فصلنامه اندازه‌گیری تربیتی
دانشگاه علامه طباطبائی
سال نهم، شماره ۳۵، بهار ۹۸، ص ۷۵ تا ۱۰۶

Quarterly of Educational Measurement
Allameh Tabataba'i University
Vol. 9, No. 35, Spring 2019

مدل‌سازی خطی سلسله‌مراتبی بین آماره‌ی برازش فرد H^T و متغیرهای زمینه‌ای دانش‌آموزان در آزمون ریاضی پایه‌ی هشتم مطالعه تیمز ۲۰۱۵[۱]

پوریا رضاسلطانی*[۲]، ابراهیم خدایی[۳]، جلیل یونسی[۴]، امین موسوی[۵]، علی مقدم‌زاده[۶]

تاریخ دریافت: ۹۷/۰۷/۲۴
تاریخ پذیرش: ۹۸/۰۲/۱۳

چکیده

سنجش برازش فرد در تضمین روایی، عدالت و تفسیر نمره‌های آزمون مفید واقع می‌شود. در این تحقیق با استفاده از آماره‌ی برازش فرد H^T به بررسی الگوی پاسخ آزمون ریاضی پایه‌ی هشتم مطالعه تیمز ۲۰۱۵ دانش‌آموزان کشورهای استرالیا، ایران و جمهوری کره پرداخته شد. جهت بررسی تأثیر متغیرهای زمینه‌ای بر مقدار آماره‌ی برازش فرد دانش‌آموزان، با توجه به ساختار سلسله مراتبی داده‌ها از مدل‌سازی خطی سلسله مراتبی استفاده شد. به استناد ضریب همبستگی درون رده‌ای، ۸۳/۷ درصد از واریانس آماره‌ی برازش فرد H^T در سطح دانش‌آموز و ۱۶/۳ درصد از آن در سطوح مدرسه و کشور است. همچنین با توجه به مدل نهایی خطی سلسله مراتبی بین مقدار آماره‌ی برازش فرد H^T و عوامل سطوح دانش‌آموزان، مدارس و کشورها، صرفاً متغیرهای میانگین پیشرفت تحصیلی کشورها، تأکید مدرسه بر موفقیت تحصیلی دانش‌آموزان، اطمینان در یادگیری ریاضیات دانش‌آموزان و برآورد توانایی دانش‌آموزان از عوامل با ضرایب معنی‌دار در مدل نهایی هستند.

واژگان کلیدی: آماره‌ی برازش فرد H^T، الگوی پاسخ، روایی نمره‌های آزمون، سنجش برازش فرد، مدل‌سازی خطی سلسله‌مراتبی.

---

۱. این مقاله از بخشی از رساله دکتری نویسنده‌ی مسئول مقاله استخراج شده است.

۲.* دانشجوی دکترای سنجش و اندازه‌گیری، دانشگاه تهران، تهران، ایران. p_rsoltani@ut.ac.ir
۳ .دانشیار گروه روش‌ها و برنامه‌های آموزشی، دانشگاه تهران، تهران، ایران.
۴. دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.
۵. استادیار گروه روان‌شناسی تربیتی و آموزش استثنایی، دانشگاه ساسکاچوان کانادا.
۶. استادیار گروه روش‌ها و برنامه‌های آموزشی و درسی، دانشگاه تهران، تهران، ایران.

مقدمه

آزمون‌ها در موقعیت‌های آموزشی، راهنمایی تحصیلی و شغلی توسط معلمان، مشاوران، روان‌شناسان بالینی و سایر متخصصان و مؤسسات برای هدف‌های مختلف مورد استفاده قرار می‌گیرند. روایی آزمون ممکن است به دلیل رفتارهای پاسخ‌دهی نابهنجار به مخاطره بیفتد. سنجش برازش فرد در تضمین روایی و عدالت استفاده و تفسیر نمره‌های آزمون، مفید واقع می‌شود (دِ لا توره و دنگ[1]، ۲۰۰۸). روش‌های سنجش برازش پاسخ‌های سؤال افراد با مدل اندازه‌گیری آزمون، اغلب به‌عنوان آماره‌های برازش فرد[2] تلقی می‌شوند. آماره‌های برازش فرد میزان تطابق بردار پاسخ فرد با الگوهای پاسخ مورد انتظار فرد تحت یک مدل خاص را ارزشیابی می‌کنند. در برخی از متون، الگوی پاسخ غیر قابل انتظار، "نابهنجار[3]" یا "برازش نامناسب[4]" تلقی می‌شوند. برازش نامناسب به عدم تطابق بین الگوهای پاسخ مشاهده‌شده و مورد انتظار افراد تحت یک مدل اندازه‌گیری معین اشاره می‌کند. برای سنجش تطابق الگوی پاسخ فرد با یک مدل اندازه‌گیری از یک آماره‌ی برازش فرد استفاده می‌شود. در اندازه‌گیری آموزشی و روان‌شناسی، آماره‌های برازش فرد متعددی برای شناسایی افراد با الگوی پاسخ نامناسب عنوان شده است (سویی و لی[5]، ۲۰۱۵). به‌طورکلی آماره‌های برازش فرد، افراد را به دو گروه افراد با بردارهای پاسخ با برازش مناسب و افراد با بردارهای پاسخ با برازش نامناسب تقسیم می‌کنند. یکی از مزایای آماره‌های برازش فرد این است که آن‌ها الگوهای پاسخ تک‌تک افرادی که در آزمون شرکت کرده‌اند را تحلیل می‌کنند. علی‌رغم وجود برچسب‌های متفاوت برای پاسخ‌دهی نابهنجار، به‌طورکلی می‌توان آن‌ها را به دودسته کلی تقسیم کرد: دسته اول رفتارهایی که منجر به نمره‌های تصنعی پایین[6] می‌شوند و دسته دوم رفتارهایی که منجر به نمره‌های تصنعی بالا[7] می‌شوند. وقتی‌که در آزمونی هر دو نوع رفتارهای فوق رخ دهد به آن نمره‌های تصنعی آمیخته[8] گویند (راپ[1]، ۲۰۱۳).

1. de la Torre & Deng
2. person fit statistics
3. aberrant
4. misfitting
5. Cui & Li
6. spuriously low scores
7. spuriously high scores
8. spuriously mixed scores

الگوی پاسخ غیر قابل انتظار در برآورد توانایی دانش‌آموزان، پارامتر سؤال‌ها و سایر ویژگی‌های روان‌سنجی آزمون بی‌تأثیر نیست. کونیجن، ایمونز و سیجتسما[2] (۲۰۱۴)، میجر[3] (۱۹۹۷) و اشمیت، کورتینا و ویتنی[4] (۱۹۹۳) بیان کردند روایی و پایایی یک آزمون ممکن است به دلیل وجود پاسخ‌های با برازش نامناسب برخی از افراد (نه همه‌ی افراد) به مخاطره بیفتد.

