

## ویژگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی دوره راهنمایی مدارس استعداد‌های درخشان و نمونه دولتی با استفاده از رویکرد اندازه‌گیری راش: با تأکید بر آزمون‌های هوش و خلاقیت

دکتر اصغر مینائی<sup>۱</sup>

تاریخ دریافت: ۹۴/۰۴/۰۱

تاریخ پذیرش: ۹۴/۰۵/۰۶

### چکیده

مرکز ملی پرورش استعداد‌های درخشان و دانش‌پژوهان جوان که به سمپاد نیز معروف است؛ به‌منظور شناسایی و گزینش دانش‌آموزان نخبه و بااستعداد جهت ورود به مدارس استعداد‌های درخشان، هر ساله آزمون‌هایی را برای دو مقطع راهنمایی و متوسطه اجرا و تعدادی از داوطلبان را انتخاب می‌نماید. پژوهش حاضر با هدف بررسی کیفیت اندازه‌گیری آزمون ورودی دوره راهنمایی بر اساس رویکرد اندازه‌گیری راش و برآورد روایی آن بر اساس هوش خلاقیت به اجرا درآمد. برای این منظور از بین کلیه داوطلبان شرکت‌کننده در آزمون ورودی مدارس استعداد‌های درخشان و نمونه دولتی دوره راهنمایی در سال تحصیلی ۹۳-۱۳۹۲ در شهر تهران با استفاده از روش نمونه‌گیری تصادفی چندمرحله‌ای، تعداد ۲۸۷۲ پاسخنامه (۱۵۴۲ پسر و ۱۳۳۰ دختر) انتخاب گردید. برای تحلیل داده‌ها در چارچوب مدل راش و برآورد ویژگی‌های روان‌سنجی سؤال‌ها از مدل ساده لوجستیک راش و نرم‌افزار Winsteps استفاده گردید. نتایج نشان داد که همه سؤال‌ها به‌غیر از سؤال ۷۴ مربوط به درس ریاضی با مدل راش برازش دارند. دامنه دشواری سؤال‌ها از ۲/۴۶- تا ۱/۸۳، به دست آمد. همچنین نتایج حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد بود. به‌طور کلی، بر اساس یافته‌های پژوهش حاضر، می‌توان گفت که آزمون سمپاد از ویژگی‌های اندازه‌گیری لازم برخوردار نیست و کیفیت لازم برای استفاده به‌عنوان ابزاری برای گزینش دانش‌آموزان برای ورود به مدارس استعداد‌های درخشان و نمونه دولتی را ندارد؛ بنابراین لازم است دست‌اندرکاران سازمان سمپاد در طراحی این آزمون‌ها دقت بیشتری نمایند تا به خاطر کیفیت ضعیف آزمون‌ها سود و زیان خاصی متوجه داوطلبان نشود و دانش‌آموزانی که واقعاً از توانایی‌ها و شایستگی‌های

۱. استادیار گروه سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، as\_minai@atu.ac.ir

بالا برخوردارند وارد نظام آموزشی سمپاد گردند. برای این منظور لازم است تا با تشکیل بانک سؤال، در تهیه این آزمون‌ها از سؤال‌هایی استفاده گردد که از کیفیت لازم برخوردارند.

واژگان کلیدی: راش، روان‌سنجی، سمپاد، دانش‌آموزان

### مقدمه

شناسایی و تشخیص تیزهوشی، قلمروی ضروری در امر آموزش و پرورش نخبگان محسوب می‌شود (کاظمی حقیقی، ۱۳۸۲). مرکز ملی پرورش استعداد‌های درخشان و دانش‌پژوهان جوان که به سمپاد نیز معروف است؛ به‌منظور شناسایی و گزینش دانش‌آموزان نخبه و بااستعداد جهت ورود به مدارس استعداد‌های درخشان، هر ساله آزمون‌هایی را برای دو مقطع راهنمایی و متوسطه اجرا و تعدادی از داوطلبان را انتخاب می‌نماید. لکن تاکنون هیچ پژوهش جدی در خصوص ویژگی‌های روان‌سنجی این آزمون‌ها صورت نگرفته است. بنابراین، هدف پژوهش حاضر بررسی کیفیت اندازه‌گیری آزمون ورودی دوره راهنمایی بر اساس رویکرد اندازه‌گیری راش و برآورد روایی آن بر اساس مقیاس تجدیدنظر شده هوش و کسلر برای کودکان (شهیم، ۱۳۷۳) و پرسشنامه خلاقیت عابدی (عابدی، ۱۳۷۲) است.

برای تحلیل سؤال‌ها و پرورش آزمون‌های معتبر، دقیق و روا مدل‌های گوناگون روان‌سنجی ارائه شده است (به‌عنوان مثال به کروکر و آلجینا، ۱۹۸۶ مراجعه کنید). از جمله مدل‌هایی که امروزه کاربرد وسیعی نیز پیدا کرده است، می‌توان به مدل راش (باند و فاکس<sup>۲</sup>، ۲۰۰۷) اشاره کرد. این مدل در مقایسه با سایر مدل‌های روان‌سنجی از مزایای زیادی برخوردار است. ویژگی منحصر به فرد مدل راش این است که این امکان را به وجود می‌آورد که به شرط برازش و انطباق داده‌ها با مدل، پارامترهای سؤال و شخص را مستقل از یکدیگر برآورد کرد (هاگ کویتس<sup>۳</sup>، ۲۰۰۱).

با توجه به اینکه برآورد ویژگی‌های روان‌سنجی یک آزمون برای یک جامعه آماری، عمدتاً به صورت اکتشافی صورت می‌گیرد، بنابراین در پژوهش حاضر به‌جای بیان فرضیه و

- 
1. Crocker & Algina
  2. Bond & Fox
  3. Hagquist

آزمون آن‌ها به بیان سؤال‌های پژوهش پرداخته شد. در واقع پژوهش حاضر به دنبال پاسخگویی به سؤال‌های زیر است:

- ۱- ضریب اعتبار شخص<sup>۱</sup> و شاخص جداسازی شخص<sup>۲</sup> سؤال‌های آزمون ورودی دوره راهنمایی مدارس استعدادهای درخشان بر اساس مدل اندازه‌گیری راش چگونه است؟
- ۲- آیا سؤال‌های آزمون ورودی دوره راهنمایی مدارس استعدادهای درخشان با مدل اندازه‌گیری راش برازش دارد؟
- ۳- سطح دشواری و قدرت تشخیص سؤال‌های آزمون ورودی دوره راهنمایی مدارس استعدادهای درخشان بر اساس مدل اندازه‌گیری راش چگونه است؟
- ۴- آیا سطح دشواری سؤال‌های آزمون ورودی دوره راهنمایی مدارس استعدادهای درخشان سطح توانایی داوطلبان تناسب و همخوانی دارد؟
- ۵- روایی این آزمون‌ها در ارتباط با آزمون هوش و کسلر برای کودکان و پرسشنامه خلاقیت عابدی چگونه است؟

## روش

### جامعه، نمونه و روش نمونه‌گیری

جامعه آماری پژوهش را کلیه داوطلبان شرکت‌کننده در آزمون ورودی مدارس استعدادهای درخشان و نمونه دولتی دوره راهنمایی در سال تحصیلی ۹۳-۱۳۹۲ در شهر تهران تشکیل می‌دهد. برای انتخاب نمونه موردنظر از روش نمونه‌گیری تصادفی چندمرحله‌ای استفاده گردید. برای این منظور ابتدا از بین حوزه‌های امتحانی شهر تهران، پنج حوزه شماره ۳، ۴، ۸، ۱۰، ۱۲ و ۱۳ به تصادف انتخاب گردید. سپس از بین پاسخنامه‌های داوطلبان شرکت‌کننده در ۵ حوزه فوق، تعداد ۲۸۷۲ پاسخنامه (۱۵۴۲ پسر و ۱۳۳۰ دختر) به صورت کاملاً تصادفی انتخاب گردید.

برای بررسی روایی آزمون سمپاد از میان داوطلبانی که قادر به ورود به مدارس استعدادهای درخشان شده بودند، تعداد ۱۲۲ دانش‌آموز (۶۶ پسر و ۵۶ دختر) به تصادف

1. Person reliability coefficient

2. Person separation index

انتخاب و مقیاس تجدیدنظر شده هوش و کسلر برای کودکان (شهیم، ۱۳۷۳) و پرسشنامه خلاقیت عابدی (عابدی، ۱۳۷۲) بر روی آن‌ها اجرا گردید.

### ابزارهای گردآوری داده‌ها

در پژوهش حاضر برای گردآوری داده‌ها از ۳ ابزار استفاده گردید که در ادامه به معرفی آن‌ها می‌پردازیم.

