

کارکرد افتراقی سؤال در چارچوب سنجش تشخیصی شناختی: مطالعه موردی

سؤال‌های حساب دیفرانسیل و انتگرال کنکور سراسری ۹۷

Differential Item Functioning in the Framework of Cognitive Diagnostic Assessment: Questions Related to the Differential and Integral Calculus of the Iranian National University Entrance Examination 2018

سمیه سعادت^۱، علی مقدم‌زاده^{۲*}، اصغر مینائی^۳، مسعود گرامی‌پور^۴

پذیرش مقاله: ۱۳۹۹/۰۲/۲۶

دریافت مقاله: ۱۳۹۸/۱۲/۲۱

چکیده

هدف: منصفانه بودن سؤال‌های آزمون‌های سراسری بنا به ادعان بسیاری از متخصصان همواره حائز اهمیت است. از آنجا که تعیین سؤال‌های دارای کارکرد افتراقی می‌تواند گامی مهم در جهت منصفانه بودن آزمون‌های سرنوشت‌ساز باشد، بنابراین در پژوهش حاضر، کارکرد افتراقی سؤال‌های ریاضی بخش حساب دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه‌های کشور در میان داوطلبان دختر و پسر با استفاده از آزمون والد تحت مدل تشخیصی شناختی دینا مورد مطالعه قرار گرفت.

روش: این پژوهش توصیفی از نوع تحلیل ثانویه در چارچوب مدلیابی تشخیصی شناختی است. نمونه شامل ۵۰۰۰ نفر از داوطلبان کنکور سال ۱۳۹۷ است که از میان ۱۴۴۴۳۷ نفر شرکت‌کننده در گروه آزمایشی ریاضی و فیزیک به روش نمونه‌گیری تصادفی انتخاب شدند. بعد از شناسایی و تعیین مهارت‌های زیربنایی هر یک از سؤال‌ها با کمک متخصصین، شاخص‌های برازش مدل و پارامترهای سؤال‌ها بر مبنای مدل غیرجبرانی دینا تحلیل شد. یافته‌ها نشان‌دهنده برازش مناسب مدل با داده‌ها بود. به‌منظور بررسی کارکرد افتراقی سؤال از آزمون والد استفاده شد.

یافته‌ها: نتایج نشان داد که از ۲۳ سؤال آزمون، ۵ سؤال دارای کارکرد افتراقی است. همه سؤال‌های دارای کارکرد افتراقی از نوع کارکرد افتراقی یکنواخت بوده و به نفع داوطلبان پسر عمل می‌کنند. همچنین ۴ سؤال از این ۵ سؤال دارای اندازه DIF با شدت کم و ۱ سؤال دارای اندازه DIF با شدت زیاد شناسایی شدند.

نتیجه‌گیری: نتایج پژوهش حاضر می‌تواند اطلاعات سودمندی برای طراحان سؤال و سازندگان آزمون سراسری از نظر تأثیر جنسیت بر عملکرد داوطلبان به‌دست دهد تا با شناسایی، اصلاح و یا حذف این سؤال‌های آزمون از منصفانه بودن و روایی آزمون نسبت به کاربرد آن اطمینان حاصل شود.

کلید واژه‌ها: کارکرد افتراقی سؤال، مدلیابی تشخیصی شناختی، مدل غیرجبرانی دینا، کنکور سراسری، ریاضی.

۱. دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه تهران

۲. استادیار، گروه روش‌ها و برنامه‌های آموزشی و درسی، دانشگاه تهران

۳. دانشیار، گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی

۴. استادیار، گروه مطالعات برنامه درسی و تحقیقات آموزشی، دانشگاه خوارزمی

۱. مقدمه

آزمون سراسری ورودی دانشگاه‌ها شناخته‌شده‌ترین آزمون سرنوشت‌ساز کشور، مهم‌ترین و تعیین‌کننده‌ترین عامل ورود به دانشگاه‌هاست و منصفانه بودن سؤال‌های این آزمون همواره یکی از دغدغه‌های اصلی سازندگان و شرکت‌کنندگان آن به‌عنوان مهم‌ترین شاخص ارزیابی دوره متوسطه تحصیلی بوده است.

منصفانه بودن آزمون به این معناست که آزمون برای تمام افراد شرکت‌کننده در آن با هر جنسیت، نژاد، سن و طبقه اجتماعی که باشند، فاقد سوگیری بوده و آزمودنی‌ها در آن از شانس موفقیت یکسانی برخوردار باشند. به‌عبارت‌دیگر یک آزمون منصفانه آزمونی است که برای همه افراد شرکت‌کننده در آن روا بوده و فرصت یکسانی را برای نشان دادن دانش و توانایی مورداندازه‌گیری آزمون فراهم آورد (پرون^۱، ۲۰۰۶). پس تأمین فرصت برابر و فراهم کردن شرایطی برای تمام آزمودنی‌ها که بتوانند دانش و سطح توانایی خود را به نمایش بگذارند برای ارائه سنجش منصفانه ضرورتی انکارناپذیر است. در نتیجه سازندگان آزمون باید نشان دهند که آزمون‌هایشان علیه گروه خاصی سوگیری ندارد (همبلتون، سوامیناتان و راجرز^۲، ۱۹۹۱). سوگیری سؤال یکی از مهم‌ترین تهدیدها برای روایی آزمون است که در بستر سنجش از آن تحت عنوان کارکرد افتراقی سؤال یاد می‌شود (میلسپ^۳، ۲۰۱۱). در واقع اصطلاح کارکرد افتراقی سؤال^۴ یا DIF به‌عنوان گواهی تجربی برای وجود یا فقدان سوگیری در سؤال مورد استفاده قرار می‌گیرد (گرامی‌پور، ۱۳۹۰). کارکرد افتراقی سؤال زمانی اتفاق می‌افتد که احتمال پاسخ درست به سؤال در آزمودنی‌هایی با سطح توانایی یکسان ولی متعلق به گروه‌های مختلف، متفاوت باشد (سوامیناتان و راجرز^۵، ۱۹۹۰). می‌توان گفت که وجود کارکرد افتراقی سؤال گواهی علیه منصفانه بودن آزمون است و طی چند دهه گذشته روش‌های آماری مختلفی در این حوزه معرفی شده‌اند. همچنین طی سالیان اخیر تغییر در اهداف سنجش باعث شده تا مدل‌های آماری جدیدی در حوزه سنجش مطرح شوند که با اهداف آموزش و پرورش به‌طور کلی و با اهداف سنجش آموزشی به‌طور خاص هماهنگ و همخوان باشند (مینائی، ۱۳۹۱). سنجش تشخیصی شناختی^۶ یکی از رویکردهای جدید در زمینه اندازه‌گیری آموزشی است که می‌تواند اطلاعات دقیق‌تری از میزان تسلط افراد بر مهارت‌های موردنیاز برای پاسخگویی به سؤال‌ها را فراهم آورد. این نوع سنجش با تلفیق نظریه‌های یادگیری روان‌شناسی شناختی و مدل‌های آماری پیشرفته به اطلاعات ریزتر و مفصلی در مورد نقاط ضعف و قوت آزمون‌شوندگان دست پیدا می‌کند (لایتون و گیرل^۷، ۲۰۰۷؛ راب، تمپلین و هنسن^۸، ۲۰۱۰).