مطالعه تیمز یکی از ملاک‌های ارزشیابی نظام‌مند درون‌دادها، فرآیندها و برون‌دادهای نظام آموزشی ایران در سطح ملی، منطقه‌ای و جهانی است. برای انجام مطالعه تیمز هزینه‌های زیادی صرف می‌شود، با استفاده از تکنیک تکمیلی برازش فرد در تحلیل نتایج این مطالعه، نتایج قابل‌اعتمادتری برای پارامترهای سؤال‌ها و توانایی دانش‌آموزان حاصل می‌شود. اصلاح در برآورد توانایی دانش‌آموزان با استفاده از تکنیک برازش فرد می‌تواند شاخص‌های متعدد مطالعه تیمز برای کشور ایران و سایر کشورها را تحت تأثیر قرار دهد و ممکن است رده‌بندی آن‌ها را نیز جابجا کند. پاسخ‌های دارای برازش نامناسب می‌تواند نشانگر عواملی باشند که در آزمون در نظر گرفته نشده باشند (ایمونز، سیجتسما و میجر، ۲۰۰۵). سنجش برازش فرد می‌تواند به‌صورت تأییدی یا اکتشافی صورت پذیرد (کونیجن، سیجتسما و ایمونز، ۲۰۱۶). بررسی عمیق‌تر الگوی پاسخ دانش‌آموزان به‌منظور شناسایی عوامل زمینه‌ای مختلف تأثیرگذار بر الگوی پاسخ دانش‌آموزان نیز قابل تأمل است. هدف کلی این تحقیق شناسایی متغیرهای زمینه‌ای مؤثر بر الگوی پاسخ دانش‌آموزان است. هدف اختصاصی این تحقیق با توجه به ساختار سلسله‌مراتبی داده‌ها، مدل‌سازی خطی سلسله‌مراتبی بین مقدار آماره‌ی برازش فرد دانش‌آموزان و متغیرهای زمینه‌ای آن‌ها است.

مبانی نظری و مروری بر مطالعات گذشته

آماره‌های برازش فرد را می‌توان با دو رویکرد اصلی طبقه‌بندی کرد: آماره‌های برازش فرد گروه- مبنا[5] و آماره‌های برازش فرد IRT[6]- مبنا. در رویکرد اول (گروه- مبنا)، آماره‌های برازش فرد بدون در نظر گرفتن مدل اندازه‌گیری خاصی و از مقایسه‌ی الگوی پاسخ

1. Rupp
2. Conijn, Emons & Sijtsma
3. Meijer
4. Schmitt, Cortina & Whitney
5. group-based
6. Item Response Theory

مشاهده‌شده یک فرد بر اساس مشخصه‌های نمره کل آزمون از روی کل نمونه محاسبه می‌شوند (کاراباتسوس[1]، ۲۰۰۳). یک آماره‌ی برازش فرد گروه- مبنا، یک الگوی پاسخ مشاهده‌شده را تحت عنوان برازش نامناسب طبقه‌بندی می‌کند؛ اگر سؤال‌های آسان به‌طور نادرست پاسخ داده شوند و سؤال‌های دشوار به‌طور درست پاسخ داده شوند (میجر و سیجتسما[2]، ۲۰۰۱). اگر نمره‌ی فردی در یک آزمون r باشد، انتظار می‌رود که فرد r تا از آسان‌ترین سؤال‌ها را به‌طور درست پاسخ دهد. یک بردار پاسخ به‌عنوان برازش نامناسب تلقی می‌شود وقتی‌که سؤال‌های به‌طور نسبی با نسبت پاسخ صحیح پایین، به‌طور درست پاسخ داده شوند ولیکن سؤال‌های به‌طور نسبی با نسبت پاسخ صحیح بالا به‌طور نادرست پاسخ داده شوند. نمونه‌هایی از آماره‌های برازش فرد گروه- مبنا شاخص احتیاط تعدیل شده C هارنیش و لین[3] (۱۹۸۱)، شاخص U3 ون در فلایر[4] (۱۹۸۲)، شاخص انطباق هنجار NCI[5] تاتسوکا و تاتسوکا[6] (۱۹۸۳) و ضریب H^T سیجتسما (۱۹۸۶) هستند.

در رویکرد دوم (IRT- مبنا)، آماره‌های برازش فرد، برازش الگوی پاسخ را با یک مدل IRT مشخص از قبیل مدل لوجستیک سه پارامتری می‌سنجند. آماره‌های برازش فرد مبتنی بر مدل با استفاده از برآورد پارامترهای سؤال و توانایی افراد، محاسبه شده و افراد را به دو گروه با برازش مناسب و برازش نامناسب طبقه‌بندی می‌کند. آماره‌های برازش فرد IRT- مبنا به‌طور ویژه برای ارزشیابی برازش نامناسب از روی الگوی پاسخ مشاهده‌شده با مدل IRT به‌وسیله‌ی محاسبه احتمال‌های پاسخ مرتبط با پارامترهای سؤال و پارامتر توانایی افراد طراحی شده‌اند. اگر مطابق مدل IRT، احتمال پاسخ درست فرد بالا باشد، فرضیه این است که فرد باید به سؤال به‌طور درست پاسخ دهد و برعکس. یک برازش نامناسب وقتی رخ می‌دهد که فرضیه فوق توسط داده‌ها تأیید نشود. نمونه‌هایی از آماره‌های برازش فرد IRT- مبنا آماره‌ی U رایت و استون[7] (۱۹۷۹)، آماره‌ی W رایت و مسترز[8] (۱۹۸۲)، آماره‌های

1. Karabatsos
2. Sijtsma
3. Harnisch & Linn
4. van der Flier
5. Norm Conformity Index
6. Tatsuoka & Tatsuoka
7. Wright & Stone
8. Wright & Masters

UW و UB اسمیت[1] (۱۹۸۵)، آماره‌ی l_z دراسگو، لوین و ویلیامز[2] (۱۹۸۵) و آماره‌ی l_z^* اسنیجرز[3] (۲۰۰۱) هستند. با توجه به پژوهش کاراباتسوس (۲۰۰۳)، آماره‌ی برازش فرد H^T، بهترین عملکرد در بین آماره‌های برازش فرد دارد. لذا در این تحقیق از آماره‌ی برازش فرد H^T استفاده و در ذیل به تشریح آن پرداخته شد.

آماره‌ی برازش فرد H^T

اغلب آماره‌های برازش فرد گروه-مبنا (ناپارامتریک)، الگوی پاسخ مشاهده‌شده با الگوی پاسخ مورد انتظار تحت مدل گاتمن[4] (۱۹۴۴) را مورد مقایسه قرار می‌دهند. مدل گاتمن عبارت است از

$$\theta_i < \delta_j \;\leftrightarrow\; P_j(\theta_i) = 0,$$
$$\theta_i \geq \delta_j \;\leftrightarrow\; P_j(\theta_i) = 1$$

که در آن $P_j(\theta_i)$ احتمال پاسخ درست دانش‌آموز i-ام با توانایی θ_i به سؤال j-ام با ضریب دشواری δ_j است. بر اساس این مدل، دانش‌آموزی با سطح توانایی θ قادر خواهد بود به r تا از سؤال‌های با ضریب دشواری کمتر یا مساوی با θ پاسخ درست بدهد و به J-r تا از سؤال‌های با ضریب دشواری بیشتر از θ پاسخ نادرست بدهد و در آن J تعداد سؤال‌هاست. چنین الگوی پاسخی به "الگوی گاتمن " یا "الگوی سازگار " مشهور است. اگر دانش‌آموزی به J-r سؤالی که ضریب دشواری بیشتر از θ دارد، پاسخ درست بدهد و در پاسخ دادن به r سؤال با ضریب دشواری کمتر یا مساوی θ ناموفق باشد، الگوی پاسخ او تحت عنوان "خطا " یا "وارونه " یا "الگوی گاتمن معکوس" در نظر گرفته می‌شود (میجر و سیجتسما، ۲۰۰۱).