آزمون ورودی مدارس استعدادهای درخشان و نمونه دولتی دوره راهنمایی

### در سال تحصیلی ۹۳-۱۳۹۲

این آزمون مداد-کاغذی که در واقع هدف پژوهش نیز بررسی ویژگی‌های روان‌سنجی، اعتبار و روایی آن است متشکل از ۸۰ سؤال چهارگزینه‌ای است که به صورت صفر یا یک نمره‌گذاری می‌شوند. آزمون مذکور که توانایی و استعداد داوطلبان را در پنج درس قرآن و تعلیمات دینی (۱۵ سؤال)، زبان و ادبیات فارسی (۱۵ سؤال)، مطالعات اجتماعی (۱۰ سؤال)، علوم تجربی (۱۵ سؤال) و ریاضیات (۲۵ سؤال) اندازه می‌گیرد در روز جمعه مورخ ۹۲/۲/۲۰ رأس ساعت ۹ برگزار گردید. زمان در نظر گرفته شده برای پاسخ‌دهی به سؤال‌ها ۸۰ دقیقه بوده و به ازای هر ۳ پاسخ غلط یک نمره منفی لحاظ می‌شود. همچنین انتخاب بیش از یک پاسخ برای هر سؤال نیز، به عنوان پاسخ نادرست محسوب می‌شود.

### مقیاس تجدیدنظر شده هوش و کسلر برای کودکان

مقیاس هوش کودکان در سال ۱۹۷۲ مورد تجدیدنظر قرار گرفت. این مقیاس، هوش کودکان گروه‌های سنی ۶ سال تا ۱۶ سال و ۱۱ ماه و ۳۰ روز را اندازه‌گیری می‌کند. این مقیاس از ۱۲ خرده‌آزمون تشکیل شده است. مقیاس کلامی کودکان شامل خرده‌آزمون‌های اطلاعات عمومی، درک و فهم، محاسبات، شباهت‌ها، خزانه لغات و حافظه اعداد می‌باشد و مقیاس عملی کودکان شامل خرده‌آزمون‌های تکمیل تصاویر، تنظیم تصاویر، طراحی با مکعب‌ها، الحاق قطعات، رمزگردانی (که همتای رمزهای عددی بزرگسالان است) و مازها می‌باشد (شریفی، ۱۳۸۴). در مقیاس هوش و کسلر کودکان، خرده‌آزمون‌های حافظه عددی و مازها به عنوان آزمون‌های تکمیلی به حساب آمده است. اگر آزماینده وقت کافی

در اختیار داشته باشد می‌تواند آن‌ها را اضافه کند یا اگر به نتیجه یک آزمون بنا به عللی اعتماد نکند، می‌تواند آن‌ها را به‌عنوان یک آزمون جانشین مورد استفاده قرار دهد (گنجی، ۱۳۷۵).

این مقیاس توسط شهیم (۱۳۷۳) بر روی یک نمونه ۱۴۰۰ نفری از دختران و پسران ۶ تا ۱۳ سال شهر شیراز انطباق و هنجاریابی شده است. در پژوهش شهیم (۱۳۷۳) دامنه ضرایب اعتبار آزمون - باز آزمون خرده مقیاس‌ها از ۰/۵۶ تا ۰/۹۲ گزارش شده است. همچنین ضرایب اعتبار خرده مقیاس‌ها به شیوه دو نیمه‌سازی از ۰/۴۳ برای خرده مقیاس‌های اطلاعات و مکعب‌ها تا ۰/۹۸ برای واژگان گزارش شده است.

### پرسشنامه خلاقیت عابدی

پرسشنامه خلاقیت عابدی دارای ۶۰ سؤال ۳ گزینه‌ای است که توسط جمال عابدی و بر اساس نظریه خلاقیت تورنس در سال ۱۳۶۳ تهیه و روی ۶۵۰ دانش‌آموز در شهر تهران اجرا و هنجاریابی شده است (عابدی، ۱۳۷۲). این پرسشنامه از چهار خرده‌مقیاس یا مؤلفه سیالی<sup>۱</sup> (سؤال ۱ تا ۲۲)، بسط<sup>۲</sup> (سؤال ۲۳ تا ۳۳)، ابتکار<sup>۳</sup> (سؤال ۳۴ تا ۴۹) و انعطاف‌پذیری<sup>۴</sup> (سؤال ۵۰ تا ۶۰) تشکیل شده است. گزینه‌ها نشان‌دهنده میزان خلاقیت پایین، متوسط و بالا می‌باشند که نمره صفر برای خلاقیت پایین، نمره یک برای خلاقیت متوسط و نمره دو برای خلاقیت بالا در نظر گرفته شده است. مجموع نمرات کسب‌شده در هر خرده‌مقیاس، نمایانگر نمره آزمودنی در آن بخش است و مجموع نمرات آزمودنی در چهار خرده‌مقیاس، نمره کلی خلاقیت او را نشان می‌دهد. دامنه نمره کل خلاقیت بین صفر و ۱۲۰ خواهد بود (قرباغی، ۱۳۸۸).

در پژوهش عابدی (۱۳۷۲) ضرایب اعتبار خرده مقیاس‌ها سیالی، بسط، ابتکار و انعطاف‌پذیری به شیوه بازآزمایی به ترتیب ۰/۸۵، ۰/۸۰، ۰/۸۲ و ۰/۸۴ گزارش شده است. ضریب همسانی درونی خرده مقیاس‌ها با استفاده از آلفای کرونباخ برای سیالی، ۰/۷۵؛

1. Fluency
2. Elaboration
3. Originality
4. Flexibility

بسط، ۰/۶۱؛ ابتکار، ۰/۶۱ و انعطاف‌پذیری، ۰/۶۶ گزارش شده است (آزمند، ویلا و عابدی، ۱۹۹۶).

### شیوه تحلیل داده‌ها

برای تحلیل داده‌ها در چارچوب رویکرد راش و برآورد ویژگی‌های روان‌سنجی، مدل ساده لوجستیک راش (SLM؛ آندریش، ۱۹۸۸) و از نسخه ۳/۷۶ نرم‌افزار Winsteps (لی‌ناکر<sup>۱</sup>، ۲۰۱۲a) استفاده گردید. در این نرم‌افزار برای برآورد پارامترها از روش بیشینه درست‌نمایی توام<sup>۲</sup> (JML) استفاده می‌گردد.

در مدل راش، برای هر سؤال براساس یک پارامتر دشواری و همچنین شاخص‌های برازش سؤال برآورد گردید. پارامتر دشواری، جایگاه سؤال را در مقیاس توانایی توصیف می‌کند (بیکر، ۲۰۰۱). به زبان فنی و ریاضی می‌توان گفت که این پارامتر نقطه عطف تابع ویژگی سؤال را توصیف می‌کند و معمولاً در دامنه بین ۲/۵- تا ۲/۵+ مقیاس پردازی می‌شود. در این مدل برای هر فرد یک پارامتر توانایی برآورد گردید. این پارامتر معمولاً در دامنه بین ۴- تا ۴+ مقیاس پردازی می‌شود.

نرم‌افزار Winsteps برای بررسی برازش سؤال‌ها با مدل، دو نوع آماره MnSq و Zstd را ارائه می‌دهد. دامنه آماره MnSq از صفر 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 سؤالات 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، درحالی‌که برای نمونه‌های بزرگتر از ۳۰۰ نفر هیچ نوع داده‌ای با مدل برازش نمی‌یابد (لی‌ناکر، ۲۰۱۲)؛ بنابراین با توجه به حجم بالای نمونه مورد استفاده در این پژوهش و همچنین با در نظر گرفتن توصیه

1. Linacre  
2. Joint maximum likelihood

لی‌ناکر (۲۰۱۲b)، جهت بررسی برازش سؤال با مدل از آماره MnSq که به دو شکل outfit و infit برآورد و گزارش می‌شود استفاده گردید. لی‌ناکر (۲۰۱۲b) آماره MnSq outfit را بر آماره MnSq infit ترجیح می‌دهد. بنابراین در پژوهش حاضر برای بررسی برازش سؤال از آماره outfit MnSq استفاده گردید. دامنه قابل قبول برای این آماره از ۰/۷ تا ۱/۳ است (لیو<sup>۱</sup>، ۲۰۱۰).

در نرم‌افزار Winsteps (لی‌ناکر، ۲۰۱۲a) ضریب اعتبار شخص<sup>۲</sup> ( $R_p$ ) که خیلی شبیه به آلفای کرونباخ و KR-20 در CTT است و همچنین شاخص جداسازی شخص<sup>۳</sup> ( $G_p$ ) (شوماکر و اسمیت، ۲۰۰۷) برآورد و گزارش می‌شود. در واقع این آماره‌ها، توانایی آزمون را در جداسازی و پراکنده ساختن افراد در طول پیوستار صفت نشان می‌دهند. دامنه آماره  $R_p$  از صفر تا ۱ و دامنه آماره  $G_p$  از صفر تا بی‌نهایت است و لذا هر چه مقدار آن‌ها بالاتر باشد بهتر است.

