

# تأثیر تعداد گزینه‌های سؤال در ویژگی‌های روان‌سنجی آزمون و توانایی برآوردشده در مدل‌های پرسش پاسخ و کلاسیک اندازه‌گیری

بهنام کریمی<sup>۱</sup>

محمد رضا فلسفی‌نژاد<sup>۲</sup>

فریبرز درتاج<sup>۳</sup>

تاریخ پذیرش: ۹۰/۱۲/۲۴

تاریخ وصول: ۹۰/۶/۱۸

## چکیده

زمینه: سهولت نمره‌گذاری، اجرا و عینیت آزمون‌های چندگزینه‌ای سبب شده که به عنوان ابزار اصلی در سنجش‌های وسیع مورد استفاده قرار گیرد. انتقادهای زیادی نسبت به سؤالات چندگزینه‌ای مطرح شده است. نظیر پوشش ندادن به تمامی اهداف تربیتی (سطوح پایین شناختی را می‌سنجند) و استفاده از عامل حدس و گمان در پاسخ به سؤال‌ها. در این میان عده‌ای نیز افزایش تعداد گزینه‌های سؤال را راهی برای مقابله با این مشکلات دانسته‌اند. هدف: هدف از پژوهش حاضر بررسی تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی آزمون‌ها و سؤالات و همچنین توانایی برآورد شده آزمودنی‌ها در نظریه کلاسیک و پرسش پاسخ بود. روش: جامعه آماری شامل کلیه دانش‌آموزان سال سوم دبیرستان‌های شهر شیراز بود که ۶۰۸ نفر از آنان به شیوه تصادفی به عنوان گروه نمونه انتخاب شدند. برای جمع‌آوری اطلاعات از دو آزمون پیشرفت تحصیلی زبان و حسابان که به همین منظور تهیه و تنظیم شده بودند استفاده شد. یافته‌ها: تجزیه و تحلیل داده‌ها نشان داد که تعداد گزینه‌ها بر پارامترهای سؤال اثر ندارد و تأثیر

---

۱- کارشناس ارشد روان‌سنجی karimi.adviser@gmail.com

۲- استاد دانشگاه علامه طباطبایی

۳- استاد دانشگاه علامه طباطبایی F\_dortaj@yahoo.com

تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی برآورد شده آزمودنی‌ها، در آزمون‌های مختلف یکسان است. همچنین بین پارامترهای برآورد شده در نظریه کلاسیک و پرسش پاسخ تفاوت وجود داشت. نتیجه‌گیری: بعد از بررسی مفروضه‌های نظریه پرسش و پاسخ مشخص شد که داده‌ها با مدل دو پارامتری برازش بهتری دارند، و تفاوتی بین تعداد گزینه‌ها و برازش با مدل مشاهده نشد. همچنین بین توانایی برآورد شده و تعداد گزینه‌ها تفاوت مشاهده شد.

**واژگان کلیدی:** آزمون‌های چندگزینه‌ای، نظریه کلاسیک اندازه‌گیری، نظریه پرسش

پاسخ.

#### مقدمه

سهولت نمره‌گذاری، اجرا و عینیت آزمون‌های چندگزینه‌ای سبب شده که به عنوان ابزار اصلی در سنجش‌های وسیع<sup>۱</sup> مورد استفاده قرار می‌گیرد (نظیر کنکور، تافل و...). انتقادهای زیادی نسبت به سؤالات چندگزینه‌ای مطرح شده است نظیر پوشش ندادن تمامی اهداف تربیتی (سطوح پایین شناختی را می‌سنجند)، و استفاده از حدس و گمان که یکی از اصلی‌ترین ایرادات نسبت به کنکور سراسری بوده که منجر به طرح حذف کنکور شده است. در این میان عده‌ای نیز افزایش تعداد گزینه‌های سؤال را راهی برای مقابله با این مشکلات دانسته‌اند.

با این همه اطلاعات محدود و ناچیزی در مورد اثربخشی رویکردهای ارائه شده وجود دارد. اگر چه شعور عادی می‌پذیرد که افزایش تعداد گزینه‌ها حدس زدن را کاهش می‌دهد، اما امکان ایجاد گزینه‌های جالب و جذاب (بیش از ۳ یا ۴) در بسیاری مواقع وجود ندارد. از طرفی تعیین تعداد مطلوب گزینه‌ها نیاز به شواهد علمی و تجربی دارد.

نوشتن سؤال‌های تستی با وجود همه تلاش‌هایی که در جهت مکانیزه و کامپیوتری کردن آن به عمل آمده است همچنان به عنوان یک هنر تلقی می‌شود. ابتکار و استادی در گنجاندن



گزینه بایستی داشته باشد تا از حداکثر پایایی برخوردار باشد همواره مورد بحث می‌باشد. تعداد گزینه‌های سؤال‌های چندگزینه‌ای (گزینه‌های درست یا انحرافی) از ۲ تا ۵ و گاه‌ا تا ۶ متغیر است و به لحاظ نظری هر چه تعداد گزینه‌ها بیشتر باشد امکان حدس زدن کمتر است اما ابتدا باید مشخص کرد در یک آزمون، چند گزینه انحرافی بیشتر، باعث افزایش پایایی سؤال می‌شود به هر حال در عمل ممکن است اضافه کردن گزینه‌های انحرافی واقعا باعث افزایش پایایی آزمون نشود زیرا پیدا کردن گزینه‌های خوب مشکل است. پایایی هر سؤال از طریق گزینه‌های انحرافی که هیچ وقت انتخاب نشده‌اند بالا نمی‌رود مطالعات نشان داده‌اند که به ندرت می‌توان سؤالاتی را یافت که بیشتر از ۳ یا ۴ گزینه انحرافی که به طور مؤثر عمل می‌کنند داشته باشند و گزینه‌های انحرافی خنثی ممکن است به پایایی آزمون لطمه وارد کنند به این دلیل که آنها زمان بیشتری را برای خواندن می‌گیرند و فضای آزمون را اشتغال می‌کنند (کاپلان و ساکوزو، به نقل از دلاور و درتاج و فرخی، ۱۳۸۶).

