

# تأثیر محل سکونت داوطلب بر کارکرد افتراقی سؤال‌های ریاضی گروه آزمایشی ریاضی و فنی کنکور سراسری ۱۳۹۶

■ علی مقدم زاده\*\*\*

■ اصغر مینائی\*\*

■ مولود علی میرزایی\*

## چکیده:

به دلیل تفاوت‌های اجتماعی، اقتصادی، فرهنگی و زبانی در استان‌های مختلف کشور، سؤال‌های آزمون سراسری ورودی دانشگاه‌ها، ممکن است برای همه استان‌ها هم‌ارز نباشد، که مقایسه‌ها را غیرمنصفانه می‌کند. هدف این مطالعه بررسی تأثیر استان محل سکونت آزمودنی به‌عنوان منبعی برای کارکرد افتراقی سؤال است. به این منظور داده‌های ماده امتحانی ریاضی گروه آزمایشی علوم ریاضی و فنی در آزمون سراسری ۱۳۹۶ مورد استفاده قرار گرفت. تقریباً نمونه‌ای به حجم ۱۰۰۰ نفر از هر استان انتخاب شد. تحلیل عاملی چندسطحی آمیخته، استان‌ها را در ۳ طبقه همگن قرار داد که نشان می‌دهد آزمودنی‌ها در استان‌ها الگوی پاسخ مشابه ندارند و بنابراین محل سکونت آزمودنی یکی از منابع کارکرد افتراقی سؤال است. مقایسه تعقیبی طبقات همگن آزمودنی‌ها در مدل چندسطحی آمیخته نشان داد که گروه‌های سنی نیز یکی از عوامل مؤثر بر عملکرد متفاوت آزمودنی‌ها در استان‌ها است. در این پژوهش، نسبت مردان و زنان در طبقه‌های مکنون یکسان است، بنابراین جنسیت آزمودنی نمی‌تواند از دلایل کارکرد افتراقی سؤال باشد. بیان اینکه آزمودنی‌ها در یک گروه (مثلاً مردان)، به سؤال دارای کارکرد افتراقی متفاوت پاسخ می‌دهند، منجر به دید کلیشه‌ای در مورد سؤال می‌شود که به نفع همه اعضای گروه (مردان) است. بنابراین، اگر ابتدا طبقه‌های مکنون تشکیل شود و سپس تفاوت ویژگی آزمودنی‌ها در طبقه‌ها مورد بررسی قرار گیرد، منجر به تحلیل آموزنده‌تری از کارکرد افتراقی می‌شود.

عدالت آموزشی، منابع کارکرد افتراقی سؤال، تحلیل عاملی چندسطحی آمیخته

کلید واژه‌ها:

□ تاریخ پذیرش مقاله: ۹۹/۱۲/۱۹

□ تاریخ شروع بررسی: ۹۹/۹/۵

□ تاریخ دریافت مقاله: ۹۹/۴/۲۸

\* دکترای سنجش و اندازه‌گیری، اصفهان، ایران ..... moloud\_alm@yahoo.com  
\*\* دانشیار گروه سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران (نویسنده مسئول) ..... asghar.minaei@yahoo.com  
\*\*\* استادیار گروه روش‌ها و برنامه‌های آموزشی و درسی دانشگاه تهران، تهران، ایران ..... amoghadamzadeh@ut.ac.ir

## مقدمه

یکی از سیاست‌های اعمال شده در زمینه جذب دانشجو، توجه به سهمیه‌بندی برای ایجاد عدالت آموزشی و رفع نیاز مناطق محروم کشور به نیروی انسانی متخصص است. از این رو، جذب دانشجو در مراکز آموزش عالی تنها بر اساس نمره عملی در آزمون صورت نمی‌گیرد، بلکه توجه به محل سکونت داوطلب نیز حائز اهمیت است. بدین ترتیب، سیاست‌گذاران آموزش عالی با رویکرد عدالت‌محوری و توجه به مناطق محروم کشور و با این استدلال که شرکت‌کنندگان در کنکور، امکانات آموزشی یکسان نداشته‌اند تا بتوانند در یک آزمون یکسان به رقابت بپردازند، طریقه سهمیه‌بندی را پیش گرفتند (صادقی جعفری، روشن و شکوری گنجوی، ۱۳۹۰). اما رودباری و فیض‌آبادی فراهانی (۱۳۸۴) بیان می‌کنند که با وجود اعمال سهمیه منطقه ۳ در کلیه بخش‌های استان سیستان و بلوچستان و سهمیه ۲ برای زاهدان مرکزی، در آزمون‌های سراسری ورودی دانشگاه‌ها در سال‌های اخیر و امکان استفاده از سهمیه بومی ۷۰٪ و همچنین سهمیه استان‌های محروم برای دانش‌آموزان، استان سیستان و بلوچستان پایین‌ترین درصد قبولی را بین استان‌های کشور دارد. سهمیه‌بندی و ناحیه و قطب‌بندی نیز نتوانسته است نیازهای این استان را به نیروهای متخصص بومی به‌طور کامل پاسخ دهد.

بنابراین، از آنجاکه دغدغه اصلی آزمون‌سازی این است که آیا استنباط‌های حاصل از نمرات آزمون در همه زیرجامعه‌های مختلف قابل‌مقایسه است، تحلیل‌های روان‌سنجی مختلفی به‌طور معمول انجام می‌شود تا بسنجد که آیا وجوه یک برنامه آزمون‌سازی ممکن است برای برخی از زیرجامعه‌های خاص، غیرمنصفانه باشد. برای مطالعه مقایسه‌پذیری نمره‌های آزمون‌ها، رویکردها و روش‌های متعددی (برای مثال ویلینگام، راگوستا، بینت، براون، راک و پاورز<sup>۱</sup>، ۱۹۸۸؛ یانگ، استینبرگ، کلاین، استون، مارتینلو، لینگ و چاو<sup>۲</sup>، ۲۰۱۰) ارائه شده است که از جمله پرکاربردترین آن‌ها می‌توان به رویکرد کارکرد افتراقی سؤال<sup>۳</sup> و به تبع آن شناسایی عوامل مرتبط و ایجادکننده آن اشاره نمود.

روش‌های تشخیص کارکرد افتراقی سؤال به‌طور بالقوه دارای این مفروضه هستند که اعضای درون گروه‌های کانونی و مرجع (که از متغیرهای آشکار مثل جنسیت تشکیل شده‌اند) همگن هستند که در این صورت تفاوت‌های درون‌گروهی نادیده گرفته می‌شود. مفهوم‌سازی<sup>۴</sup> دیگری از کارکرد افتراقی سؤال این است که گروه‌های آشکار (مثلاً زنان و مردان) به نسبت‌های مختلف در طبقه‌های مکنون از جامعه آزمودنی‌ها قرار دارند. دی ایالا، کیم، استپلتون و دایتون<sup>۵</sup> (۲۰۰۲) یک مطالعه شبیه‌سازی مونت کارلو با استفاده از شش روش شناسایی کارکرد افتراقی سؤال (نسبت درست‌نمایی، رگرسیون لوجستیک، کای اسکور لرد، روش ناپارامتری کای اسکور منتال هنزل<sup>۶</sup>، روش ناحیه علامت‌گذاری شده دقیق<sup>۷</sup> و روش آماره Q) برای بررسی این مفهوم‌سازی انجام دادند. نتایج نشان داد که هرچه نسبت متغیرهای آشکار در طبقه‌های مکنون شبیه‌تر باشد (مثلاً نسبت یکسانی از زنان و مردان در طبقه‌های مکنون وجود داشته باشد)، تشخیص کارکرد افتراقی سؤال برای تمامی شش روش سنتی شناسایی کارکرد افتراقی سؤال

دشوارتر می‌شود. همچنین، نتایج مطالعه شبیه‌سازی ساموئلسن<sup>۸</sup> (۲۰۰۵)، برای برجسته نمودن مسائل ناشی از به‌کارگیری روش‌های استاندارد تشخیص کارکرد افتراقی، نشان داد با کاهش میزان هم‌پوشانی بین گروه‌های آشکار و طبقه‌های مکنون، قدرت شناسایی صحیح سؤال‌های دارای کارکرد افتراقی کاهش می‌یابد.

