

مقایسه مدل‌های تک پارامتری و سه پارامتری تئوری سؤال پاسخ در سنجش توانایی دانش آموزان دبیرستان دخترانه توسط ماتریس‌های پیشرونده ریون

تاریخ پذیرش مقاله: ۹۰/۱۱/۲۷

تاریخ دریافت مقاله: ۹۰/۱۱/۵

کبری اسدی*

دکترحیدر علی هومن**

دکتر ریللیاقت***

چکیده

پژوهش حاضر به مقایسه تفاوت‌های بین برآورد پارامتر توانایی آزمودنی و برآورد پارامتر توانایی سؤال در مقیاس هوش پیشرونده ریون در دو مدل راش و سه پارامتری بر اساس تئوری سؤال پاسخ می‌پردازد. از میان همه دانش‌آموزان دختر پایه اول تا سوم دبیرستان‌های منطقه ۴ آموزش و پرورش شهر تهران که در سال تحصیلی ۱۳۸۶-۱۳۸۷ مشغول به تحصیل بوده‌اند تعداد ۴۹۸ دانش‌آموز به شیوه نمونه‌برداری تصادفی چند مرحله‌ای انتخاب و آزمون مذکور بر روی آنها اجرا شد. نتایج نشان می‌دهد ضریب اعتبار آزمون با استفاده از ضریب آلفای کرونباخ برابر ۰/۸۳۹ است. روایی آزمون با استفاده از روش تحلیل مولفه‌های اصلی، تعداد ۵ عامل هوش کلی، درک و استعداد ریاضی، حافظه، تجسم فضایی، و سرعت به دست آمده است. از اجرای مدل تک پارامتری و سه پارامتری بر روی آزمون و مقایسه این دو مدل این نتایج حاصل می‌شود (۱) بین برآوردهای توانایی براساس دو مدل تک پارامتری و سه پارامتری تفاوت معنادار وجود دارد اما بین نمره‌های توانایی‌ها در هر دو مدل همبستگی وجود دارد. (۲) از روی نمره‌های خام افراد می‌توان توانایی آنان را در مدل سه پارامتری پیش بینی کرد. (۳) از مقایسه نمره میانگین دو مدل این نتیجه حاصل می‌شود که مدل تک پارامتری نسبت به مدل سه پارامتری مقاوم‌تر است.

واژه‌های کلیدی: پارامتر، توانایی، اعتبار، روایی، مدل راش، مدل سه پارامتری.

Email: Asadi.Kobra@yahoo.com

Email:

info@parsianclinic.comEmail:

*کارشناس ارشد رشته سنجش و اندازه‌گیری دانشگاه آزاد اسلامی واحد تهران مرکز

**دانشیار دانشگاه آزاد اسلامی واحد تهران مرکز

Haidarali70@yahoo.com

***استاد یار دانشگاه آزاد اسلامی واحد تهران مرکز

امروزه، گسترش آزمون‌ها و روان‌سنجی موجب کاربرد روزافزون آنها شده است و در روان‌شناسی کاربردی جایگاه مهمی یافته است. اگرچه کاربردهای اختصاصی و رهنمودهای قانونی استفاده از آزمون‌ها تغییر یافته اند، ولی آزمون‌های روان‌شناختی نسبتاً ثابت مانده اند. بسیاری از آزمون‌های معروف، هرچند تا اندازه‌ای مورد تجدید نظر قرار گرفته‌اند، ولی همچنان در عمل به کار بسته می‌شوند. علاوه بر این، برخی از آزمون‌های جدید تدوین شده‌اند، به هر حال اصول بنیادین آزمون‌سازی تغییر نکرده است. اما تغییرات بنیادی زیادی در تئوری‌های روان‌سنجی به وجود آمده‌اند (امبرستون و رایس^۲، ۱۳۸۸). اگرچه تئوری کلاسیک آزمون^۳ چندین دهه به جامعه‌سنجش خدمت کرده است ولی تئوری سؤال پاسخ^۴ پیشرفت چشمگیری در پایه‌های نظری اندازه‌گیری داشته است (همبلتون و جونز^۵، ۱۹۹۳).

تئوری کلاسیک آزمون، الگوی ساده و کاملاً مفیدی است که می‌تواند نحوه تأثیر خطاهای اندازه‌گیری را بر نمره‌های مشاهده شده توصیف کند (آلن وین^۶، ۱۳۸۴). هدف اصلی تئوری کلاسیک آزمون درک و بهبود اعتبار آزمون‌های روانی است. با آنکه ساختار و معادله‌های تئوری کلاسیک آزمون بیانگر آن است که مفروضه‌های آن در سطح نمره انفرادی آزمودنی تدوین شده‌اند، اما هرگز در مدل‌های آن نمره‌های انفرادی تحلیل نمی‌شوند بلکه تمرکز اصلی بیشتر بر ویژگی‌های نمره‌های آزمون در ارتباط گروه نمونه است. به گونه کلی، مفروضه‌ها و اندازه‌هایی که بر پایه تئوری کلاسیک اندازه‌گیری به دست می‌آیند از محدودیتها و مشکلات متعددی برخورداراند (عسگری، ۱۳۸۶).

در نقطه مقابل تئوری کلاسیک، تئوری سؤال پاسخ یک روش پیشرفته و پیچیده تری برای تحلیل آزمون است. این تئوری بر الگوی صفت مکنون استوار است. در این تئوری فرض بر این است که یک صفت یا ویژگی زیربنایی وجود دارد که به شخص امکان می‌دهد تا در یک تکلیف شناختی معین موفقیت کسب نماید (سیف، ۱۳۸۶). در این تئوری، هدف اصلی برآورد سطح دشواری هر سؤال آزمون و برآورد همزمان سطح توانایی هر یک از امتحان‌شوندگان است (کوپر^۷، ۱۳۷۹). در تئوری سؤال پاسخ (IRT) برای تحلیل داده‌های به دست آمده از مقیاسها، پرسشنامه‌ها و آزمون‌های روانی از مدل‌های ریاضی استفاده می‌شود. به مدل‌های تئوری سؤال پاسخ اغلب مدل‌های خصیصه مکنون^۸ نیز گفته می‌شود. اصطلاح مکنون برای تأکید بر پاسخهای پیوسته سؤالی به کار می‌رود که در واقع نشانگر مشاهده پذیر خصائص یا رگه‌هایی است که فرض می‌شود وجود دارد و باید از پاسخها استخراج شود (عسگری، ۱۳۸۶). این مدلها از طریق تولید یک تابع (معادله) ریاضی، رابطه احتمال پاسخ درست با توانایی را تبیین می‌کند (بیکر، ۱۳۸۱). مدل‌های تئوری سؤال پاسخ همانند تئوری کلاسیک آزمون دارای مفروضاتی هستند که در صورت برقراری آنها نتایج معتبر و دقیق تری به دست خواهد آمد. این مدلها دارای سه مفروضه مهم و اساسی

^۲. Emberston & Reise

^۳. classical test theory (CTT)

^۴. item Response theory

^۵. Hambleton & Jones

^۶. Allen & Yen.

^۷. Cooper

^۸. latent attribute

هستند که عبارتند از: ۱) نخستین مفروضه IRT اشاره به این مطلب دارد که یک عامل مشترک، همه کوورایانسهای سؤال را توجیه می کند. این عامل مشترک در واقع خصیصه مکنون مورد نظر است که در روان سنجی معمولاً به دو گونه بیان می شود: الف) تک بعدی بودن: یعنی فقط یک عامل مشترک یا خصیصه مکنون وجود دارد که همبستگی - های درونی سؤال را توجیه می کند (عسگری، ۱۳۸۶، ب) استقلال مکانی: استقلال مکانی بدین معناست که هر سؤال باید مسئله کاملاً تازه ای باشد استقلال مکانی با تناسب مدل سؤال برای داده ها ارتباط دارد (کوپر، ۱۳۷۹). ۲) مفروضه دوم پیلنجر این مطلب است که بین نمره تست و خصیصه مکنون رابطه ای تکنوا^۹ وجود دارد. یعنی اینکه با بالا رفتن سطح خصیصه در نمره مورد انتظار افزایش ثابتی بوجود می آید، و جایگاه شخص در خصیصه مکنون مورد نظر نیز با بالا رفتن نمره تست افزایش می یابد (ثرندایک، ۱۳۷۵). ۳) رابطه بین خصیصه مکنون و پاسخ مشاهده شده شکل معینی دارد. خطی که خصیصه مکنون و احتمال پاسخ را به یکدیگر مربوط می سازد خم ویژه سؤال نام دارد. مدل استاندارد ریاضی که برای خم ویژه سؤال به کار می رود شکل تراکمی تابع منطقی است. این تابع، خانواده ای از خمها را تعریف می کند که به سبب سادگی کار با آن بر سایر مدلها ترجیح داده شد (بیکر، ۱۳۸۱).

