



Title Differential Item Functioning of Fourth Grade Mathematics Test of the TIMSS 2019 Based on Gender

Ensie Khalilnejadi¹, Maryam Mohsenpour², Azam Moghadam

1. Master of Educational Psychology, Alzahra University, Tehran, Iran. Email: e_kh_1370@yahoo.com

2. Assistant Professor at Department of Educational Psychology, Faculty of Education and Psychology, Alzahra University, Tehran, Iran; (Corresponding Author), Email: m.mohsenpour@alzahra.ac.ir

3. Ph.D of Measurement and Assessment, Tehran, Iran. Email: azam.moghadam@gmail.com

Article Info

Article Type:

Research Article

Received: 2023.09.26

Received in revised form: 2024.02.01

Accepted: 2024.02.26

Published online: 2024.03.20

ABSTRACT

Objective: The present study was conducted with the aim of investigating the differential item functioning of the fourth-grade mathematics items in the 2019 TIMSS according to gender.

Methods: Given that the analysis is based on the data collected from Iran in aforementioned study, this research is descriptive and secondary analysis.

Results: To investigate the differential item functioning (DIF), among the 14 booklets, two booklets (one and seven booklet) were randomly selected and the research sample consisted of 861 student. To analyzing DIF, two traditional and modern methods were used, respectively, "logistic regression method" and "Chi-2 Lord method based on Rush model". The findings showed that the results of DIF were the same with both traditional and modern methods, and out of 32 items in booklet one, four items had differential functioning which two items were in favor of girls and two items were in favor of boys. Also, out of 30 items in booklet seven, three items had a differential functioning, two items were in favor of boys and one item was in favor of girls.

Conclusion: In examining the structure of questions with DIF, it was found that, boys were more likely than girls to correctly answer numbers items and girls were more likely than boys to correctly answer geometry items. Furthermore, boys were more likely than girls to correctly answer in applying and reasoning items, and they were more likely than girls to correctly answer life related problems. Thus, it is suggested that items with differential functioning be investigated more deeply through task-based interview.

Keywords: TIMSS study – Differential Item Functioning – Mathematics Achievement Test - Gender - Fourth grade students

Cite this article: Khalilnejadi, Ensie; Mohsenpour, Maryam; Moghadam, Azam (2024). Differential Item Functioning of Fourth Grade Mathematics Test of the TIMSS 2019 Based on Gender. *Educational Measurement and Evaluation Studies*, 14 (45): 7-26 pages. DOI:10.22034/EMES.2024.2012558.2511



© The Author(s).

Publisher: National Organization of Educational Testing (NOET)



کنش افتراقی سوال‌های آزمون ریاضیات پایه چهارم مطالعه تیمز ۲۰۱۹ بر حسب جنسیت

انسبیه خلیل نژادی^۱، مریم محسن پور^۲، اعظم مقدم^۳

۱. دانش‌آموخته کارشناسی ارشد روانشناسی تربیتی، دانشگاه الزهرا (س)، تهران، ایران. رایانامه: e_kh_1370@yahoo.com
 ۲. استادیار گروه روانشناسی تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه الزهرا، تهران، ایران؛ (نویسنده مسئول)، رایانامه: m.mohsenpour@alzahra.ac.ir
 ۳. دانش‌آموخته دکتری سنجش و اندازه‌گیری، تهران، ایران. رایانامه: azam.moghadam@gmail.com

چکیده

اطلاعات مقاله

هدف: مطالعه حاضر با هدف بررسی کنش افتراقی سوالات آزمون ریاضیات مطالعه تیمز ۲۰۱۹ پایه چهارم، بر حسب جنسیت انجام شد.

نوع مقاله:
مقاله پژوهشی

روش پژوهش: باتوجه به اینکه تحلیل بر مبنای داده‌های گردآوری شده از کشور ایران در مطالعه مذکور صورت گرفت؛ این پژوهش توصیفی و از نوع تحلیل ثانویه به شمار می‌آید.

دریافت: ۱۴۰۲/۰۷/۰۴

اصلاح: ۱۴۰۲/۱۱/۱۲

پذیرش: ۱۴۰۲/۱۲/۰۷

انتشار: ۱۴۰۳/۰۱/۰۱

یافته‌ها: برای بررسی کنش افتراقی، ابتدا از میان ۱۴ دفترچه سوال، دو دفترچه یک و هفت به صورت تصادفی انتخاب و نمونه پژوهش، شامل ۸۶۱ نفر بودند. برای تحلیل از دو روش سنتی و نوین به ترتیب "روش رگرسیون لجستیک" و "روش خی‌دو لرد مبتنی بر مدل راش" استفاده شد. یافته‌ها نشان داد نتایج کنش افتراقی با هر دو روش یکسان بود و از ۳۳ سوال دفترچه یک، چهار سوال دارای کنش افتراقی بود که دو سوال به نفع دختران و دو سوال به نفع پسران بود. همچنین از ۳۰ سوال دفترچه هفت، سه سوال کنش افتراقی داشت که دو سوال به نفع پسران و یک سوال به نفع دختران بود.

نتیجه‌گیری: در بررسی ساختار سوال‌های دارای کنش افتراقی، مشخص شد در حیطه‌های محتوایی، پسران احتمالاً در اعداد و دختران در هندسه تمایل به عملکرد بهتری داشتند. همچنین باتوجه به سطوح شناختی، احتمالاً پسران تمایل به عملکرد بالاتری در به‌کار بستن و استدلال داشته و در زمینه سوال‌های مرتبط با زندگی تمایل به پاسخگویی صحیح برای پسران نسبت به دختران احتمالاً بیشتر باشد. بنابراین، پیشنهاد می‌شود، سوالات دارای کنش افتراقی از طریق مصاحبه مبتنی بر تکلیف بررسی عمیق‌تری شوند.

واژه‌های کلیدی: مطالعه تیمز، کنش افتراقی سوال، آزمون پیشرفت تحصیلی ریاضیات، جنسیت و دانش‌آموزان پایه چهارم

استناد: خلیل‌نژادی، انسبیه؛ محسن‌پور، مریم؛ مقدم، اعظم (۱۴۰۳). کنش افتراقی سوالات آزمون ریاضیات پایه چهارم مطالعه تیمز ۲۰۱۹ بر حسب جنسیت. *مطالعات*

اندازه‌گیری و ارزشیابی آموزشی، ۱۴ (شماره ۴۵)، ۲۶-۷ صفحه.

DOI:10.22034/EMES.2024.2012558.2511

حق مؤلف © نویسندگان.

ناشر: سازمان سنجش آموزش کشور



مقدمه

نظام‌های آموزشی، علاقه‌مند به ارائه آموزش با کیفیت برای ارتقا عملکرد تحصیلی دانش‌آموزان مدارس خود هستند. یکی از راه‌های موثر برای پیگیری آموزش با کیفیت، سنجش پیشرفت تحصیلی است. از نظر پاین^۱ (۲۰۰۳) سنجش تحصیلی در آموزش، بیشترین سهم را در درک فرایندهای تدریس - یادگیری دارد. به طور ویژه، داده‌های سنجش تحصیلی در تشخیص مشکلات یادگیری و ارزیابی مداخلات مبتنی بر مشکلات، کاربرد دارد. از طریق سنجش و ارزشیابی مناسب، معلمان می‌توانند دانش‌آموزان خود را طبقه‌بندی کنند، بازخورد داده و آموزش آن‌ها را ساختار دهند.

در بسیاری از نظام‌های آموزشی، مراکزی ملی جهت سنجش و ارزشیابی عملکرد تحصیلی و سنجش بازده‌های آموزشی تأسیس شده‌اند که مسئولیت ارزیابی عملکرد تحصیلی، به عهده آن‌ها است. با این حال، در نظام آموزشی ایران، به دلیل نبود مطالعات سنجش و ارزشیابی ملی در حوزه عملکرد تحصیلی، شرکت در مطالعات بین‌المللی می‌تواند فرصت مناسبی برای بررسی عملکرد نظام آموزشی در سطح ملی فراهم آورد (کیامنش، محسن‌پور، صفرخانی و اقدسی، ۱۳۹۱). یکی از آزمون‌های بین‌المللی، مطالعه بین‌المللی روند پیشرفت ریاضیات و علوم^۲ (تیمز) است که از شناخته شده‌ترین و بزرگترین مطالعات آموزش تطبیقی برای سنجش پیشرفت علوم، ریاضیات و خواندن دانش‌آموزان در راستای چارچوب برنامه درسی مدارس در کشورهای مختلف است که هر چهار سال یک بار در پایه‌های چهارم و هشتم انجام می‌شود. هدف اصلی مطالعات تیمز اندازه‌گیری عملکرد تحصیلی دانش‌آموزان در ریاضیات و علوم است به گونه‌ای که با ردیابی روندها، بتواند میزان پیشرفت یا پسرفت دانش‌آموزان کشورها را رصد کند (مولیس و مارتین^۳، ۲۰۱۷). کشور ایران نیز از سال ۱۳۷۱ با عضویت در انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی، برای اولین بار در سال ۱۹۹۵ (معادل با سال ۱۳۷۴ شمسی) در این مطالعه شرکت داشته است. با وجود شش دوره شرکت در مطالعات تیمز (۱۹۹۵، ۲۰۰۳، ۲۰۰۷، ۲۰۱۱، ۲۰۱۵، ۲۰۱۹ و ۲۰۲۳)، عملکرد دانش‌آموزان ایرانی پایه چهارم تاکنون همواره در بخش کشورهای پائین‌تر از نقطه وسط مقیاس (۵۰۰) قرار دارد و میانگین پیشرفت تحصیلی ریاضیات دانش‌آموزان ایرانی پایه چهارم ابتدایی در آخرین نتایج به دست آمده که مربوط به مطالعه ۲۰۱۹ است، ۴۴۳ بوده است (پسران با میانگین نمره ۴۴۷ و دختران با میانگین نمره ۴۳۹ که البته از لحاظ اماری معنادار نبوده است). در حالیکه کشور ایران رتبه ۵۰ را در این دوره از میان ۵۸ کشور بدست آورده است؛ کشور سنگاپور با میانگین نمره ۶۲۵ در رتبه نخست و کشور فیلیپین با میانگین نمره ۲۹۷ در رتبه پایانی قرار گرفته‌اند (مولیس و همکاران^۴، ۲۰۲۰).

