

کنش افتراقی سؤال در آزمون‌های سازمان سنجش آموزش کشور بر حسب ویژگی‌های جمعیت شناختی داوطلبان کنکور سراسری^۱

مسعود گرامی پور^۲، احمد رضایی^۳، اعظم رمضان صدر^۴، لیلا نوروزی^۵

تاریخ دریافت: ۹۵/۰۲/۲۵

تاریخ پذیرش: ۹۵/۱۰/۱۵

چکیده

کنکور سراسری بنا به اذعان بسیاری از کارشناسان سرنوشت‌سازترین آزمون در ایران است و ویژگی‌های جمعیت شناختی داوطلبان از قبیل جنسیت، منطقه (پایگاه اجتماعی-اقتصادی) و استان (زبان) می‌توانند بر عملکرد داوطلبان ورود به دانشگاه در پاسخ به سؤال‌های آزمون مؤثر باشند. اگر پاسخ به سؤال‌های آزمون با کنترل توانایی آزمودنی‌ها تابعی از ویژگی‌های جمعیت شناختی باشند، سؤال‌های آزمون نسبت به آن‌ها کنش افتراقی (DIF) دارند. هدف از پژوهش حاضر بررسی کنش افتراقی سؤال‌های آزمون سراسری سازمان سنجش آموزش کشور بر حسب ویژگی‌های جمعیت شناختی داوطلبان است. نمونه تحقیق شامل همه آزمودنی‌های یک دفترچه آزمون و برخی دروس امتحانی تخصصی گروه‌های آزمایشی کنکور سراسری از سال ۸۷ تا ۹۰ بود. به منظور تحلیل DIF از رویکرد تحلیل رگرسیون لجستیک دوجبهی استفاده شد. پس از اطمینان از تک‌بعدی بودن آزمون‌های مورد مطالعه از طریق رویکرد NOHARM، یافته‌های تحلیل DIF حاکی از این بود که به ترتیب بیشترین تعداد کنش افتراقی آشکار شده در سؤال‌های

۱. این پژوهش با حمایت سازمان سنجش آموزش کشور انجام شده است.

۲. استادیار دانشکده روانشناسی و علوم تربیتی دانشگاه خوارزمی (نویسنده مسئول) mgramipour@yahoo.com

۳. دانشجوی دکترای سنجش و اندازه‌گیری دانشگاه علامه طباطبائی

۴. کارشناس ارشد تحقیقات آموزشی دانشگاه خوارزمی

۵. کارشناس ارشد تحقیقات آموزشی دانشگاه خوارزمی

آزمون، مربوط به متغیرهای جنسیت، سهمیه مناطق (پایگاه اجتماعی-اقتصادی) و استان (زبان) هستند، اما مقادیر اندازه‌های اثر بسیار کوچک و ناچیز بودند. باین‌وجود، بر اساس یافته‌های پژوهش حاضر پیشنهاد می‌شود که کمیته‌های موضوعی تشکیل شده و بر اساس قضاوت نهایی آن‌ها، سؤال‌های سودار آزمون مشخص شوند و در طراحی سؤال در آزمون‌های آینده مدنظر قرار گیرند.

واژگان کلیدی: کنکور سراسری، کنش افتراقی سؤال، ویژگی‌های جمعیت شناختی داوطلبان، رگرسیون لجستیک

مقدمه

کنکور سراسری در ایران بدون شک سرنوشت‌سازترین آزمونی است که همه‌ساله برگزار می‌شود و تعداد زیادی از داوطلبان به امید کسب موفقیت با رعایت عدالت در آن به رقابت می‌پردازند. در طول سالیان گذشته سنجش علمی داوطلبان به‌عنوان سازوکار اصلی ورود داوطلبان به دوره‌های آموزش عالی همواره مورد مناقشه بوده است. آزمون‌ها و حرفه آزمونگری مورد توجه دقیق مردم قرار گرفته و اکنون منتشرکنندگان و معریان آزمون‌ها باید نشان دهند که آزمون‌هایشان در برابر اقلیت‌ها، عاری از سوگیری هستند (همبلتون، سوامیناتان و راجرز^۱، ۱۹۹۱؛ ترجمه فلسفی نژاد، ۱۳۸۹). از این‌رو عادلانه بودن برای آزمون‌های سرنوشت‌ساز به‌عنوان آزمون‌هایی که جهت تصمیم‌گیری برای آینده آزمودنی‌ها بکار می‌روند از اهمیت بیشتری برخوردار است (پاپهام^۲، ۲۰۰۵).

بنا به اذعان کارشناسان (کمیلی و شپارد^۳، ۱۹۹۴؛ وندربرگ^۴، ۲۰۰۲؛ برن و استوارت^۵، ۲۰۰۶؛ میلپس^۶، ۲۰۱۱) یکی از مهم‌ترین تهدیدها برای اعتبار^۷ آزمون، سوگیری سؤال^۸ است که از آن تحت عنوان کنش افتراقی سؤال (DIF)^۹ نیز یاد می‌شود. وجود DIF

1. Hambleton, Swaminathan & Rogers
2. Popham
3. Camilli & Shepard
4. Vanderberg
5. Byrne & Stewart
6. Millsap
7. Validity
8. Item Bias
9. Differential Item Functioning

گواهی علیه عادلانه بودن آزمون است، اگر پاسخ به سؤال‌های آزمون با کنترل توانایی آزمودنی‌ها تابعی از ویژگی‌های گروهی که آزمودنی‌ها به آن تعلق دارند، مانند (جنسیت، پایگاه اجتماعی و اقتصادی و زبان و...) باشند، سؤال‌های آزمون نسبت به آن‌ها کنش افتراقی (DIF) دارند. بنابراین کنش افتراقی سؤال‌های آزمون را بر اساس گروه‌بندی بررسی می‌کنند، به نحوی که گروه‌های متفاوت در ادبیات DIF اصطلاحاً گروه‌های مرجع و کانونی^۱ نامیده می‌شوند. معمولاً گروه مرجع، گروه اکثریت و گروه کانونی گروه اقلیت یا محروم در نظر گرفته می‌شوند (سوامیناتان و راجرز^۲، ۱۹۹۰). این قضیه معمولاً موقعی اتفاق می‌افتد که با کنترل توانایی، عملکرد گروه کانونی درباره یک سؤال ضعیف‌تر از گروه مرجع باشد (میلسپ، ۲۰۱۱). بنابراین هدف اصلی پژوهش حاضر مطالعه کنش افتراقی سؤال در آزمون‌های کنکور سراسری سازمان سنجش آموزش کشور بر اساس ویژگی‌های جمعیت شناختی داوطلبان است.