همانطور که گفته شد خم ویژه سؤال، تابعی است که احتمال موفقیت در سؤال را به جایگاه آزمودنی در بعد صفت بنیادی مربوط می سازد (هومن، ۱۳۸۶). خم ویژه سؤال پایه اساسی وزیر بنایی تئوری سؤال پلسخ است و دیگرسازه ها و پارامترهای این تئوری بستگی به آن دارد. خم ویژه سؤال دارای سه پارامتر است که برای توصیف آن به کار می رود که عبارتند از: پارامتر دشواری سؤال: این پارامتر معرف دشواری و سطح زیر بنایی θ است و با B نشان داده می شود. به بیان دیگر پارامتر دشواری سؤال، موقعیت مکانی سؤال را در مقیاس توانایی توصیف می کند. پارامتر دشواری سؤال (B) نقطه عطف^{۱۰} خم ویژه سؤال را توصیف می کند (هومن، ۱۳۸۶).

پارامتر قدرت تشخیص سؤال: این پارامتر معرف شیب خم ویژه سؤال است و با a نمایش داده می شود. این پارامتر نشان می دهد سؤال مورد نظر تا چه اندازه می تواند بین امتحان شوندگانی که توانایی آنها پایین تر از جایگاه سؤال است و با امتحان شوندگانی که توانایی آنها بالاتر از این جایگاه قرار دارد، تمایز ایجاد کند (عسگری، ۱۳۸۶). **پارامتر حدس^{۱۱}:** این پارامتر موفقیت در حد شانس است. (احتمال انتخاب کاملاً تصادفی گزینه درست سؤال) و با C نشان داده می شود. این پارامتر بیانگر سطح پایه خم برای سطح بسیار پایین خصیصه مکنون است. این سطح در واقع همان مجانب است که وقتی θ ، یعنی سطح توانایی فوق العاده پایین باشد به خم ویژه سؤال نزدیک است. بیشتر سؤالهای چند گزینه ای دارای پارامترهای C بزرگتر از $0/0$ است. (مقدار C برای سؤالهای ۴ گزینه ای در حدود $0/25$ و برای سؤالهای ۵ گزینه ای در حدود $0/2$ است) (هومن، ۱۳۸۶).

پارامتر آزمودنی: یعنی هر آزمودنی دارای یک نمره توانایی (مقدار پارامتر) است که جایگاه وی را در مقیاس تعیین می کند. یعنی پارامتر آزمودنی بیانگر جایگاه فرد در مقیاس توانایی است. در دنیای واقعی نمی توان مقدار پارامتر توانایی امتحان شونده (آزمودنی) را به دست آورد و در بهترین حالت تنها می توان آن را برآورد کرد (بیکر،

^۹. monotonic

^{۱۰}. Reflection point

^{۱۱}. Guessing parameter

۱۳۸۱). برآورد پارامتر آزمودنی بر اساس نمره کل فرد در انواهی گیری به دست می آید. این نمره در واقع نمره وزن دار فرد بر پایه مدلی است که شامل پارامترهای تشخیص باشد (عسگری، ۱۳۸۶).

تئوری سؤال پاسخ به عنوان یک روش مدل سازی نیرومند شناخته می شود که دارای پیش فرضهای قوی است (امبرستون و رایس، ۱۳۸۸). که این مفروضه ها قبلاً توضیح داده شد. مدل های تئوری سؤال پاسخ برای ارائه و نمایش بهتر خم و ویژه سؤال به کار می رود. اینگونه مدل ها در حقیقت برای توسعه و پیشرفت دقیق یک تئوری اندازه گیری و انتقال اطلاعات درباره ویژگی های فنی سؤالها ابزار مفیدی به شمار می آیند (بیکر، ۱۳۸۱). مدل های متعدد دی در این تئوری، با توجه به نوع آزمونها و کاربردشان ونحوه نمره گذاری آزمونها وجود دارند. معروفترین مدل های سؤال پاسخ عبارتند از:

مدل تک پارامتری یا مدل منطقی راش : یکی از عملی ترین و رایج ترین مدل های نظریه سؤال پاسخ، مدل راش است. مدل های راش در واقع روی آوردی ریاضی برای آزمون این فرضیه است که اندازه های مربوط به معنا و واحد یک سازه را می توان از ابزاری که برای آن خصیصه تهیه شده است به دست آورد. وقتی داده ها با این مدلها برازش پیدا می کنند به معنای آن است که ابزار اندازه گیری و اندازه ها در یک واحد فاصله ای مشترک مقیاس بندی شده اند و می توانند در انواع یا شکل های مختلف آن ابزار و نیز در بین نمونه های مختلف یک جامعه ثابت باقی بمانند (عسگری، ۱۳۸۷). لرد و ناولیک^{۱۲} (۱۹۶۸) و راش^{۱۳} (۱۹۶۰) اعتقاد دارند اگر قدرت تشخیص سؤالها از یکدیگر تفاوت داشته باشند اصولاً کاربرد مدل تک پارامتری راش از لحاظ فنی نادرست خواهد بود. در این مدل فقط پارامتر دشواری (b) مورد توجه قرار می گیرد و دارای دو مفروضه اساسی است: ۱) همه سؤالها از درجه تشخیص (a) ثابت و یکسانی برخوردارند. ۲) پارامتر حدس (c) معادل صفر می باشد. این مدل کلیه مزایای مدل های دو وسه پارامتری را دارد، و چنانچه داده ها با مدل برازش داشته باشد به معنی آن است که ابزار اندازه گیری و اندازه ها در یک واحد فاصله مشترک مقیاس بندی شده اند و می توانند در انواع یا شکل های مختلف آن ابزار و نیز در بین نمونه های مختلف یک جامعه ثابت باقی بمانند (ستاری، ۱۳۸۲). بنا بر نظر آندریش^{۱۴}

أ) (۲۰۰۴) دیدگاه مدل های راش به گونه بارزی با سایر مدل های اندازه گیری تفاوت دارد. در اغلب مدلها هدف اصلی توصیف مجموعه ای از داده هاست. اما هدف از به کار بردن مدل راش به دست آوردن داده هایی است که با مدل برازش داشته باشد. هومن (۱۳۷۳) در پژوهشی به مقایسه مدل تک پارامتری راش و دو پارامتری پرداخته است. نتیجه حاکی از آنست که گرچه تاثیر پارامتر شیب سؤال در برآورد پارامتر دشواری سؤال غیرقابل انکار است. ولی تاثیر آن در برآورد پارامتر توانایی افراد آن گونه نیست که مدل های تک پارامتری و دو پارامتری را از هم متمایز سازد.

مدل منطقی (لوجستیک) دو پارامتری: این مدل برای مقیاس هایی که سؤالهای آنها رابطه یکسانی با صفت مکنون ندارند مناسب است یا به بیان دیگر سؤالها به طور یکسان جایگاه شخص را روی پیوستار صفت مکنون نشان نمی دهند (برای آزمونهای تشریحی یا باز پاسخ مناسب است). (امبرستون و رایس، ۱۳۸۸). در این مدل علاوه بر پارامتر توانایی، پارامتر قدرت تشخیص و دشواری سؤال دخالت دارد. بر خلاف مدل سه پارامتری عامل حدس وجود ندارد (ستاری، ۱۳۸۲). همچنین سطح دشواری سؤال های مدل دو پارامتری تا اندازه ای با سطح دشواری سؤالهای مدل راش

^{۱۲}. Lord & Novick

^{۱۳}. Rash

^{۱۴}. Andrich

متفاوت است و این از تفاوت قدرت تشخیص سؤاها ناشی می‌شود (امبرستون و رایس، ۱۳۸۸). در مطالعه‌های اندازه-گیری آندریخ (۱۹۸۸) متوجه شد که مدل دوپارامتری تئوری سؤال پاسخ سعی دارد که این اختلاف در تشخیص ICCS را بگیرد. اشکالهای موجود در مدل این امکان را دارد که تغییرناپذیری برآورد پارامتری مشخص و سؤال را روشن کند و بنابراین در آینده قانع کننده نخواهد بود. هم‌بتون و تراپ^{۱۵} (۱۹۷۳) میزان انطباق الگوی منطقی دوپارامتری را با میزان انطباق الگوی منطقی راش برای انتخاب سؤال در قسمتهای کلامی و ریاضی آزمون استعداد تحصیلی اونتاریو^{۱۶} و آزمون کلامی استعداد تحصیلی مقایسه کردند. برای هر یک از آزمونها، الگوی دوپارامتری، از نظر انطباق با داده‌ها، بهتر از الگوی راش بود و تفاوت بین الگوها با افزایش تعداد سؤالهای آزمون کاهش یافت.