زمانی که از داده‌های به دست آمده از آزمون‌های به کار برده شده در یک سنجش ملی یا بین‌المللی برای برنامه‌ریزی نظام آموزشی استفاده می‌شود، بررسی عدالت^۵ مورد توجه پژوهشگران قرار می‌گیرد (گوکچه و همکاران^۶، ۲۰۲۱). آزمون‌ها باید برای همه دانش‌آموزان عادلانه باشند اما بعضی از آنها نسبت به جنسیت، قومیت، فرهنگ، مذهب، طبقات اقتصادی - اجتماعی، تفاوت‌های زبانی و برنامه‌های آموزشی سوگیری^۷ دارند (علی موسی^۸، ۲۰۱۹؛ آریم و ارسیکان^۹، ۲۰۱۴؛ همبلتون و راجرز^{۱۰}، ۱۹۹۴). یکی از شاخص‌های مهم در زمینه سوگیری آزمون، کنش افتراقی سؤال^{۱۱} است. کنش افتراقی یا کارکرد افتراقی سؤال، به عنوان مفهومی تعریف می‌شود که در آن آزمون‌شوندگان یک گروه به دلیل

1. payne

2. Trends in International Mathematics and Science Study (TIMSS)

3. Mullis & martin

4. Mullis, Martin, Foy, Kelly & Fishbein

5. Fairness

6. Gokce, Berberoglu & Sireci

7. Bias

8. Ali moussa

9. Arim & Ercikan

10. Hambleton & Roger

11. Differential Item Functioning (DIF)

برخی از ویژگی‌های سؤال یا وضعیت اجرای آزمون که مرتبط با هدف آن نیست، کمتر احتمال دارد نسبت به آزمون شوندگان گروه دیگر به یک سؤال به درستی پاسخ دهند (زومبو^۱، ۱۹۹۹).

مبانی نظری و پیشینه پژوهش

باتوجه به اهمیت نتایج آزمون‌های تیمز در تصمیم‌گیری‌های آموزشی، بررسی کنش افتراقی جایگاه ویژه‌ای در میان پژوهش‌های این حوزه یافته است. در بررسی مطالعات در حوزه کنش افتراقی مرتبط با مطالعات تیمز پژوهش‌های متنوعی یافت شد که نشان از اهمیت این موضوع دارد. بیشتر این مطالعات متمرکز شده است روی مقایسه روش‌های آماری برآورد کنش افتراقی با در نظر گرفتن جنسیت، زبان و فرهنگ به عنوان متغیر دوم، یافتن سوالات دارای کنش افتراقی بر حسب جنسیت، شیوه امتحان و دیدگاه اجتماعی، مقایسه کشورها با زبان‌ها و فرهنگ‌های متفاوت و همچنین بررسی منابع ایجادکننده کنش افتراقی. به‌عنوان نمونه، اینابی و دودین^۲ (۲۰۱۸) در بررسی کنش افتراقی سؤال بر حسب متغیر جنسیت در آزمون ریاضیات تیمز ۲۰۱۵ و در پایه هشتم دریافتند که از ۱۹۷ سؤال، ۶۳ سؤال دارای کنش افتراقی بودند که ۳۰ سؤال به نفع پسران و ۳۳ سؤال به نفع دختران بود. در بررسی بیشتر داده‌ها، آن‌ها دریافتند که پسران احتمالاً تمایل بیشتری نسبت به دختران در پاسخگویی به مسائل ریاضیات دشوارتر، ناآشنا تر ولی مرتبط با زندگی^۳ دارند، در حالی که، دختران بیشتر از پسران احتمالاً به سوال‌های آسانتر، کمتر دشوار و انتزاعی‌تر^۴ (نقطه مقابل مرتبط با زندگی) پاسخ صحیح می‌دهند.

در بررسی روانسنجی کنش افتراقی، در گام اولیه صرفاً شناسایی الگوهای وجود تفاوت جنسیتی با کنترل سطح پیشرفت تحصیلی مورد توجه است، با اینحال باید گام‌های بعدی در راستای علل احتمالی کنش افتراقی برداشته شود. به عنوان مثال، بالت^۵ (۲۰۰۰) معتقد است از علل بالقوه کنش افتراقی، می‌توان به قالب سؤال^۶ (باز پاسخ / چند گزینه ای)، نوع مسئله^۷ (انتزاعی / حقیقی) اشاره کرد. به علاوه، دولیتل^۸ (۱۹۸۹) معتقد است حیطه محتوایی سوالات نیز می‌تواند منبعی از کنش افتراقی جنسیتی به شمار رود چرا که پسران در هند سه بهتر از دختران عمل کرده‌اند. همچنین طبق نظر لان^۹ (۲۰۱۴)، سطح شناختی نیز از دیگر منابع شناخته شده است. پسران در مواردی با سطوح شناختی بالاتر مانند مشکلات چند مرحله ای و مشکلات کاربردی زندگی واقعی، بهتر از دختران عمل کرده‌اند، در حالی که دختران در سطوح شناختی پایین‌تر از پسران بهتر عمل کردند. همچنین در یک مطالعه، علی میرزایی و همکاران (۱۳۹۸) باهدف بررسی منابع کنش افتراقی سؤال‌ها، ۱۹ پژوهش در زمینه شناسایی منابع کنش افتراقی را ارزیابی کردند. آن‌ها در یک دسته‌بندی، منابعی چون ویژگی‌های سؤال (شامل تفاوت در ساختار منطقی محتوا، قالب سؤال‌ها، نوع مسئله، طولانی بودن نسبی سؤال و دشواری کلمات یا عبارات)، عوامل مرتبط با دانش آموز (نظیر سن، جنس، کشور، قومیت و زبان)، مسائل مربوط به انطباق و ترجمه نسخه‌های چندزبانه، ویژگی‌های فرهنگی، برنامه درسی متناسب با آزمون، عوامل مرتبط با آموزش معلم و مهارت‌های او و کلاس‌های اضافه پس از مدرسه را از جمله منابع کنش افتراقی معرفی کردند.

از طرفی بررسی مطالعات داخلی در حوزه کنش افتراقی آزمون‌های ریاضیات در مطالعه تیمز حاکی از مطالعات متنوعی است که در جدول ۱ ارائه شده است.

1. Zumbo
2. Innabi & Dodeen
3. Life-Related Problem
4. Abstract Problem
5. Bolt
6. Item format
7. problem type
8. Doolittle
9. Lan

جدول ۱. پژوهش‌های ایرانی حوزه کنش افتراقی در آزمون‌های مطالعه تیمز

پژوهشگر	عنوان	آزمون	جامعه	متغیر	روش تحلیل	یافته‌ها
سبحانی سیناب (۱۴۰۰)	کاربرد رگرسیون لجستیک و چندنمایی در شناسایی کارکرد افتراقی سؤال‌ها (DIF) و گزینه‌های (DOF) آزمون ریاضیات پایه هشتم تیمز ۲۰۱۹ بین دختران و پسران ایران	ریاضیات تیمز ۲۰۱۹ (فقط سؤال‌های چندگزینه ای)	هشتم	جنسیت	رگرسیون لجستیک چندنمایی	از ۶۲ سؤال: ۱۱ سؤال دارای کنش افتراقی است از اینها هفت سؤال به نفع پسران و چهار سؤال به نفع دختران است. نتایج نشان‌دهنده اثرگذاری کنش افتراقی سؤال‌ها و گزینه‌ها روی عملکرد بوده است.
عسکری تورزن (۱۴۰۰)	مقایسه و کاربرد روش آمیخته راش، درخت سؤال محور و لاسو در شناسایی کارکرد افتراقی سؤال‌های ریاضیات آزمون پایه هشتم تیمز ۲۰۱۵	ریاضیات تیمز ۲۰۱۵ (دفترچه سه و ده)	هشتم	جنسیت و روش	روش آمیخته راش، درخت سؤال محور و لاسو	این آزمون با روش راش آمیخته دارای کنش افتراقی است. باتوجه به روش درخت‌های سؤال محور ۵۹ سؤال از دفترچه سوم و ۶۱ سؤال از دفترچه دهم و با روش لاسو ۲۹ سؤال از دفترچه سوم و ۲۳ سؤال از دفترچه دهم دارای کنش افتراقی به نفع پسران است. همچنین در این قیاس مدل راش درختی توانمندتر و کم خطا تر بوده و گزینه بهتری برای بررسی کنش افتراقی سؤال‌ها است.
مینائی، یونسی، کاظمی دانا (۱۳۹۹)	کاربرد مدل‌های تشخیصی شناختی (CDM) در مطالعه کارکرد افتراقی سؤال‌های ریاضیات پایه هشتم تیمز ۲۰۰۷ در بین دختران و پسران	ریاضیات تیمز ۲۰۰۷ (۴۴ سؤال از دفترچه یک، دو، سه و چهارده)	هشتم	جنسیت	مدل‌های تشخیصی شناختی و مدل غیر جبرانی دینا	از ۴۴ سؤال: ۱۸ سؤال دارای کنش افتراقی است از اینها ۱۲ سؤال دارای کنش افتراقی یکنواخت و ۶ سؤال غیریکنواخت است. همچنین ۳ سؤال به نفع دختران و ۱۵ سؤال به نفع پسران است.

جدول ۱. پژوهش‌های ایرانی حوزه کنش افتراقی در آزمون‌های مطالعه تیمز

پژوهشگر	عنوان	آزمون	جامعه	متغیر	روش تحلیل	یافته‌ها
مینایی، غفاری (۱۳۹۴)	کارکرد افتراقی سؤال‌های پایه هشتم آزمون ریاضیات تیمز ۲۰۱۱ در بین دانش‌آموزان دختر و پسر با استفاده از رویکرد نظریه سؤال-پاسخ	ریاضیات تیمز ۲۰۱۱	هشتم	جنسیت	نظریه پرسش - پاسخ: نسبت درست‌نمایی	از ۲۱۹ سؤال: ۷۵ سؤال دارای کنش افتراقی است از اینها ۵۶ سؤال دارای کنش افتراقی یکنواخت به نفع پسران و ۱۹ سؤال غیریکنواخت است.
مینایی (۱۳۹۲)	سنجش مقایسه‌پذیری سازه و تحلیل کارکرد افتراقی سؤال‌ها (DIF) و بلوک‌های (DTF) آزمون علوم پایه هشتم تیمز ۲۰۰۷ در بین دانش‌آموزان ایران و آمریکا	علوم تیمز ۲۰۰۷	هشتم	کشور (ایران و آمریکا)	ترکیبی از روش‌های مبتنی بر نظریه کلاسیک و سؤال - پاسخ	از ۲۲۸ سؤال: ۱۶۹ سؤال دارای کنش افتراقی است از اینها ۱۱۸ سؤال دارای کنش افتراقی یکنواخت و ۵۱ سؤال غیریکنواخت است. همچنین ۶۴ سؤال به نفع دانش‌آموزان ایران و ۱۰۵ سؤال به نفع دانش‌آموزان آمریکا بوده است.
صادقی (۱۳۹۲)	کارکرد افتراقی (DIF) سؤال‌های آزمون ریاضیات پایه چهارم و هشتم با TIMSS 2011 استفاده از مدل‌های IRT و کلاسیک	ریاضیات تیمز ۲۰۱۱ (دفترچه ده و پنج)	چهارم و هشتم	جنسیت و روش	کلاسیک: رگرسیون لجستیک و سؤال - پاسخ: مساحت بین منحنی	روش کلاسیک: در پایه چهارم، چهار سؤال و در پایه هشتم، شش سؤال دارای سوگیری است. روش سؤال - پاسخ: در پایه چهارم، دو سؤال و در پایه هشتم، یک سؤال دارای سوگیری است. سه سؤال در هر دو روش مشترک بودند

