

## کارکرد افتراقی سؤال‌های پایه هشتم آزمون ریاضی تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر با استفاده از رویکرد نظریه سؤال - پاسخ<sup>۱</sup> (IRT)

اصغر مینایی<sup>۲</sup>

زهرا غفاری<sup>۳</sup>

تاریخ دریافت: ۹۳/۱۱/۲

تاریخ پذیرش: ۹۴/۲/۲۰

### چکیده

عمده‌ترین نگرانی که در ناعادلانه بودن آزمون‌ها مطرح می‌شود امکان وجود سوگیری یا کارکرد افتراقی است، زیرا سوگیری باعث می‌شود که روایی آزمون مورد شک و تردید قرار گیرد. در این پژوهش کارکرد افتراقی سؤال‌های ریاض پایه هشتم تیمز ۲۰۱۱ در بین دو گروه دانش‌آموزان دختر (گروه هدف) و پسر (گروه مرجع) ایرانی بر اساس رویکرد IRT مورد بررسی قرار گرفت. برای این منظور ابتدا مفروضه تک‌بعدی بودن سؤال‌های هر کدام از بلوک‌ها با استفاده از نرم افزار NOHARM بررسی گردید. در گام تحلیل‌ها نشان داد که از بین ۳ مدل یک، دو و سه پارامتری، مدل دو پارامتری از بهترین برازش با سؤال‌های هر بلوک برخوردار است و لذا این مدل به عنوان مدل پایه انتخاب گردید. سپس با استفاده از نرم افزار IRTLRDIF (تیسن، ۲۰۰۱) سؤال‌های دارای DIF شناسایی گردیدند. یافته‌ها نشان می‌دهد که از ۲۱۹ سؤال آزمون ریاضی پایه هشتم تیمز ۲۰۱۱، ۱۴۴ سؤال لنگر (فاقد DIF) و ۷۵ سؤال دارای DIF است. از بین این ۷۵ سؤال، ۵۶ سؤال دارای DIF یکنواخت و ۱۹ سؤال دارای DIF غیر یکنواخت است. در واقع یافته‌ها نشان می‌دهد که ۵۶ سؤال به ضرر دانش‌آموزان دختر عمل می‌کنند.

واژگان کلیدی: تیمز، تک بعدی، سوگیری، کارکرد افتراقی سؤال، نظریه سؤال-پاسخ

۱. این مقاله از پایان نامه کارشناسی ارشد نویسنده دوم اقتباس شده است.

۲. استادیار سنجش و اندازه‌گیری دانشگاه علامه طباطبائی (نویسنده مسئول) asghar.minaei@yahoo.com

۳. کارشناس ارشد سنجش و اندازه‌گیری دانشگاه علامه طباطبائی

## مقدمه

مساله تفاوت‌های جنسیتی در ریاضیات توجه زیادی را در محافل علمی و غیرعلمی به خود جلب کرده است. بخش عمده‌ای از این توجهات ناشی از این است که ریاضیات دروازه ورود به مشاغل و حرفه‌های پردرآمد و با پرستیژ مانند علوم، تکنولوژی، مهندسی و استادی دانشگاه است (سی و ویلیامز، ۲۰۰۷، نقل از استوت و گی‌یری<sup>۱</sup>، ۲۰۱۲). یک ایده کلیشه‌ای بسیار معروف در زمینه تفاوت‌های جنسیتی در زمینه ریاضیات این است که زنان در مقایسه با مردان از توانایی ریاضی کمتری برخوردارند (اسپنسر، استیل و کوئین<sup>۲</sup>، ۱۹۹۹). نتایج حاصل از سنجش‌های مختلف به نظر می‌رسد که از این دیدگاه حمایت می‌کنند (برای مثال به هریس و کارلتون<sup>۳</sup>، ۱۹۹۳؛ ویلدر و پاول، ۱۹۸۹، مراجعه کنید). برخی از پژوهش‌ها نشان می‌دهد که قالب سؤال بر عملکرد دختران و پسران در آزمون‌های ریاضی تاثیر می‌گذارد (بولگر و کلاهان<sup>۴</sup>، ۱۹۹۰؛ بلر و گافنی<sup>۵</sup>، ۲۰۰۰). برای مثال، بولگر و کلاهان (۱۹۹۰) دریافته‌اند که زنان در پرسش‌های باز پاسخ و مردان در سؤال‌های چند گزینه‌ای از عملکرد بهتری برخوردارند. پژوهش‌ها همچنین نشان می‌دهد که محتوای سوال‌ها عامل دیگری است که بر عملکرد دختران و پسران در ریاضیات تاثیر می‌گذارد. برای مثال، لانگفیلد<sup>۶</sup> (۱۹۹۷) دریافت که در سؤال‌های جبر، زنان بهتر از مردان عمل می‌کنند در حالی که در هندسه و حل مساله، مردان در مقایسه با زنان از عملکرد بهتر برخوردارند. در یک مطالعه دیگر، دولیتل<sup>۷</sup> (۱۹۸۹) نشان داد در حالی که عملکرد زنان در موضوعات الگوریتمی و سؤال‌های عملیات محور، همسطح مردان است، لکن در استدلال و هندسه از عملکرد پایین‌تری برخوردارند.

درباره تفاوت‌های جنسیتی در آزمون‌های ریاضی تیمز نیز مطالعات متعددی صورت گرفته است که از آن جمله می‌توان به ما<sup>۸</sup> (۱۹۹۵)، فیروس<sup>۹</sup> (۱۹۹۹)، مولیس، مارتین،

- 
1. Stoet & Geary
  2. Spencer, Steele & Quinn
  3. Harris & Carlton
  4. Bolger & Kellaghan
  5. Beller & Gaffni
  6. Langefeld
  7. Doolittle
  8. Ma
  9. Fierros

فیروس، گولدبرگ و استملر<sup>۱</sup> (۲۰۰۰) و دیندیال<sup>۲</sup> (۲۰۰۸) اشاره نمود. فیروس (۱۹۹۹) با تحلیل داده‌های تیمز دریافت که بین عملکرد دختران و پسران در ۱۲ کشور تفاوت معناداری وجود دارد. مولیس و همکاران (۲۰۰۰) بر اساس داده‌های تیمز ۱۹۹۵ عملکرد دانش‌آموزان پسر ژاپنی و کره‌ای را با عملکرد هم‌تایان دختر، مقایسه کردند. آن‌ها دریافتند که دانش‌آموزان پسر در هر دو پایه چهارم و هشتم در مقایسه با هم‌تایان دختر از عملکرد بسیار بهتری برخوردارند. با این حال در دو کشور هنگ‌کنگ و سنگاپور عملکرد دختران و پسران تفاوت معنادار نداشتند. بطور کلی آن‌ها به این نتیجه رسیدند که در حوزه‌های محتوایی در سه کشور ژاپن، کره جنوبی و هنگ‌کنگ عملکرد دانش‌آموزان پسر در هر دو پایه چهارم و هشتم بالاتر از دانش‌آموزان دختر است.

در تعدادی از مطالعات عملکرد دانش‌آموزان دختر و پسر ایران در سنجش‌های بین‌المللی تیمز مورد مقایسه قرار گرفته است. به عنوان مثال در تحلیل‌هایی که توسط مولیس، مارتین، گونزالس و کروتوسکی<sup>۳</sup> (۲۰۰۴، ص ۳۴) بر روی داده‌های تیمز ۲۰۰۳ صورت گرفت مشخص شد که بین عملکرد دانش‌آموزان پسر و دختر ایرانی در آزمون ریاضیات تفاوت وجود دارد. در واقع در حیطه‌های اعداد، جبر و هندسه عملکرد دختران ایرانی بالاتر از پسران ایرانی و در حیطه اندازه‌گیری و داده‌ها عملکرد پسران بالاتر از دختران بود. در مطالعه‌ای که توسط حجازی و نقش (۱۳۸۶) بر اساس داده‌های تیمز ۲۰۰۳ صورت گرفت تفاوتی بین عملکرد دختران و پسران پایه هشتم در درس ریاضی مشاهده نگردید.