کوهن و بالت<sup>۹</sup> (۲۰۰۵) دو مطالعه طراحی نمودند، در مطالعه اول، دو نمونه از آزمودنی‌ها انتخاب شد، در نمونه اول آزمون نسبت درست‌نمایی برای شناسایی سؤال‌های دارای کارکرد افتراقی نسبت به جنسیت مورد استفاده قرار گرفت، سپس از نتایج آن برای برآورد عضویت در دو طبقه مکنون در آزمودنی‌های نمونه دوم استفاده شد. اگر جنسیت آزمودنی دلیل کارکرد افتراقی باشد، لازم است طبقه مکنون اول شامل مردان و طبقه مکنون دوم شامل زنان باشد. ولی نتایج نشان داد که تقریباً ۵۰ درصد کل آزمودنی‌ها به طبقه‌ای مشابه با جنسیت خود تعلق یافته‌اند. بنابراین، آن سؤال‌هایی که در مرحله اول دارای کارکرد افتراقی علیه مردان (زنان) شناسایی شد، علیه تمامی مردان (زنان) نیست. همچنین، بررسی تفاوت سنی دو طبقه مکنون نشان می‌دهد آزمودنی‌های طبقه اول یک سال کوچک‌تر از طبقه دوم هستند، بنابراین در مطالعه دوم نشان داده می‌شود که اگر ابتدا آزمودنی‌ها در طبقات مکنون قرار گیرند و سپس تفاوت ویژگی آزمودنی‌ها در طبقه‌های مکنون مورد بررسی قرار گیرد، منجر به تحلیل آموزنده‌تری<sup>۱۰</sup> از کارکرد افتراقی می‌شود.

وب، کوهن و اشوانفلوگل<sup>۱۱</sup> (۲۰۰۸) داده‌های دو آزمون واژگان انگلیسی را تحلیل کردند و به این نتیجه رسیدند که دو طبقه مکنون از آزمودنی‌ها وجود دارد و هر طبقه، نسبت متفاوتی از آمریکایی‌های اروپایی تبار و آمریکایی‌های افریقایی تبار را شامل می‌شود. این یافته نشان می‌دهد که قومیت و نژاد می‌تواند عامل کارکرد افتراقی سؤال باشد. در پژوهش بالت، کوهن و وولاک<sup>۱۲</sup> (۲۰۰۱) به آزمودنی‌ها تعدادی جمله که خطا داشت، ارائه و از آن‌ها خواسته شد تا علت خطا را شناسایی کنند. با استفاده از مدل نظریه سؤال پاسخ آمیخته، دو طبقه مکنون همگن از آزمودنی‌ها شناسایی شد که از لحاظ الگوی انتخاب (انتخاب اینکه آیا گرامر یا نشانه‌گذاری به خطا منجر شده است) متفاوت بودند. در مطالعه اولیوری، ارسی کان و زامبو<sup>۱۳</sup> (۲۰۱۳) کارکرد افتراقی سؤال و منابع آن با استفاده از روش مدل‌سازی کلاس مکنون بررسی شد. نتایج نشان داد الگوهای پاسخ افتراقی که با طبقه‌های مکنون شناسایی شده است، با سطح پیشرفت تحصیلی دانش‌آموزان و متغیرهای مرتبط با معلم ارتباط بیشتر و با متغیرهایی مثل جنسیت، زبان آزمودنی و کشور که معمولاً در مطالعات مقایسه‌پذیری مقایسه می‌شود، ارتباط کمتری دارد.

روش‌هایی که برای مطالعه کارکرد افتراقی سؤال‌ها به کار می‌رود، بر مفروضه همگنی درون‌گروهی متکی است. مطالعات نشان می‌دهد در صورت برقرار نبودن مفروضه همگنی درون‌گروهی، بهتر است از رویکرد مدل‌های طبقه مکنون استفاده کرد. در این رویکرد، با توجه به پاسخ آزمودنی‌ها به سؤال‌ها،

هر آزمودنی به یکی از طبقه‌های مکنون اختصاص داده می‌شود. آزمودنی‌هایی که در یک طبقه مکنون قرار می‌گیرند، آزمودنی‌هایی با الگوی پاسخ همگن محسوب می‌شوند (میسلوی، ویلسون، ارسی کان و کدوسکی<sup>۱۴</sup>، ۲۰۰۲). برخلاف روش‌های سنتی شناسایی کارکرد افتراقی سؤال که بر متغیرهای آشکار استوارند و همه افراد موجود در یک گروه آشکار (مثلاً زنان) دارای الگوی پاسخ همگن در نظر گرفته می‌شوند، در روش‌های طبقه مکنون، گروه‌های مورد مقایسه برای کارکرد افتراقی سؤال، به صورت گروه زنان در مقابل گروه مردان تعریف نمی‌شوند، بلکه هر گروه مرکب از زنان و مردانی است که الگوی پاسخ مشابه دارند (اولیوری و همکاران، ۲۰۱۳). مدل عاملی آمیخته<sup>۱۵</sup>، ترکیبی از مدل طبقه مکنون و مدل عامل مشترک است که یک متغیر مکنون ترتیبی و یک یا چند متغیر مکنون پیوسته دارد و به عنوان ابزاری برای کشف ناهمگنی در جامعه مورداستفاده قرار می‌گیرد (لوبک و موتن<sup>۱۶</sup>، ۲۰۰۵).

در پژوهش‌های صورت گرفته در زمینه کارکرد افتراقی سؤال‌های آزمون ورودی دانشگاه‌ها، تنها دو گروه که اکثراً گروه زنان و مردان بودند، بدون در نظر گرفتن ناهمگنی جامعه، انتخاب و مورد مطالعه قرار گرفته‌اند. برای مثال، شریفی یگانه (۱۳۹۱)، بدون توجه به ناهمگنی جامعه، اثر جنسیت را بر پاسخ‌گویی به سؤال‌های آزمون ریاضی سال ۱۳۸۲ بررسی کرد و نشان داد سؤال‌هایی که به نفع دختران عمل می‌کنند، بیشتر مربوط به توابع و معادلات و سؤال‌هایی که به نفع پسران عمل می‌کنند، بیشتر مربوط به مثلثات، هندسه و احتمال است. گرامی‌پور، رضایی، رضانی صدر و نوروزی (۱۳۹۵)، نیز بدون در نظر گرفتن ناهمگنی جامعه، کارکرد افتراقی سؤال‌های برخی دروس تخصصی رشته‌های مختلف آزمون سراسری سال‌های ۱۳۸۷ تا ۱۳۹۰ را مطالعه کردند و نشان دادند که سهم جنسیت در کارکرد افتراقی سؤال، بیشتر از پایگاه اجتماعی اقتصادی و زبان داوطلب است.