### سؤال‌های تحقیق

آیا پارامترهای سؤال نسبت به تعداد گزینه‌ها حساس است؟  
 آیا تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی سؤالات و توانایی برآورد شده آزمودنی‌ها در نظریه کلاسیک و سؤال پاسخ متفاوت است؟  
 آیا توانایی برآورد شده آزمودنی‌ها از تعداد گزینه‌های سؤال تأثیر می‌پذیرد؟  
 آیا تعداد گزینه‌ها بر برآورد داده‌ها با مدل در نظریه سؤال پاسخ تأثیر می‌گذارد؟  
 یک مطالعه درباره مسائلی که به انتخاب گزینه‌های انحرافی مربوط می‌شود پیشنهاد می‌کند که معمولاً بهتر است ۳ یا ۴ گزینه انحرافی که به طور مؤثر عمل می‌کنند برای هر سؤال تدوین شود (آناستازی و ارینا، ۱۹۹۷ به نقل از دلاور، ۱۳۷۴).  
 یونسی (۱۳۸۶) به بررسی ویژگی‌های روان‌سنجی سؤال‌های آزمون فراگیر رشته روان‌شناسی دانشگاه پیام نور در سال ۱۳۸۵ پرداخته و نتایج تحقیق را این طور بیان می‌کند.

تجزیه و تحلیل داده‌ها نشان داد که گزینه‌های انحرافی تمامی سؤالات در همه آزمون‌ها هم احتمال نیستند و عملکرد معیوبی داشته‌اند. نتایج تجزیه و تحلیل حاکی از این است که در آزمون‌های ادبیات فارسی، روانشناسی و جامعه‌شناسی مدل دو پارامتری نسبت به مدل سه پارامتری برآزش بهتری با مجموعه داده‌های آزمون دارد.

معلمی اوře (۱۳۸۷) در تحقیقی به مقایسه دقت برآورد توانایی در سؤالات چندگزینه‌ای با بکارگیری مدل سازی- سؤال پاسخ دو و چند ارزشی- پرداخته است. نتایج حاصل از برآزاندن مدل دو ارزشی ۱ و ۲ و ۳ پارامتری بر داده‌های سؤال پاسخ مذکور نشان داد مدل دو پارامتری که از نظر ساختار پارامتری در میان مدل‌های دو ارزشی نزدیک‌ترین مدل به NRM<sup>۱</sup> است. در این میان از برآزندگی بهتری با داده‌های فوق‌الذکر برخوردار است و لذا این مدل به عنوان مدل دو ارزشی مناسب جهت مقایسه دقت برآورد توانایی با مدل چند ارزشی برگزیده (NRM) انتخاب شد.

هادیان (۱۳۷۶) تأثیر تعداد گزینه‌های سؤال در اعتبار تست‌های چندگزینه‌ای را مورد بررسی قرار داد. هدف از این پژوهش مطالعه تأثیر تعداد گزینه‌ها در اعتبار تست‌های چندگزینه‌ای در بین سه گروه قوی، متوسط و ضعیف است و تعیین این که چه تعداد گزینه در هر یک از گروه‌ها بیشترین اعتبار را برای تست فراهم می‌کنند. ۴۰۰ نفر از دانش‌آموزان دختر و پسر سال اول دبیرستان‌های نظام جدید مناطق ۱ و ۲ و ۱۸ آموزش و پرورش شهر تهران در سال تحصیلی (۷۶-۷۵) به شیوه نمونه‌گیری چند مرحله‌ای انتخاب شدند.

بعد از اجرای آزمون ضریب اعتبار از طریق فرمول آلفا (KR20) برای ۴ نوع سؤال ۲ تا ۵ گزینه‌ای در هر یک از سه گروه متوسط، ضعیف و قوی به دست آمد. ضرایب اعتبار به دست آمده برای هر فرم از تست مورد مقایسه قرار گرفت. نتایج نشان داد که بین ضریب اعتبار تست‌های ۲ تا ۵ گزینه‌ای برای گروه قوی و ضعیف تفاوت اساسی و معناداری وجود ندارد و در

این دو گروه ضریب اعتبار تست‌های ۲گزینه‌ای به طور قابل ملاحظه‌ای از سایر تست‌ها بیشتر بود. ولی بین ضرایب اعتبار تست‌های ۳ تا ۵گزینه‌ای برای گروه متوسط تفاوت معنادار بود بدین معنی که تست‌های ۳گزینه‌ای، نسبت به تست‌های با تعداد گزینه‌های بیشتر و کمتر از اعتبار بالاتری برخوردار بود.

بیرقی (۷۷-۱۳۷۶) به بررسی تأثیر تعداد گزینه‌ها و زمان اجرای آزمون بررسی نمره کل آزمون پرداخته است. داده‌های جمع آوری شده بر روی ۲۰۰ نفر گروه نمونه انجام شده است. آزمودنی‌ها بر اساس نمره‌های حاصل از یک آزمون ادبیات جداگانه که در ابتدای تحقیق اجرا شده است، به دو گروه تقسیم شده اند. ویکی از گروه‌ها به یک آزمون سه گزینه‌ای و گروه دوم به یک آزمون ۴گزینه‌ای کمتر از میانگین نمرات ۳گزینه‌ای است یعنی افزایش تعداد گزینه‌ها باعث کاهش نمره کل آزمون شده است.