1. Perrone
2. Hambleton, Swaminathan & Rogers
3. Millsap
4. Differential Item Functioning
5. Swaminathan & Rogers
6. Cognitive Diagnostic Assessment
7. Leighton & Gierl
8. Rupp, Templin & Henson

مدل‌های تشخیصی شناختی به‌عنوان مدل‌های متغیرمکنون گسسته، یک نیمرخ مهارت برای هر فرد ارائه می‌دهند و امکان طبقه‌بندی آزمودنی‌ها به دو گروه مسلط یا غیرمسلط در مهارت‌هایی که هر سؤال اندازه می‌گیرد را فراهم می‌کنند. کارکرد افتراقی سؤال در چارچوب سنجش تشخیصی شناختی، تأثیری است که در آن آزمودنی‌هایی با نیمرخ تسلط مهارت یکسان اما متعلق به گروه‌های مختلف، احتمال‌های متفاوت در دادن پاسخ صحیح به یک سؤال داشته باشند. به‌عبارت‌دیگر پاسخ‌های سؤال وقتی فقط روی نیمرخ مهارت شرطی شده باشند، مستقل نیستند (هو، دلاتوره و نانداکومار^۱، ۲۰۱۴). در سنجش تشخیصی شناختی زمانی که سؤال‌ها دارای کارکرد افتراقی باشند برآورد پارامترهای سؤال نامعتبر و برآورد نیمرخ مهارت فرد برای گروه‌های مورد مطالعه مشکل‌دار خواهد بود (هو و همکاران، ۲۰۱۴). در نتیجه نیمرخ‌های مهارت افراد از گروه‌های مختلف غیرقابل مقایسه شده و سنجش منصفانه و روایی^۲ آزمون مورد تهدید قرار می‌گیرد. ویژگی‌های گروهی مانند جنسیت، سن، قومیت و دین معمولاً به‌عنوان متغیرهایی که ممکن است به وجود کارکرد افتراقی بیانجامد، شناخته می‌شوند (لیو^۳ و همکاران، ۲۰۱۹). در چارچوب مدلیابی تشخیصی شناختی روش‌های مختلفی پیشنهاد شده است (ژانگ^۴، لی^۵، ۲۰۰۶؛ هو و همکاران، ۲۰۱۴؛ وانگ، گو و بیان^۶، ۲۰۱۴؛ لی و وانگ^۷، ۲۰۱۵). همچنین درصد قابل توجهی از مطالعاتی که در چارچوب مدلیابی تشخیصی شناختی انجام شده از مدل دینا^۸ برای تحلیل داده‌ها استفاده کرده‌اند که می‌توان به پژوهش‌های جانکر و سیجسما^۹، ۲۰۰۱؛ دلاتوره، ۲۰۰۴؛ ژانگ، ۲۰۰۶؛ دلاتوره، ۲۰۰۸؛ لی، ۲۰۰۸؛ دلاتوره، ۲۰۱۰، ۲۰۱۱؛ ون داویر^{۱۰}، ۲۰۱۳؛ هو و همکاران، ۲۰۱۴؛ لی و وانگ، ۲۰۱۵ اشاره کرد. یکی از روش‌های پیشنهادی برای تشخیص کارکرد افتراقی سؤال در مدلیابی تشخیصی شناختی آزمون والد^{۱۱} است که تحت مدل دینا در پژوهش هو و همکاران (۲۰۱۴) کارایی آن مورد آزمون واقع شده و از روش‌های دیگر عملکرد بهتری دارد. در مدل دینا که یک مدل غیرجبرانی است فرض می‌شود که آزمودنی باید تمام مهارت‌های احتمالی مورد نیاز را داشته باشد تا بتواند به یک سؤال پاسخ صحیح دهد. در نتیجه پاسخ آزمودنی‌ها برای هر سؤال در دو طبقه مکنون نمره‌گذاری می‌شوند؛ یکی طبقه‌ای که نشان‌دهنده پاسخ صحیح است و عدد یک به آن تعلق می‌گیرد و شامل آزمودنی‌هایی می‌شود که تمام مهارت‌های مورد نیاز برای پاسخ صحیح به سؤال را دارا هستند و طبقه دیگر که نشان‌دهنده پاسخ نادرست است و با عدد صفر

1. Hou, de la Torre & Nandakumar
2. Validity
3. Liu, Yin, Xin, Shao & Yuan
4. Zhang
5. Li
6. Wang, Guo & Bian
7. Li, & Wang
8. DINA
9. Junker & Sijtsma
10. von Davier
11. Wald Test

نمایش داده می‌شود و شامل آزمودنی‌هایی می‌شود که حداقل یکی از مهارت‌های موردنیاز برای پاسخ‌دهی صحیح به سؤال را ندارند (دلاتوره، ۲۰۱۱). در کشور ما نیز پژوهش مینائی (۱۳۹۱) به‌مثابه قدم اول در مطالعه مدل‌های تشخیصی شناختی است. بعد از آن پژوهش‌های مختلفی تحت این مدل‌ها در کشور انجام شده است (کبیری، ۱۳۹۲؛ افضلی و همکاران، ۱۳۹۳؛ محسن‌پور و همکاران، ۱۳۹۴؛ مقدم و همکاران، ۱۳۹۵). در زمینه کارکرد افتراقی سؤال با رویکردهای کلاسیک روان‌سنجی یعنی نظریه کلاسیک اندازه‌گیری و نظریه سؤال پاسخ (IRT)^۱ نیز مطالعات متعددی صورت گرفته است که از آن جمله می‌توان به پژوهش‌های رضایی و شعبانی (۱۳۸۸)، گرامی‌پور (۱۳۹۰، ۱۳۹۳)، آسیابی (۱۳۹۱)، گرامی‌پور و همکاران (۱۳۹۵، ۱۳۹۶)، استاجی (۱۳۹۷)، براتی و احمدی (۲۰۱۰)، کرمی (۲۰۱۱) و امیریان، علوی و فیدالگو (۲۰۱۴) اشاره کرد. برای مثال مطالعه براتی و احمدی (۲۰۱۰) به بررسی کارکرد افتراقی سؤال‌های آزمون زبان تخصصی انگلیسی کنکور سراسری تحت نظریه سؤال پاسخ پرداخته است. نتایج مطالعه آن‌ها نشان داد که حدود ۴۳٪ از سؤال‌ها دارای کارکرد افتراقی بوده و به‌طور کلی سؤال‌های مربوط به بخش گرامر، کارکرد زبان و درک مفهوم به نفع زنان و سؤال‌های مربوط به معنی لغات و ترتیب کلمات به نفع مردان عمل می‌کند. فلاحی سرشت (۱۳۹۴) نیز کارکرد افتراقی سؤال‌های استعداد تحصیلی آزمون دکتری سال ۱۳۹۳ در دو گروه علوم انسانی و فنی مهندسی را با کاربرد نظریه IRT و روش رگرسیون لوجستیک مورد مطالعه قرار داده است. نتایج پژوهش وی نشان داد که در علوم انسانی با روش رگرسیون لوجستیک ۱۳ سؤال دارای کارکرد افتراقی است که ۷ سؤال به نفع زنان و ۶ سؤال به نفع مردان دارای کارکرد افتراقی است و در گروه فنی مهندسی نیز ۴ سؤال دارای کارکرد افتراقی است که ۱ سؤال به نفع زنان و ۳ سؤال به نفع مردان عمل می‌کند. سؤال‌هایی که به نفع مردان عمل می‌کردند مربوط به بخش تحلیلی و منطقی و برای زنان مربوط به تجسم فضایی بوده‌اند. گرامی‌پور و همکاران (۱۳۹۶) نیز به مطالعه کارکرد افتراقی سؤال و آزمون دروس تخصصی تمام گروه‌های آزمایشی کنکور سراسری ۱۳۸۷ تا ۱۳۹۰ با استفاده از رویکرد رگرسیون لوجستیک پرداختند. نتایج نشان می‌دهد که به‌طور متوسط حدود ۱۴٪ از سؤال‌های کنکور سال‌های ۱۳۸۷ تا ۱۳۹۰ نسبت به جنسیت دارای کارکرد افتراقی با شدت کم هستند. همچنین در این مطالعه نشان داده شد که سؤال‌های درس ریاضی گروه آزمایشی ریاضی و فیزیک به‌صورت مشهودی به ضرر داوطلبان دختر دارای کارکرد افتراقی است. بررسی کارکرد افتراقی سؤال با رویکرد تشخیصی شناختی به دلیل جدیدتر بودن این رویکرد در حوزه اندازه‌گیری آموزشی کمتر مورد بررسی قرار گرفته است. برای مثال مطالعه کاظمی‌دانا (۱۳۹۵) به بررسی کارکرد افتراقی سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۰۷ در چارچوب مدل‌یابی تشخیصی شناختی پرداخته است. در این پژوهش نیز به‌منظور تحلیل داده‌ها از مدل غیرجبرانی دینا استفاده شد و نتایج نشان داد که از ۴۴ سؤال مورد مطالعه، ۲۶ سؤال لنگر و ۱۸ سؤال (۴۰٪) دارای کارکرد افتراقی بود که از میان ۱۸ سؤال دارای