پژوهشگران زیادی به پژوهش در زمینه عوامل تبیین‌کننده این تفاوت‌های جنسیتی دست زده‌اند (برای مثال، بن‌باو<sup>۴</sup>، ۱۹۸۸؛ دولیتل و کلیری، ۱۹۸۷؛ هریس و کارلتون، ۱۹۹۳). برخی از پژوهشگران (برای مثال، ویلینگهام و کول<sup>۵</sup>، ۱۹۹۷) در تبیین این تفاوت‌های جنسیتی بر عوامل بیرونی مانند برنامه‌های درسی و شیوه‌های آموزش تاکید می‌کنند. برخی دیگر از پژوهشگران، بویژه متخصصان اندازه‌گیری، نگران عامل‌های درونی، مانند امکان‌سوگیری سئوال‌های آزمون هستند (لانگفلد، ۱۹۹۷). اندازه‌گیری معتبر و روا برای انجام پژوهش، بویژه پژوهش در حوزه تفاوت‌های جنسیتی از اهمیت اساسی برخوردار است، زیرا ابزاری که

1. Mullis, Martin, Fierros, Goldberg & Stemler
2. Dindyal
3. Mullis, Martin, Gonzalez & Chrostowski
4. Benbow
5. Willingham & Cole

نمره‌های فاقد سوگیری جنسیتی فراهم می‌آورد، پیش‌نیاز انجام پژوهش‌های بیشتر در زمینه عامل‌های بیرونی تاثیرگذار بر تفاوت‌های مشاهده شده در بین دو جنس است. بطورکلی در تبیین این تفاوت‌های جنسیتی و یافته‌های متناقض در خصوص عملکرد ریاضی دختران و پسران، علاوه بر دلایل اقتصادی، اجتماعی، فرهنگی و آموزشی می‌توان به عوامل مرتبط با آزمون‌ها و از جمله امکان کارکرد افتراقی سؤال‌ها اشاره نمود.

در خصوص کارکرد افتراقی سوالهای سنجش بین‌المللی تیمز مطالعات متعددی صورت گرفته است. به عنوان مثال می‌توان به ارسی کان<sup>۱</sup> (۱۹۹۸)، ارسی کان و کاه<sup>۲</sup> (۲۰۰۵)، یان<sup>۳</sup> (۲۰۰۵)، وو، لی و زامبو<sup>۴</sup> (۲۰۰۷)، فابر (۲۰۰۸)، مینایی (۱۳۹۱) و صادقی (۱۳۹۲) اشاره کرد. ارسی کان (۱۹۹۸) دریافت که ۱۸ سؤال (۲۶٪) از ۷۰ سؤال درس علوم TIMSS سال ۹۵ دارای DIF است که تعداد ۱۰ سؤال به نفع گروه مرجع که دانش آموزان انگلیسی زبان هستند و ۸ سؤال به نفع گروه هدف که فرانسوی زبان بودند عمل می‌کنند. فابر<sup>۵</sup> (۲۰۰۸) کارکرد افتراقی سؤال‌های پایه هشتم تیمز ۲۰۰۳ را با استفاده از رویکرد اندازه‌گیری چند وجهی راش در بین دختران و پسران مورد بررسی قرار داد. نتایج نشان داد که سؤال‌های کسر و اندازه‌گیری به ترتیب ۰/۰۶ و ۰/۱۸ واحد لوجیت برای دختران دشوارتر از پسران بود. سؤال‌های هندسه و جبر به ترتیب ۰/۱۱ و ۰/۱۷ واحد لوجیت برای پسران دشوارتر از دختران بودند. در ارتباط با کارکرد افتراقی سؤال‌ها، نتایج پژوهش فابر نشان داد که از پنج حوزه محتوایی، سؤال‌های دو حوزه اندازه‌گیری و جبر از DIF برخوردارند. سؤال‌های حوزه اندازه‌گیری بر علیه دختران و سؤال‌های حوزه جبر علیه پسران دارای DIF هستند. در ارتباط با تعامل بین جنسیت و قالب سؤال، یافته‌های مطالعه فابر نشان داد که در سؤال‌های چندگزینه‌ای، تفاوت قابل ملاحظه‌ای بین دو گروه دختر و پسر وجود ندارد. سؤال‌های باز پاسخ به اندازه ۰/۲ واحد لوجیت برای دختران آسان‌تر از پسران بودند که حاکی از این است که دختران در این گونه سؤال‌ها از عملکرد بهتری برخوردارند. با این حال، یافته‌های مربوط به کارکرد افتراقی نشان داد که DIF معناداری در سؤال‌های باز پاسخ در بین دو گروه دختر و پسر مشاهده نشد.

1. Ercikan, K. & Koh
2. Koh
3. Yan
4. Wu, Li & Zumbo
5. Faber

صادقی (۱۳۹۲) کارکرد افتراقی سئوال‌های دفترچه شماره پنج درس ریاضی مربوط به هر دو پایه چهارم و هشتم تیمز ۲۰۱۱ را در بین پسران و دختران ایران مورد مطالعه قرار داد. او برای انجام تحلیل‌ها از روش‌های کلاسیک و IRT استفاده نمود. نتایج مطالعه صادقی نشان داد که بر اساس روش کلاسیک ۶ سئوال و در رویکرد IRT، یک سئوال دارای DIF است. فهم تفاوت‌های جنسیتی در ریاضیات برای پژوهشگران، معلمان و پرورش‌دهندگان آزمون‌ها جهت دستیابی به اهداف و غایت‌های آموزش و پرورش از اهمیت اساسی برخوردار است. از این رو شناسایی سوگیری یا DIF جنسیتی سئوال‌های ریاضی به پرورش‌دهندگان آزمون‌ها در پرورش سنجش‌های عادلانه و همچنین در استنتاج‌های معتبر و روا از نمره‌های آزمون‌ها کمک می‌کند. با توجه به مطالبی که در خصوص اهمیت مطالعه کارکرد افتراقی سئوال‌های آزمون‌های ریاضی پایه هشتم تیمز ۲۰۱۱ را در بین دانش‌آموزان دختر و پسر ایران مورد مطالعه قرار نداده است، بنابراین در این پژوهش قصد داریم تا کارکرد افتراقی کلیه سئوال‌های ریاضی پایه هشتم تیمز ۲۰۱۱ را در بین دختران و پسران ایران مورد بررسی قرار دهیم. در واقع در پژوهش حاضر تلاش می‌کنیم تا به سئوال‌های زیر پاسخ دهیم.

۱- آیا سئوال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱، تک‌بعدی است؟

۲- مناسب‌ترین مدل که با سئوال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ برازش دارد کدام است؟

۳- آیا سئوال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر دارای کارکرد افتراقی است؟

### روش پژوهش

**جامعه، نمونه و روش نمونه‌گیری:** جامعه آماری و گروه نمونه پژوهش حاضر همان جامعه و گروه نمونه پایه هشتم یعنی جمعیت ۲ در آزمون ریاضی تیمز ۲۰۱۱ است که در سال تحصیلی ۱۳۹۰ - ۱۳۸۹ به اجرا در آمده است. حجم نمونه کل دانش‌آموزان ایران ۶۰۲۹ نفر بوده است که از این تعداد ۲۸۱۶ نفر (۴۶/۷٪) را دختران و ۳۲۱۳ نفر (۵۳/۳٪) را پسران تشکیل می‌دهند.

