنی پارامتر توانایی) تبیین نمی‌شود حاکی از وابستگی موضعی است که یا از چند بُعدی بودن سوال‌ها و یا از وابستگی پاسخ‌های سوال‌ها ناشی می‌شود (تننت و کوناگان، ۲۰۰۷؛ ماریاس و آندریش، ۲۰۰۸).

چند بُعدی بودن نشانه وجود بیش از یک صفت مکنون در مقیاس است، در حالیکه وابستگی پاسخ حاکی از این است که پاسخ به یک سوال به پاسخ‌های برخی دیگر از سوال‌ها وابسته است. یک منبع بالقوه چند بُعدی بودن را می‌توان در مقیاس‌هایی با زیرمجموعه‌های متفاوتی از سوال‌ها که جنبه‌های متفاوتی از یک سازه را اندازه می‌گیرند مشاهده کرد. وابستگی پاسخ زمانی پیش می‌آید که پاسخ یک سوال، سرنخ‌هایی برای پاسخ سوال یا سوال‌های بعدی فراهم آورد.

شناسایی نقض استقلال موضعی گاهی اوقات دشوار است. یافته‌ها نشان می‌دهد که آماره‌های سنتی برازش که در مدل راش به کار می‌روند در شرایط خاصی به چند بُعدی بودن، حساس نیستند (اسمیت، ۲۰۰۴). یک چنین عدم حساسیتی ممکن است در ارتباط با شناسایی وابستگی پاسخ نیز رخ بدهد. از آنجائیکه وجود وابستگی پاسخ باعث افزایش اعتبار و وجود

چند بُعدی بودن باعث کاهش آن می شود (ماریاس و آندریش، ۲۰۰۸) در نتیجه با افزایش و کاهش توان آزمون های آماری برآزش، خطر نادیده گرفتن چند بُعدی بودن در مقایسه با وابستگی پاسخ بالاتر خواهد بود.

به خاطر ضعف های آماره های سنتی برآزش که در بالا ذکر گردید عاقلانه تر آن است که جهت بررسی حالت های گوناگون نقض استقلال موضعی از آزمون های آماری که مشخصاً برای شناسایی جنبه های خاصی از نقض استقلال موضعی طراحی شده اند استفاده نمود. برای مثال می توان از همبستگی های پس مانده سوال ها استفاده کرد که در این صورت، همبستگی های بالا ممکن است حاکی از چند بُعدی بودن یا وابستگی پاسخ باشد.

۶- انطباق سطح دشواری آزمون و سطح توانایی آزمودنی ها، و اعتبار

برای داشتن یک اندازه گیری خوب باید بین سطح دشواری آزمون و توانایی افراد همخوانی خوبی وجود داشته باشد. انطباق نامناسب بین سطح دشواری آزمون و توانایی آزمودنی ها، مشکلات گوناگونی بوجود می آورد که از آن جمله می توان به کاهش اعتبار اشاره نمود که می تواند مشکلاتی را در تمایز گذاری بین افراد در طول پیوستار صفت مکنون بوجود آورد. از نقطه نظر برآورد پارامترها، مکان نسبی سوالها و افراد از اهمیت زیادی برخوردار است. زیرا انطباق نامناسب بین سطح دشواری آزمون و توزیع توانایی افراد باعث برآورد نادقیق پارامترهای سوال و شخص می شود و همچنین خطر وقوع نمره های انتهایی را نیز افزایش می دهد. نمره های انتهایی از این لحاظ که از آنها نمی توان برای برآورد پارامتر توانایی استفاده کرد نامطلوب هستند. در این موارد مجبور هستیم پارامتر توانایی شخص را از طریق برون یابی برآورد کنیم (ریان، ۱۹۸۳).

۷- نحوه برخورد با برازش نامناسب داده‌ها با مدل

به طور کلی یک ویژگی مهم و متمایز مدل‌پردازی راش در مقایسه با مدل‌پردازی آماری این است که مدل راش به عنوان یک بازنمایی رسمی از اندازه‌گیری مناسب، تلقی می‌گردد که داده‌ها در برابر آن مورد بررسی قرار می‌گیرند در حالی که از مدل‌های آماری و از جمله از مدل‌های IRT برای توصیف داده‌ها استفاده می‌شود. این مطلب تلویحاً به آن معنا است که در خصوص منشاء عدم برازش یا برازش نامناسب بین داده‌ها و مدل، قضاوت‌های کاملاً متضادی صورت می‌گیرد. در رویکرد مدل‌پردازی آماری اعتقاد بر این است که برازش نامناسب بین داده‌ها و مدل ناشی از نقص و ضعف مدل در تبیین ویژگی‌های مختلف داده‌ها است و لذا از یک مدل کلی‌تر و جامع‌تر که قادر به تبیین و توصیف جنبه‌های مختلف داده‌ها است استفاده می‌شود. در حالی که در مدل‌پردازی راش، عدم برازش داده‌ها و مدل حاکی از ضعف بالقوه در داده‌ها است که از طریق گنجاندن یک پارامتر اضافی که قدرت تمایزگذاری متفاوت سوال‌ها را در نظر می‌گیرد قابل حل نیست. در عوض در مدل‌پردازی راش، عدم برازش یا برازش نامناسب حاکی از نیاز به بهبود و اصلاح اندازه‌گیری است که بایستی از طریق جمع‌آوری داده‌های جدید مجدداً مورد بررسی قرار گیرد. در واقع در رویکرد راش اعتقاد بر این است که برای دستیابی به اندازه‌های با کیفیت، این داده‌ها است که بایستی با مدل برازش داشته باشند و نه مدل با داده‌ها. در حالی که در مدل‌پردازی آماری و از جمله در مدل‌های IRT اعتقاد بر این است که این مدل است که باید با داده‌ها برازش داشته باشد و چنانچه مدلی با داده‌ها برازش نداشته باشد باید کنار گذاشته شود و از یک مدل کلی‌تر استفاده گردد (آندریش، ۲۰۰۴). با این حال، آزمون‌ها را می‌توان به صورت پس‌تجربی و از طریق حذف سوال‌های فاقد برازش یا از طریق تجزیه^۱ سوال‌های دارای DIF بهبود بخشید. از آنجائی که برازش سوال‌ها به چارچوب مرجع، وابسته است، بنابراین سوال‌های مشکل‌دار را بایستی نه به صورت همزمان بلکه بصورت تک به تک، حذف یا تجزیه نمود. از آنجائی که برازش یک سوال به سایر سوال‌های دارای

1. Resolve

برازش بستگی دارد، لذا حذف یک سوال فاقد برازش در مجموعه‌ای که دارای ۲ سوال فاقد برازش است ممکن است باعث شود تا عدم برازش سوال دوم از بین برود و این سوال به صورت تغییرناپذیری عمل نماید. برای حل مشکل فقدان تغییرناپذیری در میان گروه‌های مختلف، تجزیه سوال‌های دارای DIF عموماً بر حذف اینگونه سوال‌ها ترجیح داده می‌شود زیرا حذف سوال‌ها باعث کاهش دقت اندازه‌گیری یا اعتبار می‌شود. فرضیه زیربنایی روش DIF این است که همه سوال‌ها سازه نظری یکسانی را منعکس می‌سازند اما ممکن است در میان گروه‌های مختلف، به عنوان مثال مردان و زنان، به شیوه متفاوتی عمل نمایند (هاگ کوئیست و آندریش، ۲۰۰۴). این فرضیه را می‌توان با تجزیه سوال‌های دارای DIF مورد آزمون قرار داد. به بیان فنی و با فرض وجود DIF جنسیتی، هر سوال فاقد برازش را می‌توان به دو سوال تقسیم کرد. یک سوال برای مردان (که برای زنان عدم پاسخ در نظر گرفته می‌شود) و یک سوال برای زنان (که برای مردان عدم پاسخ در نظر گرفته می‌شود)، سپس مجموعه داده‌های جدید، مجدداً مورد تحلیل قرار می‌گیرد و چنانچه نتایج حاکی از برازش مجموعه سوال‌های جدید با مدل باشد، تاییدی است بر اینکه سوال یا سوال‌های مورد نظر دارای DIF هستند. تجزیه سوال‌های دارای DIF به جای حذف آنها مانع از کاهش دقت، اعتبار و روایی محتوایی آزمون می‌شود. اگر دو یا چند سوال دارای DIF باشند سوال‌ها به صورت متوالی تجزیه می‌شوند. به این صورت که ابتدا سوالی که دارای بیشترین مقدار DIF است، تجزیه می‌شود و سپس مجموعه سوال‌های جدید، مجدداً مورد تحلیل قرار می‌گیرد. این عمل برای سایر سوال‌های دارای DIF نیز صورت می‌گیرد. هدف از این کار این است که بتوان بین سوال‌های دارای DIF واقعی و تصنعی تمایز قائل شد (هاگ کوئیست، ۲۰۰۸).

