

ارائه مدل مطلوب برای پیوند زدن نمره‌های سوابق تحصیلی با نمره‌های آزمون سراسری ورود به دانشگاه

سمیه کاوه*
ابراهیم خدایی**
سید امین موسوی***
علی مقدم‌زاده****
جلیل یونسی*****

چکیده

هدف از نگارش این مقاله، انتخاب روشی مناسب برای مرتبط کردن نمره‌های سوابق تحصیلی به نمره‌های آزمون سراسری ورود به دانشگاه به جای استفاده از روش نرمال‌سازی بود. سپس دو روش مورد اشاره بر اساس خطای استاندارد اندازه‌گیری شرطی مقایسه شدند. برای دستیابی به این هدف، نمونه‌ای ۵۰۰۰ نفری از جامعه دانش‌آموزان دانش‌آموخته سال ۱۳۹۳ رشته ریاضی فیزیک انتخاب شد که در آزمون سراسری ورود به دانشگاه سال ۱۳۹۴ شرکت کرده بودند. پس از بررسی عامل‌های معرفی شده توسط «دوران» روی نمره‌ها و نیز نظر پمربچ و همکاران (۲۰۰۴)، تطابق برای مرتبط کردن نمره‌ها در نظر گرفته شد. سپس روش‌های میانگین، خطی و همصدک‌سازی برای تبدیل نمره‌های آزمون‌های پیشرفت تحصیلی به آزمون سراسری ورود به دانشگاه استفاده شد. نتایج نشان داد که روش همصدک‌سازی کمتر به گروه وابسته است؛ بدین‌سان، با استفاده از این روش، نمره‌های این آزمون‌ها برای تصمیم‌گیری‌ها با یکدیگر قابل مقایسه هستند. همچنین، در این پژوهش نشان داده شد که خطای استاندارد اندازه‌گیری شرطی برای روش ارائه شده بسیار کمتر از روش مورد استفاده توسط سازمان سنجش آموزش کشور است.

واژگان کلیدی: پیوند زدن، تطابق، نرمال‌سازی، خطای استاندارد اندازه‌گیری شرطی.

* دانشجوی دکتری تخصصی سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران (نویسنده مسئول: kaveh1@ut.ac.ir).

** دانشیار گروه روش‌ها و برنامه‌های آموزشی و درسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران
*** استادیار گروه روان‌شناسی تربیتی و آموزش استثنایی، دانشگاه ساسکاچوان، کانادا
**** استادیار گروه روش‌ها و برنامه‌های آموزشی و درسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه

تهران

***** دانشیار گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی

مقدمه

نمره‌های حاصل از اجرای یک آزمون اغلب می‌تواند به‌عنوان قسمتی از اطلاعات مورد نیاز برای اخذ تصمیمات مهم آموزشی برای آزمودنی به کار گرفته شود. هر چه این اطلاعات دقیق‌تر باشد، تصمیم‌گیری‌های دقیق‌تری می‌توان اتخاذ کرد. گاهی یک آزمون ممکن است در چندین زمان اجرا شود، اگر پرسش‌های آزمون در همهٔ زمان‌ها یکسان باشد، دانش‌آموزی که بیش از یک بار در آزمون شرکت می‌کند - به علت آشنایی با پرسش‌ها، تمرین و تکرار - نمره‌ای بالاتر از سطح توانایی خود کسب می‌کند؛ به همین منظور، در هر بار اجرای آزمون، مجموعه متفاوتی از پرسش‌ها طراحی می‌شود که به آن «فرم آزمون»^۱ می‌گویند. به عبارت دیگر، این مجموعه پرسش‌ها هرچند ویژگی‌های آماری و محتوایی یکسان دارند؛ اما از نظر دشواری تا حدودی با یکدیگر متفاوت هستند. این شرایط باعث می‌شود که نمره‌های حاصل از آزمون‌های مختلف با یکدیگر تفاوت داشته باشند. پس باید بتوان نمره‌های آزمون را در اجراهای مختلف با یکدیگر مقایسه کرد تا اطمینان حاصل شود که این آزمون‌ها برای تمام آزمودنی‌ها عادلانه است (کولن و برنان، ۲۰۰۴).

همترازسازی^۲ آزمون به‌صورت یک فرایند آماری در جهت مرتبط کردن نمره‌های حاصل از اجرای فرم‌های متفاوت آزمون‌ها - با دشواری‌های مختلف؛ اما ویژگی‌های آماری و محتوایی یکسان - به کار گرفته می‌شود (همان). بر این اساس، با استفاده از روش‌های آماری می‌توان نمره‌ها را با یکدیگر تطبیق داد و مرتبط کرد. نمره‌های حاصل از همترازسازی دو آزمون تعویض‌پذیر^۳ هستند. در صورتی که نمره‌های آزمون‌هایی که می‌خواهیم آنها را به یکدیگر مرتبط کنیم، محتوا و ویژگی‌های آماری متفاوت از یکدیگر داشته باشند، دیگر نمی‌توان از همترازسازی بین نمره‌های آزمون‌ها استفاده کرد. به عبارت دیگر، ارتباط دادن بین نمره‌های حاصل از آزمون‌هایی با محتوا و ویژگی‌های آماری متفاوت از یکدیگر، به روش‌های آماری دیگری نیاز دارد. به این

1. test forms

2. Kolen & Brennan

3. equating

4. interchangeable

روش‌ها برای مرتبط کردن چنین نمره‌های «پیوند زدن»^۱ می‌گویند (ین، برنان و کولن،^۲ ۲۰۰۴). این پیوند که با تبدیل نمره‌های یک آزمون به نمره‌های آزمون دیگر صورت می‌پذیرد، می‌تواند شامل نمره‌های خام یا مقیاس‌بندی شده باشد (آنگف،^۳ ۱۹۷۱). این عملیات به تفسیرپذیری نمره‌های دو آزمون متفاوت کمک می‌کند.

برخی مواقع، آزمون‌هایی که می‌خواهیم نمره‌های آنها را به یکدیگر مرتبط کنیم از نوع مجموعه آزمون (آزمون‌های ترکیبی)^۴ هستند که دربردارنده مجموعه آزمون‌هایی با حیطه‌های متفاوت از یکدیگرند؛ مانند آزمون‌های ملی معروف SAT IV+M^۵ و ACT^۶ که در ایالات متحده برای پذیرش در دانشگاه مورد استفاده قرار می‌گیرند. همچنین، آزمون سراسری ورود به دانشگاه در ایران که هر سال برای پذیرش دانش‌آموزان در دانشگاه‌ها برای پنج گروه علوم ریاضی و فنی، علوم تجربی، علوم انسانی، زبان‌های خارجی و هنر در تیرماه برگزار می‌شود، نیز از نوع آزمون‌های ترکیبی است.

پس از برگزاری آزمون سراسری، یک سری نمره‌های خام حاصل می‌شوند که دامنه تغییرات آنها برابر با ۳۳٪- تا ۱۰۰٪ است. به علت تفاوت در ماهیت درس‌های امتحانی اعم از عمومی و اختصاصی و نیز سطح دشواری درس‌ها با یکدیگر، نمره‌های خام ناهماهنگ بوده و با یکدیگر قابل مقایسه نیستند، بنابراین، برای رفع این مشکل از نمره‌های تراز استفاده می‌شود و درواقع، با استفاده از تبدیل غیرخطی مقیاس‌سازی^۷، نمره‌های خام به مقیاس نرمال‌سازی^۸ تبدیل می‌شوند. بدین‌گونه که برای تعیین نمره‌های تراز (نرمال‌سازی)، لازم است که تمامی نمره‌های خام داوطلبان در هر گروه آزمایشی و هر درس در دسترس باشند. سپس بر اساس نمره‌های هر

1. linking

2. Yin, Bernann & Kolen

3. Angoff

4. test battery

5. Scholastic Assessment Test I verbal and math

6. American College Test

7. scaling

8. normalized

درس به ترتیب نزولی، فراوانی نسبی، فراوانی تجمعی و مقادیر فراوانی تجمعی نسبی زیرسطح نمره‌ها محاسبه می‌شود. پس از آن، نمره Z متناظر با آن از روی معکوس تابع تجمعی نرمال استاندارد به دست آورده می‌شود. درنهایت، نمره‌های Z با استفاده از رابطه زیر به نمره‌های تراز (نرمال‌سازی) تبدیل می‌شوند:

$$sc(y) = \sigma(sc)z + \mu(sc) \quad (۱)$$

Sc نمره مقیاس‌بندی شده تراز (نرمال‌سازی) است. $\mu(sc)$ و $\sigma(sc)$ به ترتیب میانگین و انحراف استاندارد نمره تراز هستند که در سازمان سنجش برای دانش‌آموختگان سال ۱۳۹۳ به ترتیب برابر با ۵۰۰۰ و ۱۲۵۰ در نظر گرفته شدند. مرحله پنجم، نمره‌های به دست آمده در مرحله قبل به نزدیک‌ترین نمره صحیح گرد می‌شود (جهانی‌فر، خدایی، یونسی و موسوی، ۱۳۹۶). بدین ترتیب، نمره‌ها روی یک مقیاس مشترک قرار می‌گیرند.

پس از محاسبه نمره تراز درس‌ها، نمره کل آزمون سراسری ورود به دانشگاه داوطلب در هر زیرگروه با استفاده از میانگین وزنی نمره‌های تراز درس‌های عمومی و اختصاصی داوطلب با توجه به ضریب هر درس در سازمان سنجش محاسبه می‌شود (سازمان سنجش آموزش کشور، ۱۳۹۴).

در چند سال اخیر، برای ورود به دانشگاه‌ها، نمره کل دانش‌آموزان فقط به نمره آزمون سراسری محدود نشده است، بلکه به نمره امتحانات نهایی و سوابق تحصیلی آنها نیز بستگی دارد. بلکه نمره کل سوابق تحصیلی دانش‌آموزانی که دیپلم‌های خود را در سال ۱۳۸۴ تا ۱۳۹۳ اخذ کرده‌اند و امتحانات یک یا چند درس آنها به‌صورت نهایی و کشوری برگزار شده است نیز تأثیرگذار بوده است؛ به‌طوری که نمره کل نهایی آنها، ترکیبی از نمره آزمون و نمره سوابق تحصیلی بوده است. در غیر این

۱. طبق تعریف سازمان سنجش، سوابق تحصیلی عبارت است از نمره‌های درس‌های سال سوم متوسطه (دیپلم) که زمان امتحان آن سال ۱۳۸۴ و پس از آن بوده و به‌صورت نهایی و کشوری برگزار شده است.