در پژوهش‌هایی که ذکر شد، ناهمگنی الگوی پاسخ آزمودنی‌ها در نظر گرفته نشده است. در حقیقت، متغیرهای آشکار که برای تمایز گروه‌ها استفاده می‌شود (مثل جنسیت)، ممکن است به فهم اینکه سؤال به نفع یا زیان چه کسانی عمل می‌کند، کمک زیادی نکند. بنابراین، در پژوهش حاضر، به جای استفاده از متغیرهای آشکار در مطالعه کارکرد افتراقی سؤال، تأثیر ناهمگنی جامعه آزمودنی‌ها، به عنوان علت اصلی کارکرد افتراقی در بین بیش از دو گروه، مطالعه شد. هنگامی که گروه‌های مورد مقایسه زیاد است، مدل تحلیل عاملی چندسطحی آمیخته<sup>۱۷</sup> مناسب است. همانند تحلیل عاملی تأییدی چندگروهی، تعداد زیادی از گروه‌ها را می‌توان به راحتی با تحلیل عاملی چندسطحی آمیخته مطالعه نمود (کیم، کاو، وانگ و نووین<sup>۱۸</sup>، ۲۰۱۷). در این مدل، افراد در طبقه‌های مکنون قرار نمی‌گیرند، بلکه طبقه‌های مکنون از گروه‌های (مثلاً استان‌ها) همگن ایجاد می‌شود. یعنی استان‌هایی که شامل افرادی با الگوی پاسخ همگن هستند، در یک طبقه مکنون قرار می‌گیرند. بنابراین، مدل‌سازی عاملی آمیخته چندسطحی، افرادی را که درون گروه‌ها آشیا نه کرده‌اند در بر دارد و طبقه‌های مکنون برای گروه‌ها مشخص می‌شود. این مدل، اطلاعاتی در مورد چگونگی خوشه‌بندی

گروه‌ها فراهم می‌نماید. همچنین این روش، عضویت طبقه‌ای برای هر گروه را بر مبنای احتمال تعلق گروه به یک طبقه خاص مشخص می‌نماید. این روش اجازه می‌دهد منابع عدم تغییر ناپذیری به وسیله مدل‌بندی ارتباط بین متغیرهای کمکی مشاهده‌شده بالقوه و عضویت طبقه‌ای، بررسی شود. برای مثال، اگر طبقه‌های مکنون به لحاظ جغرافیایی متمایز باشند، متغیر منطقه می‌تواند در مدل پیش‌بینی عضویت طبقه‌ای قرار بگیرد. برای توصیف بهتر طبقه‌های مکنون ناشناخته، می‌توان از چندین متغیر کمکی استفاده نمود. مدل‌های عاملی آمیخته به داده‌ها برازش داده شد تا محتوای طبقه‌هایی از آزمودنی‌ها شناسایی شود، به طوری که آزمون برای آن‌ها بیشترین عملکرد افتراقی را داشته باشد. سپس، ویژگی‌های آزمودنی‌هایی که در این گروه‌های مکنون طبقه‌بندی شده‌اند، بررسی می‌شود تا مشخص شود که آیا وقوع کارکرد افتراقی سؤال با ویژگی ظاهری خاصی از آزمودنی‌ها (مثل سن، جنسیت) در ارتباط است. بنابراین، اهداف پژوهش به شرح زیر است:

۱. طبقه‌بندی استان‌ها در گروه‌های همگن با استفاده از مدل عاملی چندسطحی آمیخته؛
۲. شناسایی سؤال‌های دارای کارکرد افتراقی در میان استان‌ها با استفاده از گروه‌های همگن؛
۳. مقایسه طبقه‌ها بر مبنای سن و معدل کتبی دیپلم به‌عنوان عوامل مؤثر بر کارکرد افتراقی سؤال‌ها؛
۴. مقایسه طبقه‌های همگن از استان‌ها بر مبنای متغیر جنسیت به‌عنوان منبعی برای کارکرد افتراقی سؤال.

## روش

پژوهش حاضر یک مطالعه غیرآزمایشی روان‌سنجی (پومس، اسکوار و یوانو<sup>۱۹</sup>، ۲۰۱۴) است که کارکرد افتراقی سؤال‌های ریاضی آزمون اختصاصی ورودی دانشگاه گروه آزمایشی علوم ریاضی و فنی آزمون سراسری ۱۳۹۶ را در میان ۳۱ استان کشور مورد آزمون قرار داد. این ماده امتحانی ۵۵ سؤال چهارگزینه‌ای دارد.

همه شرکت‌کنندگان گروه آزمایشی علوم ریاضی و فنی در آزمون سراسری سال ۱۳۹۶، جامعه این پژوهش را تشکیل می‌دهند. از آنجاکه سازمان سنجش آموزش داده‌های تمام داوطلبان را در اختیار پژوهشگران نگذاشت، بنابراین با در نظر گرفتن ادبیات پژوهش و توصیه‌های پژوهشگران (برای مثال ایزنلو، بازگان، فرزاد، صادقی و کاوسی، ۱۳۹۳) نمونه‌ای به حجم تقریباً ۱۰۰۰ نفر از هر استان به شکل تصادفی ساده انتخاب شد. در مجموع، داده‌های مربوط به ۳۰۴۷۷ (۹۸۴۸ نفر زن و ۲۰۶۲۹ نفر مرد) مورد تحلیل قرار گرفت.

پس از دریافت داده‌ها، ابتدا پاسخ‌های داوطلبان بر اساس کلید پاسخ به‌صورت دو ارزشی نمره‌گذاری شد؛ به پاسخ درست نمره ۱ و پاسخ نادرست نمره ۰ داده شد. تک‌بعدی بودن آزمون

با استفاده از روش DIMTEST با استفاده از نرم‌افزارهای DIMPACK (استوت، فرولیچ و گاوو<sup>۲۰</sup>، ۲۰۰۱) و از رویکرد تحلیل عاملی غیرخطی در نرم‌افزار NOHARM<sup>۴</sup> (فراسر و مک‌دونالد<sup>۲۱</sup>، ۱۹۸۸) مطالعه شد. برای اجرای مدل‌های عاملی آمیخته چندسطحی، برای طبقه‌بندی استان‌ها در گروه‌های همگن برای بررسی کارکرد افتراقی سؤال، از نرم‌افزار Mplus نسخه ۷/۴ (موتن و موتن، ۲۰۱۷-۱۹۹۸) استفاده شد.

## نتایج

در این پژوهش، ۳۲ درصد از افراد گروه نمونه، زن و ۶۸ درصد مرد است. دامنه سنی افراد نمونه، ۱۶ تا ۵۹ سال با میانگین ۱۹/۰۶ و انحراف استاندارد ۱/۸۸ است. معدل کتبی دیپلم گروه نمونه از صفر تا ۲۰ با میانگین ۱۵/۲۱ و انحراف استاندارد ۳/۲۶ است. همسانی درونی سؤال‌های آزمون با استفاده از آلفای کرونباخ<sup>۲۲</sup> برابر ۰/۹۴۸ به دست آمد. جدول ۱ نشان‌دهنده ضریب پایایی<sup>۲۳</sup> سؤال‌ها در هر عامل به تفکیک استان‌ها در آزمون ریاضی است. مقدار بیشتر از ۰/۷ برای آلفای کرونباخ، نشان‌دهنده اعتبار قابل قبول است (محمدبیگی، محمدصالحی، و علی گل، ۱۳۹۳)، بنابراین آزمون ریاضی در تمام استان‌ها اعتبار قابل قبولی دارد.

جدول ۱. ضریب اعتبار هر عامل به تفکیک سؤال‌ها در آزمون ریاضی

استان‌ها	اعتبار	استان‌ها	اعتبار
آذربایجان شرقی	۰/۹۳۸	فارس	۰/۹۴۶
آذربایجان غربی	۰/۹۲۲	قزوین	۰/۹۱۳
اردبیل	۰/۹۱۰	قم	۰/۹۲۲
اصفهان	۰/۹۴۲	کردستان	۰/۹۰
ایلام	۰/۹۰۹	کرمان	۰/۹۲۵
بوشهر	۰/۹۱۵	کرمانشاه	۰/۹۰۹
تهران	۰/۹۳۳	کهگیلویه و بویراحمد	۰/۸۸۹
چهارمحال و بختیاری	۰/۹۱۵	گلستان	۰/۹۱۵

جدول ۱. (ادامه)

استان‌ها	اعتبار	استان‌ها	اعتبار
خراسان جنوبی	۰/۹۳۵	گیلان	۰/۹۱۴
خراسان رضوی	۰/۹۳۷	لرستان	۰/۹۰۸
خراسان شمالی	۰/۹۲۱	مازندران	۰/۹۳۳
خوزستان	۰/۹۲۷	مرکزی	۰/۹۱۸
زنجان	۰/۹۲۷	هرمزگان	۰/۹۱۷
سمنان	۰/۹۱۵	همدان	۰/۹۱۷
سیستان و بلوچستان	۰/۹۰۱	یزد	۰/۹۲۳