پاک نژاد (۱۳۷۷) به بررسی، تأثیر تعداد گزینه‌های سؤال و مدت زمان پاسخ دهی بر پایایی آزمون‌های چندگزینه‌ای پیشرفت تحصیلی پرداخته است. تحقیق حاضر به منظور بررسی تأثیر تعداد گزینه‌های سؤال و مدت زمان پاسخ‌دهی بر پایایی آزمون‌های چندگزینه‌ای پیشرفت تحصیلی بر روی یک نمونه ۴۰۰ نفری از دانش آموزان سال سوم دبیرستان نظام جدید آموزش متوسطه شهرستان اندیمشک که در سال تحصیل فرهنگ و معارف اسلامی انجام شده است. در این تحقیق، ضرایب پایایی آزمون‌های چهارگزینه‌ای، سه گزینه‌ای، و دوگزینه‌ای و آزمون چهارگزینه‌ای با مدت زمان پاسخ دهی کوتاه و بلند با استفاده از آزمون‌های معنی دار بودن اختلاف بین دو ضریب همبستگی در نمونه‌های همبسته و مستقل در سطوح  $\alpha=0.05$  مورد مقایسه قرار گرفته اند نتایج حاکی از آن بود که بین ضرایب پایایی آزمون‌های چهارگزینه‌ای و سه گزینه‌ای اختلاف معنی داری نمی باشد. و بین ضرایب پایایی آزمون‌های چهارگزینه‌ای و دو گزینه‌ای اختلاف معنی دار می باشد.

تام بورکار<sup>۱</sup> (۲۰۰۹) در تحقیقی بر روی آزمون‌های ورودی مدارس و دانشگاه‌ها با بیان این که باید از آزمون‌ها و سؤالات چند گزینه‌ای استفاده شود ادعا می‌کند تست‌های SAT دارای مشکلات خاص و سیستماتیک است. بورکار می‌گوید مشکلات زیربنایی، ناشی از باز پاسخ بودن بعضی از سؤالات است. همچنین ادعا می‌کند که مهارت تفکر انتقادی در این آزمون‌ها (باز پاسخ) واقعی نیست. زیرا آزمون‌گران به راحتی نمی‌توانند ادعای خود را در مورد کودکان ۶ تا ۱۱ ساله ثابت کنند. بورکار بیان می‌کند با جایگزینی آزمون‌های چند گزینه‌ای به جای باز پاسخ دیگر مشکلی نخواهیم داشت زیرا این تست‌ها به دقت دانش و توانایی را ارزیابی می‌کنند. با آزمون‌های چند گزینه‌ای می‌توان مقایسه‌های سال به سال از عملکرد دانش‌آموزان داشت. بورکار (۲۰۰۹) می‌گوید در جهان واقعی کسی فاقد دانش نیست و استفاده از سؤالات انشایی و باز پاسخ امری خطیر و فاجعه‌بار و بسیار پرهزینه برای سنجش عملکرد افراد است.

گاتس شال و کوبینجر<sup>۲</sup> (۲۰۰۸) با تحقیقی به عنوان دشواری سؤالات آزمون‌های چند گزینه‌ای، بیان داشته که آزمون‌های چند گزینه‌ای خود مشکل‌زاست زیرا اغلب به راحتی فرد به جواب درست دست پیدا می‌کند و باعث می‌شود آزمودنی یک حدس زنده خوش شانس باشد. در این تحقیق (کوبینجر و گاتس شال، ۲۰۰۸) از آزمون‌های موازی استفاده شد که از لحاظ محتوا شبیه به هم هستند اما پاسخ‌های آنها متفاوتند. ۱۷۳ تست به صورت تصادفی در کتابچه‌هایی ارائه شدند. با انجام تحلیل مدل راش بعد از حذف ۳۹ سؤال با مدل راش برآزش داشت نتیجه اولیه به این صورت بود که آزمون چندگزینه‌ای ۵ تا X گزینه‌ای تفاوت معنی‌داری با آزمون جواب آزاد نداشت. در این تحقیق بیان شده است اگر از سؤالات با فرمت ۱ گزینه صحیح از ۵ گزینه انحرافی و ۱ گزینه صحیح از ۴ گزینه انحرافی استفاده شود. چندان تفاوتی نمی‌کند.

---

1. Tom burkard  
2. Kubinger, Gottschall

در تحقیقی توسط جیمز و ماروی (۲۰۰۸) با عنوان تأثیر گزینه معیوب در امتحانات چند گزینه‌ای ارزیابی پرستاران به انتقاد از امتحانات چندگزینه‌ای پرداخته و با تاسف بیان کرده است که رواج گزینه‌های معیوب در امتحانات معلم ساخته باعث تبعات منفی زیادی شده است. در این تحقیق با یک بررسی دراز مدت بر روی دانش آموختگان بوردرپرستاری مشخص شد افراد با نمره بالا وقتی دوباره آزمون شدند نتایج به نحوی بود که انگار ارزیابی‌های قبلی، غلط بودند (دانش آموزان نمرات پایینی در آزمون جدید گرفتند). به گفته جیمز و ماروی (۲۰۰۸) ممکن است در ظاهر فقط برای دانش آموزان دوره پرستاری اتفاق افتاده باشد. ولی پر واضح است که در تمام عرصه‌ها به همین نحو است، جیمز و ماروی پیشنهاد می‌کنند عیوب سؤالات چندگزینه‌ای با دقت بیشتری بررسی شود و برای تمام سطوح موفقیت و توانایی بررسی شود.

یانگ سوک سو<sup>۱</sup> (۲۰۰۸) در پژوهش خود با عنوان استفاده از طرح آشیانه‌ای برای داده‌های آزمون‌های چندگزینه‌ای ادعا می‌کند، موقعی که راه حل و استراتژی صحیح در ممانعت از انتخاب گزینه‌های انحرافی به کار برده می‌شود، می‌تواند فرایند زیر بنایی سؤالات آزمون‌های چندگزینه‌ای را بهتر نشان دهد. این رساله مدل جدیدی برای ارزیابی سؤالات چند گزینه‌ای با استفاده از الگوریتم برآورد بیشینه درست نمایی کناری (MML) ارائه می‌دهد. به علاوه سؤالات و طبقه کاربرد اطلاعات برای هر مدل به کار رفته است. و به ارزیابی عملکرد هر مدل و مطالعات مشابه برای کشف پارامترهای سؤال، پرداخته است. در کاربرد مدل به وسیله تحلیل DIF از نسبت بیشینه تست (Likelihood Ratio) استفاده کرده است.