صورت، دانش‌آموزانی که شامل اعمال سوابق تحصیلی نشوند، نمره کل نهایی آنها برابر با نمره کل کسب شده در آزمون سراسری ورود به دانشگاه است. در آزمون سراسری ورود به دانشگاه سال ۱۳۹۴ تأثیر آن به صورت قطعی و به میزان ۲۵ درصد بوده است. این نمره‌های سوابق تحصیلی، نمره‌های حاصل از آزمون‌های پیشرفت تحصیلی است که به صورت تشریحی و با دامنه تغییرات بین ۰ تا ۲۰ برای دانش‌آموزان سال سوم دبیرستان در خرداد ماه به صورت نهایی و در سرتاسر کشور برگزار می‌شود. از آنجایی که دامنه و پراکندگی نمره‌های حاصل از سوابق تحصیلی و نمره‌های خام آزمون سراسری با یکدیگر متفاوت هستند، این نمره‌های خام نیز همانند نمره‌های خام آزمون سراسری ورود به دانشگاه به نمره‌های تراز تبدیل می‌شوند. سپس نمره کل سوابق تحصیلی در هر زیرگروه، با استفاده از میانگین وزنی نمره‌های تراز هر یک از درس‌های مؤثر در سوابق تحصیلی محاسبه می‌شود. در نهایت، نمره کل نهایی حاصل از نمره کل آزمون سراسری ورود به دانشگاه و سابقه تحصیلی در آزمون سراسری ورود به دانشگاه سال ۱۳۹۴ برای دانش‌آموزان مشمول سوابق تحصیلی به صورت زیر محاسبه می‌شود (سازمان سنجش آموزش کشور، ۱۳۹۴):

$$+ \text{نمره کل آزمون سراسری در زیرگروه} \times (\text{میزان تأثیر} - 1) = \text{نمره کل نهایی در زیرگروه} \\ \text{نمره کل سوابق تحصیلی در زیرگروه} \times \text{میزان تأثیر}$$

در سازمان سنجش آموزش کشور، در ترکیب و همسان‌سازی نمره‌های سوابق تحصیلی و آزمون سراسری ورود به دانشگاه موارد زیر در نظر گرفته نشده است:

- با تبدیل نمره‌های خام به نمره‌های تراز تنها نوعی مقیاس‌سازی، روی نمره‌ها رخ داده است که باعث قرار دادن نمره‌ها روی یک مقیاس مشترک است که این امر باعث تسهیل تفسیرپذیری نمره‌ها می‌شود. این روش را می‌توان برای یک آزمون^۱-آزمونی که شامل یک حیطة درسی است- و همچنین برای یک مجموعه آزمون به کار برد و از این طریق، قوت و ضعف آزمودنی را در

^۱. single test

آزمون‌های مختلف بررسی کرد (کولن و برنان، ۲۰۰۴). اما در روش مقیاس‌سازی نمره‌ها، دشواری و محتوای آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه لحاظ نمی‌شود، به عبارت دیگر، ویژگی‌های روان‌سنجی آزمون‌ها نادیده گرفته می‌شود که نادیده گرفتن این ویژگی‌ها می‌تواند در نمره کل آزمون - که ترکیبی از نمره‌های آزمون‌های پیشرفت تحصیلی با خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه است - تأثیرگذار باشد. به‌طور کلی، نمی‌توان تصمیم عادلانه‌ای را برای شرکت‌کنندگان در آزمون سراسری ورود به دانشگاه اخذ کرد. به‌ویژه اینکه در آزمون سراسری ورود به دانشگاه، دانش‌آموختگان سال‌های مختلف شرکت می‌کنند، حال آنکه دشواری آزمون‌های پیشرفت تحصیلی آنها با یکدیگر برابر نیست.

- مقیاس‌سازی نمره‌ها باعث تغییر توزیع نمره‌ها می‌شود. پیترسون، کولن و هوور^۱ (۱۹۸۹) چنین معتقدند؛ دلایل نظری خوبی برای به‌کارگیری تبدیل مقیاس نرمال‌سازی روی نمره‌ها وجود ندارد؛ به دلیل اینکه نمره‌های مشاهده شده، دارای توزیع نرمال نیستند و اغلب انتظار می‌رود که توزیع آنها نامتقارن باشد.

- کولن، هنسون و برنان^۲ (۱۹۹۲) نشان دادند که خطای استاندارد اندازه‌گیری شرطی، تحت تأثیر مقیاس نمره‌ها قرار می‌گیرد. در روش تبدیل نمره‌های خام آزمون‌های پیشرفت تحصیلی به نمره‌های مقیاس (تراز) خطای استاندارد اندازه‌گیری شرطی در نظر گرفته نشده است.

بنابراین، با استفاده از یکی انواع پیوند زدن بین نمره‌های آزمون‌های پیشرفت تحصیلی به نمره‌های آزمون سراسری ورود به دانشگاه می‌توان چنین مشکل‌هایی را برطرف کرد؛ به‌طوری که توزیع نمره‌های اولیه تغییر نیابد و دشواری و محتوای آزمون‌ها نیز در نظر گرفته شوند و سپس بر اساس نتایج حاصل، نمره کل دانش‌آموزان

1. Peterson, Kolen, & Hoover

2. Kolen, Hanson, & Brennan

در هر زیرگروه محاسبه می‌شوند تا بر اساس آن بتوانیم یک تصمیم بهتر درباره پذیرش آنها به دانشگاه اخذ کنیم.

از آنجایی که طیف وسیعی از چارچوب‌های پیوند زدن توسط پژوهشگران بیان شده است، بنابراین یک شرح کلی از آنها ارائه می‌شود. کولن (۲۰۰۴) این چارچوب‌های مختلف را به صورت زیر بیان کرده است.

میسلوی^۱ (۱۹۹۲) و لین^۲ (۱۹۹۳) چارچوبی از پیوند زدن را ارائه کردند که بر اساس ویژگی‌های آماری، دشواری، پایایی و شرایط اجرای یکسان به چهار رده همترازسازی، مدرج‌سازی^۳، تصویرسازی^۴ و تعدیل‌سازی^۵ تقسیم می‌شوند.

کولن و برنان (۲۰۰۴) پیوند زدن را بر اساس چهار ویژگی؛ یعنی میزان تشابه استنباط‌ها، سازه‌ها، جامعه‌ها و شرایط اندازه‌گیری، به انواع همترازسازی، مقیاس‌سازی عمودی^۶، تطابق^۷، تصویرسازی و تعدیل آماری^۸ تقسیم کردند.

دوران^۹ (۲۰۰۴) سه نوع از پیوند زدن نمره‌ها را بر اساس شباهت محتوای آزمون‌ها، به صورت همترازسازی، تطابق و پیش‌بینی عنوان کرد.

هالند و دوران^{۱۰} (۲۰۰۶)، چارچوبی از تعمیم تقسیم‌بندی میسلوی (۱۹۹۲) و لین (۱۹۹۳) را ارائه دادند و پیوند زدن را در سه طبقه همترازسازی، مقیاس‌سازی و تطابق بیان کردند.

کولن (۲۰۰۴) به طور خلاصه بیان کرده که در تمامی چارچوب‌های پیوند زدن، فرایند همترازسازی بین نمره‌های فرم‌های آزمون‌ها وجود دارد که این فرایند بین فرم‌هایی به کار برده می‌شود که دارای ویژگی‌های آماری و محتوای یکسان هستند و به

1. Mislevy

2. Linn

3. calibration

4. projection

5. moderation

6. vertical scaling

7. concordance

8. statistical moeration

9. Dorans

10. Holland & Dorans

علت سازه یکسان فرم‌ها پس از همترازسازی، نمره‌ها را می‌توان به‌طور تعویض‌پذیر مورد استفاده قرار داد. همچنین، فرایند همترازسازی تحت تأثیر گروه آزمودنی‌ها قرار نمی‌گیرد و به عبارت دیگر، با تغییر گروه آزمودنی‌ها این روابط همترازسازی بین نمره‌های فرم‌ها تغییر نخواهد کرد.

مدرج‌سازی اشاره به فرایند قرار دادن نمره‌های آزمون‌ها روی یک مقیاس مشترک است، به‌گونه‌ای که این آزمون‌ها دارای سازه‌های یکسان، اما از لحاظ دشواری و پایایی با یکدیگر متفاوت باشند. دو حالت از مدرج‌سازی معرفی شده است؛ یک حالت از مدرج‌سازی، پیوند زدن نمره‌های یک ویرایش کوتاه از یک آزمون به نمره‌های ویرایش طولانی‌تر آن است که نسبت پرسش‌ها در هر رده محتوایی ویرایش کوتاه‌تر برابر با ویرایش بلندتر باشد. حالت دیگر مدرج‌سازی، شامل فرم‌های آزمون‌ی است که برای سطوح متفاوتی کلاسی طراحی شده‌اند که این نمره‌ها روی یک مقیاس مشترک قرار داده می‌شوند. به این نوع از مدرج‌سازی، مقیاس‌سازی عمودی گفته می‌شود. به‌طور مثال، مدرج‌سازی نمره‌های آزمون پیشرفت تحصیلی ریاضی از کلاس سوم تا هشتم.

هنگامی که فرایند پیوند زدن در بین نمره‌های آزمون‌هایی رخ می‌دهد که سازه‌های مختلف را اندازه‌گیری می‌کنند، اصطلاح تصویرسازی یا پیش‌بینی برای پیوند زدن نمره‌های چنین آزمون‌هایی به کار می‌رود. با استفاده از روش‌های رگرسیون این پیوند زدن انجام می‌شود، اما در صورتی که این پیوند زدن به مشابه شدن ویژگی‌های توزیعی نمره‌های آزمون‌ها منجر شود، اصطلاح تعدیل‌سازی آماری به کار برده می‌شود. از چارچوب‌های بحث شده برای پیوند زدن، بعضی از پژوهشگران، پیوند زدن نمره‌های آزمون‌هایی که سازه‌های مشابه را اندازه‌گیری می‌کنند از آزمون‌هایی که سازه‌های متفاوت را اندازه‌گیری می‌کنند، متمایز کرده‌اند و واژه تطابق را برای پیوند زدن بین نمره‌های چنین آزمون‌هایی به کار برده‌اند.

در این پژوهش، از تقسیم‌بندی پیوند زدن دوران (۲۰۰۴) استفاده می‌شود. به همین منظور، ابتدا این سه نوع پیوند زدن به‌صورت مختصر توضیح داده می‌شوند.

همترازسازی: قوی‌ترین نوع پیوند زدن بین نمره‌های آزمون‌ها با محتوا و پایایی یکسان است. روابط به کار برده شده بین چنین نمره‌هایی، دارای ویژگی متقارن و نامتغیر بودن (روابط همترازسازی تحت تأثیر گروه آزمودنی‌ها قرار نگیرد) هستند.^۱ در این پژوهش، آزمون‌های پیشرفت تحصیلی و خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه محتوای کاملاً یکسان ندارند؛ پس این نوع پیوند زدن نمی‌تواند بین نمره‌های آزمون‌ها به کار برده شود.

تطابق: نوع دوم از پیوند زدن، بین نمره‌های آزمون‌هایی است که سازه‌های آنها تقریباً مشابه است. این روش، نیازمند همبستگی بالا بین نمره‌های آزمون‌هاست. قابل ذکر است، طرح گردآوری داده‌ها و روش‌های آماری در تطابق، همانند همترازسازی است و روابط بین نمره‌ها در تطابق نیز متقارن هستند؛ اما به دلیل اینکه سازه‌ها یکسان نیستند، همانند روش همترازسازی نمره‌ها نمی‌توانند به‌طور تعویض‌پذیر به کار برده شوند؛ پس نتایج حاصل از تطابق همانند همترازسازی استفاده نمی‌شوند. در تطابق، از توزیع نمره‌های دو آزمون برای تطبیق دادن آنها استفاده می‌شود (پمیریچ، هنسون، هریس و اسکاینینگ،^۲ ۲۰۰۴). تفاوت دیگر تطابق با همترازسازی عدم برقراری نامتغیر بودن است؛ بدین معنا که روش پیوند زدن مورد نظر، تحت تأثیر گروه آزمودنی‌ها قرار می‌گیرد؛ اما می‌توان این میزان وابستگی به گروه را محاسبه کرد. در سال‌های اخیر استفاده از تطابق برای پیوند زدن نمره‌های آزمون‌های مختلف - که ویژگی‌های روان‌سنجی متفاوت اما محتوای نسبتاً مشابه دارند - افزایش یافته است.