پیش‌فرض روش‌های مطالعه کارکرد افتراقی سؤال، تک‌بعدی بودن داده‌ها است (مازور، همبلتون و کیزر<sup>۲۴</sup>، ۱۹۹۸). ابتدا تک‌بعدی بودن داده‌ها با استفاده از روش DIMTEST که از طریق نرم‌افزار DIMPACK قابل اجراست مورد آزمون قرار گرفت. مقدار آماره آزمون T برابر ۱/۱۴۹ با سطح معناداری ۰/۱۲۵ به دست آمد که حاکی از تک‌بعدی بودن آزمون است. همچنین، نسبت ارزش ویژه اول به دوم، ۹/۰۶ به دست آمد که بسیار بزرگ‌تر از نسبت ارزش ویژه دوم به سوم، یعنی ۱/۹۹، است که نشان می‌دهد آزمون اساساً تک‌بعدی است (لرد<sup>۲۵</sup>، ۱۹۸۰). علاوه بر شاخص‌های مذکور، مقدار آماره ریشه دوم میانگین مجزورات باقیمانده<sup>۲۶</sup> برابر ۰/۰۴۴ است. مقادیر کوچک این آماره حاکی از برازش مدل با داده‌ها است و شاخص نیکویی برازش تاناکا<sup>۲۷</sup> در نرم‌افزار NOHARM برابر ۰/۹۸۷ است که شاهد دیگری بر تک‌بعدی بودن آزمون ریاضی است. مقدار ۰/۹ برای این شاخص حاکی از برازش قابل قبول و مقدار ۰/۹۵ بیانگر برازش خوب مدل با داده‌هاست (مک‌دونالد<sup>۲۸</sup>، ۱۹۹۷).

### هدف ۱: طبقه‌بندی استان‌ها در گروه‌های همگن

از آنجاکه تعداد استان‌های محل اقامت داوطلبان زیاد است، برای مطالعه تأثیر محل سکونت بر عملکرد، از مدل عاملی چندسطحی آمیخته استفاده شد. در این مدل سازی، هنگامی که تعداد گروه‌ها زیاد است، گروه‌هایی که از پارامترهای یکسان برخوردارند با هم خوشه‌بندی می‌شوند و طبقه‌های

مکنون را تشکیل می‌دهند. حداکثر تعداد طبقه‌های مکنون ممکن است برابر با تعداد گروه‌ها باشد. در این مدل، فرض می‌شود که واحدهای سطح بالاتر به طبقه‌های مکنون تعلق دارند. مدل عاملی آمیخته چند سطحی، خوشه‌بندی گروه‌ها را بر مبنای پارامترهای مدل عاملی ارائه می‌دهد که در بسیاری از کاربردها، طبیعی‌ترین راه برای توصیف تفاوت‌های گروهی نسبت به تجزیه واریانس چندسطحی مرسوم است. توجه داشته باشید که طبقه‌های مکنون از گروه‌ها (به‌عنوان مثال استان‌ها) با افراد درون گروه‌ها ایجاد نمی‌شود (کیم و همکاران، ۲۰۱۷).

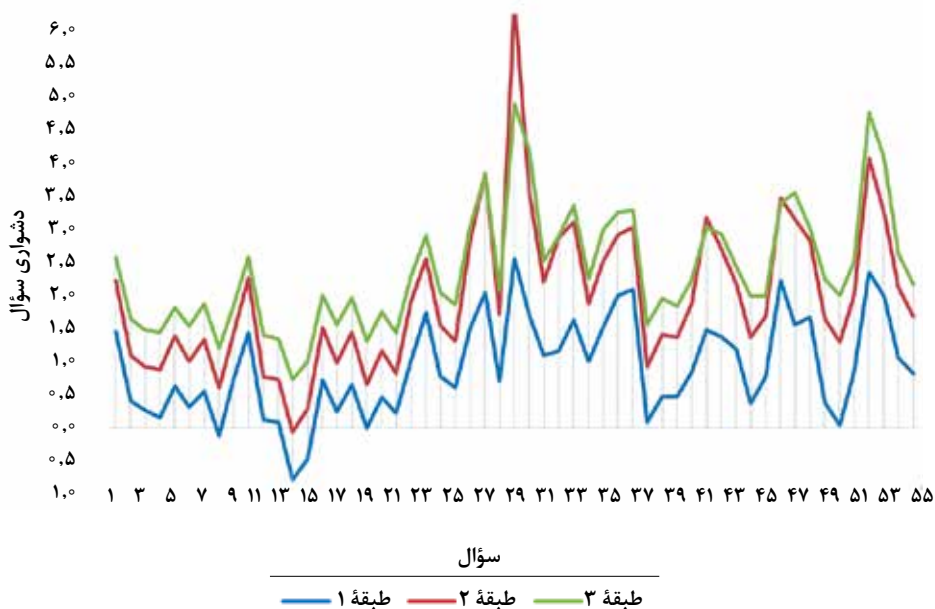
مدل عاملی چندسطحی آمیخته با افزایش تعداد طبقه‌ها به داده‌ها برازش داده شد و مقدار شاخص اطلاعات بیزی<sup>۲۹</sup> برای هر مدل به دست آمد. مقدار شاخص اطلاعات بیزی برای مدل با ۲، ۳ و ۴ طبقه مکنون به ترتیب برابر با  $۱۳۱۹۹۳۸/۶۴۳$ ،  $۱۳۱۹۰۷۸/۴۰۲$  و  $۱۳۲۱۱۷۷/۴۸۳$  به دست آمد. کم‌ترین مقدار شاخص اطلاعات بیزی در ۳ طبقه حاصل می‌شود که نشان‌دهنده برازش مناسب مدل ۳ طبقه‌ای با داده‌هاست. گروه‌بندی استان‌ها بر اساس طبقه مکنون، در جدول ۲ ارائه شده است. الگوی پاسخ داوطلبان استان‌هایی که در یک طبقه قرار دارند، تقریباً یکسان است.

جدول ۲. طبقه‌بندی استان‌ها در آزمون ریاضی با استفاده از مدل عاملی آمیخته چندسطحی

صفهان، تهران، خراسان رضوی، فارس، مازندران، یزد و البرز	طبقه ۱
آذربایجان شرقی، خوزستان، قزوین، قم، کرمان، گیلان، مرکزی، همدان	طبقه ۲
آذربایجان غربی، اردبیل، ایلام، بوشهر، چهارمحال و بختیاری، خراسان جنوبی، خراسان شمالی، زنجان، سمنان، سیستان و بلوچستان، کردستان، کرمانشاه، کهگیلویه و بویراحمد، گلستان، لرستان، هرمزگان	طبقه ۳

برآورد پارامترهای دشواری برای ۳ طبقه مکنون، در شکل ۱ نشان می‌دهد که سؤال‌های ۲۸، ۳۶، ۴۶ و ۵۲ برای طبقه اول دشوار است و اکثر سؤال‌های انتهایی آزمون برای طبقه دوم و سوم، نسبت به سؤال‌های ابتدایی، دشوارتر است. از ۲۴ سؤال اول، به‌جز سؤال ۱ (سؤال مربوط به دامنه تابع معکوس) و ۱۰ (سؤال مربوط به معادله مثلثاتی) که برای هر سه طبقه دشوارتر است، سایر سؤال‌ها برای هر طبقه دشواری کمتری دارد. سؤال ۱۳ (سؤال مربوط به دنباله‌ها) برای ۳ طبقه آسان‌ترین سؤال و سؤال ۲۸ (سؤال مربوط به زاویه‌های محیطی و محاطی از کتاب هندسه ۲) برای سه طبقه دشوارترین سؤال است.