چارچوب نمونه گیری در تیمز ۲۰۱۱ از طرح نمونه گیری خوشه‌ای طبقه‌ای دو مرحله‌ای متناسب با حجم نمونه استفاده شده است. در گام اول مدارس که حدود ۲۳۸ مدرسه بوده نمونه گیری به عمل آمده است و در گام دوم، کلاس در پایه هدف، نمونه گیری شده‌اند. در این طرح که به روش «احتمال متناسب با حجم» بیان می‌شود سهم هر یک از خوشه‌های نمونه را متناسب با حجم آن در جامعه تعیین می‌کند، برای مثال انتخاب شدن مدرسه‌ای با ۸۰ دانش آموز ۲ برابر مدرسه‌ای با ۴۰ دانش آموز است.

**ابزار اندازه گیری:** ابزار مورد استفاده در این پژوهش، آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ است. این آزمون دارای ۲۱۹ سؤال است که در ۱۴ دفترچه، که در واقع می‌توان آن‌ها را فرم‌های موازی در نظر گرفت ارائه شده‌اند. این دفترچه‌ها بصورت تصادفی در بین دانش آموزان توزیع شدند. از بین ۲۱۹ سؤال ریاضی در تیمز ۲۰۱۱ تعداد ۱۱۸ سؤال را سؤال‌های چند گزینه‌ای و مابقی سؤالات را سؤال‌های بازپاسخ یا پاسخ‌ساز تشکیل داده‌اند. سؤال‌های پاسخ‌ساز بصورت چندارزشی نمره گذاری می‌شوند. در پژوهش حاضر به منظور سهولت در تحلیل داده‌ها و به پیروی از مینایی (۱۳۹۱) سؤال‌های چندارزشی از طریق کدگذاری مجدد به سؤال‌های دوارزشی تبدیل شده است. بدین صورت که اگر فردی در یک سؤال چندارزشی بالاترین نمره را به دست آورده بود، نمره ۱ و چنانچه نمره کامل را نگرفته بود نمره صفر تعلق گرفت. همچنین سؤال‌هایی که فرد آن‌ها را خالی گذاشته بود که در چارچوب تیمز به این گونه سؤال‌ها بدون پاسخ<sup>۱</sup> گویند و همچنین سؤال‌هایی که بدلیل کمبود وقت دانش آموزان قادر به پاسخگویی به سؤال‌ها نبوده‌اند که در چارچوب تیمز سؤال‌های جا مانده<sup>۲</sup> نامیده می‌شود نمره صفر داده شد.

**طرح تحقیق:** از آنجایی که در این پژوهش از داده‌های آزمون ریاضی تیمز که در سال ۲۰۱۱ اجرا شده است استفاده می‌شود و با توجه به این که هدف پژوهش حاضر توصیف و مقایسه عملکرد سؤال‌ها در دو گروه یعنی دانش آموزان دختر و پسر ایرانی است، بنابراین می‌توان نتیجه گرفت که پژوهش حاضر جزو تحقیقات توصیفی - مقایسه‌ای از نوع تحلیل ثانویه به حساب می‌آید.

1. Omitted
2. Not reached

**روش تحلیل داده‌ها:** در تحلیل داده‌های تیمز به صورت‌های مختلفی عمل شده است. برخی از مطالعات (مانند تریسی و همکاران، ۲۰۰۷) صرفاً از روش‌های مبتنی بر IRT و برخی دیگر از پژوهشگران (مانند ارسی کان، ۱۹۹۸) از روش  $\chi^2$  متل هنسزل، و برخی نیز از تحلیل عامل تأییدی، و برخی دیگر مانند مینایی (۱۳۹۱) از هر دو روش IRT و CFA استفاده کرده‌اند.

بسیاری از متخصصان روش‌شناسی معتقدند که در تحلیل کارکرد افتراقی سؤال‌ها مطلوب است که از آزمون نسبت درستی‌مندی که متکی بر رویکرد IRT است استفاده شود. بنابراین در پژوهش حاضر از این روش‌شناسی که توسط تیسن، استنبرگ و جرارد (۱۹۸۶) و تیسن، استنبرگ و وایتر (۱۹۹۳) ارائه شده و از طریق نرم‌افزار IRTLRDIF (تیسن، ۲۰۰۱) قابل اجرا است استفاده گردید.

جهت بررسی کارکرد افتراقی سؤال‌های هر یک از بلوک‌ها علاوه بر تحلیل‌های مربوط به مفروضه ابعاد و برآورد پارامترهای سؤال‌ها، تحلیل‌های مربوط به DIF به تعداد سؤال‌های آزمون ریاضی (یعنی ۲۱۹ سؤال) تکرار گردید. در ادامه نحوه انجام این تحلیل‌ها توضیح داده می‌شود.

۱- جهت آزمون تک‌بعدی بودن سؤال‌های هر یک از بلوک‌ها و تعیین مدلی که بهترین برازش را با سؤال‌های دارد از روش تحلیل عاملی غیر خطی که از طریق نرم‌افزار NOHARM قابل اجرا است استفاده گردید.

۲- پس از اتمام مرحله ۱ و تعیین بهترین مدل، کلیه پارامترهای سؤال‌های هر کدام از بلوک‌ها بطور جداگانه در بین دو گروه مرجع<sup>۱</sup> و گروه هدف<sup>۲</sup> محدود شدند و مدل به داده‌ها برازش داده شد. هدف این مرحله شناسایی سؤال‌های لنگر، یعنی سؤال‌هایی که فاقد DIF هستند بود.

۳- پس از تعیین مجموعه سؤال‌های لنگر، بقیه سؤال‌های بلوک از لحاظ دارا بودن DIF مورد آزمون قرار گرفتند. این تحلیل برای هر کدام از بلوک‌ها بصورت جداگانه صورت گرفت. در این مرحله سؤال‌های لنگر به عنوان متغیر هم‌تاسازی بکار رفتند و دیگر از نظر DIF مورد مطالعه قرار نگرفتند. هدف این مرحله شناسایی نهایی سؤال‌های دارای DIF بود.

- 
1. Reference group
  2. Focal group

۴- پس از شناسایی سؤال‌های دارای DIF و تعیین نوع DIF (یکنواخت یا غیر یکنواخت)، یکبار دیگر پارامترهای کلیه سؤال‌ها بر اساس یک مدل دو گروهی برآورد شد. در این مرحله، اگر سؤالی بر اساس پارامتر a دارای DIF بود هر دو پارامتر a و b بین دو گروه آزاد گذاشته می‌شد و پارامترهای آن برای هر گروه بصورت جداگانه برآورد می‌شدند. اما اگر سؤالی دارای DIF یکنواخت بود صرفاً پارامتر b بین دو گروه آزاد گذاشته می‌شد و برای هر گروه به صورت جداگانه برآورد می‌گردد. هدف این مرحله، برآورد نهایی پارامترهای سؤال و توانایی افراد بود.

جهت تحلیل داده‌ها و پاسخگویی به سؤال‌های پژوهش از مدل لوجستیک ۲ پارامتری نظریه سؤال - پاسخ و برخی از شاخص‌های آمار توصیفی و استنباطی استفاده گردید. برای تحلیل داده‌ها از نرم‌افزارهای SPSS21، BILOG-MG، IRTLRF (تیسن، ۲۰۰۱)، MULTILOG و برای رسم نمودارها از EXCEL استفاده گردید.