آماره‌ی برازش فرد H^T با استفاده از فرمول ذیل محاسبه می‌شود (موسوی[5]، ۲۰۱۵).

$$H^T = \frac{\sum_{i\neq j}\sigma_{ij}}{\sum_{i\neq j}\sigma_{ij}^{max}}$$

که در آن σ_{ij} کوواریانس بین بردارهای پاسخ دانش‌آموزان i و j است و σ_{ij}^{max} ماکزیمم ممکن کوواریانس بین بردارهای پاسخ دانش‌آموزان i و j است. با توجه به فرمول فوق،

1. Smith
2. Drasgow, Levine & Williams
3. Snijders
4. Guttman
5. Mousavi

آماره‌ی برازش فرد H^T همبستگی بین بردار پاسخ مشاهده‌شده‌ی دانش‌آموز i با سایر دانش‌آموزان را می‌سنجد. مقدار آماره‌ی برازش فرد H^T می‌تواند بین ۱± باشد. به عبارت دیگر، آماره‌ی برازش فرد H^T بردار پاسخ یک دانش‌آموز را با بردار پاسخ سایر دانش‌آموزان نمونه مقایسه می‌کند. اگر بردار پاسخ مشاهده‌شده‌ی یک دانش‌آموز، سازگار با الگوهای پاسخ سایر دانش‌آموزان باشد، آنگاه صورت کسر آماره‌ی برازش فرد H^T و درنتیجه مقدار آماره‌ی برازش فرد H^T مثبت می‌شود. به همین ترتیب، اگر بردار پاسخ مشاهده‌شده‌ی یک دانش‌آموز، سازگار با الگوهای پاسخ سایر دانش‌آموزان نباشد، آنگاه صورت کسر آماره‌ی برازش فرد H^T و درنتیجه مقدار آماره‌ی برازش فرد H^T منفی می‌شود. بردار پاسخ تصادفی که با سایر الگوهای پاسخ ناهمبسته باشد، منجر به مقدار صفر برای آماره‌ی برازش فرد H^T می‌شود (میجر و سیجتسما، ۲۰۰۱).

تبیین الگوهای پاسخ با برازش مناسب

هرچند آماره‌های برازش فرد به الگوهای پاسخ با برازش نامناسب حساس هستند، یافتن این الگوها تبیینی برای این‌که به چه دلیلی این افراد به این نحو خاص پاسخ داده‌اند، وجود ندارد (میجر و سیجتسما، ۱۹۹۵). تحقیقات برازش فرد که بر استفاده‌ی آماره‌های برازش فرد همراه با سایر روش‌های آماری به‌منظور ارائه‌ی برازش نامناسب در زمینه‌های مختلف متمرکز شده‌اند در بخش بعدی ارائه شد.

رادنر، بریسی و اسکیج[1] (۱۹۹۶) داده‌های سنجش ملی پیشرفت تحصیلی (NAEP[2]) سال ۱۹۹۰ را تحلیل کردند و دریافتند که بین الگوی پاسخ و متغیرهای زمینه‌ای دانش‌آموزان رابطه‌ای وجود ندارد. لامپریانو و بویل[3] (۲۰۰۴) درستی اندازه‌گیری را با استفاده از داده‌های آزمون برنامه درسی ملی ریاضی (MNCT[4]) برای دانش‌آموزان اقلیّت قومی و دانش‌آموزانی که به زبان انگلیسی به‌عنوان زبان دوم صحبت می‌کنند، ارزشیابی کردند و دریافتند که این نوع دانش‌آموزان در مقایسه با سایر دانش‌آموزان با احتمال بیشتری پاسخ‌های با برازش نامناسب دارند و در مورد نتایج حاصل از آزمون هشدار دادند.

1. Rudner, Braccy & Skagg
2. National Assessment of Educational Progress
3. Lamprianou & Boyle
4. Mathematics National Curriculum Test

پتریدو و ویلیامز[1] (۲۰۰۷) از مدل‌سازی خطی سلسله‌مراتبی (HLM[2]) برای شناسایی الگوهای پاسخ با برازش نامناسب در یک سنجش ریاضی استفاده کردند و نسبت‌های معنی‌داری از برازش نامناسب فرد ناشی از متغیرهای سطح کلاس و سطح دانش‌آموزان پیدا کردند. یافته‌های آن‌ها نشان داد که همه‌ی متغیرهای سطح دانش‌آموزان به‌استثنای جنسیت (توانایی، زبان، اضطراب و انگیزش) تأثیر معنی‌داری بر الگوی پاسخ دانش‌آموزان دارند؛ همچنین یافته‌ها نشان داد که دانشجویان با توانایی بیشتر، دانشجویانی که با یک زبان دیگر در خانه صحبت می‌کنند، دانش‌آموزان با اضطراب کمتر و دانش‌آموزان با انگیزش کمتر، احتمال بیشتری برای داشتن الگوی پاسخ با برازش نامناسب دارند. آن‌ها همچنین به این نتیجه رسیدند که عامل‌های سطح کلاس از قبیل مشکلات اجرایی عمومی، شیوه‌های اجرایی غیراستاندارد از قبیل تفسیر سؤال‌ها می‌تواند منجر به الگوهای پاسخ با برازش نامناسب شود.

در مطالعه‌ی دودن و دارابی[3] (۲۰۰۹) چهار آزمون شخصیت برای ریاضیات (نگرش نسبت به ریاضیات، سطح اضطراب ریاضیات، سطح اضطراب امتحان و سطح انگیزش برای یادگیری ریاضیات) و روابط آن‌ها با آماره‌ی برازش فرد l_z مورد بررسی قرار گرفته است. نتایج نشان داد، انگیزش برای یادگیری ریاضیات رابطه‌ی قوی‌ای با آماره‌ی برازش فرد دارد. در حالی‌که توانایی اندازه‌گیری شده‌ی دانش‌آموزان به‌وسیله‌ی نمره‌ی کل آزمون تأثیری بر آماره‌ی برازش فرد ندارد، اضطراب ریاضیات و نگرش نسبت به ریاضیات به نظر می‌رسد بر آماره‌ی برازش فرد تأثیر اندکی داشته باشند.

در پژوهش سویی و موسوی (۲۰۱۵) در آزمون پیشرفت تحصیلی کانادا، با استفاده از مدل‌سازی خطی سلسله‌مراتبی به شناسایی عوامل مؤثر سطوح دانش‌آموزان و کلاس آن‌ها بر الگوی پاسخ دانش‌آموزان پرداخته شد. نتایج این پژوهش نشان داد که علی‌رغم معنی‌داری ضرایب عوامل پیشگویی در سطح دانش‌آموز و کلاس، برای مقدار آماره‌ی برازش فرد l_z، مقدار ناچیزی (۳/۶۵ درصد) از تغییرات این آماره‌ی برازش فرد (l_z) توسط این عوامل تبیین می‌شود. مطالعاتی مشابه مطالعات پتریدو و ویلیامز (۲۰۰۷)، لامپریانو و

1. Petridou & Williams
2. Hierarchical Linear Modeling
3. Dodeen & Darabi

بویل (۲۰۰۴)، دودن و دارابی (۲۰۰۹) و سویی و موسوی (۲۰۱۵) می‌تواند بنیان ارزشمندی برای تبیین رفتارهای آزمودن دانش‌آموزان ارائه دهد.