مدل سه پارامتری مدل سه پارامتری دارای یک پارامتر دقیق‌تر از مدل تک و دو پارامتری است که سبب می‌شود انتهای خم ویژه سؤال روی نقطه صفر قرار نگیرد، پارامتر حدس (C) نام دارد (امبرستون و رایس، ۱۳۸۸). پارامتر حدس بیانگر احتمال به دست آوردن پاسخ درست سؤال از طریق حدس محض است. مقدار C به عنوان تابعی از سطح توانایی تغییر نمی‌کند. از این رو احتمال به دست آوردن پاسخ درست از طریق حدس بر آزمودنی‌هایی که دارای بالاترین و پایین‌ترین سطح توانایی باشند، یکسان است (یکسان استیپیکر، ۱۳۸۱). یوری (۱۹۷۱) (نقل از ستاری، ۱۳۸۴) مدلهای تک، دو و سه پارامتری را با هم مقایسه کرد نتایج نشان داد که مدل سه پارامتری بهترین توصیف را آزمونهای چند گزینگی با تئای واقعی آزمون به دست می‌دهد. این مدل بیشتر در مورد سؤالهای چند گزینگی کاربرد دارد زیرا عامل حدس در این گونه سؤالها دخیل است و در برآورد نمره واقعی مؤثر است (فراهانی، ۱۳۷۵) در پژوهشی به مقایسه مدلهای انولگیری (کلاسیک و سؤال - پاسخ) از لحاظ برآورد پارامترهای سؤال و توانایی پرداخته است نتایج نشان داد که مدلهای اندازه‌گیری سؤال پاسخ از مدل کلاسیک آزمون برای برآورد پارامترهای سؤال و توانایی مناسبتی دقیق‌تر هستند، و حتی در بین مدلهای تئوری سؤال پاسخ، مدل سه پارامتری برتری خاصی نسبت به دو مدل دیگر در برآورد پارامترها دارد

پژوهش حاضر به مقایسه تفاوت‌های بین برآورد پارامتر توانایی آزمودنی و برآورد پارامتر توانایی سؤال در مقیاس هوش پیشرونده ریون در دو مدل راش و سه پارامتری بر اساس تئوری سؤال پاسخ می‌پردازد و درصدد پاسخگویی به این پرسشهاست: (۱) آیا بین سؤالهای آزمون هوش پیشرونده ریون هماهنگی وجود دارد؟ (۲) آیا آزمون از اعتبار کافی برخوردار است؟ (۳) آیا بین برآورد پارامترهای توانایی آزمودنی در آزمون هوش پیشرونده ریون برپایه مدل راش و مدل سه پارامتری تفاوت اساسی وجود دارد؟

روش

جامعه آماری مورد مطالعه را همه دانش آموزان دختر پایه اول تا سوم دبیرستان دخترانه منطقه آموزش و پرورش شهر تهران که در سال تحصیلی ۱۳۸۷-۱۳۸۶ مشغول به تحصیل بودند؛ تشکیل می‌دهد. نمونه‌ای به حجم ۴۹۸ دانش آموز دختر با روش نمونه‌برداری چند مرحله‌ای انتخاب و در نهایت آزمون هوش پیشرونده ریون بر روی آنها به گونه یکسان اجرا گردید این آزمون شامل ۶۰ سؤال (ماتریس) است. سوالها شامل توالی الگوهای اشکال یا نمودارهایی است که بر اساس منطق خاص تنظیم شده‌اند. آزمودنی باید هر یک از اشکال یا نمودارها را با انتخاب یک شکل از بین چند شکل ارائه شده در پائین صفحه کامل کند یافتن پاسخ هر سؤال مستلزم آن است که آزمودنی رابطه بین اجزای

^{۱۵}. Hembleton and Traub

^{۱۶}. Ontario Scholastic Aptitude test

نمودارها را کشف کند سپس قاعده‌ی را که کشف کرده است برای یافتن جزء ناقص به کار برد. آزمون هم به صورت گروهی و هم به صورت فردی قابل اجرا است (شریفی، ۱۳۸۳). این مقیاس از ویژگی‌های روانسنجی (اعتبار و روایی) مناسبی برخوردار است. در این پژوهش، تجزیه و تحلیل داده‌ها بر اساس هدف پژوهش به ترتیب زیر انجام گرفته است.

(۱) برای تعیین هماهنگی درونی و میزان اعتبار پرسشنامه از ضریب آلفای کرونباخ استفاده گردید (۲). برای بررسی روایی سازه آزمون، روش تحلیل مؤلفه‌های اصلی (PC) اجرا شد. (۳) رابطه احتمالی بین پاسخهای مشاهده شده آزمودنی به آزمون هوش پیشرونده ریون و سطح توانایی آزمودنی‌ها در صفت مکنون برپایه مدل‌های IRT (تک و سه پارامتری) مشخص گردید. (۴) مقایسه دو مدل تک و سه پارامتری از لحاظ برآورد پارامتر توانایی و دشواری سؤال از طریق اجرای آزمون t.

یافته‌ها

تجزیه و تحلیل اطلاعات و نتایج به دست آمده بر روی ۴۹۸ دانش‌آموز دختر پایه اول تا سوم دبیرستان در فاصله سنی ۱۴ تا ۱۷ سال در چهار بخش کلی ارائه می‌شود.

۱. بررسی اعتبار آزمون: برای برآورد ضریب اعتبار آزمون هوش پیشرونده ریون از فرمول کلی ضریب آلفای کرونباخ استفاده شده و نتایج آن برای کل مجموعه ۶۰ سؤالی در جدول ۱ ارائه شده است.

جدول ۱: میانگین، انحراف استاندارد، ضریب همبستگی هر سؤال با نمره کل و ضریب آلفا در صورت حذف هر سؤال در مجموعه ۶۰ سؤالی آزمون هوش پیشرونده ریون

سؤال	میانگین سؤال	انحراف استاندارد	همبستگی با کل تست	آلفا در صورت حذف سؤال	سؤال	میانگین سؤال	انحراف استاندارد	همبستگی با کل تست	آلفا در صورت حذف سؤال
۱	۰/۹۸	۰/۱۳	۰/۸۳۹	۰/۰۹۸	۳۱	۰/۸۴	۰/۳۷	۰/۸۳۸	۰/۲۱۷
۲	۰/۹۸	۰/۱۵	۰/۸۳۹	۰/۱۰۶	۳۲	۰/۸۶	۰/۳۵	۰/۸۳۶	۰/۳۰۶
۳	۰/۹۹	۰/۱۲	۰/۸۳۹	۰/۱۱۶	۳۳	۰/۸۷	۰/۳۳	۰/۸۳۸	۰/۲۲۹
۴	۰/۹۹	۰/۱۰	۰/۸۳۹	۰/۰۵۹	۳۴	۰/۷۹	۰/۴۱	۰/۸۳۴	۰/۳۸۰
۵	۰/۹۸	۰/۱۵	۰/۸۳۹	۰/۱۲۸	۳۵	۰/۷۷	۰/۴۲	۰/۸۳۳	۰/۴۱۱
۶	۰/۹۹	۰/۱۱	۰/۸۳۹	۰/۱۲۲	۳۶	۰/۵۸	۰/۴۹	۰/۸۳۵	۰/۳۳۷
۷	۰/۹۷	۰/۱۸	۰/۸۳۶	۰/۳۹۵	۳۷	۰/۹۳	۰/۲۵	۰/۸۳۸	۰/۱۷۱
۸	۰/۹۱	۰/۲۸	۰/۸۳۸	۰/۲۲۸	۳۸	۰/۹۱	۰/۲۹	۰/۸۳۸	۰/۱۷۴
۹	۰/۹۸	۰/۱۵	۰/۸۳۷	۰/۳۱۰	۳۹	۰/۹۱	۰/۲۹	۰/۸۴۰	۰/۱۰۶
۱۰	۰/۹۵	۰/۲۱	۰/۸۳۷	۰/۲۷۴	۴۰	۰/۸۹	۰/۳۲	۰/۸۳۷	۰/۲۴۴
۱۱	۰/۹۲	۰/۲۷	۰/۸۳۷	۰/۲۶۵	۴۱	۰/۸۲	۰/۳۸	۰/۸۳۸	۰/۲۰۷
۱۲	۰/۷۷	۰/۴۲	۰/۸۳۷	۰/۲۷۲	۴۲	۰/۹۴	۰/۲۴	۰/۸۳۷	۰/۲۸۰
۱۳	۰/۹۸	۰/۱۵	۰/۸۴۰	۰/۰۵۸	۴۳	۰/۹۱	۰/۲۹	۰/۸۳۵	۰/۳۵۶
۱۴	۰/۹۸	۰/۱۵	۰/۸۴۰	۰/۰۵۱	۴۴	۰/۹۳	۰/۲۵	۰/۸۳۷	۰/۲۷۳
۱۵	۰/۹۷	۰/۱۶	۰/۸۳۸	۰/۲۲۵	۴۵	۰/۸۷	۰/۳۴	۰/۸۳۴	۰/۴۰۸
۱۶	۰/۹۵	۰/۲۱	۰/۸۳۷	۰/۲۵۶	۴۶	۰/۹۱	۰/۲۸	۰/۸۳۷	۰/۲۶۹