کارکرد افتراقی، ۱۲ سؤال آن دارای DIF یکنواخت و ۶ سؤال دیگر دارای DIF غیریکنواخت است. کاربردهای اولیه مدل‌های تشخیصی شناختی بیشتر برای داده‌های ریاضی بوده، چرا که در ریاضیات بیشتر امکان تقسیم راه‌حل‌ها به مراحل مختلف و اجزای خردتر وجود دارد. اگر فرد آزمون‌شونده بتواند تمامی مراحل را با موفقیت انجام دهد، می‌تواند به پاسخ صحیح دست یابد. در نتیجه به‌طور کلی این‌طور به نظر می‌رسد که در استفاده از مدل‌های غیرجبرانی برای درس ریاضی توافق وجود دارد (روسوس، تمپلین و هنس، ۲۰۰۷). خاطرنشان می‌شود تا جایی که پژوهشگر اطلاع دارد به نظر می‌رسد که تا به حال در زمینه کارکرد افتراقی سؤال‌های ریاضی آزمون سراسری ورودی دانشگاه‌های کشور تحت مدل‌های تشخیصی شناختی، پژوهشی صورت نپذیرفته است. از آنجایی که تحلیل کارکرد افتراقی برای رواسازی آزمون‌های سرنوشت‌ساز اهمیت ویژه‌ای دارد و امکان وجود سؤال‌های دارای کارکرد افتراقی در چنین آزمونی وجود دارد، در نتیجه می‌توان گفت بررسی کارکرد افتراقی این آزمون تحت چنین رویکردی لازم و ضروری به نظر می‌رسد. با توجه به نقش سرنوشت‌ساز آزمون سراسری ورودی دانشگاه در زندگی داوطلبان، نتایج این تحقیق می‌تواند برای متخصصان و سازندگان آزمون بسیار سودمند بوده و راهگشای تصمیم‌گیرندگان و دست‌اندرکاران ذی‌صلاح قرار گیرد. از این‌رو، هدف پژوهش حاضر بررسی کارکرد افتراقی سؤال‌های حساب‌دیفرانسیل و انتگرال آزمون سراسری ۱۳۹۷ در میان داوطلبان دختر و پسر با به کار بردن آزمون والد تحت مدل تشخیصی شناختی دینا است. لذا با توجه به اهداف پژوهش، سؤال‌های تحقیق به شرح زیر مطرح می‌شوند:

- ۱- آیا در چارچوب سنجش تشخیصی شناختی و با استفاده از آزمون والد، سؤال‌های حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ در میان داوطلبان دختر و پسر کارکرد افتراقی دارند؟
- ۲- شدت DIF سؤال‌های دارای کارکرد افتراقی درس حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ چه مقدار است؟
- ۳- نوع DIF سؤال‌هایی که در درس حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ دارای کارکرد افتراقی هستند، به چه صورت است؟

۲. روش پژوهش

از آنجاکه پژوهش حاضر با داده‌های آزمون سراسری ورودی دانشگاه‌های کشور انجام شده است؛ جزو پژوهش‌های توصیفی از نوع تحلیل ثانویه در چارچوب مدلیابی تشخیصی شناختی به‌شمار می‌آید. جامعه آماری پژوهش کلیه داوطلبان شرکت‌کننده در آزمون سراسری ورودی دانشگاه‌های کشور گروه آزمایشی ریاضی و فیزیک سال ۱۳۹۷ است. طبق آمار منتشر شده توسط سازمان سنجش کشور تعداد ۱۴۴۴۳۷ نفر در این آزمون شرکت کرده‌اند (دفتر طرح و آمار سازمان سنجش، ۱۳۹۷). از آنجاکه اکثر

مدل‌های تشخیصی شناختی به حجم نمونه حساس بوده و نیازمند نمونه‌های بزرگ هستند (لی، ۲۰۱۱)، لذا نمونه پژوهش شامل ۵۰۰۰ نفر از داوطلبان شرکت‌کننده در کنکور گروه آزمایشی ریاضی و فیزیک می‌شود که در تیرماه سال ۱۳۹۷ در آزمون سراسری شرکت کرده‌اند. از آنجایی که نمونه موردنظر با مراجعه پژوهشگر به سازمان سنجش و آموزش کشور تهیه شده و پژوهشگر هیچ دخالتی در انتخاب افراد نداشته و سازمان سنجش نیز به شکل تصادفی نمونه درخواستی پژوهشگر را از میان کلیه داده‌های موجود انتخاب کرده است؛ لذا روش نمونه‌گیری پژوهش را می‌توان جزو نمونه‌گیری تصادفی ساده قلمداد کرد. از این ۵۰۰۰ نفر افراد گروه نمونه، ۱۸۰۲ نفر دختر (۳۶٪) و ۳۱۹۶ نفر پسر (۶۴٪) بودند. دفترچه سؤال‌های اختصاصی گروه آزمایشی ریاضی و فیزیک شامل ۵۵ سؤال ریاضی است که ۲۴ سؤال آن به درس حساب دیفرانسیل و انتگرال اختصاص دارد و شامل سؤال‌های ۱۰۱ تا ۱۲۴ می‌شود که در سال ۱۳۹۷ سؤال شماره ۲۰ به دلیل اشتباهی که در طرح آن رخ داده از مجموع سؤال‌ها حذف شده است. در نتیجه در مجموع ۲۳ سؤال این درس مورد مطالعه قرار گرفت. روش نمره‌گذاری آزمون به صورت دوازده‌گانه است. چنانچه داوطلبی به یک سؤال پاسخ صحیح داده باشد عدد یک و چنانچه به سؤال پاسخ نادرست داده باشد یا به آن سؤال پاسخ نداده باشد عدد صفر به وی تعلق می‌گیرد. به منظور تحلیل آماری سؤال‌ها از نرم‌افزار رایگان R استفاده شد. در این نرم‌افزار برای بررسی برازش داده‌ها با مدل دینا، از بسته^۱ CDM و به منظور بررسی کارکرد افتراقی سؤال با آزمون والد، از بسته^۲ GDINA استفاده شد (ما و دلاتوره^۳، ۲۰۱۷). برای بررسی برازش مدل با داده‌ها از شاخص‌های نسبی^۴ AIC، BIC^۵ و شاخص‌های مطلق^۶ RMSEA، MADcor^۷، MADRESIDCov^۸، MADQ3^۹، SRMSR^{۱۰} و مجذورکای^{۱۱} استفاده شده است. به جز شاخص مجذورکای، برای شاخص‌های دیگر هرچه میزان شاخص کوچک‌تر و به صفر نزدیک‌تر باشد، داده‌ها برازش بهتری با مدل دارند (راوند و روبیج^{۱۱}، ۲۰۱۵). از آنجایی که در مورد این شاخص‌ها توافق وجود ندارد، بهتر است بررسی برازش مدل به هر دو روش صورت پذیرد (میدو^{۱۲}، ۲۰۱۳ به نقل از کاظمی‌دانا، ۱۳۹۵). همچنین از شاخص^{۱۳} UA برای شناسایی میزان بزرگی و کوچکی DIF استفاده شده است. برای بررسی نوع DIF نیز ابتدا باید آن را تعریف کرد کارکرد افتراقی یکنواخت زمانی به وجود می‌آید که

1. Cognitive Diagnostic Modeling
2. Generalized DINA Model Framework
3. Ma & de la Torre
4. Akaike Information Criterion
5. Bayesian Information Criterion
6. Root Mean Square Error of Approximation
7. Mean Absolute Deviation Correlation
8. Mean of Absolute Deviations of Residual Covariance
9. Mean of Absolute Difference of Q3
10. Standardized Root Mean Square Residual
11. Ravand & Robitzsch
12. Maydeu
13. Unsigned Area

احتمال‌های پاسخ صحیح به سؤال صرف‌نظر از نیمرخ تسلط مهارت همواره برای یک گروه بالاتر یا پایین‌تر باشد و کارکرد افتراقی غیریکنواخت نیز زمانی به وجود می‌آید که احتمال‌های پاسخ صحیح به سؤال برای یک گروه در بعضی از نیمرخ‌های تسلط مکنون بالاتر و در بعضی از نیمرخ‌ها پایین‌تر باشد (هو و همکاران، ۲۰۱۴).

