

ویژگی‌های روان‌سنجی درس ریاضی آزمون سراسری گروه ریاضی - فیزیک سال ۱۳۹۱

Psychometrics properties of math questions of university entrance examination 2012/2013 academic year of mathematics and physics participants in Iranمرتضی حسینی^۱، علی مقدم زاده^۲، مصطفی رستگارآگاه^۳**Abstract:**

The purpose of this study was to examine psychometric properties of mathematics items of university entrance examination 2012/2013 academic year of mathematics and physics students in Iran. The method of this study was descriptive and applied. To do so, 1000 math university participants were randomly selected out of all of participants. For data analysis, SPSS-19 and Bilog-mg was used. All items analyzed based on CTT and IRT. Then Item difficulty and item parameters were calculated in terms of both IRT and CTT. CTT results showed that math examination was so hard with good reliability and validity. As well, some items had low discrimination index. Three of items had negative discrimination index that had major problems based on CTT. The assumptions and the goodness of fit analysis were checked for IRT. The results showed that this is a Unidimensional test with good Local independence. The items that had fit with none of models are: 1-2-3-4-6-12-14-15-17-19-21-26-29-38-40-41-43-44-45-51-54. Also, the most appropriate model which fit the data was achieved by 2pl logistic IRT model. IR analysis based on 2pl showed that none of items had negative discrimination index. Also, 17 items had not good discrimination index. Item difficulty of items in IRT such as CTT showed that math questions were so hard. This results were showed test information curve.

Key words: High-stakes tests, CTT, IRT**چکیده**

هدف این پژوهش توصیفی و کاربردی؛ تعیین ویژگی‌های روان‌سنجی درس ریاضی آزمون سراسری گروه ریاضی - فیزیک سال ۹۱ بر اساس هر دو نظریه کلاسیک و سوال - پاسخ بود. به همین منظور از بین کلیه شرکت‌کنندگان این گروه، یک نمونه ۱۰۰۰ نفری به طور تصادفی انتخاب و برای تجزیه و تحلیل داده‌ها نیز از نرم‌افزارهای SPSS و Bilog استفاده گردید. نتایج تحلیل‌های کلاسیک نشان داد؛ این آزمون دارای روایی و اعتبار خوبی است. بیشترین و کمترین ضریب تشخیص به ترتیب به سوال‌های ۳۷ و ۵۱ تعلق داشت. ۳ سوال دارای ضریب تشخیص منفی و اشکال اساسی بودند. بررسی مفروضه‌های نظریه سوال - پاسخ نشان داد که این آزمون تک‌بعدی بوده و از استقلال موضعی نیز برخوردار است. ۲۱ سوال با هیچ‌کدام از مدل‌های سوال - پاسخ برازش نداشتند. مجموعه داده‌ها با مدل دو پارامتری برازش داشت. نتایج تحلیل‌های سوال - پاسخ (بر خلاف کلاسیک) نشان داد، ضریب تشخیص هیچ سوالی منفی نیست. ۱۷ سوال نیز دارای ضریب تشخیص نامناسب و سوالات ۱۷ و ۴۲ به ترتیب دارای بیشترین و کمترین ضریب تشخیص بودند. باتوجه به منحنی آگاهی و سطوح دشواری سوالات در تحلیل‌های سوال - پاسخ (همانند نتایج کلاسیک)، سوالات این درس مربوط به یک آزمون دشوار بوده است. در نهایت، این آزمون به دلیل برازش با مدل ۲ پارامتری دارای مزیت، و بعلا داشتن آیت‌های دشوار و ضریب تشخیص منفی در نظریه کلاسیک دارای اشکال است. ارجحیت نظریه سوال - پاسخ نسبت به کلاسیک نیز در این تحقیق به اثبات رسید.

واژگان کلیدی: آزمون‌های سرنوشت‌ساز، اندازه‌گیری، نظریه کلاسیک، نظریه

سوال - پاسخ

^۱کاندیدای دکتری روان‌سنجی (سنجش و اندازه‌گیری)، دانشگاه علامه طباطبایی، گروه سنجش و اندازه‌گیری

(۰۹۱۲۸۸۹۷۲۵۰@sanjesh.hm@gmail.com)

^۲دکتری سنجش و اندازه‌گیری، عضو هیات علمی دانشگاه فارابی تهران.^۳دکتری روان‌سنجی (سنجش و اندازه‌گیری)، دانشگاه علامه طباطبایی، گروه سنجش و اندازه‌گیری.

مقدمه

آزمون‌های سرنوشت‌ساز (خطیر)^۱، آزمون‌هایی هستند که در مقیاس وسیع اجرا شده و بر اساس نتایج آنها تصمیماتی خطیر در امور افراد اتخاذ می‌شود. چنین تصمیماتی دارای تبعات شخصی^۲، اجتماعی^۳ و سیاسی^۴ هستند؛ بنابراین، اعتبار تصمیم‌گیری و کارایی مدل‌های سنجش در این مورد از اهمیت فوق‌العاده‌ای برخوردار بود و از سالیان گذشته جزء مهم‌ترین چالش‌های روان‌سنجی^۵ باقی‌مانده‌اند (پاپهام^۶، ۲۰۰۵، به نقل از گرامی پور و همکاران، ۱۳۹۱).

ابزاری که ضعیف ساخته شده باشد نه تنها مفید نیست، بلکه خطرناک نیز است؛ بنابراین آزمون‌های سرنوشت‌ساز باید به‌دقت ساخته، اجرا و نمره‌گذاری شوند. علاوه بر این حتی ابزاری که به‌خوبی ساخته شده باشد، می‌تواند به طور مؤثر یا نامؤثر به کار گرفته شود و یا حتی مورد سوءاستفاده قرار گیرد. اعتقاد بر این است که برخی از آزمون‌های سرنوشت‌ساز بهتر از بقیه ساخته می‌شوند و برخی از آن‌ها به طور مؤثرتری از بقیه مورد استفاده قرار می‌گیرند (نیکولز، گلاس و برلینر^۷، ۲۰۰۵). هدف اصلی توسعه^۸ و ساخت تست، ساختن یک آزمون کیفیت بالا^۹ توسط انتخاب سؤالات مناسب است. دو نظریه اصلی، نظریه کلاسیک آزمون^{۱۰} و نظریه سؤال – پاسخ^{۱۱} برای ساخت آزمون می‌تواند مورد استفاده قرار بگیرد (نسه گلر، گلدن کایا یویانیک، گولدن تاسدلن تیکر^{۱۲}، ۲۰۱۴). نظریه کلاسیک در طول بیشتر سال‌های قرن بیستم (دمیرتاش^{۱۳}، ۲۰۰۲؛ تراوب^{۱۴}، ۱۹۹۷) و حتی هم‌اکنون، برای ساخت آزمون مورد استفاده قرار گرفته است (بچگار^{۱۵} و همکاران، ۲۰۰۳). مزیت اصلی آن مفروضه‌های نظری ضعیف است که کاربرد آن را در شرایط آزمون‌گیری آسان کرده است (همبلتون و جونز^{۱۶}، ۱۹۹۳). اگر چه تمرکز اصلی نظریه کلاسیک در آگاهی‌دهندگی در سطح آزمون^{۱۷} است؛ اما آماره‌های سؤال^{۱۸} (مانند سطح دشواری و ضریب تشخیص) نیز یکی از مهم‌ترین بخش‌های این مدل است (فن^{۱۹}، ۱۹۹۸). نظریه کلاسیک برای سال‌های متمادی در ساخت آزمون خدمات خوبی ارائه داد؛ اما دارای برخی محدودیت‌های کاربردی است. محدودیت‌های اصلی آن نشئت‌گرفته از وابستگی به گروه نمونه^{۲۰} و آماره‌های آزمون است (همبلتون و سوامیناتان^{۲۱}، ۱۹۸۵). به بیانی

-
- 1- High stakes tests
 - 2- Personal
 - 3- Social
 - 4- Political
 - 5- Psychometric
 - 6- Popham
 - 7- Nichols , Glass , Berliner
 - 8 -Development
 - 9- Desired quality
 - 10- Classical test theory
 - 11- Item response theory
 - 12- Neşe Güler, Gllden Kaya Uyanic, Glşen Taşdelen Teker
 - 13- Demirtash
 - 14- Traub
 - 15- Bechger
 - 16- Hambleton & Jones
 - 17- Test-level information
 - 18- Item statistics
 - 19- Fan
 - 20- Group dependent
 - 21- Hambleton & Swaminathan

دیگر، در این نظریه، آماره آزمودنی (برای مثال نمره مشاهده شده) وابسته به سؤال است و آماره‌های سؤال مثل سطح دشواری و ضریب تشخیص، وابسته به آزمودنی^۱ است و این برخی دشواری‌های نظری^۲ در کاربرد نظریه کلاسیک در شرایط اندازه‌گیری ایجاد می‌کند؛ مانند هم‌ترازسازی آزمون^۳، سنجش انطباقی رایانه‌ای^۴، تشخیص سوگیری سؤالات، پیوند و ساخت بانک‌های سؤال (دمیرتاش، ۲۰۰۲؛ فن، ۱۹۹۸).