### یافته‌های پژوهش

جدول ۱ اطلاعات جمعیت‌شناختی دانش‌آموزان شرکت‌کننده در آزمون ریاضی تیمز ۲۰۱۱ را به تفکیک جنسیت و بلوک نشان می‌دهد.

جدول ۱. اطلاعات جمعیت‌شناختی نمونه

جنسیت					
پسر			دختر		
تعداد کل	درصد	فراوانی	درصد	فراوانی	بلوک
۸۵۵	۵۳/۳	۴۵۶	۴۶/۶	۳۹۹	۱
۸۶۰	۵۲/۳	۴۵۰	۴۷/۶	۴۱۰	۲
۸۷۳	۵۳	۴۶۳	۴۶/۹	۴۱۰	۳
۸۶۵	۵۲/۹	۴۵۸	۴۷	۴۰۷	۴
۸۵۹	۵۳/۲	۴۵۷	۴۶/۷	۴۰۲	۵
۸۴۷	۵۳/۸	۴۵۶	۴۶/۱	۳۹۱	۶
۸۴۱	۵۳	۴۴۶	۴۶/۹	۳۹۵	۷
۸۶۹	۵۳/۱	۴۶۲	۴۶/۸	۴۰۷	۸
۸۷۳	۵۲/۸	۴۶۱	۴۷/۱	۴۱۲	۹
۸۶۹	۵۲/۷	۴۵۸	۴۷/۲	۴۱۱	۱۰
۸۶۱	۵۳/۵	۴۶۱	۴۶/۴	۴۰۰	۱۱

۸۵۰	۵۳/۵	۴۵۵	۴۶/۴	۳۹۵	۱۲
۸۵۸	۵۳/۶	۴۶۰	۴۶/۳	۳۹۸	۱۳
۸۵۷	۵۳/۹	۴۶۲	۴۶/۱	۳۹۵	۱۴

جدول ۲ تعداد سئوال‌ها و ضرایب همسانی درونی حیطه‌های محتوایی را که با استفاده از آلفای کرونباخ برآورد شده‌اند، نشان می‌دهد. حیطه محتوایی هندسه و داده‌ها و اطلاعات بلوک ۱۴ دارای یک سئوال است و همچنین حیطه محتوایی اعداد در بلوک ۶ نیز دارای یک سئوال است، لذا امکان برآورد ضریب آلفای کرونباخ وجود نداشت. بالا بودن ضریب همسانی درونی اکثر حیطه‌ها ناشی از تعداد کم سئوال‌ها است.

جدول ۲. تعداد سئوال‌ها و ضرایب همسانی درونی حیطه‌ها به تفکیک دختر و پسر

بلوک	حیطه محتوایی	تعداد سئوال	آلفای کرونباخ	
			دختر	پسر
۱	اعداد	۵	۰/۵۹	۰/۶۴
	جبر	۳	۰/۳۲	۰/۲۷
	هندسه	۴	۰/۵۶	۰/۵۸
	داده و اطلاعات	۳	۰/۳۱	۰/۳۳
۲	اعداد	۴	۰/۵۳	۰/۵۸
	جبر	۶	۰/۵۹	۰/۵۴
	هندسه	۳	۰/۴۳	۰/۳۲
	داده و اطلاعات	۴	۰/۴۴	۰/۴۳
۳	اعداد	۵	۰/۴۸	۰/۶۰
	جبر	۴	۰/۴۶	۰/۴۳
	هندسه	۳	۰/۲۷	۰/۳۹
	داده و اطلاعات	۳	۰/۲۶	۰/۲۱
۴	اعداد	۳	۰/۳۱	۰/۳۷
	جبر	۳	۰/۲۶	۰/۵۱
	هندسه	۶	۰/۳۶	۰/۵۱
	داده و اطلاعات	۲	۰/۳۸	۰/۳۸
۵	اعداد	۴	۰/۴۲	۰/۴۴
	جبر	۶	۰/۶۶	۰/۶۵
	هندسه	۴	۰/۷۹	۰/۷۹
	داده و اطلاعات	۴	۰/۳۸	۰/۴۱
۶	اعداد	۱	--	-

۰/۵۰	۰/۵۱	۶	جبر	
۰/۴۷	۰/۳۹	۴	هندسه	
۰/۳۸	۰/۲۸	۴	داده و اطلاعات	
۰/۵۷	۰/۶۴	۶	اعداد	
۰/۵۳	۰/۶۶	۴	جبر	۷
۰/۳۶	۰/۴۶	۲	هندسه	
۰/۴۲	۰/۴۱	۳	داده و اطلاعات	
۰/۴۹	۰/۵۴	۵	اعداد	
۰/۷۵	۰/۷۷	۹	جبر	۸
۰/۵۱	۰/۴۷	۳	هندسه	
۰/۴۱	۰/۳۵	۲	داده و اطلاعات	
۰/۶۶	۰/۶۱	۵	اعداد	
۰/۴۵	۰/۴۹	۴	جبر	۹
۰/۳۸	۰/۴۳	۳	هندسه	
۰/۴۷	۰/۴۸	۳	داده و اطلاعات	
۰/۶۹	۰/۶۹	۴	اعداد	
۰/۴۳	۰/۵۶	۶	جبر	۱۰
۰/۲۵	۰/۲۹	۳	هندسه	
۰/۳۸	۰/۲۹	۴	داده و اطلاعات	
۰/۴۴	۰/۴۶	۵	اعداد	
۰/۳۱	۰/۳۵	۳	جبر	۱۱
۰/۴۷	۰/۳۲	۳	هندسه	
۰/۵۲	۰/۴۶	۴	داده و اطلاعات	
۰/۵۱	۰/۵۹	۶	اعداد	
۰/۳۲	۰/۴۸	۶	جبر	۱۲
۰/۲۸	۰/۳۲	۲	هندسه	
۰/۴۴	۰/۵۲	۳	داده و اطلاعات	
۰/۵۷	۰/۵۳	۵	اعداد	
۰/۴۷	۰/۴۴	۵	جبر	۱۳
۰/۰۷	۰/۱۷	۳	هندسه	
۰/۳۱	۰/۲۶	۳	داده و اطلاعات	
۰/۵۱	۰/۴۷	۳	اعداد	
۰/۶۸	۰/۶۵	۶	جبر	۱۴
-	-	۱	هندسه	

-	-	۱	داده و اطلاعات
---	---	---	----------------

در جدول ۳ آماره‌های توصیفی حیطه‌های محتوایی بلوک‌ها به تفکیک جنسیت ارائه شده است. این آماره‌ها شامل میانگین، انحراف استاندارد، چولگی و کشیدگی است. بررسی داده‌های جدول نشان می‌دهد که توزیع نمره‌های اکثر حیطه‌های محتوایی از چولگی و کشیدگی زیادی برخوردار است و توزیع نمره‌ها به شکل نرمال نیست.