برای شناسایی مهارت‌های احتمالی موردنیاز سؤال‌های آزمون و تدوین ماتریس کیو از روش تحلیل محتوا و پروتکل تفکر بلند و مصاحبه‌های کانونی با حضور متخصصان ریاضی در سازمان سنجش و آموزش کشور استفاده شد. به همین منظور دو متخصص دارای مدرک دکترای ریاضی و با حداقل ده سال سابقه آزمون‌سازی برای تحلیل محتوا و سه متخصص با مدرک کارشناسی‌ارشد ریاضی برای بخش پروتکل تفکر بلند انتخاب شدند. بر مبنای جلسات اولیه گروه‌های کانونی ابتدا فهرست اولیه مهارت‌های شناختی لازم برای پاسخگویی به سؤال‌های حساب‌دیفرانسیل و انتگرال با کمک متخصصین تعیین و نام‌گذاری شدند، سپس این فهرست برای اساتید روان‌شناسی و همچنین مدرسان ریاضی در آموزشگاه‌های کنکور ارسال شد و از آن‌ها خواسته شد تا در مورد این مهارت‌ها نظراتی ارائه دهند. با جمع‌بندی این نظرات، فرم نهایی مهارت‌ها تدوین و در جلسات گروه کانونی نهایی شد. سپس لیست نهایی به افراد با مدرک کارشناسی‌ارشد ارائه و از آن‌ها درخواست شد تا به سؤال‌های حساب‌دیفرانسیل و انتگرال کنکور سال ۱۳۹۷ پاسخ دهند و مهارت‌های موردنیاز برای پاسخگویی به هر سؤال را تعیین کنند. در جلسات بعدی با حضور متخصصین دکترا اختلاف‌های به‌وجود آمده به بحث گذاشته شد و مهارت‌های انتخاب‌شده برای هر سؤال به تفصیل مورد بررسی و توافق نهایی قرار گرفت. مهارت‌های شناسایی شده برای پاسخگویی به سؤال‌های حساب‌دیفرانسیل و انتگرال در جدول ۱ نشان داده شده است.

جدول ۱: مهارت‌های زیربنایی شناسایی شده برای پاسخ به سؤال‌های حساب‌دیفرانسیل و انتگرال

۱- تعاریف و دانش مفاهیم و قواعد
۲- مدل‌سازی تنه سؤال
۳- کاربرد اعمال جبری مقدماتی سریع
۴- کاربرد تکنیک‌های متداول
۵- کاربرد اعمال جبری پیشرفته
۶- کاربرد قضایا
۷- استفاده از اشکال و نمودارها
۸- تجزیه مفاهیم
۹- تحلیل مفاهیم
۱۰- ترکیب مفاهیم

۳. یافته‌های پژوهش

همان‌طور که در بخش قبل اشاره شد، عدم معناداری شاخص مجذور کای، نشان‌دهنده برازش مناسب مدل با داده‌هاست؛ اما سایر شاخص‌ها هرچه به صفر نزدیک‌تر باشند؛ مدل برازش بهتری دارد و به‌طور معمول مقادیر کمتر از ۰/۰۵ بهترین برازش با مدل را نشان می‌دهند (دی بلو، روسوس و استات، ۲۰۰۷ به نقل از راوند و رویچ، ۲۰۱۵). نتایج به‌دست‌آمده در جدول ۲ نشان می‌دهد که اکثر شاخص‌ها به‌جز، MADRESIDCOV نشان از برازش خوب داده‌ها با مدل دارند. البته چون داده‌ها مربوط به ۵۰۰۰ آزمون‌شونده است و شاخص مجذور کای نیز به حجم نمونه حساس است، این شاخص معنادار شده است.

جدول ۲: شاخص‌های برازش مدل با داده‌ها

مقدار	آماره
۱۰۵۳۷۶	AIC
۱۱۲۳۴۳	BIC
۰/۰۴۶	RMSEA
۰/۰۵۰	SRMSR
۱۲۲/۹۶ (p=۰/۰۰۱)	χ^2
۰/۰۴۷	MADcor
۰/۰۳۲	MADQ3
۰/۷۱۳	MADRESIDCOV

پس از بررسی برازش مدل، به برآورد پارامترهای مدل پرداخته شد. در مدل دینا برای هر سؤال دو پارامتر محاسبه می‌شود که پارامتر لغزش و حدس نام دارند. به‌طور کلی هر چه مقدار این پارامترها کمتر باشند نشان‌دهنده برازش بهتر میان طرح سنجش تشخیصی، داده‌ها و مدل شناختی دینا است (راوند، براتی و ویدھیارسو، ۲۰۱۲). با توجه به جدول ۳ می‌توان دریافت که میانگین پارامتر لغزش و حدس در داوطلبان دختر به ترتیب ۰/۵۱۴ و ۰/۰۹۷ است. متوسط مقادیر لغزش و حدس در پسران نیز به ترتیب ۰/۴۳۸ و ۰/۱۲۳ است. این مقدار بیانگر آن است که از داوطلبان دختری که در همه مهارت‌های موردنیاز برای دادن پاسخ صحیح به سؤال به تسلط رسیده‌اند، به‌طور متوسط ۵۱٪ شانس پاسخ غلط به سؤال را دارند. این احتمال برای پسران ۴۳٪ است. همچنین از داوطلبان دختری که حداقل در یک مهارت به حد تسلط نرسیده‌اند به‌طور متوسط ۹/۷٪ شانس دادن پاسخ صحیح به سؤال برایشان وجود دارد. برای داوطلبان پسری که حداقل در یک مهارت به تسلط نرسیده‌اند این شانس به‌طور متوسط ۱۲٪ است. همچنین مشاهده می‌شود میزان پارامتر حدس در دختران کمتر از پسران است. در سؤال‌های ۲، ۱۰، ۱۲، ۲۰ و ۲۱ مقدار پارامتر حدس در دختران بسیار کم است. درحالی‌که

1. DiBello, Rouso & Stout
2. Ravand, Barati & Widhiarso

این مقدار برای پسران تنها در سؤال‌های ۲، ۱۲ و ۲۰ بسیار ناچیز است. پارامتر لغزش نیز در اکثر سؤال‌ها به خصوص سؤال‌های ۲، ۱۶ و ۲۰ زیاد است. کمتر بودن احتمال حدس به دلیل داشتن نمره منفی در آزمون سراسری قابل توجیه است. در مقابل احتمال لغزش زیاد برای افرادی که علی‌رغم رسیدن به تسلط به سؤال پاسخ اشتباه داده‌اند نیز می‌تواند حاکی از دشوار بودن آزمون باشد.