آماره‌های برازش فرد برای مقاصد دیگری نیز مورد استفاده قرار می‌گیرند. برای نمونه، فینکلمن و کیم[1] (۲۰۰۷)، آماره‌های برازش فرد را در رویکرد تنظیم استاندارد BOW[2] به‌کار برده‌اند. در این رویکرد، شرکت‌کنندگان در میزگرد، الگوهای پاسخ واقعی دانش‌آموزان را تحلیل می‌کنند و هر یک از دانش‌آموزان را در یکی از سطوح عملکردی چندگانه طبقه‌بندی می‌کنند؛ درنهایت پس از چندین دور شرکت‌کنندگان در میزگرد، مقادیر برش برای مجموعه‌های پاسخ دانش‌آموزان تنظیم می‌کنند. فینکلمن و کیم (۲۰۰۷) استدلال کردند که شمول الگوهای پاسخ با برازش نامناسب ممکن است در رویکرد طبقه‌بندی به علت پاسخ‌دهی تصادفی یا بی‌هدف، روایی رویکرد تنظیم استاندارد را مختل کند؛ بنابراین آن‌ها از آماره‌های برازش فرد برای انتخاب کار دانش‌آموز با الگوی پاسخ با برازش مناسب به‌عنوان خزانه‌ی[3] کار دانشجویان برای BOW استفاده کردند. آن‌ها از آماره‌ی برازش فرد l_z (ترابین و ویس[4]، ۱۹۸۳) و فرم گروه_مبنای پیشنهاد شده‌ی آن که همانند فرم استاندارد شده‌ی U3 است، استفاده کردند (ون در فلایر، ۱۹۸۲). نتایج آن‌ها نشان داد که آماره‌های برازش فرد در کاهش انتخاب دانش‌آموزانی که الگوهای پاسخی داشتند که برای رویکرد BOW ناکافی بودند، موفقیت‌آمیز بودند.

وودس[5] و همکاران (۲۰۰۸) رگرسیون لوجستیک دوسطحی را به‌عنوان روش تشخیصی برازش فرد برای ۱۳ مقیاس شخصیت مورد استفاده برای استخدام در ارتش به‌کار برده‌اند. یافته‌های این پژوهش نشان داد که پنج مقیاس از ۱۳ مقیاس جدید، برای برخی از داوطلبان قوه‌ی تمیز بیشتری داشتند. تحقیقات بیشتر الگوهای پاسخ با برازش نامناسب نشان داد که نابهنجاری در الگوی پاسخ ممکن است به دلیل آسیب‌شناسی شخصیت باشد که این آسیب‌شناسی شخصیت در مراحل اولیه غربالگری تشخیص داده نشده باشد. همچنین بررسی عملکرد افتراقی آزمون شواهدی مربوط به جنسیت، قومیت یا هر دوی آن‌ها را نشان داد. وودس و همکاران (۲۰۰۸) به این نتیجه رسیدند که سنجش برازش فرد می‌تواند

1. Finkelman & Kim
2. Body of Work
3. pool
4. Trabin & Weiss
5. Woods

یک ابزار مفید جهت بهبود اندازه‌گیری روان‌شناسی به‌منظور شناسایی افرادی باشد که آن افراد ممکن است داده‌های (نمره‌های) نامعتبری ایجاد کنند یا این‌که آن‌ها به توجه بیشتری نیاز داشته باشند.

لی‌یو و یو[1] (۲۰۱۱)، از آماره‌های برازش فرد برای تشخیص یادگیری با برازش نامناسب در ارتباط با یادگیری الکترونیکی شخصی استفاده کردند. آن‌ها برآورد توانایی با استفاده از تئوری سؤال- پاسخ را برای تشخیص یادگیری با برازش نامناسب به‌کار گرفتند. سیستم شخصی، مسیر یادگیری برای هر یادگیرنده را بر اساس دشواری دوره و توانایی برآورد شده‌ی یادگیرنده توسعه می‌دهد. پس از هر دوره، برآورد توانایی یادگیرنده از طریق اطلاعات گردآوری شده در طول دوره و مقادیر محاسبه شده برای شاخص‌ها، بهنگام می‌شود. اگر نتایج برازش نامناسب را تشخیص داد، ابتدا یک مشاور مجازی یادگیرنده را به تلاش مجدد تشویق می‌کند، اگر این استراتژی کارساز نبود، در مرحله‌ی دوم یک مشاور انسانی ممکن است برای کمک به چنین یادگیرنده‌ای (یادگیرنده با الگوی با برازش نامناسب) وارد عمل شود. با استفاده از یک طرح آزمایشی، نتایج نشان داد که این سیستم، کارآیی یادگیری را بهتر و تجربه‌ی یادگیری برای گروهی که از سیستم یادگیری الکترونیکی شخصی استفاده کردند، بهبود پیدا می‌کند.

آماره‌های برازش فرد همچنین برای یک آزمون شخصیت به‌منظور سنجش ثبات پاسخ‌دهی به آزمون به‌کار برده شده است. فراندو[2] (۲۰۱۲) آماره‌ی برازش فرد l_z را برای دو اندازه‌گیری شخصیت (روان‌رنجوری[3] و برون‌گرایی[4]) به‌کار برد و (۴ درصد) الگوهای پاسخ با برازش نامناسب را با استفاده از مقدار (۲-) به‌عنوان نقطه‌ی برش هر دو اندازه‌گیری به دست آورد. یافته‌ها نشان داد که تفسیر شخصی[5] برخی سؤال‌ها و اعتمادبه‌نفس پایین افراد منابع اصلی نابهنجاری برای هر دو اندازه‌گیری بودند. فراندو (۲۰۱۲) همچنین استدلال کرد که کاربرد بالقوه‌ی سنجش برازش فرد در تشخیص رفتار پاسخ‌دهی بی‌ثبات در مطالعات بالینی بسیار مرتبط و مهم است.

1. Liu & Yu
2. Ferrando
3. neuroticism
4. extraversion
5. idiosyncratic interpretation

تندیرو[1] و همکاران (۲۰۱۳) از آماره‌های تجمعی (CUSUM[2]) برای تشخیص بی‌ثباتی در آزمودن اینترنتی بدون مراقب (UIT[3]) استفاده کردند. این آماره یک متدولوژی برازش فرد حاصل از کنترل فرآیند آماری است که در شرایطی کاربرد دارد که در آن دنباله‌ای از پاسخ‌ها مانند آزمودن انطباقی کامپیوتری[4] وجود دارد (آرمسترانگ و شی[5]، ۲۰۰۹). همان‌طور که توسط تندیرو و همکاران (۲۰۱۳) ذکر شده است، UIT در حال تبدیل‌شدن به یکی از محبوب‌ترین روش‌ها در انتخاب افراد به دلیل انعطاف‌پذیری آن می‌باشد. با این‌حال، UIT همچنان مشکلاتی از قبیل نقص امنیت کافی یا امکان تقلب را دارد. رایج‌ترین روش غلبه بر نواقص مذکور UIT، اجرای یک آزمون تعقیبی در محیطی امن و تحت نظارت است. تندیرو و همکاران (۲۰۱۳)، آماره‌های CUSUM را به‌منظور شناسایی نقایص UIT و به دست آوردن آزمونی مورد تأیید، به‌کار بردند. آن‌ها از آماره‌ی برازش فرد l_z و آماره‌ی z (گو[6] و دراسگو، ۲۰۱۰)، اندازه‌ای خاص برای شناسایی تقلب در UIT، به‌منظور تشخیص برازش نامناسب استفاده کردند. نتایج نشان داد که متدولوژی CUSUM، رویکردی امیدوارکننده در تشخیص برازش نامناسب در آزمودن اینترنتی بدون مراقب می‌باشد.