پیش‌بینی: این نوع از پیوند زدن، برخلاف همترازسازی و تطابق، بین نمره‌های آزمون‌هایی با سازه‌های کاملاً متفاوت است؛ در نتیجه نمی‌تواند رابطه مقارنی بین نمره‌های آزمون‌ها ایجاد کند. یکی از روش‌های متداول در این نوع، «رگرسیون» است.

1. Invariant

۲. متقارن بودن بدین معناست که؛ اگر تابعی برای تبدیل نمره‌های آزمون X به Y به کار برده شود، عکس آن تابع نیز برای تبدیل نمره‌های آزمون Y به X مورد استفاده قرار می‌گیرد و همچنین نامتغیر بودن بدین معناست که روابط همترازسازی تحت تأثیر گروه آزمودنی‌ها قرار نگیرد.

3. Pommerich, Hanson, Harris, & Scanning

در این پژوهش، آزمون‌های پیشرفت تحصیلی و خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه از لحاظ ویژگی‌های آماری با یکدیگر یکسان نیستند، اما شباهت محتوایی بین سازه‌های مورد اندازه‌گیری در هر دو آزمون وجود دارد. بنابراین از روش تطابق برای پیوند زدن بین نمره‌های این دو آزمون باید استفاده شود؛ به همین دلیل، به‌طور مختصر، به پیشینه پژوهشی برای پیوند زدن آزمون‌هایی با ویژگی‌های مشابه اشاره می‌شود.

مارکو، عبدالفتاح و بارون^۱ (۱۹۹۲)، نمره‌های SAT و ACT را از طریق تطابق به یکدیگر، مرتبط کردند و اظهار داشتند نتایج حاصل از تطابق، تحت تأثیر نمونه‌های به کار گرفته شده است. دوران، لیو، پمریچ و هوستون^۲ (۱۹۹۷) به پیوند زدن نمره‌های مجموع SAT و ACT اقدام کردند. آنها بیان کردند که به علت تفاوت در محتوا و ویژگی‌های دو آزمون، روش مناسب برای برقراری رابطه بین آنها تطابق است و ایجاد جدول‌های تطابق بین دو مجموعه آزمون الزامی و ضروری است.

ین و همکاران (۲۰۰۴) در پژوهش خود اظهار کردند از آنجایی که آزمون‌های ITED و ACT محتوا و ویژگی‌های آماری متفاوت دارند و همچنین برای هدف‌های متفاوتی استفاده می‌شوند، برای مرتبط کردن نمره‌های آزمون ITED به ACT از تطابق استفاده شود. آنها برای تصمیم‌گیری بهتر، اقدام به طراحی جدول تطابق بین نمره‌های این دو مجموعه آزمون‌ها کردند، این کار مقایسه بین این نمره‌ها را تسهیل و آسان کرد.

پمریچ و همکاران (۲۰۰۴) پیوند زدن نمره‌های ACT به SAT را با استفاده از تطابق انجام دادند و گفتند که برای تطابق بین نمره‌های حاصل از آزمون‌هایی با ویژگی‌های متفاوت نیز می‌توان از روش‌های همترازسازی استفاده کرد. آنها اشاره کردند که اگرچه می‌توان از همترازسازی برای برقراری ارتباط بین دو آزمون با ویژگی‌های متفاوت استفاده کرد، اما یافته‌های آنها نشان داده است که فرایندها، تصمیم‌گیری و کاربردهای متناسب با همترازسازی را نمی‌توان به همان قطعیت در

1. Macro, Abdel -Fattah & Baron

2. Dorans, Lyu, Pommerich & Houston

3. Iowa Tests of Educational Development (ITED)

مورد تطابق به کار گرفت. آنها همچنین اظهار کردند؛ یکی از راه‌های بررسی تعمیم‌پذیری تطابق، نامتغیر بودن است؛ یعنی نتایج تطابق در گروه‌های مختلف مانند گروه‌های جنسیتی یا گروه‌های نژادی / قومیت چقدر یکسان هستند. در صورتی که نتایج تطابق در میان گروه‌ها خیلی از یکدیگر متفاوت باشند، آنها معتقدند که برای هر گروه باید به صورت جداگانه تطابق انجام گیرد.

تاکنون در ایران، روی داده‌های آزمون سراسری ورود به دانشگاه، پیوند زدن از طریق تطابق صورت نگرفته است. اما مطالعه و پژوهش‌هایی درباره همترازسازی و مزیت‌ها و محدودیت‌های آن انجام گرفته است که به برخی از این پژوهش‌ها اشاره می‌شود:

سلیمی جهرمی (۱۳۷۷)، روش‌های همترازسازی و کاربرد آنها در همترازسازی آزمون‌های ورودی دانشگاه‌های کشور را بررسی کرد. در پژوهش وی، نمره‌های آزمون‌های مربوط به مرحله دوم آزمون سراسری ورود به دانشگاه‌ها در سال ۷۵ و ۷۶ (شرکت‌کنندگان نظام قدیم متوسطه)، با استفاده از روش‌های خطی^۱ و همصدک‌سازی^۲ مورد همترازسازی قرار گرفتند^۳ و نتایج گویای این بود که روش همصدک‌سازی مناسب‌تر از خطی است، به‌ویژه اینکه هر چه شکل توزیع فراوانی دو فرم بیشتر از یکدیگر متفاوت باشند استفاده از روش همصدک‌سازی ضروری‌تر است. فرجی (۱۳۸۹)، همترازسازی نمره‌های فرم A و B آزمون هوشی کتل (مقیاس ۳) در دانش‌آموزان پیش‌دانشگاهی شهر سقز را انجام داد تا بر اساس آن بتواند معادل هر یک از نمره‌های هوشی فرم A را از روی فرم B پیدا کند.

رضوانی‌فر (۱۳۹۱)، نمره‌های درس‌های ریاضی و فیزیک رشته علوم تجربی آزمون سراسری ورود به دانشگاه سراسری سال‌های ۱۳۸۸ و ۱۳۸۹ بر اساس روش‌های همترازسازی با یکدیگر مورد مقایسه قرار داد و معادل هر یک از نمره‌های درس‌های ریاضی و فیزیک سال ۱۳۸۸ را نسبت به سال ۱۳۸۹ ارائه کرد.

1. linear

2. equipercentile

۳. در قسمت تحلیل داده‌ها این روش‌ها توضیح داده شده‌اند.

مقدم‌زاده (۱۳۹۲)، نیز به همترازسازی نمره‌های آزمون‌های مهارت زبان انگلیسی در ایران همچون تولیمو^۱ اقدام کرد. از آنجایی که در این آزمون‌ها فرم‌های مختلفی از یک آزمون تهیه می‌شود که دشواری‌های متفاوت دارند و در زمان‌های متفاوتی هم اجرا می‌شوند، پس برای مقایسه‌پذیر کردن چنین نمره‌هایی، از روش همترازسازی استفاده شد. به عبارت دیگر، با استفاده از این روش، نمره‌های حاصل از فرم‌های مختلف را به نمره‌های یکی از فرم‌ها یا به نمره‌های یک فرم مرجع به گونه‌ای تبدیل کرد که گویی همه افراد یک فرم را امتحان داده‌اند و نمره‌های همتراز شده تحت تأثیر تفاوت درجه دشواری فرم‌ها قرار نگرفته‌اند. همچنین در این پژوهش، عواملی که می‌تواند بر روش‌های همترازسازی تأثیر داشته باشد، بررسی شد.

بنابراین به علت اینکه هیچ پژوهشی روی نمره‌های آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه، انجام نگرفته است؛ این پژوهش، با هدف‌های زیر اجرا شد:

- انتخاب شیوه پیوند زدن مناسب بین نمره‌های آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه
- مقایسه روش ارائه شده با روش مورد استفاده در سازمان سنجش آموزش کشور با استفاده از خطای استاندارد اندازه‌گیری شرطی

روش پژوهش

در این پژوهش، به دنبال روشی برای پیوند زدن بین نمره‌های سوابق تحصیلی هر درس با خرده‌آزمون متناظر آزمون سراسری ورود به دانشگاه بودیم؛ به طوری که باعث تغییر نیافتن توزیع نمره‌ها همراه با در نظر گرفتن ویژگی‌های روان‌سنجی و دارای خطای استاندارد اندازه‌گیری شرطی کمتری نسبت به روش مورد استفاده سازمان سنجش آموزش کشور (روش تراز کردن نمره‌ها) شد. این روش از روش‌های توصیفی غیرآزمایشی محسوب می‌شود. از آنجایی که در این پژوهش، برای پیوند زدن بین نمره‌های آزمون‌های پیشرفت تحصیلی با نمره‌های آزمون سراسری ورود به

^۱, TOLIMO

دانشگاه، از بانک اطلاعاتی سازمان سنجش آموزش کشور استفاده شد، پس روش این پژوهش، تحلیل داده‌های ثانویه نیز هست.

جامعه آماری و گروه نمونه

به‌منظور دستیابی به هدف‌های این پژوهش، از نمونه ۵۰۰۰ تایی از شرکت‌کنندگان گروه علوم ریاضی و فنی آزمون سراسری ورود به دانشگاه سال ۱۳۹۴ در سازمان سنجش آموزش کشور به‌صورت تصادفی ساده انتخاب شد.^۱ از این تعداد، ۳۹۴۳ نفر دانش‌آموزانی بودند که آزمون‌های پیشرفت تحصیلی‌شان در خردادماه سال ۱۳۹۳ به‌صورت جداگانه و در سرتاسر کشور برگزار شده بود. با توجه به اینکه یک گروه ثابت از دانش‌آموزان به هر دو آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه پاسخ داده‌اند، طرح به کار برده شده در این پژوهش به‌منظور مرتبط کردن نمره‌های دو آزمون، طرح تک‌گروهی^۲ است (کولن و برنان، ۲۰۰۴). در این طرح با دو محدودیت اثر ترتیب^۳ و بازه زمانی بین دو آزمون مواجه شدیم که می‌تواند بر نتایج تطابق تأثیرگذار باشد؛ اما به علت ماهیت آزمون‌ها مجبور به نادیده گرفتن آنها شدیم.

داده‌های پژوهش و شیوه تحلیل داده‌ها

داده‌های این پژوهش به شرح زیر از بانک اطلاعات سازمان سنجش آموزش کشور دریافت شد:

الف- نمره‌های خام و تراز شده سوابق تحصیلی دانش‌آموزان رشته ریاضی فیزیک؛ یعنی درس‌های «فیزیک»، «شیمی»، «جبر و احتمال»، «هندسه ۲»، «حسابان»، «فرهنگ و معارف اسلامی»، «زبان فارسی»، «ادبیات فارسی»، «زبان عربی» و «زبان خارجی» سال سوم دبیرستان بودند که در سال ۱۳۹۳ دانش‌آموخته شدند. این آزمون‌ها

۱. به علت مسائل امنیتی سازمان سنجش، امکان دسترسی به نمونه‌های با حجم بیشتر فراهم نبود.