چاین چی (۲۰۰۷) رساله دکتری خود را با عنوان تأثیر حدس در ارزیابی ابعاد (تک بعدی بودن و چند بعدی بودن) آزمون‌های چندگزینه‌ای ارائه کرده است. در این تحقیق ۴ شاخص برای ارزیابی تک بعدی بودن مورد استفاده قرار گرفته است نسبت واریانس، تحلیل‌های موازی، تبدیل RMSR و تفاوت خی دو.

1. Young suk suh



وای مین هوانگ<sup>۱</sup> (۲۰۰۴) بیان می‌کند که آزمون‌های چندگزینه‌ای بیشترین تعداد را در میان تنوع بسیار زیاد آزمون‌ها به خود اختصاص داده است. آزمون‌های چندگزینه‌ای عمومی برای مواردی مثل تعیین پیشرفت، تعیین پایه دانش، دانشجویان ابقا یا ارتقا درجه علمی مورد استفاده قرار می‌گیرد. بر همین اساس می‌گوید علی‌رغم نفوذ آزمون‌های چندگزینه‌ای، شواهد تجربی حاکی از آن است که چگونگی نوشتن آزمون‌های چندگزینه‌ای دچار کاستی است.

آلن ام<sup>۲</sup> (۲۰۰۸) بیان می‌کند ابهام در آیتم‌های تست باعث ورود واریانس‌های نامربوط در آزمون‌های چندگزینه‌ای می‌شود.

کارین سوودفورد و پیتر بانکرافت (۲۰۰۹) در مقاله‌ای تحت‌عنوان آزمون‌های چندگزینه‌ای را سخت تصور نکنید به بررسی آزمون‌های چندگزینه‌ای پرداخته است. و چگونگی ساخت، اجزا و محدودیت‌های این آزمون‌ها را مورد بررسی قرار داده است. در بخشی از این مقاله به تعداد گزینه‌ها در آزمون‌های چندگزینه‌ای پرداخته و اظهار می‌دارد که در سال‌های اخیر بیشترین اظهارنظرها به آزمون‌های ۳ و ۴ و ۵ گزینه‌ای معطوف بوده است. و بیان می‌کند که ادعای آنها که ۵ گزینه‌ای را انتخاب می‌کنند بیشتر به خاطر از بین بردن حدس است. این پژوهشگران استفاده از ۳ گزینه‌ای یا ۴ یا ۵ گزینه‌ای را معطوف به مقتضی زمان کاربرد می‌دانند تا از اشتباهات دوری شود.

تسو هیتوشیزوکا<sup>۳</sup> (۲۰۰۶) در تحقیقی در دانشگاه کانسای ژاپن آثار کاهش تعداد گزینه‌ها سؤالات بر روی مشخصه‌های روان‌سنجی آزمون ورودی دانشگاه کانسای را مورد بررسی قرار داد. این آزمون به صورت یک سری سؤالات ۴ گزینه‌ای بود که برای غربالگری مورد استفاده قرار می‌گرفت که با حذف یک گزینه به یک آزمون سه گزینه‌ای تبدیل شد. و بر روی گروه دیگری اجرا شد. پاسخ به دو آزمون و مقایسه دو آزمون نشان داد که درجه سهولت

1. Yi min Huang

2. Oleen-m

3. Tetsuhito shizuka

و پارامتر تشخیص به طور معناداری تغییر نکرد. نتایج این تحقیق پیشنهاد می کند که سؤالات سه گزینه‌ای همان کارایی سؤالات ۴ گزینه‌ای را دارد و در اصل می توان به جای ۴ گزینه‌ای از سه گزینه‌ای استفاده کرد.

مایکل سی<sup>۱</sup> (۲۰۰۵) در تحقیقی بیان می کند که سؤالات چندگزینه‌ای بهترین کاربرد را در زمینه آزمون‌های پیشرفت تحصیلی دارند. و برای به دست آوردن سؤالاتی با محتوای عالی و با کیفیت سؤالات سه گزینه‌ای می تواند از سؤالات ۴ یا ۵ گزینه‌ای برتری داشته باشد. به گفته مایکل سی (۲۰۰۵) محققان در طی ۸۰ سال تحقیق مدارک تجربی به دست آورده‌اند که سؤالات ۳ گزینه‌ای بهترین کاربرد را دارد و نیاز به این سؤالات روز به روز باید مورد توجه قرار گیرد.

رالف جی استراتون<sup>۲</sup> (۱۹۸۰) به مقایسه آزمون سه گزینه‌ای و چهارگزینه‌ای پرداخته و بیان داشته است که هر چند بیشتر کتاب‌های اندازه گیری به طور معمول سؤالات ۵ یا ۴ گزینه‌ای را توصیه می کنند، اما مطالعات تجربی نشان داده است که سؤالات سه گزینه‌ای در شرایط خاص بهینه تر است. در این تحقیق پایایی و خطای استاندارد اندازه گیری آزمون سه گزینه‌ای برابر یا حتی بهتر از سؤالات ۴ گزینه‌ای و یا ۲ گزینه‌ای بود و این نتایج با در نظر گرفتن زمان آزمون بود. رالف (۱۹۸۰) پیشنهاد می کند از سؤالات سه گزینه‌ای در کلاس‌های عمومی استفاده شود.

## روش

باتوجه به هدف اصلی پژوهش که به بررسی تأثیر تعداد گزینه‌های سؤال در ویژگی‌های روان‌سنجی آزمون و توانایی برآورد شده می‌پردازد، بهترین روش تحقیق روش تجربی می‌باشد. جامعه آماری تحقیق حاضر شامل کلیه دانش آموزان دبیرستان‌های شهر شیراز و بخش کربال فارس می‌باشد.