مدل‌سازی خطی سلسله‌مراتبی

به‌عنوان مقدمه باید اشاره کرد که به‌طورکلی رفتار افراد در متن «ساختار اجتماعی» تحلیل می‌شود. افراد جدا از محیطشان و به‌اصطلاح معلّق در فضای اجتماعی بی‌شکل بررسی نمی‌شوند. عواملی که در تحلیل و تبیین مسائل مورد استفاده قرار می‌گیرند هم شامل ویژگی‌های افراد مورد مطالعه و هم ویژگی‌های محیط‌های اجتماعی آن‌ها می‌شوند و از خانواده تا دوستان و زمینه‌های اجتماعی وسیع‌تر همانند مدرسه را در بر می‌گیرد (کلمن[7] و همکاران، ۱۹۶۶). در اینجا محیط اجتماعی نه با یک متغیر بلکه با مجموعه‌ای از متغیرها شناسایی خواهند شد که در آن روابط متقابل متغیرهای اجتماعی و نیز روابطی که آن‌ها با

1. Tendeiro
2. Cumulative SUM
3. Unprotected Internet Testing
4. computerized adaptive testing
5. Armstrong & Shi
6. Guo
7. Coleman

متغیرهای فردی دارند از یکدیگر تفکیک می‌گردند (بودون[1]، ۱۳۷۳). به‌طورکلی، در این بررسی‌ها افراد و گروه‌های اجتماعی به‌مثابه یک نظام سلسله‌مراتبی[2] مفهوم‌سازی می‌شوند. چنین نظام‌هایی می‌توانند در سطوح مختلف سلسله‌مراتب مشاهده شوند و درنتیجه ممکن است دارای متغیرهای مشخص و تعریف شده در هر سطح باشند. این امر به تحقیق درباره‌ی تعامل بین متغیرهایی که افراد را توصیف و متغیرهایی که گروه‌های اجتماعی را تشریح می‌کنند منتهی می‌شود. نوعی از تحقیق که با عنوان «تحقیق چندسطحی[3]» از آن یاد می‌شود (هاکس[4]، ۲۰۰۲).

مدل‌های خطی سلسله‌مراتبی (مدل‌های چندسطحی، مدل‌های ضریب تصادفی) در تحقیق‌های علوم انسانی و اجتماعی برای برخورد با شرایطی به کار می‌رود که در آن داده‌ها از موقعیت‌ها یا شرایط خوشه‌بندی شده یا آشیانه‌ای جمع‌آوری می‌شوند که اغلب اوقات منجر به جمع‌آوری داده‌هایی می‌شود که در درون این خوشه‌ها همبستگی وجود دارد. مثالی از این موقعیت که در آن مشاهدات وابسته در درون هر گروه وجود دارد، موقعیت‌های آموزشی است که داده‌ها از دانش‌آموزان جمع‌آوری می‌شود و این دانش‌آموزان در درون هر کلاس آشیانه گرفته‌اند. دانش‌آموزانی که یک کلاس مشترک دارند، دارای محیط مشترکی هستند و احتمالاً معلم یکسانی نیز دارند. داده‌هایی که صرفاً از یک دانش‌آموز در یک کلاس به‌خصوص جمع‌آوری می‌شوند را نمی‌توان کاملاً از داده‌هایی که از دیگر دانش‌آموزان در همان کلاس جمع‌آوری می‌شود مستقل دانست. نادیده گرفتن این وابستگی در تحلیل آماری این داده‌ها اغلب به تأثیرات متورم[5] و خطاهای استاندارد کوچک تصنعی در برآورد پارامترها منجر خواهد شد.

ازنظر سابقه تحلیل داده‌ها گفتنی است که دو رویکرد برای تحلیل متغیرها از سطوح مختلف در یک سطح منفرد وجود دارد که هر دو مورد انتقاد قرار گرفته است. در رویکرد اول تمامی متغیرهای سطوح بالاتر در یک سطح منفرد تجزیه[6] می‌شود؛ یعنی، داده‌های حاصل از سطوح بالاتر به تعداد خیلی بیشتری از واحدها در سطح یک اختصاص می‌یابد.

1. Boudon
2. Hierarchy system
3. Multilevel research
4. Hox
5. Inflated effects
6. disaggregate

در چنین شرایطی است که فرض می‌شود تمامی مقادیر تجزیه‌شده مستقل از یکدیگر هستند که نوعی کج‌فهمی و سوء برداشت است که روایی شواهد حاصل را تهدید می‌کند. در رویکرد دوم، مشاهدات سطح یک در سطح بالاتر ترکیب[1] می‌شوند. درنتیجه، تمامی اطلاعات درون گروه‌ها[2] از دست می‌رود. ارتباط بین متغیرهای ترکیب‌شده می‌تواند قوی‌تر باشد و از روابط بین متغیرهای غیرترکیب‌شده متفاوت باشد. اسنیجرز و بوسکر[3] (۱۹۹۹) چشم‌انداز جامعی از خطاهای آماری بالقوه زمانی که ساختار خوشه‌بندی داده‌ها نادیده گرفته می‌شود ارائه داده‌اند.

مدل‌های آماری بسیاری با نام‌های متعددی به وجود آمده‌اند که از طریق آن‌ها می‌توان سطوح مختلف داده‌ها را مورد تجزیه‌وتحلیل قرار داد. این مدل‌ها واریانس‌های بین آزمودنی و درون آزمودنی را از طریق مدل‌سازی داده‌ها در دو مرحله موردنظر قرار می‌دهد. در مرحله‌ی اول، یک تابع رگرسیونی برای مشاهدات حاصل از هر آزمودنی تعیین می‌شود. هر آزمودنی تابع رگرسیون مربوط به خود را دارد. کوواریت‌های یکسانی برای آزمودنی‌های مختلف مورد استفاده قرار می‌گیرد ولی امکان تغییر ضرایب رگرسیون وجود دارد. در مرحله‌ی دوم، با ضرایب رگرسیون که می‌توانستند تغییر کنند به‌عنوان متغیرهای وابسته تصادفی برخورد می‌شود. این متغیرهای وابسته تصادفی، اثرات رگرسیون تصادفی یا ضرایب تصادفی نامیده می‌شوند. با این شیوه می‌توان عبارت مدل‌های ضرایب تصادفی را تبیین کرد که اغلب در تحقیقات جامعه‌شناختی و اقتصادسنجی مورد استفاده قرار می‌گیرد (دی‌لیو و کرفت[4]، ۱۹۸۶؛ لانگفورد[5]، ۱۹۹۳). به هرحال نام‌های متعددی در پیشینه برای توصیف این مدل‌ها وجود دارد. در تحقیقات آموزشی و جامعه‌شناختی، به این مدل‌ها اغلب "مدل‌های چندسطحی " گفته می‌شود (گلدشتاین[6]، ۲۰۰۳؛ اسنیجرز و بوسکر، ۱۹۹۹). در تحقیقات زیست‌سنجی، اغلب مدل‌ها، مدل اثرات آمیخته یا مدل اثرات تصادفی هستند (لیرد و وار[7]، ۱۹۸۲؛ لانگفورد، ۱۹۸۷). دیگر نام‌های معمول مدل‌های مؤلفه‌های

1. Aggregate
2. Within-group information
3. Snijders & Bosker
4. De Leeuw & Kreft
5. Longford
6. Goldstein
7. Laird & Ware

واریانس[1] (دمپستر، روبین و تیسوتاکاوا[2]، ۱۹۸۱) و مدل‌های خطی سلسله‌مراتبی، رگرسیون خطی سلسله‌مراتبی[3]، مدل ضرایب تصادفی[4]، مدل خطی آمیخته سلسله‌مراتبی[5] است (رادن بوش و بریک[6]، ۲۰۰۲؛ اسینجرز و بوسکر، ۱۹۹۹).