جدول ۱. پژوهش‌های ایرانی حوزه کنش افتراقی در آزمون‌های مطالعه تیمز

پژوهشگر	عنوان	آزمون	جامعه	متغیر	روش تحلیل	یافته‌ها
هاشمی (۱۳۸۹)	بررسی سوگیری سوال ها آزمون تیمز ۲۰۰۷ در مقطع سوم راهنمایی در درس علوم باتوجه به متغیرهای جنسیت و قومیت و همچنین مقایسه نظریه‌های کلاسیک و سوال - پاسخ	علوم تیمز ۲۰۰۷ (دفترچه یک)	هشتم	جنسیت و روش	کلاسیک: روش خی دو کامل: دو سؤال روش متتل هنزل وخی دو کامل سؤال پاسخ: مساحت بین دو منحنی: سه سؤال سودار تشخیص داده شد. مقایسه پارامترهای سؤال	روش خی دو کامل: دو سؤال روش متتل هنزل: چهار سؤال روش مقایسه پارامترهای سؤال: یک سؤال مساحت بین دو منحنی: سه سؤال سودار تشخیص داده شد. دقت روش‌های کلاسیک در بررسی پارامترهای سؤال بیشتر است و روش مقایسه پارامترهای سؤال دقت کمتری دارد.

طبق جدول ۱، از میان هفت پژوهش انجام شده در حوزه کنش افتراقی سوال‌های مطالعه تیمز، پنج پژوهش به بررسی سوال‌های آزمون ریاضیات و دو پژوهش سوال‌های آزمون علوم را مورد بررسی قرار داده‌اند. همچنین شناسایی سوال‌های دارای کنش افتراقی در هفت پژوهش برای پایه هشتم و تنها یک پژوهش برای پایه چهارم اجرا شده است. به علاوه، تمرکز اصلی کنش افتراقی بر روی متغیر جنسیت بوده است و تنها یک پژوهش بر روی متغیر کشور متمرکز شده است. مقایسه روش‌های تحلیل کنش افتراقی در سه پژوهش و مقایسه عملکرد کشورها در یک پژوهش به کار برده شده است. به علاوه، نتایج جدول ۱ نشان می‌دهد که در همه پژوهش‌ها، کنش افتراقی وجود داشته است.

بررسی پیشینه داخلی در مورد کارکرد افتراقی جنسیتی نشان از تمرکز پژوهش‌ها در مقطع هشتم داشته و مطالعات کمتری در مقطع ابتدایی انجام شده است. باتوجه به اهمیت آموزش در پایه‌های ابتدایی، پژوهش حاضر قصد دارد جهت بررسی عدالت آزمون بین‌المللی تیمز به دلیل اهمیت و نفوذ زیاد در تصمیم‌گیری‌های آموزشی، به مطالعه کارکرد افتراقی آخرین داده‌های منتشر شده از مطالعه تیمز یعنی تیمز ۲۰۱۹ در پایه چهارم ابتدایی و در بین دانش‌آموزان ایرانی بپردازد. بنابراین این پژوهش دنبال پاسخگویی به این سوال است آیا سوال‌های آزمون ریاضیات تیمز ۲۰۱۹ پایه چهارم در بین دختران و پسران دارای کنش افتراقی است؟ و در صورت وجود کنش افتراقی، شدت اثر در سوال‌های شناسایی شده دارای کنش افتراقی به چه میزان است؟

روش پژوهش

در این پژوهش کنش افتراقی سوال‌ها با دو روش مبتنی بر نظریه کلاسیک و نظریه سؤال - پاسخ به ترتیب از روش رگرسیون لجستیک و روش خی دو لرد مبتنی بر مدل راش استفاده شده است. از آنجا که تحلیل‌های این مطالعه براساس داده‌های به دست آمد از نتایج آزمون ریاضیات کشور ایران در مطالعه تیمز ۲۰۱۹ صورت گرفته است؛ لذا جزء پژوهش‌های توصیفی از نوع تحلیل ثانویه به شمار می‌آید. در مطالعه تیمز ۲۰۱۹، جامعه آماری شامل ۱۳۳۴۲۵۰ دانش‌آموز پایه چهارم از ۳۸۶۴۵ مدرسه بود که در سال تحصیلی ۹۸-۱۳۹۷ در کشور ایران در حال تحصیل بوده‌اند. نمونه‌گیری مطالعات تیمز به روش خوشه‌ای دومرحله‌ای (مدرسه و کلاس) است که در سال ۲۰۱۹، ۶۰۱۰ دانش‌آموز پایه چهارم از ۲۲۴ مدرسه انتخاب شده‌اند (مولیس و همکاران، ۲۰۲۰).

باتوجه به گستردگی محتوای مورد پوشش در مطالعات تیمز، سوال‌ها آزمون ریاضیات پایه چهارم در مجموعه‌ای از ۱۴ دفترچه^۱ دسته بندی شده‌اند. سوال‌های ریاضیات در سه حیطه محتوایی شامل اعداد، اندازه‌گیری و هندسه و داده و در سه حیطه شناختی مختلف شامل دانستن، به کار بستن و استدلال و از نوع چندگزینه‌ای و پاسخ ساز هستند. در این مطالعه به طور تصادفی دو دفترچه یک (با ۳۲ سؤال) و هفت (با ۳۰ سؤال) برای تحلیل انتخاب شدند و نمونه مورد بررسی در پژوهش حاضر شامل ۸۶۱ نفر (۴۳۲ دختر و ۴۲۹ پسر) بود. به منظور بررسی همه سوال‌های یک دفترچه، سوال‌های پاسخ ساز مشابه سوال‌های چندگزینه‌ای به حالت دوارزشی کدگذاری مجدد شدند. در این روش به پاسخ کامل نمره یک و باقی حالات نمره صفر داده شد. جهت تحلیل داده‌ها از نرم‌افزار آر^۲ و بسته دیف آر^۳ (مگیس^۴ و همکاران، ۲۰۲۲) استفاده شد. در خصوص اندازه اثر، جودوین و گیرل^۵ (۲۰۰۱) نیز معیار طبقه‌بندی اندازه اثر سؤال را در روش رگرسیون این‌گونه شرح دادند که اگر بین ۰ و ۰/۰۳۵ باشد اندازه اثر ناچیز است؛ اگر بین ۰/۰۳۵ و ۰/۰۷ باشد اندازه اثر متوسط است و اگر بین ۰/۰۷ و ۱ باشد اندازه اثر زیاد است. هالند و تایر^۶ (۱۹۸۵) مقیاس دلتای ای تی اس^۷ را برای اندازه اثر در روش لرد بدین صورت طبقه‌بندی کردند که اگر قدر مطلق دلتا کمتر یا مساوی یک باشد ناچیز به شمار می‌رود، اگر بین ۱ و ۱/۵ باشد متوسط است و اگر بزرگتر و مساوی ۱/۵ باشد اندازه اثر زیاد است.

یافته‌ها

جدول ۲ اطلاعات مربوط به سوال‌ها آزمون ریاضیات پایه چهارم و عملکرد آزمودنی‌ها در دفترچه‌های یک و هفت را نشان می‌دهد.

جدول ۲. اطلاعات توصیفی دفترچه‌های یک و هفت

دفترچه	تعداد سؤال	حیطه محتوایی	حیطه شناختی	نوع سؤال	جنسیت	فراوانی	میانگین نمره
دفترچه یک	۳۲	اعداد (۱۳)	دانستن (۹)	پاسخ ساز (۲۳)	دختر	۲۱۵	۱۰/۸۴
		اندازه‌گیری و هندسه (۱۱)	به کار بستن (۱۶)	چندگزینه‌ای (۹)	پسر	۲۱۲	۱۱/۲۷
		داده (۸)	استدلال (۷)				
دفترچه هفت	۳۰	اعداد (۱۱)	دانستن (۶)	پاسخ ساز (۱۶)	دختر	۲۱۷	۱۰
		اندازه‌گیری و هندسه (۷)	به کار بستن (۱۵)	چندگزینه‌ای (۱۴)	پسر	۲۱۷	۱۰/۱۸
		داده (۱۲)	استدلال (۹)				

طبق جدول ۲، متوسط عملکرد دختران و پسران در هر یک از دو دفترچه تفاوت قابل چشمگیری ندارد. در ادامه و در جداول ۳، ۴، ۵ و ۶ کنش افتراقی سوالات دفترچه یک و هفت را به ترتیب با روش رگرسیون لجستیک و روش خی دو مبتنی بر مدل راش^۸ ارائه شده است.

1. Booklet

2. R

3. DifR

4. Magis

5. Jodoin & Gierl

6. Holland, P. W., & Thayer

7. ETS

۸. گفتنی است، قبل از استفاده از مدل راش، ابتدا مفروضه تک‌بعدی بودن از روش تحلیل موازی و برازش سوال‌ها مورد بررسی قرار گرفت.

همچنین، نمودارهای ۱، ۲، ۳ و ۴ کنش افتراقی دفترچه یک و هفت را به ترتیب با روش لجستیک و روش خی‌دو لرد مبتنی بر مدل راش نشان می‌دهد.