۲. single group

۳. order effect

در خردادماه سال ۱۳۹۳ هر کدام به‌صورت جداگانه و به شکل سراسری در کل کشور برگزار شده‌اند.

ب- نمره‌های خام و تراز شده آزمون سراسری ورود به دانشگاه گروه علوم ریاضی و فنی سال ۱۳۹۴ بودند. این آزمون در اواسط تیرماه سال ۱۳۹۴ اجرا شد و در واقع، آزمون‌ی برای پذیرش در دانشگاه‌ها برای مقطع کارشناسی است. در این آزمون، هر یک از داوطلبان دو دفترچه سؤال دارند که ابتدا دفترچه عمومی و پس از آن، دفترچه تخصصی بین داوطلبان پخش می‌شود. دفترچه عمومی آزمون سراسری ورود به دانشگاه، به ترتیب شامل خرده‌آزمون‌های «زبان و ادبیات فارسی»، «زبان عربی»، «فرهنگ و معارف اسلامی» و «زبان انگلیسی» است و دفترچه تخصصی، به ترتیب شامل خرده‌آزمون «ریاضیات»، «فیزیک» و «شیمی» است.

در این پژوهش، از نرم‌افزار R برای نوشتن کدها و تجزیه و تحلیل داده‌ها استفاده شد. بدین صورت که با استفاده از بسته *equate*، نمره‌های سوابق تحصیلی به نمره‌های آزمون سراسری ورود به دانشگاه مرتبط شدند (آلبانو، ۲۰۱۶) و بر اساس بسته *SEASic*، ارزیابی نامتغیر بودن روش‌های پیوند زدن در بین نمره‌های آزمون‌ها صورت گرفت (هوگینز-مانلی و ویتاکر، ۲۰۱۵). سرانجام، با استفاده از بسته *VGAM*، بین روش ارائه شده در این پژوهش (پیوند زدن نمره خام آزمون‌های پیشرفت تحصیلی به خرده‌آزمون‌های متناظر آزمون سراسری ورود به دانشگاه) و روش مورد استفاده در سازمان سنجش آموزش کشور (تبدیل نمره خام آزمون‌های پیشرفت تحصیلی به نمره تراز) بر اساس معیار خطای استاندارد اندازه‌گیری شرطی مقایسه صورت گرفت (یی، ۲۰۱۸).

مقایسه شباهت محتوا در آزمون‌های مرتبط شده: به نظر دوران (۲۰۰۴) انتخاب یک نوع پیوند زدن مناسب به میزان شباهت محتوا در میان آزمون‌ها بستگی دارد. از این‌رو، ارزیابی دقیق محتوای آزمون‌ها الزامی است و به مواردی از قبیل حیطه‌های هر درس، تعداد پرسش‌ها و حدود نمره باید توجه شود. کولن (۲۰۰۴) روش‌های آماری

1. Albano

2. Huggins-Manley & Whitaker

3. Yee

از قبیل برآورد همبستگی بین نمره‌های مشاهده‌شده، معادلات ساختاری^۱ و تحلیل‌های نظریه تعمیم‌پذیری^۲ را نیز در نظر گرفت، اما همان‌گونه که دوران (۲۰۰۴) بیان کرده است، در صورت در دسترس نبودن نمره تک‌تک سؤال‌ها، نمی‌توان از روش‌های ذکر شده، استفاده کرد؛ بنابراین در پژوهش حاضر، برای بررسی شباهت محتوای بین آزمون‌ها، به‌طور مفصل به ارزیابی حیطه‌های سؤال‌های آزمون‌ها اقدام شد.

قدرت ارتباط بین نمره‌ها در آزمون‌ها: این مرحله، شامل ارزیابی قدرت رابطه بین نمره‌های آزمون‌هایی است که می‌خواهیم نمره‌های آنها با یکدیگر مرتبط شوند. دوران و واکر^۳ (۲۰۰۷) بیان کردند که کمترین راه برای ارزیابی پیوند زدن بین نمره‌های آزمون‌ها محاسبه همبستگی است، حتی در بیشتر مواقع ضریب همبستگی بسیار مهم‌تر از پایایی است. به نظر دوران و هالند (۲۰۰۰) نیز میزان ضریب همبستگی بین نمره‌های آزمون‌ها تعیین‌کننده این است که چه میزان آزمون‌ها سازه‌های مشابه را اندازه‌گیری می‌کنند؟ آنها بر این باور بودند که میزان ضریب همبستگی می‌تواند در کل جامعه زیاد باشد؛ اما در بین زیرجامعه‌ها مقدار متفاوتی داشته باشد؛ بنابراین در این پژوهش، مقدار ضریب همبستگی بین نمره‌های آزمون‌های پیشرفت تحصیلی با آزمون سراسری ورود به دانشگاه برای کل دانش‌آموزان و نیز گروه آزمودنی‌های دختران و پسران محاسبه می‌شود.

نامتغیر بودن: پیوند زدن به رابطه بین نمره‌های حاصل از آزمون‌ها در یک جامعه گفته می‌شود و کیفیت پیوند زدن بین این نمره‌های به خصوصیات جامعه‌ای بستگی دارد که نمره‌ها باید در آن به یکدیگر مرتبط شوند (دوران و واکر، ۲۰۰۷). ایجاد پیوند زدن بدون وابستگی به خصوصیت جامعه را نامتغیر بودن می‌گویند که از لحاظ تئوری امکان‌پذیر است؛ اما در کاربرد و عمل، دستیابی به چنین ویژگی‌ای، اگرچه غیرممکن، اما دشوار است. آنچه مورد توجه است این است که تا چه اندازه پیوند زدن به جامعه وابسته است. از دیدگاه دوران (۲۰۰۴) وجه تمایز همترازسازی با تطابق در ویژگی نامتغیر بودن است. آنگف (۱۹۷۱) نیز معتقد بود بیشتر توابع پیوند زدن، شرط نامتغیر

1. Structural equation models

2. Generalizability theory

3. Dorans & Walker

بودن را نقض می‌کنند؛ اما مطلوب است که میزان وابستگی تابع پیوند زدن در جامعه، به‌اندازه کافی کم باشد. برای بررسی کردن مقدار این وابستگی به جامعه می‌توان از نمره‌های معادل حاصل از پیوند زدن در میان زیر جامعه‌ها استفاده کرد. به گفته دوران (۲۰۰۴) بدیهی‌ترین انتخاب زیرجامعه برای بررسی نامتغیر بودن برحسب جنسیت است؛ بنابراین در این پژوهش، برای بررسی میزان نامتغیر بودن تطابق در نمونه انتخاب شده از دانش‌آموزان، آنها به دو گروه^۱ دختران و پسران تقسیم شدند. در صورتی که نتایج حاصل در گروه‌ها خیلی با یکدیگر متفاوت باشند؛ می‌توان چنین نتیجه گرفت که نتایج حاصل از تطابق به گروه‌ها وابسته است.

در مطالعات زیادی به روش‌های محاسبه میزان و شدت ویژگی نامتغیر بودن اشاره شده است که می‌توان به روش‌های زیر اشاره کرد:

الف- دوران (۲۰۰۴) نامتغیر بودن بین نمره‌های دو آزمون SAT و ACT را بر اساس تفاضل میانگین استاندارد شده^۲ بین دو جنسیت بررسی کرد که به‌صورت زیر محاسبه می‌شود:

$$SMD = \frac{\mu_1 - \mu_2}{\sigma} \quad (2)$$

μ_1 ، μ_2 و σ به ترتیب میانگین دختران، پسران و انحراف معیار کل گروه هستند. مطالعات او نشان می‌دهد که تا حدی در تابع پیوند زدن، وابستگی به گروه وجود دارد.^۳

ب- دوران و هالند (۲۰۰۰) و دوران (۲۰۰۴) جذر میانگین مربع تفاضل^۴ را معرفی کردند. در واقع، این معیار استاندارد شده^۲ جذر میانگین مربع تفاضل‌های توابع

^۱ قابل توجه است، از آنجایی که نمره‌های حاصل از آزمون‌ها برای نمونه انتخاب شده از جامعه دانش‌آموزان دانش‌آموخته رشته ریاضی فیزیک ۱۳۹۳ است، پس در این حالت به جای کلمه «زیرجامعه» از «گروه» استفاده می‌شود.

^۲ Standardized Mean Difference (SMD)

^۳ در صورت در دست داشتن نمونه از جامعه مورد نظر به جای میانگین‌ها و انحراف معیار از برآورد آنها استفاده می‌شود.

^۴ Root means square difference (RMSD)

مرتبط شده در هر زیرجامعه از جامعه کل به ازای هر نمره آزمون است که به صورت زیر محاسبه می‌شود:

$$\text{RMSD}(x) = \frac{\sqrt{\sum_j w_j [e_{p_j}(x) - e_p(x)]^2}}{\sigma_p(Y)} \quad (۳)$$

معادله‌های حاصل از پیوند زدن آزمون X به آزمون Y در زیر جامعه، معادله‌های حاصل از پیوند زدن آزمون X به آزمون Y در جامعه و انحراف استاندارد نمره‌های آزمون Y در جامعه P هستند. همان‌گونه که در کولن (۲۰۰۴) شرح داده شده است، این معادله‌های حاصل از پیوند زدن، می‌تواند بر اساس یکی از روش‌های زیر محاسبه می‌شوند:

- میانگین: در این نوع پیوند زدن فرض بر این است که آزمون X و آزمون Y در تمام طول مقیاس نمره‌ها به اندازه یک مقدار ثابت از نظر دشواری با یکدیگر تفاوت دارند.

دو آزمون X و Y را در نظر بگیرید، اگر x و y به ترتیب دو نمره خاص از این دو آزمون باشند، آنگاه طبق قاعده پیوند زدن از طریق میانگین، نمره معادل x از آزمون X در آزمون Y برای کل جامعه با $m_Y(x)$ نشان داده شده و برابر است با:

$$m_Y(x) = x - \mu(X) - \mu(Y) \quad (۴)$$

به ترتیب میانگین‌های نمره‌های روی آزمون‌های X و Y در نظر گرفته می‌شوند.

معادله‌های پیوند زدن از طریق میانگین در هر زیر جامعه نیز به صورت زیر محاسبه می‌شود:

$$m_{Y_j}(x) = x - \mu_j(X) + \mu_j(Y) \quad (۵)$$

$\mu_j(X)$ و $\mu_j(Y)$ به ترتیب میانگین‌های نمره‌ها روی آزمون‌های X و Y در زیر جامعه J در نظر گرفته می‌شوند.