برای مطالعه تناسب یا همخوانی بین سطح دشواری سؤال‌ها با سطح توانایی افراد از نمودار رایت (لی‌ناکر، ۲۰۱۲b) استفاده گردید. در نرم‌افزار Winsteps میانگین دشواری سؤال‌ها در مقیاس لوجیت برابر با صفر قرار داده می‌شود. 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 حالی که مقادیر منفی عکس این مطلب را نشان می‌دهد.

در مدل‌های خانواده راش، قدرت تشخیص تمام سؤال‌ها یکسان و برابر با ۱ فرض می‌شود. با این حال، برنامه Winsteps قدرت تشخیص سؤال‌ها را به‌عنوان نوعی آماره برازش، بر اساس داده‌های تجربی برآورد و گزارش می‌دهد. میزان انحراف پارامتر تشخیص یک سؤال از ۱ نشان‌دهنده میزان عدم برازش آن سؤال با مدل است. اگر پارامتر تشخیص سؤال بالاتر از ۱ باشد به آن معنا است که قدرت تمایزگذاری آن سؤال بین افراد قوی و

- 
1. Liu
  2. Person reliability coefficient
  3. Person separation index

ضعیف، فراتر از آن است که از سؤالی با این درجه دشواری انتظار می‌رود. برعکس، اگر ضریب تشخیص سؤالی کمتر از ۱ باشد نشان می‌دهد که توانایی آن سؤال در تمایزگذاری بین افراد قوی و ضعیف، کمتر از آن است که از سؤالی با این درجه دشواری انتظار می‌رود. علاوه بر روش‌ها و مدل‌های فوق، از آماره‌های توصیفی مانند میانگین و انحراف استاندارد برای توصیف عملکرد افراد گروه استفاده گردید. همچنین برای بررسی روایی آزمون ورودی مدارس استعدادهای درخشان و مدارس دولتی دوره راهنمایی، همبستگی نمرات کل آزمون و همبستگی نمرات دروس آن با نمرات مقیاس تجدیدنظر شده هوش و کسلر برای کودکان و با نمرات کل پرسشنامه خلاقیت عابدی و چهار خرده‌مقیاس آن از ضریب همبستگی گشتاوری پیرسون استفاده گردید.

## نتایج

جدول ۱ میانگین و انحراف استاندارد هر یک از دروس آزمون سمپاد را به تفکیک جنسیت و کل گروه نشان می‌دهد. در جدول مذکور، اعداد داخل پرانتز، میانگین و اعداد خارج پرانتز، انحراف استاندارد را نشان می‌دهد.

جدول ۱. میانگین و انحراف استاندارد دروس آزمون سمپاد به تفکیک جنسیت و کل گروه

کل گروه	جنسیت		دروس
	پسر	دختر	
۴/۲۶(۲/۳۰)	۴/۰۳(۲/۳۳)	۴/۵۳(۲/۲۳)	قرآن و تعلیمات دینی
۵/۱۶(۲/۵۱)	۴/۹۵(۲/۶۰)	۵/۴۲(۲/۳۸)	زبان و ادبیات فارسی
۳/۵۴(۱/۵۹)	۳/۶۰(۱/۶۱)	۳/۴۹(۱/۵۵)	مطالعات اجتماعی
۵/۳۹(۳/۴۰)	۵/۴۷(۲/۴۲)	۵/۳۰(۲/۳۸)	علوم تجربی
۶/۶۷(۳/۵۴)	۶/۸۲(۳/۷۵)	۶/۵۰(۳/۲۷)	ریاضیات
۲۵/۰۲(۸/۳۶)	۲۴/۸۵(۸/۸۲)	۲۵/۲۲(۷/۸۰)	کل آزمون

بررسی میانگین‌های دروس و کل آزمون، بیانگر این است که آزمون در مقایسه با سطح دانش و توانایی داوطلبان، بسیار دشوار است که البته با هدف آزمون که جداسازی دانش‌آموزان قوی برای ورود به مدارس سمپاد است همخوانی دارد.



ویژگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۱۹۹

سؤال یک: ضرایب اعتبار شخص و شاخص جداسازی شخص سؤال‌های آزمون ورودی مدارس استعداد‌های درخشان و نمونه دولتی دوره راهنمایی بر اساس مدل اندازه‌گیری راش چگونه است؟

جدول ۲ ضرایب اعتبار شخص<sup>۱</sup> و شاخص جداسازی شخص<sup>۲</sup> را برای سؤال‌های هر یک از دروس آزمون سمپاد به تفکیک جنسیت و کل گروه بر اساس مدل راش نشان می‌دهد.

جدول ۲. ضریب اعتبار شخص و شاخص جداسازی شخص برای دروس به تفکیک جنسیت و کل گروه

کل گروه	جنسیت						تعداد سؤال	دروس
	پسر		دختر					
	$G_p$	$R_p$	$G_p$	$R_p$	$G_p$	$R_p$		
۰/۸۱	۰/۴۰	۰/۸۰	۰/۳۹	۰/۸۰	۰/۳۹	۱۵	قرآن و تعلیمات	
۰/۹۴	۰/۴۷	۰/۹۸	۰/۴۹	۰/۸۶	۰/۴۲	۱۵	زبان و ادبیات	
۰/۳۶	۰/۱۲	۰/۳۹	۰/۱۳	۰/۳۳	۰/۱۰	۱۰	مطالعات اجتماعی	
۰/۸۷	۰/۴۳	۰/۹۰	۰/۴۵	۰/۸۵	۰/۴۲	۱۵	علوم تجربی	
۰/۸۱	۰/۴۰	۱/۳۲	۰/۶۳	۱/۱۱	۰/۵۵	۲۵	ریاضیات	
۱/۹۰	۰/۷۸	۲/۰۳	۰/۸۰	۱/۷۴	۰/۷۵	۸۰	کل آزمون	

سؤال دوم: آیا داده‌های حاصل از آزمون ورودی دوره راهنمایی مدارس استعداد‌های درخشان با رویکرد اندازه‌گیری راش برازش دارد؟

برآوردهای مربوط به دشواری سؤال‌ها همراه با خطای استاندارد مدل، پارامتر تشخیص و همچنین مقادیر آماره‌های  $infit$  MnSq و  $outfit$  MnSq در جدول ۳ ارائه شده است. اگرچه در این جدول، مقادیر آماره  $infit$  MnSq نیز ارائه شده است، لکن در ارزیابی برازش سؤال‌ها صرفاً از آماره  $outfit$  MnSq استفاده گردید. در برآورد این آماره‌ها سؤال‌های هر یک از دروس به صورت جداگانه با استفاده از نرم‌افزار Winsteps مورد تحلیل قرار گرفتند. دامنه مقادیر آماره  $outfit$  MnSq از ۰/۷۷ تا ۱/۲۳ است که حاکی از این است که همه سؤال‌ها به غیر از سؤال ۷۴ مربوط به درس ریاضی که همه داوطلبان به آن پاسخ درست داده‌اند با مدل راش برازش دارند.

1. Person reliability coefficient
2. Person separation index

جدول ۳. پارامتر دشواری، خطای استاندارد، پارامتر تشخیص و شاخص های برازش سؤال در مدل راش