به منظور درک بهتر DIF لازم است به روش‌های مورداستفاده در تشخیص آن اشاره شود. روش‌های آماری مختلفی جهت تعیین DIF وجود دارد که می‌توان آن‌ها را به دو دسته کلی روش‌های مبتنی بر نظریه پاسخ سؤال (IRT)^۳ و کلاسیک تقسیم کرد. روش‌های موجود در گروه اول به منظور مقایسه پارامترهای سؤال و مساحت میان منحنی‌های ویژگی سؤال (ICC)^۴ مورداستفاده قرار گرفته (همبلتون و سوامیناتان، ۱۹۸۵) و روش‌های گروه دوم مبتنی بر نظریه اندازه‌گیری کلاسیک (CTT)^۵، تحلیل عاملی و تحلیل داده‌های طبقه‌ای از جمله روش رگرسیون لجستیک هستند (گومز و ناواس^۶، ۲۰۰۰). در شرایطی که حجم نمونه بزرگ باشد رویکرد رگرسیون لجستیک کارایی زیادی دارد (گرامی‌پور، ۱۳۹۳)، که در پژوهش حاضر نیز این رویکرد آماری نسبت به DIF مدنظر

1. Reference and focal groups
2. Rogers
3. Item Response Theory
4. Item Characteristic Curve
5. Classical Test Theory
6. Gomez & Navas

قرار گرفته است. DIF معمولاً به خاطر اثر چندبعدی بودن^۱ به وجود می‌آید که به آن مؤلفه مزاحم ثانوی گفته می‌شود. بنابراین اگر تدوین آزمونی تک‌بعدی^۲ مدنظر است، باید استفاده از سؤال‌های سودار مورد پرهیز قرار گیرد (شیلی و استات^۳، ۱۹۹۳). بنابراین آشکارسازی DIF یک فرایند آماری است که در این فرایند سؤال‌هایی که کنش افتراقی دارند مشخص شده و برای آن اندازه اثر^۴ محاسبه می‌شود. اما این فرایند در این مرحله پایان نمی‌یابد، بلکه گام آخر در تشخیص سوگیری سؤال بر اساس نظرات کیفی و محتوایی اعضای گروهی از متخصصان موضوع درسی و طراحی سؤال خواهد بود (کمیلی و شپارد، ۱۹۹۴).

یکی از نمونه‌های شواهد تجربی در دستیابی به عدالت آزمون، کشف سوگیری در گروه‌های آزمودنی بر اساس برخی ویژگی‌های جمعیت شناختی، به‌عنوان مثال، جنسیت، زبان، و یا وضعیت قومی است (کول و زیکی^۵، ۲۰۰۱ و مک نامارا و روور^۶، ۲۰۰۶؛ به نقل از سانگ، چنگ و کلینجر^۷، ۲۰۱۵). سایر یافته‌ها نشان داده است که سرمایه فرهنگی، اجتماعی و اقتصادی خانواده سهم مثبت و معنی‌داری در به دست آوردن رتبه‌های خوب و رشته‌های ممتاز کسب‌شده از سوی دانشجویان داشته است (نوربخش، ۱۳۸۹). به اعتقاد متخصصان سنجش و اندازه‌گیری، ویژگی‌های دیگری هم به‌جز جنسیت و زبان مادری، که تأثیر بسزایی در عملکرد افتراقی دارند، در این زمینه وجود دارند که می‌توان آن‌ها را در نظر گرفت. در این زمینه می‌توان به رشته تحصیلی و آشنایی با یک فرهنگ قومی خاص اشاره کرد (لئو، شدل، مالوی و کانگ^۸، ۲۰۱۱؛ به نقل از یانگ، مورگان، ریبنسکی^۹ و همکاران، ۲۰۱۳).

1. Multidimensionality
2. Unidimensional
3. Shealy & Stout
4. Effect size
5. Cole & Zieky
6. McNamara & Roever
7. Song, Cheng & Klinger
8. Liu, Schedl, Malloy & Kong
9. Young, Morgan & Rybinski

تحقیقات چندی پیرامون موضوع کارکرد افتراقی سؤال‌های آزمون انجام شده است. یافته‌ها در بیشتر موارد نشان از وجود عملکرد افتراقی در سؤال‌های آزمون بر اساس ویژگی‌های متفاوت است. نوربخش (۱۳۸۹) و خدایی (۱۳۸۸، ب) موفقیت تحصیلی دانشجویان را متأثر از سرمایه‌های فرهنگی، اجتماعی و اقتصادی خانواده می‌دانند. مطالعه پیشینه‌های مرتبط با موضوع نشان می‌دهد که وجود کنش افتراقی در آزمون‌ها متصور است، به نحوی که سانگ، چنج و کلینگر^۱ (۲۰۱۵) کنش افتراقی سؤال بر اساس گروه‌های جنسیتی و سابقه دانشگاهی در آزمون بزرگ مقیاس را بررسی کرده و نشان داده‌اند که درصد بالایی از سؤال‌های آزمون ورودی تحصیلات تکمیلی نسبت به مؤلفه‌های گروهی مورد بررسی کنش افتراقی دارند.

از جمله عواملی که در ایجاد کنش افتراقی مؤثر است، جنسیت است که تحقیقات زیادی پیرامون آن انجام شده است. از جمله اوگرتمن^۲ (۲۰۱۵) نشان داد، منحنی‌های ویژگی سؤال در خرده آزمون درک مطلب زبان به لحاظ جنسیتی دارای DIF ناهماهنگ هستند. امیریان، علوی و فیدلگو^۳ (۲۰۱۴) در پژوهش خویش نشان دادند که ۲۸٪ از سؤال‌های آزمون دارای کنش افتراقی جنسیتی هستند، هرچند که اندازه‌های اثر کوچک بودند. براتی و احمدی^۴ (۲۰۱۰) اثر جنسیت و موضوع درسی را در کنش افتراقی تأیید کردند. عبدالعزیز^۵ (۲۰۱۰) کنش افتراقی سؤال‌های آزمون ریاضی و رضایی و شعبانی^۶ (۲۰۱۰) کنش افتراقی سؤال‌های آزمون مهارت انگلیسی دانشگاه تهران را بر اساس جنسیت بررسی کردند، آن‌ها نشان دادند که جنسیت در عملکرد آزمودنی‌ها در پاسخ به سؤال‌های آزمون مؤثر است. دودین و انابی^۷ (۲۰۰۸) در کنش افتراقی سؤال‌های آزمون تیمز نشان دادند که بیش از ۲۰ درصد از سؤال‌ها در هر دفترچه آزمون دارای DIF مرتبط با جنسیت

1. Song, Cheng & Klinger
2. Öğretmen
3. Amirian, Alavi & Fidalgo
4. Barati & Ahmadi
5. Abedalaziz
6. Rezaee & shabani
7. Doudeen & Annabi

هستند. لی^۱ (۱۹۹۹) در یک مطالعه طولی و های تاکالا و کافتانوا^۲ (۲۰۰۰)، کنش افتراقی جنسیتی را در سؤال‌های آزمون مورد مطالعه قرار دادند. آن‌ها به این نتیجه رسیدند که زنان و مردان در سؤال‌هایی که در مورد نقش‌های قالبی جنسیت خودشان است عملکرد بهتری نشان می‌دهند، مثلاً مردان سؤال‌هایی که محتوای ورزشی و زنان سؤال‌هایی که محتوای خرید داشت را بهتر درک می‌کردند.