۱۷	۰/۸۴	۰/۳۶	۰/۸۴۰	۰/۱۰۱	۴۷	۰/۷۳	۰/۴۵	۰/۸۳۸	۰/۲۴۹
۱۸	۰/۹۵	۰/۲۳	۰/۸۳۸	۰/۲۰۴	۴۸	۰/۵۱	۰/۵۰	۰/۸۳۹	۰/۲۲۷
۱۹	۰/۹۱	۰/۲۹	۰/۸۳۶	۰/۲۹۲	۴۹	۰/۸۸	۰/۳۲	۰/۸۳۸	۰/۲۱۸
۲۰	۰/۹۳	۰/۲۶	۰/۸۳۷	۰/۲۷۶	۵۰	۰/۸۲	۰/۳۸	۰/۸۳۵	۰/۳۴۳
۲۱	۰/۹۲	۰/۲۷	۰/۸۳۷	۰/۲۸۴	۵۱	۰/۷۲	۰/۴۵	۰/۸۳۷	۰/۲۷۲
۲۲	۰/۹۸	۰/۱۵	۰/۸۳۷	۰/۳۱۰	۵۲	۰/۸۵	۰/۳۶	۰/۸۳۲	۰/۴۹۷
۲۳	۰/۸۹	۰/۳۱	۰/۸۳۸	۰/۲۱۲	۵۳	۰/۸۱	۰/۳۹	۰/۸۳۲	۰/۴۶۵
۲۴	۰/۸۹	۰/۳۱	۰/۸۳۳	۰/۴۵۸	۴	۰/۸۰	۰/۴۰	۰/۸۳۲	۰/۴۵۵
۲۵	۰/۸۴	۰/۳۷	۰/۸۴۱	۰/۰۹۰	۵۵	۰/۶۸	۰/۴۷	۰/۸۳۴	۰/۳۶۹
۲۶	۰/۹۶	۰/۱۹	۰/۸۳۷	۰/۳۰۸	۵۶	۰/۶۸	۰/۴۷	۰/۸۳۳	۰/۴۱۱
۲۷	۰/۹۶	۰/۱۹	۰/۸۳۷	۰/۲۷۰	۵۷	۰/۶۲	۰/۴۹	۰/۸۳۴	۰/۳۷۰
۲۸	۰/۸۳	۰/۳۷	۰/۸۳۹	۰/۱۶۲	۵۸	۰/۵۳	۰/۵۰	۰/۸۳۵	۰/۳۳۸
۲۹	۰/۹۲	۰/۲۷	۰/۸۳۶	۰/۳۱۲	۵۹	۰/۳۷	۰/۴۸	۰/۸۳۹	۰/۱۹۹
۳۰	۰/۹۰	۰/۳۷	۰/۸۳۵	۰/۳۷۷	۶۰	۰/۳۷	۰/۴۸	۰/۸۳۹	۰/۲۱۳

در جدول ۱، ملاحظه می‌شود که سؤالهای ۳، ۴ و ۶ دارای بالاترین سطح دشواری (میانگین) و سؤالهای ۵۹ و ۶۰ دارای کمترین سطح دشواری است. سؤالهای ۳۶، ۴۸ و ۵۸ دارای بیشترین و سؤالهای ۴ و ۶ دارای کمترین مقدار انحراف استاندارد هستند. تمامی سؤالها همبستگی مثبت با نمره کل آزمون دارند. سؤالهای ۵۳ و ۵۴ با

ضریب ۰/۸۳۲ کمترین همبستگی و سؤال ۲۵ با ضریب همبستگی ۰/۸۴۱ بیشترین مقدار همبستگی با کل مقیاس را داراست. اعتبار کل مقیاس به دست آمده برای این مقیاس با استفاده از روش آماری آلفای کرونباخ برابر ۰/۸۳۹ است. باتوجه به این که حذف سؤالها، تأثیر اندکی (افزایش ۰/۰۰۱ سطح آلفا) بر روی اعتبار آزمون دارد (که بسیار نامحسوس است) و نیز جهت بررسی ساختار عاملی، از حذف این سؤالها صرف نظر شده است.

۲. بررسی روایی سازه آزمون: در نتیجه تحلیل عاملی مواد آزمون که با روش تحلیل مؤلفه های اصلی (PC)

بدست آمده است، ۵ عامل استخراج شده، که فرایند عاملیابی را توجیه می‌کند. ارزشهای ویژه ۵ عامل معناداری که استخراج شده، بیشترین مقدار ارزش ویژه برابر با ۶/۶۰۲ مربوط به عامل یکم و کمترین مقدار آن برابر با ۱/۹۱۲ مربوط به عامل پنجم است. عامل یکم به تنهایی ۱۱/۰ درصد از واریانس و ۵ عامل روی هم ۳۰/۰ درصد از کل واریانس متغیرها را تبیین می‌کند. برپایه نتایج چرخش واریانس و ماتریس ساختار عاملی موجود از این تحلیل به شرح زیر و به ترتیب بیشترین بار عاملی استخراج و نامگذاری گردید و این نتایج استنباط می‌شود: (۱) تعداد ۴ پرسش دارای پیچیدگی و بار آنها متمرکز بر عامل ۲ است. (۲) سؤالهای دیگر ناب و فاقد پیچیدگی است. (۳) حداکثر ضریب در ماتریس ساختار عاملی (همبستگی هر سؤال با عامل) ۰/۷۸۶، مربوط به سؤال ۲ و کمترین ضریب برابر ۰/۳۰۱، مربوط به سؤال ۴۹ است. (۴) عامل یکم با ۳۵ سؤال همبستگی قوی دارد، که بیانگر هوش کلی (g) است. (۵) عامل دوم با ۲۰ سؤال همبستگی قوی دارد که بیانگر درک استعداد ریاضی است. (۶) عامل سوم با ۷ سؤال همبستگی قوی دارد، که بیانگر حافظه است. (۷) عامل چهارم با ۸ سؤال همبستگی قوی دارد که بیانگر تجسم فضایی است. (۸) عامل پنجم با ۵ سؤال همبستگی قوی دارد، که بیانگر سرعت است.

۳. اجرای مدل تک پارامتری و سه پارامتری: تحلیل داده ها به کمک مدل تک پارامتری راش و آزمون برازندگی تک تک سؤالها و مجموع آزمون در جدول ۲ ملاحظه می‌شود.