جدول ۳. کنش افتراقی یکنواخت سوال‌ها در دفترچه یک مبتنی بر رگرسیون لجستیک

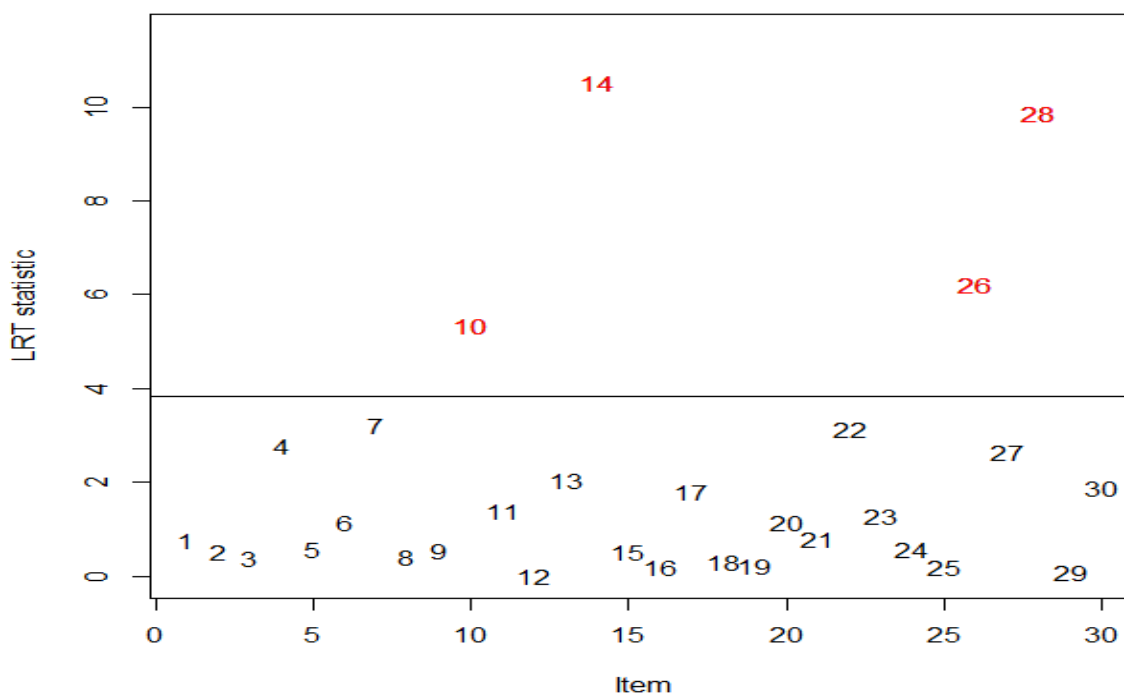
ردیف	سؤال	حیطه محتوایی	حیطه شناختی	LRT آماره	P	R ²	مقدار اندازه اثر
۱	MP51043	عدد	دانستن	۰/۷۸	۰/۳۷	۰/۰۰۲	-/۰۰۰۲۲
۲	MP51040	عدد	دانستن	۰/۵۲	۰/۴۷	۰/۰۰۲	-/۰۰۰۲۴
۳	MP51008	عدد	استدلال	۰/۳۶	۰/۵۵	۰/۰۰۰۱	-/۰۰۰۱۳
۴	MP51031A	عدد	به کار بستن	۲/۶۸	۰/۱۰	۰/۰۰۶	-/۰۰۰۶۰
۵	MP51031B	عدد	به کار بستن	۰/۵۳	۰/۴۶	۰/۰۰۱	-/۰۰۰۱۲
۶	MP51508	عدد	به کار بستن	۱/۱۶	۰/۲۸	۰/۰۰۳	-/۰۰۰۳۰
۷	MP51216A	اندازه‌گیری و هندسه	به کار بستن	۳/۲۷	۰/۰۷	۰/۰۰۹	-/۰۰۰۸۹
۸	MP51216B	اندازه‌گیری و هندسه	دانستن	۰/۴۲	۰/۵۲	۰/۰۰۱	-/۰۰۰۱۴
۹	MP51221	اندازه‌گیری و هندسه	دانستن	۰/۵۳	۰/۴۶	۰/۰۰۱	-/۰۰۰۱۵
*۱۰	MP51115	اندازه‌گیری و هندسه	استدلال	۵/۳۵	۰/۰۲	۰/۰۲	-/۰۰۰۹۵
۱۱	MP51507A	داده	دانستن	۱/۳۷	۰/۲۴	۰/۰۰۴	-/۰۰۰۴۲
۱۲	MP51507B	داده	به کار بستن	۰/۰۰۰۵	۰/۹۸	۰/۰۰۰۰	-/۰۰۰۰۰
۱۳	MP71219	عدد	دانستن	۲/۰۸	۰/۱۵	۰/۰۰۷	-/۰۰۰۶۸
*۱۴	MP71021	عدد	به کار بستن	۱۰/۵۴	۰/۰۰۱	۰/۰۳	-/۰۰۰۲۶۱
۱۵	MP71167	عدد	دانستن	۰/۵۳	۰/۴۷	۰/۰۰۲	-/۰۰۰۲۴
۱۶	MP71041	عدد	به کار بستن	۰/۱۹	۰/۶۶	۰/۰۰۰۴	-/۰۰۰۰۴
۱۷	MP71162A	عدد	استدلال	۱/۸۲	۰/۱۷	۰/۰۰۹	-/۰۰۰۹۴
۱۸	MP71162B	عدد	استدلال	۰/۳۲	۰/۵۷	۰/۰۰۲	-/۰۰۰۱۸
۱۹	MP71162	عدد	استدلال	۰/۱۹	۰/۶۶	۰/۰۰۱	-/۰۰۰۱۵
۲۰	MP71078	اندازه‌گیری و هندسه	دانستن	۱/۱۴	۰/۲۸	۰/۰۰۳	-/۰۰۰۳۲
۲۱	MP71090	اندازه‌گیری و هندسه	استدلال	۰/۷۹	۰/۳۷	۰/۰۰۲	-/۰۰۰۲۳
۲۲	MP71151A	اندازه‌گیری و هندسه	به کار بستن	۳/۴۲	۰/۰۶	۰/۰۱۱	-/۰۰۰۱۱۸
۲۳	MP71151B	اندازه‌گیری و هندسه	به کار بستن	۱/۵۰	۰/۲۲	۰/۰۰۴	-/۰۰۰۴۶
۲۴	MP71151C	اندازه‌گیری و هندسه	به کار بستن	۰/۴۳	۰/۵۱	۰/۰۰۱	-/۰۰۰۱۶
۲۵	MP71151	اندازه‌گیری و هندسه	به کار بستن	۰/۳۲	۰/۵۶	۰/۰۰۱	-/۰۰۰۱۳
*۲۶	MP71119	اندازه‌گیری و هندسه	به کار بستن	۶/۲۳	۰/۰۱	۰/۰۲	-/۰۰۰۲۱۳
۲۷	MP71217A	داده	دانستن	۲/۶۹	۰/۱۰	۰/۰۰۸	-/۰۰۰۸۵
*۲۸	MP71142	داده	استدلال	۹/۸۴	۰/۰۰۱	۰/۰۲۴	-/۰۰۰۲۴۴
۲۹	MP71204A	داده	به کار بستن	۰/۰۰۰۲	۰/۹۹	۰/۰۰۰۰	-/۰۰۰۰۰
۳۰	MP71204B	داده	به کار بستن	۱/۷۰	۰/۱۹	۰/۰۰۷	-/۰۰۰۶۸
۳۱	MP71204C	داده	به کار بستن	۰/۱۵	۰/۷۰	۰/۰۰۰۶	-/۰۰۰۰۶

جدول ۳. کنش افتراقی یکنواخت سوال‌ها در دفترچه یک مبتنی بر رگرسیون لجستیک

ردیف	سؤال	حیطه محتوایی	حیطه شناختی	LRT آماره	P	R ²	مقدار اندازه اثر
۳۲	MP71204	داده	به کاربردن	۰/۳۱	۰/۵۸	۰/۰۰۱	۰/۰۰۱۵

همان‌طور که در جدول ۳ ملاحظه می‌شود مقدار معناداری آماره آزمون نسبت درستی^۱ در سوال‌های ۱۰، ۱۴، ۲۶ و ۲۸ دفترچه یک، کم‌تر از ۰/۰۵ است ($p < 0.05$) و بنابراین در این سوال‌ها کنش افتراقی وجود دارد و عملکرد دخترها و پسرها علی‌رغم داشتن توانایی یکسان، متفاوت است. در سوال ۱۰ با اندازه اثر ۰/۰۱۹۵ و سوال ۲۶ با اندازه اثر ۰/۰۲۱۳ کنش افتراقی به نفع دخترها و در سوال ۱۴ با اندازه اثر ۰/۰۲۶۱ و سوال ۲۸ با اندازه اثر ۰/۰۲۴۴ کنش افتراقی به نفع پسرها است. باینحال، در ستون مربوط به اندازه اثر جدول ۳، طبق معیار جودوین و گیرل (۲۰۰۱)، شاخص کنش افتراقی، حاکی از ناچیز بودن این شاخص در سوالات دارای کنش افتراقی است.

Logistic regression (LRT statistic)



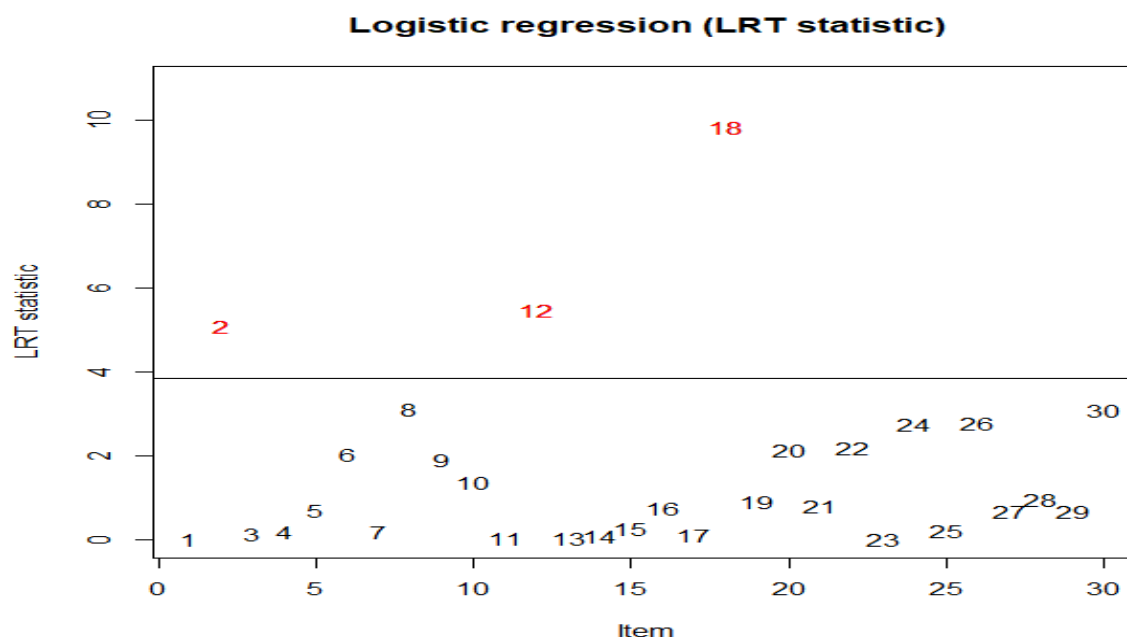
نمودار ۱- کنش افتراقی یکنواخت سوال‌ها آزمون ریاضیات پایه چهارم برحسب جنسیت در دفترچه یک با رگرسیون لجستیک