1. Michael. c  
2. Ralph G. straton

گروه نمونه این تحقیق عبارت است از ۶۰۸ نفر دانش آموز سال سوم دبیرستان رشته ریاضی، به صورت تصادفی خوشه‌ای به علت وسعت منطقه جغرافیایی و پراکندگی جامعه آماری و عدم دسترسی به همه آنان صورت گرفت. برای همگون کردن آزمودنی‌ها در تحلیل، آزمودنی‌های دختر و پسر از هم تفکیک نشدند.

با توجه به این که کاربرد موفقیت آمیز مدل‌های نظریه سؤال پاسخ نیاز به حجم نمونه کافی و طول مناسب آزمون برای برآورد دقیق پارامترهای توانایی و پارامترهای سؤال دارد، بر طبق تحقیقات و پژوهش‌های انجام شده در مورد مدل‌های IRT، نمونه به صورت زیر انتخاب خواهد شد ۲۰ سؤال و ۲۰۰ آزمودنی برای مدل تک پارامتری، ۳۰ سؤال و ۵۰۰ آزمودنی برای مدل دو پارامتری و ۶۰ سؤال و ۱۰۰۰ آزمودنی برای مدل سه پارامتری توصیه می‌شود از طرف دیگر بعضی مطالعات حتی حجم نمونه ۲۰۰ نفر و ۲۰ سؤال را برای کسب برآوردهای قابل قبول کافی دانسته‌اند (همبلتون و کوک، ۱۹۸۳).

داده‌ها در تحقیق حاضر از طریق آزمون پیشرفت تحصیلی درس حسابان، و زبان عمومی در رشته ریاضی فیزیک سال سوم دبیرستان جمع آوری شد.

آزمون‌های موردنظر توسط جمعی از دبیران حسابان و هندسه و زبان عمومی شهرستان خرامه از سؤالات کنکور سال‌های قبل، ۲۰ سؤال تهیه شد. در ساخت این آزمون‌ها به نکات زیر توجه شد:

با توجه به زمان سپری شده از سال تحصیلی این آزمون‌ها از یک سوم کتاب تدریس شده تهیه شد. و به تایید گروه ریاضی و زبان آموزش پرورش شیراز و شهرستان خرامه رسید. بعد از ساخت سؤالات به صورت ۴گزینه‌ای توسط دبیران گزینه‌ای دیگر به سؤالات ۴گزینه‌ای اضافه شد و سؤالات به صورت ۵گزینه‌ای، از سؤالات ۴گزینه‌ای یک گزینه غلط (گزینه انحرافی) حذف گردید و سؤالات ۳گزینه‌ای تهیه شد. دراصل در تمام فرمت‌های ۳ و ۴ و ۵گزینه‌ای گزینه صحیح و سؤال مشابه بود. روایی آزمون به صورت محتوایی و ظاهری توسط دبیران حسابان

وزبان شهرستان خرامه بررسی شد و در یک مدرسه به عنوان نمونه اجرا شد و پایایی سؤالات با روش آلفای کرونباخ مشخص و سؤالات مناسب جایگزین سؤالات نامناسب شد.

### یافته‌ها

به منظور ارزیابی تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی آزمون‌ها و سؤالات و همچنین توانایی برآورد شده آزمودنی‌ها در آزمون حسابان و زبان در نظریه کلاسیک اندازه گیری و پرسش پاسخ، داده‌ها با استفاده از دو آزمون حسابان و زبان که هر کدام با فرمت ۳ و ۴ و ۵ گزینه‌ای ساخته شده بودند، از گروه نمونه‌ای که از ۱۲۰ دختر و ۴۹۱ پسر (۶۰۸ نفر به آزمون زبان و ۶۱۱ نفر به آزمون حسابان) تشکیل شده بودند و از جامعه به صورت تصادفی انتخاب شده بودند جمع آوری شد. ابتدا پایایی آزمون‌ها از طریق آلفای کرونباخ سنجیده شد و سپس با استفاده از تحلیل عاملی به بررسی مفروضه تک بعدی بودن آزمون پرداختیم. بعد از بررسی مفروضات اولیه به تحلیل نتایج پرداخته و ضرایب تمیز و دشواری و حدس آزمون‌ها محاسبه شد و با استفاده از Z پارامترها با هم مقایسه شدند. برای مقایسه پارامترها از منطق DIF و از آنالیز واریانس آمیخته نیز استفاده شد.

### جدول ۱. جدول پایایی کل آزمون‌های حسابان و زبان

تعداد سؤالات	آزمون و تعداد گزینه‌ها	آلفای کرونباخ (کودریچارد سون)
۱۷	حسابان ۳ گزینه‌ای	۰.۷۵۵
۱۷	حسابان ۴ گزینه‌ای	۰.۷۵۳
۱۷	حسابان ۵ گزینه‌ای	۰.۷۹۸
۱۷	زبان ۳ گزینه‌ای	۰.۷۷۴
۱۷	زبان ۴ گزینه‌ای	۰.۷۸۰
۱۷	زبان ۵ گزینه‌ای	۰.۷۶۲

به منظور بررسی پیش فرض‌های نظریه پرسش پاسخ از تحلیل عاملی برای بررسی تک بعدی بودن آزمون‌ها استفاده شد با توجه به ارزش‌های ویژه به دست آمده و اسکرین‌های به دست آمده از نرم افزار SPSS آزمون‌های مورد استفاده تک بعدی بودند. همپلتون (۱۹۸۹) ذکر می‌کند که چنانکه فرض تک بعدی بودن برقرار باشد. مفروضه استقلال موضعی نیز برقرار است. با توجه به این مطلب در آزمون مذکور مفروضه ی استقلال موضعی نیز برقرار است و آزمون‌ها مناسب برای تحلیل با نظریه پرسش پاسخ می باشد.

بعد از بررسی پیش فرض‌های آماری داده‌ها با نرم افزار BILOG-MG تحلیل شد و ضرایب تشخیص و تمیز در دو نظریه کلاسیک و IRT محاسبه شد.