جدول ۲: آزمون برازش سؤالها بر اساس مدل راش

سؤال	دشواری	خطای استاندارد (b)	مجذور کای	نمره مقیاس استاندارد	سؤال	دشواری	خطای استاندارد (b)	مجذور کای	نمره مقیاس استاندارد
۱	-۴/۸۰۷	۰/۳۶۸	۱۴/۳۹۸	۲۸	۳۱	-۱/۹۳۷	۰/۱۲۹	۱۸/۴۶۶	۷۱
۲	-۴/۴۴۳	۰/۳۱۶	۱۳/۷۲۴	۳۳	۳۲	-۲/۱۶	۰/۱۳۷	۵/۴۶۷	۶۸
۳	-۴/۹۵۸	۰/۳۹۲	۲۵/۰۰۳	۲۶	۳۳	-۲/۳۰۴	۰/۱۴۳	۲۶/۱۵۱	۶۵
۴	-۵/۳۳۶	۰/۴۶۱	۱۰/۸۳۲	۲۰	۳۴	-۱/۶	۰/۱۱۹	۱۹/۵۳	۷۶
۵	-۴/۴۴۳	۰/۳۱۶	۹/۱۹	۳۳	۳۵	-۱/۴۱۵	۰/۱۱۴	۲۹/۱۱۷	۷۹
۶	-۵/۱۳۲	۰/۴۲۳	۸/۲۰۹	۲۳	۳۶	-۰/۳۴۴	۰/۰۹۹	۱۳/۱۳۴	۹۵
۷	-۳/۹۳۶	۰/۲۵۷	۹/۱۲۲	۴۱	۳۷	-۳/۱۳۸	۰/۱۸۹	۱۹/۱۳۹	۵۳
۸	-۲/۸۰۵	۰/۱۶۸	۱۵/۹۲۹	۵۸	۳۸	-۲/۷۱۹	۰/۱۶۳	۱۹/۹۹۴	۵۹
۹	-۴/۳۴۳	۰/۳۰۳	۷/۳۳۲	۳۵	۳۹	-۲/۷۴۷	۰/۱۶۵	۴۰/۸۳۹	۵۹
۱۰	-۳/۵۷۷	۰/۲۲۳	۱۳/۶۹۴	۴۶	۴۰	-۲/۴۳۸	۰/۱۴۹	۲۱/۷۹۳	۶۳
۱۱	-۲/۸۹۷	۰/۱۷۴	۱۳/۸۵۴	۵۷	۴۱	-۱/۸۱۹	۰/۱۲۵	۲۸/۳۶۷	۷۳
۱۲	-۱/۴۴۳	۰/۱۱۵	۳۵/۶۶۴	۷۸	۴۲	-۳/۲۱۵	۰/۱۹۵	۹/۵۷۱	۵۲
۱۳	-۴/۳۴۳	۰/۳۰۳	۲۲/۷۹۷	۳۵	۴۳	-۲/۶۹۱	۰/۱۶۲	۱۵/۶۳۷	۶۰
۱۴	-۴/۳۴۳	۰/۳۰۳	۲۲/۲۸۴	۳۵	۴۴	-۳/۱۰۱	۰/۱۸۷	۹/۹۲۹	۵۳
۱۵	-۴/۲۵	۰/۲۹۲	۷/۵۴۱	۳۶	۴۵	-۲/۲۶۲	۰/۱۴۱	۲۱/۶۴۲	۶۶
۱۶	-۳/۵۷۷	۰/۲۲۳	۵/۲۱۳	۴۶	۴۶	-۲/۷۷۶	۰/۱۶۷	۸/۳۰۸	۵۸
۱۷	-۲/۰۰۸	۰/۱۳۲	۳۹/۷۳۳	۷۰	۴۷	-۱/۱۵۵	۰/۱۰۹	۲۲/۳۹۵	۸۳
۱۸	-۳/۳۸۴	۰/۲۰۷	۱۸/۰۵	۴۹	۴۸	-۰/۰۳۲	۰/۰۹۸	۴۹/۳۲۳	۱۰۰
۱۹	-۲/۶۹۱	۰/۱۶۲	۳۰/۲۵۱	۶۰	۴۹	-۲/۳۹۲	۰/۱۴۷	۲۴/۷۷۸	۶۴
۲۰	-۳/۰۲۹	۰/۱۸۲	۸/۲۴۷	۵۵	۵۰	-۱/۸۳۵	۰/۱۲۶	۳۵/۸۶۹	۷۲
۲۱	-۲/۹۶۲	۰/۱۷۸	۲۵/۴۲۴	۵۶	۵۱	-۱/۱۱۸	۰/۱۰۸	۶/۹۷۳	۸۳
۲۲	-۴/۳۴۳	۰/۳۰۳	۵/۲۳۵	۳۵	۵۲	-۲/۰۴۵	۰/۱۳۳	۳۱/۳۹۶	۶۹
۲۳	-۲/۵۳۵	۰/۱۵۴	۲۷/۳۴۶	۶۲	۵۳	-۱/۷۷	۰/۱۲۴	۲۲/۸۲۴	۷۳
۲۴	-۲/۵۱	۰/۱۵۳	۱۷/۷۹	۶۲	۵۴	-۱/۶۱۵	۰/۱۱۹	۲۰/۲۷۷	۷۶
۲۵	-۱/۹۵۵	۰/۱۳	۴۱/۹۸۴	۷۱	۵۵	-۰/۸۷۱	۰/۱۰۴	۱۸/۹۲۲	۸۷
۲۶	-۳/۸۶۹	۰/۲۵	۹/۲۶۳	۴۲	۵۶	-۰/۸۷۱	۰/۱۰۴	۱۷/۸۱۱	۸۷
۲۷	-۳/۸۶۹	۰/۲۵	۳۷/۱۷۸	۴۲	۵۷	-۰/۵۳۲	۰/۱	۱۴/۴۰۷	۹۲
۲۸	-۱/۹۲	۰/۱۲۹	۲۳/۴۱۷	۷۱	۵۸	-۰/۰۸۹	۰/۰۹۸	۱۰/۱۶۵	۹۹
۲۹	-۲/۹۶۲	۰/۱۷۸	۲۷/۸۱۸	۵۶	۵۹	۰/۷۲۷	۰/۱۰۲	۱۹/۳۶۸	۱۱۱
۳۰	-۲/۶۱۱	۰/۱۵۸	۶/۹۴۴	۶۱	۶۰	۰/۷۱۶	۰/۱۰۲	۱۹/۹۴۹	۱۱۱

در این جدول مدل تک پارامتری (راش) نمایش داده شده است که احتمال شانس در آن صفر ($cg=0$) و ضریب

تشخیص یا شیب خط در تمامی سؤاها ثابت و برابر (۰/۹۵۵) است. به عبارتی سؤال ها از لحاظ شکل و محتوا همگون است. تنها پارامتر قابل محاسبه در این مدل، پارامتر دشواری است. مقادیر برآورد شده دشواری سؤاها ۳، ۴، ۶ نشان دهنده آسانترین و سؤاها ۵۹ و ۶۰ نشان دهنده دشوارترین سؤال است. سؤاها ۴ و ۶ دارای بیشترین خطای استاندارد دشواری و سؤال ۳۶ دارای کمترین خطای استاندارد دشواری است. به منظور بررسی انطباق هر سؤال با مدل از لحاظ تفکیک سطوح توانایی لازم است محاسبه مجذورکای صورت گیرد. چنانچه مقدار مجذورکای حاصل بر df ، بیش از ۲ باشد نشان دهنده نامناسب بودن سؤال ها است. سؤاها ۱۲، ۱۷، ۲۵، ۲۷، ۳۹، ۴۸، ۵۰ از جمله مواردی هستند که X^2 بحرانی کوچکتر است. در این جدول، ارزش مقیاسی با انتقال خطی به مقیاس استاندارد دیگری تبدیل شده است (مقیاس وکسلر با میانگین ۱۰۰ و انحراف استاندارد ۱۵) سؤال ۴۸ با درجه دشواری نزدیک به صفر، در مقیاس مذکور عدد ۱۰۰ را نشان می دهد. سؤالات ۵۹ و ۶۰ در بالای میانگین قرار دارد. در مقیاسه برآورد دشواری سؤال و توانایی، چنانچه افرادی با توانایی $1/06$ - به سؤالی با درجه دشواری $1/15$ - مواجه شوند، ۵۰ درصد احتمال پاسخ درست وجود دارد.

در جدول ۳ نتایج حاصل از تحلیل داده ها به کمک سه پارامتری و آزمون برازندگی تک تک سؤاها و مجموع آزمون ارائه شده است.

جدول ۳: آزمون برازش سؤاها بر اساس مدل سه پارامتری

سؤال	قدرت تشخیص	دشواری	عامل حدس	مجذور کای	سؤال	قدرت تشخیص	دشواری	عامل حدس	مجذور کای
۱	۰/۷۹۸	-۳	۰/۱۶	۱۵/۱۶۱	۳۱	۰/۴۰	-۲/۲۷۳	۰/۱۵	۱۹/۴۵۳
۲	۰/۷۳۲	-۳	۰/۱۶	۱۹/۶۷۳	۳۲	۰/۵۴۴	-۲/۰۶۲	۰/۱۳	۱۹/۰۰۹
۳	۰/۸۷۵	-۳	۰/۱۵	۴۸/۹۷۲	۳۳	۰/۴۱۲	-۲/۷۵۷	۰/۱۳	۲۶/۰۱۸
۴	۰/۹۲۸	-۳	۰/۱۶	۲۴/۵۹۷	۳۴	۱/۵۲۲	-۰/۶۹۱	۰/۱۹	۲۳/۹۹۲
۵	۰/۷۳۲	-۳	۰/۱۶	۲۳/۲۲	۳۵	۰/۹۶۶	-۰/۸۹۶	۰/۱	۲۵/۱۳
۶	۰/۹۶۱	-۳	۰/۱۵	۱۶/۳۳۴	۳۶	۰/۸۳۴	-۰/۱۸۳	۰/۰۶	۲۷/۵۸۵
۷	۰/۸۷۹	-۲/۸۷۱	۰/۱۲	۳۳/۶۴۱	۳۷	۰/۴۴۸	-۳	۰/۱۶	۳۱/۳۰۲
۸	۰/۶۲	-۲/۴۹۸	۰/۱۳	۲۷/۵۸۱	۳۸	۰/۴۲۲	-۳	۰/۱۵	۲۰/۵۷۸
۹	۰/۸۹۴	-۳	۰/۱۳	۱۳/۲۴	۳۹	۰/۴۰	-۳	۰/۱۸	۲۸/۲۰۲
۱۰	۰/۷۳۹	-۲/۸۸۵	۰/۱۲	۱۸/۰۶۵	۴۰	۰/۴۲۵	-۲/۸۱۵	۰/۱۴	۱۵/۱۵۱
۱۱	۰/۵۹	-۲/۶۷۶	۰/۱۳	۲۱/۷۰۸	۴۱	۰/۴۰	-۲/۱۰۷	۰/۱۵	۳۲/۷۱۴
۱۲	۰/۴۵۴	-۱/۵۳	۰/۱۱	۴۷/۲۴۳	۴۲	۰/۵۵۶	-۳	۰/۱۴	۱۲/۰۰۲
۱۳	۰/۶۵۲	-۳	۰/۱۷	۲۴/۰۹۸	۴۳	۰/۷۲۳	-۲/۱۵۳	۰/۱۳	۱۶/۲۳۷
۱۴	۰/۶۳۳	-۳	۰/۱۷	۲۳/۳۷۸	۴۴	۰/۵۹۳	-۲/۸۶۶	۰/۱۳	۱۵/۷۸۴
۱۵	۰/۷۴	-۳	۰/۱۵	۱۲/۷۳۲	۴۵	۰/۷۰۸	-۱/۸۳۹	۰/۱۱	۲۹/۲۹۱
۱۶	۰/۵۸۳	-۳	۰/۱۵	۱۶/۰۱۷	۴۶	۰/۵۸۴	-۲/۵۸۴	۰/۱۳	۱۹/۶۳۵
۱۷	۰/۴۰	-۲/۳۳	۰/۱۶	۵۷/۴۷۸	۴۷	۰/۵۶۳	-۰/۹۷۵	۰/۱۲	۱۵/۷۳۲
۱۸	۰/۵۵۲	-۳	۰/۱۵	۲۵/۹۶	۴۸	۰/۵۵۶	-۰/۱۰۷	۰/۰۵	۲۷/۳۷۳
۱۹	۰/۶۳۴	-۲/۳۴۷	۰/۱۳	۳۰/۲۶۶	۴۹	۰/۴۱۸	-۲/۸۵۱	۰/۱۳	۳۰/۸۸۴