^۱. Likelihood Ratio Test

جدول ۴. کنش افتراقی یکنواخت سوال‌ها در دفترچه هفت مبتنی بر رگرسیون لجستیک

ردیف	سؤال	حیطه محتوایی	حیطه شناختی	آماره LRT	P	R ²	مقدار اندازه اثر
۱	MP51401	عدد	به کار بستن	۰/۰۰۲	۰/۹۷	۰/۰۰	۰/۰۰۰۰
*۲	MP51075	عدد	دانستن	۵/۰۸	۰/۰۲	۰/۰۲	۰/۰۱۶۲
۳	MP51402	عدد	به کار بستن	۰/۱۳	۰/۷۲	۰/۰۰۰۳	۰/۰۰۰۰۳
۴	MP51226	عدد	به کار بستن	۰/۱۸	۰/۶۷	۰/۰۰۰۵	۰/۰۰۰۰۵
۵	MP51131	عدد	دانستن	۰/۶۹	۰/۴۰	۰/۰۰۰۲	۰/۰۰۰۲۱
۶	MP51103	عدد	دانستن	۲/۰۳	۰/۱۵	۰/۰۰۰۶	۰/۰۰۰۵۷
۷	MP51217	اندازه‌گیری و هندسه	دانستن	۰/۱۸	۰/۶۷	۰/۰۰۰۵	۰/۰۰۰۰۵
۸	MP51079	اندازه‌گیری و هندسه	به کار بستن	۳/۱۰	۰/۰۸	۰/۰۰۰۷	۰/۰۰۰۷۲
۹	MP51211	اندازه‌گیری و هندسه	استدلال	۱/۹۱	۰/۱۷	۰/۰۰۰۵	۰/۰۰۰۵۵
۱۰	MP51102	اندازه‌گیری و هندسه	به کار بستن	۱/۳۷	۰/۲۴	۰/۰۰۰۳	۰/۰۰۰۳۵
۱۱	MP51009	داده	به کار بستن	۰/۰۲	۰/۸۹	۰/۰۰۰۰	۰/۰۰۰۰۰
*۱۲	MP51100	داده	استدلال	۵/۴۷	۰/۰۲	۰/۰۲	۰/۰۱۵۷
۱۳	MP71018	عدد	به کار بستن	۰/۰۳	۰/۸۶	۰/۰۰۰۱	۰/۰۰۰۰۱
۱۴	MP71009	عدد	به کار بستن	۰/۰۷	۰/۷۹	۰/۰۰۰۲	۰/۰۰۰۰۲
۱۵	MP71037	عدد	به کار بستن	۰/۲۷	۰/۶۰	۰/۰۰۰۱	۰/۰۰۰۱۰
۱۶	MP71051	عدد	به کار بستن	۰/۷۴	۰/۳۹	۰/۰۰۰۲	۰/۰۰۰۲۳
۱۷	MP71064	عدد	دانستن	۰/۱۱	۰/۷۳	۰/۰۰۰۴	۰/۰۰۰۰۴
*۱۸	MP71169	اندازه‌گیری و هندسه	دانستن	۹/۸۶	۰/۰۰۲	۰/۰۰۴	۰/۰۳۶۳
۱۹	MP71083	اندازه‌گیری و هندسه	استدلال	۰/۹۲	۰/۳۴	۰/۰۰۰۳	۰/۰۰۰۳۲
۲۰	MP71184	اندازه‌گیری و هندسه	استدلال	۰۲/۱۳	۰/۱۴	۰/۰۰۰۷	۰/۰۰۰۷۳
۲۱	MP71141A	داده	استدلال	۰/۷۹	۰/۳۷	۰/۰۰۰۵	۰/۰۰۰۴۸
۲۲	MP71141B	داده	استدلال	۲/۱۸	۰/۱۴	۰/۰۰۰۹	۰/۰۰۰۸۸
۲۳	MP71141C	داده	استدلال	۰/۰۰۳	۰/۹۶	۰/۰۰۰۰	۰/۰۰۰۰۰
۲۴	MP71141D	داده	استدلال	۲/۷۵	۰/۰۹	۰/۰۱	۰/۰۱۴۵
۲۵	MP71141	داده	استدلال	۰/۱۹	۰/۶۵	۰/۰۰۰۸	۰/۰۰۰۰۸
۲۶	MP71194	داده	به کار بستن	۲/۷۹	۰/۰۹	۰/۰۰۰۹	۰/۰۰۰۹۳
۲۷	MP71193A	داده	به کار بستن	۰/۶۷	۰/۴۱	۰/۰۰۰۲	۰/۰۰۰۲۴
۲۸	MP71193B	داده	به کار بستن	۰/۹۴	۰/۳۲	۰/۰۰۰۳	۰/۰۰۰۳۲
۲۹	MP71193	داده	به کار بستن	۰/۶۸	۰/۴۱	۰/۰۰۰۲	۰/۰۰۰۲۵
۳۰	MP71192	داده	به کار بستن	۳/۰۸	۰/۰۸	۰/۰۲	۰/۰۱۸۳

همان‌طور که در جدول ۴ ملاحظه می‌شود مقدار معناداری آماره مورد بررسی در سوال‌های ۲، ۱۲ و ۱۸ کم‌تر از ۰/۰۵ است ($p < 0.05$). بنابراین در این سوال‌ها کنش افتراقی وجود دارد و عملکرد دخترها و پسرها علی‌رغم داشتن توانایی یکسان، متفاوت است. سوال ۲ با اندازه اثر ۰/۰۱۶ به نفع دختران و سوالات ۱۲ با اندازه اثر ۰/۰۱۵ و ۱۸ با اندازه اثر ۰/۰۳۶ به نفع پسران است. باتوجه به جدول ۴ و در ستون مربوط به اندازه اثر، طبق معیار جودوین و گیرل (۲۰۰۱)، شاخص کنش افتراقی، حاکی از ناچیز بودن این شاخص در همه سوالات دارای کنش افتراقی به جز سوال ۱۸ (دارای اندازه اثر متوسط) است.



نمودار ۲- کنش افتراقی یکنواخت سوال‌ها آزمون ریاضیات پایه چهارم برحسب جنسیت در دفترچه هفت با رگرسیون لجستیک

جدول ۵. کنش افتراقی سوال‌ها در دفترچه یک با روش خی دو لرد مبتنی بر مدل راش

ردیف	سؤال	دشواری سؤال (خطای استاندارد)	دشواری سؤال پسرها (خطای استاندارد)	دشواری سؤال دخترها (خطای استاندارد)	خی دو لرد	معناداری	مقدار اندازه اثر
۱	MP51043	۰/۶۹ (۰/۱۳)	۰/۶۴ (۰/۱۷)	۰/۷۵ (۰/۱۸)	۰/۹۳	۰/۳۳	۰/۵۷
۲	MP51040	-۲/۶۸ (۰/۱۹)	-۲/۸۷ (۰/۲۸)	-۲/۵۲ (۰/۲۵)	۰/۶۴	۰/۲۰	۱/۱۴
۳	MP51008	۲/۲۸ (۱۶)	۲/۵۰ (۰/۲۵)	۲/۰۸ (۰/۲۲)	۰/۷۴	۰/۳۹	-۰/۶۶
۴	MP51031A	-۰/۴۹ (۰/۱۲)	-۰/۶۴ (۰/۱۷)	-۰/۳۴ (۰/۱۸)	۳/۰۸	۰/۰۸	۱/۰۲
۵	MP51031B	-۰/۲۱ (۰/۱۲)	-۰/۲۵ (۰/۱۷)	-۰/۱۶ (۰/۱۸)	۰/۹۱	۰/۳۴	۰/۵۵
۶	MP51508	۰/۴۱ (۰/۱۳)	۰/۵۹ (۰/۱۸)	۰/۲۱ (۰/۱۸)	۳/۷۵	۰/۰۵	-۰/۵۷
۷	MP51216A	۰/۱۷ (۰/۱۲)	۰/۰۸ (۰/۱۷)	۰/۳۴ (۰/۱۷)	۰/۸۶	۰/۳۵	۱/۱۱

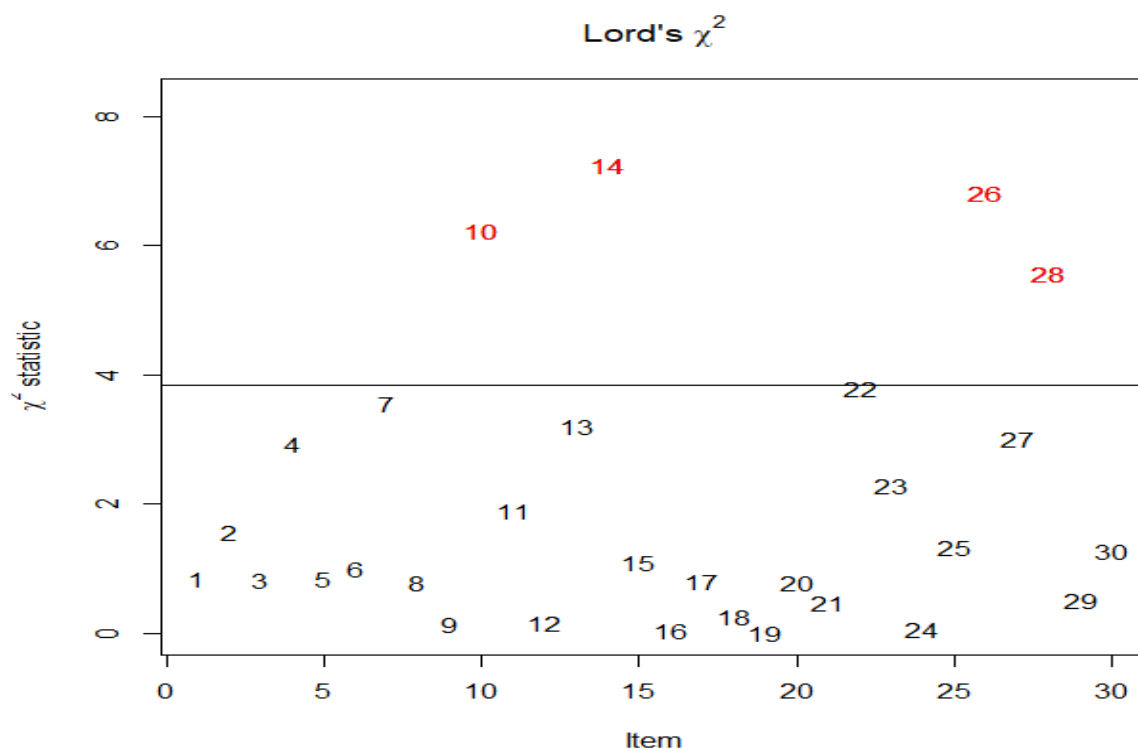
جدول ۵. کنش افتراقی سوال‌ها در دفترچه یک با روش خی دو لرد مبتنی بر مدل راش