از طرفی یک نظریه آزمون باید ویژگی‌های مطلوب زیر را داشته باشد: الف) ویژگی‌های سؤالی را به دست دهد که وابسته به گروه نباشد. ب) نمراتی را ایجاد نماید که مهارت آزمودنی‌ها را توصیف نماید و وابسته به آزمون نباشد. ج) مدلی باشد که در سطح سؤال بیان گردد و نه در سطح آزمون (د) مدلی ارائه نماید که نیازمند آزمون‌های کاملاً موازی برای برآورد اعتبار نباشد و ح) برای هر نمره توانایی شاخصی از دقت را فراهم سازد. نشان داده شده است که چنین ویژگی‌هایی می‌تواند در چارچوب یک نظریه آزمون که به نظریه سؤال - پاسخ^۵ معروف است، حاصل شود (همبلتون، ۱۹۸۳، لرد، ۱۹۸۰، رایت و استون^۶، ۱۹۷۹، به نقل از سوامیناتان و همکاران، ۱۳۸۹، ص ۷). در نظریه سؤال-پاسخ باور بر این است که اگر چگونگی ارتباط هر سؤال آزمون با فرد مورد مطالعه درک شود، می‌توان نمره واقعی یا صفت واقعی مورد نظر را به صورت مستقیم برآورد کرد (برنان، ۲۰۰۶، نانالی و برنشتاین؛ ۱۹۹۴؛ کلاین، ۱۹۹۸؛ وندرلیندن و همبلتون، ۱۹۹۷؛ لرد، ۱۹۸۰، به نقل از یونسی، ۱۳۹۱، ص ۹). بر خلاف مدل کلاسیک نمره واقعی، مدل‌های نظریه سؤال-پاسخ مدل‌هایی ابطال‌پذیرند^۷. همچنین، سنجش برازش مدل با داده‌ها در تمام کاربردهای نظریه سؤال-پاسخ ضروری است. وقتی یک مدل فرضی نظریه سؤال-پاسخ با داده‌های آزمون مورد نظر برازش دارد، چندین ویژگی مطلوب حاصل می‌شود. توانایی برآورد شده آزمودنی‌ها وابسته به آزمون نیست و ویژگی‌های سؤال وابسته به گروه نمی‌باشد. توانایی‌های برآورد شده‌ای که از مجموعه‌های متفاوت سؤال به دست می‌آیند، مشابه خواهند بود (مگر به دلیل خطاهای اندازه‌گیری) و ویژگی‌های سؤالی از گروه‌های مختلف آزمودنی به دست می‌آیند یکسان‌اند (مگر به دلیل خطاهای اندازه‌گیری). گفته می‌شود که در نظریه سؤال - پاسخ پارامترهای سؤال و توانایی تغییرناپذیرند^۸. ویژگی تغییرناپذیری پارامترهای سؤال و توانایی از طریق دخالت دادن اطلاعات سؤال در فرایند برآورد توانایی و دخالت دادن اطلاعات توانایی آزمودنی‌ها در فرایند برآورد پارامترهای سؤال، حاصل می‌گردد. آزمودنی‌هایی که توانایی یکسانی دارند از احتمال مشابهی در ارائه پاسخ صحیح برخوردارند. به دلیل آنکه احتمال موفقیت آزمودنی با توانایی مفروض، به وسیله پارامترهای سؤال تعیین می‌شود، پارامترهای سؤال نیز باید برای دو گروه یکسان باشد. علاوه بر این نظریه سؤال - پاسخ در هر سطح توانایی برآوردی از خطای استاندارد فراهم می‌کند. این در حالی است که در نظریه کلاسیک آزمون تنها یک برآورد خطا برای تمامی آزمودنی‌ها (و در تمامی سطوح توانایی) ارائه می‌گردد (همبلتون و همکاران، ۱۳۸۹، ص ۱۲ و ۱۳). مشهورترین سه مدل تک‌بعدی سؤال - پاسخ، مدل‌های لجستیک یک، دو و سه پارامتری هستند. این مدل‌ها بر اساس تعداد پارامترهایی که در بردارنده نام‌گذاری شده‌اند و برای سؤالاتی با داده‌های دوارزشی مناسب‌اند (همان منبع، ص ۱۸).

- 1- Examine dependent
- 2- Theoretical
- 3- Test equating
- 4- Computerized adaptive testing
- 5- Item-response theory
- 6- Wright & Stone
- 7- Falsifiable
- 8- Invariant

مدل‌های رایج و پرکاربرد سؤال-پاسخ بر ۳ مفروضه اصلی تک‌بعدی بودن^۱، تکنوا بودن^۲ و استقلال موضعی^۳ یا استقلال شرطی استوار هستند. تکنوا بودن به این معناست که احتمال پاسخ درست به سؤال با بالارفتن توانایی افزایش یابد؛ در استقلال موضعی فرض بر این است که پاسخ‌های سؤال‌های مختلف یک آزمون، برای یک سطح خاص از توانایی (θ)، مستقل از یکدیگرند؛ و تک‌بعدی بودن به این معناست که تمام سؤال‌های آزمون، تنها یک صفت یا ویژگی مکنون را اندازه می‌گیرند (مینائی و فلسفی‌نژاد، ۱۳۸۹).

منحنی ویژه سؤال^۴ که دو ویژگی اساسی دارد: سطح دشواری و ضریب تشخیص (شریفی و شریفی، ۱۳۹۱، ص ۴۹۶)؛ پایه اساسی و زیر بنایی نظریه سؤال - پاسخ است و همه ی سازه های دیگر این نظریه بستگی به آن دارد، و به همین دلیل نقش آن در نظریه مذکور به گونه قابل ملاحظه ای مورد توجه قرار گرفته است (بیکر، ۱۳۸۱، ص ۴). سطح دشواری شاخص مکانی^۵ منحنی ویژه سؤال است. بدین معنا که در سؤال‌های آسان سطح دشواری یا شاخص مکانی در بین آزمودنی‌هایی قرار دارد که سطح توانایی پایینی دارند. برعکس، در سؤال‌های دشوار سطح دشواری در بین افراد دارای توانایی بالا قرار می‌گیرد (شریفی و شریفی، ۱۳۹۱، ص ۴۹۶). خم ویژه تست^۶ (TCC) نیز بواقع رابط تابعی^۷ بین نمره حقیقی و مقیاس توانایی است و می‌توان از طریق آن نمره حقیقی متناظر با هر یک از سطوح توانایی را پیدا کرد. نقش اساسی خم ویژه سؤال در نظریه سؤال - پاسخ تهیه ابزاری است که از طریق آن بتوانیم نمره های توانایی را به نمره های حقیقی تبدیل کنیم (بیکر، ۱۳۸۱، ص ۵۷-۵۱). همچنین، یکی دیگر از مفاهیم بنیادی نظریه سؤال - پاسخ تابع آگاهی^۸ است. وقتی صحبت از داشتن آگاهی می‌شود، معنای ضمنی آن این است که شما اطلاعاتی در مورد یک موضوع خاص دارید (بارکر، ۲۰۰۱، به نقل از دلاور و زهراکار، ۱۳۹۲، ص ۲۶۸).

در بحث بررسی ویژگی‌های روان‌سنجی آزمون‌های سرنوشت‌ساز تحقیقاتی هر چند مختصر صورت گرفته است که می‌توان به پژوهش یونسی (۱۳۸۶) تحت عنوان بررسی ویژگی‌های روان‌سنجی سؤالات آزمون‌های فراگیر رشته روانشناسی دانشگاه پیام نور در سال ۱۳۸۵؛ پژوهش واشقانی فراهانی (۱۳۸۰) تحت عنوان کاربرد روش هم‌ترازسازی همصدک در معادلسازی نمرات آزمونهای ورودی دانشگاه (کنکور سال ۷۸)؛ پژوهش آبکار (۱۳۹۱) تحت عنوان بررسی ویژگی‌های روان‌سنجی سؤالات کنکور سراسری در رشته علوم تجربی سال ۱۳۸۹ از نظر تئوری سؤال - پاسخ؛ پژوهش محمودیان (۱۳۹۰) تحت عنوان تنظیم بانک سؤال درس ریاضی کنکور کارشناسی رشته ریاضی سال ۸۹ بر اساس نظریه سؤال - پاسخ؛ تحقیق احمدی آذر (۱۳۸۷) در خصوص نمره کل سازی با استفاده از تکنیک های نظریه کلاسیک و مقایسه آن با مدل‌های نظریه سؤال - پاسخ در کنکور کارشناسی رشته ریاضی سال ۱۳۸۷ و همچنین مقالاتی در خصوص مقایسه نظریات کلاسیک و سؤال - پاسخ از نظر تئوری و عملی، به‌عنوان نمونه هایی از تحقیقات صورت گرفته در خصوص آزمون‌های سرنوشت‌ساز اشاره کرد. این در حالی است که در خارج از کشور سالانه چندین آزمون استاندارد بین المللی مانند تافل^۹ و آیلتس با استفاده از نظریه سؤال - پاسخ طراحی و اجرا می‌گردد. در راس مؤسسات بین المللی ساخت

-
- 1- Unidimensionality
 - 2- Monotonicity
 - 3- Local independence
 - 4- Item characteristic curve
 - 5- Location index
 - 6- Test characteristic curve
 - 7- Functional relation
 - 8- Information function
 - 9- TOEFL and IELTS

آزمون، مؤسسه خدمات آزمون‌های آموزشی (ETS)^۱ قرار دارد که سالانه بیش از ۲۰ میلیون آزمون در امریکا و ۱۸۰ کشور دیگر برگزار می‌نماید.

شایان‌ذکر است که محققین مؤسسه خدمات آزمون‌های آموزشی، بیشتر از اشخاص دیگر مؤسسات، از نظریه سؤال - پاسخ استفاده کاربردی می‌نمایند (کارلسون و داویر^۲، ۲۰۱۳). سالانه محققین این مؤسسه گزارشی از روند تحقیقات خود ارائه می‌نمایند. برای مثال، کارلسون و داویر (۲۰۱۳) در گزارش هشتم به نقش محققین این مؤسسه در بسط نظریه سؤال - پاسخ (که بیشتر در برنامه‌ها و فرایندهای ارزیابی جهانی در مقیاس وسیع استفاده می‌شود) پرداخته‌اند. به‌طور کلی می‌توان گفت که محققین این مؤسسه در خط مقدم بسط و گسترش نظریه سؤال - پاسخ بوده‌اند طوریکه این گسترش با تحقیقات تاکر و لرد^۳ در دهه‌های ۴۰ تا ۵۰ که به زمینه‌های این نظریه کمک کرده‌اند، آغاز گردیده است (همان منبع). همچنین، در کشور همسایه ما پاکستان، از سال ۱۹۸۴ حداقل ۱۹ ارزیابی مقدماتی آموزشی انجام گرفته است. این تلاش‌ها بعد از اعلامیه جهانی در مورد آموزش در ۱۹۹۰ شدت گرفت و دوباره با پیوستن به ابتکار جهانی یونیسف - یونسکو^۴ برای نظارت^۵ موفقیت یادگیری در ۱۹۹۶ ادامه یافت (اندرابی، داس و خوواج^۶، ۲۰۰۲).