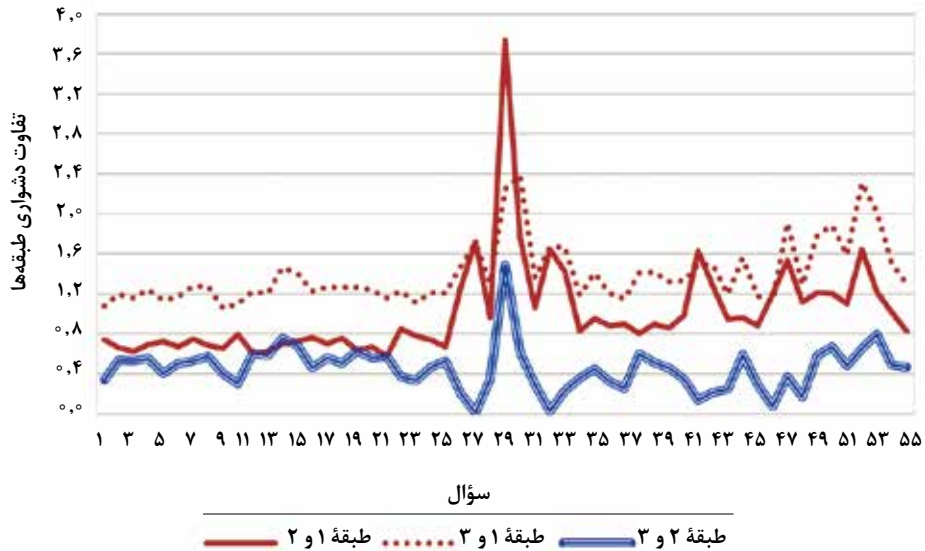


شکل ۱. دشواری سؤال‌ها در طبقه‌های مکون استان‌ها

## هدف ۲: شناسایی سؤال‌های دارای کارکرد افتراقی

برای شناسایی سؤال‌های دارای کارکرد افتراقی در میان استان‌ها با استفاده از گروه‌های همگن، از قواعد ساموئلسن مطابق با مطالعه اولیوری و همکاران (۲۰۱۳) استفاده شد. تفاوت کمتر از ۰/۴ در پارامتر دشواری سؤال‌ها، نشان‌دهنده فقدان کارکرد افتراقی، تفاوت بین ۰/۴ تا ۰/۸ نشان‌دهنده کارکرد افتراقی کوچک، تفاوت بین ۰/۸ تا ۱/۲ نشان‌دهنده کارکرد افتراقی متوسط، و تفاوت بزرگ‌تر از ۱/۲ نشان‌دهنده کارکرد افتراقی بزرگ است. تفاوت پارامتر دشواری در میان طبقه‌ها در شکل قرار دارد. سؤال‌های ۱، ۱۰، ۲۱، ۲۲، ۲۵، ۲۶، ۲۷، ۳۱، ۳۲، ۳۳، ۳۵، ۳۶، ۴۰، ۴۱، ۴۲، ۴۳، ۴۵، ۴۶، ۴۷، ۴۸ در بین طبقه ۲ و ۳ کارکرد افتراقی ندارد. بنابراین، سکونت در استان‌های آذربایجان شرقی، خوزستان، قزوین، قم، کرمان، گیلان، مرکزی، همدان، آذربایجان غربی، اردبیل، ایلام، بوشهر، چهارمحال و بختیاری، خراسان جنوبی، خراسان شمالی، زنجان، سمنان، سیستان و بلوچستان، کردستان، کرمانشاه، کهگیلویه و بویراحمد، گلستان، لرستان، هرمزگان تأثیری بر پاسخ‌گویی به این سؤال‌ها ندارد. تمامی سؤال‌ها برای طبقه اول و سوم دارای کارکرد افتراقی متوسط یا زیاد است و سؤال‌های ۲۵ تا ۵۵ (سؤال‌های مربوط به هندسه، ریاضیات گسسته، جبر و احتمال) نیز برای طبقه

دوم و اول دارای کارکرد افتراقی متوسط یا زیاد است. بنابراین، سکونت در استان‌های اصفهان، تهران، خراسان رضوی، فارس، مازندران، یزد و البرز باعث کارکرد افتراقی در پاسخ‌گویی به سؤال‌های آزمون ریاضی می‌شود.



شکل ۲. تفاوت پارامتر دشواری سؤال‌ها در طبقات مکتون استان‌ها

### هدف ۳: مقایسه طبقه‌ها بر مبنای سن، آزمون ریاضی کنکور و معدل دیپلم

برای مقایسه طبقه‌ها بر مبنای سن و معدل کتبی دیپلم به‌عنوان عوامل مؤثر بر کارکرد افتراقی سؤال‌ها، میانگین و انحراف استاندارد افراد طبقه‌های مکتون در متغیرهای سن، آزمون ریاضی کنکور، و معدل دیپلم محاسبه شد. نتایج در جدول ۳ ارائه شده است. همان‌طور که مشاهده می‌کنید، میانگین نمره‌های افراد طبقه مکتون ۱ در امتحان ریاضی کنکور و همچنین میانگین معدل دیپلم آن‌ها در مقایسه با دو طبقه دیگر بالاتر است، این در حالی است که میانگین سن آن‌ها از دو طبقه دیگر پایین‌تر است. اگر طبقه‌ها با توجه به متغیرهای پیش‌زمینه تفاوت زیادی داشته باشند، می‌تواند در تحلیل‌های بعدی به‌عنوان متغیر کمکی در مدل گنجانده شود (لوبک و موتن، ۲۰۰۵). بنابراین، سن و معدل کتبی دیپلم در پژوهش‌های آینده می‌توانند به‌عنوان متغیر کمکی وارد مدل تحلیل عاملی آمیخته چندسطحی شود.

جدول ۳. مشخصه‌های آزمودنی‌ها در طبقه‌های همگن از استان‌ها

انحراف استاندارد معدل دیپلم	میانگین معدل دیپلم	انحراف استاندارد سن	میانگین سن	انحراف استاندارد آزمون	میانگین نمره آزمون ریاضی	فراوانی	
۲/۶	۱۷/۳۱	۱/۶۸	۱۸/۷۳	۲۲/۲۹	۲۶/۹۷	۶۹۹۹	طبقه ۱
۲/۸۹	۱۵/۷۷	۱/۸۳	۱۸/۹۲	۱۴/۹۱	۱۳/۱۷	۶۹۹۹	طبقه ۲
۳/۱۶	۱۴/۰۸	۱/۹۶	۱۹/۲۴	۱۱/۶۷	۷/۰۱	۱۶۴۷۹	طبقه ۳

#### هدف ۴: مقایسه طبقه‌های مکنون همگن بر مبنای جنسیت

طبقه‌های همگن از استان‌ها بر مبنای متغیر جنسیت، به‌عنوان منبعی برای کارکرد افتراقی سؤال، موردبررسی قرار گرفت که مشخص شود آیا عضویت در این طبقه‌های مکنون با ویژگی‌های ظاهری مثل جنسیت در ارتباط است. آزمون کای دو می‌تواند برای بررسی ارتباط با توجه به سایر ویژگی‌های آزمودنی انجام شود، ولی با توجه به اینکه حجم نمونه زیاد است، آزمون‌های ارتباط میان متغیرها همواره معنادار می‌شود. به همین منظور، میزان همبستگی بین متغیرها موردبررسی قرار می‌گیرد. همبستگی جنسیت با عضویت در طبقه‌های مکنون استان‌ها برابر ۰/۰۰۸ است، بنابراین ارتباطی بین جنسیت و عضویت گروهی بسیار ضعیف است. همان‌طور که در جدول ۴ مشاهده می‌شود، درصد مردان و زنان در کل نمونه تقریباً ۳۲ و ۶۸ درصد است. این نسبت در ۳ طبقه مکنون از استان‌ها نیز تقریباً یکسان است. ترکیب مردان و زنان در طبقه‌ها، مشابه با کل نمونه است که نشان می‌دهد جنسیت تأثیری در طبقه‌بندی افراد نداشته است.