ردیف	سؤال	دشواری سؤال (خطای استاندارد)	دشواری سؤال پسرها (خطای استاندارد)	دشواری سؤال دخترها (خطای استاندارد)	خی دو لرد	معناداری	مقدار اندازه اثر
۸	MP51216B	(۰/۱۴) - ۱/۴۹	(۰/۱۹) - ۱/۵۵	(۰/۱۹) - ۱/۴۳	۰/۷۴	۰/۳۹	۰/۶۰
۹	MP51221	(۰/۱۲) - ۰/۳۸	(۰/۱۷) - ۰/۲۸	(۰/۱۷) - ۰/۴۹	۰/۱۱	۰/۷۴	-۰/۱۹
۱۰*	MP51115	(۰/۱۵) ۱/۶۹	(۰/۱۹) ۱/۴۱	(۰/۲۲) ۲/۰۲	۶/۴۶	۰/۰۱	۱/۷۸
۱۱	MP51507A	(۰/۱۲) - ۰/۴۵	(۰/۱۷) - ۰/۵۶	(۰/۱۸) - ۰/۳۴	۲/۰۱	۰/۱۵	-۰/۸۳
۱۲	MP51507B	(۰/۲۰) ۲/۸۱	(۰/۲۹) ۲/۹۵	(۰/۲۷) ۲/۶۷	۰/۱۳	۰/۷۱	-۰/۳۴
۱۳	MP71219	(۰/۱۴) - ۱/۶۳	(۰/۲۱) - ۱/۸۳	(۰/۱۹) - ۱/۴۴	۳/۳۵	۰/۰۶	۱/۲۵
۱۴*	MP71021	(۰/۱۳) ۰/۴۷	(۰/۱۹) ۰/۸۹	(۰/۱۸) ۰/۰۶	۷/۰۹	۰/۰۰۷	-۱/۶۳
۱۵	MP71167	(۰/۱۹) ۲/۶۵	(۰/۲۸) ۲/۹۳	(۰/۲۵) ۲/۴۱	۱/۰۰۹	۰/۳۱	-۰/۸۹
۱۶	MP71041	(۰/۱۲) - ۰/۰۷	(۰/۱۷) ۰/۰۲	(۰/۱۸) - ۰/۱۶	۰/۰۳	۰/۸۶	-۰/۱۰
۱۷	MP71162A	(۰/۲۰) ۲/۶۸	(۰/۲۸) ۲/۵۷	(۰/۲۹) ۲/۸۰	۰/۸۵	۰/۳۵	-۰/۸۷
۱۸	MP71162B	(۰/۲۰) ۲/۵۶	(۰/۳۱) ۲/۷۴	(۰/۲۷) ۲/۴۱	۰/۲۱	۰/۶۴	-۰/۴۵
۱۹	MP71162	(۰/۱۶) ۳/۴۸	(۰/۳۸) ۳/۵۷	(۰/۳۵) ۳/۹۵	۰/۰۰۰۴	۰/۹۸	-۰/۰۸
۲۰	MP71078	(۰/۱۲) ۰/۱۷	(۰/۱۷) ۰/۳۴	(۰/۱۸) - ۰/۰۰۹	۰/۷۰	۰/۴۰	-۰/۴۹
۲۱	MP71090	(۰/۱۲) ۰/۰۵	(۰/۱۸) ۰/۱۹	(۰/۱۸) - ۰/۱۰	۰/۴۱	۰/۵۲	-۰/۳۸
۲۲	MP71151A	(۰/۱۶) ۱/۴۶	(۰/۲۴) ۱/۸۶	(۰/۲۱) ۱/۰۹	۳/۷۳	۰/۰۵	-۱/۴۶
۲۳	MP71151B	(۰/۱۶) ۱/۱۶	(۰/۲۲) ۱/۴۶	(۰/۲۱) ۰/۸۷	۲/۱۸	۰/۱۴	-۱/۰۸
۲۴	MP71151C	(۰/۱۷) ۱/۶۲	(۰/۲۴) ۱/۶۵	(۰/۲۳) ۱/۵۹	۰/۰۶	۰/۸۱	-۰/۱۹
۲۵	MP71151	(۰/۱۸) ۲/۲۶	(۰/۲۸) ۲/۵۷	(۰/۲۴) ۲/۰۰۵	۱/۲۷	۰/۲۶	-۰/۹۹
۲۶*	MP71119	(۰/۱۵) - ۱/۷۷	(۰/۲۴) - ۲/۱۴	(۰/۲۰) - ۱/۴۴	۶/۹۹	۰/۰۰۸	۱/۹۷
۲۷	MP71217A	(۰/۱۴) - ۰/۷۴	(۰/۱۹) - ۰/۹۵	(۰/۱۹) ۰/۶۱	۳/۱۷	۰/۰۷	۱/۱۵
۲۸*	MP71142	(۰/۱۳) - ۰/۲۸	(۰/۱۸) ۰/۰۸	(۰/۱۹) - ۰/۶۷	۵/۳۸	۰/۰۲	-۱/۴۵
۲۹	MP71204A	(۰/۲۰) ۲/۱۳	(۰/۳۱) ۲/۳۶	(۰/۱۹) ۱/۹۵	۰/۴۵	۰/۵۰	-۰/۶۴
۳۰	MP71204B	(۰/۱۶) ۰/۰۹	(۰/۲۲) ۰/۳۲	(۰/۲۳) - ۰/۱۷	۱/۱۸	۰/۲۸	-۰/۸۲
۳۱	MP71204C	(۰/۱۶) ۰/۴۶	(۰/۲۲) ۰/۴۹	(۰/۲۳) ۰/۴۴	۰/۰۷	۰/۷۹	-۰/۲۰
۳۲	MP71204	(۰/۲۰) ۲/۳۴	(۰/۳۱) ۲/۶۵	(۰/۲۶) ۲/۰۸	۱/۱۵	۰/۲۸	-۱/۰۱

از آنجا که در روش خی دو مبتنی بر مدل راش، آستانه (نمره برش^۱) برای طبقه‌بندی سؤال‌های دارای کنش افتراقی، $3/84$ است (مگیس، بلند و رایچه، ۲۰۲۲)، در جدول ۵، سوال‌های ۱۰، ۱۴، ۲۶ و ۲۸ دارای کنش افتراقی هستند. سوال‌های ۱۰ و ۲۶ به نفع پسرها و سوال‌ها ۱۴ و ۲۸

¹. Threshold (cut-score)

به نفع دخترها است. همچنین، طبق معیار مقیاس دلتای ای تی اس^۱ (هالند و تایر، ۱۹۸۵)، سوالات ۱۰، ۱۴ و ۲۶ دارای اندازه اثر زیاد و سوال ۲۸ دارای اندازه اثر متوسط است.



نمودار ۳- کنش افتراقی یکنواخت سوال‌ها آزمون ریاضیات پایه چهارم برحسب جنسیت در دفترچه یک با استفاده از روش خی دو لرد

جدول ۶. کنش افتراقی سوال‌ها در دفترچه هفت با روش خی دو لرد مبتنی بر مدل راش

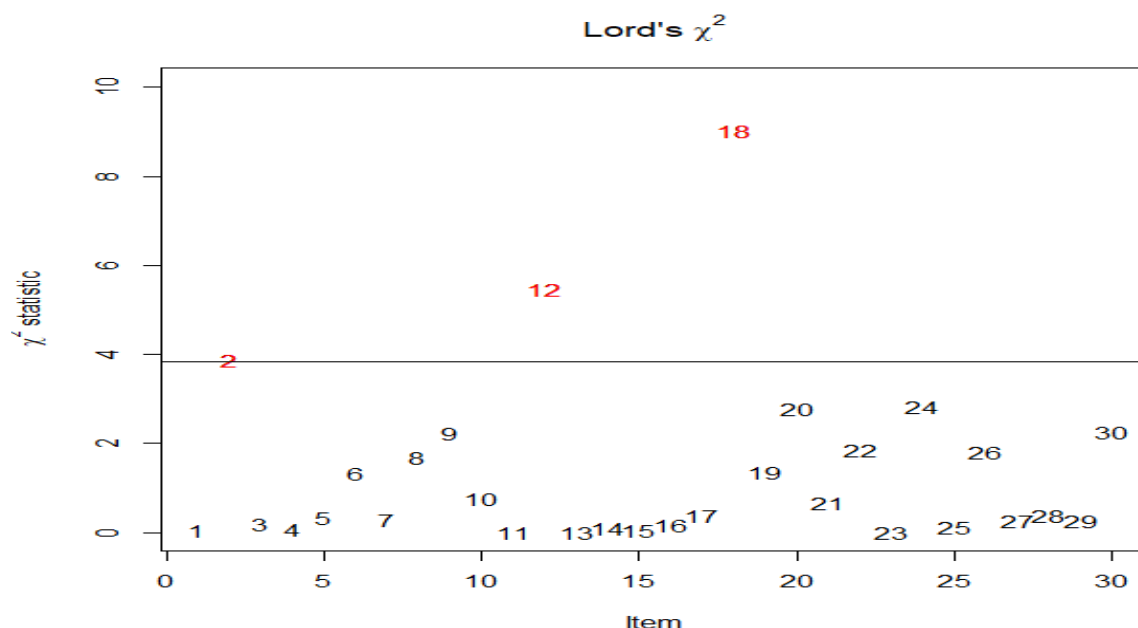
ردیف	سؤال	دشواری سؤال (خطا استاندارد)	دشواری سؤال پسرها (خطا استاندارد)	دشواری سؤال دخترها (خطا استاندارد)	خی دو لرد	معناداری	مقدارانداز ه اثر
۱	MP51401	۰/۷۶ (۰/۱۳)	۰/۷۸ (۰/۱۸)	۰/۷۵ (۰/۱۸)	۰/۰۶	۰/۸۱	-۰/۱۴۳۸
*۲	MP51075	-۱/۵۱ (۰/۱۳)	-۱/۸۱ (۰/۲۱)	-۱/۲۵ (۰/۱۸)	۳/۸۷	۰/۰۴۹	۱/۲۷۲۵
۳	MP51402	۰/۸۹ (۰/۱۳)	۰/۹۵ (۰/۱۹)	۰/۸۶ (۰/۱۸)	۰/۱۹	۰/۶۵	-۰/۲۶۹۳
۴	MP51226	-۰/۳۶ (۰/۱۲)	-۰/۴۱ (۰/۱۷)	-۰/۳۲ (۰/۱۷)	۰/۰۸	۰/۷۸	۰/۱۵۷۹