بنا بر نیاز جامعه آموزشی ما، رسیدن به معیارهای جهانی در آزمون‌سازی و کاربردی نمودن نظریه‌های جدید اندازه‌گیری، ما بر آن شدیم که ویژگی‌های روان‌سنجی درس ریاضی سال ۱۳۹۱ گروه ریاضی - فیزیک (به‌عنوان نمونه‌ای از آزمون‌های برگزار شده در سطح ملی) را بر اساس هر دو نظریه کلاسیک و سؤال - پاسخ برآورد نمائیم. درس ریاضی از کاربردی‌ترین دروس پایه نظری است که هر دانش‌آموزی به اهمیت آن آگاه است؛ اما متأسفانه بیشتر آن‌ها توانایی پاسخ به سؤالات کنکوری این درس را ندارند. در این راستا، سؤال‌های پژوهشی ذیل مطرح می‌شود:

- ۱- آیا آزمون ریاضی سال ۹۱ گروه ریاضی - فیزیک از اعتبار مناسبی برخوردار است؟
- ۲- ویژگی‌های سؤالات آزمون ریاضی سال ۹۱ گروه ریاضی - فیزیک بر اساس نظریه کلاسیک چگونه است؟
- ۳- کدام‌یک از مدل‌های نظریه سؤال-پاسخ، بهترین برآزش را با این مجموعه از داده‌ها دارد؟
- ۴- ویژگی‌های سؤالات آزمون ریاضی سال ۹۱ گروه ریاضی - فیزیک بر اساس نظریه سؤال - پاسخ چگونه است؟

روش پژوهش

این تحقیق از نوع توصیفی^۷ و کاربردی^۸ است. جامعه آماری این پژوهش شامل کلیه داوطلبان گروه ریاضی - فیزیک آزمون سراسری سال ۱۳۹۱ بود. نمونه آماری نیز باتوجه‌به در دسترس بودن فهرست تصادفی داوطلبان ورود به دانشگاه‌ها در سال ۹۱، از روش سیستماتیک برای انتخاب گروه نمونه استفاده گردید. حجم نمونه به دلیل اینکه نظریه‌های جدید اندازه‌گیری، نظریه‌های نمونه

1- ETS (Educational Testing Service)
 2- Carlson and Davier
 3- Tucker and Lord
 4- UNESCO-UNICEF
 5- Monitoring
 6- Andrabi, Das and Khwaja
 7- Descriptive research
 8- Applied

بزرگ هستند، ۱۰۰۰ نفر از جامعه به صورت کاملاً تصادفی انتخاب گردید. ابزار مورد استفاده در این تحقیق از نوع پرسش نامه است. بدین منظور از درس ریاضی گروه ریاضی - فیزیک سال ۱۳۹۱ که شامل ۵۵ سؤال است، استفاده گردیده است. سؤال ۳۹ به دلیل داشتن دو گزینه جواب، از تحلیل ها کنار گذاشته شد؛ بنابراین از ۵۴ سؤال باقی مانده استفاده گردید. جهت تجزیه و تحلیل داده ها از نرم افزارهای SPSS و پیرایش ۱۹ و BILOG و برای پاسخ به سؤالات پژوهشی، از روش های آلفای کرونباخ برای بررسی اعتبار آزمون به روش همسانی درونی، روایی محتوایی و روش های تحلیل سؤال بر اساس هر دو نظریه کلاسیک و سؤال - پاسخ استفاده گردید. همچنین جهت تجزیه و تحلیل آزمون درس ریاضی و سؤالات آن بر اساس نظریه سؤال - پاسخ، ابتدا مفروضه های تک بعدی بودن و استقلال موضعی بررسی گردید. همچنین، شایان ذکر است که برای تعیین تعداد ابعاد آزمون ها می توان از نسبت حاصل ارزش ویژه^۱ عامل اول به ارزش ویژه عامل دوم استفاده کرد. یکی دیگر از روشهای بررسی تک بعدی بودن، نمودار اسکری^۲ است.

شایان ذکر است که استقلال موضعی، به طور خودکار از تک بعدی بودن پیروی می کند و مفروضه ای جداگانه به حساب نمی آید (لرد، ۱۹۸۰؛ همبلتون، ۱۹۸۹، به نقل از یونسی، دلاور و فلسفی نژاد، ۱۳۸۹). در این تحقیق جهت تعیین تک بعدی بودن آزمون ها، از روش تحلیل عاملی با نرم افزار SSPS استفاده شد.

یافته ها

* ویژگی های روان سنجی آزمون ریاضی بر اساس نظریه کلاسیک

جهت پاسخ به سؤال اول پژوهشی و تعیین اعتبار (پایایی) مجموعه داده ها از روش آلفای کرونباخ استفاده گردید. ضریب آلفای کرونباخ پرسش نامه ۵۴ سؤالی آزمون ریاضی، برابر ۰/۸۴۶ به دست آمد که ضریب بالایی است و می توان نتیجه گرفت که سؤالات پرسش نامه از همسانی درونی (اعتبار) خوبی برخوردارند. همچنین، حذف تک تک آیتم ها، افزایش قابل توجهی در مقدار ضریب آلفای کل به وجود نیاورد. شایان ذکر است که آزمون ریاضی استفاده شده در این تحقیق، از نوع آزمون های درسی است که از منابع درسی تهیه شده و روایی^۳ محتوایی آن توسط طراحان سؤال سازمان سنجش مورد بررسی و تأیید قرار می گیرد؛ بنابراین، آزمون ریاضی از روایی محتوایی مناسبی برخوردار است.

* تحلیل سؤالات آزمون ریاضی بر اساس نظریه کلاسیک

از آنجایی که ملاک مناسب بودن سؤال از نظر درجه دشواری، ضریب بین ۰/۳ تا ۰/۷ است؛ بنابراین، سؤالات دارای ضریب کمتر از ۰/۳ و بیشتر از ۰/۷ نامناسب تلقی خواهند شد. همچنین، در این تحقیق، ملاک نامناسب بودن سؤال از نظر ضریب تشخیص سؤال در نظریه کلاسیک، ضریب کمتر از ۰/۲ در نظر گرفته شد. ضریب تشخیص منفی نیز نشان می دهد که در آن سؤال، گروه قوی بدتر از گروه ضعیف عمل کرده است. این گونه سؤالات دارای معایب اساسی هستند که یا باید به کلی کنار گذاشته شوند یا در آن ها تجدید نظر اساسی صورت پذیرد. اگر سؤال دارای اشکال فنی نباشد، دلیل چنین ضریبی می تواند این باشد که افراد گروه بالا یا آن سؤال را

1- eigenvalue
2- Scree plot
3- Validity

به طور کامل یاد نگرفته اند یا آن را به غلط آموخته اند؛ بنابراین، تمامی ضریب های تشخیص سؤال های آزمون باید مثبت باشند (سیف، ۱۳۹۰، ص ۳۱۳).

جهت پاسخ به سؤال دوم پژوهشی و بر اساس جدول ۱ که از تحلیل سؤالات بر اساس نظریه کلاسیک با نرم افزار بایلوگ استخراج گردیده؛ تعداد افرادی که به سؤال پاسخ درست داده‌اند، دشواری، ضریب همبستگی پیرسون و ضریب تشخیص سؤالات نمایش داده شده است.

جدول ۱. جدول تحلیل سؤالات آزمون ریاضی بر اساس نظریه کلاسیک آزمون

شماره سؤال	تعداد افراد پاسخ‌دهنده به سؤال	تعداد افرادی که به سؤال پاسخ درست داده‌اند	دشواری سؤال	ضریب همبستگی پیرسون	ضریب تشخیص
۱	۱۰۰۰	۲۲۲	۰/۲۲۲	۰/۱۵۶	۰/۲۱۸
۲	۱۰۰۰	۴۲۲	۰/۴۲۲	۰/۲۸۵	۰/۳۶۰
۳	۱۰۰۰	۶۱۴	۰/۶۱۴	۰/۱۶۸	۰/۲۱۴
۴	۱۰۰۰	۲۷۱	۰/۲۷۱	۰/۵۰۵	۰/۶۷۸
۵	۱۰۰۰	۲۰۶	۰/۲۰۶	۰/۲۸۶	۰/۴۰۵
۶	۱۰۰۰	۲۵۷	۰/۲۵۷	۰/۱۰۳	۰/۱۴
۷	۱۰۰۰	۱۴۹	۰/۱۴۹	۰/۴۴۲	۰/۶۷۸
۸	۱۰۰۰	۳۱۴	۰/۳۱۴	۰/۳۳۵	۰/۴۳۹
۹	۱۰۰۰	۱۵۴	۰/۱۵۴	۰/۴۰۱	۰/۶۱
۱۰	۱۰۰۰	۲۶۳	۰/۲۶۳	۰/۲۱۱	۰/۲۸۵
۱۱	۱۰۰۰	۱۹۴	۰/۱۹۴	۰/۳۸۶	۰/۵۵۶
۱۲	۱۰۰۰	۱۷۳	۰/۱۷۳	۰/۴۶۷	۰/۶۹۱
۱۳	۱۰۰۰	۱۵۶	۰/۱۵۶	۰/۲۴۷	۰/۳۷۴
۱۴	۱۰۰۰	۲۷۹	۰/۲۷۹	۰/۰۲۳	۰/۰۳۱
۱۵	۱۰۰۰	۳۶۱	۰/۳۶۱	۰/۰۸۶	۰/۱۱
۱۶	۱۰۰۰	۲۳۲	۰/۲۳۲	۰/۴۸	۰/۶۶۳
۱۷	۱۰۰۰	۲۵۴	۰/۲۵۴	۰/۵۶۸	۰/۷۷۲
۱۸	۱۰۰۰	۲۱۴	۰/۲۱۴	۰/۵۳۹	۰/۷۵۸
۱۹	۱۰۰۰	۲۳۱	۰/۲۳۱	۰/۴۱۶	۰/۵۷۶
۲۰	۱۰۰۰	۷۱	۰/۰۷۱	۰/۳۰۵	۰/۵۷۶
۲۱	۱۰۰۰	۱۳۴	۰/۱۳۴	۰/۳۳۷	۰/۵۳۲
۲۲	۱۰۰۰	۶۵	۰/۰۶۵	۰/۴۳	۰/۸۳۶
۲۳	۱۰۰۰	۱۱۱	۰/۱۱۱	۰/۴۸۱	۰/۷۹۹
۲۴	۱۰۰۰	۷۲	۰/۰۷۲	۰/۳۱	۰/۵۸۴
۲۵	۱۰۰۰	۴۹	۰/۰۴۹	۰/۲۹۲	۰/۶۲
۲۶	۱۰۰۰	۲۰۲	۰/۲۰۲	-۰/۰۱۸	-۰/۰۲۵
۲۷	۱۰۰۰	۳۸	۰/۰۳۸	۰/۱۲۵	۰/۲۸۹
۲۸	۱۰۰۰	۲۰۱	۰/۲۰۱	۰/۵۸۳	۰/۸۳۱