برای بررسی حساسیت پارامترهای سؤال نسبت به تعداد گزینه‌ها از ۳ روش، برای محاسبه

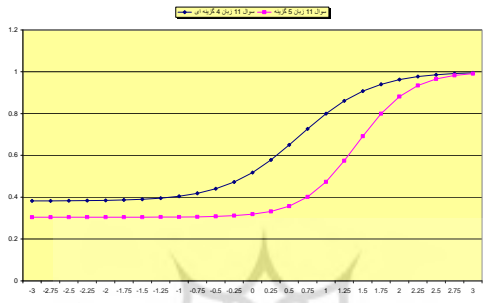
استفاده شد:

$$Z = \frac{|b_1 - b_2|}{\sqrt{S_{E_1}^2 - S_{E_2}^2}} \quad Z \text{ - اولین روش استفاده از فرمول}$$

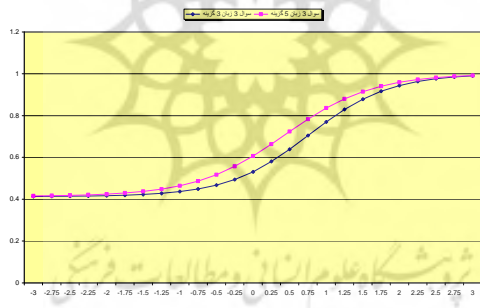
- دوم استفاده از منطق DIF در تحلیل پارامترهای سؤال

تعریف اساسی نظریه سؤال پاسخ از کار کرد افتراقی سؤال، روش بسیار ساده‌ای فراهم می‌کند که براساس آن می‌توان کارکرد افتراقی سؤال را مورد بررسی قرار داد. یعنی، انتخاب نمونه بزرگی از آزمودنی‌ها از دو جامعه، اجرای مقیاس موردنظر (منظور تبدیل داده‌ها به مقیاسی مشترک می‌باشد) و برآورد پارامترهای سؤال به طور جداگانه برای هر دو گروه و سپس مقایسه دیداری منحنی‌های سؤال پاسخ. تنها در صورتی پارامترهای سؤال (یا نمره‌های به دست آمده از سطوح صفت) که به طور جداگانه برای دو یا چند گروه برآورد شده‌اند قابل مقایسه‌اند که محقق این پیش فرض را بپذیرد که گروه‌ها در متغیر صفت مکنون توزیع یکسانی دارند. به طور اخص، سؤال در صورتی کارکرد افتراقی را نشان می‌دهد که منحنی سؤال پاسخ آن در بین گروه‌های مختلف متفاوت باشند و یا اینکه بین هریک از پارامترهای سؤال در بین

گروه ها تفاوت و وجود داشته باشد. (فلسفی نژاد، ۱۳۸۸) ملاک DIF در این مورد تفاوت ICC ها و قدر مطلق  $|P(\theta)_i - P(\theta)_{i2}|$  می باشد.  
 - سوم استفاده از آنالیز واریانس آمیخته  
 در زیر نمونه‌ای از ICC هایی که به روش منطق DIF بررسی شده اند آورده شده است.



نمودار ۱. بیشترین تفاوت در سؤال ۱۱ زبان ۴ و ۵ گزینه‌ای



نمودار ۲. کمترین تفاوت در سؤال ۲ زبان ۳ و ۵ گزینه‌ای

نتایج بررسی ICC ها نشان داد تفاوت فقط در چند سؤال قابل توجه است و اکثر ICC سؤالات تفاوت زیادی با هم ندارند. این نتایج با استفاده از تبدیل پارامترها به Z نیز تأیید شد.

جدول ۲. استفاده از قدر مطلق Z برای بررسی تفاوت ضرایب دشواری در آزمون زبان

زبان	Z ضرایب دشواری ۳ و ۴ گزینه‌ای	Z ضرایب دشواری ۴ و ۵ گزینه‌ای	Z ضرایب دشواری ۳ و ۵ گزینه‌ای
سؤال ۱	۱.۱۶۲	۲.۰۲۶	۰.۷۲۰۲
سؤال ۲	۲.۰۷۲۸	۰.۵۵۷۷	۰.۷۵۴۸
سؤال ۳	۱.۲۹۴	۱.۲۸۹	۰.۸۳۹
سؤال ۴	۱.۰۳۲	۰.۶۵۲۴	۰.۷۳۶۶
سؤال ۵	۲.۱۰۹	۱.۲۶۸۵	۰.۹۴۲۵
سؤال ۶	۱.۲۹۳۶	۱.۵۱۳۰	۰.۹۹۷۰
سؤال ۷	۰.۱۳۳۴	۱.۳۱۹۸	۰.۹۵۶۹
سؤال ۸	۰.۸۴۱۱	۰.۵۵۲۹	۰.۸۵۱۰
سؤال ۹	۱.۱۵۲	۱.۲۴۱	۰.۹۶۸۲
سؤال ۱۰	۲.۰۸۷	۰.۴۴۳۳	۰.۸۴۸۹
سؤال ۱۱	۱.۲۶۸۴	۰.۶۰۵۱	۰.۴۲۴۹
سؤال ۱۲	۱.۰۲۱۱	۱.۲۲۷۷	۰.۸۵۵۳
سؤال ۱۳	۱.۰۶۸۱	۰.۵۲۴۰	۰.۸۷۳۰
سؤال ۱۴	۲.۴۵۷	۶۳۸۷.۱	۰.۸۶۸۸
سؤال ۱۵	۱.۱۰۰۶	۱.۰۳۶۹	۰.۹۶۸۲
سؤال ۱۶	۰.۰۲۹۹	۱.۰۷۳	۰.۷۷۷
سؤال ۱۷	۱.۳۶۵۴	۱.۲۱۹۹	۰.۹۴۸۵

جدول، تفاوت ضرایب دشواری آزمون‌های زبان ۳ و ۴ و ۵ گزینه‌ای را نشان می‌دهد. ستون اول سمت راست تفاوت ضرایب دشواری زبان ۳ و ۵ گزینه‌ای و در ستون دوم زبان ۴ و ۵ گزینه‌ای و در ستون سوم زبان ۳ و ۴ گزینه‌ای با استفاده از Z محاسبه شد. با توجه به Z جدول در سطح معنی داری ۰.۰۵ تفاوت بین ضرایب دشواری آزمون زبان ۳ و ۴ و ۵ گزینه‌ای فقط در سؤالات ۲ و ۱۴ آزمون زبان ۳ و ۴ گزینه‌ای و سؤال ۱ از آزمون ۴ و ۵ گزینه‌ای مشاهده شد.