^۱. ETS delta scale

جدول ۶. کنش افتراقی سوال‌ها در دفترچه هفت با روش خی دو لرد مبتنی بر مدل راش

ردیف	سؤال	دشواری سؤال (خطا استاندارد)	دشواری سؤال پسرها (خطا استاندارد)	دشواری سؤال دخترها (خطا استاندارد)	خی دو لرد	معناداری	مقدارانداز ه اثر
۵	MP51131	۰/۰۲- (۰/۱۳)	۰/۱۰- (۰/۱۹)	۰/۰۷ (۰/۱۸)	۰/۳۴	۰/۵۶	۰/۳۵۴۶
۶	MP51103	۰/۳۱- (۱۲)	۰/۴۷- (۰/۱۷)	۰/۱۷- (۰/۱۷)	۱/۳۴	۰/۲۵	۰/۶۴۸۸
۷	MP51217	۱/۵۶ (۰/۱۴)	۱/۶۲ (۰/۲۰)	۱/۴۹ (۰/۱۹)	۰/۲۸	۰/۶۰	-۰/۳۴۴۵
۸	MP51079	۰/۰۶- (۰/۱۲)	۰/۱۷- (۰/۱۷)	۱/۵۵ (۰/۱۶)	۱/۶۸	۰/۱۹	۰/۷۱۵۸
۹	MP51211	۰/۵۷- (۱۲)	۰/۳۹- (۰/۱۷)	۰/۷۴- (۰/۱۷)	۲/۲۳	۰/۱۳	-۰/۸۴۴۴
۱۰	MP51102	۰/۱۹ (۱۲)	۰/۰۷ (۰/۱۷)	۰/۳۰ (۰/۱۷)	۰/۷۷	۰/۳۸	۰/۴۹۴۲
۱۱	MP51009	۰/۲۹ (۰/۱۲)	۰/۲۸ (۰/۱۷)	۰/۳۱ (۰/۱۷)	۰/۰۰۰	۰/۹۹	-۰/۰۰۱۴
۱۲*	MP51100	۰/۰۷- (۰/۱۲)	۰/۲۱ (۰/۱۷)	۰/۳۳- (۰/۱۷)	۵/۴۵	۰/۰۲	-۱/۳۳۵۳
۱۳	MP71018	۰/۴۴- (۱۲)	۰/۴۵- (۰/۱۷)	۰/۴۲ (۰/۱۷)	۰/۰۰۰۳	۰/۹۹	۰/۰۰۸۹
۱۴	MP71009	۰/۹۵ (۱۳)	۰/۹۸ (۰/۱۹)	۰/۹۱ (۰/۱۸)	۰/۱۱	۰/۷۴	-۰/۲۰۴۹
۱۵	MP71037	۱/۶۰ (۰/۱۵)	۱/۵۶ (۰/۲۲)	۱/۶۵ (۰/۲۲)	۰/۰۵	۰/۸۲	۰/۱۶۴۰
۱۶	MP71051	۱/۸۶ (۱۵)	۱/۷۹ (۰/۲۱)	۱/۹۴ (۰/۲۲)	۰/۱۸	۰/۶۷	۰/۳۰۳۲
۱۷	MP71064	۰/۹۲ (۰/۱۳)	۰/۹۹ (۰/۱۸)	۰/۸۶ (۰/۱۸)	۰/۳۸	۰/۵۳	-۰/۳۷۳۴
۱۸*	MP71169	۲/۱۰ (۰/۱۶)	۲/۶۶ (۰/۲۶)	۱/۶۸ (۰/۲۰)	۹/۰۳	۰/۰۰۳	-۲/۳۶۶۰
۱۹	MP71083	۱/۰۴ (۰/۱۴)	۱/۱۹ (۰/۲۰)	۰/۸۹ (۰/۱۹)	۱/۳۶	۰/۲۴	-۰/۷۵۷۹
۲۰	MP71184	۰/۹۳ (۰/۱۴)	۱/۱۶ (۰/۲۰)	۰/۷۳ (۰/۱۹)	۲/۷۷	۰/۰۹	-۱/۰۷۴۲
۲۱	MP71141A	۰/۳۱- (۰/۲۰)	۰/۵۰- (۰/۳۱)	۰/۱۴- (۰/۲۸)	۰/۶۶	۰/۴۲	۰/۷۸۸۰
۲۲	MP71141B	۰/۰۴- (۰/۱۵)	۰/۲۸- (۰/۲۳)	۰/۱۶ (۰/۲۱)	۱/۸۶	۰/۱۷	۰/۹۹۲۹
۲۳	MP71141C	۰/۷۲ (۰/۱۶)	۰/۷۰ (۰/۲۳)	۰/۷۳ (۰/۲۲)	۰/۰۰۰۱	۰/۹۹	۰/۰۰۷۱
۲۴	MP71141D	۰/۴۲- (۰/۱۹)	۰/۱۳- (۰/۲۵)	۰/۷۳- (۰/۲۷)	۲/۸۳	۰/۰۹	-۱/۴۶۸۵
۲۵	MP71141	۱/۹۳ (۰/۱۸)	۱/۹۸ (۰/۲۶)	۱/۸۹ (۰/۲۴)	۰/۱۲	۰/۷۲	-۰/۲۹۴۹
۲۶	MP71194	۰/۵۰- (۰/۱۴)	۰/۷۱- (۰/۱۹)	۰/۳۱- (۰/۱۹)	۱/۸۱	۰/۱۸	۰/۸۷۴۴
۲۷	MP71193A	۱/۵۷ (۰/۱۶)	۱/۴۸ (۰/۲۲)	۱/۶۶ (۰/۲۲)	۰/۲۵	۰/۶۱	۰/۳۷۴۸
۲۸	MP71193B	۱/۲۱ (۰/۱۵)	۱/۱۱ (۰/۲۱)	۱/۳۱ (۰/۲۱)	۰/۳۷	۰/۵۴	۰/۴۳۰۰
۲۹	MP71193	۱/۹۷ (۰/۱۷)	۱/۸۸ (۰/۲۴)	۲/۰۷ (۰/۲۴)	۰/۲۵	۰/۶۱	۰/۴۰۴۴
۳۰	MP71192	۲/۳۷ (۰/۲۰)	۲/۰۷ (۰/۲۷)	۲/۷۲ (۰/۳۲)	۲/۲۷	۰/۱۳	۱/۴۸۷۶

طبق جدول ۶ در سوال‌های ۲، ۱۲ و ۱۸ کنش افتراقی سؤال وجود دارد. سؤال ۲ به نفع دخترها و سوال‌های ۱۲ و ۱۸ به نفع پسرها است. طبق معیار هالند و تایر (۱۹۸۵)، سوال ۱۸ دارای اندازه اثر زیاد و سوالات ۲ و ۱۲ دارای اندازه اثر متوسط هستند.



نمودار ۴- کنش افتراقی یکنواخت سوال‌ها آزمون ریاضیات پایه چهارم برحسب جنسیت در دفترچه هفت با استفاده از روش خی دو لرد

در جدول شماره ۷، برخی ویژگی‌های سوال‌هایی که دارای کنش افتراقی هستند، ارائه شده است.

جدول ۷. برخی ویژگی‌های سوالات شناسایی شده با کنش افتراقی

دفترچه	سوال	کد سوال	حیطه محتوا	حیطه شناختی	زمینه سوال	نوع سوال	به نفع
۱	۱۰	MP51115	اندازه‌گیری و هندسه (هندسه)	استدلال	انتزاعی	چند گزینه‌ای	دختر
۱	۱۴	MP71021	عدد (عدد صحیح)	به‌کار بستن	مرتبط با زندگی	چند گزینه‌ای	پسر
۱	۲۶	MP71119	اندازه‌گیری و هندسه (هندسه)	به‌کار بستن	انتزاعی	پاسخ ساز	دختر
۱	۲۸	MP71142	داده (خواندن، تفسیر و ارائه)	استدلال	مرتبط با زندگی	پاسخ ساز	پسر
۷	۲	MP51075	عدد (کسر و اعشار)	دانستن	انتزاعی	چند گزینه‌ای	دختر
۷	۱۲	MP51100	داده (خواندن، تفسیر و ارائه)	استدلال	مرتبط با زندگی	چند گزینه‌ای	پسر
۷	۱۸	MP71169	اندازه‌گیری و هندسه (اندازه‌گیری)	به‌کار بستن	غیرقابل انتشار	چند گزینه‌ای	پسر

طبق جدول ۷، از میان هفت سوال شناسایی شده که دارای کنش افتراقی بودند، شش سوال قابلیت انتشار داشتند که در پیوست این مقاله، سوال ۱۴ از دفترچه یک (زمینه مرتبط با زندگی واقعی) و سوال ۲ از دفترچه هفت (زمینه انتزاعی) ارایه شده‌اند.

بحث

یکی از ویژگی‌های آزمون‌های معتبر، عادلانه بودن سؤال‌های آزمون است، بدین معنا که افراد از گروه‌های متفاوت مانند جنسیت و طبقه اجتماعی - اقتصادی ولی با توانایی‌های مشابه، عملکرد مشابهی داشته باشند، این ویژگی از طریق کنش افتراقی بررسی می‌شود. در این