- خطی: در این روش، فرض می‌شود که نمره‌های استاندارد شده آزمون‌های X و Y با هم برابر هستند؛ یعنی:

$$\frac{x - \mu(X)}{\sigma(X)} = \frac{y - \mu(Y)}{\sigma(Y)} \quad (۶)$$

$\mu(X)$ و $\sigma(X)$ به ترتیب میانگین و انحراف استاندارد نمره‌های آزمون X و $\mu(Y)$ و $\sigma(Y)$ به ترتیب میانگین و انحراف استاندارد نمره‌های آزمون Y در نظر گرفته می‌شوند. بنابراین، تبدیلات خطی به کار گرفته شده برای کل جامعه برابر است با:

$$L_Y(x) = \mu(Y) + \frac{\sigma(Y)}{\sigma(X)} [x - \mu(X)] \quad (۷)$$

تطابق نمره‌های سوابق تحصیلی به آزمون سراسری ورود به دانشگاه برای هر گروه مانند جنسیت نیز به صورت زیر محاسبه می‌شود:

$$L_{Y_j}(x) = \mu_j(Y) + \frac{\sigma_j(Y)}{\sigma_j(X)} [x - \mu_j(X)] \quad (۸)$$

- همصدک‌سازی: در صورتی که به ازای تابع همترازسازی، توزیع نمره‌های تبدیل شده آزمون X به مقیاس آزمون Y با توزیع نمره‌های آزمون Y یکسان باشد، به این شیوه تبدیل «همصدک‌سازی» گفته می‌شود. به عبارت دیگر، تابع همصدک‌سازی نمره‌هایی را روی فرم X مشخص می‌کند که رتبه‌های صدکی آنها (درصد آموذنی‌های که نمره آنها کمتر یا مساوی با نمره مورد نظر باشند) برابر با نمره‌های روی آزمون Y است که همان رتبه صدکی را داشته باشند. تابع تبدیل بین نمره‌های آزمون‌های X و Y به صورت زیر است:

$$eq_Y(x) = G^{-1}[F(x)] \quad (۹)$$

F و G^{-1} به ترتیب تابع توزیع تجمعی نمره‌های X و معکوس تابع توزیع تجمعی Y هستند. معادله‌های پیوند زدن از طریق همصدک‌سازی برای هر زیرجامعه به صورت زیر است:

$$eq_{Y_j}(x) = G_j^{-1}[F_j(x)] \quad (10)$$

برای بررسی نامتغیر بودن، کل دانش‌آموزان به دو گروه دختران و پسران تقسیم شدند. نخست، از معیار تفاضل میانگین استاندارد شده دوران (۲۰۰۴)، نامتغیر بودن بین نمره‌های آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه بررسی شد تا مشخص شود که میزان تفاضل میانگین بین دو گروه نسبت به انحراف استاندارد کل دانش‌آموزان در هر آزمون چه مقدار با یکدیگر تفاوت دارند. سپس، با استفاده از روش‌های میانگین، خطی و همصدک‌سازی، نمره‌های حاصل از پیوند زدن آزمون‌ها محاسبه شدند. علاوه بر آن، با استفاده از جذر میانگین مربع تفاضل، مقایسه بین این روش‌ها صورت گرفت و در هر کدام از روش‌ها که میزان اختلاف بین نمره‌های حاصل کل دانش‌آموزان، دختران و پسران کمتر بود، آن روش برای پیوند زدن بین نمره‌ها انتخاب شد.

خطای استاندارد اندازگیری شرطی: در این پژوهش، بعد از اینکه یک شیوه پیوند زدن مناسب بین نمره‌های آزمون پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه انتخاب شد، با روش مورد استفاده در سازمان سنجش آموزش کشور، یعنی نمره‌های تراز (نرمال‌سازی) مقایسه شد؛ به عبارت دیگر، در این پژوهش بررسی شد که کدام یک از دو روش دارای خطای کمتری است؟

برای محاسبه این خطا، از معیار خطای استاندارد اندازه‌گیری شرطی استفاده شد. یکی از روش‌هایی که برنان و لی^۱ (۱۹۹۹) برای محاسبه این معیار ارائه داده‌اند، بر اساس مفروضات توزیع دوجمله‌ای است. بدین گونه که فرض می‌شود آزمون‌ها دارای k سؤال هستند که به صورت دو وضعیتی (صحیح و غلط) نمره‌گذاری شده‌اند

^۱. Brennan & Lee

و احتمال اینکه یک آزمودنی به Y سؤال از این آزمون پاسخ صحیح دهد برابر است با:

$$p(Y = y|k, \pi) = \binom{k}{y} \pi^y (1 - \pi)^{k-y} \quad y = 0, 1, \dots, k \quad (11)$$

π نسبت واقعی پاسخ‌های درست برای یک آزمودنی لحاظ می‌شود که برآورد آن برابر است با تعداد پاسخ‌های درست به تعداد کل آنها؛ یعنی $\bar{X} = \frac{x}{k}$. قابل توجه است که Y و X به ترتیب برابر با نمره‌ای است که قصد داریم خطای استاندارد اندازه‌گیری شرطی را برای آن محاسبه کنیم، اما Y متغیری با توزیع دوجمله‌ای است که مقادیر ۰ تا k را با احتمال دوجمله‌ای اختیار می‌کند. برنان و لی (۱۹۹۹) خطای استاندارد اندازه‌گیری شرطی به ازای تبدیل هر نمره Y به $f(y)$ برای هر نمره خاص X را به صورت زیر بیان کردند:

$$\hat{\sigma}_{E(S|X)} = c_k \sqrt{\sum_{y=0}^k [f(y)]^2 P(y|\bar{x}, k) - \left[\sum_{y=0}^k [f(y)] P(y|\bar{x}, k) \right]^2} \quad (12)$$

$$c_k = \sqrt{k/(k-1)} \quad (13)$$

c_k میزانی است که به ازای آن $\hat{\sigma}_{E(S|X)}^2$ به یک برآورد نااریب تبدیل می‌شود. بر اساس این معیار، خطای استاندارد اندازه‌گیری شرطی برای هر نمره آزمون پیشرفت تحصیلی که به معادل‌های حاصل از پیوند زدن و نرمال‌سازی تبدیل شده، محاسبه شد.

یافته‌های پژوهش

در این پژوهش، برای انتخاب تطابق بین نمره‌های سوابق تحصیلی دانش‌آموزان رشته ریاضی فیزیک به نمره‌های آزمون سراسری ورود به دانشگاه، سه عامل معرفی شده توسط دوران (۲۰۰۴) به شرح زیر بررسی شد:

۱. **مقایسه شباهت محتوا:** شباهت بین محتوای آزمون‌های مورد نظر، نقش مهمی را در تعیین انتخاب نوع پیوند زدن ایفا می‌کند. بنابراین، نخست برای بررسی مقایسه

محتوای آزمون‌های پیشرفت تحصیلی دبیرستان سال ۱۳۹۳ با آزمون سراسری ورود به دانشگاه ۱۳۹۴ برای گروه ریاضی فیزیک، هر درس، از لحاظ تعداد، نوع و مدت زمان پاسخگویی به پرسش‌ها نظیر به نظیر بررسی شد. این موارد در جدول (۱) به‌طور کامل مشخص شده است.

جدول (۱) مقایسه بین آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه از لحاظ تعداد، نوع و مدت پاسخگویی به پرسش‌ها

آزمون سراسری ورود به دانشگاه				آزمون‌های پیشرفت تحصیلی			
نوع پرسش	مدت زمان (دقیقه)	تعداد سؤال‌ها	درس	نوع سؤال	مدت زمان (دقیقه)	تعداد پرسش‌ها	درس
چهارگزینه‌ای	۱۸	۲۱	ادبیات فارسی	تشریحی	۸۰	۴۹	ادبیات فارسی
		۴	زبان فارسی	تشریحی	۹۰	۲۶ ۲۴	نگارش دستور
چهارگزینه‌ای	۲۰	۲۵	زبان عربی	تشریحی	۹۰	۵۸	عربی
چهارگزینه‌ای	۱۷	۲۵	فرهنگ و معارف اسلامی	تشریحی	۹۰	۲۵	فرهنگ و معارف اسلامی
چهارگزینه‌ای	۲۰	۲۵	زبان انگلیسی	تشریحی	۱۰۰	۶۷	زبان انگلیسی
چهارگزینه‌ای	۸۵	۲۳	حسابان	تشریحی	۱۳۵	۱۷	حسابان
		۱۷	هندسه	تشریحی	۱۳۵	۱۷	هندسه ۲
		۸	جبر و احتمال				
		۷	سایر ^۱	تشریحی	۱۳۵	۱۸	جبر و احتمال
چهارگزینه‌ای	۵۵	۴۵	فیزیک	تشریحی	۱۲۰	۱۸	فیزیک
چهارگزینه‌ای	۳۵	۳۵	شیمی	تشریحی	۱۱۰	۲۵	شیمی

۱. بعضی آزمون‌ها از چند بخش تشکیل شده‌اند که هر بخش دارای تعدادی پرسش است؛ به همین خاطر تعداد پرسش بعضی آزمون‌ها زیاد شده‌اند.

۲. خرده‌آزمون ریاضیات شامل درس‌های دیگر از قبیل؛ آمار و مدل‌سازی، ریاضیات ۲، حساب و دیفرانسل و انتگرال ۱ و ۲، جبر خطی و ریاضیات گسسته است.

پس از توصیف کلی از آزمون‌ها، شباهت محتوای بین آزمون‌ها بررسی شد. به همین منظور، آزمون‌های پیشرفت تحصیلی هر درس و خرده‌آزمون متناظر آزمون سراسری ورود به دانشگاه به شش معلم^۱ که در سال سوم دبیرستان و دوره پیش‌دانشگاهی آن درس مورد نظر سابقه تدریس داشته‌اند، ارائه شد و از آنها درخواست شد که مشخص کنند:

(۱) هر کدام از پرسش‌ها آزمون پیشرفت تحصیلی متعلق به کدام حیطه کتاب درسی است؟

(۲) چه تعداد از پرسش‌ها خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه از منابع سال سوم دبیرستان است؟

(۳) چه تعداد و چند درصد از پرسش‌هایی که از سال سوم در خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه طراحی شده‌اند، با محتوای پرسش‌های آزمون پیشرفت تحصیلی شباهت دارد؟

در درس «زبان انگلیسی»، پس از بررسی نتایج به‌صورت زیر حاصل شد. آزمون پیشرفت تحصیلی زبان انگلیسی از چند بخش تشکیل شده و هر بخش دارای تعدادی پرسش است. آزمون جمعاً شامل ۶۷ پرسش تشریحی و مدت پاسخگویی به آن ۱۰۰ دقیقه است. خرده‌آزمون زبان انگلیسی آزمون سراسری ورود به دانشگاه که به‌عنوان خرده‌آزمونی از آزمون‌های آزمون سراسری ورود به دانشگاه است، از ۲۵ پرسش چهارگزینه‌ای تشکیل شده و مدت‌زمان پاسخگویی آن ۲۰ دقیقه در نظر گرفته شده بود. درجدول (۲)، تقسیم‌بندی محتوای آزمون‌ها بر اساس چهار محتوا: دیکته، گرامر، واژگان و درک مطلب مشخص شد. ۱۸ پرسش از ۲۵ پرسش خرده‌آزمون زبان انگلیسی آزمون سراسری ورود به دانشگاه یعنی، ۳ پرسش گرامر، ۳ پرسش واژگان و ۹ پرسش درک مطلب^۲ دارای محتوای مشترک با آزمون پیشرفت تحصیلی بودند؛ پس

۱. بعضی معلمان محتوای بیش از یک آزمون را بررسی کردند.

۲. قابل ذکر است که بعضی موضوعات بین کتاب‌های درسی دبیرستان مشترک است؛ اما به دلیل اینکه این پرسش‌ها از کتاب‌های غیر از سال سوم دبیرستان مطرح شده است، تعداد آنها در نظر گرفته نشده است.

به عبارت دیگر، می‌توان بیان کرد که در حدود ۶۰ درصد از محتوای این دو آزمون مشترک است.