بر اساس ادبیات مرتبط با تحقیق علاوه بر تأثیر جنسیت، متغیرهای اجتماعی دیگری هم می‌توانند منبع کنش افتراقی باشند. شاموگام و سالان^۳ (۲۰۱۴) نشان دادند که زبان در کنش افتراقی سؤال‌های آزمون مؤثر است. یونگ، مورگان، ریبنسکی و همکاران (۲۰۱۳) در آزمون کنش افتراقی سؤال‌های آزمون تافل نشان دادند که داده‌های مربوط به متغیر گروهی زبان دوم، DIF سؤال‌های آزمون را تبیین می‌کند. هاردینگ^۴ (۲۰۱۲) نقش لهجه، زبان اول و زبان دوم در کنش افتراقی سؤال‌های آزمون را نشان دادند. همچنین بارنز و ولز^۵ (۲۰۰۹) نقش جنسیت و نژاد را در کنش افتراقی بررسی نموده و آشکار کردند که زنان و دانشجویان رنگین‌پوست در پاسخ به سؤال‌های آزمون متفاوت عمل می‌کنند. بوروئر^۶ (۲۰۰۵) کنش افتراقی سؤال‌های آزمون ورودی تحصیلات تکمیلی را مطالعه کردند. آن‌ها در پژوهش خود نشان دادند که نژاد و زبان دوم باعث ایجاد DIF با شدت متوسط در سؤال‌های آزمون می‌شوند.

سرانجام برخی صاحب‌نظران حوزه سوگیری سؤال‌های آزمون از جمله اون^۷ (۱۹۹۲)، کمیلی و شپارد (۱۹۹۴)، ون دی وجیجور^۸ (۱۹۹۸) و ون زیل و ویسر^۹ (۱۹۹۸) به صورت متفق‌القول معتقدند که به جز عوامل فوق‌الذکر، پایگاه اجتماعی و اقتصادی آزمودنی‌ها نیز

-
1. Ie
 2. takala & Kaftandjjeva
 3. Shanmugam & Saw Lan
 4. Harding
 5. Barnes & Wells
 6. Broer
 7. Owen
 8. Van de Vijver
 9. Van Zyl & Visser

می‌تواند عامل دیگری در کنش افتراقی سؤال‌های آزمون باشد. در این رابطه، شاتلورث-جوردن^۱ (۱۹۹۶) در پژوهشی در مورد کاربرد آزمون‌های شناختی در کشور آفریقای جنوبی نشان داد که یکی از عوامل مؤثر در سوگیری سؤال‌های آزمون، پایگاه اجتماعی و اقتصادی آزمودنی‌ها است که از نظر ایشان این امر ممکن است به واسطه بافت متغیر جامعه آفریقایی به لحاظ پایگاه اجتماعی و اقتصادی باشد. همچنین در یک پژوهش تجربی دیگر دینیس داکاستا و آروجو^۲ (۲۰۱۲) در یک مطالعه کنش افتراقی سؤال در داده‌های برنامه سنجش بین‌المللی دانش آموزان (PISA^۳) نشان دادند که برحسب کشورهای مختلف، دانش آموزان مهاجر که دارای پایگاه اجتماعی مختلف هستند در پاسخ به سؤال‌های آزمون خواندن عملکرد افتراقی دارند.

مرور تحقیقات گذشته نشان می‌دهد که هنوز در ایران در رابطه با کنش افتراقی آزمون‌های استاندارد ملی که آزمون سراسری ورود به دانشگاه‌ها در صدر آن قرار دارد، به‌جز چند پژوهش محدود در زمینه زبان انگلیسی تخصصی کنکور سراسری (براتی و احمدی، ۲۰۱۰؛ ایزانلو و حبیبی، ۱۳۸۹)، پژوهشی انجام نشده است. همچنین در تحقیقات گذشته داخلی و خارجی، کنش افتراقی سؤال‌های آزمون تنها برحسب یک ویژگی و آن‌هم بیشتر به لحاظ جنسیت مورد مطالعه قرار گرفته است. اما در پژوهش حاضر اطلاعات در دسترس داوطلبان کنکور سراسری مورد استفاده قرار می‌گیرد و سهمیه‌بندی ۲ منطقه اول در کنکور سراسری برحسب شهرهای بزرگ (منطقه ۱)، و به نسبت تعداد داوطلبان، شهرهای کوچک‌تر (منطقه ۲) مدنظر قرار گرفته است که می‌توان به‌طور نسبی آن را شاخصی برای پایگاه اجتماعی و اقتصادی دانست. بعلاوه، به واسطه تفاوت‌های زبانی اکثریت داوطلبان در برخی استان‌های خاص، می‌توان این متغیر را شاخصی نسبی برای مقایسه زبان تکلم در برخی استان‌ها دانست. لذا جای خالی پژوهشی در رابطه با ویژگی‌های چندگانه مؤثر بر کنش افتراقی در آزمون سراسری ضروری به نظر می‌رسد.

1. Shuttleworth-Jordan
2. Dinis da Costa & Araújo
3. Programme for International Student Assessment

از این رو، در پژوهش حاضر سعی بر آن است که اعتبار آزمون سراسری کارشناسی از منظر کنش افتراقی سؤال‌های آزمون (DIF) در برخی دروس تخصصی گروه‌های آزمایشی کنکور سراسری، مورد بررسی و تحلیل قرار گیرد. بنابراین اهدافی که این پژوهش دنبال می‌کند، شامل مطالعه کنش افتراقی سؤال‌های آزمون بر اساس برخی ویژگی‌های جمعیت شناختی شامل جنسیت، منطقه و استان در برخی دروس تخصصی رشته‌های مختلف کنکور سراسری و مقایسه DIF در گروه‌های آزمایشی داوطلبان کنکور سراسری و مطالعه اندازه اثر آن‌ها از سال ۱۳۸۷ تا ۱۳۹۰ است.

روش پژوهش

پژوهش حاضر از نظر هدف کاربردی و از نظر روش، توصیفی موسوم به تحقیقات روان‌سنجی است.

جامعه و نمونه تحقیق شامل همه آزمودنی‌های یک دفترچه آزمون گروه‌های آزمایشی کنکور سراسری از سال ۸۷ تا ۱۳۹۰ بود. جدول ۱ خلاصه اطلاعات نمونه آزمودنی‌ها بر اساس ویژگی‌های جمعیت شناختی مورد مطالعه به تفکیک سال‌های برگزاری کنکور سراسری را نشان می‌دهد.

جدول ۱. نمونه تحقیق به تفکیک سال‌های برگزاری آزمون و ویژگی‌های جمعیت شناختی

سال و رشته	متغیر		استان		منطقه		جنسیت
	اصفهان	آذربایجان	منطقه ۲	منطقه ۱	مرد	زن	
سال ۸۷	ریاضی	۲۳۰۱	۱۴۱۷	۱۲۸۲۳	۹۱۴۴	۱۵۶۱۷	۱۲۴۶۵
	تجربی	۲۶۷۲	۲۱۰۹	۱۷۶۳۶	۷۶۲۷	۱۲۰۷۷	۲۹۶۳۴
	انسانی	۲۴۶۹	۱۹۴۶	۱۶۱۹۳	۶۴۴۹	۱۵۴۷۷	۳۰۵۱۷
	هنر	۵۱۵	۱۵۹	۲۰۰۶	۱۷۰۲	۱۴۳۳	۳۷۴۰
	زبان	۹۲۶	۵۷۷	۴۷۹۹	۳۷۰۷	۳۹۳۶	۸۳۰۷
سال ۸۸	ریاضی	۲۲۷۱	۱۳۸۱	۱۲۵۸۰	۸۷۴۹	۱۵۶۲۱	۱۱۸۸۸
	تجربی	۲۴۷۴	۱۸۵۶	۱۶۹۱۷	۶۸۹۱	۱۱۹۷۵	۲۷۷۲۲
	انسانی	۲۱۴۷	۱۷۲۳	۱۴۶۵۶	۵۶۰۱	۱۴۳۰۸	۲۶۳۵۸
	هنر	۴۴۴	۱۴۳	۱۷۲۰	۱۵۶۶	۱۲۹۶	۳۲۳۹
	زبان	۷۹۳	۴۸۲	۴۲۱۴	۳۱۵۵	۳۳۱۰	۷۲۳۲