پژوهش با هدف شناسایی سوالات دارای کنش افتراقی در آزمون ریاضیات تیمز ۲۰۱۹ برحسب جنسیت، از دو مدل لجستیک و χ^2 دو مبتنی بر مدل راش استفاده شد. یافته‌ها نشان داد از ۳۲ سوال دفترچه یک، چهار سوال طبق هر دو مدل تحلیلی، دارای کنش افتراقی بود که دو سؤال به نفع دختران و دو سؤال به نفع پسران بود. همچنین از ۳۰ سؤال دفترچه هفت، سه سؤال طبق هر دو مدل تحلیلی، دارای کنش افتراقی شد که دو سؤال به نفع پسران و یک سؤال به نفع دختران بود. در دفترچه یک، سؤال ۱۰ با ضریب دشواری مبتنی بر مدل راش $۱/۶۹$ در حیطه محتوایی "اندازه‌گیری و هندسه" و موضوع "هندسه" و در سطح شناختی استدلال به صورت چندگزینه‌ای و به نفع دختران بود. سؤال ۱۴ ضریب دشواری مبتنی بر مدل راش $۰/۴۷$ داشت و مربوط به حیطه محتوایی اعداد و موضوع "اعداد صحیح" در سطح شناختی کاربرد و به صورت چندگزینه‌ای و به نفع پسران بود. سؤال ۲۶ نیز با ضریب دشواری مبتنی بر مدل راش $۱/۷۷$ - مجدداً حیطه محتوایی "اندازه‌گیری و هندسه" را با موضوعی "هندسه" ولی در سطح شناختی کاربرد؛ به صورت پاسخ ساز بوده و به نفع دختران بود. در نهایت سؤال ۲۸ با ضریب دشواری مبتنی بر مدل راش $۰/۲۸$ - با محتوای "داده" و موضوع "خواندن، تفسیر و ارائه" در سطح شناختی استدلال به صورت پاسخ ساز و به نفع پسران بود. همچنین در دفترچه هفت، سؤال ۲ با ضریب دشواری مبتنی بر مدل راش $۱/۵۱$ - در حیطه محتوایی "اعداد" و موضوع "کسر و اعشار" در سطح شناختی دانش، به صورت چندگزینه‌ای و به نفع دختران بود. سؤال ۱۲ نیز با ضریب دشواری مبتنی بر مدل راش $۰/۰۷$ - حیطه محتوایی "داده" با موضوعیت "خواندن، تفسیر و ارائه" در سطح شناختی استدلال، چندگزینه‌ای و به نفع پسران بود. سؤال ۱۸ با ضریب دشواری مبتنی بر مدل راش $۲/۱۰$ محتوای "اندازه‌گیری و هندسه" با موضوع "اندازه‌گیری" در سطح شناختی کاربرد به صورت چهارگزینه‌ای و به نفع پسران بود. به علاوه، از دیگر یافته‌های این پژوهش شناسایی شدت اثر سوال‌ها دارای کنش افتراقی بود. اگرچه نتایج معناداری آماری حاکی از تشابه نتایج به دست آمده از دو مدل تحلیلی رگرسیون لجستیک و مدل χ^2 دو لرد مبتنی بر راش بود، اما نتایج معناداری عملی باتوجه به اندازه اثر نشان داد که با روش رگرسیون لجستیک سوال‌ها ۱۰، ۱۴، ۲۶ و ۲۸ از دفترچه یک و ۲ و ۱۲ از دفترچه هفت از آنجا که مقدار اندازه اثرشان کمتر از $۰/۰۳۵$ بود، کنش افتراقی ناچیز و قابل اغماض داشتند، ولی سؤال ۱۸ از دفترچه هفت بین مقدار $۰/۰۳۵$ و $۰/۰۷$ قرار داشت و دارای اندازه اثر متوسط بود. اما، با روش χ^2 دو لرد مبتنی بر مدل راش سؤال ۲۸ از دفترچه یک و سؤال‌ها ۲ و ۱۲ از دفترچه هفت دارای اندازه اثر متوسط بودند، چراکه مقدار اندازه اثرشان بین ۱ تا $۱/۵$ قرار داشت. سؤال‌ها ۱۰، ۱۴، ۲۶ از دفترچه یک و سؤال ۱۸ از دفترچه هفت با مقدار اندازه اثر بیشتر از $۱/۵$ دارای اندازه اثر زیاد بودند. به طور کلی یافته‌های این پژوهش با پژوهش سبحانی سیناب (۱۴۰۰) که با روش رگرسیون لجستیک دریافت، از ۶۲ سؤال چندگزینه‌ای ریاضیات پایه هشتم تیمز ۲۰۱۹، ۱۱ سؤال با کنش افتراقی جنسیتی وجود دارد، همخوانی دارد. همچنین با پژوهش‌های عسکری تورزن (۱۴۰۰)، مینایی و همکاران (۱۳۹۹)، مینایی و غفاری (۱۳۹۴) و صادقی (۱۳۹۲) نیز که در بررسی کنش افتراقی آزمون ریاضیات تیمز در دوره‌های مختلف دریافتند که کنش افتراقی در سوال‌ها وجود دارد، همسو است.

نتیجه‌گیری

به طور کلی روش‌های آماری کنش افتراقی براساس جنسیت، به عنوان یک تکنیک روانسنجی و در راستای شناسایی الگوها در سوال‌ها با سرخ‌گرایی به نفع دختران یا پسران، کاربرد دارد. چنین اطلاعاتی می‌تواند در خصوص درک بهتر تفاوت در یادگیری ریاضی دختران و پسران کمک کند. نتایج این پژوهش نشان داد با کنترل سطح پیشرفت عمومی، تفاوت جنسیتی در هفت سوال در دو دفترچه وجود دارد. طبق یافته‌ها، سرخ‌گرایی به نفع دختران یا پسران باتوجه به دشواری سوال‌ها یافت نشد. همچنین، تأثیر نوع سوال (باز پاسخ / چندگزینه‌ای) به عنوان علت احتمالی کنش افتراقی شناسایی نشد، چون از میان سوال‌های دارای کنش افتراقی، دو سوال پاسخ ساز (یک سوال به نفع دختران / یک سوال به نفع پسران) و پنج سوال چندگزینه‌ای (سه سوال به نفع پسران / دو سوال به نفع دختران) یافت شد. در زمینه تأثیر حیطه محتوایی نتایج اندکی متفاوت بود. در حیطه عدد، یک سوال به نفع پسران و یک سوال به نفع دختران بود. اما در حیطه اندازه‌گیری و هندسه، دو سوال به نفع دختران و یک سوال به نفع پسران بود و این برخلاف نظر دولیتل (۱۹۸۹) مبنی بر برتری پسران در هندسه نسبت به دختران است. همچنین در حیطه داده، کنش افتراقی هر دو سوال، به نفع پسران بود. به علاوه، طبق پیشینه پژوهشی یکی دیگر از عوامل احتمالی کنش افتراقی، سطوح شناختی

است که در این پژوهش یافته‌ها نشان داد، در سطح دانستن به سوال به نفع دختران، در سطح به‌کار بستن، دو سوال به نفع پسران و یک سوال به نفع دختران بود و در سطح استدلال، مشابه سطح به‌کار بستن، دو سوال به نفع پسران و یک سوال به نفع دختران بود و این یافته با نتایج پژوهش (لان، ۲۰۱۴) همخوانی دارد که پسران در سطوح بالاتر شناختی، عملکرد بهتری دارند. در بررسی دقیق‌تر زمینه سوال‌ها، مشخص شد چنانچه ساختار سوال‌ها را به دو دسته سوال‌های نمادین^۱ (انتزاعی) و سوال‌های مرتبط با زندگی واقعی تقسیم نمود، در این پژوهش، دختران در سوال‌های انتزاعی و پسران در سوال‌های مرتبط با زندگی واقعی عملکرد بهتری داشتند. این یافته با نتایج به دست آمده از پژوهش اینابی و دودین (۲۰۱۸) همخوانی دارد. باتوجه به نتایج این پژوهش بایستی برای فهم عمیق‌تر ریشه تفاوت‌های عملکرد دانش‌آموزان دختر و پسر، در مطالعات آتی از مصاحبه مبتنی بر تکلیف با دانش‌آموزان دو جنس برای شنا سایی علل کنش افتراقی استفاده شود تا با برنامه‌ریزی متناسب با توانایی‌های دو جنس، جهت پیشرفت و ارتقا ریاضیات دانش‌آموزان بدون سوگیری جنسیتی بتوان گام برداشت.

References

- Alimirzaie, M., Mogadam, Z. A., Minaei, A., Ezanloo, B., & Salehi, K. (2019). Sources of the differential item functioning and its application in education. *Journal of Research in Teaching*, 7(1), 129-149.
- Ali moussa, M. (2019). Item bias approaches comparison of TIMSS test among Egyption environment. *International journal of research in educational sciences*, 2(4), 501-540.
- Arim, R. G., & Ercikan, K. (2014). Comparability between the American and Turkish Versions of the TIMSS Mathematics Test Results. *Education & Science/Egitim ve Bilim*, 39(172), 33-48.
- Askari Turzen, D. (2021). *Comparison and application of the mixed method of rush, question-oriented tree and lasso in identifying the differential item functioning of mathematical questions of the 2015 TIMSS 8th grade test*. Unpublished Master's thesis, Faculty of Psychology and Education, Allameh Tabatabai University.
- Bolt, D. M. (2000). A SIBTEST approach to testing DIF hypotheses using experimentally designed test items. *Journal of Educational Measurement*, 37(4), 307-327.
- Doolittle, A. E. (1989). Gender differences in performance on mathematics achievement items. *Applied Measurement in Education*, 2, 161-177.
- Gökçe, S., Berberoğlu, G., Wells, C. S., & Sireci, S. G. (2021). Linguistic Distance and Translation Differential Item Functioning on Trends in International Mathematics and Science Study Mathematics Assessment Items. *Journal of Psychoeducational Assessment*, 39(6), 728-745.
- Hambleton, R., & Rogers, J. (1994). Item bias review. *Practical Assessment, Research, and Evaluation*, 4(1), Article 6. Available at: <https://scholarworks.umass.edu/pare/vol4/iss1/6>
- Hashemi, M. (2010). Examining the bias of the 2007 TIMSS test questions in the third grade of middle school in the science course with regard to gender and ethnicity variables, as well as comparing classical theories and question-answer. Unpublished Master's Thesis, Faculty of Psychology and Education, Allameh Tabatabai University.
- Holland, P. W., & Thayer, D. T. (1985). An alternate definition of the ETS delta scale of item difficulty. *ETS Research Report Series*, 1985(2), 1-10.
- Innabi, H., & Dodeen, H. (2018). Gender differences in mathematics achievement in Jordan: A differential item functioning analysis of the 2015 TIMSS. *School Science and Mathematics*, 118(3-4), 127-137.
- Jodoin, M. G. and Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329-349.
- Kiamanesh, A., Mohsenpour, M., Safarkhani, M., Aghdasi, S. (2012). The trend of change in mathematics performance of grade 8 Iranian students between 1999 to 2007 based on the TIMSS findings in Iran and

¹. Symbolic Problems

- other countries of the region considering the 20 years vision document of Iran, *Journal of curriculum studies*, 6(24), 59-82.
- Lan, M. C. (2014). Exploring gender differential item functioning (DIF) on eighth grade mathematics items for the United States and Taiwan (Doctoral dissertation).
- Magis, D., Sebastien Beland, S., and Raiche, G. (2022). Collection of Methods to Detect Dichotomous Differential Item Functioning (DIF), difR package, Version 5.1.
- Minaei, A. (2012). Measuring the comparability of the structure and analyzing the differential item functioning of questions (DIF) and blocks (DTF) of the 2007 TIMSS 8th grade science test among Iranian and American students. *Educational Measurement*.4(11), 113-151.
- Minaei, A. and Ghaffari, Z. (2014). The differential item functioning of the eighth grade questions of the 2011 TIMSS math test among male and female students using the question-answer theory (IRT) approach. *Educational measurement*. 6(21), 21-40.
- Minaei, A., Younesi, J. and Kazemi Dana, B. (2019). The application of cognitive diagnostic models (CDM) in the study of the differential item functioning of 2007 TIMSS 8th grade mathematics questions among boys and girls. *Educational Measurement and Studies*. 10(32), 205-234.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center.
- Payne, D., A. (2003). *Applied Educational Assessment*, 2nd ed, Wadsworth.
- Sadeghi, M. (2012). Differential Item Functioning (DIF) of TIMSS 2011 4th and 8th grade math test questions using IRT and classical models. Unpublished Master's Thesis, Faculty of Social Sciences, Islamic Azad University.
- Sobhani Sinab, L. (2021). The application of logistic and multinomial regression in identifying the differential Item Functioning (DIF) and options (DOF) of the 2019 TIMSS 8th grade math test between girls and boys in Iran. Unpublished Master's Thesis, Faculty of Psychology and Education. Allameh Tabatabai University.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.