۰/۲۰۶	۰/۱۱۵	۰/۰۸۵	۸۵	۱۰۰۰	۲۹
۰/۲۸۶	۰/۱۴۵	۰/۰۶۲	۶۲	۱۰۰۰	۳۰
۰/۷۲۳	۰/۴۰۶	۰/۰۸۷	۸۷	۱۰۰۰	۳۱
۰/۲۷۲	۰/۲۱۷	۰/۴۷۲	۴۷۲	۱۰۰۰	۳۲
۰/۲۷۵	۰/۲۱۲	۰/۳۳۴	۳۳۴	۱۰۰۰	۳۳
۰/۴۷۸	۰/۲۹۸	۰/۱۲۶	۱۲۶	۱۰۰۰	۳۴
۰/۷۱۶	۰/۴۷۲	۰/۱۵۶	۱۵۶	۱۰۰۰	۳۵
۰/۸۱۹	۰/۴۷۸	۰/۰۹۹	۹۹	۱۰۰۰	۳۶
۰/۹۳	۰/۵۳۸	۰/۰۹۶	۹۶	۱۰۰۰	۳۷
۰/۶۱۶	۰/۳۹۷	۰/۱۴۳	۱۴۳	۱۰۰۰	۳۸
۰/۱۳۸	۰/۱	۰/۲۳	۲۳۰	۱۰۰۰	۴۰
۰/۷۵۹	۰/۴۹۲	۰/۱۴۶	۱۴۶	۱۰۰۰	۴۱
-۰/۱۱۱	-۰/۰۸۲	۰/۲۶	۲۶۰	۱۰۰۰	۴۲
۰/۲۸۹	۰/۲۱۷	۰/۲۸۶	۲۸۶	۱۰۰۰	۴۳
۰/۸۱	۰/۵۴۵	۰/۱۶۹	۱۶۹	۱۰۰۰	۴۴
۰/۳۳۶	۰/۱۸۸	۰/۰۸۵	۸۵	۱۰۰۰	۴۵
۰/۳۶	۰/۲۰۹	۰/۰۹۷	۹۷	۱۰۰۰	۴۶
۰/۳۳۲	۰/۱۴۳	۰/۰۳۷	۳۷	۱۰۰۰	۴۷
۰/۴۶۲	۰/۲۲۷	۰/۰۵۶	۵۶	۱۰۰۰	۴۸
۰/۲۰۱	۰/۰۸۲	۰/۰۳۱	۳۱	۱۰۰۰	۴۹
۰/۶۸۵	۰/۵۱۱	۰/۲۷۳	۲۷۳	۱۰۰۰	۵۰
۰/۰۱۴	۰/۰۱	۰/۲۳۱	۲۳۱	۱۰۰۰	۵۱
۰/۷۲	۰/۴۹۴	۰/۱۸۴	۱۸۴	۱۰۰۰	۵۲
۰/۴۶۹	۰/۲	۰/۰۳۶	۳۶	۱۰۰۰	۵۳
۰/۱۹۴	۰/۱۴۳	۰/۲۵۸	۲۵۸	۱۰۰۰	۵۴
-۰/۰۴۳	-۰/۰۳۲	۰/۲۶۰	۲۶۰	۱۰۰۰	۵۵

باتوجه به نتایج جدول ۱، می‌توان گفت که بغیر از سؤالات ۲-۳-۸-۱۵-۳۲-۳۳ بقیه سؤالات دارای درجه دشواری کمتر از ۰/۳ می باشند و این بدین معناست که بغیر از سؤالات مذکور، مابقی سؤالات خیلی دشوار بوده اند و در نتیجه اطلاعات زیادی را درباره تفاوت های آزمودنی‌ها ارائه نمی کنند. با بررسی سطوح دشواری سؤالات به این نکته پی می بریم که آیتم‌ها به ترتیب از آسان به دشوار مرتب نشده اند. همچنین، سؤالات ۶-۱۴-۱۵-۲۶-۴۰-۴۲-۵۱-۵۴-۵۵ از نظر ضریب تشخیص، نامناسب شناخته می شوند. سؤال ۳۷ و ۵۱ به ترتیب، دارای قوی ترین و ضعیف ترین ضریب تشخیص در بین سؤالات آزمون است. سؤالات ۲۶، ۴۲ و ۵۵ نیز دارای ضریب تشخیص منفی هستند.

*ویژگی‌های روان‌سنجی آزمون ریاضی بر اساس نظریه سؤال-پاسخ

**بررسی مفروضه‌های نظریه سؤال - پاسخ

الف- تک‌بعدی بودن و استقلال موضعی: جهت بررسی تک‌بعدی بودن آزمون ریاضی، تحلیل عاملی اکتشافی با نرم‌افزار SPSS صورت گرفت. قبل از انجام تحلیل عاملی، باید از اعتبار داده‌ها اطمینان حاصل کرد. جهت بررسی اعتبار داده‌ها از آزمون کیزر - مایر - الکین (KMO) که روشی برای بررسی صحت نمونه‌گیری است، استفاده شده است.

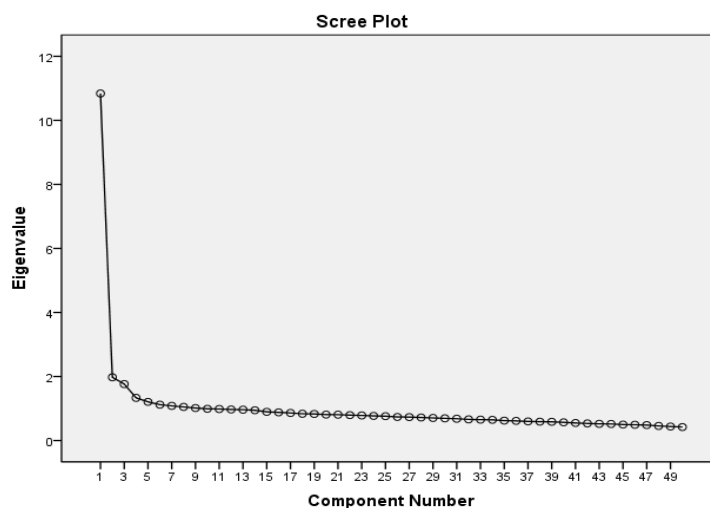
جدول ۲. آزمون کیزر - مایر - الکین و کرویت بارتلت

۰/۹۲	آزمون کفایت نمونه‌گیری کیزر - مایر - اولکین	
۱۲۳۹۹/۷	مقدار خی دو	
۱۴۳۱	درجات آزادی	آزمون کرویت بارتلت
۰/۰۰۰	سطح معناداری	

بر اساس نتایجی که در جدول ۲ نشان داده شده است، ملاک کفایت نمونه‌گیری کیزر - مایر - الکین (۰/۹۲)، نشان می‌دهد که داده‌ها برای تحلیل مولفه‌های اصلی مناسب‌اند. همچنین، آزمون کرویت بارتلت معنادار بود که نشان می‌دهد برای اجرای تحلیل بین متغیرها همبستگی کافی وجود دارد. در جدول ۴ نیز ارزش‌های ویژه دو عامل استخراجی اول نشان داده شده است.

جدول ۳. نتایج بررسی مفروضه‌های نظریه سؤال-پاسخ در آزمون درس ریاضی

عامل	مقادیر ویژه	درصد واریانس
اول	۱۰/۸	۲۱/۷
دوم	۱/۹	۳/۹۵



نمودار ۱

همان‌طور که در جدول ۳ نشان داده شده است، تحلیل عاملی اکتشافی داده‌ها نشان می‌دهد که عامل اول با ارزش ویژه ۱۰/۸ حدود ۲۱/۷ درصد واریانس عامل‌ها را تبیین کرده است (خیلی بیشتر از ۳ برابر عامل دوم با ارزش ویژه ۱/۹ و ۳/۹۵ درصد واریانس). این عامل غالب توسط نمودار اسکری (نمودار ۱) نیز تأیید شده است؛ بنابراین، می‌توان نتیجه گرفت که این آزمون تک‌بعدی است. از آنجایی که استقلال موضعی، به طور خودکار از تک‌بعدی بودن پیروی می‌کند و مفروضه‌های جداگانه به حساب نمی‌آید (لرد، ۱۹۸۰؛ همبلتون، ۱۹۸۹، به نقل از یونسی، دلاور و فلسفی‌نژاد، ۱۳۸۹)، بنابراین این آزمون از استقلال موضعی نیز برخوردار است.