روش

طرح پژوهش

پژوهش حاضر یک مطالعه مقطعی است که با استفاده از داده‌هایی که قبلاً توسط فرهبد و مینائی (۱۳۸۳) گردآوری شده بود به ارزشیابی ویژگی‌های روانسنجی آزمون TVMS-R با استفاده از مدل راش می‌پردازد. در واقع پژوهش حاضر در زمره تحقیقات تحلیل ثانویه قرار می‌گیرد.

جامعه، نمونه و روش نمونه‌گیری

جامعه آماری و گروه نمونه پژوهش حاضر همان جامعه آماری و گروه نمونه پژوهش فرهبد و مینائی (۱۳۸۳) است که در سال تحصیلی ۸۲-۱۳۸۱ در شهر تهران به اجرا درآمد و با استفاده از روش نمونه‌گیری چند مرحله‌ای تعداد ۱۲۸۱ دانش‌آموز (۶۵۵ پسر و ۶۲۶ دختر) در مهدهای کودک، مراکز پیش دبستانی و مدارس ابتدائی و راهنمائی دخترانه و پسرانه شهر تهران به طور تصادفی انتخاب و آزمون TVMS-R بر روی آنها اجرا گردید. ویژگی‌های جمعیت‌شناختی نمونه مورد مطالعه در جدول ۲ ارائه شده است.

ابزار

آزمون TVMS-R (گاردنر^۱، ۱۹۹۵) با هدف سنجش توانایی کودکان در انتقال ادراکات و برداشت‌های بصری خود به دستشان و به دست آوردن تصویری از ضعف‌ها و قوت‌های کودک در یکپارچه‌سازی توانایی‌های دیداری-حرکتی طراحی گردید. این آزمون که توسط فرهبد و مینائی (۱۳۸۳) برای کودکان و دانش‌آموزان محدوده سنی ۳ تا ۱۳ سال و ۱۱ ماه، استاندارد و هنجاریابی شده است از ۲۳ طرح هندسی تشکیل یافته است که بر اساس سطح دشواری مرتب شده‌اند و کودک بایستی هر یک از این طرح‌ها را کپی نماید. آزمون مذکور،

1. Gardner

که به صورت فردی و گروهی قابل اجرا است، محدودیت زمانی ندارد اما زمان اجرا به طور معمول ۱۰ دقیقه است که البته به سن فرد نیز بستگی دارد.

این ۲۳ طرح هندسی بر اساس ۸ ملاک یا مشخصه^۱، نمره گذاری می شوند. این ۸ ملاک عبارت است از: ۱- بسته بودن؛ ۲- زوایا؛ ۳- قطع کردن و یا رویهم قرار گرفتن خطوط؛ ۴- اندازه کل طرح یا بخشی از آن؛ ۵- چرخش یا وارونه شدن؛ ۶- طول خط؛ ۷- بیرون زدگی یا فقدان بیرون زدگی؛ و ۸- تغییر طرح (فرهد و مینائی، ۱۳۸۳).

تمام این ۸ ملاک نمره گذاری در مورد همه ۲۳ طرح کاربرد ندارد. در برخی از طرح ها از یک ملاک نمره گذاری بیش از یک بار استفاده می شود. برای مثال، طرح ۱ که شکل یک دایره است بر اساس ۳ تا از این ۸ ملاک، نمره گذاری می شود در حالیکه طرح ۲۳ که شکل یک پروانه است بر اساس ۷ تا از این ملاک ها نمره گذاری می شود. طرح ۳ که شکل یک بیضی افقی است دو بار بر اساس ملاک ۸ نمره گذاری می شود و طرح ۷ که یک مثلث متساوی الاضلاع است ۳ بار بر اساس ملاک ۲ نمره گذاری می شود. از این رو، دامنه نمره برای هر یک از ۲۳ طرح از ۳ (طرح بر اساس ۳ ملاک نمره گذاری می شود) تا ۱۰ (طرح بر اساس ۱۰ ملاک نمره گذاری می شود) که متناظر با تعداد ملاک های مورد استفاده در نمره گذاری است، متغیر است.

در مجموع، ۲۳ طرح بر اساس ۱۴۲ ملاک نمره گذاری می شود. هر یک از این ۱۴۲ مشخصه یا ملاک که به صورت ۰ یا ۱ نمره گذاری می شود به عنوان یک سوال مجزا در نظر گرفته می شود. بر این اساس، کمترین و بالاترین نمره ممکن در این آزمون، به ترتیب ۰ و ۱۴۲ است. یعنی هر فرد می تواند نمره ای بین ۰ تا ۱۴۲ به دست آورد.

تحلیل داده‌ها

برای تحلیل داده‌ها در چارچوب رویکرد راش و برآورد ویژگی‌های روانسنجی آزمون TVMS-R از مدل ساده لوجستیک راش (SLM؛ آندریش، ۱۹۸۸) و از نسخه ۳/۷۶ نرم افزار Winsteps (لی‌ناکر، ۲۰۱۲a) استفاده گردید. در این نرم‌افزار برای برآورد پارامترها از روش بیشینه درستنمایی توام^۲ (JML) استفاده می‌گردد.

نرم‌افزار Winsteps برای بررسی برازش سوال‌ها با مدل، دو نوع آماره MnSq و Zstd را ارائه می‌دهد. دامنه آماره MnSq از ۰ تا بی‌نهایت و مقدار مورد انتظار آن ۱ است. مقادیر بالاتر از ۱ بیانگر انحراف از تک بُعدی بودن، و مقادیر پایین‌تر از ۱ حاکی از بیش برازش الگوهای پاسخ با مدل است که به معنای وجود وابستگی در میان پاسخ‌ها یا سوالات است. آماره Zstd که از تبدیل آماره MnSq به یک آماره t به دست می‌آید در نمونه‌های بزرگ دارای توزیع z است. این آماره در واقع، معناداری آماری انحراف داده‌ها از مدل راش را آزمون می‌کند. این آماره به حجم نمونه بسیار حساس است. برای نمونه‌هایی با حجم کوچکتر از ۳۰ نفر هر نوع داده‌ای با مدل برازش می‌یابد، در حالی که برای نمونه‌های بزرگتر از ۳۰۰ نفر هیچ نوع داده‌ای با مدل برازش نمی‌یابد (لی‌ناکر، ۲۰۱۲). بنابر این با توجه به حجم بالای نمونه مورد استفاده در این پژوهش و همچنین با در نظر گرفتن توصیه لی‌ناکر (۲۰۱۲b)، جهت بررسی برازش سوال با مدل از آماره MnSq که به دو شکل *infit* و *outfit* برآورد و گزارش می‌شود استفاده گردید. لی‌ناکر (۲۰۱۲b) آماره *outfit* MnSq را بر آماره *infit* MnSq ترجیح می‌دهد. بنابراین در پژوهش حاضر برای بررسی برازش سوال از آماره *outfit* MnSq استفاده گردید. دامنه قابل قبول برای این آماره از ۰/۷ تا ۱/۳ است (لیو^۳، ۲۰۱۰).