جدول ۳. آماره‌های توصیفی حیطه‌های محتوایی بلوک‌ها به تفکیک جنسیت

بلوک	حیطه	پسران				دختران			
		میانگین	انحراف استاندارد	چولگی	کشیدگی	میانگین	انحراف استاندارد	چولگی	کشیدگی
۱	اعداد	۱/۷۵	۱/۴۳	۰/۶۲	-۰/۴۹	۱/۶۵	۱/۵۰	۰/۸۲	-۰/۳۰
		۰/۶۷	۰/۶۵	۰/۸۲	۱/۲۳	۰/۶۱	۰/۶۸	۰/۹۵	+۰/۸۸
	هندسه	۱/۱۴	۱/۱۶	۰/۸۰	-۰/۳۰	۱/۲۳	۱/۲۱	۰/۷۶	-۰/۴۴
		داده و اطلاعات	۰/۵۴	۰/۷۲	۱/۳۸	۱/۷۴	۰/۵۸	۰/۷۲	۱/۱۹
	اعداد	۱/۴۵	۱/۱۷	۰/۶۸	-۰/۳۰	۱/۳۹	۱/۲۲	۰/۶۵	-۰/۴۶
		جبر	۱/۴۵	۱/۴	۰/۸۹	۰/۱۶	۱/۱۹	۱/۲۶	۱/۰۶
	هندسه	۰/۹۶	۰/۹۶	۰/۶۹	-۰/۵۲	۱	۰/۹۲	۰/۵۸	+۰/۵۶
		داده و اطلاعات	۱/۷۲	۱/۱۶	۰/۲۵	-۰/۷۵	۱/۷۴	۱/۱۵	۰/۱۷
	اعداد	۱/۳	۱/۱۷	۱/۰۲	۰/۷۷	۱/۴۹	۱/۳۷	۱/۰۴	+۰/۴۹
		جبر	۱/۰۶	۱/۰۴	۰/۹۷	۰/۵۲	۱/۰۹	۱/۰۳	۰/۹۰
	هندسه	۰/۸۴	۰/۷۹	۰/۵۳	-۰/۵۳	۰/۹۰	۰/۸۶	۰/۵۶	-۰/۵۸
		داده و اطلاعات	۰/۸۳	۰/۸۳	۰/۷۴	-۰/۱۰	۰/۸۶	۰/۸۱	۰/۶۸
اعداد	۰/۸۶	۰/۷۹	۰/۶۵	-۰/۱۱	۰/۸۴	۰/۸۶	۰/۷۸	-۰/۹۲	
	جبر	۰/۸۲	۰/۸۵	۰/۸۲	۰/۰۰۴	۰/۹۳	۰/۹۷	۰/۷۴	-۰/۵۲
هندسه	۲/۲۸	۱/۳۸	۰/۴۴	-۰/۲۷	۲/۳۶	۱/۵۷	۰/۵۰	-۰/۳۲	

	داده و اطلاعات	۰/۷۵	۰/۷۶	۰/۴۴	-۱/۱۴	۰/۷۴	۱/۷۶	۰/۴۸	-۱/۱۲
	اعداد	۱/۶۲	۱/۰۵	۰/۳	-۰/۴۹	۱/۶۸	۱/۱۱	۰/۲۷	-۰/۶۳
	جبر	۲/۱۵	۱/۶۲	۰/۶۵	-۰/۱۹	۲/۲۶	۱/۶۷	۰/۵۶	-۰/۵۰
۵	هندسه	۱/۴۹	۱/۴۶	۰/۶۹	-۰/۹۱	۱/۳۱	۰/۴۱	۰/۹۰	-۰/۵۰
	داده و اطلاعات	۱/۰۷	۰/۹۰	۰/۵۶	۰/۱۴	۱/۰۵	۰/۹۱	۰/۵۵	۰/۰۴
	اعداد	۰/۰۵	۰/۲۳	۳/۹۷	۱۳/۸۶	۰/۰۵	۰/۲۲	۳/۹۲	۱۳/۴۶
	جبر	۲/۷۱	۱/۴۹	۰/۲۳	-۰/۵۰	۲/۷۳	۱/۵۲	۰/۲۴	-۰/۶۶
۶	هندسه	۱/۴۰	۱/۱۱	۰/۵۵	-۰/۴۹	۱/۳۴	۱/۱۶	۰/۶۵	-۰/۳۴
	داده و اطلاعات	۰/۹۴	۰/۹۰	۰/۹۶	۰/۸۰	۱/۱۶	۱/۰۴	۰/۶۸	-۰/۲۰
	اعداد	۱/۸۲	۱/۵۷	۰/۸۱	-۰/۰۷	۱/۶۸	۱/۴۳	۰/۷۳	-۰/۲۳
	جبر	۰/۸۵	۱/۱۲	۱/۲۱	۰/۵۴	۰/۶۷	۰/۹۳	۱/۳۴	۱/۰۴
۷	هندسه	۱/۰۱	۰/۷۸	-۰/۰۲	-۱/۳۵	۰/۹۷	۰/۷۵	۰/۰۴	-۱/۲۴
	داده و اطلاعات	۱/۱۲	۰/۸۶	۰/۲۳	-۰/۸۲	۱/۰۶	۰/۸۷	۰/۳۲	-۰/۷۹
	اعداد	۱/۰۹	۱/۱۵	۱/۰۵	۰/۵۶	۱/۱۴	۱/۰۹	۰/۹۴	۰/۷۲
	جبر	۱/۴۸	۱/۸۸	۲/۱۵	۵/۰۲	۱/۴۸	۱/۸۳	۱/۹۳	۳/۹۹
۸	هندسه	۱/۳۶	۰/۹۳	۰/۱۸	-۰/۵۳	۱/۳۵	۰/۹۶	۰/۱۵	-۰/۹۵
	داده و اطلاعات	۰/۶۸	۰/۷۰	۰/۵۲	-۰/۸۶	۰/۷۸	۰/۷۵	۰/۳۷	-۱/۱۵
	اعداد	۱/۰۹	۱/۲۱	۱/۳۸	۱/۷۹	۱/۰۷	۱/۳۰	۱/۳۳	۱/۰۷
	جبر	۱/۱۹	۱/۸۳	۰/۶۸	-۰/۳۰	۱/۱۵	۱/۰۵	۰/۶۹	-۰/۲۶
۹	هندسه	۱/۰۵	۰/۹۷	۰/۵۰	-۰/۸۲	۱/۲۸	۰/۹۵	۰/۵۸	-۰/۶۲
	داده و اطلاعات	۱/۳۴	۱/۰۱	۰/۲۱	-۱/۰۴	۱/۴۰	۰/۹۹	۰/۱۵	-۱/۰۳
	اعداد	۱/۵۳	۱/۳۰	۰/۶۵	-۰/۶۲	۱/۹۲	۱/۳۹	۰/۱۸	-۱/۲
	جبر	۱/۲۴	۱/۲۹	۱/۴۴	۲/۲۲	۱/۱۶	۱/۱۲	۱/۱۷	-۱/۷۲
۱۰	هندسه	۱/۰۷	۰/۸۸	۰/۴۳	-۰/۵۸	۱/۱۳	۰/۹۰	۰/۳۳	-۰/۷۴
	داده و اطلاعات	۱/۶۹	۱/۰۱	۰/۲۸	-۰/۳۲	۱/۸۴	۱/۰۷	۰/۲۲	-۰/۶۲
	اعداد	۲/۰۷	۱/۲۸	۰/۳۷	-۰/۴۶	۲/۰۹	۱/۲۹	۰/۳۸	-۰/۵۱
۱۱	جبر	۰/۸۸	۰/۸۰	۰/۷۳	۰/۱۸	۰/۷۳	۰/۷۶	۰/۸۶	۰/۳۹