ب- برازش مدل - داده‌ها^۱

جهت برازش مدل - داده‌ها از مقدار $-2\log \text{likelihood}$ در آخرین چرخش استفاده می‌شود. بر اساس پاسخ به سؤال سوم پژوهشی، جهت بررسی برازش داده‌ها با بهترین مدل، تحلیل داده‌ها با هر سه مدل یک، دو و سه پارامتری انجام گردید. از آنجاکه نتایج تحلیل کلاسیک، نشان داد آزمون ریاضی از سری سؤالات سخت تشکیل یافته و احتمال حدس بسیار پائین بوده و همچنین، نرسیدن چرخش‌ها به ملاک همگرایی با در نظر گرفتن تحلیل داده‌ها بر اساس مدل ۳ پارامتری نظریه سؤال - پاسخ؛ بنابراین، می‌توان نتیجه گرفت که مجموعه داده‌های این تحقیق با مدل ۳ پارامتری برازش ندارند.

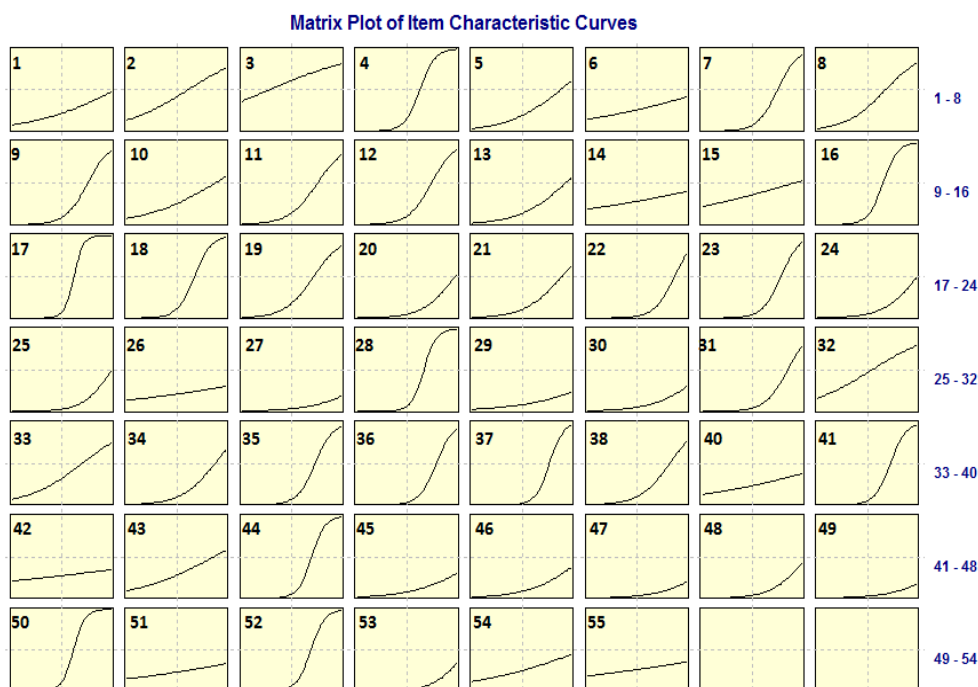
در نهایت، جهت بررسی برازش مدل - داده‌ها (مقایسه مدل ۱ و ۲ پارامتری) از آزمون خی دو^۲ استفاده شد. مقدار $-2\log \text{likelihood}$ برای مدل‌های ۱ و ۲ پارامتری به ترتیب برابر است با $۴۴۷۹۰/۸۳۳۷$ و $۴۲۳۹۶/۰۴۴۵$. خی دو محاسبه شده از خی دو بحرانی در هر دو سطح اطمینان ۹۹ و ۹۵ درصد ($۷۶/۱۵$ و $۶۷/۵$) با درجات آزادی ۵۴، بزرگ‌تر است؛ بنابراین، مدل دو پارامتری برازش بهتری با مجموعه داده‌های آزمون دارد. شایان ذکر است که شاخص ریشه دوم میانگین مربع‌ها^۳ (RMS)، به‌عنوان شاخصی برای قضاوت در مورد برازش کلی مدل استفاده می‌شود و نباید بیشتر از ۲ باشد (امبرتسون و رایس، ۱۳۸۸، ص ۴۹۳). در این پژوهش با توجه به برازش داده‌ها با مدل دو پارامتری؛ چون مقدار RMS کمتر از ۲ است، می‌توان نتیجه گرفت که مدل دو پارامتری کاملاً با داده‌ها برازش دارد.

باتوجه به برازش داده‌ها با مدل دو پارامتری، جهت تحلیل سؤالات بر اساس نظریه سؤال - پاسخ، از این مدل استفاده نمودیم. بر اساس پاسخ به سؤال چهارم پژوهشی؛ ابتدا خم‌های ویژه سؤالات که از مهم‌ترین قسمت‌های خروجی برنامه بایلوگ است، در نمودار ۲ نشان داده شده است.

1- Model – data fit

۲: خی دو محاسبه شده بر اساس این فرمول برآورد می‌گردد: $\chi^2 = [(-2\log \text{likelihood } 1) - (-2\log \text{likelihood } 2)]$

3- root – mean – square (RMS)



نمودار ۲. منحنی‌های ویژگی تمامی سؤالات آزمون ریاضی در مدل دو پارامتری

نمودار ۲، شمای کلی از آزمون را از لحاظ سطح دشواری و قدرت تشخیص سؤالات طبق مدل ۲ پارامتری سؤال - پاسخ نشان می‌دهد. برای مثال، سؤال ۴ دارای شیب و ضریب تشخیصی مناسب است، به طوری‌که به خوبی توانسته گروه قوی و ضعیف را از هم تفکیک کند. در جدول ۴ نیز ضریب دشواری و تشخیص، بار عاملی سؤال با عامل زیربنایی، ضریب خی دو، سطح معناداری و درجه دشواری هر سؤال مشخص شده است.

جدول ۴. جدول تحلیل سؤالات آزمون ریاضی بر اساس نظریه سؤال - پاسخ

شماره سؤال	قدرت تشخیص سؤال (a)	دشواری سؤال (B)	بار عاملی سؤال با عامل زیربنایی	خی دو	سطح معناداری	درجه آزادی
۱	۰/۲۳۵	۳/۲۲	۰/۲۲	۳۱/۴	۰/۰۰۰۱	۷
۲	۰/۲۹۸	۰/۶۵۸	۰/۲۸۶	۵۹	۰/۰۰	۷
۳	۰/۲۰۱	-۱/۳۸۳	۰/۱۹۷	۳۷/۷	۰/۰۰	۸
۴	۱/۲۹۷	۰/۷۶۷	۰/۷۹۲	۱۶/۶	۰/۰۱۰۷	۶
۵	۰/۳۷۴	۲/۲۹۹	۰/۳۵	۲۹/۲	۰/۰۰۰۱	۷
۶	۰/۱۴۱	۴/۳۹۵	۰/۱۴۰	۵۳/۳	۰/۰۰۰	۷
۷	۰/۹۸۷	۱/۵۰۶	۰/۷۰۲	۲/۴	۰/۸۷۶	۶
۸	۰/۴۹۲	۱/۰۷۲	۰/۴۴۲	۱۷/۴	۰/۰۱۴۹	۷

۶	۰/۲۶۱۳	۷/۷	۰/۶۶۵	۱/۵۵۳	۰/۸۹۱	۹
۷	۰/۱۰۷۳	۱۱/۸	۰/۲۶۸	۲/۲۸۲	۰/۲۷۸	۱۰
۶	۰/۵۲۸۱	۵/۱	۰/۵۶۴	۱/۵۳۳	۰/۶۸۳	۱۱
۶	۰/۰۰۴۳	۱۸/۹	۰/۶۵۵	۱/۴۵۴	۰/۸۶۶	۱۲
۷	۰/۱۹۴۹	۹/۹	۰/۳۸۷	۲/۵۹۴	۰/۴۲	۱۳
۸	۰/۰۰	۷۲/۷	۰/۱۰۳	۵/۲۵۱	۰/۱۰۴	۱۴
۸	۰/۰۰	۵۶/۵	۰/۱۳۶	۲/۴۴۳	۰/۱۳۷	۱۵
۶	۰/۱۵۳۱	۹/۴	۰/۷۹۶	۰/۹۲۳	۱/۳۱۶	۱۶
۴	۰/۰۰۶۶	۱۴/۲	۰/۹۰۳	۰/۷۲۴	۲/۰۹۸	۱۷
۶	۰/۱۰۲۷	۱۰/۶	۰/۷۶	۱/۰۵۱	۱/۱۶۸	۱۸
۷	۰/۰۰۱۵	۲۳/۴	۰/۵۶۵	۱/۲۹۸	۰/۶۸۴	۱۹
۶	۰/۳	۷/۲	۰/۵۲۷	۲/۸۶۸	۰/۶۲	۲۰
۶	۰/۰۴۴۷	۱۲/۹	۰/۴۷۱	۲/۳۶۵	۰/۵۳۴	۲۱
۵	۰/۴۲۳۳	۴/۹	۰/۷۰۳	۲/۲۳	۰/۹۸۸	۲۲
۶	۰/۵۳۲۷	۵/۱	۰/۷۳۲	۱/۷۰۹	۱/۰۷۴	۲۳
۶	۰/۰۰۷۸	۱۷/۴	۰/۵۰۲	۲/۹۹۵	۰/۵۸	۲۴
۶	۰/۴۱۷	۶/۱	۰/۵۷۶	۲/۹۹	۰/۷۰۵	۲۵
۸	۰/۰۰۰	۵۷/۵	۰/۱۰۱	۷/۷۱۱	۰/۱۰۱	۲۶
۶	۰/۶۱۱۷	۴/۵	۰/۳۵۵	۵/۲۴۲	۰/۳۷۹	۲۷
۵	۰/۸۱۲۳	۲/۳	۰/۸۴۵	۱/۰۰۲	۱/۵۷۷	۲۸
۶	۰/۰۰۰۰	۴۸/۲	۰/۲۳۵	۵/۸۶۱	۰/۲۴۲	۲۹
۶	۰/۷۴۶۶	۳/۵	۰/۳۷۷	۴/۲۰۳	۰/۴۰۷	۳۰
۶	۰/۸۵۱۶	۲/۶	۰/۶۶۱	۲/۱۱۴	۰/۸۸۱	۳۱
۷	۰/۰۱۵۲	۱۷/۴	۰/۲۸۵	۰/۲۳۶	۰/۲۹۷	۳۲
۷	۰/۴۴۹۶	۶/۸	۰/۳۴۱	۱/۲۱۲	۰/۳۶۳	۳۳
۶	۰/۱۷۰۳	۹/۱	۰/۵۰۶	۲/۲۸۵	۰/۵۸۷	۳۴
۶	۰/۵۹۷۵	۴/۶	۰/۷۳۹	۱/۳۹	۱/۰۹۸	۳۵
۵	۰/۴۴۷۱	۴/۷	۰/۷۵۵	۱/۷۵۲	۱/۱۵۱	۳۶
۴	۰/۲۸۷	۵	۰/۸۱۹	۱/۶۳۹	۱/۴۳۰	۳۷
۶	۰/۰۰۰۱	۲۷/۴	۰/۵۵۶	۱/۹۳۸	۰/۶۶۸	۳۸
۸	۰/۰۰۰۰	۰/۵۳	۰/۱۳۸	۵/۰۵۷	۰/۱۳۹	۴۰
۵	۰/۰۰۰۲	۲۴/۸	۰/۷۸۳	۱/۳۷۲	۱/۲۵۹	۴۱
۸	۰/۰۰۰۰	۱۳۴/۹	۰/۰۶۹	۸/۳۰۱	۰/۰۶۹	۴۲
۸	۰/۰۰۰۵	۲۷/۹	۰/۲۴۸	۲/۱۸۲	۰/۲۵۶	۴۳
۵	۰/۰۰۲۳	۱۸/۵	۰/۸۱۹	۱/۱۸۹	۱/۴۲۸	۴۴
۶	۰/۰۱۵۴	۱۵/۷	۰/۳۰۸	۴/۵۰۹	۰/۳۲۴	۴۵
۷	۰/۱۶۹۸	۱۰/۳	۰/۳۴۹	۳/۷۵۶	۰/۳۷۲	۴۶
۶	۰/۰۱۵۳	۱۵/۷	۰/۳۷۶	۴/۹۹۱	۰/۴۰۶	۴۷
۶	۰/۰۸۷۱	۱۱	۰/۵۰۴	۳/۲۷۳	۰/۵۸۳	۴۸
۵	۰/۰۱۴۵	۱۴/۲	۰/۳۷۶	۵/۲۳۹	۰/۴۰۶	۴۹
۵	۰/۱۵۶	۸	۰/۸۵۸	۰/۶۹۵	۱/۶۷۴	۵۰
۸	۰/۰۰۰۰	۷۷/۶	۰/۱۰۱	۶/۷۳۱	۰/۱۰۲	۵۱