-
1. Linacre
 2. Joint Maximum Likelihood
 3. Liu

در رویکرد اندازه‌گیری راش برای بررسی تک بُعدی بودن سوال‌ها از روش‌های متعدد (تنت و پالانت^۱، ۲۰۰۶) از جمله آماره‌های برازش که در بالا ذکر شد، استفاده می‌شود. با این حال یافته‌های پژوهشی نشان می‌دهد که این آماره‌ها از حساسیت لازم برای شناسایی چند بُعدی بودن برخوردار نیستند. بنابراین منطقی است که در کنار آماره‌های برازش از تحلیل مؤلفه‌های اصلی^۲ روی داده‌های خام یا پس‌مانده‌ها نیز استفاده کرد (اسمیت^۳، ۲۰۰۴). در پژوهش حاضر برای بررسی تک بُعدی بودن سوال‌ها از روش تحلیل مؤلفه‌های اصلی (PCA) روی پس‌مانده‌ها (لی‌ناکر، ۲۰۱۲b) و آزمون t (اسمیت، ۲۰۰۲) استفاده گردید. در روش PCA برای تعیین تک بُعدی بودن سوال‌ها از ملاک‌های پیشنهادی لی‌ناکر (۲۰۰۸) استفاده گردید: ۱- عامل راش (یعنی صفت مکنون زیربنایی مورد اندازه‌گیری) حداقل ۶۰٪ از واریانس را تبیین نماید؛ ۲- ارزش ویژه اولین مقابله^۴ (یعنی اولین مؤلفه پس‌مانده) کمتر از ۳ باشد و کمتر از ۵٪ از واریانس را تبیین نماید.

در روش اسمیت (۲۰۰۲) ابتدا پارامترهای سوال برای تمام سوال‌ها برآورد گردید که ما آن را مرحله A می‌نامیم. سپس سوال‌هایی که مقدار آماره outfit MnSq آنها بالاتر از ۱/۳ بود شناسایی و با انجام یک تحلیل دیگر، پارامترهای این سوال‌ها به صورت جداگانه برآورد گردید که ما آن را مرحله B می‌نامیم. آنگاه تفاوت بین پارامتر دشواری این سوال‌ها در مرحله B از پارامتر دشواری آنها در مرحله A به دست آمد و میانگین این تفاوت‌ها محاسبه گردید که ثابت تعدیل‌کننده^۵ (اسمیت، ۲۰۰۲) نامیده می‌شود. در گام بعد پارامتر توانایی افراد یکبار براساس کل سوال‌های آزمون (سوال‌های مرحله A) و یکبار نیز براساس صرفاً سوال‌های مرحله B برآورد گردید. سپس مقدار ثابت تعدیل‌کننده، که قبلاً به دست آمده بود به پارامتر توانایی

-
1. Tennant & Pallant
 2. Principal Component Analysis
 3. Smith
 4. Contrast
 5. Shift Constant

افراد در سوال‌های مرحله B اضافه گردید. در گام آخر تفاوت بین پارامتر توانایی هر فرد در کل سوال‌ها و سوال‌های مرحله B با آزمون t مستقل مورد آزمون معناداری آماری قرار گرفت. در تعیین معناداری آزمون t سطح خطای نوع اول ($\alpha=0/05$) با استفاده از روش بنجامینی-هاچبرگ^۱ (۱۹۹۵) مورد تعدیل قرار گرفت. اسمیت (۲۰۰۲) معتقد است اگر تعداد آزمون‌های معنادار t بیش از ۵ درصد باشد حاکی از نقض استقلال موضعی و تک بُعدی بودن سوال‌ها است.

تغییرناپذیری یا فقدان کارکرد افتراقی سوال‌ها (DIF) نیز در بین دو گروه دختران و پسران مورد آزمون قرار گرفت. برای آزمون کارکرد افتراقی سوال معمولاً تفاوت بین پارامتر دشواری سوال در دو گروه با استفاده از آزمون t مورد آزمون معناداری آماری قرار می‌گیرد. در پژوهش حاضر با توجه به حجم بالای گروه‌های مورد مطالعه و با توجه با اینکه آزمون‌های معناداری آماری تحت تاثیر حجم نمونه قرار می‌گیرند، بنابراین با استناد به توصیه لی‌ناکر (۲۰۱۲b) و کورت، گرین‌لند و مارگرین^۲ (۲۰۱۰) تصمیم گرفته شد تا وجود حداقل ۰/۵ لوجیت تفاوت بین دشواری سوال در دو گروه دختران و پسران به عنوان معیاری برای تعیین DIF مورد استفاده قرار گیرد. یعنی اگر بین دشواری سوالی در دو گروه دختران و پسران، حداقل ۰/۵ لوجیت تفاوت وجود داشت، آن سوال به عنوان سوالی که از DIF برخوردار است در نظر گرفته می‌شد.

در نرم‌افزار Winsteps (لی‌ناکر، ۲۰۱۲a) ضریب اعتبار شخص^۳ (R_p) که خیلی شبیه به آلفای کرونباخ و KR-20 در CTT است و همچنین شاخص جداسازی شخص^۴ (G_p) (شوماکر و اسمیت، ۲۰۰۷) برآورد و گزارش می‌شود. در واقع این آماره‌ها، توانایی آزمون را در

-
1. Benjamini- Hochberg
 2. Court, Greenland & Margrain
 3. Person Reliability Coefficient
 4. Person Separation Index

جداسازی و پراکنده ساختن افراد در طول پیوستار صفت نشان می دهند. دامنه آماره R_p از ۰ تا ۱ و دامنه آماره G_p از ۰ تا بی نهایت است و لذا هر چه مقدار آنها بالاتر باشد بهتر است. برای مطالعه تناسب یا همخوانی بین سطح دشواری سوالها با سطح توانایی افراد از نمودار رایت (لی ناگر، ۲۰۱۲b) استفاده گردید. در نرم افزار Winsteps میانگین دشواری سوالها در مقیاس لجیت برابر با صفر قرار داده می شود. اگر آزمونی از همخوانی خوبی با سطح توانایی افراد برخوردار باشد (یعنی نه خیلی آسان و نه خیلی دشوار باشد) میانگین توانایی افراد باید به صفر نزدیک باشد. مقادیر مثبت حاکی از این است که گروه نمونه، در مجموع در مقایسه با سطح دشواری آزمون از توانایی بالاتری برخوردار است، در حالیکه مقادیر منفی عکس این مطلب را نشان می دهد.

در مدل های خانواده راش، قدرت تشخیص تمام سوالها یکسان و برابر با ۱ فرض می شود. با این حال، برنامه Winsteps قدرت تشخیص سوالها را، به عنوان نوعی آماره برازش، براساس داده های تجربی برآورد و گزارش می دهد. میزان انحراف پارامتر تشخیص یک سوال از ۱ نشان دهنده میزان عدم برازش آن سوال با مدل است. اگر پارامتر تشخیص سوالی بالاتر از ۱ باشد به آن معنا است که قدرت تمایزگذاری آن سوال بین افراد قوی و ضعیف، فراتر از آن است که از سوالی با این درجه دشواری انتظار می رود. بر عکس، اگر ضریب تشخیص سوالی کمتر از ۱ باشد نشان می دهد که توانایی آن سوال در تمایزگذاری بین افراد قوی و ضعیف، کمتر از آن است که از سوالی با این درجه دشواری انتظار می رود.

نتایج

جدول ۲ ویژگی های جمعیت شناختی گروه نمونه مورد مطالعه را نشان می دهد. گروه نمونه مورد مطالعه مرکب از ۶۵۵ پسر و ۶۲۶ دختر با میانگین سنی ۸/۰۹ و انحراف استاندارد ۳/۲۰ بود.