۱۲۰۴۴	۱۵۶۹۸	۸۶۴۵	۱۲۶۴۲	۱۲۹۸	۲۲۳۵	ریاضی	
۲۸۲۰۰	۱۲۵۵۷	۶۸۷۵	۱۷۳۸۰	۱۸۸۱	۲۴۱۱	تجربی	
۲۵۰۵۵	۱۴۱۴۸	۵۴۴۴	۱۴۰۰۹	۱۵۷۰	۲۰۹۴	انسانی	سال ۸۹
۲۹۹۳	۱۳۸۷	۱۴۰۹	۱۶۴۱	۱۳۴	۴۴۴	هنر	
۸۲۵۹	۴۰۳۴	۳۵۴۸	۴۹۲۲	۵۱۰	۹۲۰	زبان	
۱۰۷۷۸۴	۱۴۴۲۹۶	۷۸۰۷۸	۱۱۴۱۲۵	۱۱۵۶۵	۲۰۳۳۲	ریاضی	
۱۶۷۰۵۶	۶۵۶۵۰	۶۵۳۰۶	۱۶۷۴۰۰	۱۱۵۶۴	۲۰۷۵۷	تجربی	
۲۰۶۳۰۲	۱۱۷۰۰۷	۴۴۱۷۷	۱۱۶۶۴۶	۱۱۷۳۱	۱۷۸۷۲	انسانی	سال ۹۰
۲۷۶۳۶	۱۲۵۶۴	۱۳۰۴۳	۱۵۳۶۱	۱۲۰۷	۴۲۷۳	هنر	
۷۲۹۰۷	۳۴۵۵۸	۳۱۷۶۹	۴۳۳۵۴	۴۲۱۷	۸۵۷۳	زبان	

ابزارها: ابتدا دروس امتحانی یک دفترچه آزمون گروه‌های آزمایشی کنکور سراسری از سال ۸۷ تا ۱۳۹۰ انتخاب شدند. سپس از میان دروس برای هر گروه آزمایشی دروس تخصصی انتخاب شدند. دروس انتخاب شده شامل درس ریاضی (شامل ۵۰ سؤال)، فیزیک (۴۵ سؤال) و شیمی (۳۵ سؤال) از گروه ریاضی و فنی، زیست‌شناسی (۵۰ سؤال) و شیمی (۳۵ سؤال) از گروه علوم تجربی، ادبیات فارسی (۳۰ سؤال) از گروه علوم انسانی، درک عمومی هنر (۳۰ سؤال)، خلاقیت تصویری و تجسمی (۲۰ سؤال) و خلاقیت نمایشی (۲۰ سؤال) از گروه هنر و زبان انگلیسی تخصصی (۷۰ سؤال) از گروه زبان بودند.

روش تجزیه و تحلیل داده‌ها: جهت تشخیص کنش افتراقی سؤال‌های آزمون از رویکرد رگرسیون لجستیک دووجهی^۱ و نرم‌افزارهای SPSS و بسته difR در نرم‌افزار R استفاده شد. این رویکرد شامل سه گام در قالب ۳ معادله است که در آن معادله‌ها، متغیر تابع پاسخ (درست یا غلط) به سؤال مورد تحلیل است، ۱- در معادله اول متغیر پیش‌بین (نمره کل آزمون) وارد معادله می‌شود، ۲- سپس در معادله دوم نشانگر عضویت گروهی (جنسیت، استان اخذ دیپلم و منطقه) وارد معادله می‌شود، ۳- و در معادله سوم متغیر تعامل دو متغیر وارد شده قبلی وارد معادله رگرسیون لجستیک می‌شود. با استفاده از لگاریتم نسبت درستنمایی در آزمونی با $df=2$ ، مدل مرحله ۱ و ۲ و مدل مرحله ۲ و ۳ مقایسه می‌شوند. این آزمون، دو آزمون دارای $df=1$ را ترکیب می‌کند. آزمون اول مدل مرحله ۱ را با مدل

1. binary

مرحله ۲ مقایسه می‌کند که در این آزمون DIF هماهنگی را مورد بررسی قرار می‌گیرد. رد فرض صفر به معنی وجود DIF هماهنگی است. آزمون دوم نیز با استفاده از لگاریتم نسبت درستنمایی، مدل مرحله ۲ را با مدل مرحله ۳ مقایسه می‌کند. این مقایسه، DIF ناهماهنگی را ارزیابی می‌کند (گرامی پور، ۱۳۹۳). بعلاوه، یکی از پیش‌فرض‌های تحلیل داده‌ها به‌منظور آشکارسازی DIF، تک‌بعدی بودن داده‌ها است (میزور، همبلتون و کلوزر، ۱۹۹۸). جهت آزمون تک‌بعدی بودن داده‌ها از نرم‌افزار NOHARM^۳ استفاده شد و برای تأیید آن، ۳ شاخص مجموع مجذورات باقی‌مانده‌ها، ریشه دوم مجذور میانگین باقی‌مانده‌ها و شاخص تاناکا^۴ و تاناکا مورد استناد قرار گرفت.

نتایج

نتایج ۳ شاخص مورد استناد در NOHARM حاکی از تک‌بعدی بودن همه آزمون‌های مورد مطالعه بود، بدین ترتیب که ۱- مجموع مجذورات باقی‌مانده برای همه آزمون‌های مورد مطالعه کوچک‌تر یا مساوی با ۰/۰۰۵ بودند، ۲- مجذور میانگین ریشه باقی‌مانده برای همه آزمون‌ها کوچک‌تر مساوی با ۰/۰۱ بودند و ۳- مقدار شاخص تاناکا و تاناکا که یک شاخص پیشینه درستنمایی است، برای همه آزمون‌ها بزرگ‌تر مساوی با ۰/۹۵ بود.

نتایج رویکرد تحلیل رگرسیون لجستیک با استفاده از نرم‌افزارهای SPSS و R بر اساس آماره نسبت درستنمایی نشان داد که هیچ‌گونه DIF ناهماهنگی معنی‌داری در سؤال‌های آزمون‌های کنکور سراسری وجود ندارد ($LRT^5 < 1, P > 0/05$). اما بر اساس آماره آزمون نسبت درستنمایی، برخی از سؤال‌های آزمون‌های کنکور سراسری برحسب ویژگی‌های جمعیت شناختی به شرح جداول ۲ تا ۵ دارای DIF هماهنگی معنی‌داری بودند ($LRT > 4, P < 0/05$). نتایج آشکارسازی DIF برحسب ویژگی‌های جمعیت شناختی در قالب دستیابی به اهداف پژوهش به شرح زیر هستند:

1. Mazor
2. Clauser
3. Normal Ogive Harmonic Model
4. Tanaka
5. Likelihood Ratio Test

نتایج آشکارسازی DIF برحسب جنسیت، سال‌های برگزاری آزمون و گروه‌های آزمایشی در جدول ۲ ارائه شده است.