جدول ۳: برآوردهای پارامتر حدس و لغزش سؤال‌ها به تفکیک جنسیت

سؤال	دختران		پسران	
	پارامتر لغزش	پارامتر حدس	پارامتر لغزش	پارامتر حدس
۱	۰/۴۴۳	۰/۰۷۴	۰/۴۶۱	۰/۱۳۶
۲	۰/۹۷۵	۰/۰۰۰	۰/۹۰۲	۰/۰۰۳
۳	۰/۱۰۱	۰/۳۹۲	۰/۰۷۵	۰/۴۶۶
۴	۰/۳۲۱	۰/۰۹۴	۰/۱۵۶	۰/۱۵۸
۵	۰/۳۶۹	۰/۰۴۳	۰/۳۲۵	۰/۰۷۵
۶	۰/۴۸۰	۰/۰۵۱	۰/۲۴۴	۰/۱۲۸
۷	۰/۶۳۹	۰/۰۸۶	۰/۳۶۴	۰/۱۶۱
۸	۰/۸۱۲	۰/۰۲۷	۰/۶۷۴	۰/۰۵۶
۹	۰/۴۴۳	۰/۱۳۲	۰/۳۶۴	۰/۱۰۰
۱۰	۰/۷۵۲	۰/۰۰۹	۰/۷۳۶	۰/۰۲۱
۱۱	۰/۰۵۷	۰/۱۸۵	۰/۱۰۷	۰/۲۱۲
۱۲	۰/۷۷۹	۰/۰۰۲	۰/۶۱۵	۰/۰۰۰
۱۳	۰/۷۶۷	۰/۰۴۲	۰/۶۹۱	۰/۰۶۸
۱۴	۰/۳۷۰	۰/۲۲۶	۰/۲۳۱	۰/۲۸۲
۱۵	۰/۰۸۶	۰/۳۰۹	۰/۰۳۲	۰/۳۴۸
۱۶	۰/۹۲۳	۰/۰۱۴	۰/۸۷۸	۰/۰۲۹
۱۷	۰/۶۳۹	۰/۰۵۴	۰/۴۵۰	۰/۰۷۶
۱۸	۰/۳۴۳	۰/۱۳۲	۰/۳۴۲	۰/۱۶۱
۱۹	۰/۴۶۸	۰/۱۱۴	۰/۴۶۶	۰/۱۱۶
۲۰	۰/۸۴۴	۰/۰۰۳	۰/۷۸۷	۰/۰۰۱
۲۱	۰/۷۵۱	۰/۰۰۷	۰/۶۴۶	۰/۰۲۲
۲۲	۰/۱۷۳	۰/۱۶۱	۰/۱۴۸	۰/۱۳۷
۲۳	۰/۳۸۲	۰/۰۷۰	۰/۳۸۱	۰/۰۷۷
میانگین	۰/۵۱۴	۰/۰۹۷	۰/۴۳۸	۰/۱۲۳

پرسش اول: آیا در چارچوب سنجش تشخیصی شناختی و با استفاده از آزمون والد سؤال‌های حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ در میان داوطلبان دختر و پسر کارکرد افتراقی دارند؟

پس از بررسی برازش داده‌ها با مدل دینا، برای شناسایی سؤال‌های دارای کارکرد افتراقی از آزمون والد استفاده گردید. آماره والد دارای توزیع مجذورکای با درجه آزادی ۲ است و به بررسی معناداری

تفاوت میان پارامترهای حدس و لغزش در دو گروه مرجع و کانونی می‌پردازد (هو و همکاران، ۲۰۱۴). فرض صفر این آزمون به این معناست که پارامتر حدس و پارامتر لغزش در دو گروه مرجع و کانونی با هم برابر هستند یا به عبارت دیگر میان دو پارامتر سؤال در دو گروه تفاوت معناداری وجود ندارد. فرض خلاف نیز به این صورت است که حداقل یکی از پارامترهای سؤال میان دو گروه متفاوت باشد. در صورتی که اختلاف میان پارامترهای سؤال در دو گروه معنادار باشد ($p < 0.05$)، سؤال دارای کارکرد افتراقی تشخیص داده می‌شود. در جدول ۴ نتایج آشکارسازی سؤال‌های دارای DIF با استفاده از روش والد ارائه شده است.

جدول ۴: شاخص‌های مربوط به کارکرد افتراقی سؤال‌ها

سؤال	آماره والد	سطح اطمینان تصحیح شده هولم	درجه آزادی	اندازه اثر UA	کارکرد افتراقی
۱	۵/۸۱	۰/۹۲۹	۲	۰/۰۲۶	-
۲	۱۸/۴۰	۰/۰۰۲	۲	۰/۰۰۸	دارد
۳	۰/۵۶	۱/۰۰۰	۲	۰/۰۲۶	-
۴	۳/۴۹	۱/۰۰۰	۲	۰/۰۲۸	-
۵	۱/۳۰	۱/۰۰۰	۲	۰/۰۱۳	-
۶	۱۱/۹۸	۰/۰۴۷	۲	۰/۰۴۷	دارد
۷	۵۰/۸۸	۰/۰۰۰	۲	۰/۱۴۵	دارد
۸	۱۶/۵۲	۰/۰۰۵	۲	۰/۰۴۱	دارد
۹	۵/۵۴	۰/۹۹۹	۲	۰/۰۲۸	-
۱۰	۰/۵۸	۱/۰۰۰	۲	۰/۰۰۵	-
۱۱	-/۲۳	۱/۰۰۰	۲	۰/۰۰۷	-
۱۲	۳/۴۷	۱/۰۰۰	۲	۰/۰۱۵	-
۱۳	۴/۹۸	۱/۰۰۰	۲	۰/۰۴۲	-
۱۴	۱/۳۴	۱/۰۰۰	۲	۰/۰۱۶	-
۱۵	۱/۰۴	۱/۰۰۰	۲	۰/۰۱۷	-
۱۶	۱۳/۶۰	۰/۰۲۲	۲	۰/۰۱۷	دارد
۱۷	۳/۶۴	۱/۰۰۰	۲	۰/۰۷۸	-
۱۸	۸/۴۷	۰/۲۶۰	۲	۰/۰۹۴	-
۱۹	۱/۵۶	۱/۰۰۰	۲	۰/۰۱۷	-
۲۰	۱/۴۸	۱/۰۰۰	۲	۰/۰۰۴	-
۲۱	۰/۴۸	۱/۰۰۰	۲	۰/۰۰۵	-
۲۲	۱/۵۲	۱/۰۰۰	۲	۰/۰۱۴	-
۲۳	۰/۰۰۶	۱/۰۰۰	۲	۰/۰۰۰	-

یافته‌های جدول ۴ نشان می‌دهد که از ۲۳ سؤال مورد مطالعه در بخش حساب‌دیفرانسیل و انتگرال آزمون سراسری ۱۳۹۷، ۵ سؤال ۲، ۶، ۷، ۸ و ۱۶ یعنی ۲۲٪ سؤال‌ها نسبت به جنسیت دارای کارکرد افتراقی بودند.

پرسش دوم: شدت DIF سؤال‌های دارای کارکرد افتراقی درس حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ چه مقدار است؟

برای پاسخگویی به این پرسش، از شاخص UA برای شناسایی میزان بزرگی و کوچکی اندازه DIF استفاده شد. راوند و روبیچ (۲۰۱۵) برای این شاخص نقطه برش تعیین کرده‌اند به گونه‌ای که اندازه‌های کمتر از ۰/۰۵۹ به عنوان DIF کوچک، مقادیر میان ۰/۰۵۹ و ۰/۰۸۸ به عنوان DIF متوسط و اندازه‌های بیشتر از ۰/۰۸۸ دارای DIF بزرگ تفسیر می‌شوند. با توجه به میزان UA در جدول ۴، از میان ۵ سؤالی که دارای کارکرد افتراقی تشخیص داده شده‌اند، ۴ سؤال ۲، ۶، ۷ و ۸ دارای اندازه DIF کوچک و یک سؤال (سؤال ۷) دارای اندازه DIF بزرگ بودند.