روش تحقیق

به استناد طبقه‌بندی کلی تحقیقات به دو گروه تحقیقات آزمایشی و غیرآزمایشی (توصیفی)، این تحقیق جزو تحقیقات غیرآزمایشی (توصیفی) محسوب می‌شود. تحقیقات توصیفی شامل مجموعه روش‌هایی است که هدف آن‌ها توصیف کردن شرایط یا پدیده-های مورد بررسی است؛ اجرای تحقیق توصیفی می‌تواند صرفاً برای شناخت بیشتر شرایط موجود یا یاری دادن به فرآیند تصمیم‌گیری باشد (سرمد، بازرگان و حجازی، ۱۳۸۴). از منظر تقسیم‌بندی تحلیل‌ها به دو نوع اولیه و ثانویه، تحلیل‌های این تحقیق از نوع تحلیل ثانویه محسوب می‌شوند.

جامعه آماری این تحقیق، شامل تمامی دانش‌آموزان ثبت‌نام کرده در پایه‌ی هشتم کشورهای ایران، جمهوری کره و استرالیا در سال ۲۰۱۵ میلادی (۱۳۹۴ شمسی) است. متوسط سن دانش‌آموزان در زمان آزمون حداقل ۱۳/۵ سال است. دانش‌آموزان ثبت‌نام کرده در مدارس دولتی و غیردولتی، جزء جامعه آماری تحقیق هستند. یک سری از دانش‌آموزان که در مدارس خاصی مشغول به تحصیل هستند، جزء جامعه آماری نیستند. دانش‌آموزانی که در مدارسی مشغول به تحصیل هستند که ازلحاظ موقعیت جغرافیایی، دسترسی به آن‌ها آسان نیست. دانش‌آموزانی که در مدارسی مشغول به تحصیل هستند که تعداد دانش‌آموزان پایه‌ی هشتم آن مدارس خیلی کم (چهار یا کمتر) باشد. دانش‌آموزانی که در مدارسی مشغول به تحصیل هستند که ازلحاظ ساختار، برنامه درسی و ... از سیستم مدارس عمومی کاملاً متفاوت هستند. دانش‌آموزان با ناتوانی عملکردی، معلولیت ذهنی و زبان غیربومی نیز جزء جامعه آماری محسوب نمی‌شوند (مارتین، مولیس و هوپر[7]، ۲۰۱۶).

1. Variance Component Models
2. Dempster, Rubin & Tsutakawa
3. Hierarchical Linear Regression
4. R&om Coefficients Model
5. Hierarchical Mixed Linear Model
6. Raudenbush & Bryk
7. Martin, Mullis & Hooper

در مطالعه تیمز به دلیل ماهیت ذاتی سلسله‌مراتبی دانش‌آموزان در مدارس، از روش نمونه‌گیری خوشه‌ای دومرحله‌ای استفاده می‌شود. در مرحله‌ی اول نمونه‌گیری، مدارس با احتمال متناسب با اندازه (PPS[1]) از لیست که شامل دانش‌آموزان واجد شرایط هستند، انتخاب می‌شوند. در مرحله‌ی دوم نمونه‌گیری، یک یا چند کلاس هشتم از هر یک از مدارس انتخاب شده در مرحله‌ی اول نمونه‌گیری انتخاب می‌شود. نمونه‌گیری از کلاس‌ها به‌وسیله‌ی هماهنگ‌کننده‌های ملی تحقیق (NRC[2]) با استفاده از نرم‌افزار نمونه‌گیری درون مدرسه (WinW3S) تولید شده توسط IEA[3] DPC[4] و مرکز آمار کانادا انجام می‌شود (مارتین و همکاران، ۲۰۱۶).

ازآنجایی‌که مطالعه تیمز اساساً یک مطالعه پیشرفت تحصیلی دانش‌آموزان است، دقت برآوردهای پیشرفت تحصیلی دانش‌آموزان از اهمیت بالایی برخوردار است. برای تأمین کردن استانداردهای مطالعه تیمز در مورد دقت نمونه‌گیری، تعداد نمونه‌های هر کشور باید طوری باشد که خطای استاندارد[5] میانگین نمره پیشرفت تحصیلی دانش‌آموزان آن کشور بزرگ‌تر از ۰/۰۳۵ انحراف استاندارد[6] نمره پیشرفت تحصیلی دانش‌آموزان آن کشور نباشد. برای اغلب کشورهای شرکت‌کننده در مطالعه تیمز ۲۰۱۵، انتخاب حداقل ۱۵۰ مدرسه و ۴۰۰۰ دانش‌آموز پایه‌ی هشتم، استانداردهای لازم مطالعه تیمز را محقق می‌سازد. در مطالعه تیمز ۲۰۱۵، تعداد مدارس انتخاب شده برای کشورهای ایران، استرالیا و جمهوری کره به ترتیب ۲۵۰، ۲۸۵ و ۱۵۰ مدرسه است. اندازه نمونه در پایه‌ی هشتم برای کشورهای ایران، استرالیا و جمهوری کره به ترتیب ۶۱۳۰، ۱۰۳۳۸ و ۵۳۰۹ دانش‌آموز است (مارتین و همکاران، ۲۰۱۶).

ابزارهای گردآوری داده‌های مطالعه تیمز ۲۰۱۵، از دو بخش آزمون پیشرفت تحصیلی و پرسشنامه‌های زمینه‌ای، تشکیل می‌شود. آزمون پیشرفت تحصیلی ریاضی پایه‌ی هشتم، شامل ۲۱۲ سؤال به‌منظور اندازه‌گیری دانش و مهارت‌های ریاضی دانش‌آموزان پایه‌ی هشتم است. همچنین سه پرسشنامه زمینه‌ای نیز شامل پرسشنامه دانش‌آموز، پرسشنامه دبیر

1. Probability Proportional to their Size
2. National Research Coordinator
3. International Association for Evaluation of Educational Achievement
4. Data Processing & research Center
5. standard error
6. standard deviation

ریاضی و پرسشنامه مدیر مدرسه است (کبیری، کریمی و بخشعلی زاده، ۱۳۹۵). روایی محتوایی تمامی پرسشنامه‌ها به‌دقت بررسی شده است. برای بررسی روایی ملاکی مقیاس‌های مختلف پرسشنامه‌های زمینه‌ای از ضریب همبستگی بین آن مقیاس‌ها با نمره آزمون پیشرفت تحصیلی ریاضی استفاده شده است؛ به‌طور مثال ضریب همبستگی پیرسون بین مقیاس اعتمادبه‌نفس دانشجویان در ریاضی و نمره آزمون پیشرفت تحصیلی ریاضی برای کشورهای ایران، استرالیا و جمهوری کره به‌ترتیب ۰/۴۲، ۰/۵۱ و ۰/۵۲ است (مارتین و همکاران، ۲۰۱۶: ۱۵,۲۱۲). برای بررسی پایایی نیز ضریب پایایی آلفای کرونباخ آزمون پیشرفت تحصیلی ریاضی پایه‌ی هشتم مطالعه تیمز ۲۰۱۵ برای کشورهای ایران، استرالیا و جمهوری کره به‌ترتیب عبارت است از ۰/۸۷، ۰/۸۹ و ۰/۹۱ است (بخش‌های ۱۱ و ۱۶ مارتین و همکاران، ۲۰۱۶). ضریب پایایی آلفای کرونباخ مقیاس‌های مختلف پرسشنامه‌های زمینه‌ای در سطح قابل قبولی قرار دارند، تقریباً همه بالای ۰/۷ و بسیاری از آن‌ها بالای ۰/۸ است (مارتین و همکاران، ۲۰۱۶: ۱۵,۱۰).