جدول ۱. ویژگی‌های جمعیت‌شناختی گروه نمونه (N=۱۲۸۱)

ویژگی	فراوانی	درصد
جنسیت		
پسر	۶۵۵	۵۱
دختر	۶۲۶	۴۹
پایه تحصیلی		
مهد کودک/پیش دبستان	۳۶۶	۲۸/۶
اول	۱۲۳	۹/۶
دوم	۱۱۶	۹/۱
سوم	۱۱۰	۸/۶
چهارم	۱۱۵	۹
پنجم	۱۱۵	۹
اول راهنمایی	۱۰۹	۸/۵
دوم راهنمایی	۱۱۶	۹/۱
سوم راهنمایی	۱۱۱	۸/۷
سن		
۳ سال	۱۰۳	۸
۴ سال	۱۲۱	۹/۴
۵ سال	۱۳۴	۱۰/۴
۶ سال	۱۱۶	۹
۷ سال	۱۰۹	۸/۵
۸ سال	۱۱۲	۸/۷
۹ سال	۱۱۱	۸/۶
۱۰ سال	۱۱۷	۹/۱
۱۱ سال	۱۱۲	۸/۷
۱۲ سال	۱۱۷	۹/۱
۱۳ سال	۱۲۹	۱۰

پارامتر دشواری و برازش سوال‌ها

برآوردهای مربوط به دشواری سوال‌ها همراه با خطای استاندارد مدل، پارامتر تشخیص و همچنین مقادیر آماره‌های $infit\ MnSq$ و $outfit\ MnSq$ در جدول ۳ ارائه شده است. دامنه دشواری سوال‌ها از $-۳/۶۸$ تا $۳/۸۰$ با میانگین ۰ و انحراف استاندارد $۱/۴۴$ است.

اگرچه در جدول ۳ مقادیر آماره MnSq infit نیز ارائه شده است، لکن در ارزیابی برازش سوال‌ها صرفاً از آماره MnSq outfit استفاده گردید. دامنه مقادیر این آماره از ۰/۲۶ تا ۷/۴۱ است که حاکی از این است که برخی از سوال‌ها با مدل راش برازش ندارند. سوال‌های فاقد برازش به شکل پر رنگ‌تر مشخص شده‌اند. بررسی آماره MnSq outfit نشان می‌دهد که تعداد ۶۴ سوال (یعنی ۴۵ درصد از سوال‌ها) فاقد برازش با مدل راش هستند.

جدول ۳. پارامتر دشواری، خطای استاندارد، پارامتر تشخیص و شاخص‌های برازش

شماره سوال در تحلیل	پارامتر TVMS-R	پارامتر تشخیص	MnSq		خطای استاندارد مدل	پارامتر دشواری
			Infit	Outfit		
۸۵	b ۱۵/۲	۱	۱/۰۱	۰/۹۸	۰/۰۸	۳/۸۰
۱۱۴	e ۱۹/۲	۰/۹۶	۱/۰۴	۰/۹۱	۰/۰۷	۳/۵۱
۶۴	۱۲/۲	۰/۷۴	۱/۰۸	۲/۴۹	۰/۰۷	۳/۱۶
۱۱	b۳/۸	-۰/۳۶	۱/۴۹	۷/۴۱	۰/۰۷	۳/۱۰
۱۲۵	۲۱/۲	۱/۳۷	۰/۸۲	۰/۶۹	۰/۰۷	۲/۹۷
۱۳۹	c۲۳/۲	۱/۲۱	۰/۹۱	۰/۷۶	۰/۰۷	۲/۸۹
۳۴	c۷/۲	۱/۱۰	۰/۹۷	۰/۸۱	۰/۰۷	۲/۶۹
۱۳۸	b۲۳/۲	۱/۱۵	۰/۹۳	۰/۸۷	۰/۰۷	۲/۶۱
۱۰۵	c۱۸/۸	۱/۱۰	۰/۹۶	۰/۹۰	۰/۰۷	۲/۴۰
۹۶	a۱۷/۸	۱/۱۲	۰/۹۵	۰/۸۴	۰/۰۷	۲/۲۱
۱۲۰	b۲۰/۷	۱/۳۲	۰/۸۳	۰/۹۴	۰/۰۷	۲/۱۴
۶۹	۱۳/۶	۰/۸۷	۱/۰۷	۱/۰۴	۰/۰۷	۲/۰۷
۱۰۶	a۱۹/۲	۱/۰۱	۰/۹۹	۰/۹۶	۰/۰۷	۱/۹۴
۵۴	b۱۰/۲	۰/۸۸	۱/۰۶	۱/۰۵	۰/۰۷	۱/۸۹
۵۰	a۱۰/۲	۱/۲۱	۰/۸۹	۰/۷۶	۰/۰۷	۱/۶۸
۱۲۶	c۲۱/۳	۰/۹۱	۱/۰۳	۱/۳۲	۰/۰۷	۱/۶۸
۲	۱/۸	-۰/۷۱	۱/۸۲	۵/۹۰	۰/۰۷	۱/۶۲
۱۳۷	۲۳/۵	۱/۱۶	۰/۹۲	۰/۸۰	۰/۰۷	۱/۵۷
۹	۳/۱	-۰/۲۷	۱/۶۶	۴/۶۴	۰/۰۷	۱/۵۳
۱۰۰	a۱۸/۸	۱/۰۷	۰/۹۵	۰/۹۲	۰/۰۷	۱/۵۰
۱۱۸	۲۰/۱	۰/۲۴	۱/۴۴	۱/۸۸	۰/۰۷	۱/۴۷
۱۰۳	b۱۸/۸	۱/۲۹	۰/۸۳	۰/۶۷	۰/۰۷	۱/۴۶
۶۷	a۱۳/۲	۱/۰۲	۱/۰۱	۰/۹۰	۰/۰۷	۱/۳۸

۱۱۷	۲۰/۲	۱/۱۳	۰/۹۳	۰/۸۷	۰/۰۷	۱/۳۶
۱۱۲	d۱۹/۲	۱/۲۱	۰/۸۸	۰/۷۲	۰/۰۷	۱/۲۹
۱۲۸	b۲/۲۲	۱/۱۷	۰/۹۰	۰/۷۶	۰/۰۷	۱/۲۸
۱۲۷	a۲۲/۲	۰/۷۳	۱/۲۰	۱/۱۶	۰/۰۷	۱/۲۵
۸۷	a۱۶/۲	۰/۸۷	۱/۱۱	۱/۰۳	۰/۰۷	۱/۲۲
۲۵	a۶/۲	۱/۱۴	۰/۹۲	۰/۸۴	۰/۰۷	۱/۲۰
۱۰۲	۱۸/۳	۰/۵۵	۱/۲۹	۱/۶۰	۰/۰۷	۱/۱۷
۴۰	a۸/۱	۰/۲۲	۱/۵۰	۲/۳۵	۰/۰۷	۱/۱۴
۱۳۶	a۲۳/۲	۱/۱۹	۰/۸۸	۰/۷۲	۰/۰۷	۱/۱۰
۸۰	a۱۵/۲	۱/۱۲	۰/۹۳	۰/۷۸	۰/۰۷	۱/۰۹
۱۳۰	b۲۲/۱	۰/۶۲	۱/۲۷	۱/۳۷	۰/۰۷	۱/۰۹
۹۱	d۱۶/۳	۱/۳۲	۰/۷۷	۰/۶۰	۰/۰۷	۱/۰۶
۸۴	b۱۵/۱	۱/۰۴	۰/۹۸	۰/۸۹	۰/۰۷	۱/۰۳
۶۸	b۱۳/۲	۱/۳۲	۰/۷۷	۰/۶۰	۰/۰۷	۱/۰۲
۷۹	۱۴/۴	۰/۷۸	۱/۱۷	۱/۱۹	۰/۰۷	۱
۷۸	b۱۴/۲	۱/۰۶	۰/۹۶	۰/۹۴	۰/۰۷	۰/۹۹
۸۲	b۱۵/۷	۱/۳۵	۰/۷۱	۰/۵۶	۰/۰۸	۰/۸۴
۲۷	b۶/۲	۰/۷۸	۱/۲۰	۱/۱۷	۰/۰۸	۰/۷۹
۱۰۹	a۱۹/۱	۰/۹۰	۱/۰۷	۱/۱۵	۰/۰۸	۰/۷۹
۱	۱/۱	-۰/۳۳	۱/۹۸	۴/۱۰	۰/۰۸	۰/۷۶
۱۰۷	b۱۹/۲	۰/۸۶	۱/۱۵	۱/۰۳	۰/۰۸	۰/۷۴
۷۴	a۱۴/۲	۱/۲۳	۰/۸۱	۰/۶۷	۰/۰۸	۰/۷۰
۴۶	۹/۲	۰/۸۲	۱/۱۸	۱/۱۰	۰/۰۸	۰/۶۹
۱۰۸	c۱۹/۲	۱/۴۱	۰/۶۴	۱/۴۸	۰/۰۸	۰/۶۷
۱۲۲	۲۱/۸	۱/۱۸	۰/۸۶	۰/۷۰	۰/۰۸	۰/۶۵
۱۴۲	b۲۳/۱	۱/۱۵	۰/۸۶	۰/۸۰	۰/۰۸	۰/۵۹
۱۲۳	a۲۱/۳	۱/۱۵	۰/۸۷	۰/۷۳	۰/۰۸	۰/۴۶
۹۴	16.1	۰/۹۶	۱/۰۳	۱/۰۶	۰/۰۸	۰/۴۴
۳۰	a۷/۲	۱/۰۳	۰/۹۸	۰/۸۶	۰/۰۸	۰/۳۸
۳۱	a۷/۱	۰/۹۴	۱/۰۵	۱/۰۵	۰/۰۸	۰/۳۶
۱۵	۴/۲	۰/۷۹	۱/۲۵	۱/۱۱	۰/۰۸	۰/۳۶
۹۳	b۱۶/۲	۱/۱۷	۰/۸۳	۰/۷۷	۰/۰۸	۰/۳۶
۵۳	b۱۰/۱	۰/۸۸	۱/۱۰	۱/۲۵	۰/۰۸	۰/۳۱
۱۲۱	۲۰/۴	۱/۰۱	۱/۰۱	۰/۸۷	۰/۰۸	۰/۲۹
۸۳	a۱۵/۱	۱/۱۱	۰/۸۷	۰/۸۴	۰/۰۸	۰/۲۳