پارامتر تشخیص	MnSq		خطای استاندارد	پارامتر دشواری	دروس / سؤال	پارامتر تشخیص	MnSq		خطای استاندارد	پارامتر دشواری	دروس / سؤال
	Infi t	Outfi t					Infi t	Outfi t			
علوم تجربی						قرآن و تعلیمات دینی					
۰/۸۹	۱/۰۳	۱/۰۴	۰/۰۴	-۰/۸۵	۴۱	۰/۹۴	۱/۰۴	۱/۰۹	۰/۰۵	۰/۲۸	۱
۱/۰۶	۰/۹۸	۰/۹۷	۰/۰۴	-۰/۱۲	۴۲	۱/۰۰	۰/۹۹	۱/۰۳	۰/۰۸	۱/۸۳	۲
۱/۰۰	۱/۰۰	۱/۰۱	۰/۰۵	۰/۸۶	۴۳	۰/۹۲	۱/۰۲	۱/۰۶	۰/۰۴	۰/۸۰	۳
۰/۹۴	۱/۰۴	۱/۲۳	۰/۰۶	۱/۳۳	۴۴	۰/۹۷	۱/۰۲	۱/۰۳	۰/۰۵	۰/۲۶	۴
۰/۸۴	۱/۰۴	۱/۰۵	۰/۰۴	-۰/۳۷	۴۵	۰/۹۹	۱/۰۱	۰/۹۹	۰/۰۵	۰/۳۷	۵
۱/۰۳	۰/۹۹	۰/۹۹	۰/۰۴	-۱/۰۱	۴۶	۱/۰۲	۰/۹۹	۰/۹۵	۰/۰۵	۰/۳۲	۶
۱/۴۰	۰/۹۰	۰/۸۷	۰/۰۴	-۰/۷۲	۴۷	۱/۰۲	۰/۹۹	۱/۰۱	۰/۰۴	۱/۴۶	۷
۱/۰۰	۰/۹۹	۱/۰۲	۰/۰۵	۱/۲۰	۴۸	۰/۹۸	۱/۰۱	۱/۰۰	۰/۰۴	۱/۶۱	۸
۱/۰۰	۱/۰۰	۰/۹۹	۰/۰۴	۰/۳۲	۴۹	۱/۱۹	۰/۹۵	۰/۹۳	۰/۰۴	۱/۴۳	۹
۰/۸۵	۱/۰۵	۱/۰۶	۰/۰۴	-۰/۱۴	۵۰	۰/۹۷	۱/۰۳	۱/۰۵	۰/۰۷	۱/۴۱	۱۰
۱/۱۷	۰/۹۴	۰/۹۲	۰/۰۴	۰/۰۲	۵۱	۱/۰۷	۰/۹۵	۰/۹۱	۰/۰۵	۰/۴۲	۱۱
۰/۹۳	۱/۰۷	۱/۱۰	۰/۰۵	۱/۱۴	۵۲	۰/۹۹	۱/۰۱	۱/۰۲	۰/۰۷	۱/۳۶	۱۲
۱/۰۴	۰/۹۸	۰/۹۶	۰/۰۴	-۱/۷۰	۵۳	۱/۱۴	۰/۹۵	۰/۹۳	۰/۰۴	۰/۶۰	۱۳
۱/۱۷	۰/۹۵	۰/۹۴	۰/۰۴	-۰/۳۷	۵۴	۰/۸۴	۱/۰۴	۱/۰۹	۰/۰۴	۱/۱۹	۱۴
۰/۹۲	۱/۰۴	۱/۰۶	۰/۰۴	۰/۴۲	۵۵	۱/۰۱	۰/۹۹	۰/۹۹	۰/۰۶	۰/۸۵	۱۵
ریاضیات						زبان و ادبیات فارسی					
۱/۱۸	۰/۹۵	۰/۹۴	۰/۰۴	۱/۰۴۴	۵۶	۰/۹۶	۱/۰۳	۱/۰۲	۰/۰۵	۰/۵۵	۱۶
۰/۸۳	۱/۰۴	۱/۰۵	۰/۰۴	۱/۰۳۹	۵۷	۰/۹۷	۱/۰۱	۱/۰۱	۰/۰۴	۰/۰۵	۱۷
۱/۰۲	۰/۹۹	۰/۹۹	۰/۰۴	-۰/۶۳	۵۸	۱/۰۸	۰/۹۶	۰/۹۸	۰/۰۴	۰/۱۰	۱۸
۱/۱۶	۰/۹۴	۰/۹۱	۰/۰۴	-۰/۴۹	۵۹	۰/۹۵	۱/۰۴	۱/۰۴	۰/۰۵	۰/۶۷	۱۹
۱/۱۱	۰/۹۷	۰/۹۶	۰/۰۴	-۰/۷۷	۶۰	۱/۰۴	۰/۹۴	۰/۹۶	۰/۰۷	۱/۷۱	۲۰
۰/۹۵	۱/۰۲	۱/۰۷	۰/۰۴	-۰/۰۹	۶۱	۱/۰۷	۰/۹۵	۰/۹۱	۰/۰۵	۰/۶۶	۲۱
۱/۰۰	۱/۰۰	۰/۹۸	۰/۰۶	-۰/۸۶	۶۲	۰/۸۷	۱/۰۳	۱/۰۵	۰/۰۴	۰/۸۶	۲۲
۱/۱۳	۰/۹۶	۰/۹۴	۰/۰۴	-۰/۶۹	۶۳	۰/۹۰	۱/۰۳	۱/۰۴	۰/۰۴	۱/۲۰	۲۳
۰/۹۱	۱/۰۳	۱/۰۴	۰/۰۴	-۰/۷۰	۶۴	۱/۱۴	۰/۹۶	۰/۹۳	۰/۰۴	۰/۲۵	۲۴
۰/۷۶	۱/۱۲	۱/۱۷	۰/۰۴	-۰/۰۹	۶۵	۰/۹۶	۱/۰۲	۱/۰۳	۰/۰۵	۰/۴۱	۲۵
۰/۹۹	۱/۰۱	۱/۰۰	۰/۰۵	۰/۵۹	۶۶	۱/۱۰	۰/۹۶	۰/۹۶	۰/۰۴	۱/۳۵	۲۶
۱/۰۲	۰/۹۹	۰/۹۹	۰/۰۴	-۰/۵۳	۶۷	۱/۰۹	۰/۹۵	۰/۹۲	۰/۰۴	۰/۲۹	۲۷
۱/۲۱	۰/۸۶	۰/۷۷	۰/۰۵	۰/۲۵	۶۸	۰/۹۷	۱/۰۱	۱/۰۳	۰/۰۴	۰/۳۰	۲۸
۱/۱۷	۰/۹۴	۰/۹۱	۰/۰۴	-۰/۵۶	۶۹	۱/۰۴	۰/۹۸	۱/۰۲	۰/۰۴	۱/۱۴	۲۹
۱/۰۲	۰/۹۹	۱/۰۰	۰/۰۴	-۰/۹۳	۷۰	۰/۸۶	۱/۰۷	۱/۱۵	۰/۰۵	۰/۳۶	۳۰
تعلیمات اجتماعی						تعلیمات اجتماعی					
۱/۰۳	۰/۹۶	۰/۹۶	۰/۰۵	۰/۷۴	۷۱	۰/۹۲	۱/۰۴	۱/۱۰	۰/۰۵	۰/۵۹	۳۱
۱/۰۲	۱/۰۰	۰/۹۸	۰/۰۴	-۰/۵۰	۷۲	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۴	۰/۹۲	۳۲

## ویژگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۲۰۱

-	-	-	۱/۸۳	۸/۳۵	۷۴	۱/۰۲	۰/۹۹	۰/۹۹	۰/۰۴	۰/۱۵	۳۳
۰/۹۲	۱/۱۰	۱/۱۶	۰/۰۷	۱/۳۰	۷۵	۱/۰۳	۰/۹۷	۰/۹۷	۰/۰۵	۰/۹۷	۳۴
۰/۹۶	۱/۰۵	۱/۰۵	۰/۰۶	۰/۹۶	۷۶	۱/۰۲	۰/۹۹	۰/۹۵	۰/۰۵	۰/۶۱	۳۵
۰/۹۳	۱/۰۴	۱/۰۴	۰/۰۵	-۰/۰۲	۷۷	۰/۹۷	۱/۰۲	۱/۰۷	۰/۰۵	۲/۴۶	۳۶
۰/۹۹	۱/۰۰	۱/۰۷	۰/۰۶	۱/۱۷	۷۸	۱/۰۰	۰/۹۹	۱/۰۳	۰/۰۶	۱/۷۱	۳۷
۰/۹۹	۱/۰۲	۱/۰۱	۰/۰۵	۰/۶۹	۷۹	۱/۱۱	۰/۹۷	۰/۹۶	۰/۰۴	۰/۸۵	۳۸
۰/۹۵	۱/۰۵	۱/۱۸	۰/۰۷	۱/۳۲	۸۰	۰/۹۳	۱/۰۴	۱/۰۷	۰/۰۵	۰/۵۷	۳۹
						۱/۰۸	۰/۹۷	۰/۹۸	۰/۰۴	۰/۳۷	۴۰

سؤال سوم: سطح دشواری و قدرت تشخیص سؤال‌های آزمون ورودی دوره راهنمایی

مدارس استعداد‌های درخشان بر اساس مدل اندازه‌گیری راش چگونه است؟

درجه دشواری و پارامتر تشخیص سؤال‌های آزمون سمپاد بر اساس مدل راش به تفکیک جنسیت و کل گروه در جدول ۳ ارائه شده است. دامنه دشواری سؤال‌ها از ۲/۴۶- تا ۱/۸۳، است. قدرت تشخیص سؤال‌ها به‌غیر از دو سؤال ۷۴ و ۶۵ مطابق با انتظارات مدل راش است که شاهد دیگری بر برازش سؤال‌ها با مدل راش است.