برای بررسی تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی سؤالات و توانایی برآورد شده آزمودنی‌ها در نظریه کلاسیک و سؤال پاسخ از تحلیل واریانس آمیخته (mixed) استفاده شد که نتایج تحلیل آزمون زبان در جداول زیر آورده شده است.

جدول ۳. خروجی آنالیز واریانس آمیخته برای بررسی تفاوت ضریب دشواری زبان در IRT,CTT

sig	F	ms	Df	Ss	منابع تغییر	
۰.۴۱۸	۱.۵۶۱	۰.۵۱۴	۲	۱.۰۲۸	تعداد گزینه‌ها	درون گروهی
۰.۰۴۲	۳.۳۴۰	۱.۱	۲	۲.۲	تعامل	
		۰.۳۲۹	۶۴	۲۱.۰۸۰	خطا	
۰.۰۰۰	۴۱.۴۰۷	۲۰.۲۰۷	۱	۲۰.۲۰۷	مدل	بیرون گروهی
		۰.۴۸۸	۳۲	۱۵.۶۱۷	خطا	

جدول ۴. خروجی آنالیز واریانس آمیخته برای بررسی تفاوت ضریب تمیز زبان در IRT,CTT

ضریب تمیز زبان در CTT,IRT

sig	f	ms	df	Ss	منابع تغییر	
۰.۴۹۵	۰.۷۱۲	۰.۸۲۷	۲	۱.۶۵۳	تعداد گزینه‌ها	درون گروهی
۰.۴۵۲	۰.۸۰۴	۰.۹۳۴	۲	۱.۸۶۷	تعامل	
		۱.۱۶۱	۶۴	۷۴.۳۳	خطا	
۰.۰۰۰	۲۶.۴۹۸	۳۰.۷۶۱	۱	۳۰.۷۶۱	مدل	بیرون گروهی
		۱.۱۶۱	۳۲	۱۷.۱۹۹	خطا	





مقایسه پارامترها با استفاده از فرمول  $Z$ : ضریب تمیز، ضریب دشواری، ضریب حدس، هر سؤال دو به دو به  $Z$  تبدیل شده و با هم مقایسه شدند. با توجه به  $Z$  محاسبه شده و  $Z$  بحرانی در سطح ۰.۰۵ تفاوت معنی داری بین ضرایب دشواری و تمیز و حدس در آزمون‌های ۳ و ۴ و ۵ گزینه‌ای مشاهده نشد.

نتایج تحلیل واریانس آمیخته نیز نشان داد که بین ضرایب تمیز و دشواری و آزمون‌های ۳ و ۴ و ۵ گزینه‌ای تفاوت معنی داری وجود ندارد. برای بهتر نشان دادن این موضوع از منطقی DIF استفاده شد تفاوت مشاهده شده بین ICCها، فقط در تعداد اندکی از سؤالات قابل توجه بودند و بقیه سؤالات تفاوت زیادی باهم نداشتند.

نتایج این تحقیق با تحقیقات «تسو هیتو شیزوکا (۲۰۰۶) مبنی بر این که بین ضرایب دشواری و تمیز سؤالات ۳ و ۴ گزینه‌ای تفاوتی وجود ندارد» همسو می‌باشد.

سؤال دوم: آیا تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی سؤالات و توانایی برآورد شده آزمودنی‌ها در نظریه کلاسیک و سؤال پاسخ متفاوت است؟  
برای بررسی تأثیر تعداد گزینه‌ها بر ویژگی‌های روان‌سنجی سؤالات و توانایی برآورد شده آزمودنی‌ها در نظریه کلاسیک و سؤال پاسخ از تحلیل واریانس آمیخته (MIXED) استفاده شد. نتایج تحلیل نشان داد که پارامترهای محاسبه شده در نظریه کلاسیک و پرسش پاسخ باهم متفاوت بودند.

این نتایج با تحقیق احمدی آذر (۱۳۸۷)، مبنی بر «تفاوت در برآورد پارامترهای سؤال در نظریه سؤال پاسخ و کلاسیک اندازه گیری» همسو می‌باشد.

سؤال ۳: آیا توانایی برآورد شده آزمودنی‌ها از تعداد گزینه‌های سؤال تأثیر می‌پذیرد؟  
برای بررسی این سؤال توانایی بدست آمده از هر سه فرمت آزمون، از خی ۲ برای بررسی تفاوت استفاده شد. خی ۲ محاسبه شده معنی دار بود و تفاوت در برآورد توانایی در فرمت‌های آزمون‌های ۳ و ۴ و ۵ گزینه‌ای مشاهده شد. با توجه به نمودارهای فراوانی توانایی به دست آمده و درصد پاسخ‌های صحیح افراد و توابع آگاهی آزمون‌های مختلف نتایج زیر نیز حاصل شد.

افراد با توانایی کم و متوسط در آزمون‌های ۳ گزینه‌ای و ۴ گزینه‌ای عملکرد بهتری داشتند و افراد با توانایی ضعیف در آزمون‌های ۵ گزینه‌ای عملکرد بدتری داشتند. بررسی درصد پاسخ‌های صحیح افراد نشان داد در آزمون‌های ۳ گزینه‌ای درصد پاسخ‌های صحیح به مراتب بالاتر از فرمت‌های دیگر بود همچنین در بررسی توابع آگاهی فرمت‌های مختلف آزمون مشخص شد آزمون‌های ۳ گزینه‌ای به اندازه آزمون‌های ۴ گزینه‌ای آگاهی دهنده توانایی افراد است. شاید بتوان این را مطرح کرد که آزمون‌های ۳ گزینه‌ای به همان مقدار آزمون‌های ۴ گزینه‌ای، توانایی را می‌سنجند و می‌توان به جای یکدیگر به کار برد.