شایان ذکر است در مطالعه تیمز برای پوشش تمام و کمال برنامه‌ی درسی ریاضی از روش نمونه‌گیری ماتریسی سؤال‌ها استفاده می‌شود. به عبارتی هر یک از دانش‌آموزان تمامی سؤال‌های خزانه‌ی سؤال‌ها[1] را پاسخ نمی‌دهد، بلکه زیرمجموعه‌ای از سؤال‌ها را پاسخ می‌دهند (اولسون[2]، مارتین و مولیس، ۲۰۰۸). یکی از رویکردها برای پوشش تمام و کمال برنامه‌ی درسی و درعین‌حال به حداقل رساندن زمان آزمون برای دانش‌آموزان، استفاده از روش نمونه‌گیری ماتریسی سؤال‌ها است. روش نمونه‌گیری ماتریسی سؤال‌ها مشتمل بر توسعه‌ی مجموعه کاملی از سؤال‌های در نظر گرفته برای پوشش برنامه درسی است؛ در این روش سؤال‌ها به زیرمجموعه‌هایی تقسیم می‌شوند و هر یک از دانش‌آموزان صرفاً به یکی از زیرمجموعه‌های سؤال‌ها پاسخ می‌دهند. نمونه‌گیری ماتریسی با محدود کردن تعداد سؤال‌هایی که هر دانش‌آموز پاسخ می‌دهد، زمان موردنیاز آزمون را نیز محدود می‌کند؛ این در حالی است که پوشش تمام و کمال محتوای آزمون برای تمامی دانش‌آموزان تعبیه شده است (چایلدس و ژاسیو[3]، ۲۰۰۳).

1. Item pool
2. Olson
3. Childs & Jaciw

روش تجزیه‌وتحلیل داده‌ها

برای تجزیه‌وتحلیل داده‌ها، ابتدا با استفاده از نرم‌افزار تحلیل پایگاه داده‌های بین‌المللی IEA IDB[1] (IEA, 2017) داده‌های اولیه موردنیاز برای این تحقیق از پایگاه داده‌های مطالعه تیمز ۲۰۱۵ در قالب فایل SPSS استخراج شده است. برای محاسبه‌ی مقادیر آماره‌های برازش فرد از بسته نرم‌افزاری PerFit موجود در نرم‌افزار R استفاده می‌شود (موسوی، تندیرو و یونسی[2]، ۲۰۱۶). در تجزیه‌وتحلیل داده‌ها و مدل‌سازی خطی سلسله‌مراتبی بین مقدار آماره‌های برازش فرد و متغیرهای زمینه‌ای از نرم‌افزار HLM برای برآورد پارامترهای ثابت و تصادفی مدل‌های خطی سلسله‌مراتبی و آزمون فرض آن‌ها استفاده شده است.

٤. یافته‌های تحقیق

برای به دست آوردن مدل خطی بین مقادیر آماره‌ی برازش فرد H^T دانش‌آموزان و متغیرهای زمینه‌ای آن‌ها و با توجه به ساختار آشیانه‌ای دانش‌آموزان در مدارس و مدارس در کشورها، از مدل خطی سلسله‌مراتبی استفاده خواهد شد. در این بخش ابتدا، مدل خطی سلسله‌مراتبی سه سطحی مقدار آماره‌ی برازش فرد H^T، شامل صرفاً عرض از مبدأ بررسی می‌شود.

سطح اول:

$$y_{ijk} = \pi_{0jk} + e_{ijk}$$

سطح دوم:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

سطح سوم:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

مدل آمیخته:

$$y_{ijk} = \gamma_{000} + u_{00k} + r_{0jk} + e_{ijk}$$

که در آن‌ها

y_{ijk}: مقدار آماره‌ی برازش فرد H^T دانش‌آموز i–ام، مدرسه‌ی j–ام کشور k–ام است.

π_{0jk}: میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان مدرسه‌ی j–ام کشور k–ام است.

1. International Database Analyzer
2. Younesi

e_{ijk}: اثر تصادفی، انحراف مقدار آماره‌ی برازش فرد H^T دانش‌آموز i–ام، مدرسه‌ی j–ام کشور k–ام از میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان مدرسه‌ی j–ام کشور k–ام است $e_{ijk} \sim N(0, \sigma^2)$.

β_{00k}: میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان کشور k–ام است.

r_{0jk}: اثر تصادفی، انحراف میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان مدرسه‌ی j–ام کشور k–ام از میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان کشور k–ام است $r_{0jk} \sim N(0, \tau_\pi)$.

γ_{000}: میانگین مقدار آماره‌ی برازش فرد H^T تمامی دانش‌آموزان است.

u_{00k}: اثر تصادفی، انحراف میانگین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان کشور k–ام از میانگین مقدار آماره‌ی برازش فرد H^T تمامی دانش‌آموزان است $u_{00k} \sim N(0, \tau_\beta)$.

جدول ۱. برآورد اثرات ثابت و تصادفی مدل خطی سه سطحی مقدار آماره‌ی برازش فرد H^T، شامل صرفاً عرض از مبدأ

اثر ثابت						
		برآورد	خطای استاندارد	آماره‌ی آزمون (t)	درجه‌ی آزادی تقریبی	مقدار احتمال (p-value)
γ_{000}		۰/۳۴۷	۰/۰۲۶	۱۳/۱۶۸	۲	۰/۰۰۶
اثر تصادفی						
		انحراف استاندارد	مؤلفه واریانس	آماره‌ی آزمون (x^2)	درجه‌ی آزادی	مقدار احتمال (p-value)
سطح (u_{00k})	کشور	۰/۰۴۵	۰/۰۰۲۰۳	۱۱۹/۱۷۰	۲	<۰/۰۰۱
سطح (r_{0jk})	مدرسه	۰/۰۳۸	۰/۰۰۱۴۴	۶۹۲/۵۸۰	۵۸۸	۰/۰۰۲
سطح (e_{ijk})	دانش‌آموز	۰/۱۳۴	۰/۰۱۷۸۴			

اولین مرحله در تحلیل سلسله‌مراتبی معنی‌داری اثر تصادفی است. با توجه به جدول ۱ و مقدار احتمال (p-value) کمتر از ۰/۰۰۱ برای سطح کشور و مقدار احتمال (p-value) ۰/۰۰۲ برای سطح مدرسه، لذا انجام تحلیل سلسله‌مراتبی جایز است.