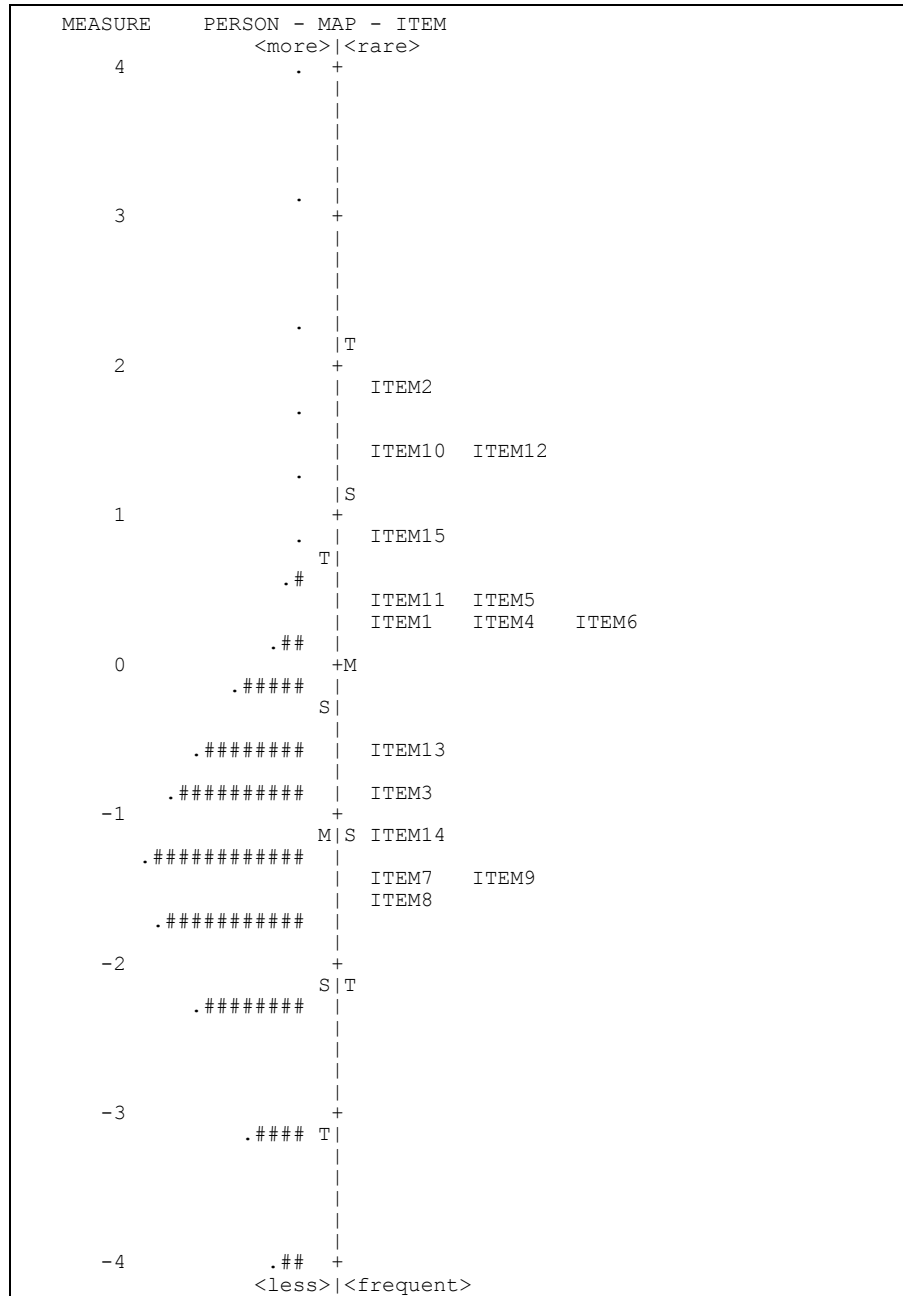
سؤال چهارم: آیا سطح دشواری سؤال‌های آزمون ورودی دوره راهنمایی مدارس

استعداد‌های درخشان سطح توانایی داوطلبان تناسب و همخوانی دارد؟

در مدل اندازه‌گیری راش از آنجا که پارامتر توانایی افراد و دشواری سؤال‌ها در یک مقیاس مشترک قرار دارند؛ لذا می‌توان میانگین و انحراف استاندارد این دو پارامتر به‌صورت مستقیم با یکدیگر مقایسه نمود. در ادامه، تناسب و همخوانی سؤال‌های هر یک از دروس به‌صورت جداگانه مورد بررسی قرار می‌گیرد.

**قرآن و تعلیمات دینی:** دامنه پارامتر دشواری سؤال‌های این درس از ۱/۶۱- تا ۱/۸۳ با

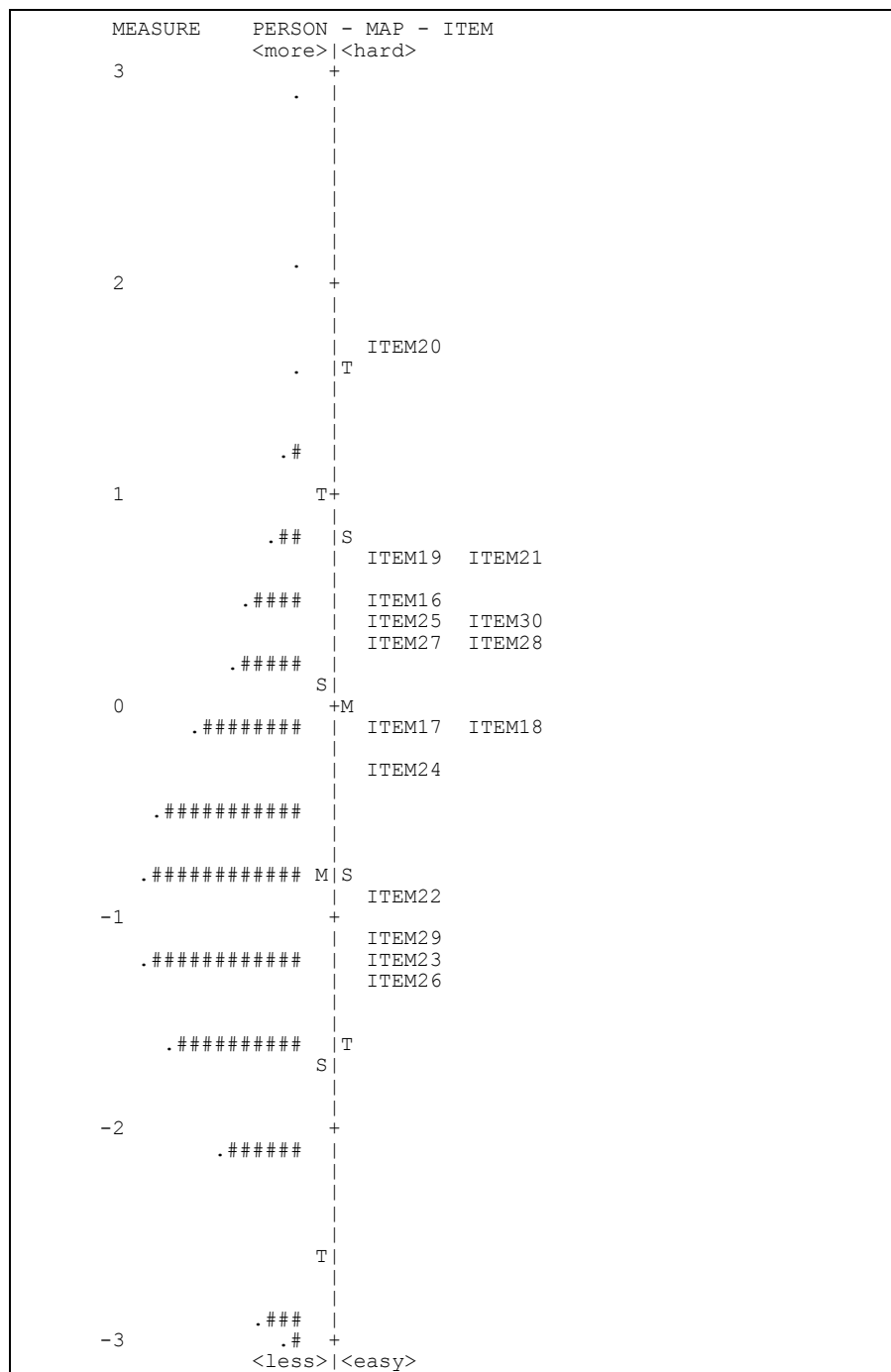
میانگین صفر و انحراف استاندارد ۱/۰۸ و دامنه پارامتر توانایی افراد از ۴/۴۰- تا ۴/۴۰ با میانگین ۱/۳۳- و انحراف استاندارد ۱/۱۲ به دست آمد. این یافته بیانگر آن است که افراد در مقایسه با سؤال‌ها از سطح توانایی پایین‌تر برخوردارند. این یافته به‌صورت تصویری در شکل ۱ که به نمودار رایت نیز معروف است نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که سطح دشواری ۹ سؤال از ۱۵ سؤال، بیشتر از سطح توانایی اکثر افراد است و سطح توانایی اکثریت افراد بسیار پایین‌تر از سطح دشواری سؤال‌ها است.



ویژگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۲۰۳

**زبان و ادبیات فارسی:** دامنه پارامتر دشواری سؤال‌های این درس از ۱/۳۵- تا ۱/۷۱ با میانگین صفر و انحراف استاندارد ۰/۸۱ و دامنه پارامتر توانایی افراد از ۴/۲۰- تا ۲/۹۱ با میانگین ۰/۸۵- و انحراف استاندارد ۱/۰۱ به دست آمد. این یافته نشان می‌دهد که افراد در مقایسه با سؤال‌ها از سطح توانایی پایین‌تر برخوردارند. این یافته به صورت تصویری در شکل ۲ نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که سطح دشواری ۹ سؤال از ۱۵ سؤال، بالاتر از متوسط سطح توانایی افراد است و سطح توانایی اکثریت افراد بسیار پایین‌تر از سطح دشواری سؤال‌ها است.

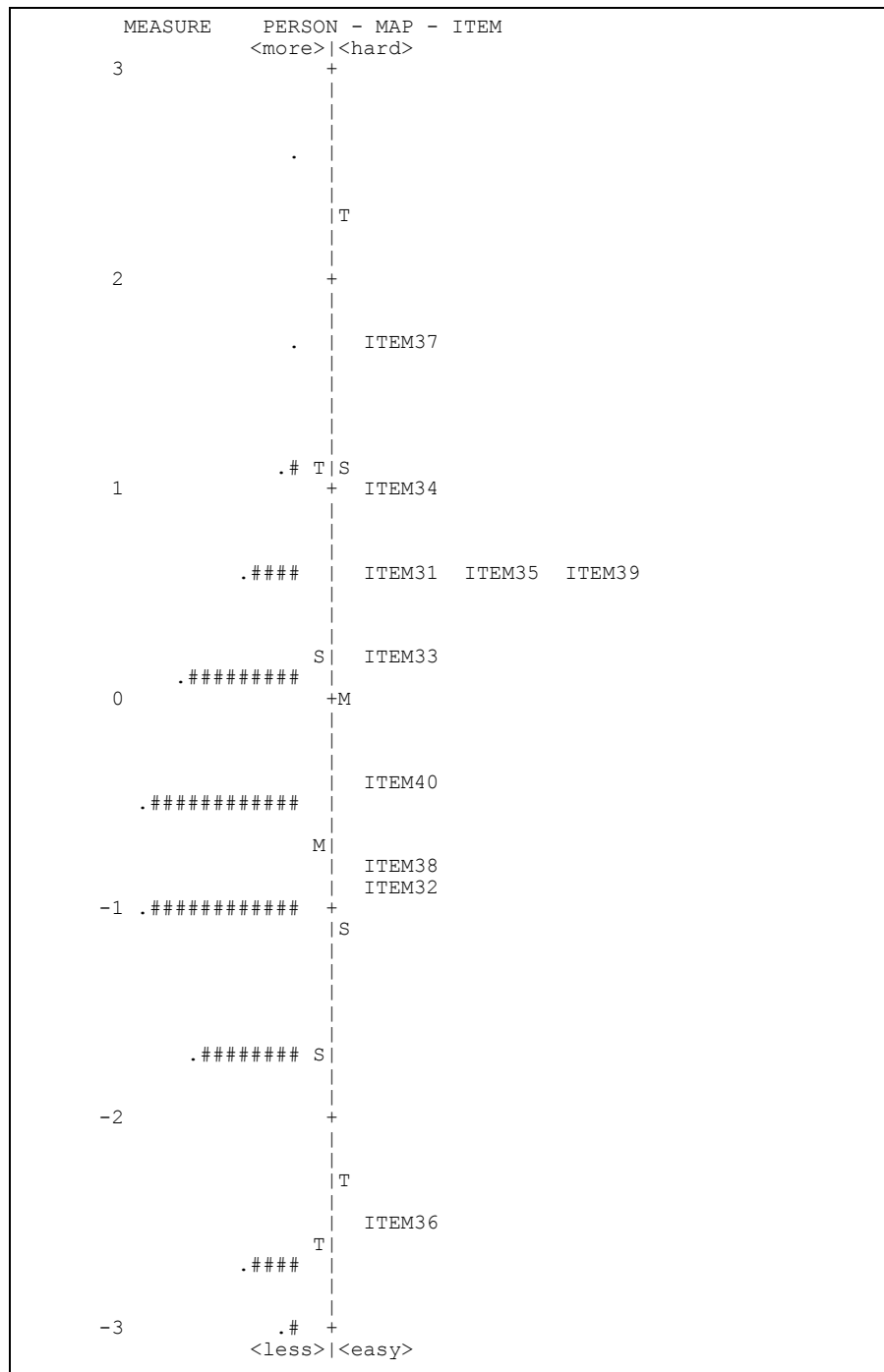
**تعلیمات اجتماعی:** دامنه پارامتر دشواری سؤال‌های این درس از ۲/۴۶- تا ۱/۷۱ با میانگین صفر و انحراف استاندارد ۱/۱۳ و دامنه پارامتر توانایی افراد از ۴/۱۵- تا ۲/۶۰ با میانگین ۰/۸۳- و انحراف استاندارد ۱/۰۸ به دست آمد. این یافته نشان می‌دهد که افراد در مقایسه با سؤال‌ها از سطح توانایی پایین‌تر برخوردارند. این یافته به صورت تصویری در شکل ۳ نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که سطح دشواری ۶ سؤال از ۱۰ سؤال این درس، بالاتر از متوسط سطح توانایی افراد است و سطح توانایی اکثریت افراد بسیار پایین‌تر از متوسط سطح دشواری سؤال‌ها است.



شکل ۲. نمودار دشواری سؤال‌ها و توانایی افراد در درس زبان و ادبیات فارسی (نمودار رایت)

هر "#" بیانگر ۳۶ نفر و هر "۰" بیانگر ۱ تا ۳۵ نفر است.

ویزگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۲۰۵



شکل ۳. نمودار دشواری سؤال‌ها و توانایی افراد در درس تعلیمات اجتماعی (نمودار رایت)