۱۲۴	b۲۱/۳	۱/۰۵	۰/۹۷	۰/۸۱	۰/۰۸	۰/۲۲
۱۳۱	a۲۲/۷	۱/۰۸	۰/۹۲	۰/۸۳	۰/۰۸	۰/۱۸
۴۸	۹/۱	۰/۵۰	۱/۴۸	۱/۹۱	۰/۰۸	۰/۱۷
۹۹	c۱۷/۸	۱/۰۳	۰/۹۸	۰/۸۹	۰/۰۹	۰/۱۲
۱۳۵	b۲۲/۴	۱/۰۵	۰/۹۷	۰/۸۳	۰/۰۹	۰/۱۰
۹۸	b۱۷/۸	۰/۹۴	۱/۰۸	۱/۰۱	۰/۰۹	۰/۱۰
۱۴۰	۲۳/۳	۱/۲۶	۰/۸۳	۰/۵۵	۰/۰۹	۰/۱۰
۷۷	b۱۴/۱	۱/۰۹	۰/۸۹	۰/۹۳	۰/۰۹	۰/۰۳
۱۱۵	۱۹/۴	۱/۱۲	۰/۸۷	۰/۸۰	۰/۰۹	۰/۰۱
۵۵	a۱۰/۴	۱/۱۹	۰/۸۱	۰/۶۰	۰/۰۹	-۰/۰۲
۱۷	b۴/۳	۰/۷۸	۱/۲۳	۱/۴۷	۰/۰۹	-۰/۰۲
۷۰	c۱۳/۲	۱/۳۵	۰/۶۰	۰/۴۶	۰/۰۹	-۰/۰۳
۵۹	۱۱/۶	۱/۲۳	۰/۷۶	۰/۵۵	۰/۰۹	-۰/۰۵
۱۱۱	۱۹/۷	۱/۲۴	۰/۷۴	۰/۵۳	۰/۰۹	-۰/۰۶
۷	b۲/۲	۰/۹۳	۱/۱۲	۰/۹۷	۰/۰۹	-۰/۰۷
۱۰۴	b۱۸/۵	۱/۲۳	۰/۷۵	۰/۵۶	۰/۰۹	-۰/۰۷
۳۶	۷/۶	۱/۱۴	۰/۸۳	۰/۷۶	۰/۰۹	-۰/۰۹
۱۴۱	a۲۳/۱	۱/۲۰	۰/۷۸	۰/۶۲	۰/۰۹	-۰/۱۲
۸۹	b۱۶/۳	۱/۲۵	۰/۷۲	۰/۵۲	۰/۰۹	-۰/۱۴
۲۶	۶/۶	۰/۹۸	۱/۰۲	۰/۹۶	۰/۰۹	-۰/۱۷
۷۵	۱۴/۵	۱/۲۴	۰/۷۵	۰/۴۹	۰/۰۹	-۰/۲۰
۹۲	e۱۶/۳	۱/۲۰	۰/۷۷	۰/۶۴	۰/۰۹	-۰/۲۱
۵۷	b۱۰/۴	۱/۱۴	۰/۸۳	۰/۷۷	۰/۰۹	-۰/۲۲
۸۶	۱۵/۵	۱/۲۸	۰/۶۹	۰/۴۶	۰/۰۹	-۰/۲۲
۱۲۹	a۲۲/۱	۰/۸۳	۱/۱۸	۱/۳۶	۰/۰۹	-۰/۲۲
۲۳	۲/۵	۰/۷۱	۱/۳۹	۱/۳۰	۰/۰۹	-۰/۲۴
۸۸	a۱۶/۳	۰/۹۵	۱/۰۸	۰/۹۸	۰/۰۹	-۰/۲۶
۳۳	b۷/۲	۰/۹۸	۱/۰۱	۱	۰/۰۹	-۰/۲۹
۳۹	a۸/۴	۰/۹۲	۱/۱۰	۱/۰۶	۰/۰۹	-۰/۳۴
۴۵	c۹/۳	۱/۰۵	۰/۹۶	۰/۷۸	۰/۰۹	-۰/۳۶
۶۲	۱۲/۵	۱/۰۷	۰/۹۳	۰/۸۴	۰/۰۹	-۰/۳۷
۷۱	a۱۳/۱	۱/۱۸	۰/۷۴	۰/۷۹	۰/۱۰	-۰/۵۳
۷۶	a۱۴/۱	۱/۲۳	۰/۷۲	۰/۵۳	۰/۱۰	-۰/۵۴
۱۱۰	b۱۹/۱	۱/۱۵	۰/۸۰	۰/۷۳	۰/۱۰	-۰/۵۸
۲۹	b۶/۱	۰/۷۴	۱/۲۸	۱/۶۸	۰/۱۰	-۰/۶۰