۲۰	۰/۶۰۸	-۲/۷۴	۰/۱۳	۲۰/۷۲۳	۵۰	۰/۵۸۴	-۱/۶۴۹	۰/۱۲	۳۴/۰۴۱
۲۱	۰/۶۰۶	-۲/۶۸۳	۰/۱۳	۱۷/۱۱۵	۵۱	۰/۴۰	-۱/۱۷	۰/۱۴	۲۷/۹۸۴
۲۲	۰/۹۰	-۳	۰/۱۳	۱۹/۴۸۴	۵۲	۱/۳۳۱	-۱/۱۳۹	۰/۱۳	۱۶/۰۶۷
۲۳	۰/۴۳۷	-۲/۹۰۱	۰/۱۴	۲۰/۴۴۸	۵۳	۱/۱۲۳	-۱/۰۹۶	۰/۰۹	۳۱/۱۲۴
۲۴	۱/۰۶۷	-۱/۵۸۷	۰/۱۴	۱۸/۰۳۷	۵۴	۱/۰۵۲	-۱/۰۱۱	۰/۱۰	۱۰/۳۰۸
۲۵	۰/۴۰	-۲/۲۶۱	۰/۱۶	۵۷/۳۷۹	۵۵	۱/۰۰۲	-۰/۴۸۵	۰/۱۰	۲۸/۴۷۵
۲۶	۰/۷۶۷	-۳	۰/۱۳	۹/۰۶۶	۵۶	۲/۲۹۷	-۰/۱۹۶	۰/۱۹	۱۹/۰۴۴
۲۷	۰/۶۴۷	-۳	۰/۱۵	۳۴/۲۹۶	۵۷	۱/۱۷	-۰/۱۶	۰/۱۳	۱۳/۳۵۵
۲۸	۰/۴۰	-۲/۲۱۴	۰/۱۶	۳۴/۸۷۴	۵۸	۱/۲۸۵	۰/۱۱۲	۰/۱۱	۱۹
۲۹	۰/۵۷۹	-۲/۷۷۷	۰/۱۳	۱۹/۶۷۵	۵۹	۲/۲۲۱	۰/۹۴۴	۰/۲۲	۰/۸۶۱
۳۰	۰/۷۴	-۲/۰۵۱	۰/۱۳	۲۲/۸۵۵	۶۰	۲/۲۵۷	۰/۹۵۱	۰/۲۲	۲۴/۴۶۶

در این جدول با توجه به مقادیر دشواری (bg)، سؤالهای آسان و سؤالهای با دشواری بالاتر دیده می شود. در آزمون ریون که متشکل از ۵ گروه سؤال ۱۲ تایی است، برآورد مقادیر دشواری سؤالها نشان می دهد که هر گروه با سؤالهای سطح پایه (آسان) شروع می شود و با سؤالات دشوار خاتمه می یابد. به طوری که سؤال های ۱، ۱۳، ۲۶، ۳۷ از پائین ترین درجه دشواری (آسانترین) برخوردار است، و سؤالهای هر گروه شامل سؤال ۱۲، ۳۶، ۴۸، ۶۰ جزو سؤالهای دشوار محسوب می گردد. کمترین قدرت تشخیص متعلق به سؤالهای ۲۵، ۲۸، ۳۱، ۳۹، ۴۱ و ۵۱ برابر ۰/۴ و بیشترین قدرت تشخیص متعلق به سؤال ۵۶ با ۲/۲۹۷ است. مقادیر برآورد شده عامل حدس، سؤال ۴۸ دارای کمترین عامل حدس و سؤالهای ۵۹ و ۶۰ دارای بیشترین عامل حدس است. یعنی آزمودنیها به این سؤال بیشتر از همه سؤالها به صورت تصادفی پاسخ داده اند. آزمون مجذور کای نشان می دهد مجذور کای سؤال ۱۷ (۵۷/۴۷) نسبت به درجه آزادی (۱۷) زیاد است و با مدل تک پارامتری برازش ندارد.

۴. مقایسه دو مدل تک و سه پارامتری از لحاظ برآورد پارامتر توانایی و دشواری سؤال: به منظور مقایسه دو مدل تک و سه پارامتری اطلاعات مربوط به برازش آزمون و مجذور کای در جدول ۴ نشان داده شده است.

جدول ۴: مقدار مجذور کای و درجه آزادی در مدل تک پارامتری و سه پارامتری

سؤالها	مقدار مجذور کای در سه		درجه آزادی
	پارامتری	مجدور کای در تک پارامتری	
۳	۴۸/۹۷۲	۲۵/۰۰۳	۱۴
۱۲	۴۷/۲۴۳	۳۵/۶۶۴	۱۴
۱۷	۵۷/۴۷۸	۳۹/۷۳۳	۱۴
۲۵	۵۷/۳۷۹	۴۱/۹۸۴	۱۴
۲۷	۳۴/۲۹۶	۳۷/۱۷۸	۱۴
۳۹	۳۸/۲۰۲	۴۰/۸۳۹	۱۴
۴۸	۲۷/۲۷۳	۴۹/۳۲۳	۱۴
۵۰	۳۴/۰۴۱	۳۵/۸۶۹	۱۴

جدول ۴ اطلاعات مربوط به برازش و آزمون مجذور کای چند سؤال را بر اساس هر دو مدل نمایش می دهد. در

این جدول سؤالاتی که مجذور کای آنها از مقدار بحرانی بیشتر است برپایه هر دو مدل تک پارامتری و سه پارامتری آمده است. ۳ سؤال ۲۵، ۱۷، ۱۲ با هیچ یک از دو روش برازش ندارند سایر سؤالا با هر دو آزمون برازش دارند.

دشواری هر سؤال براساس هر دو مدل برآورد شده، در جدول ۵ ارائه شده است. شیب خط در این سؤالا بالای ۰/۷ است. دامنه شیب بین ۰/۵ تا ۲/۵ تغییر می کند.

جدول ۵: مقایسه دشواری سؤال در دو مدل تک و سه پارامتری

سؤالا	مدل تک پارامتری	مدل سه پارامتری	سؤالا	مدل تک پارامتری	مدل سه پارامتری
۱	-۴/۸۰۷	-۳	۳۱	-۲/۶۱۱	-۲/۲۷۳
۲	-۴/۴۴۳	-۳	۳۲	-۱/۹۳۷	-۲/۰۶۲
۳	-۴/۹۵۸	-۳	۳۳	-۲/۱۶۰	-۲/۷۵۷
۴	-۵/۳۳۶	-۳	۳۴	-۲/۳۰۴	-۰/۶۹۱
۵	-۴/۴۴۳	-۳	۳۵	-۱/۶۰۰	-۰/۸۹۶
۶	-۵/۱۳۲	-۳	۳۶	-۱/۴۱۵	-۰/۱۸۳
۷	-۳/۹۳۶	-۲/۸۷۱	۳۷	-۰/۳۴۴	-۳
۸	-۲/۸۰۵	-۲/۴۹۸	۳۸	-۳/۱۳۸	-۳
۹	-۴/۳۴۳	-۳	۳۹	-۲/۷۱۹	-۳
۱۰	-۳/۵۷۷	-۲/۸۸۵	۴۰	-۲/۷۴۷	-۲/۸۱۵
۱۱	-۲/۸۹۸	-۲/۶۷۶	۴۱	-۲/۴۳۸	-۲/۱۰۷
۱۲	-۱/۴۴۳	-۱/۵۳	۴۲	-۱/۸۱۹	-۳
۱۳	-۴/۳۴۳	-۳	۴۳	-۳/۲۱۵	-۲/۱۵۳
۱۴	-۴/۳۴۳	-۳	۴۴	-۲/۶۹۱	-۲/۸۶۶
۱۵	-۴/۲۵۰	-۳	۴۵	-۳/۱۰۱	-۱/۸۳۹
۱۶	-۳/۵۷۷	-۳	۴۶	-۲/۲۶۲	-۲/۵۸۴
۱۷	-۲/۰۰۸	-۲/۳۳	۴۷	-۲/۷۷۶	-۰/۹۷۵
۱۸	-۳/۳۸۴	-۳	۴۸	-۱/۱۵۵	-۰/۱۰۷
۱۹	-۲/۶۹۱	-۲/۳۴۷	۴۹	۰/۰۳۲	-۲/۸۵۱
۲۰	-۳/۰۲۹	-۲/۷۴۸	۵۰	-۲/۳۹۲	-۱/۶۴۹
۲۱	-۲/۹۶۲	-۲/۶۸۳	۵۱	-۱/۸۳۵	-۱/۱۷
۲۲	-۴/۳۴۳	-۳	۵۲	-۱/۱۱۸	-۱/۱۳۹
۲۳	-۲/۵۳۵	-۲/۹۰۱	۵۳	-۲/۰۴۵	-۱/۰۹۶
۲۴	-۲/۵۱۰	-۱/۵۸۷	۵۴	-۱/۷۷۰	-۱/۰۱۱
۲۵	-۱/۹۵۵	-۲/۲۶۱	۵۵	-۱/۶۱۵	-۰/۴۸۵
۲۶	-۳/۸۶۹	-۳	۵۶	-۰/۸۷۱	-۰/۱۹۶
۲۷	-۳/۸۶۹	-۳	۵۷	-۰/۸۷۱	-۰/۱۶
۲۸	-۱/۹۲۰	-۲/۲۱۴	۵۸	-۰/۵۳۲	۰/۱۱۲
۲۹	-۲/۹۶۲	-۲/۷۷۷	۵۹	-۰/۰۸۹	۰/۹۴۴