هندسه	۰/۵۵	۰/۷۴	۱/۲۵	۰/۹۹	۰/۶۸	۰/۸۶	۱/۰۹	۰/۳۰
داده و اطلاعات	۰/۹۷	۱/۰۱	۰/۸۶	-۰/۰۶	۱/۱۶	۱/۱۱	۰/۶۸	-۰/۴۶
اعداد	۱/۲۶	۱/۲۱	۱/۱۸	۱/۴۳	۱/۲۱	۱/۱۱	۱/۲۹	۲/۷۲
جبر	۲/۱۶	۱/۴۳	۰/۶۰	-۰/۲۲	۲/۰۵	۱/۲۸	۰/۶۷	۰/۳۸
هندسه	۰/۴۹	۰/۶۶	۱/۰۲	-۰/۱۶	۰/۴۸	۰/۶۵	۱/۰۳	-۰/۰۷
داده و اطلاعات	۰/۷۳	۰/۸۳	۱/۱۲	۰/۸۲	۰/۶۶	۰/۷۹	۱/۰۵	۰/۵۳
اعداد	۱/۵۴	۱/۲۴	۰/۷۳	۰/۰۲	۱/۶۸	۱/۳۵	۰/۶۳	-۰/۳۱
جبر	۱/۶۸	۱/۱۵	۰/۶۸	۰/۲۶	۱/۶۱	۱/۲۰	۰/۶۴	۰/۱۸
هندسه	۰/۷۶	۰/۷۸	۰/۸۰	۰/۱۱	۰/۷۸	۰/۷۶	۰/۶۶	-۰/۱۷
داده و اطلاعات	۱/۰۳	۰/۸۶	۰/۵۳	-۰/۳۶	۱/۰۹	۰/۹۱	۰/۳۹	-۰/۷۴
اعداد	۱/۲	۱/۰۲	۰/۳۳	۱/۰۴	۱/۲۵	۱/۰۵	۰/۳۵	-۱/۰۹
جبر	۳۲۱	۱/۳۸	۱/۳۳	۱/۴۹	۱/۳۳	۱/۴۲	۱/۳۶	۱/۴۲
هندسه	۰/۲۳	۰/۴۲	۱/۳۰	-۰/۳۰	۰/۲۳	۰/۴۱	۱/۳۲	-۰/۲۵
داده و اطلاعات	۰/۳۱	۰/۴۶	۰/۸۴	-۱/۲۹	۰/۳۲	۰/۴۶	۰/۷۹	-۱/۳۷

انجام تحلیل‌های DIF مستلزم انتخاب یکی از مدل‌های دوارزشی IRT به عنوان مدل پایه است. منظور از مدل پایه، مدلی است که با داده‌های ما برازش بهتری دارد. مهمترین مفروضه مدل‌های دوارزشی IRT، مفروضه تک‌بودن سؤال‌هاست. یکی از مناسب‌ترین روش‌های آزمون این مفروضه، انجام تحلیل‌عاملی در سطح سؤال است (مینایی و فلسفی نژاد، ۱۳۹۱). بنابراین پیش از انتخاب یکی از مدل‌های ۱، ۲ و ۳ پارامتری بعنوان مدل پایه، سؤال‌های هر یک از بلوک‌ها با استفاده از تحلیل‌عاملی غیرخطی که از طریق برنامه NOHARM (فریزر<sup>۱</sup>، ۱۹۸۸) قابل اجرا است مورد تحلیل قرار گرفت. به منظور افزایش حجم نمونه‌ها سؤال‌های مشترک در دفترچه‌ها با یکدیگر ادغام شدند و تبدیل به ۱۴ بلوک شدند. نتایج حاصل از این تحلیل‌ها نشان داد که شواهد کافی برای رد فرضیه تک‌بعدی بودن بلوک‌های آزمون ریاضی تیمز ۲۰۱۱ وجود ندارد. به دلیل کمبود فضا از آوردن جدول خودداری شده است. این یافته حاکی از این است که سؤال‌های تمام ۱۴ بلوک در آزمون

1. Fraser

ریاضی تیمز ۲۰۱۱ صرفاً یک سازه یا عامل مکنون را می‌سنجد و مفروضه تک‌بعدی بودن در مورد تمام آن‌ها صادق است.

به منظور انتخاب مدل پایه، مدل‌های لوجستیک ۱، ۲ و ۳ پارامتری به هر یک از بلوک‌ها و بصورت جداگانه با داده‌های آزمون ریاضی تیمز ۲۰۱۱ برازش داده شده و با استفاده از آماره نسبت درست‌نمایی ( $-2LL$ ) که با  $G^2$  نشان داده می‌شود، میزان برازش این مدل‌ها با یکدیگر مورد مقایسه قرار گرفت. نتایج حاصل از این تحلیل‌ها که با استفاده از نرم‌افزار BILOG-MG صورت گرفت نشان داد که مدل دوپارامتری در مقایسه با دو مدل دیگر از برازش بهتری با داده‌ها برخوردار است. به منظور جلوگیری از طولانی شدن مقاله، یافته‌های مربوط به این تحلیل در این جا ارائه نشده است.

پس از تعیین مدل پایه، به منظور شناسایی سؤال‌های دارای DIF، هر یک از بلوک‌ها با استفاده از نرم‌افزار IRTL RDIF بصورت جداگانه مورد تحلیل قرار گرفتند. شناسایی سؤال‌های دارای DIF در سه مرحله صورت گرفت. در واقع، برنامه IRTL RDIF در مورد هر بلوک سه بار اجرا گردید. در مرحله اول که هدف آن شناسایی سؤال‌های لنگر اولیه است، برنامه IRTL RDIF هر یک از سؤال‌ها را بعنوان سؤال مورد مطالعه و بقیه سؤال‌ها را به عنوان سؤال لنگر در نظر می‌گیرد و بر این اساس تعیین می‌کند که آیا سؤال مورد مطالعه، دارای DIF است یا نه. با اجرای این مرحله، سؤال‌های لنگر اولیه شناسایی گردید. آنگاه سایر سؤال‌ها که دارای DIF تشخیص داده شدند از فایل داده حذف گردیدند. بعبارت دقیق‌تر، پس از اجرای مرحله اول، برای هر یک از بلوک‌ها فایل جدیدی تشکیل گردید که صرفاً سؤال‌های لنگر اولیه را شامل می‌شد. آنگاه این فایل‌های داده مجدداً با برنامه IRTL RDIF مورد تحلیل قرار گرفتند. هدف از این کار، خالص سازی متغیر همتاسازی به منظور شناسایی سؤال‌های لنگر نهایی بود. پس از شناسایی سؤال‌های لنگر نهایی، برنامه IRTL RDIF یک بار دیگر در مورد فایل داده‌های اولیه هر یک از بلوک‌ها اجرا گردید. در این قسمت، سؤال‌هایی که به عنوان سؤال لنگر نهایی شناسایی شده بودند، به عنوان سؤال لنگر به برنامه معرفی شدند و همچنین سؤال‌های دارای DIF اولیه نیز به برنامه معرفی شدند. پس از اجرای این مرحله که سومین مرحله به حساب می‌آید سؤال‌های دارای DIF شناسایی گردیدند و همچنین نوع DIF بر اساس یکنواخت بودن و غیریکنواخت بودن در این مرحله شناسایی گردیدند. در اجرای برنامه IRTL RDIF چون برخی از بلوک‌ها که ۳ پارامتری به