جدول ۴. درصد افراد هر طبقه به تفکیک جنسیت

طبقه		جنسیت		
کل	۳	۲	۱	
۳۲/۳	۳۲/۴	۳۲/۶	۳۱/۷	زن
۶۷/۷	۶۷/۶	۶۷/۴	۶۸/۳	مرد

(علامت درصد حذف شده است)

## ■ بحث و نتیجه‌گیری ■

هرساله، مباحثی در زمینه عملکرد متفاوت آزمودنی‌ها در استان‌ها و عدالت آموزشی مطرح می‌شود. استان‌های ایران از نظر میزان توسعه‌یافتگی و نابرابری در توزیع امکانات و بهره‌مندی از آموزش، متفاوت هستند (دربان آستانه، طهماسبی و رضایی، ۱۳۹۴؛ میرغفوری، طهراری مهرجردی و بابایی، ۱۳۸۹). این پژوهش‌ها نشان می‌دهد که برخی از استان‌های مرزی مانند سیستان و بلوچستان، و بوشهر نسبت به استان‌های مرکزی امکانات آموزشی بسیار ضعیف‌تری دارند، درحالی‌که استان‌هایی مانند تهران، یزد و آذربایجان شرقی، بهترین وضعیت را در زمینه شاخص‌های کتابخانه‌ای دارند که به‌عنوان شاخصی برای توسعه‌یافتگی در نظر گرفته می‌شود. صفائی‌پور و مودت (۱۳۹۲) برای بررسی توسعه و فقر در استان‌های ایران از شاخص‌های اجتماعی اقتصادی و توسعه انسانی استفاده کردند که نشان داد میزان توسعه استان‌ها تفاوت‌های فاحشی دارد.

مقایسه درصد رتبه‌های برتر یا درصد قبولی‌ها در هر استان، اطلاعات توصیفی ارائه می‌نماید، اما نشان نمی‌دهد که تفاوت‌های عملکردی آزمودنی‌ها در استان‌ها اصولاً مربوط به مهارت‌ها و توانایی‌های تحت‌سنجش، یا عوامل نامرتبط است. تفاوت‌های معنادار است که توانایی موردنظر در گروه‌ها کنترل شده باشد. برای سازندگان آزمون‌ها، یکی از اهداف اصلی در تحقیق کارکرد افتراقی سؤال، درک بهتر علل کارکرد افتراقی در سؤال‌های آزمون است. منابع متعددی برای کارکرد افتراقی سؤال وجود دارد و ممکن است بسته به نوع و هدف گروه‌های موردسنجش و مقایسه، معنا و منابع کارکرد افتراقی سؤال متفاوت باشد. استفاده از روش طبقه‌های مکنون برای شناسایی منابع کارکرد افتراقی سؤال مفید است. تلاش برای درک اینکه چرا برخی آزمودنی‌ها به‌طور افتراقی به سؤال‌ها پاسخ می‌دهند، اغلب در خارج از چارچوب نظریه سؤال پاسخ انجام می‌شود. دی‌ایالا و همکاران (۲۰۰۲) و ساموئلسن (۲۰۰۵) و کوهن و بالت (۲۰۰۵) نشان دادند با کاهش میزان هم‌پوشانی بین گروه‌های آشکار و طبقه‌های مکنون، قدرت شناسایی صحیح سؤال‌های دارای کارکرد افتراقی توسط روش‌های استاندارد کاهش می‌یابد و بنابراین مشخص می‌شود که چگونه مدل‌های آمیخته می‌تواند به‌عنوان پایه‌ای برای درک علت کارکرد افتراقی سؤال، استفاده شود. به همین دلیل، در پژوهش

حاضر تأثیر ناهمگنی جامعه آزمودنی‌ها، به‌عنوان علت اصلی کارکرد افتراقی در بین بیش از دو گروه، مطالعه شد. هنگامی که گروه‌های مورد مقایسه زیاد است (تعداد استان‌ها)، مدل تحلیل عاملی چندسطحی آمیخته مناسب است. آزمون ورودی آموزش عالی در ایران در تعیین سرنوشت دانش‌آموزان اهمیت فراوانی دارد. ۵۵ سؤال درس ریاضی این آزمون، به‌عنوان نمونه برای بررسی اهداف پژوهش حاضر مورداستفاده قرار گرفت.

یافته‌های پژوهش برای بررسی هدف اول، طبقه‌بندی استان‌ها در گروه‌های همگن با استفاده از مدل عاملی چندسطحی آمیخته نشان داد که استان‌های ایران در سه طبقه مکنون قرار می‌گیرند که آزمودنی‌ها در این طبقه‌ها دارای الگوی مشابهی برای پاسخ به سؤال‌ها هستند، بنابراین سکونت در استان‌های مختلف ایران باعث تفاوت عملکرد آزمودنی‌ها در آزمون ریاضی شده است. سؤال‌های انتهایی آزمون که مربوط به هندسه، جبر و احتمال است، برای استان‌های طبقه دوم و سوم دشوارتر است. کوهن و بالت (۲۰۰۵) نیز به توصیف مدل‌های آمیخته به‌عنوان ابرازی برای کشف ناهمگنی جامعه پرداخته است که از روش مدل‌های آمیخته نظریه سؤال پاسخ برای شناسایی ابعاد ثانویه مزاحم استفاده شده است. یافته‌های این تحقیق سه طبقه مکنون را نشان می‌دهد ماهیت سؤال‌های آزمون ریاضی باعث شد که برخی سؤال‌های آزمون برای طبقه مکنون خاصی به‌طور افتراقی دشوارتر باشد. سؤال‌های جبر برای طبقه اول، هندسه یا اندازه‌گیری برای طبقه دوم، فاکتورگیری و ساده‌سازی جبری برای طبقه سوم دشوارتر بود. همچنین، تحقیقات کوهن و بالت (۲۰۰۵)، وب و همکاران، (۲۰۰۸) و اولیوری و همکاران (۲۰۱۳) تأثیر ناهمگنی جامعه را در گروه‌های جنسیتی، قومی، و زبانی نشان دادند. نتایج مطالعات بالت و همکاران (۲۰۰۱) نیز شواهدی در مورد ناهمگنی درون گروهی نشان داد. این مطالعه از مدل نظریه سؤال پاسخ آمیخته برای مقایسه پیشرفت تحصیلی دانش‌آموزان بر مبنای انتخاب آزمون چندگزینه‌ای دانشگاهی، استفاده کرد. دو طبقه مکنون همگن از آزمودنی‌ها با استفاده از مدل نظریه سؤال پاسخ آمیخته، شناسایی شد که از لحاظ الگوی انتخاب (انتخاب اینکه آیا گرامر یا نشانه‌گذاری، منجر به خطا شده است) متفاوت بودند. ناهمگنی درون گروهی همچنین در مطالعه وب و همکاران (۲۰۰۸) دو طبقه مکنون وجود دارد که نشان‌دهنده ناهمگنی درون گروهی

قومیتی است. سهمیه آزمون سراسری در بخش‌های مختلف کشور بر اساس میزان شاخص‌های توسعه‌یافتگی تعیین می‌شود. استفاده از قابلیت مدل‌های آمیخته در طبقه‌بندی آزمودنی‌ها در طبقات همگن بر مبنای الگوی پاسخ به سؤال‌ها، یک راهکار مناسب آماری برای تعیین سهمیه مناطق در آزمون سراسری است.