پس از چرخش هر دو مدل تک پارامتری و سه پارامتری، سؤال ۴۸ در نقطه میانگین قرار دارد و نمرات ۵۸ و ۶۰ بالاتر از میانگین، ولی در تعیین درجه‌های دشواری بین دو مدل تفاوت‌هایی دیده می‌شود، به طوری که، سؤال‌های ۱۶، ۱۵، ۱۴، ۱۳، ۶، ۵، ۴، ۳، ۲، ۱ در مدل سه پارامتری از یک درجه دشواری برخوردار است ولی در مدل تک پارامتری محدودی از درجه دشواری یکسان برخوردار است، مانند دشواری یکسان دو سؤال ۵، ۲ که متفاوت است از ۴ سؤال دیگر با درجه دشواری یکسان ۲۲، ۱۴، ۹، ۱۳ است. مقایسه توانایی برآورد شده از مدل تک پارامتری و سه پارامتری مستلزم توجه به میانگین توانایی‌های دو مدل و انحراف استاندارد و مقدار کشیدگی منحنی توزیع توانایی برآورد شده در هر مدل است. برای تعیین معناداری این تفاوت ناگزیر از آزمون تفاوت بین توانایی‌های برآورد شده برپایه دو مدل، در جدول ۶ ارائه شده است.

جدول ۶: نتایج آزمون t برای مقایسه میانگین پارامتر توانایی آزمودنی‌ها در هر دو مدل تک و سه پارامتری

مدل	میانگین	انحراف استاندارد	خطای استاندارد	t	p
تک پارامتری	-۰/۲۰	۱/۶۵	۰/۰۷	۲/۶۱	۰/۰۰۹
سه پارامتری	-۰/۰۱	۰/۸۰	۰/۰۴		

توجه: درجه آزادی برابر با ۴۶۷ است.

چنانکه در جدول ۶ دیده می‌شود، نتیجه آزمون t مربوط به مقایسه میانگین دو مدل تک پارامتری و سه پارامتری برابر با ۲/۶۱ و از لحاظ آماری در سطح ۰/۰۰۹ معنادار است. با توجه به ارقام جدول میانگین مدل تک پارامتری $\bar{X} = -۰/۲۰$ بالاتر از میانگین مدل سه پارامتری -۰/۰۱ است. این یافته بیانگر آن است که مدل تک پارامتری نسبت به مدل سه پارامتری توانمندتر است.

چنانچه تفاوت دو مدل و برآوردهای انجام شده در مدل تک پارامتری را بتوان با درصد خطائی قبول کرد، می‌توان در محاسبات و مصرف زمان صرفه جویی کرد. با این هدف، آزمون همبستگی بین توانایی‌های برآورد شده ناشی از اجرای دو مدل تک پارامتری و سه پارامتری و محاسبه ضریب رگرسیون صورت گرفت. همبستگی بین توانایی‌ها برابر با ۰/۲۸۹ است. این مقدار هرچند زیاد نیست اما از لحاظ آماری معنادار است و ضریب رگرسیون مربوط نیز نشانگر رابطه خطی بین توانایی‌های برآورد شده برپایه این دو مدل است که قابلیت پیش بینی بین این دو مدل را نشان می‌دهد.

از محاسبات مربوط به همبستگی و آزمون t این نتایج استنتاج می‌شود: (۱) بین برآوردهای توانایی براساس دو مدل تک پارامتری و سه پارامتری تفاوت معنادار وجود دارد اما بین نمره‌های توانایی‌ها در هر دو مدل همبستگی وجود دارد. (۲) از روی نمره‌های خام افراد می‌توان توانایی آنان را در مدل سه پارامتری پیش بینی کرد. (۳) از مقایسه میانگین دو مدل این نتیجه حاصل می‌شود که مدل تک پارامتری نسبت به مدل سه پارامتری مقاوم تر است.

بحث

با در نظر گرفتن هدف اصلی پژوهش و با توجه به یافته‌های حاصل از این پژوهش، می‌توان سؤال‌های پژوهش را مورد بررسی قرار داد. در بررسی سؤال یکم و دوم پژوهش که «آیا بین سؤال‌های آزمون هوش پیشرونده ریون هماهنگی وجود دارد؟» و «آیا مقیاس از اعتبار کافی برخوردار است؟» برای برآورد اعتبار مقیاس از ضریب آلفای کرونباخ و برای تعیین میزان هماهنگی بین سؤال‌های آزمون از تحلیل مؤلفه‌های اصلی (PC) استفاده شد. اعتبار به دست آمده برای این آزمون با استفاده از روش آماری آلفای کرونباخ برابر ۰/۸۳۹ است. این برآورد اعتبار نشان می‌دهد

که آزمون از اعتبار مناسبی برخوردار است. با استفاده از روش تحلیل مولفه‌های اصلی، تعداد ۵ عامل هوش کلی، درک و استعداد ریاضی، حافظه، تجسم فضایی، و سرعت به دست آمده است. با استفاده از تجزیه و تحلیل به روش کلاسیک نشان می‌دهد که آزمون از ویژگیهای روانسنجی مناسبی برخوردار است.

در بررسی سؤال سوم پژوهش که «آیا بین برآورد پارامترهای توانایی آزمودنی در آزمون هوش پشرونده ریون برپایه مدل راش و مدل سه پارامتری تفاوت اساسی وجود دارد؟» با توجه به نتایج حاصل از اجرای تحلیل داده‌ها به کمک مدل تک پارامتری راش و آزمون برازندگی تک تک سؤالها و مجموع آزمون می‌توان نتیجه گرفت از مجموع ۶۰ سؤال مورد مطالعه تنها سؤال ۶۰ در این گروه سنی با مجموعه سؤالها هماهنگ نیست. نتایج حاصل از اجرای آزمون برازندگی برای کل مجموعه سؤالها از آن است که مجموع مقیاس با مدل راش هماهنگی لازم دارد. با آنکه مقصود اصلی پژوهش حاضر تعیین چگونگی برازش سؤالها با مدل راش نبوده است ولی نتیجه فرعی پژوهش در تأیید این فرضیه است که مجموعه سؤالها، یک خصیصه تک بعدی را می‌سنجد.

هدف اصلی پژوهش حاضر، مقایسه بررسی توانمندی مدل راش و سه پارامتری است. بنابراین با مقایسه پارامترهای دشواری و توانایی برآورد شده به کمک دو مدل، هم تأثیر قدرت تشخیص سؤالها برای برآورد پارامترهای معلوم شده و هم میزان مقاوم بودن مدل راش را در شرایطی که مفروضه هایش رعایت نشده است روشن می‌سازد. مقایسه دو مدل نشان می‌دهد میزان پارامتر شیب سؤال (a) در برآورد پارامتر دشواری سؤال (b) مؤثر است و تأثیر پارامتر a بر روی b غیرقابل انکار است، به کمک مدل مجذور کای برای مقایسه فراوانی پارامترهای برآورد شده توانایی در دو مدل، نتایج حاکی از آن است که بین دو مدل از نظر برآورد پارامتر توانایی تفاوت معنادار وجود دارد. به عبارت دیگر میزان پارامتر a در برآورد پارامتر توانایی افراد تأثیرگذار است. می‌توان نتیجه گرفت که مدل راش برای برآورد پارامتر دشواری سؤال وقتی قدرت تشخیص سؤالها متفاوت است، برآورد کننده خوبی نیست ولی برای برآورد توانایی افراد برآورد کننده نسبتاً مناسبی است. البته بدیهی است در برآورد پارامتر توانایی، میزان تفاوت‌های قدرت تشخیص سؤالها و حجم نمونه نیز مؤثر است و چنانچه به این دو مسئله توجه نشود، پارامتر توانایی نیز با خطای زیاد برآورد خواهد شد. لرد و ناویک (۱۹۶۸) و راش (۱۹۶۰) اعتقاد دارند اگر قدرت تشخیص سؤالها از یکدیگر تفاوت داشته باشند اصولاً کار برد مدل تک پارامتری راش از لحاظ فنی نادرست خواهد بود. در این مدل فقط پارامتر دشواری (b) مورد توجه قرار می‌گیرد و دارای دو فرض اساسی است. ۱- همه سؤالها از درجه تشخیص (a) ثابت و یکسانی برخوردارند. ۲- پارامتر حدس (c) معادل صفر می‌باشد (ستاری، ۱۳۸۲).