جدول (۲) بررسی محتوای بین آزمون پیشرفت تحصیلی و خرده‌آزمون

زبان انگلیسی		
تعداد پرسش‌ها		
محتوا	آزمون سراسری ورود به دانشگاه	آزمون پیشرفت تحصیلی
دیکنه	-	۸
گرامر	۴	۱۰
واژگان	۸	۲۰
درک مطلب	۹	۱۲
مابقی مباحث	۴	۱۷
تعداد کل سؤال‌ها	۲۵	۶۷

به صورت کلی، می‌توان چنین گفت که محتوای این دو آزمون تا اندازه‌ای با یکدیگر شباهت دارند. به همین صورت، برای تک‌تک درس‌ها این مراحل انجام گرفت که به علت محدودیت حجم مقاله، از ذکر آنها خودداری شده است. بنابراین، پس از مقایسه محتوای آزمون‌ها می‌توان چنین نتیجه گرفت که با وجود اینکه محتوای آزمون‌ها با یکدیگر کاملاً یکسان نبودند، اما همان‌گونه که در جدول (۳) مشخص شده است تا حدی دارای محتوای شبیه به هم هستند و برخی از پرسش‌هایشان از حیطه‌های مشترک انتخاب شده‌اند.

جدول (۳) درصد همخوانی محتوای بین آزمون‌های پیشرفت تحصیلی و خرده‌آزمون‌ها

درصد همخوانی محتوا	خرده‌آزمون	آزمون پیشرفت تحصیلی
۴۰ درصد	«زبان و ادبیات فارسی»	«ادبیات فارسی» و «زبان فارسی»
۶۸ درصد	«زبان عربی»	«زبان عربی»
۲۴ درصد	«فرهنگ و معارف اسلامی»	«فرهنگ و معارف اسلامی»
۶۰ درصد	«زبان انگلیسی»	«زبان انگلیسی»
۳۳ درصد	«ریاضیات»	«حسابان»، «هندسه ۲» و «جبر و احتمال»
۲۴ درصد	«فیزیک»	«فیزیک»
۳۷ درصد	«شیمی»	«شیمی»

۲. قدرت ارتباط بین نمره‌ها در آزمون‌ها: پس از بررسی شباهت محتوای آزمون‌ها، باید قدرت رابطه نمره‌های آزمون‌های پیشرفت تحصیلی ۱۳۹۳ با آزمون سراسری ورود به دانشگاه ۱۳۹۴ ارزیابی شود. حداقل راه برای ارزیابی پیوند بین نمره‌های آزمون‌ها محاسبه ضریب همبستگی است. همان‌گونه که دوران و هالند (۲۰۰۰) بیان کردند، باید برای گروه‌های دختران و پسران نیز ضریب همبستگی محاسبه شود تا مشخص شود که این مقادیر خیلی با ضریب همبستگی برای کل دانش‌آموزان متفاوت نیست. به همین منظور، ضریب همبستگی پیرسون بین نمره‌های خرده‌آزمون آزمون سراسری ورود به دانشگاه با آزمون‌های پیشرفت تحصیلی هر درس برای کل دانش‌آموزان، گروه دختران و پسران بررسی شد.

جدول (۴)، نشان می‌دهد که این مقادیر برای کل دانش‌آموزان گروه دختران و پسران با یکدیگر تفاوت قابل توجهی ندارند. همچنین این میزان همبستگی تحت تأثیر نمونه منتخب از جامعه دانش‌آموزان است. بنابراین رسیدن به مقادیر ایدئال ممکن است امکان‌پذیر نباشد و این مقادیر برای برقراری پیوند بین نمره‌ها مناسب است.

جدول (۴) ضریب همبستگی بین نمره‌های خرده‌آزمون با آزمون‌های پیشرفت تحصیلی

پسران	دختران	کل دانش‌آموزان	درس
۰/۶۲۰*	۰/۵۶۲*	۰/۵۹۹*	ادبیات و زبان فارسی
۰/۶۱۰*	۰/۵۳۰*	۰/۵۸۰*	زبان عربی
۰/۵۸۰*	۰/۵۴۵*	۰/۵۶۹*	فرهنگ و معارف اسلامی
۰/۷۱۷*	۰/۶۷۹*	۰/۶۹۹*	زبان انگلیسی
۰/۶۱۹*	۰/۵۸۰*	۰/۵۹۳*	ریاضیات
۰/۶۹۴*	۰/۶۲۲*	۰/۶۶۲*	فیزیک
۰/۵۱۱*	۰/۴۴۳*	۰/۴۷۳*	شیمی

*در سطح ۵٪ معنادار هستند.

پمیریچ و همکاران (۲۰۰۴) معتقد بودند که علاوه بر بررسی عامل‌های معرفی‌شده توسط دوران (۲۰۰۴)، باید هدف از استفاده نتایج حاصل از پیوند زدن را نیز در نظر گرفت. همچنین به نظر آنها در صورتی که نتایج حاصل از پیوند زدن به صورت قابل مقایسه در نظر گرفته و نیز توزیع نمره‌هایشان حفظ شوند، روش تطابق برای پیوند زدن بین نمره‌ها معقول است. از آنجایی که هدف، انتخاب روشی در بین نمره‌های

آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه بود، به طوری که نمره‌ها با یکدیگر قابل مقایسه شوند و نیز توزیع آنها حفظ شود، بنابراین برای پیوند زدن بین نمره‌های آزمون‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه از تطابق استفاده شد.

۳. **سنجش نامتغیر بودن تابع پیوند زدن انتخاب شده:** در این پژوهش، طبق شاخص‌های بیان شده دوران (۲۰۰۴)، «تطابق» برای مرتبط کردن نمره‌های آزمون‌های پیشرفت تحصیلی با آزمون سراسری ورود به دانشگاه انتخاب شد. از آنجایی که در تطابق شرط نامتغیر بودن برقرار نیست، پس باید یکی از روش‌های مرتبط کردن نمره‌ها (میانگین، خطی و هم‌مصدک‌سازی) به گونه‌ای انتخاب شود که میزان وابستگی آن به گروه کمتر باشد. پس بر اساس نمونه انتخاب شده از دانش‌آموختگان رشته ریاضی فیزیک سال ۱۳۹۳ در میان گروه‌های دختران و پسران، میزان نامتغیر بودن بر اساس معیارهای زیر بررسی شد:

۳-۱. محاسبه میانگین‌ها و انحراف استانداردها

در نمونه مورد پژوهش، ۳۷ درصد آزمودنی‌ها دختران و ۶۳ درصد پسران بودند. ویژگی‌های توصیفی نمره‌های خرده‌آزمون در آزمون سراسری ورود به دانشگاه و نمره‌های حاصل از آزمون پیشرفت تحصیلی دانش‌آموختگان سال ۱۳۹۳ بررسی شد. نتایج جدول (۵) برای آزمون «زبان انگلیسی» است.

جدول (۵) آمار توصیفی نمره‌های آزمون پیشرفت تحصیلی و خرده‌آزمون زبان انگلیسی برای دختران، پسران و کل دانش‌آموختگان

تعداد	بیشترین نمره	کمترین نمره	برجستگی	چولگی	انحراف استاندارد	میانگین	نمره زبان انگلیسی دانش‌آموختگان
۳۹۴۳	۱۰۰	-۲۰	۰/۰۱۲	۱/۰۳۸	۲۶/۷۷۵	۲۲/۶۸۱	کل
۱۴۴۷	۱۰۰	-۱۸/۶۰	۰/۰۴۱	۱/۰۲۱	۲۶/۴۲۲	۲۳/۳۰۲	دختران
۲۴۹۶	۱۰۰	-۲۰	۰/۰۰۲	۱/۰۴۹	۲۶/۹۷۶	۲۲/۳۲۱	پسران
۳۹۴۳	۲۰	۰	-۰/۷۶۳	-/۰۵۰۶	۳/۹۳۸	۱۴/۴۳۲	کل
۱۴۴۷	۲۰	۴	-۰/۴۵۱	-۰/۷۲۰	۳/۷۴۴	۱۵/۱۶۸	دختران
۲۴۹۶	۲۰	۰	-۰/۸۶۶	-۰/۳۹۱	۳/۹۸۶	۱۴/۰۰۴	پسران

همان‌گونه که در جدول (۵) دیده می‌شود، نمره‌های آزمون پیشرفت تحصیلی دارای چولگی چپ و نمره‌های خرده‌آزمون دارای چولگی راست هستند. بین نمره‌های دختران و پسران تفاوت وجود دارد؛ اما به دلیل مقیاس‌های متفاوت نمره‌های آزمون پیشرفت تحصیلی و خرده‌آزمون به راحتی نمی‌توان بررسی کرد که این تفاوت برای نمره‌های کدام آزمون بیشتر است؛ به همین منظور از تفاضل میانگین‌های استاندارد شده، استفاده شد.

۲-۳. تفاضل میانگین‌های استاندارد شده

کمیت‌های تفاضل میانگین استاندارد شده برای آزمون پیشرفت تحصیلی و خرده‌آزمون در جدول (۶) نشان می‌دهد که تفاوت بین گروه‌های دختران و پسران در نمره‌های سوابق تحصیلی بیشتر است. می‌توان گفت تطابق نمره‌های آزمون پیشرفت تحصیلی ۱۳۹۳ به نمره‌های آزمون سراسری ورود به دانشگاه ۱۳۹۴ نسبت به جنسیت نامتغیر نیستند.

جدول (۶) تفاضل میانگین استاندارد شده آزمون پیشرفت تحصیلی و خرده‌آزمون دانش‌آموختگان

تفاضل میانگین استاندارد شده	خرده‌آزمون	آزمون پیشرفت تحصیلی
زبان و ادبیات فارسی	۰/۱۸۹	۰/۵۴۸
عربی	۰/۱۴۸	۰/۵۲۵
فرهنگ و معارف اسلامی	۰/۱۹۳	۰/۵۸۰
زبان انگلیسی	۰/۰۳۷	۰/۳۰۰
ریاضیات	-۰/۰۷۹	۰/۳۰۲
فیزیک	-۰/۰۷۳	۰/۲۱۸
شیمی	-۰/۰۹۵	۰/۴۰۲

۳،۳. جذر میانگین مربع تفاضل

برای بررسی نامتغیر بودن با استفاده از روش جذر میانگین مربع تفاضل، ابتدا تطابق نمره‌های آزمون‌های زبان انگلیسی پیشرفت تحصیلی به نمره‌های خرده‌آزمون در بین دختران، پسران و کل دانش‌آموختگان ۱۳۹۳ با استفاده از سه روش میانگین، خطی و همصداک‌سازی صورت گرفت که نتایج در جدول (۷) مشخص شده است.