بررسی هدف دوم شناسایی سؤال‌های دارای کارکرد افتراقی در میان استان‌ها با استفاده از گروه‌های همگن، با استفاده از محاسبه تفاوت دشواری پارامترها در گروه‌های همگن سؤال‌های فاقد کارکرد افتراقی و سؤال‌های با کارکرد افتراقی متوسط و زیاد را در میان استان‌های ایران مشخص نمود. برای بررسی هدف سوم، مقایسه طبقه‌ها بر مبنای سن و معدل کتبی دیپلم، به‌عنوان عوامل مؤثر بر کارکرد افتراقی سؤال‌ها، نشان داد که آزمودنی‌های ساکن در طبقه ۱، به‌طور میانگین، سن کمتر و معدل کتبی دیپلم بالاتری دارند و نمره بالاتری در آزمون ریاضی کسب نموده‌اند. بنابراین، سن و پیشینه تحصیلی آزمودنی‌ها از عوامل تأثیرگذار بر کارکرد افتراقی است و پیشنهاد می‌شود، در مطالعات آینده این متغیرها به‌عنوان متغیر کمکی در مدل تحلیل عاملی چندسطحی آمیخته مورد استفاده قرار گیرد. به‌طور کلی، عملکرد آزمودنی‌ها در استان‌های (طبقه اول) اصفهان، تهران، خراسان رضوی، فارس، مازندران، یزد و البرز در اکثر سؤال‌ها نسبت به سایر آزمودنی‌ها بهتر است. این استان‌ها، برخوردار و توسعه‌یافته هستند و آزمودنی‌ها در این استان‌ها، به‌طور میانگین، معدل کتبی دیپلم بالاتری نسبت به سایر استان‌ها دارند و درصد پاسخ درست به سؤال‌ها برای این آزمودنی‌ها بالاتر است و سؤال‌ها برای آن‌ها اغلب ساده‌تر است. از آنجاکه آزمون سراسری ورودی دانشگاه‌ها لازم است تمامی سطوح توانایی را پوشش دهد، نمی‌توان از سؤال‌ها یا محتواهایی که در میان این ۷ استان با سایر استان‌ها دارای کارکرد افتراقی است، حذف نمود. قرار دادن سؤال‌های متفاوت برای مناطق آموزشی مختلف یا وزن‌دهی مختلف به سؤال‌ها بر اساس میزان کارکرد افتراقی در میان استان‌ها، می‌تواند در این زمینه راه‌گشا باشد.

از طرفی با اختصاص سهمیه مناطق بومی برای استان‌های محروم بازهم درصد پذیرش در آزمون سراسری در این استان‌ها نسبت به استان‌های برخوردار کمتر است، بنابراین نمی‌توان بیان کرد که استان محل سکونت

آزمودنی تنها دلیل برای کارکرد افتراقی آزمودنی‌هاست. بررسی اطلاعات توصیفی آزمودنی‌ها در استان‌های طبقه ۳، مثل سیستان و بلوچستان، آذربایجان غربی، ایلام و کهگیلویه و بویراحمد، نشان داد که آزمودنی‌ها در این استان‌ها میانگین سنی بالاتر و معدل کتبی دیپلم پایین‌تری دارند. بنابراین، عامل کارکرد افتراقی در این سؤال‌ها تنها محل سکونت آزمودنی نیست، بلکه سن آزمودنی‌ها و پیشینه تحصیلی ضعیف‌تر (معدل کتبی دیپلم پایین‌تر) باعث کارکرد افتراقی شده است. داوطلبان آزمون سراسری با پیشینه تحصیلی ضعیف و توانایی پایین‌تر در استان‌های محروم در آزمون سراسری شرکت می‌نمایند و پذیرفته نمی‌شوند. پیشنهاد می‌شود، در پژوهش‌های آینده با استفاده از روش‌هایی مثل تفسیر سؤال‌ها توسط شرکت‌کنندگان در آزمون و قوانین تفکر با صدای بلند، چگونگی تأثیر سن آزمودنی بر پاسخ‌گویی به سؤال‌ها مورد بررسی قرار گیرد. همچنین، پژوهشی کیفی و آسیب‌شناسانه می‌تواند در زمینه شناسایی سایر عوامل تأثیرگذار بر کارکرد افتراقی یاری‌رسان باشد.

نتایج بررسی هدف چهارم، مقایسه طبقه‌های همگن از استان‌ها بر مبنای متغیر جنسیت، به‌عنوان منبعی برای کارکرد افتراقی سؤال، نشان داد که ترکیب زنان و مردان در طبقات، مشابه ترکیب زنان و مردان در نمونه است، بنابراین جنسیت عامل مهمی در تمایز طبقات نیست و زنان و مردان الگوی متفاوتی در پاسخ به سؤال‌ها ندارند که با نتایج تحقیق کوهن و بالت (۲۰۰۵) که نشان داد جنسیت ارتباط کاملاً ضعیفی با عضویت در طبقه‌های مکنون دارد و نتایج پژوهش اولیوری و همکاران (۲۰۱۳) نشان داد که زنان و مردان کم‌وبیش به‌طور مساوی در هر طبقه توزیع شده‌اند، همخوان است و با تحقیق گرامی‌پور و همکاران (۱۳۹۵)، شریفی یگانه (۱۳۹۱)، گرامی‌پور، رضایی، نوروزی و مختاریان (۱۳۹۶) که بدون بررسی ناهمگنی نشان دادند سؤال‌های آزمون ورودی دانشگاه‌ها نسبت به جنسیت، دارای کارکرد افتراقی است، ناهمخوان است. بیان این مطلب که آزمودنی‌ها در یک گروه مثلاً زنان، به‌طور کلی در مقایسه با گروه مردان، به سؤال دارای کارکرد افتراقی، متفاوت پاسخ می‌دهند، منجر به یک دید کلیشه‌ای در مورد یک سؤال می‌شود که به نفع همه اعضای یک گروه (مردان یا زنان) است، درحالی‌که ناهمگنی واقعی در هر گروه نادیده گرفته می‌شود. چنین دیدگاهی

منجر به ایده لغو اثر کارکرد افتراقی شود که در آن تأثیر کارکرد افتراقی یک سؤال به نفع یک گروه با وارد کردن سؤال‌ها به نفع گروه دیگر حذف می‌شود. سؤال‌هایی که دارای کارکرد افتراقی جنسیت شناسایی شده‌اند، به این صورت نیست که برای تمامی مردان و زنان دارای کارکرد افتراقی باشد. در حقیقت، در این پژوهش علت سؤال‌هایی که دارای کارکرد افتراقی شناسایی شده است، محتوا و موضوع سؤال است که آن سؤال را برای گروهی از آزمودنی‌ها نسبت به سایرین دشوارتر می‌سازد.

محدودیت مطالعه کارکرد افتراقی سؤال در این پژوهش، این است که آن‌ها صرفاً بر مبنای نتایج حاصل از آزمون ورودی دانشگاه‌هاست و در نتیجه تنها به استنباط‌هایی مربوط به رابطه بین محتوای سؤال و کارکرد افتراقی سؤال پرداخته شد. با توجه به اینکه تاکنون بررسی درزمینه منابع کارکرد افتراقی در آزمون‌های ورودی دانشگاه انجام نگرفته است، تنها از رویکرد آماری مدل تحلیل عاملی چندسطحی آمیخته، برای بررسی منابع احتمالی کارکرد افتراقی سؤال استفاده شد. طبقه‌بندی آزمودنی‌ها در طبقات مکنون با استفاده از مدل‌های آمیخته، به‌عنوان پایه‌ای برای درک علت کارکرد افتراقی سؤال استفاده می‌شود. همچنین، با توجه به اطلاعات محدودی که از آزمودنی‌های در دسترس پژوهشگر قرار گرفت، مثل سن و جنسیت، تنها تأثیر این متغیرها به‌عنوان منبعی برای کارکرد افتراقی سؤال در این پژوهش مورد بررسی قرار گرفت. محدودیت مطالعه منابع کارکرد افتراقی سؤال، فقدان اظهارات صریح در مورد مهارت‌ها و توانایی‌هایی زیربنایی برای پاسخ‌گویی به سؤال‌ها در آزمون است. ابعاد اولیه و ثانویه برای پاسخ‌گویی به سؤال‌ها توسط طراحان سؤال مشخص نشده است. اگر ابعاد و مهارت‌های زیربنایی سؤال مشخص بود، محقق می‌توانست فرضیه‌های لازم را در سؤال‌های چندبعدی، به‌عنوان منبعی برای کارکرد افتراقی سؤال، بررسی نماید.