جدول ۲. نتایج تحلیل سؤال‌های برحسب جنسیت و گروه آزمایشی در سال‌های ۱۳۸۷ تا ۱۳۹۰

گروه آزمایشی	سال		۱۳۸۸		۱۳۸۹		۱۳۹۰		جمع	درصد
	مرد	زن	مرد	زن	مرد	زن	مرد	زن		
ریاضی	۹	۴	۸	۱۰	۱۲	۹	۱۶	۹	۸۳	۱۵٪
جمع	۱۳	۱۸	۲۷	۲۵						
علوم تجربی	۸	۳	۶	۳	۷	۶	۹	۷	۴۹	۱۴٪
جمع	۱۱	۹	۱۳	۱۶						
علوم انسانی	۲	۱	۲	۳	۴	۱	۱	۲	۱۶	۱۳٪
جمع	۳	۵	۵	۳						
هنر	۱	۳	۱	۴	۲	۱	۷	۵	۲۴	۸٪
جمع	۴	۵	۳	۱۲						
زبان	۴	۱	۶	۲	۴	۳	۸	۳	۲۹	۱۰٪
جمع	۵	۸	۵	۱۱						
جمع کل	۳۶	۴۵	۵۳	۶۷					۲۰۱	۱۳٪
درصد DIF در سؤال‌های آزمون	۹٪	۱۲٪	۱۴٪	۱۷٪						

همان‌طور که در جدول ۲ ملاحظه می‌شود، در مجموع، تعداد سؤال‌های DIF مرتبط با جنسیت در آزمون‌های مورد مطالعه در سال ۸۷ کمتر از ۱۰ درصد، در سال‌های ۸۸ و ۸۹ کمتر از ۱۵ درصد و در سال ۱۳۹۰ کمتر از ۲۰ درصد است، در حالی که هر سال تعداد سؤال‌های DIF در دروس مورد مطالعه نسبت به سال قبل به صورت تقریباً یکنواخت افزایش یافته است. نسبت کنش افتراقی در رشته‌ها هم نشان می‌دهد که رشته ریاضی با بیشترین میزان کنش افتراقی سؤال‌های آزمون برحسب متغیر جنسیت را دارد و بعد از آن به ترتیب رشته‌های علوم تجربی (۱۴٪)، علوم انسانی (۱۳٪)، زبان (۱۰٪) و هنر (۸٪) قرار گرفته‌اند. به طور میانگین، میزان کنش افتراقی سؤال در آزمون‌های مورد مطالعه برحسب متغیر جنسیت برای ۴ سال متوالی ۱۳٪ است.

نتایج آشکارسازی DIF برحسب استان، که با هدف مقایسه یک دو استان با زبان متفاوت انجام شده است و در جدول ۳ ملاحظه می‌شود.

جدول ۳. نتایج تحلیل سؤال برحسب استان و گروه آزمایشی در سال‌های ۱۳۸۷ تا ۱۳۹۰

گروه آزمایشی	سال		۱۳۸۷		۱۳۸۸		۱۳۸۹		۱۳۹۰		جمع درصد
	اصفهان	آذربایجان	اصفهان	آذربایجان	اصفهان	آذربایجان	اصفهان	آذربایجان	اصفهان	آذربایجان	
ریاضی	۱	۲	۱	۱	۳	۰	۳	۰	۴	۴	۱۸٪
جمع	۳	۳	۴	۳	۸	۸	۸	۸	۸	۸	
علوم تجربی	۲	۱	۲	۲	۰	۱	۲	۳	۴	۳	۱۵٪
جمع	۳	۳	۲	۳	۷	۷	۷	۷	۷	۷	
علوم انسانی	۱	۱	۲	۰	۰	۰	۱	۱	۱	۱	۶٪
جمع	۱	۱	۲	۲	۲	۱	۲	۲	۲	۲	
هنر	۰	۰	۱	۰	۰	۰	۳	۰	۴	۱	۹٪
جمع	۰	۰	۱	۱	۵	۵	۵	۵	۵	۵	
زبان	۰	۱	۰	۰	۰	۰	۲	۰	۰	۰	۳٪
جمع	۱	۱	۰	۰	۲	۲	۲	۲	۲	۲	
جمع کل	۸	۸	۹	۹	۱۲	۱۲	۱۲	۱۲	۱۲	۱۲	۵۱٪
درصد سؤال‌های DIF	٪۲	٪۲	٪۲	٪۲	٪۳	٪۳	٪۳	٪۳	٪۳	٪۳	

جدول ۳ میزان شیوع کنش افتراقی سؤال را در بین استان اخذ دیپلم (زبان) در گروه‌های مختلف امتحانی به تفکیک سال‌های برگزاری آزمون نشان می‌دهد. تعداد سؤال‌های DIF مرتبط با استان در آزمون‌های مورد مطالعه در سال‌های ۸۷ تا ۸۹ کمتر از ۵ درصد و در سال ۱۳۹۰ کمتر از ۱۰ درصد است، درحالی که در این مؤلفه مورد بررسی نیز هرساله سؤال‌های DIF در دروس مورد مطالعه افزایش یافته‌اند. البته نرخ DIF مرتبط با زبان در سؤال‌های آزمون‌های مورد مطالعه نسبت به جنسیت بسیار کمتر است. درصد کنش افتراقی سؤال‌های آزمون به نسبت رشته و استان (زبان تکلم) هم نشان می‌دهد که بیشترین میزان کنش افتراقی سؤال‌های مورد مطالعه در رشته علوم انسانی با ۰.۵٪ و بعد از آن رشته‌های علوم تجربی (۰.۴٪)، ریاضی (۰.۳٪)، هنر (۰.۳٪) و زبان (۰.۱٪) قرار دارند. میانگین کنش افتراقی سؤال در دروس مورد مطالعه برحسب مؤلفه استان (زبان) در چهار سال متوالی معادل با ۰.۳٪ است.

نتایج آشکارسازی DIF برحسب مناطق (۱ و ۲)، به‌عنوان یکی از مؤلفه‌های عمده مبتنی بر پایگاه اجتماعی-اقتصادی که سازمان سنجش آموزش کشور گزینش را بر اساس آن انجام می‌دهد، در جدول ۴ ارائه شده است.

جدول ۴: نتایج تحلیل سؤال‌های گروه‌های آزمایشی برحسب سهمیه مناطق و گروه آزمایشی در

سال‌های ۱۳۸۷ تا ۱۳۹۰

گروه آزمایشی	سال		۱۳۸۷		۱۳۸۸		۱۳۸۹		۱۳۹۰	
	۱	۲	۱	۲	۱	۲	۱	۲	۱	۲
ریاضی	۶	۶	۸	۷	۹	۳	۷	۱۶	۶۲	۱۱٪
جمع	۱۲	۱۵	۱۲	۲۳						
علوم تجربی	۱	۳	۱	۲	۴	۰	۲	۸	۲۱	۶٪
جمع	۴	۳	۴	۱۰						
علوم انسانی	۲	۱	۲	۱	۰	۱	۲	۳	۱۲	۱۰٪
جمع	۳	۳	۳	۵						
هنر	۴	۰	۲	۱	۱	۰	۴	۲	۱۴	۵٪
جمع	۴	۳	۴	۶						
زبان	۰	۱	۰	۲	۵	۰	۵	۲	۱۵	۵٪
جمع	۱	۲	۱	۷						
جمع کل	۲۴	۲۶	۲۳	۵۱					۱۲۴	۸٪
درصد DIF در سؤال‌های آزمون	۶٪	۷٪	۶٪	۱۳٪						