همصدک‌سازی				خطی				میانگین				زبان انگلیسی ۹۳ نمره‌های
جدر میانگین مربع تفاضل	معادل در زیر گروه پسران	معادل در زیر گروه دختران	معادل در کل	جدر میانگین مربع تفاضل	معادل در زیر گروه پسران	معادل در زیر گروه دختران	معادل در کل	جدر میانگین مربع تفاضل	معادل در زیر گروه پسران	معادل در زیر گروه دختران	معادل در کل	
۰/۳۶	۵۰/۰۰	۳۰/۰۰	۴۳/۷۹	۰/۰۶	۴۷/۱۵	۴۳/۶۰	۴۵/۷۴	۰/۱۵	۵۱/۲۱	۴۳/۰۹	۴۸/۳۶	۱۸
۰/۲۶	۶۷/۶۰	۵۴/۰۰	۶۰/۲۵	۰/۰۵	۵۳/۰۶	۵۰/۴۰	۵۱/۸۴	۰/۱۵	۵۷/۸۶	۴۹/۷۴	۵۵/۰۱	۱۹
۰/۰۸	۸۴/۰۷	۸۰/۱۷	۸۳/۸۳	۰/۰۳	۵۸/۹۷	۵۷/۲۰	۵۷/۹۴	۰/۱۵	۶۴/۵۱	۵۶/۳۹	۶۱/۶۶	۲۰

در جدول (۷) دیده می‌شود که هر نمره آزمون پیشرفت تحصیلی زبان انگلیسی با استفاده از سه روش میانگین، خطی و همصدک‌سازی برای کل دانش‌آموزان، گروه دختران و پسران معادل با چه نمره‌ای از خرده‌آزمون زبان انگلیسی آزمون سراسری ورود به دانشگاه است. همچنین در ستون‌های جدر میانگین مربع تفاضل، میزان نامتغیر بودن معادل‌ها برای کل دانش‌آموزان نسبت به گروه جنسیت بررسی شده است؛ به‌طور مثال، نمره صفر آزمون زبان انگلیسی برای کل دانش‌آموختگان، دختران و پسران به ترتیب برابر با $۷۱/۳۴-$ ، $۷۶/۶۱-$ و $۶۸/۴۹-$ است، اما به دلیل اینکه این مقادیر کمتر از حد پایین نمره خرده‌آزمون است، پس در موقع گزارش نمره‌ها به جای آنها $۳۳-$ در نظر گرفته می‌شود^۱ و مقدار جدر میانگین مربع تفاضل برابر با $۰/۱۵$ است.

۱. نمره‌هایی که در جدول (۷) پررنگ شده‌اند، مقادیرشان کمتر از حد پایین نمره خرده‌آزمون است.

همان‌طور که در جدول (۸) مشخص شده، متوسط این کمیت برای روش همصدک‌سازی از روش‌های دیگر کمتر است.

جدول (۸) آمار توصیفی جذر میانگین مربع تفاضل‌های درس زبان انگلیسی برای سه روش پیوند زدن میانگین، خطی و همصدک‌سازی دانش‌آموختگان

تعداد	بیشترین	کمترین	انحراف استاندارد	میانگین	جذر میانگین مربع تفاضل
۲۱	۰/۱۵۰	۰/۱۵۰	۰	۰/۱۵۰	میانگین
۲۱	۰/۳۶۰	۰/۰۳۰	۰/۱۰۰	۰/۲۰۰	خطی
۲۱	۰/۵۷۰	۰	۰/۱۵۰	۰/۱۰۰	همصدک‌سازی

دوران (۲۰۰۴) برای بررسی نامتغیر بودن با استفاده از جذر میانگین مربع تفاضل، از مقدار ثابت «تفاوت قابل اهمیت» استفاده کرد که برابر با نصف واحد نمره گزارش شده است. منطق «تفاوت قابل اهمیت» این است: در صورتی که به ازای هر نمره، تفاوت پیوند زدن حاصل از گروه‌بندی نسبت به کل (بدون گروه‌بندی) کمتر از نصف واحد نمره گزارش شده باشد، این تفاوت قابل چشم‌پوشی است. در این حالت، چون مقیاس نمره‌های آزمون سراسری ورود به دانشگاه به صورت اعداد صحیح در نظر گرفته شده است پس مقدار تفاوت قابل اهمیت برابر با ۰/۵ است. چون در این حالت می‌خواهیم به‌عنوان معیاری برای ارزیابی جذر میانگین مربع تفاضل استفاده شود، مقدار آن استاندارد می‌شود؛ یعنی ۰/۵ تقسیم بر انحراف معیار نمره‌های خرده‌آزمون زبان انگلیسی می‌شود:

$$\frac{0/5}{26/775} = 0/019$$

از مقایسه جذر میانگین مربع تفاضل برای هر نمره در جدول (۸) با تفاوت قابل اهمیت استاندارد می‌توان چنین نتیجه گرفت که تطابق نمره‌های آزمون‌ها با استفاده از روش همصدک‌سازی به‌جز در ۵ نمره ۳، ۱۷، ۱۸، ۱۹ و ۲۰ نسبت به روش‌های دیگر خیلی بهتر است. این و همکاران (۲۰۰۴) نیز معتقد بودند، در صورتی که از لحاظ

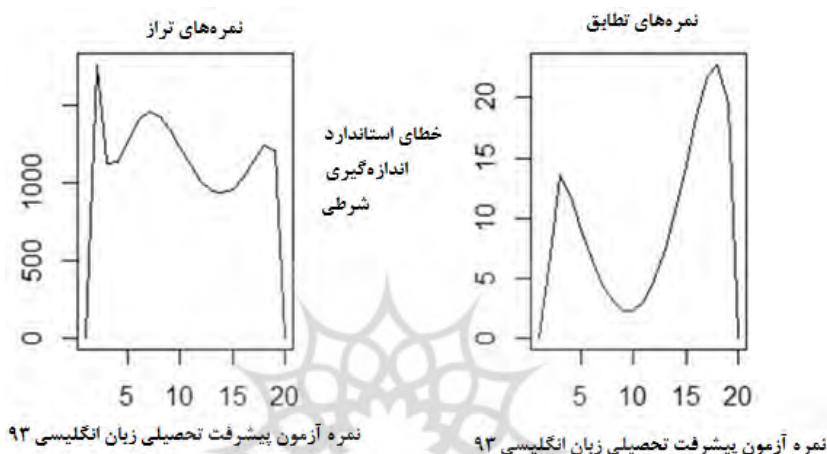
¹. Difference That Matters (DTM)

ویژگی‌های محتوایی، آزمون‌ها خیلی با یکدیگر مشابه نباشند، روش همصدک‌سازی از سایر روش‌ها مناسب‌تر است و برقراری ویژگی نامتغیر بودن با استفاده از این روش، نسبت به روش‌های دیگر بهتر است. بنابراین در پژوهش حاضر، روش مناسب برای تطابق نمره‌های آزمون پیشرفت تحصیلی به خرده‌آزمون زبان انگلیسی آزمون سراسری ورود به دانشگاه روش همصدک‌سازی است.

مقایسه روش سازمان سنجش آموزش کشور با روش تطابق

در این پژوهش، روش همصدک‌سازی از بین روش‌های تطابق برای پیوند زدن نمره‌های آزمون‌های پیشرفت تحصیلی با خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه پیشنهاد داده شد. اکنون این روش با روش مورد استفاده در سازمان سنجش آموزش کشور؛ یعنی نمره‌های تراز (نرمال‌سازی) مقایسه می‌شود. بدون کاسته شدن از کلیت فرض شرط توزیع دوجمله‌ای، پرسش‌ها را به صورت ساختگی و به شکل دووضعیتی در نظر می‌گیریم؛ یعنی هر آزمودنی به پرسش، درست یا غلط پاسخ می‌دهد. از ابتدا برای تفسیر بهتر، نتایج تطابق نمره‌های آزمون پیشرفت تحصیلی را به صورت گرد شده در نظر می‌گیریم و مقادیر k ، \bar{X} ، γ و X در رابطه (۱۲) به ترتیب برابر با کل نمره کسب شده (یعنی ۲۰)، نسبت نمره آزمودنی به کل نمره، γ و X هر دو نمره کسب شده آزمودنی از آزمون هستند. با این تفاوت که X نمره‌ای است که می‌خواهیم به ازای آن خطای استاندارد اندازه‌گیری شرطی را به دست بیاوریم؛ اما γ مقادیر ۰ تا ۲۰ است که هر یک از این مقادیر طبق توزیع احتمالاتی توزیع دوجمله‌ای اختیار می‌شود. بنابراین، مقدار خطای استاندارد اندازه‌گیری شرطی با دو روش تطابق و تراز (نرمال‌سازی) نمره‌ها برآورد می‌شود. در شکل (۱) برآورد خطای استاندارد اندازه‌گیری شرطی برای تبدیل نمره‌های آزمون پیشرفت تحصیلی زبان انگلیسی به هر یک از این دو روش رسم شده است و نشان می‌دهد که خطای استاندارد اندازه‌گیری شرطی با هر دو تبدیل، روی نمره‌های آزمون پیشرفت تحصیلی زبان انگلیسی در حدود پایین و بالا نمره‌های به‌طور نسبی کم و در حدود میانی نمره‌ها نسبتاً دارای مقدار زیاد است. به‌طور کلی، خطای استاندارد اندازه‌گیری شرطی برای نمره‌های تراز خیلی بیشتر از نمره‌های تطابق است.

بنابراین می‌توان چنین نتیجه گرفت که تراز کردن (نرمال‌سازی) برای نمره‌ها، نسبت به تطابق دارای خطای اندازه‌گیری بیشتری است و روش تطابق روش مناسب‌تری است.



شکل (۱) خطای استاندارد اندازه‌گیری شرطی برای نمره‌های حاصل از تطابق و نمره‌های تراز

به همین صورت، تمام مراحل فوق که برای نمره‌های آزمون پیشرفت تحصیلی و خرده‌آزمون زبان انگلیسی آزمون سراسری ورود به دانشگاه اجرا شد، برای درس‌های دیگر نیز انجام گرفت. در بقیه درس‌ها نیز از هم‌صداک‌سازی برای تطابق بین نمره‌های آزمون پیشرفت تحصیلی با خرده‌آزمون متناظر آزمون سراسری ورود به دانشگاه استفاده شد؛ زیرا از لحاظ ویژگی نامتغیر بودن نسبت به دو روش دیگر؛ یعنی میانگین و خطی مناسب‌تر بود. پس از تطابق بین نمره‌های آزمون‌ها، خطای استاندارد اندازه‌گیری شرطی تحت هر دو تبدیل، روی نمره‌های آزمون‌های پیشرفت تحصیلی محاسبه و مشخص شد که همانند زبان انگلیسی، خطای استاندارد اندازه‌گیری شرطی برای نمره‌های تراز (نرمال‌سازی) خیلی بیشتر از نمره‌های تطابق است.

بحث و نتیجه‌گیری

در سازمان سنجش آموزش کشور برای اینکه نمره‌های خام آزمون‌های پیشرفت تحصیلی (سوابق تحصیلی) با نمره‌های خام آزمون سراسری ورود به دانشگاه

مقایسه‌پذیر شوند، از روش تراز کردن (نرمال‌سازی) روی نمره‌های این آزمون‌ها استفاده می‌شود، اما این روش تنها باعث می‌شود حدود نمره‌های حاصل از آزمون‌ها در یک مقیاس یکسان قرار بگیرند. در حالی که ویژگی‌های روان‌سنجی آزمون‌ها از قبیل دشواری و محتوای آزمون‌ها لحاظ نمی‌شود. همچنین این روش باعث تغییر توزیع نمره‌ها به نرمال می‌شود، به عبارت دیگر، توزیع نمره‌ها در این حالت تغییر می‌کنند. به همین دلیل در این پژوهش، به دنبال روش مناسبی برای پیوند زدن نمره‌های آزمون‌های پیشرفت تحصیلی هر درس به نمره‌های خرده آزمون متناظر آزمون سراسری ورود به دانشگاه بودیم؛ به طوری که علاوه بر هم مقیاس‌شدن نمره‌ها، دشواری و محتوای آنها لحاظ شود و در عین حال توزیع نمره‌ها نیز تغییر نکند.