۵	۰/۰۰۰۲	۲۳/۹	۰/۸۰۹	۱/۱۲۷	۱/۳۷۷	۵۲
۵	۰/۰۶۷۴	۱۰/۳	۰/۵۲۵	۳/۶۰۵	۰/۶۱۷	۵۳
۸	۰/۰۰۰۰	۴۴/۴	۰/۱۷	۳/۶۲۶	۰/۱۷۳	۵۴
۸	۰/۰۰۰۰	۸۵/۱	۰/۰۸۶	۶/۸۴۸	۰/۰۸۶	۵۵

باتوجه به مقادیر جدول ۴، ضریب تشخیص هیچ سؤالی منفی نیست و این بدان معناست که در تمام سؤالات آزمون، احتمال پاسخ درست با افزایش سطح توانایی، افزایش می‌یابد (در خم‌های ویژه سؤالات و جدول فوق به وضوح مشخص است به طریقه هیچ نموداری شیب منفی ندارد و هیچ‌کدام از مقادیر b ها نیز منفی نیست).

همچنین، در این تحقیق ملاک نامناسب بودن سؤال بر اساس نظریه سؤال-پاسخ، ضریب تشخیص کمتر از ۰/۳۵ که ضریب خیلی ضعیفی است، در نظر گرفته شد. سؤالات دارای ضریب تشخیص پائین‌تر از ۰/۳۵ عبارت‌اند از؛ ۱-۲-۳-۶-۱۰-۱۴-۱۵-۲۶-۲۹-۳۲-۴۰-۴۲-۴۳-۴۵-۵۱-۵۴-۵۵. مابقی سؤالات دارای ضریب تشخیص متوسط یا بالا است. سؤال ۱۷ دارای بیشترین و سؤال ۴۲ دارای کمترین مقدار ضریب تشخیص آزمون ریاضی بر اساس نظریه سؤال - پاسخ است.

مقادیر b (سطح دشواری) به طور معمول از (۲+ تا ۲-) تغییر می‌کند. مقادیر نزدیک به ۲- به سؤالاتی مربوط می‌شود که خیلی آسانند و مقادیر نزدیک ۲ به سؤالاتی مربوط می‌شوند که برای آزمودنی‌ها خیلی سخت می‌باشند (همبلتون و همکاران، ۱۳۸۹). بر اساس نتایج تحلیل‌های کلاسیک و سؤال - پاسخ می‌توان نتیجه گرفت که سؤالات آزمون ریاضی ۹۱، جزء سؤالات سخت محسوب می‌شوند. همچنین، سؤال ۴۲ با کمترین ضریب تشخیص و بیشترین ضریب دشواری (۸/۳)، سخت‌ترین و سؤال ۳ نیز، آسان‌ترین سؤال این آزمون بر اساس نظریه سؤال - پاسخ است.

سؤال‌هایی که سطح معناداری آن‌ها از مقدار آلفا کوچک‌تر باشد، عدم برازش آن سؤال با مدل مربوطه را نشان خواهد داد (امبرسون و رایس، ۱۳۸۸، ص ۲۹۹). طبق نتایج جدول ۴، تنها ۲۴ سؤال با مدل ۲ پارامتری برازش دارند. ۴۰ سؤال آزمون با مدل ۱ پارامتری و ۴۷ سؤال نیز با مدل ۳ پارامتری برازش ندارند. سؤالاتی که با هیچ‌کدام از مدل‌ها برازش ندارند نیز بایستی حذف شوند. آیتم‌هایی که با هر سه مدل نظریه سؤال - پاسخ، برازش ندارند؛ عبارت‌اند از: ۱-۲-۳-۴-۶-۱۲-۱۴-۱۵-۱۷-۱۹-۲۱-۲۶-۲۹-۳۸-۴۰-۴۱-۴۳-۴۴-۴۵-۵۱-۵۴.

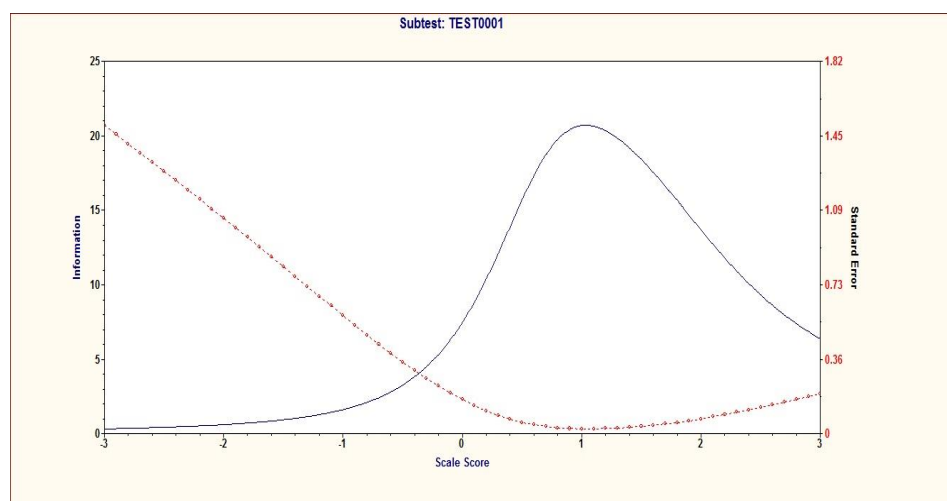
در جدول ۵ نیز بیشینه آگاهی سؤالات و همچنین نقطه‌ای که بیشینه آگاهی در آن به وجود می‌آید را نشان داده ایم. در مدل دو پارامتری، نقطه‌ای که بیشینه آگاهی در آن به وجود می‌آید، برابر با مقدار سطح دشواری سؤال است و بیشینه آگاهی سؤال نیز با فرمول $1/4.D^2.a^2$ یا $(.72a^2)$ محاسبه می‌گردد.

جدول ۵. بیشینه آگاهی و نقاط بیشینه آگاهی سؤالات در مدل 2pl درس ریاضی

سؤال	نقطه بیشینه آگاهی	بیشینه سؤال	نقطه بیشینه آگاهی	بیشینه سؤال	نقطه بیشینه آگاهی	بیشینه سؤال	نقطه بیشینه آگاهی	بیشینه سؤال	نقطه بیشینه آگاهی		
۱	۳/۲۲	۰/۰۳۹	۱۵	۲/۴۴۳	۰/۰۱۴	۲۹	۵/۸۶۱	۰/۰۴۳	۴۴	۱/۱۸۹	۱/۴۶۹