همان‌طور که در جدول ۴ مشاهده می‌شود، تعداد سؤال‌های DIF مرتبط با منطقه در آزمون‌های مورد مطالعه در سال‌های ۸۷ تا ۸۹ کمتر از ۱۰ درصد و در سال ۱۳۹۰ کمتر از ۱۵ درصد است. در این مؤلفه نیز افزایش سؤال‌های DIF علی‌الخصوص در سال ۱۳۹۰ در درس مورد مطالعه مشهود است. درصد کنش افتراقی سؤال‌های آزمون به نسبت سهمیه مناطق هم نشان می‌دهد که بیشترین میزان کنش افتراقی سؤال‌های آزمون در رشته ریاضی با ۱۱٪ و بعد از آن رشته‌های علوم انسانی (۱۰٪)، علوم تجربی (۶٪)، هنر (۵٪) و زبان (۵٪) قرار دارند. میانگین کنش افتراقی سؤال‌های مورد مطالعه برحسب مؤلفه منطقه (پایگاه

اجتماعی-اقتصادی) در چهار سال متوالی معادل با ۸٪ است. همچنین همان‌طور که ملاحظه می‌شود نرخ DIF در مناطق تفاوت مشهودی ندارد.

نتایج تحلیل سؤال‌های آزمون‌های مورد مطالعه در گروه‌های مختلف آزمایشی کنکور سراسری برحسب ویژگی‌های جمعیت شناختی در سال‌های ۱۳۸۷ تا ۱۳۹۰ در جدول ۵ ملاحظه می‌شود.

جدول ۵. نتایج تحلیل سؤال‌های آزمون‌های مورد مطالعه در گروه‌های امتحانی برحسب ویژگی‌های جمعیت شناختی در سال‌های ۱۳۸۷ تا ۱۳۹۰

گروه امتحانی	مؤلفه	سال				درصد
		۱۳۸۷	۱۳۸۸	۱۳۸۹	۱۳۹۰	
گروه ریاضی	جنسیت	۱۳	۱۸	۲۷	۲۵	۸۳٪
	منطقه	۱۱	۱۵	۱۲	۲۲	۶۰٪
	استان	۳	۴	۳	۸	۱۸٪
گروه علوم تجربی	جنسیت	۱۱	۱۰	۱۳	۱۶	۵۰٪
	منطقه	۴	۳	۴	۱۰	۲۱٪
	استان	۳	۲	۳	۷	۱۵٪
گروه علوم انسانی	جنسیت	۳	۵	۵	۳	۱۶٪
	منطقه	۳	۳	۱	۵	۱۲٪
	استان	۱	۲	۱	۲	۶٪
گروه هنر	جنسیت	۴	۵	۳	۱۲	۲۴٪
	منطقه	۴	۳	۲	۶	۱۵٪
	استان	۰	۱	۳	۵	۹٪
گروه زبان	جنسیت	۵	۸	۸	۱۱	۳۲٪
	منطقه	۱	۲	۵	۷	۱۵٪
	استان	۱	۱	۲	۰	۴٪

جدول ۵ نشان از افزایش میزان کنش افتراقی در دروس مورد مطالعه در بیشتر گروه‌های آزمایشی در گذر زمان دارد، به طوری که در سال ۱۳۸۹ و ۱۳۹۰ تعداد سؤال‌های بیشتری دارای کنش افتراقی هستند. نتایج تحلیل سؤال‌های گروه ریاضی از سال ۱۳۸۷ تا

۱۳۹۰ نشان می‌دهد که بیشترین میزان سوگیری آماری سؤال‌ها در این دروس به ترتیب در جنسیت با ۱۵٪، سهمیه مناطق با ۱۱٪ و استان با ۴٪ است. نتایج تحلیل سؤال‌های گروه علوم تجربی هم از سال ۱۳۸۷ تا ۱۳۹۰ نشان داد که بیشترین میزان DIF سؤال‌ها در دروس مورد مطالعه به ترتیب در مؤلفه- های جنسیت با ۱۵٪، سهمیه مناطق با ۶٪ و استان با ۵٪ هستند. در گروه علوم انسانی به‌طور کلی بیشترین میزان کنش افتراقی سؤال بر حسب جنسیت با ۱۳٪، سپس منطقه با ۱۰٪ و سرانجام در استان با ۵٪ بوده است. نتایج تحلیل رگرسیون لجستیک جهت آشکارسازی DIF در رشته هنر تقریباً شبیه به سایر گروه‌های آزمایشی است. به‌طور کلی در دروس مورد مطالعه گروه هنر، بیشترین میزان کنش افتراقی در متغیر جنسیت با ۹٪، سپس منطقه با ۵٪ و استان با ۳٪ است. بیشترین میزان کنش افتراقی در دروس مورد مطالعه گروه زبان در متغیر جنسیت با ۱۱٪، سپس منطقه با ۵٪ و استان با ۱٫۴٪ است.

سرانجام، مقادیر اندازه‌های اثر برای مقادیر DIF هماهنگی که معنی‌دار بودند، شامل ضریب تبیین (R^2) ناگلاکرک بود که مقادیر آن‌ها برای همه سؤال‌های مورد مطالعه کمتر از ۰/۰۱ بودند. بنا بر نظر زومبو و توماس^۱ (۱۹۹۷) مقادیر اندازه‌های اثر حدود ۰/۱۳ در حد متوسط و مقادیر نزدیک به صفر ناچیز محسوب می‌شوند، بنابراین می‌توان اعلام کرد که مقادیر اندازه‌های اثر DIF معنی‌دار در سؤال‌های مورد مطالعه در پژوهش حاضر بسیار ناچیز هستند.

بحث

هدف از پژوهش حاضر، بررسی کنش افتراقی سؤال در آزمون‌های سراسری سازمان سنجش آموزش کشور بر حسب ویژگی‌های جمعیت شناختی بود. یافته‌های پژوهش نشان داد که سؤال‌های آزمون‌های مورد مطالعه در هر سه متغیر جمعیت شناختی مورد مطالعه دارای DIF هستند، اما درصد سؤال‌های DIF کم هستند، به طوری که بیشترین نرخ DIF مشاهده شده کمتر از ۲۰ درصد است، این در حالی است که اندازه‌های اثر محاسبه شده نیز

1. Thomas

بسیار ناچیز ($R^2 < 0.01$) هستند. بنابراین شانس وجود قطعی سؤال‌های سودار در کنکورهای سراسری بر اساس قضاوت نهایی ممکن است بسیار کم باشد. همچنین یافته‌های پژوهش نشان داد که سهم DIF مرتبط با جنسیت نسبت به سایر مؤلفه‌های منطقه (پایگاه اجتماعی-اقتصادی) و استان (زبان) بیشتر است. این یافته نشان می‌دهد که در هر ۵ رشته امتحانی به ترتیب بیشترین میزان کنش افتراقی مرتبط با جنسیت، منطقه و زبان است. همچنین یافته‌های پژوهش نشان داد که در میان گروه‌های آزمایشی، دروس مورد مطالعه گروه ریاضی دارای بیشترین سؤال‌های DIF و گروه زبان دارای کمترین نرخ سؤال‌های DIF است که به نوعی نشانگر دقت بیشتر طراحی سؤال‌ها در گروه زبان است. آماره‌های به دست آمده نشان می‌دهد، به طور میانگین در طی ۴ سال متوالی در تمام رشته‌های امتحانی ۱۳٪ سؤال‌های کنکور سراسری از لحاظ متغیر جنسیت دارای کنش افتراقی هستند، اما این آماره در رابطه با متغیر استان ۳٪ و در متغیر منطقه ۸٪ است. سرانجام یافته‌های پژوهش نشان داد که نرخ سؤال‌های DIF در آزمون‌های مورد مطالعه از سال ۱۳۸۷ تا ۱۳۹۰ نرخ صعودی دارد. افزایش نرخ DIF مخصوصاً در آزمون‌های مورد مطالعه در سال ۱۳۹۰ مشهود است.