۶۳	a1۲/۳	۱/۰۵	۰/۹۸	۰/۸۲	۰/۱۰	-۰/۶۱
۸۱	a۱۵/۷	۱/۱۴	۰/۸۲	۰/۶۷	۰/۱۰	-۰/۶۴
۲۲	b۵/۶	۰/۷۰	۱/۴۰	۱/۵۴	۰/۱۰	-۰/۶۹
۴۲	b۸/۸	۰/۸۸	۱/۱۴	۱/۳۱	۰/۱۰	-۰/۷۱
۱۳۴	b۲۲/۷	۱/۱۳	۰/۸۳	۰/۷۰	۰/۱۰	-۰/۷۱
۱۱۳	۱۹/۵	۱/۲۴	۰/۷۱	۰/۴۳	۰/۱۰	-۰/۷۴
۹۵	۱۷/۲	۱/۲۳	۰/۷۱	۰/۴۶	۰/۱۰	-۰/۸۰
۴۱	b۸/۱	۰/۷۶	۱/۳۱	۱/۵۸	۰/۱۰	-۰/۸۱
۱۱۹	a۲۰/۷	۰/۶۹	۱/۳۷	۱/۹۵	۰/۱۰	-۰/۸۶
۹۷	۱۷/۵	۱/۱۸	۰/۷۵	۰/۷۳	۰/۱۰	-۰/۸۸
۶	۲/۶	۰/۸۵	۱/۲۲	۱/۱۲	۰/۱۰	-۰/۸۹
۷۳	۱۳/۵	۱/۲۷	۰/۶۶	۰/۳۱	۰/۱۰	-۰/۹۲
۳۵	۷/۵	۱/۳۰	۰/۶۲	۰/۳۰	۰/۱۰	-۰/۹۵
۴۴	b۹/۳	۰/۸۵	۱/۲۱	۱/۳۴	۰/۱۰	-۰/۹۹
۷۲	b۱۳/۱	۱/۳۱	۰/۵۹	۰/۴۰	۰/۱۰	-۰/۹۹
۳۲	b۷/۱	۱/۲۳	۰/۷۰	۰/۴۳	۰/۱۱	-۱/۰۳
۵۲	a۱۰/۱	۱/۱۵	۰/۷۷	۰/۸۱	۰/۱۱	-۱/۰۳
۵۶	۱۰/۶	۱/۲۴	۰/۶۶	۰/۴۰	۰/۱۱	-۱/۲۳
۳۷	a۸/۸	۱	۰/۹۴	۱/۵۲	۰/۱۱	-۱/۲۹
۴	a۲/۲	۰/۸۷	۱/۱۴	۱/۴۲	۰/۱۱	-۱/۳۲
۲۰	a۵/۶	۰/۸۶	۱/۲۳	۱/۱۷	۰/۱۱	-۱/۳۴
۴۹	۹/۵	۱/۰۵	۰/۹۳	۰/۷۸	۰/۱۱	-۱/۳۵
۱۰۱	a۱۸/۵	۱/۰۳	۱/۰۱	۰/۶۸	۰/۱۱	-۱/۳۵
۱۰	a۳/۸	۰/۷۹	۱/۳۵	۱/۲۸	۰/۱۱	-۱/۳۷
۱۲	۳/۵	۰/۸۲	۱/۲۴	۱/۷۵	۰/۱۱	-۱/۴۱
۱۳۲	a۲۲/۴	۱/۱۲	۰/۸۱	۰/۹۲	۰/۱۱	-۱/۴۶
۵	۲/۵	۱/۱۴	۰/۸۷	۰/۳۴	۰/۱۲	-۱/۵۵
۳۸	۸/۷	۱/۱۰	۰/۸۲	۰/۹۸	۰/۱۲	-۱/۵۸
۱۳۳	۲۲/۵	۱/۲۳	۰/۶۸	۰/۲۸	۰/۱۲	-۱/۵۹
۱۳	۴/۵	۱/۰۳	۰/۹۸	۰/۶۸	۰/۱۲	-۱/۶۷
۲۱	b۵/۳	۰/۸۷	۱/۲۴	۱/۰۳	۰/۱۲	-۱/۶۷
۲۸	a۶/۱	۱/۱۳	۰/۸۰	۰/۹۲	۰/۱۲	-۱/۶۷
۹۰	۱۶/۵	۱/۲۴	۰/۶۷	۰/۲۶	۰/۱۲	-۱/۶۹
۱۱۶	۲۰/۵	۱/۱۸	۰/۷۴	۰/۵۳	۰/۱۲	-۱/۸۳
۶۶	b۱۲/۳	۰/۹۶	۱/۰۶	۱/۱۲	۰/۱۲	-۱/۸۴

۵۱	۱۰/۵	۱/۱۷	۰/۷۶	۰/۴۲	۰/۱۳	-۱/۸۹
۱۸	۲۵/۳	۰/۹۳	۱/۱۲	۱/۰۳	۰/۱۳	-۱/۹۲
۳	۱/۴	۰/۵۸	۱/۵۷	۲/۶۵	۰/۱۳	-۲
۸	۲/۱	۰/۹۷	۱/۰۱	۱/۴۲	۰/۱۳	-۲/۰۴
۱۶	۲۴/۳	۰/۹۲	۱/۰۱	۱/۸۱	۰/۱۳	-۲/۰۷
۲۴	۶/۵	۱/۱۵	۰/۸۰	۰/۵۹	۰/۱۳	-۲/۱۲
۴۷	۹/۸	۱/۱۴	۰/۷۵	۰/۶۱	۰/۱۳	-۲/۱۹
۶۱	۱۱/۸	۱/۰۱	۱/۰۳	۰/۷۷	۰/۱۳	-۲/۲۱
۶۵	۱۲/۸	۱/۰۵	۰/۹۲	۰/۹۰	۰/۱۴	-۲/۲۵
۴۳	۲۹/۳	۱/۱۶	۰/۷۱	۰/۷۶	۰/۱۴	-۲/۴۶
۱۹	۵/۵	۱/۰۲	۱	۰/۵۶	۰/۱۵	-۲/۶۱
۱۴	۴/۸	۱/۱۶	۰/۷۰	۰/۵۷	۰/۱۶	-۳/۰۱
۶۰	۱۱/۵	۰/۹۹	۱/۰۴	۰/۶۹	۰/۱۶	-۳/۰۱
۵۸	۱۱/۷	۰/۹۷	۰/۹۶	۱/۹۳	۰/۱۹	-۳/۶۸
M	-	-	۰/۹۸	۱/۰۶	۰/۰۹	۰/۰۰
SD	-	-	۰/۲۴	۰/۸۹	۰/۰۲	۱/۴۴

بُعدیت آزمون

برای آزمون تک بُعدی بودن سوال‌ها علاوه بر آماره‌های برازش سوال از تحلیل مؤلفه‌های اصلی روی پس‌مانده‌های استاندارد شده و همچنین آزمون t (اسمیت، ۲۰۰۲) استفاده گردید. نتایج حاصل از تحلیل مؤلفه‌های اصلی پس‌مانده‌ها نشان داد که عامل راش (یعنی سازه مورد اندازه‌گیری آزمون) ۴۷/۶ درصد از واریانس را تبیین می‌کند که بسیار پایین‌تر از ملاک پیشنهادی لی‌ناکر (۲۰۰۸) قرار دارد. همچنین ارزش ویژه اولین تا سومین عامل پس‌مانده، به ترتیب ۵/۳، ۴/۳، و ۴ بود که بالاتر از مقدار پیشنهادی لی‌ناکر (۲۰۰۸) قرار دارد.

جهت اجرای آزمون t (اسمیت، ۲۰۰۲) بررسی آماره outfit MnSq نشان داد که مقدار این آماره برای ۲۷ سوال بیش از ۱/۳۰ است. این ۲۷ سوال به عنوان سوال‌هایی که منجر به نقض تک بُعدی بودن می‌شوند در نظر گرفته شدند و مراحل محاسبه آزمون‌های t به شکلی که در بخش روش توضیح داده شد صورت گرفت. مجموع تفاوت‌های بین پارامتر دشواری این ۲۷

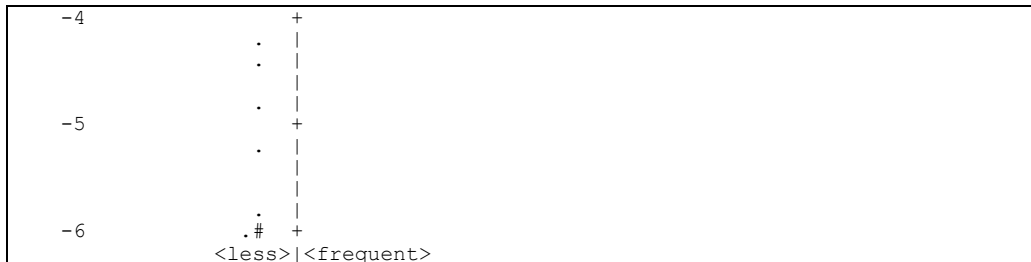
سوال در دو مرحله A و B، ۱/۱۳- بود که با تقسیم آن بر تعداد سوال‌ها (یعنی ۲۷ سوال) مقدار ۰/۰۴-، که همانطور که گفته شد ثابت تعدیل‌کننده نامیده می‌شود، به دست آمد. این ثابت تعدیل‌کننده به پارامتر توانایی افراد در سوال‌های مرحله B اضافه گردید و آزمون‌های معناداری t اجرا شد و سطح معناداری آنها با استفاده از روش بنجامینی- هاجبرگ اصلاح گردید. از ۱۲۸۱ آزمون t تعداد ۱۵۴ مورد (یعنی ۱۲٪) به لحاظ آماری معنادار بود که بسیار بالاتر از حد مجاز قابل قبول، یعنی ۵ درصد (اسمیت، ۲۰۰۲) قرار دارد که حاکی از نقض استقلال موضعی سوال‌ها و تک بُعدی بودن است.