در این پژوهش، مراحل پیوند زدن نمره‌های زبان انگلیسی آزمون پیشرفت تحصیلی به خرده‌آزمون انگلیسی آزمون سراسری ورود به دانشگاه به‌طور مفصل و برای نمره‌های دیگر به‌طور مختصر بیان شد. انواع پیوند زدن بر اساس دوران (۲۰۰۴)؛ یعنی همترازسازی، تطابق و پیش‌بینی در نظر گرفته شد و پیوند مناسب بر اساس سه ابزار معرفی شده دوران (۲۰۰۴)؛ یعنی شباهت محتوای آزمون‌ها، قدرت رابطه بین نمره‌های آزمون‌ها و نامتغیر بودن انتخاب شد. برای بررسی شباهت محتوا، از معلم‌هایی که در سال سوم دبیرستان و پیش‌دانشگاهی سابقه تدریس داشتند، درخواست شد که محتوای آزمون پیشرفت تحصیلی سال سوم دبیرستان زبان انگلیسی را مشخص کنند. سپس، قدرت رابطه نمره‌های این دو آزمون با استفاده از ضریب همبستگی پیرسون به دست آمد که تقریباً اندازه‌ای معقول برای ایجاد پیوند زدن بین نمره‌های بود. بنابراین چنین نتیجه گرفته شد که چون محتوای آزمون‌ها تا اندازه‌ای مشابه‌اند و همچنین قدرت رابطه بین نمره‌ها مناسب است، پس می‌توان از تطابق برای پیوند زدن بین آزمون‌ها استفاده کرد.

برای بررسی نامتغیر بودن نتایج حاصل از تطابق، از روش‌های کمیت‌های توصیفی، تفاضل میانگین‌های استاندارد شده و جذر میانگین مربع تفاضل در بین دو گروه دختران و پسران استفاده شد و چنین نتیجه گرفته شد که: روش همصدک‌سازی برای تطابق نمره‌های زبان انگلیسی آزمون پیشرفت تحصیلی به نمره‌های خرده‌آزمون در آزمون سراسری ورود به دانشگاه از لحاظ ویژگی نامتغیر بودن بهتر است. اما از آنجایی که نتایج حاصل از تطابق برای کل آزمودنی‌ها با گروه‌های دختران و پسران متفاوت است؛

باید تصمیم‌گیری شود که یک جدول تطابق برای کل آزمودنی‌ها به کار برده شود یا جدول‌های تطابق به ازای هر گروه.

ین و همکاران (۲۰۰۴) معتقد بودند به کارگیری جدول‌های جداگانه از نظر متخصصان روان‌سنجی مطلوب‌تر است؛ اما باعث سردرگمی کاربران می‌شود. بنابراین اظهار کردند که اگر بخواهیم جدول تطابق یکسانی برای گروه‌های متفاوت به کار ببریم، باید تابع‌های تبدیل برای هر گروه تا حد امکان یکسان باشد که این انتخاب روش‌های آماری برای تطابق، وابسته به ماهیت آزمون‌ها و هدف از انجام تطابق است. در صورتی که هدف از انجام تطابق برای تفسیر آزمون‌های وابسته به هنجار باشد، آنگاه روش همصدک‌سازی بهتر است؛ اما در صورتی که هدف از تفسیر نمره‌ها بر اساس نمره برش باشد، آنگاه روش‌های خطی مناسب‌تر هستند. در این پژوهش، به دلیل اینکه تفسیر نمره‌ها به صورت وابسته به هنجار بود، یک جدول تطابق ارائه شد.

پس از مشخص کردن یک روش مناسب برای پیوند زدن بین نمره‌ها، این روش با روش سازمان سنجش آموزش کشور؛ یعنی تبدیل نمره‌های خام زبان انگلیسی به نمره‌های تراز (نرمال‌سازی) با استفاده از خطای استاندارد اندازه‌گیری شرطی برنان و لی (۱۹۹۹) مقایسه شدند و نشان داده شد که خطای استاندارد اندازه‌گیری شرطی برای روش ارائه شده در این پژوهش بسیار کمتر است.

به‌طورکلی، می‌توان نتیجه گرفت که با استفاده از تطابق، می‌توان رابطه و نسبت بین دشواری آزمون‌ها و توزیع آنها را علیرغم تفاوت در ویژگی‌های روان‌سنجی حفظ کرد. همچنین روش تطابق دارای خطای استاندارد اندازه‌گیری شرطی کمتری نسبت به نرمال‌سازی نمره‌ها است؛ پس می‌توان تصمیم‌های دقیق‌تر و عادلانه‌تری در مورد داوطلبان اتخاذ کرد.

محدودیت‌های زیادی در این پژوهش وجود داشت که می‌تواند بر دقت نتایج تأثیرگذار باشد؛ اول اینکه، حجم نمونه کم بود که به علت مسائل امنیتی سازمان سنجش آموزش کشور، بیشتر از ۵۰۰۰ نمونه از شرکت‌کنندگان گروه علوم ریاضی و فنی آزمون سراسری ورود به دانشگاه سال ۱۳۹۴ در اختیار گذاشته نشد. محدودیت دیگری که در این پژوهش وجود داشت، در اختیار نداشتن نمره سؤال‌های آزمون‌های پیشرفت تحصیلی بود؛ در صورت در اختیار داشتن این نمره‌ها می‌توانستیم علاوه بر روش‌های همترازسازی نظریه کلاسیک از روش‌های همترازسازی نظریه سؤال-پاسخ هم استفاده

کنیم. همچنین می‌توانستیم شباهت محتوای آزمون‌ها را از روش‌های دیگر از جمله معادلات ساختاری یا تحلیل‌های نظریه تعمیم‌پذیری بررسی کنیم. در این پژوهش، تطابق بین نمره‌های آزمون‌های پیشرفت تحصیلی هر درس به خرده‌آزمون متناظر آزمون سراسری ورود به دانشگاه انجام گرفت؛ شاید بتوان در پژوهش‌های آینده، تطابق ترکیبی از نمره‌های آزمون‌های پیشرفت تحصیلی به ترکیبی از نمره‌های خرده‌آزمون‌های آزمون سراسری ورود به دانشگاه را بررسی کرد. نکته دیگر، در مورد خطای استاندارد اندازه‌گیری شرطی است. در این پژوهش، بدون کاسته شدن از کلیت فرض در نظر گرفتیم نمره سؤال‌های آزمون‌های پیشرفت تحصیلی به صورت دو وضعیتی نمره‌گذاری شده‌اند و همچنین فرض شد، نمره آزمون هر آزمودنی دارای توزیع دوجمله‌ای است. پژوهش‌های آینده می‌توانند تطابق بین نمره‌های پیشرفت تحصیلی و آزمون سراسری ورود به دانشگاه را با دیگر توزیع‌های مفروض برای نمره‌های آزمون پیشرفت تحصیلی بررسی کنند.

منابع

- جهانی‌فر، مجتبی؛ خدایی، ابراهیم؛ یونسی، جلیل و موسوی، امین (۱۳۹۶). نقش هموارسازی در تبدیل غیرخطی نمره‌های خام به نمره‌های مقیاس نرمال. فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، ۷ (۲۰)، ۱۰۳-۱۳۲.
- رضوانی‌فر، شیرین (۱۳۹۱). همترازسازی نمرات دروس ریاضی و فیزیک رشته علوم تجربی آزمون کنکور سراسری سال‌های ۱۳۸۸ و ۱۳۸۹ بر اساس نظریه کلاسیک و جدید اندازه‌گیری. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- سازمان سنجش آموزش کشور (۱۳۹۴). آشنایی با روش نمره کل‌سازی در آزمون سراسری سال ۱۳۹۴. هفته‌نامه خبری و اطلاع‌رسانی سازمان سنجش آموزش کشور (ویژه‌نامه پیک سنجش)، ۱۹-۲۷.
- سلیمی جهرمی، محمدحسین (۱۳۷۷). بررسی روش‌های همترازسازی آزمون‌ها در نظریه کلاسیک و *IRT* و کاربرد آنها در همترازسازی آزمون‌های ورودی دانشگاه‌های کشور. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- فرجی، حیدر (۱۳۸۹). همترازسازی نمره‌های فرم *A* و *B* آزمون هوش کتل (مقیاس ۳) در دانش‌آموزان پیش‌دانشگاهی شهر سقز با استفاده از مدل‌های کلاسیک و *IRT*. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- مقدمزاده، علی (۱۳۹۲). روش بهینه همترازسازی با توجه به ویژگی‌های بومی آزمون‌های ملی ایران: مورد مطالعه آزمون تولیمو و آزمون‌های جامع کنکورهای آزمایشی سازمان سنجش آموزش کشور. رساله دکتری، دانشگاه علامه طباطبایی.
- Albano, D. A. (2016). *Equate: Observed-Score Linking and Equating. R package version 2.0.6*. Retrieved from <https://CRAN.R-project.org/package=equate>
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 508-600). Washington, DC: American Council Education.
- Brennan, R. L. & Lee, W. C. (1999). Conditional scale-score standard errors of measurement under binomial and compound binomial assumptions. *Educational & Psychological Measurement*, 59 (1), 5-24.
- Dorans, N. J. (2004). Equating, concordance and expectation. *Applied Psychological Measurement*, 28 (4), 227-246.
- Dorans, N. J. & Holland, P. W. (2000). Population invariance and the Equatability of Tests: Basic Theory and the Linear Case. *Journal of Educational Measurement*, 37 (4), 281-306.

- Dorans, N. J.; Lyu, C. F.; Pommerich, M. & Houston, W. M. (1997). Concordance between ACT Assessment and Recentered SAT I sum scores. *College & University*, 73 (2), 24-34.
- Dorans, J. N. & Walker, M. E. (2007). Sizing Up Linkages. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 191-198). New York: Springer.
- Holland, P. & Dorans, N. J. (2006). Linking and Equating. In R. L. Brennan (Ed.), *Educational Measurement* (4th e., pp. 187-220). Westport, CT: Praeger Publishers.
- Huggins-Manley, A. C. & Whitaker, D. (2015). *SEAsic: Score Equity Assessment- summary index computation. R package version 0.1*. Retrieved from <https://cran.r-project.org/web/packages/SEAsic/index.html>
- Kolen, M. J.; Hanson, B. A. & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, 29 (4), 285-307.
- Kolen, M. J. (2004). Linking Assessment: Concept and History. *Journal of Applied Measurement*, 28 (4), 219-226.
- Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (3th Ed.). New York: Springer.
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6 (1), 83-102.
- Marco, G. L.; Abdel-fattah, A. A. & Baron, P. A. (1992). *Methods used to establish score comparability on the enhanced ACT assessment and the SAT* (College Board Rep. No. 92-3). New York: The College Board.
- Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: ETS Policy Information Center.
- Petersen, N. S.; Kolen, M. J. & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed. pp.221-262). New York: Macmillan.
- Pommerich, M.; Hanson, D. J.; Harris, D. J. & Sconning, J. A. (2004). Issues in Conducting Linkages between Distinct tests. *Applied Psychological Measurement*, 28 (4), 247-273.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from <https://www.R-project.org>
- Yin, P.; Brennan, R. L. & Kolen, M. J. (2004). Concordance between ACT and ITED scores from different populations. *Applied Psychological Measurement*, 28 (4), 274-289.