یافته‌های تحقیق حاضر از منظر ویژگی‌های جمعیت شناختی با مطالعات لوپس (۲۰۰۰) همسو است. یافته‌های حاصل شده در مطالعه متغیر جنسیت در کنش افتراقی سؤال‌های آزمون با تحقیق سانگک و همکاران (۲۰۱۵)، امیریان و همکاران (۲۰۱۴) و با بخشی از یافته‌های براتی و احمدی (۲۰۱۰)، عبدالعزیز (۲۰۱۰)، رضایی و شعبانی (۲۰۰۹)، دودین و انابی (۲۰۰۸) و تاکالا (۲۰۰۰) که تحقیقاتی را در رابطه با اثر جنسیت و نژاد بر کنش افتراقی سؤال انجام داده‌اند، و با یافته‌های بارنز و ولز (۲۰۰۹) مبنی بر تأثیر جنسیت و قومیت (نژاد) در کنش افتراقی در دانشگاه ماساچوست همسو است.

یافته دیگری از این تحقیق نشان می‌دهد که میزان کنش افتراقی در سؤال‌های آزمون به لحاظ متغیر استان (زبان) برخلاف ۲ متغیر دیگر در رشته‌های علوم انسانی و تجربی بیشتر از سایر رشته‌ها است که تأثیر سازه‌های زبان را می‌توان در آن دخیل دانست. این یافته تحقیق به نوعی با یافته‌های شاموگام و لان (۲۰۱۴)، یونگ و همکاران (۲۰۱۳) و لوک هاردینگ

(۲۰۱۲) به لحاظ تأثیر مؤلفه زبان در کنش افتراقی سؤال همسو است. سرانجام یافته‌های تحلیل اندازه اثر مربوط به مقادیر DIF نشان داد که مقادیر شدت سوگیری آماری بسیار ناچیز هستند، این یافته کاملاً با نتایج پژوهش امیریان و همکاران (۲۰۱۴) همخوانی دارد. به نظر می‌رسد در آزمون‌های سرنوشت‌ساز و بزرگ‌مقیاس رسیدن به مقادیر بزرگ اندازه اثر DIF احتمال کمی دارد، به‌رحال این ادعا هنوز مستلزم پژوهش بیشتری است. همچنین پژوهش حاضر DIF منظمی را نسبت به منطقه ۱ و ۲ نشان نداد و تقریباً نرخ DIF در سال‌های مختلف به‌صورت تناوبی در هر دو منطقه به یک‌میزان مشاهده شد.

دلایل احتمالی وقوع DIF برحسب ویژگی‌های جمعیت شناختی ممکن است در طراحی محتوای سؤال‌های کنکور سراسری نهفته باشد که باعث سوگیری نسبت به جنسیت، زبان و یا پایگاه اجتماعی و اقتصادی خاصی می‌شود. به‌طور مثال اگر پسران با محتوای سؤال خاصی (به‌طور مثال یک ورزش که بیشتر مختص پسران است) بیشتر از دختران آشنا باشند، درک آن سؤال برای آن‌ها بهتر خواهد بود و بالعکس دختران در درک محتوای آن سؤال مشکل خواهند داشت و یا اگر محتوای یک سؤال برای قومیتی با زبانی متفاوت مشکل‌تر از قوم دیگر باشد، این امر می‌تواند موجب سوگیری سؤال نسبت به آن قومیت شود. همچنین اگر محتوای سؤال شامل کلمات، فعالیت‌ها و یا تجاربی باشد که قشری با پایگاه اجتماعی و اقتصادی بالاتر بیشتر با آن آشنا باشند، این محتوا نسبت به اقشاری با پایگاه اجتماعی و اقتصادی پایین‌تر سوگیری دارد. البته از آنجایی که در پژوهش حاضر با در نظر گرفتن سال‌های برگزاری کنکور، نرخ DIF در دو منطقه موردبررسی تقریباً به یک‌میزان است، نمی‌توان به‌طور قطع در مورد نقش پایگاه اجتماعی و اقتصادی در این دو منطقه و سوگیری قطعی نسبت به یک منطقه بحث کرد. البته هنوز اختلاف این دو منطقه با منطقه ۳ می‌تواند موضوع جذابی برای پژوهش‌های آینده DIF و نقش مناطق محروم در این رابطه باشد. روی‌هم‌رفته همان‌طور که قبلاً نیز اشاره شد، قضاوت نهایی در مورد وجود چنین سوگیری‌هایی مستلزم تشکیل گروهی از متخصصان موضوع درسی و طراحی سؤال است که با در دست داشتن یافته‌های پژوهش حاضر بتوانند در مورد سوگیری سؤال‌های کنکور سراسری قضاوت نهایی داشته باشند. البته با توجه به یافته‌های

پژوهش حاضر، همچنان باید ناچیز بودن مقادیر اندازه اثر DIF در قضاوت نهایی برای اعلام سوگیری سؤال مدنظر قرار گرفته و با احتیاط بسیار زیاد حکم نهایی در مورد سوگیری سؤال صادر شود.

از جمله محدودیت‌های تحقیق حاضر را می‌توان در بررسی میزان کنش افتراقی بر اساس متغیرهای منطقه و استان دانست. از آنجایی که اطلاعاتی از آزمودنی‌ها در رابطه با زبان تکلم آن‌ها وجود نداشت؛ بنابراین آذربایجان به‌عنوان استانی متفاوت از اصفهان به لحاظ زبان تکلم در نظر گرفته شد، هرچند که در استان‌های ذکر شده نمی‌توان به‌طور کامل به این هدف دست یافت، اما در این مورد اکثریت و طیف در نظر گرفته شد. در مورد منطقه هم با توجه به تعداد متناسب شرکت‌کنندگان در منطقه ۱ و ۲، این دو منطقه به‌عنوان متغیر گروهی برای DIF در نظر گرفته شد، البته همان‌طور که اشاره شد پژوهش‌هایی نیز متناسب با تعداد شرکت‌کنندگان در منطقه ۳ و مقایسه آن با سایر مناطق مورد نیاز است.