منابع

- ایزانلو، بلال، بازرگان، عباس، فرزاد، ولی ا.، صادقی، ناهید و کاوسی، امیر. (۱۳۹۳). تفکیک ابعاد متعامد از خوشه‌های سؤال بر اساس هشت روش تعیین بعد در داده‌های دوارزشی: مورد آزمون ریاضی رشته ریاضی فیزیک کنکور ۹۲-۹۱. *فصلنامه اندازه‌گیری تربیتی*، ۵(۱۸)، ۲۰۷-۲۴۰.
- دربان آستانه، علیرضا، طهماسبی، سیامک و رضایی، پانید. (۱۳۹۵). تحلیل الگوی نابرابری فضای آموزشی شهرستان‌های کشور. *مطالعات برنامه‌ریزی آموزشی* ۵(۹)، ۳۱-۵۰.
- رودباری، مسعود و فیض‌آبادی فراهانی، زهرا. (۱۳۸۶). فراوانی و متغیرهای مرتبط در نتیجه دانش‌آموزان دوره پیش‌دانشگاهی در کنکور سال ۱۳۸۴. *مجله علوم تربیتی و روان‌شناسی*، ۴(۶)، ۸۴-۶۵.
- میرغفوری، سید حبیب‌الله، طهارت مهرجردی، محمدحسین و بابایی، حمید. (۱۳۸۹). شناسایی وضعیت توسعه‌یافتگی و رتبه‌بندی استان‌های کشور از لحاظ دسترسی به شاخص‌های بخش کتابخانه‌ای. *فصلنامه کتابداری و اطلاع‌رسانی*، ۱۳(۳).
- صفائی‌پور، مسعود و مودت، الیاس. (۱۳۹۲). ارزیابی استان‌های ایران با تأکید بر شاخص‌های اجتماعی-اقتصادی و شاخص ترکیبی توسعه انسانی با استفاده از تکنیک TOPSIS و GIS. *مطالعات ساختار و کارکرد شهری*، ۱(۳)، ۲۷-۱۱.
- شریفی یگانه، نگار. (۱۳۹۱). ارزیابی کارکرد افتراقی جنسیتی سؤال‌های آزمون ریاضی با استفاده از دو روش مانتل - هنزل و نظریه سؤال - پاسخ. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۱(۲)، ۷۶-۵۳.
- صادقی جعفری، جواد، روشن، محمد و شکوری گنجوی، حامد. (۱۳۹۰). بررسی مقایسه‌ای عملکرد تحصیلی دانشجویان سهمیه‌های ایثارگری و سهمیه مناطق در دوره‌های روزانه دانشگاه‌های دولتی. *نامه آموزش عالی*، ۳(۱۰)، ۷۵-۵۱.
- گرامی پور، مسعود، رضایی، احمد، رضائی صدر، اعظم و نوروزی، لیلا. (۱۳۹۵). کنش افتراقی سؤال در آزمون‌های سازمان سنجش آموزش کشور بر حسب ویژگی‌های جمعیت‌شناختی داوطلبان کنکور سراسری. *فصلنامه اندازه‌گیری تربیتی*، ۷(۲۶)، ۱۲۲-۱۰۱.
- گرامی پور، مسعود، رضائی صدر، اعظم، رضایی، احمد، نوروزی، لیلا و مختاریان، فرانک. (۱۳۹۶). کنش افتراقی سؤال (DIF) و آزمون (DTF) مرتبط با جنسیت در آزمون‌های کنکور سراسری سازمان سنجش آموزش کشور. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۷(۱۹)، ۷۳-۴۱.
- محمدبیگی، ابوالفضل، محمدصالحی، نرگس و علی گل، محمد. (۱۳۹۳). روایی و پایایی ابزارهای و روش‌های مختلف اندازه‌گیری آن‌ها در پژوهش‌های کاربردی در سلامت. *مجله دانشگاه علوم پزشکی رفسنجان*، ۱۳(۱۲)، ۱۱۷۰-۱۱۵۳.
- Bolt, D., M., Cohen, A. S., & Wollack, J. A. (2001). A mixture model for multiple choice data. *Journal of Educational and Behavioral Statistics*, 26, 381-409.
- Cohen, A., & Bolt, D. (2005). A Mixture Model Analysis of Differential Item Functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- De Ayala, R. J., Kim, S.-H., Stapleton, L. M., & Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing*, 2, 243-276.
- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.
- Kim, E. S., Cao, CH., Wang, Y., & Nguyen, D. T. (2017) Measurement Invariance Testing with Many Groups: A Comparison of Five Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544. doi: 1080/10/2017/10705511.1304822
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10, 21-39. doi:1037/10/1082-989X.1/10.21
- Mazor, M. M., Hambleton, R. K., and Clauser, B. E. (1998). Multidimensional DIF analysis: the effects of matching on unidimensional subtest scores. *Applied Psychological Measurement*, 22, 357-367. doi: 10.1177/014662169802200404

- McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 258-269). New York: Springer Verlag.
- Mislavy, R., Wilson, M., Ercikan, K., & Chudowsky, N. (2002). Psychometric principles in student evaluation. In D. Nevo & D. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 478-520). Dordrecht, The Netherlands: Kluwer.
- Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide* (Eighth Ed). Los Angeles, CA: Muthén & Muthén.
- Oliveri, M., Ercikan, K. & Zumbo, B. (2013). Analysis of Sources of Latent Class Differential Item Functioning in International Assessments. *International Journal of Testing*, 13(3), 272-293. doi:13.10.1080/15305058.2012.738266
- Pomes, M., Squires, J., & Yovanoff, P. (2014). Psychometric examination of a Spanish translation of a developmental screening instrument. *Journal of Early Childhood Research*, 14. doi: 10.1177/1476718X14529279
- Samuelsen, K. M. (2005). *Examining differential item functioning from a latent class perspective* (Unpublished Doctoral dissertation). University of Maryland, College Park.
- Stout, W., Froelich, A., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 357-376). New York, NY: Springer-Verlag.
- Webb, M. L., Cohen, A. S., & Schwanenflugel, P. J. (2008). Latent class analysis of differential item functioning on the Peabody Picture Vocabulary Test-III. *Educational and Psychological Measurement*, 68, 335-351.
- Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). *Testing handicapped people*. Needham, MA: Allyn & Bacon.
- Young, J. W., Steinberg, J., Cline, F., Stone, E., Martiniello, M., Ling, G., & Cho, Y. (2010). Validity and fairness of standards-based assessments for initially fluent students and former English language learners. *Educational Assessment*, 15, 87-106.

## پی‌نوشت‌ها

- |   |   |
|---|---|
| <ol style="list-style-type: none"> <li>1. Willingham, Ragosta, Bennett, Braun, Rock &amp; Powers</li> <li>2. Young, Steinberg, Cline, Stone, Martiniello, Ling &amp; Cho</li> <li>3. differential item functioning</li> <li>4. conceptualization</li> <li>5. De Ayala, Kim, Stapleton, Dayton</li> <li>6. Mantel-Haenszel Chi-Square</li> <li>7. Exact Signed Area</li> <li>8. Samuelsen</li> <li>9. Cohen &amp; Bolt</li> <li>10. informative</li> <li>11. Webb, Cohen &amp; Schwanenflugel</li> <li>12. Bolt, Cohen, &amp; Wollack</li> <li>13. Oliveri, Ercikan &amp; Zumbo</li> <li>14. Mislavy, Wilson, Ercikan &amp; Chudowsky</li> </ol> | <ol style="list-style-type: none"> <li>15. Factor mixture model</li> <li>16. Lubke &amp; Muthén</li> <li>17. multilevel factor mixture model</li> <li>18. Kim, Cao, Wang &amp; Nguyen</li> <li>19. Pomes., Squires &amp; Yovanoff</li> <li>20. Stout, Froelich &amp; Gao</li> <li>21. Fraser &amp; McDonald</li> <li>22. Cronbach's Alpha</li> <li>23. Reliability</li> <li>24. Mazor, Hambleton &amp; Clauser</li> <li>25. Lord</li> <li>26. Root Mean Square of Residuals(RMSR)</li> <li>27. Tanaka</li> <li>28. McDonald</li> <li>29. Bayesian Information Criteria (BIC)</li> </ol> |
|---|---|