پرسش سوم: نوع DIF سؤال‌هایی که در درس حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه سال ۱۳۹۷ دارای کارکرد افتراقی هستند، به چه صورت است؟

بررسی مقدار تفاوت میان پارامترهای حدس و لغزش دو گروه در جدول ۵ برای تعیین یکنواخت و غیریکنواخت بودن سؤال‌های دارای کارکرد افتراقی نیز نشان داد که همه سؤال‌های دارای کارکرد افتراقی از نوع یکنواخت بودند. به این صورت که در همه ۵ سؤال، پارامتر لغزش گروه کانونی (دختران) بزرگ‌تر و پارامتر حدس آن‌ها کوچک‌تر بود که نشان می‌دهد علیرغم نیمرخ تسلط صفت یکسان، احتمال دادن پاسخ صحیح به سؤال برای داوطلبان دختر نسبت به پسران کمتر است؛ یعنی همه ۵ سؤالی که دارای کارکرد افتراقی شناسایی شده‌اند، به نفع پسران عمل کرده‌اند.

جدول ۵: تفاوت پارامترهای حدس و لغزش در سؤال‌های دارای کارکرد افتراقی

نوع کارکرد افتراقی	تفاوت پارامترهای حدس <i>GF-GR</i>	تفاوت پارامترهای لغزش <i>SR-SF</i>	سؤال دارای DIF
یکنواخت	-۰/۰۰۳	-۰/۰۷۳	۲
یکنواخت	-۰/۰۷۷	-۰/۲۳۶	۶
یکنواخت	-۰/۰۷۵	-۰/۲۷۵	۷
یکنواخت	-۰/۰۲۹	-۰/۱۳۸	۸
یکنواخت	-۰/۰۱۵	-۰/۰۴۵	۱۶

بحث و نتیجه‌گیری

این ادعا که پسران در درس ریاضی نسبت به دختران عملکرد بهتری دارند توسط محققان مختلف همواره مورد مطالعه قرار گرفته است (لیو و ویلسون، ۲۰۰۹). عادلانه بودن آزمون برای همه کسانی

که آزمون بر روی آن‌ها اجرا می‌شود، یکی از شروط اصلی استفاده از آزمون‌ها در امر تصمیم‌گیری است. این امر بخصوص زمانی که یک آزمون سرنوشت‌ساز داشته باشیم اهمیت دوچندان پیدا می‌کند. لذا منصفانه بودن آزمون سراسری ورودی دانشگاه‌ها به‌عنوان مهم‌ترین آزمون سرنوشت‌ساز کشور همواره مورد توجه پژوهشگران می‌باشد. از آنجایی که وجود کارکرد افتراقی سؤال در این آزمون محتمل است، مطالعه حاضر به بررسی کارکرد افتراقی سؤال‌های حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه‌های سال ۱۳۹۷ تحت مدل‌یابی تشخیصی شناختی می‌پردازد. اهمیت تشخیص کارکرد افتراقی از این جهت است که از یک طرف وجود کارکرد افتراقی در سنجش تشخیصی شناختی می‌تواند پیامدهای جدی از جمله برآورد نادقیق نیمرخ تسلط سؤال و مهارت، در پی داشته باشد و از طرف دیگر شناسایی، اصلاح و یا حذف سؤال‌های دارای کارکرد افتراقی در آزمون باعث می‌شود که از عادلانه بودن و روایی آزمون نسبت به کاربرد آن اطمینان حاصل شود. همچنین مدل‌های تشخیصی شناختی نویدبخش روش‌شناسی‌ای است که می‌تواند با ارائه اطلاعات دقیق‌تر، در امر یادگیری و آموزش بهتر یاری‌رسان باشد. به دلیل وجود داده‌های مربوط به درس ریاضی، در این مطالعه از مدل غیرجبرانی دینا استفاده شده و به‌منظور بررسی برازش مدل با داده‌ها شاخص‌های برازش مطلق و نسبی به کار گرفته شده است. نتایج نشان می‌دهد که داده‌ها با مدل برازش مناسب دارند. همچنین داوطلبان دختر و پسر در متوسط پارامترهای لغزش و حدس شرایط تقریباً مشابهی داشتند که این یافته‌ها با پژوهش مینائی (۱۳۹۱)، افضلی (۱۳۹۳) و کاظمی‌دانا (۱۳۹۵) همسو است. برای پاسخگویی به پرسش اول پژوهش و تشخیص کارکرد افتراقی از آزمون والد استفاده شده است. نتایج نشان می‌دهد که ۵ سؤال از ۲۳ سؤال درس حساب‌دیفرانسیل و انتگرال آزمون سراسری سال ۱۳۹۷، نسبت به جنسیت دارای کارکرد افتراقی است. به بیان دیگر سؤال‌های درس حساب‌دیفرانسیل و انتگرال آزمون سراسری سال ۱۳۹۷ برای داوطلبان دختر و پسر به یک اندازه منصفانه نیست. برای پاسخگویی به پرسش دوم پژوهش و بررسی شدت DIF، از شاخص UA استفاده شده است که با توجه به نقطه برش تعیین‌شده از میان سؤال‌های دارای کارکرد افتراقی چهار سؤال دارای کارکرد افتراقی با شدت ضعیف (کمتر از ۰/۰۵۹) و یک سؤال دارای کارکرد افتراقی با شدت زیاد (بیشتر از ۰/۰۸۸) تفسیر شده‌اند که از حیث کوچک بودن اندازه با پژوهش‌های گذشته (گرامی‌پور، ۱۳۹۵) همسو است. همچنین با بررسی تفاوت پارامترهای لغزش و حدس در دو گروه دختران و پسران برای تعیین نوع DIF به پرسش سوم پژوهش پاسخ داده شده و نتایج حاکی از آن است که تمام سؤال‌ها دارای کارکرد افتراقی یکنواخت به نفع پسران می‌باشند. در واقع با توجه به پرسش اول پژوهش می‌توان گفت که ۲۲٪ سؤال‌ها به نفع داوطلبان پسران عمل می‌کند. البته گفتنی است که شناسایی کارکرد افتراقی سؤال اولین گام در مطالعه امکان‌سودار بودن سؤال است (ارسی‌کان، ۲۰۰۲)، در نتیجه صرفاً وجود کارکرد افتراقی نمی‌تواند نشان‌دهنده وجود سوگیری در آزمون باشد. گام بعدی بررسی منشأ کارکرد افتراقی است.

به عبارت دیگر زمانی که مشخص می‌شود سؤال‌های یک آزمون نسبت به یک گروه، مثلاً در پژوهش حاضر نسبت به پسران، دارای کارکرد افتراقی است بایست سازوکاری که این افتراق بر عملکرد افراد می‌گذارد، شناخته شود. مزیت استفاده از مدل‌یابی تشخیصی شناختی این است که می‌تواند علاوه بر بررسی مهارت ریاضی به‌عنوان یک سازه کلی، امکان بررسی تسلط افراد در مهارت‌های اندازه‌گیری شده در هر سؤال را فراهم آورد و اطلاعات تشخیصی مناسبی در این زمینه ارائه دهد. در واقع، با مطالعه و تحلیل فرایندی که در زمان پاسخ دادن به سؤال در ذهن فرد انجام می‌گیرد، می‌توان به ترکیبات پیچیده‌ای از مهارت‌ها، راهبردها و ساختارهای دانش دست پیدا کرد و توانایی و عملکرد افراد را برای دادن پاسخ صحیح به یک سؤال بررسی کرد (بجار^۱، ۲۰۰۸). در پژوهش حاضر اکثر سؤال‌هایی که دارای DIF هستند بخصوص سؤال ۷ که دارای DIF با شدت بزرگ است بیشتر به مهارت‌های دانش قواعد و مفاهیم، مدل‌سازی تنه سؤال و استفاده از اشکال و نمودارها مربوط می‌شود. چنانچه پیشتر گفته شد در پژوهش‌های گذشته (فلاحی‌سرشت، ۱۳۹۴) نیز سؤال‌هایی که برای پاسخگویی به آن‌ها نیاز به استفاده از راهبردهای متعددی است و سؤال‌هایی که مستلزم استفاده از نمودار و شکل‌ها و جداول است به نفع پسران عمل می‌کند و یافته‌های پژوهش حاضر از این حیث با پژوهش‌های قبلی همسو است.