بر اساس یافته‌های پژوهش حاضر پیشنهاد می‌شود که سؤال‌های دارای DIF در آزمون‌های مورد مطالعه در پانل‌های خبرگان و متخصصان موضوعی و محتوایی مورد مطالعه قرار گیرند و سؤال‌هایی که نسبت ویژگی‌های جمعیت شناختی مورد مطالعه سوگیری محتوایی دارند، مشخص شوند. بعلاوه عواملی که در طراحی سؤال نسبت به ویژگی‌های جمعیت شناختی باعث ایجاد چنین سوگیری‌هایی می‌شوند مشخص شده و در طراحی سؤال‌های کنکور سراسری مدنظر قرار گیرند. ضمناً برای مطالعه بیشتر نیز پیشنهاد می‌شود آزمون‌های دروس عمومی و سایر دروس تخصصی کنکور سراسری به لحاظ DIF برحسب ویژگی‌های جمعیت شناختی مورد بررسی و مقایسه قرار گیرند تا DIF در آزمون‌های سراسری به‌طور جامع‌تری مورد بررسی قرار گیرد.

منابع

- ایزانلو، ب. و حبیبی، م. (۱۳۸۹). شناسایی کارکرد متفاوت سؤال بر اساس نظریه سؤال-پاسخ: کاربرد الگوی تک پارامتری با استفاده از نرم‌افزار بای لوگ-ام جی. فصلنامه روان‌شناسی کاربردی، (۲) ۴، ۳۱-۲۰.
- خدایی، ابراهیم، (۱۳۸۸) الف. بررسی عوامل مؤثر بر قبولی در آزمون کارشناسی ارشد. فصلنامه پژوهش و برنامه‌ریزی در آموزش عالی، (۵۴) ۳۴-۱۹.
- خدایی، ابراهیم (۱۳۸۸) ب. بررسی رابطه سرمایه اقتصادی و فرهنگی والدین دانش‌آموزان با احتمال قبولی آن‌ها در آزمون سراسری سال تحصیلی ۱۳۸۵. فصلنامه انجمن آموزش عالی ایران، (۱) ۴، ۸۴-۶۵.
- گرامی پور، مسعود (۱۳۹۳). ارزیابی توان آماری تحلیل رگرسیون لجستیک در آشکارسازی کنش افتراقی سؤال‌های آزمون. فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، (۸) ۴، ۲۱۱-۱۸۷.
- نوربخش، سید مرتضی (۱۳۸۹). نقش سرمایه‌های فرهنگی، اجتماعی و اقتصادی خانواده در موفقیت داوطلبان آزمون سراسری. فصلنامه برنامه‌ریزی رفاه و توسعه اجتماعی، (۱۳۵) ۴-۹۴.
- همبلتون، رونالد ک؛ سوامیناتان، اچ؛ راجرز، اچ. جین (۱۹۹۱). مبانی نظریه پرسش-پاسخ. ترجمه محمدرضا فلسفی نژاد (۱۳۸۹). تهران: دانشگاه علامه طباطبایی.
- Abedalaziz, N. (2010). A gender-related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Assessment*, 5, 101- 116.
- Amirian, S. M. R., Alavi, S. M., & Fidalgo, A. M. (2014). Detecting Gender DIF with an English Proficiency Test in EFL Context. *Iranian Journal of Language Testing*, 4(2).
- Barati, H., & Ahmadi, A. R. (2010). Gender-based DIF across the Subject Area: A Study of the Iranian National University Entrance Exam. *Journal of Teaching Language Skills*, 2(3), 1-26.
- Barnes, B. J., & Wells, C. S. (2009). Differential item functional analysis by gender and race of the national doctoral program survey. *International Journal of Doctoral Studies*, 4, 77-96.

- Broer, M. (2005). Ensuring the fairness of GRE writing prompts: Assessing differential difficulty. *ETS Research Report Series*, 2005(1), i-41.
- Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, 13, 287-321.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications.
- Cole, N.S., Zieky, M.J. (2001). The New Faces of Fairness. *Journal of Educational Measurement*, 38, no. 4: 369-382.
- Dinis da Costa, P., & Araújo, L. (2012). Differential Item Functioning (DIF): What Functions Differently for Immigrant Students in PISA 2009 Reading Items?. Luxembourg: *Publications Office of the European Union, Report EUR 25565 EN*.
- Gomez-Benito, J., & Navas-Ara, M. J. (2000). A comparison of χ^2 , RFA and IRT based procedures in the detection of DIF. *Quality and Quantity*, 34(1), 17-31.
- Hambleton, R.K., Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer-Nijhoff, Boston.
- Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, 29(2), 163-180. DOI:10.1177/0265532211421161
- Le, V. N. (1999). *Identifying Differential Item Functioning on the NELS: 88 History Achievement Test*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.
- Lewis, Anne (April 2000). *High-stakes testing: Trends and issues*. Mid Continent Research for Education and Learning.
- Mazor, K. M., Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Applied Psychological Measurement*. Vol 22(4), 357-367.
- Doudeen, H., & Annabi, H. (2008). Sex-Related Differential Item Functioning (DIF) Analysis of TIMSS. *Educational Sciences*, 35, p697.
- Millsap, R.E (2011). *Statistical Approaches to Measurement Invariance*. New York:NY, Routledge
- Ogretmen, T. (2015). DIF analysis across genders for reading comprehension part of English language achievement exam as a foreign language. *Educational Research and Reviews*, 10(11), 1505-1513.
- Owen, K. (1992b). *Test-item bias: Methods, findings and recommendations*. Pretoria: Human Sciences Research Council.

- Popham, W.J. (2005). HIGH-STAKES TESTS: HARMFUL, PERMANENT, FIXABLE, *American Educational Research Journal*, 6, p85.
- Rezaee, A., & Shabani, E. (2009). Gender Differential Item Functioning Analysis of the University of Tehran English Proficiency Test. *Pazhuhesh-e Zabanha-ye Khareji*, 56, 89-108.
- Rezaee, A., & Shabani, E. (2010). Gender differential item functioning analysis of the University of Tehran English Proficiency Test. *Pazhuhesh-e Zabanha-ye Khareji*, 56, 89, 108.
- Shanmugam, S. K. S., & Lan, O. S. (2014). THE VALIDITY OF ADMINISTERING BILINGUAL MATHEMATICS TEST AMONG MALAYSIAN BILINGUAL STUDENTS USING DIFFERENTIAL ITEM FUNCTIONING (DIF). *Asia Pacific Journal of Educators and Education*, 29, 1-18.
- Shuttleworth-Jordan, A.B. (1996). On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology*, 26 (2), 96- 102.
- Shealy, R.T., Stout, W.F. (1993). A model based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika* 58, 197-239.
- Song, X., Cheng, L., & Klinger, D. DIF investigations across groups of gender and academic background in a large-scale high-stakes language test. *Language Testing and Assessment* 4, 1, p97.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361- 370.
- Takala, S., & Kaftandjieva, F. (2000). Test fairness: A DIF analysis of an L2 vocabulary test. *Language Testing*, 17(3), 323-340.
- Vanderberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5, 139-158.
- Van de Vijver, F. & Poortinga, Y.H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13 (1), 29-37.
- Van Zyl, E. & Visser, D. (1998). Differential item functioning in the figure classification test. *Journal of Industrial Psychology*, 24 (2), 25-33.
- Young, J. W., Morgan, R., Rybinski, P., Steinberg, J., & Wang, Y. (2013). Assessing the Test Information Function and Differential Item Functioning for the TOEFL Junior® Standard Test. *ETS Research Report Series*, 2013(1), i-27
- Zumbo, B. D. and Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada:

University of Northern British Columbia, Edgeworth Laboratory for
Quantitative Behavioral Science.

Archive of SID