تغییر ناپذیری یا فقدان کارکرد افتراقی سوال

در پژوهش حاضر، کارکرد افتراقی سوال‌ها (DIF) در میان دو گروه دختران و پسران مورد مطالعه قرار گرفت. نتایج تحلیل کارکرد افتراقی سوال نشان داد که از بین ۱۴۲ سوال فقط ۷ سوال ۱/۱، ۱/۸، ۳/۸، ۴/۵، ۹/۱، ۹/۸، و ۲۰/۵ (یعنی تقریباً ۵ درصد) از DIF برخوردارند.

همخوانی بین سطح دشواری سوال‌ها و توانایی آزمودنی‌ها، و اعتبار

دامنه پارامتر دشواری سوال‌ها از ۳/۶۸- تا ۳/۸۰+ با میانگین ۰ و انحراف استاندارد ۱/۴۴ و دامنه پارامتر توانایی افراد از ۷/۰۸- تا ۵/۹۸+ با میانگین ۱/۶۰ و انحراف استاندارد ۲/۱۲ به دست آمد. این یافته بیانگر آن است که افراد در مقایسه با سوال‌ها از توانایی بالاتری برخوردارند. این یافته به صورت تصویری در شکل ۱ که به نمودار رایت نیز معروف است نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سوال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که اکثر افراد از سطح توانایی بالاتر از ۱ قرار دارند، این در حالی است که اکثر سوال‌ها از دشواری پایین‌تر از ۱ برخوردارند. در واقع می‌توان نتیجه گرفت که آزمون TVMS-R برای اندازه‌گیری سطوح بالای توانایی یکپارچگی دیداری- حرکتی چندان مناسب



*هر "#" بیانگر ۷ نفر و هر "۰" بیانگر ۱ تا ۶ نفر است.

شکل ۱. نمودار دشواری سوال‌ها و توانایی افراد (نمودار رایت)

ترتیب سلسله مراتبی دشواری سوال‌ها

ترتیب سوال‌ها بر حسب دشواری آنها در مقیاس لوجیت در جدول ۳ ارائه شده است. در زیر، سوال‌ها از لحاظ دشواری از آسان به دشوار به ۱۴ طبقه، گروه‌بندی شده‌اند.

(a) 9.8; (b) 11.8, 12.8; (c) 9.3a, 5.5, 4.8, 11.5, 11.7; (d) 4.5, 5.3b, 6.1a, 16.5, 20.5, 12.3, 10.5, 5.3a, 1.4, 2.1, 4.3a, 6.5; (e) 10.6, 8.8a, 2.2a, 5.6a, 9.5, 18.5, 3.8a, 3.5, 22.4, 2.5, 8.7, 22.5; (f) 5.6b, 8.8b, 22.7, 19.5, 17.2, 8.1b, 20.7, 17.5, 2.6, 13.5, 7.5, 9.3b, 13.1, 7.1b, 10.1; (g) 6.6, 14.5, 16.3, 10.4, 15.5, 22.1, 5.2, 16.3, 7.2b, 8.4a, 9.3c, 12.5, 13.1, 14.1, 19.1, 6.1b, 12.3, 15.7; (i) 10.1, 20.4, 15.1, 21.3, 22.7, 9.1, 17.8, 22.4, 17.8, 23.3, 14.1, 19.4, 10.4, 4.3b, 13.2, 11.6, 19.7, 2.2b, 18.5, 7.6, 23.1, 16.3; (j) 6.2b, 19.1, 1.1, 19.2, 14.2, 9.2, 19.2, 21.8, 23.1, 21.3, 16.1, 7.2a, 7.1a, 4.2, 16.2; (k) 19.2, 22.2, 22.2, 16.2, 6.2a, 18.3, 8.1a, 23.2, 15.2, 22.1, 16.3, 15.1, 13.2, 14.4, 14.2, 15.7; (l) 10.2, 21.3, 1.8, 23.5, 3.1, 18.8, 20.1, 18.8, 13.2, 20.2; (m) 17.8, 20.7, 13.6, 19.2, 10.2; (n) 7.2c, 23.2, 18.8; (o) 12.2, 3.8b, 21.2, 23.2; (p) 15.2, 19.2

این گروه‌بندی نشان می‌دهد در برخی از طبقات، تعداد سوال‌ها بسیار زیاد است. در واقع می‌توان گفت که در این طبقات برخی از سوال‌ها زائد هستند و بنابراین می‌توان آنها را حذف کرد. در عوض لازم است به برخی از طبقات تعدادی سوال اضافه گردد. با نگاه به این طبقه‌بندی می‌توان دریافت که تعداد سوال‌ها در دو انتهای آسان و دشوار پیوستار صفت، اندک

است که به آن معنا است که آزمون TVMS-R توانایی یکپارچگی دیداری- حرکتی افرادی که در دو انتهای پیوستار قرار دارند را با دقت و به شکل مناسب اندازه گیری نمی کند.

بحث و نتیجه گیری

هدف مقاله حاضر معرفی مدل راش و چگونگی کاربرد آن در ساخت و بهبود کیفیت روان سنجی ابزارهای اندازه گیری روانی تربیتی است. مدل راش، شناسایی و کشف مشکلات و معضلات اندازه گیری، به عنوان مثال تغییرناپذیری را که ممکن است شناسایی آنها از طریق تحلیل های سنتی آسان نباشد تسهیل می کند.

در واقع در پژوهش حاضر، تک بعدی بودن، کارکرد افتراقی، برازش و تناسب سطح دشواری سوال های آزمون TVMS-R با سطح توانایی آزمودنی ها با استفاده از رویکرد مدل پردازی راش مورد بررسی قرار گرفت. نتایج حاصل از تحلیل ها نشان داد که ۴۵ درصد از سوال های آزمون مورد نظر با مدل راش برازش ندارد. همچنین یافته های حاصل از اجرای تحلیل مولفه های اصلی روی پس مانده های استاندارد شده و آزمون t نشان داد که مفروضه استقلال موضعی و تک بعدی بودن برقرار نیست. به عبارت دیگر، براساس نتایج آماره outfit و MnSq و تحلیل مولفه های اصلی و روش t می توان چنین نتیجه گرفت که آزمون TVMS-R، بر خلاف ادعای مؤلف آن (گاردنر، ۱۹۹۵)، تک بعدی نیست و صرفاً سازه کلی یکپارچگی دیداری- حرکتی را نمی سنجد که بیانگر این است که آزمون مورد بحث، چند بعدی است. این یافته با نتایج حاصل از پژوهش براون و آنس ورث (۲۰۰۹) و براون، آنس ورث و لیونز (۲۰۰۹) از همخوانی خوبی برخوردار است. براون و آنس ورث (۲۰۰۹) با استفاده از مدل راش، آزمون TVMS-R را مورد ارزیابی قرار دادند. آنها نیز به این نتیجه رسیدند که سازه یکپارچگی دیداری- حرکتی که توسط این آزمون اندازه گیری می شود، یک سازه تک بعدی نیست و دارای ابعاد مختلفی است. در یک پژوهش دیگر، آنس ورث و همکاران (۲۰۰۹) با استفاده از تحلیل عاملی به شناسایی ساختار زیربنایی آزمون TVMS-R پرداختند. نتایج حاصل از مطالعه

آنها نشان داد که این آزمون از ۴ بُعد یا عامل شکل گرفته است که جمعاً ۴۴/۰۸ درصد از واریانس کل را تبیین می‌کنند.