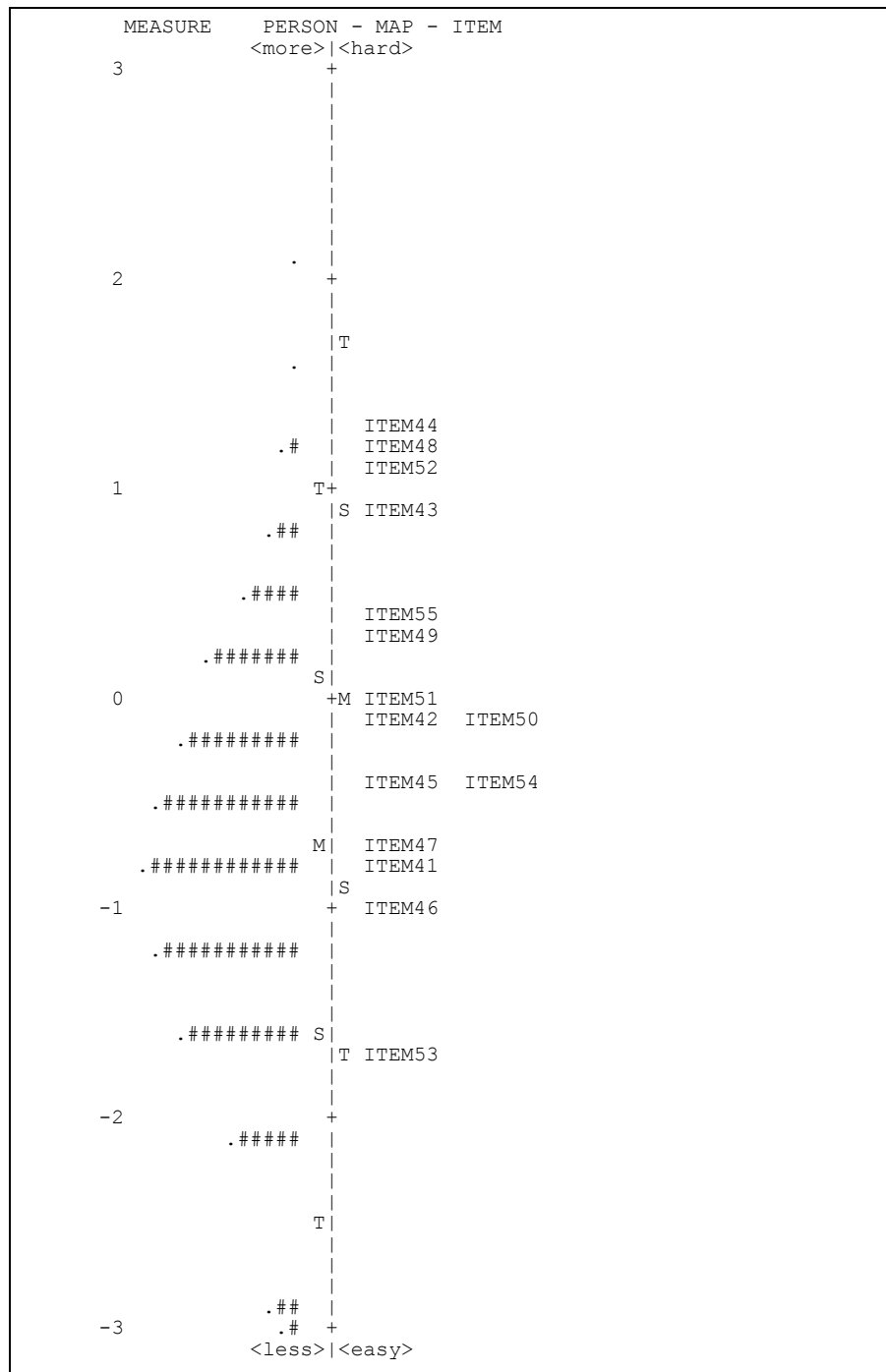
هر "#" بیانگر ۵۳ نفر و هر "۰" بیانگر ۱ تا ۵۲ نفر است.

**علوم تجربی:** دامنه پارامتر دشواری سؤال‌های این درس از  $۱/۷۰$  تا  $۱/۳۳$  با میانگین صفر و انحراف استاندارد  $۰/۸۶$  و دامنه پارامتر توانایی افراد از  $۴/۲۳$  تا  $۲/۱۲$  با میانگین  $۰/۷۷$  - و انحراف استاندارد  $۰/۹۶$  به دست آمد. این یافته نشان می‌دهد که افراد در مقایسه با سؤال‌ها از سطح توانایی پایین‌تر برخوردارند. این مطلب به صورت تصویری در شکل ۴ نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که سطح دشواری ۱۲ سؤال از ۱۵ سؤال این درس، بالاتر از متوسط سطح توانایی افراد است و سطح توانایی اکثریت افراد بسیار پایین‌تر از متوسط سطح دشواری سؤال‌ها است.

**ریاضیات:** دامنه پارامتر دشواری سؤال‌های این درس، به‌غیر از سؤال شماره ۷۴ که به دلیل اینکه هیچ کدام از داوطلبان به آن پاسخ درست نداده‌اند با مدل برازش ندارد، از  $۱/۳۹$  - تا  $۱/۳۲$  با میانگین صفر و انحراف استاندارد  $۰/۱۷$  و دامنه پارامتر توانایی افراد از  $۴/۶۸$  - تا  $۲/۲۰$  با میانگین  $۱/۲۶$  - و انحراف استاندارد  $۱/۰۰$  به دست آمد. این یافته نشان می‌دهد که افراد در مقایسه با سؤال‌ها از سطح توانایی بسیار پایین‌تری برخوردارند. این موضوع به صورت تصویری در شکل ۵ نشان داده شده است. این نمودار حاکی از عدم تناسب و همخوانی بین سطح دشواری سؤال‌ها و سطح توانایی افراد است. این نمودار نشان می‌دهد که سطح دشواری ۲۲ سؤال از ۲۵ سؤال این درس، بالاتر از متوسط سطح توانایی افراد است و سطح توانایی اکثریت افراد بسیار پایین‌تر از متوسط سطح دشواری سؤال‌ها است.

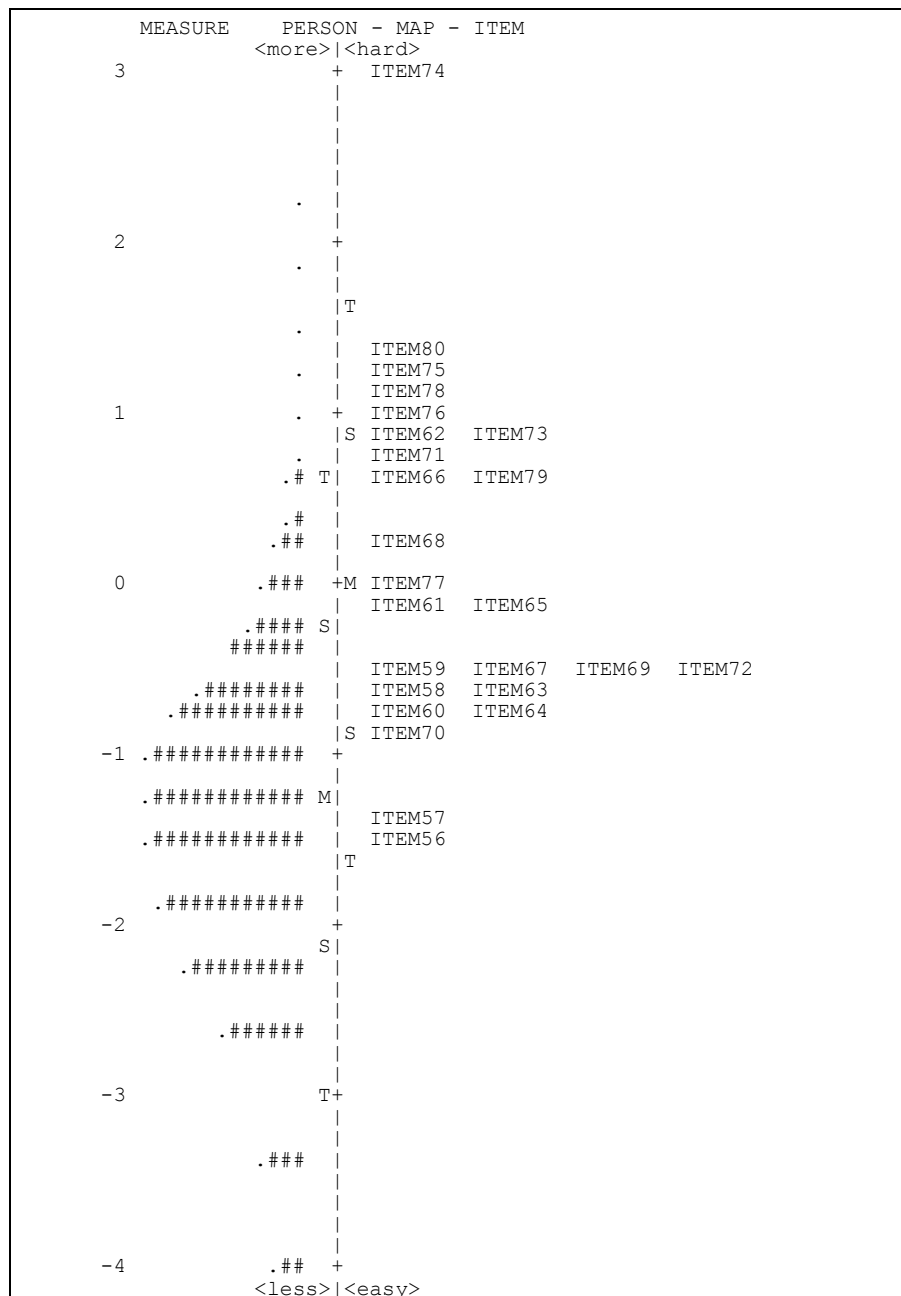


ویزگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۲۰۷



شکل ۴. نمودار دشواری سؤال‌ها و توانایی افراد در درس علوم تجربی (نمودار رایت)

هر "#" بیانگر ۳۶ نفر و هر "۰" بیانگر ۱ تا ۳۵ نفر است.



شکل ۵. نمودار دشواری سؤال‌ها و توانایی افراد در درس ریاضیات (نمودار رایت)

هر "#" بیانگر ۲۶ نفر و هر "۰" بیانگر ۱ تا ۲۵ نفر است.