سرانجام، از یافته‌های حاصل می‌توان نتیجه‌گیری کرد که جنسیت نقش مهمی در پاسخگویی به سؤال‌های آزمون سراسری ورودی دانشگاه‌ها ایفا می‌کند و کارکرد افتراقی سؤال‌ها می‌تواند به‌عنوان عاملی در قبولی یا عدم‌قبولی در آزمون‌های سرنوشت‌ساز نقش داشته باشد. از آنجایی که آزمون سراسری ورودی دانشگاه‌ها عرصه‌ای است که جوانان برای ورود به دانشگاه و رسیدن به رشته موردعلاقه خود با یکدیگر رقابت می‌کنند، لازم است چنین سؤال‌هایی شناسایی شوند. نتایج این پژوهش می‌تواند در مرحله اول برای معلمان ریاضی مدارس و مدرسان آموزشگاه‌ها سودمند باشد از این جهت که بتوانند بر نقاط ضعف دانش‌آموزان تمرکز کرده و تمرین‌ها و راهبردهایی به‌منظور تقویت کاستی‌های آن‌ها ارائه دهند. در مرحله بعد، نتایج این مطالعه می‌تواند برای سازندگان آزمون و طراحان سؤال‌های ریاضی کنکور، برای بودجه‌بندی سؤال‌ها و همچنین بازنگری سؤال‌های موجود در بانک سؤال که احتمال وجود کارکرد افتراقی در آن‌ها بیشتر است، قبل از اجرای آزمون کنکور سراسری مفید واقع شود. البته نبود پژوهش‌هایی که به بررسی کارکرد افتراقی سؤال‌های حساب‌دیفرانسیل و انتگرال آزمون سراسری ورودی دانشگاه‌ها با استفاده از رویکردهای دیگر پرداخته باشند، امکان مقایسه روش والد با سایر روش‌ها را برای پژوهشگر محدود می‌کند. همچنین به دلیل جدید بودن مدل‌های تشخیصی شناختی، کمبود مطالعات پژوهشی به‌خصوص در زمینه شناسایی کارکرد افتراقی سؤال تحت این مدل‌ها وجود دارد که این کمبود پژوهشی لزوم انجام پژوهش‌های بیشتر و جامع‌تر را در این رویکرد نمایان می‌سازد.

از جمله محدودیت‌های پژوهش نیز می‌توان به این موضوع اشاره کرد که یافته‌های حاصل از پژوهش تنها برای داده‌های دوارزشی کاربرد دارد و برای داده‌های چند ارزشی قابل استفاده نیست. همچنین یافته‌ها صرفاً قابل تعمیم به آزمون‌های شناختی است و در آزمون‌ها و موقعیت‌های غیرشناختی قابل به‌کارگیری نیست یا باید با احتیاط مورداستفاده قرار گیرد. به‌علاوه، دشواری آزمون سراسری از یک‌طرف و وجود نمره منفی از طرفی دیگر، باعث می‌شود که بسیاری از داوطلبان این آزمون به بسیاری از سؤال‌های ریاضی پاسخ ندهند و وجود سؤال‌های بی‌پاسخ در مدل‌های تشخیصی شناختی می‌تواند برآورد پارامترهای سؤال و همچنین روایی ماتریس کیو را مخدوش نماید. از آنجایی که علاوه بر جنسیت، ویژگی‌های دیگری همچون قومیت و زبان نیز می‌توانند بر کارکرد افتراقی سؤال تأثیر داشته باشند، لذا پیشنهاد می‌شود کارکرد افتراقی سؤال‌های آزمون با چنین متغیرهایی نیز بررسی شود. به‌علاوه، استفاده از روش‌های دیگر تشخیص کارکرد افتراقی سؤال تحت مدل‌یابی تشخیصی شناختی نیز می‌تواند موضوع مناسبی برای پژوهش‌های آتی باشد.

تقدیر و تشکر

نویسندگان بر خود لازم می‌دانند از مساعدت‌های سازمان سنجش و آموزش کشور که در قالب یک کارگروهی تحت هدایت آقای دکتر محمدباقر مهرلطیفان برای گردآوری داده‌ها و تعیین مهارت‌های زیربنایی سؤال‌ها به عمل آمد، صمیمانه تشکر و قدردانی نمایند.

منابع

- آسیایی، مینا. (۱۳۹۱). *ارزشیابی آزمون کارشناسی ارشد رشته جغرافیای سیاسی با استفاده از مدل‌های جدید اندازه‌گیری و تعیین سوگیری جنسیتی در آن*، پایان‌نامه کارشناسی ارشد. دانشگاه علامه طباطبایی.
- استاجی، معصومه و بابانژاد کفشگر، نگار. (۱۳۹۷). «کارکرد افتراقی جنسیتی سؤال‌های آزمون کنکور سراسری کارشناسی ارشد رشته‌ی آموزش زبان انگلیسی در ایران: مقایسه روش‌های آماری رگرسیون لجیستیک و منتل-هنسزل»، *فصلنامه اندازه‌گیری تربیتی*، ۸(۳۲)، ۴۱-۶۲.
- افضلی، افشین؛ دلاور، علی؛ فلسفی‌نژاد، محمدرضا؛ فرخی، نورعلی و برجعلی، احمد. (۱۳۹۳). «کاربرد مدل‌های تشخیصی شناختی در تعیین ماهیت تفاوت در عملکرد دانش‌آموزان دختر و پسر پایه اول دبیرستان در ریاضیات». *مجله دستاوردهای روانشناختی*، ۴(۲)، ۸۹-۱۰۴.
- رضایی، عباسعلی و شعبانی، عنایت‌اله. (۱۳۸۸). «تحلیل کارکرد افتراقی جنسیتی آزمون سنجش توانش عمومی زبان دانشگاه تهران». *پژوهش زبان‌های خارجی*، ۵۶، ۸۹-۱۰۸.
- فلاحی سرشت، شیوا. (۱۳۹۴). *بررسی کارکرد افتراقی سؤال‌های استعداد تحصیلی آزمون نیمه‌متمرکز دکتری سال ۹۳ با کاربرد نظریه سؤال پاسخ و رگرسیون لجیستیک*. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- کاظمی دانا، بهروز. (۱۳۹۵). *کارکرد افتراقی سؤال‌های آزمون ریاضی پایه هشتم تیمز ۲۰۰۷ در بین دانش‌آموزان دختر و پسر با استفاده از مدل‌های تشخیصی شناختی*، پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- کبیری، مسعود. (۱۳۹۲). *کاربرد سنجش تشخیصی شناختی به‌منظور تعیین مهارت‌های کسب‌شده علوم تجربی در دانش‌آموزان*، پایان‌نامه دکتری دانشگاه تهران.
- گرامی‌پور، مسعود. (۱۳۹۰). *مقایسه قدرت آزمون نسبت درستمایی مبتنی بر مدل سؤال-پاسخ با روش‌های تحلیل عاملی تأییدی و رگرسیون لجیستیک در شناسایی کارکرد افتراقی سؤال به‌منظور اطمینان از عادلانه بودن سنجش آزمون‌های سرنوشت‌ساز*، رساله دکتری دانشگاه علامه طباطبایی.
- گرامی‌پور، مسعود. (۱۳۹۳). «ارزیابی توان آماری تحلیل رگرسیون لجیستیک در آشکارسازی کارکرد افتراقی سؤال‌های آزمون». *فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی* ۴(۸)، ۱۸۷-۲۱۱.
- گرامی‌پور، مسعود؛ رضایی، احمد؛ رضانی‌صدر، اعظم؛ نوروزی، لیلا. (۱۳۹۵). «کنش افتراقی سؤال در آزمون‌های سازمان سنجش آموزش کشور برحسب ویژگی‌های جمعیت‌شناختی داوطلبان کنکور سراسری». *فصلنامه اندازه‌گیری تربیتی*، ۷(۲۶)، ۱۰۱-۱۲۲.
- گرامی‌پور، مسعود؛ رضانی‌صدر، اعظم؛ رضایی، احمد؛ نوروزی، لیلا و مختاریان، فرانک. (۱۳۹۶). «کارکرد افتراقی سؤال و آزمون مرتبط با جنسیت در آزمون‌های کنکور سراسری سازمان سنجش آموزش کشور». *فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۷(۱۹)، ۴۱-۷۳.
- محسن‌پور، مریم؛ گویا، زهرا؛ شکوهی یکتا، محسن؛ کیامنش، علیرضا و بازرگان، عباس. (۱۳۹۴). «سنجش تشخیصی صلاحیت‌های سواد ریاضی». *فصلنامه نوآوری‌های آموزشی*، ۱۴(۵۳)، ۷-۳۳.