برنامه معرفی شده بودند دارای پارامتر حدس بالایی بودند و ما نمی‌توانستیم نوع DIF را بیان کنیم، تمام بلوک‌ها را با استفاده از برنامه IRTLRDIF به صورت ۲ پارامتری اجرا کردیم. پس از شناسایی سؤال‌های لنگر و سؤال‌های دارای DIF، سؤال‌های هر یک از بلوک‌ها بصورت جداگانه و از طریق برنامه MULTILOG مورد تحلیل قرار گرفتند. در اجرای این تحلیل که به شیوه چند گروهی صورت گرفت پارامترهای سؤال‌های لنگر بین دو گروه دختر و پسر یکسان قرار داده شدند، در حالی که پارامترهای سؤال‌های واجد DIF برای دو گروه بصورت جداگانه برآورد گردید. همچنین پارامتر  $a$  سؤال‌هایی که دارای DIF یکنواخت بودند نیز بین دو گروه یکسان قرار داده شدند. به عبارتی، در مرحله برآورد نهایی پارامترهای سؤال و توانایی، پارامتر  $a$  و  $b$  سؤال‌های لنگر و همچنین پارامتر سؤال‌های دارای DIF یکنواخت بین دو گروه یکسان قرار داده شده و پارامتر  $a$  و  $b$  سؤال‌های دارای DIF غیریکنواخت و همچنین پارامتر  $b$  سؤال‌های دارای DIF یکنواخت بین دو گروه آزاد گذاشته شد و برای هر گروه بصورت جداگانه برآورد گردید. به دلیل زیاد بودن تعداد سؤال‌ها (۲۱۹ سؤال) و به منظور جلوگیری از طولانی شدن مقاله از ارائه نتیجه مربوط به تک تک سؤال‌ها در اینجا خوداری شده است. برای آگاهی از نتیجه مربوط به هر کدام از سؤال‌ها می‌توانید به غفاری (۱۳۹۳) مراجعه کنید. نتایج حاصل از این تحلیل در جدول ۴ ارائه شده است در این جدول علاوه بر برآوردهای نهایی پارامترهای سؤال‌ها، لنگر یا DIF بودن سؤال‌ها، نوع DIF نیز ارائه شده است. بطور کلی، نتیجه تحلیل‌ها نشان داد که از بین ۲۱۹ سوال مورد مطالعه، ۱۴۴ سوال فاقد DIF (۶۵/۷۵٪) و ۷۵ سؤال (۳۴/۲۵٪) واجد DIF در بین گروه دختران و پسران هستند.

در جدول ۴، تعداد و درصد سؤال‌های دارای DIF یکنواخت و غیر یکنواخت و همچنین تعداد و درصد کل سؤال‌های دارای کارکرد افتراقی به تفکیک حیطه‌های محتوایی نشان داده شده است.

جدول ۴. تعداد سؤال‌های دارای DIF یکنواخت، غیر یکنواخت و تعداد کل در هر یک از حیطه‌ها

حیطه محتوایی	DIF یکنواخت		DIF غیر یکنواخت		تعداد کل
	درصد	فراوانی	درصد	فراوانی	
اعداد	۳۹/۲۸٪	۲۲	۳۱/۵۸٪	۲۸	۳۷/۳۳٪
جبر	۳۳/۹۳٪	۱۹	۳۶/۸۴٪	۲۶	۳۴/۶۷٪

هندسه	۱۰	% ۱۷/۸۶	۲	% ۱۰/۵۲	۱۲	% ۱۶
داده و اطلاعات	۵	% ۸/۹۴	۴	% ۲۱/۰۵	۹	% ۱۲
کل	۵۶	% ۱۰۰	۱۹	% ۱۰۰	۷۵	% ۱۰۰

### بحث و نتیجه گیری

با توجه به مباحث مربوط به اهمیت تفاوت‌های جنسیتی در ریاضیات و بررسی علل این تفاوت‌ها و با توجه به یافته‌های مربوط به کارکرد افتراقی سؤال‌های ریاضیات و علوم تیمز، در پژوهش حاضر تلاش شد تا کارکرد افتراقی سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر ایرانی مورد بررسی قرار گیرد. در ادامه یافته‌های مربوط به هر کدام از سؤال‌های پژوهش به تفکیک مورد بحث و بررسی قرار می‌گیرد.

سؤال اول: آیا سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ تک بعدی است؟

برای پاسخگویی به این سؤال از روش تحلیل عاملی غیر خطی که از طریق برنامه NOHARM (فریزر، ۱۹۸۸) قابل اجرا است استفاده گردید. قبل از انجام تحلیل عاملی، به منظور افزایش دقت برآوردها و تکرارپذیری نتایج، سؤالات مشترک دفترچه‌های مجاور با یکدیگر ادغام شدند و تبدیل به ۱۴ بلوک شدند. پس از شکل‌گیری بلوک‌ها، سؤال‌های هر بلوک بصورت مجزا با روش تحلیل عاملی غیر خطی مورد تحلیل قرار گرفتند. نتایج نشان داد که تمام ۱۴ بلوک صرفاً یک سازه یا عامل مکنون را می‌سنجد و مفروضه تک بعدی بودن صادق است. این یافته گرچه به آزمون ریاضی تیمز اشاره دارد لیکن با نتایج حاصل از پژوهش مینایی (۱۳۹۱) که مورد آزمون علوم تیمز ۲۰۰۷ انجام گردیده، از همخوانی برخوردار است.

سؤال دوم: مناسب‌ترین مدل که با سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ برازش

دارد کدام است؟

به منظور تعیین مناسب‌ترین مدل، مدل‌های لوجستیک ۱، ۲ و ۳ پارامتری به صورت جداگانه به هر یک از بلوک‌ها برازش داده شدند و میزان برازش آن‌ها با یکدیگر مورد مقایسه قرار گرفت. برای مقایسه مدل‌ها از آماره نسبت درستمایی  $G^2$  استفاده گردید. تفاضل بین آماره  $G^2$  مربوط به دو مدل که با  $\Delta G^2$  نمایش داده می‌شود دارای توزیع  $\chi^2$  است و درجات آزادی آن برابر با تفاضل بین تعداد پارامترها در مدل کامل و کاهش یافته است. اگر  $\Delta G^2$  معنادار نباشد، بدین معنا است که برازش مدل کامل با داده‌ها بهتر از برازش مدل

کاهش یافته نیست. نتایج حاصل از این تحلیل‌ها نشان داد که اکثر بلوک‌ها با مدل لوجستیک دو پارامتری از برآزش بهتری برخوردارند.

سؤال سوم: آیا سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر دارای کارکرد افتراقی (DIF) است؟

برای پاسخگویی به این سؤال پس از تعیین مدل پایه، به منظور شناسایی سؤال‌های دارای DIF، هر کدام از بلوک‌ها به صورت جداگانه با استفاده از نرم‌افزار IRTLRDIF (تیسن، ۲۰۰۱) مورد تحلیل قرار گرفتند. یافته‌ها نشان داد که از ۲۱۹ سؤال مورد مطالعه ۲۰۱۱، ۱۴۴ سؤال لنگر (فاقد DIF) و ۷۵ سؤال دارای DIF است. به عبارت دیگر ۶۵/۷۵٪ از سؤال‌ها فاقد کارکرد افتراقی و ۳۴/۲۵٪ از سؤال‌ها دارای DIF هستند. این یافته با پژوهش مینایی (۱۳۹۱) که در مورد آزمون علوم تیمز ۲۰۰۷ انجام گرفت همخوانی دارد. وی به این نتیجه رسید که از ۲۲۸ سؤال، ۱۶۹ سؤال لنگر و ۵۹ سؤال دارای DIF است. تحلیل‌های بیشتر نشان داد از بین ۷۵ سؤال دارای DIF تعداد ۵۶ سؤال دارای DIF یکنواخت و ۱۹ سؤال دارای DIF غیریکنواخت است.