ویژگی‌های روان‌سنجی (اعتبار و روایی) آزمون‌های ورودی ... / ۲۰۹

سؤال پنجم: روایی آزمون ورودی مدارس استعداد‌های درخشان و نمونه دولتی ورودی راهنمایی در ارتباط با مقیاس تجدیدنظر شده هوش و کسلر برای کودکان و پرسشنامه خلاقیت عابدی چگونه است؟

ضرایب همبستگی آزمون سمپاد با مقیاس هوش و کسلر و خرده مقیاس‌های آن در جدول ارائه شده است. ضرایب همبستگی که در سطح آلفای ۰/۰۵ به لحاظ آماری معنادار هستند با علامت \* مشخص شده است.

جدول ۴. همبستگی‌های نمرات دروس و نمره کل آزمون سمپاد با خرده مقیاس‌ها و هوشبهر

کلامی، عملی و کلی آزمون و کسلر (n=۱۲۲)

نمره کل	دروس					خرده مقیاس‌ها
	ریاضیات	علوم تجربی	مطالعات اجتماعی	زبان و ادبیات	قرآن	
۰/۰۷	۰/۰۲	۰/۰۲	۰/۰۲	۰/۰۹	۰/۰۷	اطلاعات
-۰/۱۶	-۰/۱۰	-۰/۰۲	-۰/۱۰	۰/۰۰۷	*-۰/۲۴	شباهت‌ها
۰/۱۸	*۰/۲۳	۰/۰۲	۰/۰۸	۰/۰۴	۰/۰۶	حساب
-۰/۰۴	-۰/۰۵	-۰/۰۷	۰/۰۳	۰/۰۶	-۰/۰۷	واژه‌ها
-۰/۰۸	-۰/۰۲	-۰/۰۲	-۰/۰۴	-۰/۰۶	-۰/۱۱	ادراک
*۰/۲۹	*۰/۲۵	۰/۱۷	۰/۲۰	۰/۱۹	-۰/۰۲	حافظه ارقام
۰/۱۰	۰/۱۵	۰/۱۵	-۰/۰۰۹	۰/۰۳	-۰/۰۸	تکمیل تصاویر
*۰/۲۰	۰/۱۸	۰/۰۵	۰/۱۵	۰/۱۰	۰/۰۷	ترتیب تصاویر
*۰/۲۹	*۰/۲۹	۰/۱۰	۰/۰۳	۰/۲۰	۰/۱۱	مکعب‌ها
۰/۱۱	۰/۱۱	۰/۰۴	۰/۰۲	۰/۲۰	-۰/۰۷	تنظیم قطعات
۰/۱۴	۰/۰۸	-۰/۰۷	۰/۱۱	۰/۱۷	۰/۱۲	تطبیق علائم
۰/۱۱	۰/۱۷	۰/۰۷	۰/۰۵	-۰/۰۳	-۰/۰۰۴	مازها
۰/۱۴	*۰/۲۴	-۰/۰۲	۰/۰۸	۰/۲۱	-۰/۱۷	هوشبهر کلامی
*۰/۳۲	*۰/۳۱	۰/۱۱	۰/۱۲	۰/۲۵	۰/۰۶	هوشبهر عملی
*۰/۳۴	*۰/۴۱	۰/۱۳	۰/۱۰	۰/۲۴	-۰/۰۵	هوشبهر کلی

جدول ۵ ضرایب همبستگی‌های آزمون سمپاد را با پرسشنامه خلاقیت عابدی و مؤلفه‌های آن نشان می‌دهد. نتایج نشان داد که هیچ کدام از ضرایب همبستگی آزمون سمپاد با پرسشنامه خلاقیت عابدی و مؤلفه‌های آن به لحاظ آماری معنادار نیست.

جدول ۵. همبستگی‌های خرده مقیاس‌ها و نمره کل پرسشنامه خلاقیت با نمرات دروس و نمره کل آزمون

سمپاد (n=۱۲۲)

نمره کل	دروس آزمون سمپاد					خرده مقیاس‌ها
	ریاضیات	علوم تجربی	مطالعات اجتماعی	زبان و ادبیات	قرآن	
-۰/۱۰	-۰/۰۳	-۰/۱۲	-۰/۲۰	-۰/۰۰۹	۰/۰۳	سیالی
۰/۰۳	۰/۰۵	۰/۰۰۵	-۰/۰۱	۰/۰۳	-۰/۰۱	بسط
۰/۰۶	۰/۰۲	۰/۰۳	-۰/۰۰۳	۰/۰۸	۰/۰۵	ابتکار
-۰/۰۰۵	۰/۰۳	-۰/۰۴	-۰/۰۱۴	-۰/۰۱	۰/۰۹	انعطاف‌پذیری
-۰/۰۵	۰/۰۲	-۰/۰۵	-۰/۱۳	۰/۰۳	۰/۰۵	نمره کل

### بحث و نتیجه‌گیری

بر اساس یافته‌های پژوهش حاضر، به‌طور کلی می‌توان گفت که آزمون سمپاد از ویژگی‌های اندازه‌گیری لازم برخوردار نیست و کیفیت لازم برای استفاده به‌عنوان ابزاری برای گزینش دانش‌آموزان برای ورود به مدارس استعدادهای درخشان و نمونه دولتی را ندارد. آزمون در مقایسه با سطح دانش و توانایی افراد بسیار دشوار است. همین مسئله داوطلبان را بی‌انگیزه نموده و باعث می‌شود تا برای دادن پاسخ درست به سؤال‌ها دست به حدس بزنند که منجر به افزایش خطا و کاهش دقت اندازه‌گیری می‌شود. آزمون سمپاد بر اساس آزمون‌های هوش و خلاقیت فاقد روایی کافی است. در واقع، اکثریت ضرایب همبستگی‌های نمره کل و نمره دروس آزمون سمپاد با آزمون هوش و کسلر و خلاقیت بسیار ناچیز و حتی منفی است؛ بنابراین لازم است که دست‌اندرکاران سازمان سمپاد در طراحی این آزمون‌ها دقت بیشتری نمایند تا به خاطر کیفیت ضعیف آزمون‌ها سود و زیان خاصی متوجه داوطلبان نشود و دانش‌آموزانی که واقعاً از توانایی‌ها و شایستگی‌های بالا برخوردارند وارد نظام آموزشی سمپاد گردند. برای این منظور لازم است تا با تشکیل بانک سؤال، در تهیه این آزمون‌ها از سؤال‌هایی استفاده گردد که از کیفیت لازم برخوردارند.

### منابع

- شریفی، حسن پاشا (۱۳۸۴). *اصول روان‌سنجی و روان‌آزمایی*. تهران: رشد.
- شهیم، سیما (۱۳۷۲). *مقیاس تجدیدنظر شده هوشی و کسلر برای کودکان: انطباق و هنجاریابی*. شیراز: انتشارات دانشگاه شیراز.

عابدی، جمال (۱۳۷۲). خلاقیت و شیوه‌ای نو در اندازه‌گیری آن. پژوهش‌های روان‌شناختی، ۳، ۴۶-۵۴.

قرباغی، حسن (۱۳۸۸). بررسی رابطه بین خلاقیت با خودکارآمدی رایانه‌ای در دانشجویان کارشناسی رشته تکنولوژی آموزشی دانشکده روانشناسی و علوم تربیتی دانشگاه علامه طباطبائی. پایان‌نامه کارشناسی ارشد. دانشگاه علامه طباطبائی.

کاظمی حقیقی، ناصرالدین (۱۳۸۲). ابهامات تشخیص و نخبگی تحصیلی. استعداد‌های درخشان، ۴۶، ۱۴۸-۱۶۱.

گنجی، حمزه (۱۳۷۵). آزمون‌های روانی. مشهد: دانشگاه امام رضا.

Andrich, D. (1988). *Rasch Models for Measurement*. Beverly Hills: Sage Publications.

Auzmendi, E. Villa, A. & Abedi, J. (1996). Reliability and validity of a newly-constructed multiple-choice creativity instrument. *Creativity Research Journal*, 9 (1), 89-95.

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation.

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. CENGAGE learning

Hagquist, C. (2001). Evaluating composite measures using Rasch modeling: an illustrative example. *Social and Preventive Medicine* 46, 369-378.

Linacre, J. M. (2012a). *Winsteps[Rasch measurement computer program, version 3.76]*. Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2012b). *A User's Guide to Winsteps: Rasch – Model Computer Program*. Beaverton, Oregon: Winsteps.com.

Liu, X. (2010). Using and developing measurement instruments in science education: A Rasch modeling approach. North Carolina: IAP, Inc.

Rasch, G. (1980). Probabilistic Models for Some Intelligence and Attainment Tests (First published 1960 by the Danish Institute for Educational Research). Chicago: MESA Press.

Schumacker, R. E. & Smith, E. V. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 3, 394-409.