- مقدم، اعظم؛ فلسفی نژاد، محمدرضا؛ فرخی، نورعلی؛ استاجی، معصومه. (۱۳۹۵). «تحلیل تشخیصی سؤالات بخش درک مطلب زبان انگلیسی عمومی آزمون ورودی دوره‌های دکتری با استفاده از مدل غیر جبرانی فیوژن». *فصلنامه اندازه‌گیری تربیتی*، ۶(۲۲)، ۴۱-۶۸.
- مینائی، اصغر. (۱۳۹۱). مدل‌پردازی تشخیصی شناختی سؤال‌های تیمز ۲۰۰۷ در دانش‌آموزان پایه هشتم ایران با استفاده از مدل یکپارچه با پارامترپردازی مجدد (RUM) و مقایسه مهارت‌های ریاضی دانش‌آموزان دختر و پسر، پایان‌نامه دکتری دانشگاه علامه طباطبائی.
- Amirian, S. M.R., Alavi, S. M., Fidalgo, A. M. (2014). "Analyzing Gender Differences with an English Proficiency Test in EFL Context". *Iranian Journal of Language Testing*, 4(2), 187-203.
- Barati, H., & Ahmadi, A. R. (2010). "Gender-based DIF across the subject area: A study of the Iranian National University Entrance Exam". *The Journal of Teaching Language Skills*, 2(3), 1-22.
- Bejar, I. I. (2008). "Standard setting: What is it? Why is it important?", *R&D Connection*, 7, 1-6.
- De la Torre, J., & Douglas, J. (2004). "Higher order latent trait models for cognitive diagnosis". *Psychometrika*, 69(3), 333-353. <https://doi.org/10.1007/BF02295640>
- De la Torre, J. (2008). "An empirically based method of Q-matrix validation for the DINA model: development and applications". *Journal of Educational Measurement*, 45(4), 343-362. <https://doi.org/10.1111/j.1745-3984.2008.00069.x>
- De la Torre, J., Lee, Y. S. (2010). "A note on the invariance of the DINA model parameters". *Journal of Educational Measurement*, 47(1), 115-127. <https://doi.org/10.1111/j.1745-3984.2009.00102.x>
- De la Torre, J. (2011). "The generalized DINA model framework". *Psychometrika*, 76(2), 179-199. <https://doi.org/10.1007/s11336-011-9207-7>
- DiBello, L. V., Roussos, L. A., Stout, W. F. (2007). "Review of cognitively diagnostic assessment and summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.)", *Handbook of statistics. Volume 26: Psychometrics* (pp. 979-1030). Amsterdam, The Netherlands: Elsevier.
- Ercikan, K. (2002). "Disentangling sources of differential item functioning in multilanguage assessments", *International Journal of Testing*, 2(3-4), 199-215. <https://doi.org/10.1080/15305058.2002.9669493>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Hou, L., de la Torre, J., Nandakumar, R. (2014). "Differential item functioning assessment in cognitive diagnosis modeling: Applying Wald test to investigate DIF for DINA model". *Journal of Educational Measurement*, 51(1), 98-125. <https://doi.org/10.1111/jedm.12036>
- Junker, B. W., Sijtsma, K. (2001). "Cognitive assessment models with few assumptions, and connections with nonparametric item response theory". *Applied Psychological Measurement*, 25(3), 258-272. <https://doi.org/10.1177/01466210122032064>.
- Karami, H. (2011). "Detecting gender bias in a language proficiency test". *International Journal of Language Studies*, 5(2), 27-38.

- Leighton, J. P., & Gierl, M.J. (Eds.) (2007). *Cognitive diagnostic assessment for education: Theory and applications* (pp. 146-172). Cambridge University Press.
- Li, F. M. (2008). *A modified higher-order DINA model for detecting differential item functioning and differential attribute functioning*. Unpublished doctoral dissertation, University of Georgia.
- Li, X., Wang, W. C. (2015). "Assessment of differential item functioning under cognitive diagnosis models: the DINA model example". *Journal of Educational Measurement*, 52(1), 28-54. <https://doi.org/10.1111/jedm.12061>
- Liu, O. L., and Wilson, M. (2009). "Gender differences and similarities in PISA 2003 mathematics: A comparison between the United States and Hong Kong". *International Journal of Testing*, 9(1), 20-40. <https://doi.org/10.1080/15305050902733547>.
- Liu Y, Yin H, Xin T, Shao L, Yuan L. (2019). "A Comparison of Differential Item Functioning Detection Methods in Cognitive Diagnostic Models". *Frontiers in Psychology*, 10, 1137. <https://doi.org/10.3389/fpsyg.2019.01137>
- Ma, W., de la Torre, J. (2017). Available at: <https://CRAN.R-project.org/package=GDINA>.
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York: Routledge, Taylor and Francis group.
- Perrone, M. (2006). "Differential item functioning and item bias: Critical considerations in test fairness". *Columbia University Working Papers in TESOL & Applied Linguistics*. 6(2), 1-3.
- Ravand, H., Barati, H., Widhiarso, W. (2012). "Exploring Diagnostic Capacity of a High Stakes Reading Comprehension Test: A pedagogical demonstration". *Iranian Journal of Language Testing*, 3(1), 11-37.
- Ravand, H., Robitzsch, A. (2015). "Cognitive Diagnostic Modeling Using R". *Practical Assessment, Research & Evaluation*, 20(11), 1-12. <https://doi.org/10.7275/5g6f-ak15>.
- Roussos, L.A., Templin, J. L., Henson, R. A. (2007). "Skills diagnosis using IRT-based latent class models". *Journal of Educational Measurement*, 44(4), 293-311. <https://doi.org/10.1111/j.1745-3984.2007.00040.x>
- Rupp, A. A., Templin, J., Henson, R. A. (2010) *Diagnostic measurement: theory, methods, and applications*. Guilford, New York.
- Swaminathan, H., Rogers, H. J. (1990). "Detecting differential item functioning using logistic regression procedures". *Journal of Educational Measurement*, 27(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Von Davier, M. (2013). "The DINA model as a constrained general diagnostic model: Two variants of a model equivalency". *British Journal of Mathematical and Statistical Psychology*, 67(1), 49-71. <https://doi.org/10.1111/bmsp.12003>
- Wang, Z. R., Guo, L., Bian, Y. F. (2014). "Comparison of DIF detecting methods in cognitive diagnostic test". *Acta Psychologica Sinica*, 46(12), 1923-1932. <https://doi.org/10.3724/sp.j.1041.2014.01923>.
- Zhang, W. (2006). *Detecting differential item functioning using the DINA model*. Unpublished doctoral dissertation, University of North Carolina at Greensboro.