با توجه به یافته‌های پژوهش حاضر می‌توان گفت که اولاً باید این تصور و ایده قالب که در ریاضیات دختران در مقایسه با پسران از توانایی کمتری برخوردارند کنار گذاشته شود. این ادعا در مطالعات دیگر نیز، از جمله توسط دیندیال (۲۰۰۸) مورد تأیید و تأکید قرار گرفته است. معلمان و مربیان تعلیم و تربیت باید دانش‌آموزان دختر را تشویق به مشارکت در موضوعات و فعالیت‌های مرتبط با ریاضی نمایند تا بدینوسیله اعتماد به نفس و علاقه بیشتری به ریاضیات پیدا کنند. پژوهش‌ها نشان می‌دهد اعتماد به نفس و علاقه به ریاضیات منجر به مشارکت مداوم و مستمر دختران و زنان در کلاس‌های ریاضی و حضور بیشتر آنان در فعالیت‌ها و مشاغل مرتبط با ریاضیات می‌شود.

ثانیاً با توجه به این که اکثر سؤال‌های دارای DIF به دو حیطه محتوایی اعداد و جبر مربوط می‌شوند (به جدول ۴ نگاه کنید) توصیه می‌شود معلمان باید تلاش کنند تا تفاوت‌های جنسیتی در این دو حیطه را با فعالیت‌هایی نظیر صرف زمان بیشتر برای مسائل مرتبط با این دو حیطه، دادن تمرین‌های اضافی و برگزاری کلاس‌های فوق برنامه، به حداقل برسانند.

ثالثا با توجه به این که پژوهش‌های متعدد نشان داده است که قالب سؤال می‌تواند تفاوت‌های جنسیتی را کاهش یا افزایش دهد، لذا پرورش‌دهندگان آزمون‌ها باید تلاش کنند تا سؤال‌ها را به گونه‌ای طراحی نمایند که تفاوت‌های جنسیتی را به حداقل برسانند. دو نکته دیگر که پژوهشگران باید به آن‌ها پردازند این است که تحلیل‌های DIF جنسیتی در مورد سؤال‌های سایر آزمون‌های ریاضی مانند امتحانات کنکور سراسری و امتحانات نهایی وزارت آموزش و پرورش در مقاطع مختلف انجام گیرد. علاوه بر این توصیه می‌شود عوامل بیرونی که ممکن است منجر به تفاوت‌های جنسیتی در ریاضیات گردد نیز مورد مطالعه و شناسایی قرار گیرد.

با شناسایی عوامل تاثیرگذار بر تفاوت‌های جنسیتی، مربیان، معلمان و سیاست‌گزاران تعلیم و تربیت کشور خواهند توانست برنامه‌های درسی جدید و خلاقانه طراحی و تدوین نمایند یا برنامه‌های موجود را بهبود بخشند تا نظام و محیط آموزشی کشورمان بالنده و غنی گردد.

### منابع

- حجازی، الهه و نقش، زهرا. (۱۳۸۶). رابطه خودکارآمدی ریاضی، سودمندی ادراک شده و راهبردهای خود تنظیمی با پیشرفت دانش آموزان: یک مقایسه جنسیتی. *مجله علمی و پژوهشی مطالعات زنان* (۲)، ۸۴-۱۰۲.
- غفاری، زهرا. (۱۳۹۳). کارکرد افتراقی سؤال‌های پایه هشتم آزمون ریاضی تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر با استفاده از رویکرد نظریه سؤال - پاسخ (IRT). پایان‌نامه کارشناسی ارشد. تهران: دانشگاه آزاد اسلامی، واحد تهران مرکز.
- صادقی، مریم. (۱۳۹۲). کارکرد افتراقی (DIF) سؤال‌های آزمون ریاضی پایه چهارم و هشتم تیمز ۲۰۱۱ با استفاده از مدل‌های IRT و کلاسیک. پایان‌نامه کارشناسی ارشد. تهران: دانشگاه آزاد اسلامی واحد تهران مرکز.
- مینایی، اصغر. (۱۳۹۱). سنجش مقایسه‌پذیری سازه و تحلیل کارکرد افتراقی سؤال‌ها (DIF) و بلوک‌ها (DTF) آزمون علوم پایه هشتم ۲۰۰۷ در بین دانش‌آموزان ایران و آمریکا. تهران: پژوهشگاه آموزش و پرورش.
- مینایی، اصغر و فلسفی نژاد، محمد رضا. (۱۳۸۹). روش‌های سنجش تک‌بعدی بودن سؤال‌ها در مدل‌های دوارزشی IRT. *فصلنامه اندازه‌گیری تربیتی*، ۳، صص ۷۱-۹۹.

- Beller, M. & Gafni, N. (2000). Can item format(multiple-choice vs. open-ended) account for gender differences in mathematics achievement?, *Sex Roles*, 42 (1/2), 1-21.
- Benbow, C. P. (1989). Sex differences in mathematical reasoning ability in intellectually talented preadolescent: Their nature, effects, and possible causes. *Behavioral and Brain Science*, 11, 169-183.
- Bolger, N. & Kellaghan, D. (1990). Method of measurement and gender differences in scholastics achievement. *Journal of Educational Measurement*, 27, 165-174.
- Dindyal, J. (2008). An overview of the gender factor in mathematics in TIMSS – 2003 for the Asia – Pacific region. *Mathematics Education*, 40, 993 – 1005.
- Doolittle, A. E. (1989). Gender differences in performance on mathematics achievement items. *Applied Measurement in Education*, 2(2), 161-177.
- Doolittle, A. E. & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement*, 24, 157-166.
- Ercikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research*, 29, 543-553.
- Ercikan, K. & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, 5, 23-35.
- Faber, R. (2008). *Gender bias in the Trends in Mathematics and Science Study 2003 (TIMSS) for Canadian Students* (Master's thesis). Retrieved from <https://dr.library.brocku.ca/handle/10464/1469>
- Fierros, E. G. (1999). *Examining gender differences in mathematics achievement on the Third International Mathematics and Science Study (TIMSS)*. In Annual meeting of the American Educational Research Association (ERIC Document No. 431602).
- Fraser, C. (1988). *NOHARM: An IBM PC program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory*. Armidale, Australia: The University of New England.
- Harris, A. M. & Carlton, S. T. (1993). Patterns of gender differences on mathematics itemson the scholastic aptitude test. *Applied Measurement in Education*, 6(2), 137-151.
- Langefeld, T. E. (1997). Test fairness: Internal and external investigations of gender bias in mathematics testing. *Educational Measurement: Issues and Practice*, 16(1), 20-26.
- Ma, X. (1995). Gender differences in mathematics achievement between Canadian and Asian education systems. *The Journal of Educational Research* 89(2), 118–127.
- Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). Gender differences in achievement:
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report*. Boston: International Study Center.

- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Stote, G., Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, 16, 93-102.
- Thissen, D. (2001). *IRTLRDIF(version 2.02b): Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning [Computer program]*. Chapel Hill, NC: L. L. Thurstone Psychometric Laboratory.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99, 118-128.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* ( pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum.
- Wilder, G. Z., & Powell, K. (1989). *Sex differences in test difference: A survey of the literature*. College Board Report, No. 89-93.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Association.
- Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical assessment Research & Evaluation*, 12, 1-26. Available Online: <http://pareonline.net/pdf/v12n2.pdf>
- Yan, S. (2005). *Gender – Related Differential Item Functioning in Mathematics Assessment on the Third International and Science Study – Repeat(TIMSS-R)*. (Doctoral dissertation). Available from ProQuest Dissertation and Theses database.(UMI No. 3177610).