نتایج اجرای آزمون t نیز نشان می‌دهد که بین توانایی های افراد برپایه دو مدل تک پارامتری و سه پارامتری تفاوت وجود دارد. این یافته بیانگر آن است که مدل تک پارامتری نسبت به مدل سه پارامتری توانمندتر است. از محاسبات مربوط به همبستگی و آزمون این نتایج به دست آمده است: (۱) بین برآوردهای توانایی براساس دو مدل تک پارامتری و سه پارامتری تفاوت وجود دارد اما بین نمره های توانایی ها در هر دو مدل همبستگی وجود دارد

(۲) از روی نمره‌های خام افراد می‌توان توانایی آنان را در مدل سه پارامتری پیش بینی کرد. ۳- از مقایسه میانگین دو مدل این نتیجه حاصل می‌شود که مدل تک پارامتری نسبت به مدل سه پارامتری مقاوم‌تر است. این مدل کلیه مزایای مدل‌های دو و سه پارامتری را دارد و مفروضه‌های قویتری در مقایسه با سایر مدلها دارد. چنانچه داده‌ها با مدل برازش داشته باشد، استفاده فایده عملی دیگری از آن حاصل می‌شود. (آلن و وین، ۱۳۸۴).

محدودیت‌های پژوهش

۱. نمونه مورد آزمون فقط از منطقه تهران انتخاب شده و آزمون در یک محدوده سنی کم و فقط در دختران دبیرستانی انجام گرفته است.
۲. حجم نمونه در برآورد پارامترهای دشواری سؤال به ویژه در پارامتر شیب تأثیر اساسی دارد. وقتی قدرت تشخیص سؤالها از گروههای کوچک برآورد شود، اشتباه برآوردهای مقدار تشخیص سؤال، ممکن است به حدی بزرگ شود که کاربرد سایر برآوردها را غیر کارآمد سازد.
۳. از مفروضه‌های اصلی مدل راش، یکسان بودن پارامتر شیب همه سؤالهاست، چون ماهیت آزمون هوش پیشرونده ریون به گونه ای است که باید در هر مجموعه آن هم سؤالهای دشوار و هم سؤالهای ساده وجود داشته باشد. بدون تردید این مسئله در برآورد ارزش مقیاس مربوط به سؤالها و نمره ها تأثیر گذاشته است.
۴. یکی از مفروضه‌های اصلی هر دو مدل، تک عاملی بودن تست موردنظر است. اما نتایج تحلیل عاملی نشان می‌دهد که ماتریس‌های پیشرونده ریون، دست کم براساس مطالعه حاضر، از پنج عامل اشباع شده است.

پیشنهاد‌های پژوهش

۱. برای دست‌یابی به نمونه ای دقیق و معرف، حجم نمونه‌ها افزایش یافته و با توجه به هدف اساسی تهیه مقیاس، اجرای مواد و نرم‌یابی لازم بر روی کل جامعه ایران صورت گیرد.
۲. با توجه به عدم تساوی مقدار شیب سؤالها (قدرت تشخیص)، ایجاب می‌کند که در مرحله اجرایی تجربی مقیاس در سؤالهای سقفی و پایه از نظر تعداد و کیفیت آنها تجدیدنظر صورت گیرد.
۳. از آن جا که می‌توان پارامترهای دو مدل را با مقدار خطای قابل قبول کمتری برآورد نمود، توصیه می‌شود موقعی که لازم است بدانیم فرد واقعاً در کجای خصیصه مکنون قرار دارد، از مدل چند پارامتری استفاده شود ولی در مواقعی که لازم است بدانیم فرد در مقایسه با دیگران چگونه است، برای صرفه جویی در زمان می‌توان مدل تک پارامتری راش را برای برآورد توانایی وی به کار برد.

منابع

- آلن، مری جی و ین، وندی ام. (۱۳۸۴). مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی). (علی دلاور، مترجم). (چاپ دوم). تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت). تاریخ انتشار به زبان اصلی، ۱۹۷۹).
- امبرستون، سوزان ای و رایس، استیون پی. (۱۳۸۸). نظریه‌های جدید روان‌سنجی برای روان‌شناسان (IRT). (حسن پاشا)
- شریفی، ولی‌الله فرزاد، مجتبی حبیبی عسگرآباد و بلال ایزانلو، مترجمان. (چاپ اول). تهران: انتشارات رشد. تاریخ انتشار به زبان اصلی، ۲۰۰۰).
- بی‌بیکر، فرانک. (۱۳۸۱). پایه‌های اساسی تئوری سؤال پاسخ (نظریه جدید روان‌سنجی). (حیدرعلی هومن و علی عسگری، مترجمان) (چاپ سوم). تهران: انتشارات دانشگاه تهران. تاریخ انتشار به زبان اصلی، ۱۹۸۲).
- ثرندایک، آر.ال. (۱۳۷۵). روان‌سنجی کاربردی. (حیدرعلی هومن، مترجم) (چاپ سوم). تهران: انتشارات دانشگاه تهران. تاریخ انتشار به زبان اصلی، ۱۹۸۲).
- ستاری، بهزاد. (۱۳۸۲). روان‌سنجی پیشرفته کاربردی. نظریه پاسخ پرسش (IRT) رویکردی جدید در روان‌سنجی (مدلهای سه پارامتری، دو پارامتری و تک پارامتری). از نظریه تا کاربرد. مشهد: موسسه به‌نشر (انتشارات آستان قدس رضوی).
- سیف، علی‌اکبر. (۱۳۸۶). اندازه‌گیری، سنجش و ارزشیابی آموزشی. تهران: نشر دوران.
- شریفی، حسن پاشا. (۱۳۸۳). اصول روان‌سنجی و روان‌آزمایی. تهران: انتشارات رشد.
- عسگری، علی. (۱۳۸۶). روی آوردهای نوین در روان‌سنجی. مجله روان‌شناسان ایرانی، دانشگاه آزاد اسلامی، سال سوم شماره ۱۱ و ۱۲ (قسمت اول و دوم).
- عسگری، علی. (۱۳۸۷). روی آوردهای نوین در روان‌سنجی. مجله روان‌شناسان ایرانی، دانشگاه آزاد اسلامی، سال چهارم شماره ۱۳ (قسمت سوم).
- کوپر، کالین. (۱۳۷۹). تفاوت‌های فردی (نظریه و سنجش). (حسن پاشا شریفی و جعفر نجفی زند، مترجمان). (چاپ ؟). تهران:
- انتشارات سخن. (تاریخ انتشار به زبان اصلی، ۱۹۹۸). (تاریخ چاپ به زبان اصلی، ؟).
- فراهانی، مهدی. (۱۳۷۵). مقایسه مدلهای اندازه‌گیری (کلاسیک و سؤال پاسخ) از لحاظ برآورد پارامترهای سؤال و توانایی. پایان‌نامه کارشناسی ارشد. دانشگاه علامه طباطبایی. هومن، حیدرعلی. (۱۳۸۶). اندازه‌گیری‌های روانی و تربیتی (فن تهیه تست). (چاپ پانزدهم). تهران: انتشارات پیک فرهنگ.
- هومن، علی. (۱۳۷۳). مقایسه مدل تک پارامتری راش و مدل دو پارامتری. پایان‌نامه کارشناسی ارشد. دانشگاه علامه طباطبایی.

Anderson. E. B.(1973). a goodness of fit test for the Rasch model. *Psychometrical* 1973, 38, 123 – 140
Andrich, D.(2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42, 1-16.

Hambleton, R. K., & Traub, R. E.(1973).Analysis of empirical data using two logistic latent trait models. *British Journal of Mathematical and Statistical Psychology*, 1973, 26, 195 – 211.

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), 3847.

Lord, F.M., & Novick,M,R.(1968). *Statistical Theories of Mental Test Score*.Menlo Park, Calif: Addison.Wesley.

Rash, G.(1960).*Probabilistic Models for some Intelligence and Attainment Test*. Copenhagen,Denmark: Danmarks pædagogiske Institut.

Whitely, S. E., & Dawis, R.(1984)The nature of objectivity with the Rash model. *Journal of Educational Measurement*, 1974, 11, 163 – 178.