۰/۰۷۶	۴/۵۰۹	۴۵	۰/۱۲۰	۴/۲۰۳	۳۰	۱/۲۵	۰/۹۲۳	۱۶	۰/۰۶۳	۰/۶۵۸	۲
۰/۰۹۹۷	۳/۷۵۶	۴۶	۰/۱۵۵۹	۲/۱۱۴	۳۱	۳/۱۷	۰/۷۲۴	۱۷	۰/۰۲۹	-۱/۳۸۳	۳
۰/۱۱۹	۴/۹۹۱	۴۷	۰/۰۶۴	۰/۲۳۶	۳۲	۰/۹۸۳	۱/۰۵۱	۱۸	۱/۲۲	۰/۷۶۷	۴
۰/۲۴۵	۳/۲۷۳	۴۸	۰/۰۹۵	۱/۲۱۲	۳۳	۰/۳۳۷	۱/۲۹۸	۱۹	۰/۱	۲/۲۹۹	۵
۰/۱۳۱	۵/۲۳۹	۴۹	۰/۲۴۹	۲/۲۸۵	۳۴	۰/۲۷۷	۲/۸۶۸	۲۰	۰/۰۱۵	۴/۳۹۵	۶
۲/۰۱۸	۰/۶۹۵	۵۰	۰/۱۸۶۹	۱/۳۹	۳۵	۰/۲۱	۲/۳۶۵	۲۱	۰/۷۱	۱/۵۰۶	۷
۰/۰۰۸	۶/۷۳۱	۵۱	۰/۹۵۴	۱/۷۵۲	۳۶	۰/۷۰۳	۲/۲۳	۲۲	۰/۱۷۵	۱/۰۷۲	۸
۱/۳۶۶	۱/۱۲۷	۵۲	۱/۴۷۳	۱/۶۳۹	۳۷	۰/۸۳۱	۱/۷۰۹	۲۳	۰/۵۷۲	۱/۵۵۳	۹
۰/۲۷۵	۳/۶۰۵	۵۳	۰/۳۲۲	۱/۹۳۸	۳۸	۰/۲۴۳	۲/۹۹۵	۲۴	۰/۰۵۶	۲/۲۸۲	۱۰
۰/۰۲۲	۳/۶۲۶	۵۴	۰/۰۱۴	۵/۰۵۷	۴۰	۰/۳۵۸	۲/۹۹	۲۵	۰/۳۳۶	۱/۵۳۳	۱۱
۰/۰۰۵	۶/۸۴۸	۵۵	۱/۱۴۲	۱/۳۷۲	۴۱	۰/۰۰۸	۷/۷۱۱	۲۶	۰/۵۴	۱/۴۵۴	۱۲
			۰/۰۰۴	۸/۳۰۱	۴۲	۰/۱۰۴	۵/۲۴۲	۲۷	۰/۱۲۸	۲/۵۹۴	۱۳
			۰/۰۴۸	۲/۱۸۲	۴۳	۱/۷۹۱	۱/۰۰۲	۲۸	۰/۰۰۸	۵/۲۵۱	۱۴

در نهایت نمودار ۳، منحنی آگاهی و خطای استاندارد آزمون ریاضی را نشان می دهد. توجه داشته باشید که خطوط نقطه چین بیانگر منحنی خطای استاندارد آزمون است.



نمودار ۳. منحنی آگاهی و خطای استاندارد آزمون ریاضی

همان‌طور که مشاهده می‌کنید، نمودار ۳ به‌خوبی گویای منحنی آگاهی یک آزمون دشوار است، به‌طوری‌که افراد با توانایی پائین کمتر می‌توانند به سؤالات پاسخ درست بدهند. بیشترین میزان آگاهی دهندگی این آزمون، در فاصله تتای ۰/۵ تا ۱/۵ است. آزمودنی‌هایی که توانایی کمتر از ۰ دارند نمی‌توانند در این آزمون موفق باشند و خطای بیشتری در پاسخگویی به سؤالات مرتکب می‌شوند. همچنین، خطای استاندارد آزمون در نقاطی که میزان آگاهی دهندگی بیشینه است، به کمترین میزان خود می‌رسد.

ضریب آلفای کرونباخ آزمون ریاضی کنکور گروه ریاضی - فیزیک، ۰/۸۴۶ برآورد شد به طوریکه میزان همسانی درونی خوب سؤالات را آشکار می‌سازد. حذف تک‌تک سؤالات نیز تأثیر چندانی در افزایش این ضریب ایجاد نکرد. همچنین، آزمون‌های سراسری که در سازمان سنجش تهیه می‌گردند از لحاظ روایی مورد تأیید طراحان و متخصصان است؛ بنابراین، می‌توان نتیجه گرفت که این آزمون از روایی محتوایی و اعتبار مناسبی برخوردار است.

تجزیه و تحلیل داده‌ها بر اساس نظریه کلاسیک آزمون برای پاسخ به سؤال دوم پژوهش، نشان داد که به‌غیر از سؤالات ۲-۳-۸-۱۵-۳۲-۳۳، بقیه سؤالات دارای درجه دشواری کمتر از ۰/۳ هستند و این بدین معناست که اکثر سؤالات آزمون ریاضی، جزء سؤالات دشوار بوده‌اند. در ضمن، چیدمان سؤالات از ساده به دشوار صورت نگرفته بود. ۹ سؤال از نظر ضریب تشخیص، نامناسب شناخته شدند. سؤالات ۴۲-۲۶ و ۵۵ نیز دارای ضریب تشخیص منفی بوده‌اند که بر اساس نظریه کلاسیک دارای اشکال اساسی می‌باشند.

باتوجه به تحقیقات صورت‌گرفته در چند دهه گذشته، ارجحیت نظریه سؤال - پاسخ بر کلاسیک بارها ثابت شده است. نتایج تحقیق مطیعی (۱۳۸۱) به منظور مقایسه دو نظریه کلاسیک و سؤال - پاسخ از نظر تأثیر خصوصیات آزمودنی‌ها بر ویژگی‌های سؤال و بالعکس نیز نشان داد؛ ویژگی‌های دو گروه نمونه که از نظر توانایی با یکدیگر تفاوت دارند، بر برآورد شاخص‌های سؤال در نظریه کلاسیک تأثیر می‌گذارد؛ ولی بر برآورد پارامترهای سؤال در نظریه سؤال - پاسخ بی‌تأثیر است. به دلیل ارجحیت نظریه سؤال - پاسخ بر کلاسیک داده‌های این پژوهش با هر دو نظریه تحلیل گردید. از مهم‌ترین مفروضه‌های نظریه سؤال - پاسخ، تک‌بعدی بودن است که در اینجا از روش تحلیل عاملی اکتشافی جهت بررسی عامل غالب استفاده شد. باتوجه به مقادیر ویژه برآورد شده، نتیجه گرفتیم که عامل اول دارای ارزش ویژه بیش از ۳ برابر عامل دوم بود. نمودار اسکری نیز تأییدکننده تک‌بعدی بودن این آزمون بود؛ بنابراین، آزمون ریاضی سال ۹۱ تک‌بعدی و دارای استقلال موضعی است. از دیگر مفروضه‌های نظریه سؤال - پاسخ، بررسی برآزش مدل - داده‌هاست. تحلیل با هر سه مدل سؤال - پاسخ انجام گردید. بعد از بررسی‌های انجام شده نتیجه گرفتیم که مدل دو پارامتری برآزش بهتری با مجموعه داده‌های تحقیق دارد. به‌طور کلی می‌توان گفت که مدل‌های دو و سه پارامتری نظریه سؤال - پاسخ از مدل‌های کلاسیک و راش برای برآورد پارامترهای سؤال و توانایی، دقیق‌تر و باثبات‌تر هستند (مظفری، ۱۳۸۱). یافته‌های این پژوهش نیز با یافته‌های آبکار (۱۳۸۹)، مظفری (۱۳۸۱)، احمدی آذر (۱۳۸۷) و یونسی، دلاور و فلسفی‌نژاد (۱۳۸۹) در برآزش مدل - داده‌ها همسو است.

از مزیت‌های آزمون ریاضی سال ۹۱، عدم برآزش با مدل ۳ پارامتری سؤال - پاسخ و برآزش با مدل ۲ پارامتری است، زیرا عامل حدس در این آزمون نقشی نداشته است. از آنجایی که طراحان سؤالات آزمون‌های سرنوشت‌ساز سعی می‌کنند سؤالاتی بسازند که قابل حدس نباشد، بنابراین از این بابت، این آزمون از ویژگی‌های روان‌سنجی خوبی برخوردار بوده است. در ادامه داده‌ها با استفاده از مدل ۲ پارامتری سؤال - پاسخ مورد تجزیه و تحلیل قرار گرفت. نتایج این تحلیل‌ها نشان داد؛ ضریب تشخیص هیچ سؤالی منفی نیست. تحلیل کلاسیک آزمون نشان داد که سؤالات ۴۲-۲۶ و ۵۵ دارای ضریب تشخیص منفی بوده‌اند. همچنین، در نظریه سؤال - پاسخ سؤالات ۱-۲-۳-۶-۱۰-۱۴-۱۵-۲۶-۲۹-۳۲-۴۰-۴۲-۴۳-۴۵-۵۱-۵۴-۵۵ دارای ضریب تشخیص نامناسب بودند. سؤال ۱۷ دارای بیشترین و سؤال ۴۲ دارای کمترین ضریب تشخیص آزمون بر اساس نظریه سؤال - پاسخ بودند. سطوح دشواری سؤالات در تحلیل‌های نظریه سؤال - پاسخ مانند کلاسیک نشان داد که سؤالات آزمون ریاضی سال ۹۱، جزء سؤالات بسیار سخت محسوب می‌شوند.

حال اگر بخواهیم نتیجه‌گیری کنیم که آزمون ریاضی ۱۳۹۱ به چه صورت عمل نموده است، باید به این نکته توجه نمائیم که یا سؤالات برای آزمودنی‌ها بسیار مشکل بوده، یا نظام آموزشی ما دچار مشکل است. یعنی، احتمالاً پاسخ دادن به این سؤالات سخت بالاتر از سطوح یادگیری آزمودنی‌ها بوده است. تحلیل‌های روان‌سنجی نشان می‌دهد، برخی از سؤالات (مثل سؤال ۴۹ که تنها ۳۱ نفر از ۱۰۰۰ نفر توانسته‌اند به آن پاسخ درست بدهند) بسیار دشوار بوده و درصد بالایی از افراد توانایی پاسخ‌گویی به آن را نداشته‌اند. به‌طورکلی، می‌توان گفت که اکثر دانش‌آموزان و دانشجویان ما از درس ریاضی چندان رضایت نداشته و بیشتر درصدد پاس کردن این درس کاربردی هستند تا یادگیری آن. بدین سان نیاز اساسی نظام آموزشی ایجاد می‌کند در روش‌های تدریس ریاضی تجدید نظر و دانش‌آموزان را علاقه‌مند به یادگیری آن نمائیم. این نتیجه می‌تواند نتیجه یک شکاف بین سازمان‌سنجش به‌عنوان نهادی تاثیرگذار و آموزش و پرورش باشد؛ بنابراین، تعامل بین سازمان‌سنجش و وزارت آموزش و پرورش می‌تواند در تسریع این امر مؤثر باشد.