سؤال ۴: آیا تعداد گزینه‌ها بر برازش با مدل در نظریه سؤال پاسخ تأثیر می‌گذارد؟

برای بررسی برازش مدل در نظریه پرسش پاسخ از نرم افزار BILOG-MG مقدار -2Loglikelihood در آخرین چرخش استخراج کرده و برای هر سه مدل آن را با خی ۲ بحرانی با درجه آزادی برابر با تعداد سؤال‌ها مقایسه کردیم.

(اگر خی ۲ مشاهده شده از خی ۲ بحرانی کوچکتر باشد، معلوم می‌شود که به لحاظ آماری بین برازش مدل‌ها تفاوت معناداری وجود ندارد و میتوان هریک را به جای دیگری به اختیار گزارش کرد. ولی براساس اصل امساک بهتر است مدل دارای پارامترهای کمتر را انتخاب کرد.)

با توجه به  $-2\log \text{likelihood}$  به دست آمده و خی ۲ جدول با درجه آزادی ۱۷ که برابر است با ۲۷/۵۹ نتایج زیر به دست آمد: آزمون‌های زبان و حسابان در فرمت‌های ۳ و ۴ و ۵ گزینه‌ای با مدل دو پارامتری برازش بهتری داشت. این نشان دهنده این است که می‌توان از پارامتر حدس در آزمون‌های چند گزینه‌ای صرف نظر کرد.

## منابع فارسی

- احمدی، آذر. (۱۳۸۷). نمره کل سازی با استفاده از تکنیک‌های نظریه کلاسیک و مقایسه آن با مدل‌های نظریه سؤال پاسخ در کنکور کارشناسی رشته ریاضی. پایان نامه کارشناسی ارشد. تهران: دانشگاه علامه طباطبایی.
- آلن ام، جی وین دبلیو ام. (۱۳۸۴). مقدمه‌ای بر نظریه‌های اندازه گیری (روان سنجی)، ترجمه علی، دلاور، تهران: سمت (تاریخ انتشار به زبان اصلی ۱۹۸۲).
- بیرقی، ابولفضل. (۱۳۷۷). بررسی اثر تعداد گزینه‌ها و زمان بر روی نمره کل در آزمون‌های پیشرفت تحصیلی. پایان نامه کارشناسی ارشد. تهران: دانشگاه علامه طباطبایی.
- ثرنادیک، برت ال. (۱۳۶۹). روان سنجی کاربردی. ترجمه هومن، حیدرعلی دانشگاه تهران.
- سیده مؤمنی، سید طاهره. (۱۳۷۹). تأثیر عامل حدس بر نمره‌های آزمون‌های چندگزینه‌ای پیشرفت تحصیلی. پایان نامه کارشناسی ارشد. تهران: دانشگاه علامه طباطبایی.
- سیف، علی اکبر. (۱۳۶۵). اندازه گیری پیشرفت تحصیلی، تهران: انتشارات آگاه.
- شریفی حسن پاشا، اصول روان سنجی در روان آزمایی، انتشارات رشد (۱۳۸۴)
- فتوحی، لیلا. (۱۳۸۷). بررسی کارکرد افتراقی سؤالات کنکور کارشناسی ارشد رشته روان‌شناسی. فلسفی نژاد، محمد رضا. (۱۳۸۸). مبانی نظری پرسش پاسخ. زیر چاپ.
- کاپلان، روبرت ام؛ ساکوز، دنیس پ. (۲۰۰۴). روان آزمایی، مترجم علی، دلاور، فریبرز درتاج، نورعلی فرخی، تهران: نشر ارسباران، ۱۳۸۶.
- معلمی اوره، مهرناز. (۱۳۸۷). مقایسه دقت برآورد توانایی در سؤالات چند گزینه‌ای با بکارگیری مدل سازی - سؤال پاسخ دو و چند ارزشی. پایان نامه کارشناسی ارشد. تهران: دانشگاه علامه طباطبایی.
- هادیان، مینا. (۱۳۷۵-۷۶). تأثیر گزینه‌های سؤال در اعتبار تست‌های چندگزینه‌ای. کتابنامه ص ۱۱۱-۱۰۸.



## منابع لاتین

- Chien-chi-yeh (2007). *The effect of gussing on assessing dimentionality in multiple choice test,university of Pittsburgh.*
- Educational and psychological measurement ,vol,40,no,2,357-365/(1980) SAGE publication /a comparison of the two ,three and four choice item test given/fixed total number of choice /RALPH G,STRATON
- James ware&marve(2008).*impact of item-writing flaws in multiple choice question on student achievement in high-stakes nursing,assessment.42:198-20*
- Karyn woodfard ,peter Bancroft,(2009).*multiple chioice questions not considered harmful,queensland university of technology.*
- Kobinger&Christian.h.ghottschall(2008).*item difficulty of multiple choice test dependant on different ietm response formats,university of Vienna.*
- Olleen m,(2008).heffernan.university of Alberta.
- Michael c rodriguez (2005),*tree option are optimal for multiple -choice item/uni/washin.vol;24,lss,2;pg3,11pgs.*
- TOM burkard.uk.cps,januray (2009).*multiplechoice.*
- Youngsuk suh (2008),*nested logit models for multiple choice item response data.university of Wisconsin-madison..*
- Tetsuhito shizuka and coworkers(2006)*a comparison of three and four option English test for university entrance selection,purposes in japan.language testing (2006);23;35.*
- YI MIN huang.(2004),*the impact of the all-of-the above option and student ability in multiple choice testing,Washington state university college of education.*