نتیجه حاصل از تحلیل کارکرد افتراقی سوال نیز نشان داد که تنها ۷ سوال از ۱۴۲ سوال در بین دو گروه دختران و پسران از DIF برخوردارند. در پژوهش براون و آنس‌ورث (۲۰۰۹) نیز مشخص شد که تنها ۶ سوال در بین دختران و پسران دارای DIF هستند.

در ارتباط با تناسب سطح دشواری سوال‌ها با سطح توانایی افراد، نتایج پژوهش حاضر و پژوهش براون و آنس‌ورث (۲۰۰۹) نشان می‌دهد که سطح دشواری سوال‌ها همخوانی خوبی با سطح توانایی افراد ندارد و در واقع می‌توان گفت که سوال‌ها در مقایسه با توانایی افراد از دشواری پایین‌تری برخوردارند. نمودار رایت که در شکل ۱ ارائه شده است نشان می‌دهد که دشواری اکثریت سوال‌ها در دامنه بین ۲- تا ۳ قرار دارد، این در حالی است که دامنه توانایی اکثریت افراد بین ۳/۵- تا ۴/۵ است. در واقع می‌توان گفت که سوال‌های این آزمون از لحاظ دشواری در حد متوسط قرار دارند و می‌توانند افرادی را که از توانایی متوسط برخوردارند به خوبی اندازه‌گیری نمایند در حالیکه برای اندازه‌گیری افرادی که در دو انتهای پیوستار صفت قرار گرفته‌اند چندان مناسب نیستند.

نتیجه حاصل از پژوهش حاضر و مطالعه براون و آنس‌ورث (۲۰۰۹) ادعای گاردنر (۱۹۹۵) را مبنی بر اینکه سوال‌های آزمون از لحاظ دشواری به شکل پیشرونده‌ای مرتب شده‌اند حمایت نمی‌کند. بنابراین، لازم است در اصلاحات و تجدیدنظرهایی که در آینده روی این آزمون صورت می‌گیرد، ترتیب سوال‌ها از لحاظ دشواری اصلاح گردد.

در پایان پیشنهاد می‌شود در پژوهش‌های آتی در مورد آزمون TVMS-R، ۱- ساختار زیربنایی آن با استفاده از روش‌های مناسب مورد بررسی قرار گیرد؛ ۲- تلاش گردد تا در تجدید نظرهایی که روی آزمون به عمل می‌آید از رویکرد مدل‌پردازی راش استفاده گردد؛ ۳- ویژگی‌های اندازه‌گیری آزمون TVMS-R به طور مستمر مورد ارزشیابی قرار گیرد تا شواهد کافی در خصوص سودمندی آن گردآوری گردد.

منابع فارسی

فرهبد، مژگان؛ مینائی، اصغر. (۱۳۸۳). انطباق و هنجاریابی نسخه بازبینی شده آزمون مهارت‌های بینایی - حرکتی، *توانبخشی*، ۱۶ و ۱۷، ۳۹-۴۸.

منابع لاتین

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D.(1982). Using latent trait measurement models to analyse attitudinal data: A synthesis of viewpoints. In: Spearitt, D. (Ed.), *The Improvement of Measurement in Education and Psychology*, (pp. 89-126). ACER, Hawthorn: Victoria.
- Andrich, D.(1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological Methodology*, (pp. 33-80). San Fransisco: Jossey-Bass.
- Andrich, D.(1988). *Rasch Models for Measurement*. Beverly Hills: Sage Publications.
- Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms, In E. V. Smith, Jr., & R. M. Smith (Eds), *Introduction to Rasch Measurement*, (pp. 143-166). Maple Grove, Minnesota: JAM Press.
- Barrett, P.T. (2003). Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 3, 421-439.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling for the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society, Series B*, 57, 189-300.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Borsboom, D., & Mellenbergh, G.J. (2004). Why psychometrics is not pathological. *Theory & Psychology*, 14, 105-120.
- Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch Models. *Journal of Applied Measurement*, 1, 87-100.

- Brown, T., & Unsworth, C. (2009). Construct validity of the Test of Visual-Motor Skills –Revised (TVMS-R): An evaluation using Rasch Measurement Model. *Scandinavian Journal of Occupational Therapy*, 16, 133-145.
- Brown, T., Unsworth, C., & Lyons, C. (2009). Factor structure of four visual-motor instruments commonly used to evaluate school-age children. *American Journal of Occupational Therapy*, 6, 710–723.
- Court, H., Greenland, K., & Margrain, T. H. (2010). Measuring Patient Anxiety in Primary Care: Rasch Analysis of the 6-item Spielberger State Anxiety Scale. *Value IN Health*, 6, 813-819.
- Embretson, S. E., and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fischer, G. H. (1995). Derivations of the Rasch model. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications*, (pp. 15-38). New York: Springer-Verlag.
- Gardner, M. F. (1995). *Test of Visual-Motor Skills-Revised*. Hydesville, CA: Psychological and Educational Publications.
- Green, K. E. (1986). Fundamental measurement: A review and application of additive conjoint measurement in educational testing. *Journal of Experimental Education*, 19, 141-147.
- Hagquist, C.(2001). Evaluating composite measures using Rasch modeling: an illustrative example. *Social and Preventive Medicine* 46, 369–378.
- Hagquist, C. (2008). Psychometric properties of the PsychoSomatic Problems scale: A Rasch analysis on adolescent data. *Social Indicators Research* 86, 511–523.
- Hagquist, C., & Andrich, D. (2004). Is the sense of coherence instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personality and Individual Differences* 36, 955–968.
- Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, 46, 380–393.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, 2, 389-423.
- Kline, P. (1998). *The New Psychometrics: Science, Psychology and Measurement*. London: Routledge.
- Kline, P. (2000). *Handbook of Psychological Testing*. (2nd). London: Routledge.
- Linacre, J. M. (2012). *Winsteps[Rasch measurement computer program, version 3.76]*. Beaverton, Oregon: Winsteps.com.

- Linacre, J. M. (2012). *A User's Guide to Winsteps: Rasch – Model Computer Program*. Beaverton, Oregon: Winsteps.com.
- Linacre, J. M. (2008). *A User's Guide to Winsteps: Rasch – Model Computer Program*. Chicago: Winsteps.com.
- Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. North Carolina: IAP, Inc.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement* 9, 200–215.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355-383.
- Perline, R., Wright, B.D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 2, 237–255.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests* (First published 1960 by the Danish Institute for Educational Research). Chicago: MESA Press.
- Ryan, J.P.(1983). Introduction to latent trait analysis and item response theory. In W. E. Hathaway (Ed.), *Testing in the Schools: New Directions for Testing and Measurement*, 19. Jossey-Bass, San Fransisco, pp. 49–65.
- Schumacker, R. E., & Smith, E. V. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 3, 394-409.
- Smith, E.V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231.
- Smith, R. M. (2004). Fit analysis in latent trait measurement theory. In E. V. Smith, Jr., & R. M. Smith (Eds), *Introduction to Rasch Measurement*, (pp. 73-92). Maple Grove, Minnesota: JAM Press.
- Tennant, A., & Conaghan, P.G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism* 57, 1358–1362.
- Tennant, A., & Pallant, J. (2006). Unidimensionality matters: A tale of two Smiths?. *Rasch Measurement Transactions*, 20, 1048-1051.
- Wright, B. D. (1985). Additivity in psychological measurement. In E. Roskam (Ed.), *Measurement and personality assessment*. North Holland: Elsevier Science.

- Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know*, (pp. 65-104). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Wright, B.D., & Masters, G.N.(1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.
- Wright, B. D. (1999). Rasch measurement models. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment*, (pp. 85-97). New York: Pergamon.
- Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70, 857-860.