از طرفی اکثر ریاضی‌دانان هوش اجتماعی بالایی دارند و کوشش فراوان آنان حکایت از اعتمادبه‌نفس بالای آنان است، طوریکه فلسفه و مفاهیم انتزاعی را به‌خوبی درک می‌کنند و تفکر باعث می‌شود هوش معنوی سرشاری داشته باشند. ابداع و ابتکار آنان در حل مسائل حاکی از هوش خلاق و یادگیری دقیق و به‌دوراز خطا شاخصه هوش شناختی بالای آنان است؛ بنابراین باید در جهت ارتقای انواع هوش نه فقط هوش شناختی گام برداشت، این پیشگیری بهتر از درمان مشکلاتی نظیر عدم یادگیری، افت تحصیلی و نظایر آن است (قاسم‌زاده، ۱۳۸۶)؛ بنابراین، برای ریشه‌یابی بی‌علاقگی اکثر دانش‌آموزان به درس بسیار کاربردی ریاضی باید اقدامات اصولی و علمی و تحقیقات نظری و کاربردی انجام گردد.

از جمله محدودیت‌های این پژوهش می‌توان به عدم تعمیم نتایج این تحقیق به دروس یا گروه‌های دیگر آزمایشی اشاره کرد. همچنین، ضریب اعتبار به‌دست‌آمده در این پژوهش با استفاده از روش آلفای کرونباخ به‌دست‌آمده که نشان‌دهنده هماهنگی درونی مجموعه سؤالات آزمون ریاضی است و مفاهیمی مانند پایایی و تکرارپذیری را در بر نمی‌گیرد. تأثیر زبان و فرهنگ‌های مختلف ایران بر روی نحوه پاسخگویی و توانایی آزمودنی‌ها در این پژوهش مدنظر نبوده است. از مهم‌ترین پیشنهادها پژوهشی آینده می‌توان به انجام این تحقیق در دیگر گروه‌های آزمایشی، و بررسی کنش افتراقی سؤالات جهت مقایسه توانایی آزمودنی‌های شهرهای مختلف ایران و نحوه پاسخگویی آزمودنی‌ها بر اساس تفاوت فرهنگی و تفاوت امکانات آموزشی اشاره کرد.

منابع و مأخذ

منابع فارسی

- آلن، مری جی؛ ین، وندی ام. (۱۳۹۰). مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی)، ترجمه علی دل‌اور، چاپ پنجم، تهران، انتشارات سمت
- آبکار، کبری. (۱۳۹۱). بررسی ویژگی‌های روان‌سنجی سؤالات کنکور سراسری در رشته علوم تجربی سال ۱۳۸۹ از نظر تئوری سؤال و پاسخ (IRT)، پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی واحد تهران مرکزی
- امیرتسون، سوزان ای. پی راپس، استیون. (۱۳۸۸). نظریه‌های جدید روان‌سنجی برای روان‌شناسان، ترجمه حسن پاشا شریفی و همکاران، تهران، انتشارات رشد

- احمدی آذر، راحله. (۱۳۸۷). نمره کل سازی با استفاده از تکنیک های نظریه کلاسیک و مقایسه آن با مدل های نظریه سؤال - پاسخ در کنکور کارشناسی رشته ریاضی سال ۱۳۸۷، پایان نامه کارشناسی ارشد رشته سنجش و اندازه گیری (روان سنجی)، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- بست، جان دبلیو. (۱۳۹۰). روش های تحقیق در علوم تربیتی و رفتاری، ترجمه حسن پاشا شریفی و نرگس طالقانی، ویرایش دوم، چاپ یازدهم، تهران، انتشارات رشد
- بیکر، فرانک بی. (۱۳۸۱). پایه های اساسی تئوری سؤال - پاسخ. ترجمه حیدر علی هومن و علی عسگری، چاپ اول، تهران، انتشارات پارسا
- دلاور، علی، زهرا کار، کیانوش. (۱۳۹۲). سنجش و اندازه گیری در روانشناسی، مشاوره و علوم تربیتی، چاپ چهارم، (ویراست دوم)، تهران، نشر ارسباران.
- سیف، علی اکبر. (۱۳۹۰). اندازه گیری، سنجش و ارزشیابی آموزشی، ویرایش پنجم، چاپ سی و سوم، تهران، نشر دوران
- شریفی، حسن پاشا؛ شریفی، نسترن (۱۳۹۱). اصول روان سنجی و روان آزمایی، ویراست دوم (با اصلاحات و افزوده ها)، تهران، انتشارات رشد
- قاسم زاده دببگی، شاهرخ. (۱۳۸۶). ذهن، هوش و آموزش ریاضی. مجله ریاضیات کاربردی واحد لاهیجان، سال چهارم، شماره ۱۳، تابستان ۱۳۸۶.
- گرامی پور، مسعود و همکاران. (۱۳۹۱). مقایسه روش های تحلیل عاملی تأییدی (CFA) و نسبت در دستنمایی مبتنی بر مدل سؤال - پاسخ (IRT) در رد گیری کنش افتراقی سؤالات آزمون های سرنوشت ساز، فصلنامه اندازه گیری تربیتی شماره ۱۹، سال سوم، پاییز ۹۱.
- محمودیان، حسن. (۱۳۹۰). تنظیم بانک سؤال درس ریاضی کنکور کارشناسی رشته ریاضی سال ۸۹ بر اساس نظریه سؤال - پاسخ، پایان نامه کارشناسی ارشد رشته سنجش و اندازه گیری، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- مقدم زاده، علی. (۱۳۸۶). مقایسه مزایا و محدودیت مدل غیر پارامتریک سؤال - پاسخ با مدل پارامتریک سؤال - پاسخ، پایان نامه کارشناسی ارشد، رشته سنجش و اندازه گیری (روان سنجی)، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- مطیعی لنگرودی، سیده طیبه. (۱۳۸۱). بررسی مزایا و محدودیت های دو روش اندازه گیری کلاسیک و سؤال پاسخ از لحاظ نظری و عملی، پایان نامه کارشناسی ارشد رشته تحقیقات آموزشی، دانشگاه تهران، دانشکده روانشناسی و علوم تربیتی
- مظفری، زینب. (۱۳۸۱). مقایسه برآورد پارامترهای توانایی / سؤال بین مدل های دو و سه پارامتری نظریه سؤال - پاسخ با مدل راش نظریه سؤال - پاسخ و نظریه کلاسیک اندازه گیری، پایان نامه کارشناسی ارشد، رشته سنجش و اندازه گیری (روان سنجی)، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- مینایی، اصغر؛ فلسفی نژاد، محمدرضا. (۱۳۸۹). روش های سنجش تک بعدی بودن سؤال ها در مدل های دو ارزشی IRT، فصلنامه اندازه گیری تربیتی، شماره ۳

- **واشقانی فراهانی، مریم.** (۱۳۸۰). کاربرد روش هم‌ترازسازی همصدک در معادل‌سازی نمرات آزمون‌های ورودی دانشگاه (کنکور سال ۱۳۷۸)، پایان نامه کارشناسی ارشد روان‌سنجی، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- **همبلتون، ک. رونالد، سوامیناتان، اچ، راجرز، جین.** (۱۳۸۹). مبانی نظریه سؤال - پاسخ، ترجمه دکتر محمدرضا فلسفی‌نژاد، تهران، انتشارات دانشگاه علامه طباطبائی
- **یونسی، جلیل.** (۱۳۹۱). تحلیل داده‌های آزمون تیمز پیشرفته (۲۰۰۸): توانمندی رویکرد بیزی مدل IRT چند سطحی، پایان نامه برای دریافت درجه دکتری، رشته سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- **یونسی، جلیل.** (۱۳۸۶). بررسی ویژگی‌های روان‌سنجی سؤالات آزمون‌های فراگیر رشته روانشناسی دانشگاه پیام نور در سال ۱۳۸۵، پایان نامه کارشناسی ارشد رشته روان‌سنجی، دانشگاه علامه طباطبائی، دانشکده روانشناسی و علوم تربیتی
- **یونسی، جلیل، دلاور، علی، فلسفی‌نژاد، محمدرضا.** (۱۳۸۹). بررسی ویژگی‌های روان‌سنجی سؤالات تخصصی آزمون‌های فراگیر رشته روانشناسی دانشگاه پیام نور در سال ۱۳۸۵، فصلنامه اندازه‌گیری تربیتی، ۱۳۱-۱۶۱.

Reference

- **Andrabi, Tahir, Das, Jishnu and Khwaja, Asim Ijaz.** (2002). Test Feasibility Survey PAKISTAN: Education Sector. *Education in Pakistan*, Overview Document, October 29, 2002.
- **Bechger, T.M., Maris, G., Verstralen, H.H.F.M., & Beguin, A.A.** (2003). Using classical test theory in combination with item response theory. *Applied Psychological Measurement*, 27 (5), 319-334.
- **Carlson, James E., and Matthias von Daviar.** (2013). Item Response Theory. Educational Testing Service, Princeton, New Jersey. *ETS Research Report No. RR-13-28*.
- **Demirtaşlı, N.Ç.** (2002). A study of raven standard progressive matrices test's item measures under classical and item response models: An empirical comparison. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 35 (2), 71-79.
- **Fan, X.** (1998). Item response theory and classical test theory: An empirical comparison of their item/person parameters. *Educational and Psychological Measurement*, 58, 357-381.
- **Hambleton, R. K., & Jones, R. W.** (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 253-262.
- **Hambleton, R. K., & Swaminathan, H.** (1985). *Item Response Theory: Principles and Applications*. USA: Kluwer-Nijhoff Publishing.
- **Neşe Güler, Gülşen Kaya Uyanic, Gülşen Taşdelen Teker** (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of research on Education*, 2(1), 1-6.
- **Nichols SL & Berliner. D.C** (2005). The inevitable corruption of indicators & educators through high-stakes testing. East Landing, MI: *the great lakes center for educational research practice*.