خاطر نشان می‌سازد برای مدل فوق ضریب همبستگی درون رده‌ای (ICC[1]) عبارت است از

$$\mathrm{ICC} = \frac{0.00203 + 0.00144}{0.00203 + 0.00144 + 0.01784} = 0.163$$

ضریب همبستگی درون رده‌ای، بیانگر آن است که ۸۳/۷ درصد از واریانس آماره‌ی برازش فرد H^T، در سطح دانش‌آموز و ۱۶/۳ درصد از آن در سطوح مدرسه و کشور است. در مرحله‌ی بعد، مدل خطی سلسله‌مراتبی مقدار آماره‌ی برازش فرد H^T، شامل صرفاً متغیرهای سطح دانش‌آموز برازش داده می‌شود و صرفاً متغیرهایی که ضرایب معنی‌داری دارد، نگه داشته و سایر متغیرها از مدل حذف می‌شوند. خاطر نشان می‌سازد در ابتدا ضرایب متغیرهای پیش‌بین "اطمینان در یادگیری ریاضیات" و "برآورد توانایی" تصادفی در نظر گرفته شدند، ولیکن با عنایت به میزان مؤلفه واریانس هر یک از آن‌ها، ضرایبی که مؤلفه واریانس آن‌ها معنی‌دار نبود، آن ضرایب ثابت در نظر گرفته شدند.

سطح اول:

$$y_{ijk} = \pi_{0jk} + \pi_{1jk} * (\mathrm{BSBGSCM}_{ijk}) + \pi_{2jk} * (\mathrm{ABILITY}_{ijk}) + e_{ijk}$$

سطح دوم:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$

سطح سوم:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$

مدل آمیخته:

$$y_{ijk} = \gamma_{000} + \gamma_{100} * (\mathrm{BSBGSCM}_{ijk}) + \gamma_{200} * (\mathrm{ABILITY}_{ijk}) + r_{0jk} + r_{2jk} * (\mathrm{ABILITY}_{ijk}) + u_{00k} + u_{20k} * (\mathrm{ABILITY}_{ijk}) + e_{ijk}$$

که در آن‌ها علاوه بر موارد فوق

$\mathrm{BSBGSCM}_{ijk}$: مقدار متغیر "اطمینان در یادگیری ریاضیات" دانش‌آموز i–ام، مدرسه‌ی j–ام کشور k–ام است.

$\mathrm{ABILITY}_{ijk}$: مقدار متغیر "برآورد توانایی" دانش‌آموز i–ام، مدرسه‌ی j–ام کشور k–ام است.

1. Intraclass Correlation Coefficient

جدول ۲. برآورد اثرات ثابت و تصادفی معنی‌دار مدل خطی سه سطحی مقدار آماره‌ی برازش فرد H^T، شامل صرفاً متغیرهای سطح دانش‌آموز

اثر ثابت					
	برآورد	خطای استاندارد	آماره‌ی آزمون (t)	درجه‌ی آزادی تقریبی	مقدار احتمال (p-value)
γ_{000} اطمینان دانش‌آموزان در یادگیری ریاضیات (سطح دانش‌آموز (γ_{100}))	۰/۳۴۶	۰/۰۲۶	۱۲/۶۵۶	۲	۰/۰۰۶
برآورد توانایی دانش‌آموزان (سطح دانش‌آموز (γ_{200}))	۰/۰۰۵	۰/۰۰۲	۲/۰۸۴	۷۶۷	۰/۰۳۸
	۰/۰۴۵	۰/۰۰۶	۷/۸۴۴	۷۹۶	<۰/۰۰۱
اثر تصادفی					
	انحراف استاندارد	مؤلفه واریانس	آماره‌ی آزمون (x^2)	درجه‌ی آزادی	مقدار احتمال (p-value)
u_{00k}	۰/۰۴۶	۰/۰۰۲۰۹	۱۲۳/۹۱۶	۲	<۰/۰۰۱
r_{0jk}	۰/۰۵۵	۰/۰۰۳۰۳	۷۱۴/۵۰۰	۴۵۷	<۰/۰۰۱
r_{2jk}	۰/۰۵۶	۰/۰۰۳۰۹	۵۹۴/۱۱۰	۴۵۷	<۰/۰۰۱
u_{20k}	۰/۰۴۵	۰/۰۰۲۰۱	۴۷/۲۴۹	۲	<۰/۰۰۱
e_{ijk}	۰/۱۱۷	۰/۰۱۳۷۸			

در جدول ۲ برآورد پارامترهای مدل اخیر و آزمون فرض آن پارامترها ارائه شده است. نظر به این که مقادیر کمتر آماره‌ی برازش فرد H^T نشانگر انحراف بیشتر الگوی پاسخ دانش‌آموز از الگوی پاسخ مناسب است؛ برآوردهای مثبت ضرایب متغیرهای "اطمینان دانش‌آموزان در یادگیری ریاضیات" و "برآورد توانایی دانش‌آموزان" نشانگر آن است که هر چه دانش‌آموزان اطمینان بیشتری در یادگیری ریاضیات داشته و توانایی بیشتری در فراگیری ریاضیات داشته باشند، احتمال بیشتری دارد که الگوی پاسخ مناسبی داشته باشند. به‌طور برعکس نیز هر چه دانش‌آموزان اطمینان کمتری در یادگیری ریاضیات داشته و توانایی کمتری در فراگیری ریاضیات داشته باشند، احتمال بیشتری دارد که الگوی پاسخ

نامناسبی داشته باشند. اندازه‌ی اثر[1] مدل اخیر در مقایسه با مدل اولیه بدون هیچ متغیر پیش‌بین برابر ۲۲/۷۶ درصد است. شایان ذکر است توانایی دانش‌آموزان، در مقیاسی با میانگین صفر و انحراف استاندارد یک، برآورد شده است (دوتویت[2]، ۲۰۰۳).

در مرحله‌ی بعد، متغیرهای سطح مدرسه به مدل خطی سلسله‌مراتبی قبلی اضافه می‌شود.

سطح اول:

$$y_{ijk} = \pi_{0jk} + \pi_{1jk} * (BSBGSCM_{ijk}) + \pi_{2jk} * (ABILITY_{ijk}) + e_{ijk}$$

سطح دوم:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} * (BTBGEAS_{jk}) + \beta_{02k} * (BTBGSCR_{jk}) + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k}$$
$$\pi_{2jk} = \beta_{20k} + \beta_{21k} * (BTBGEAS_{jk}) + \beta_{22k} * (BTBGSCR_{jk}) + r_{2jk}$$

سطح سوم:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{01K} = \gamma_{010}$$
$$\beta_{02K} = \gamma_{020}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$
$$\beta_{21K} = \gamma_{210}$$
$$\beta_{22K} = \gamma_{220}$$

مدل آمیخته:

$$\begin{aligned} y_{ijk} = {} & \gamma_{000} + \gamma_{010} * (BTBGEAS_{jk}) + \gamma_{020} * (BTBGSCR_{jk}) + \gamma_{100} \\ & * (BSBGSCM_{ijk}) + \gamma_{200} * (ABILITY_{ijk}) + \gamma_{210} \\ & * (ABILITY_{ijk}) * (BTBGEAS_{jk}) + \gamma_{220} * (ABILITY_{ijk}) \\ & * (BTBGSCR_{jk}) + r_{0jk} + r_{2jk} * (ABILITY_{ijk}) + u_{00k} + u_{20k} \\ & * (ABILITY_{ijk}) + e_{ijk} \end{aligned}$$

که در آن‌ها علاوه بر موارد قبلی

$BTBGEAS_{jk}$: مقدار متغیر "تأکید مدرسه بر موفقیت تحصیلی" مدرسه‌ی j–ام کشور k–ام است.

$BTBGSCR_{jk}$: مقدار متغیر "شرایط و منابع مدرسه" مدرسه‌ی j–ام کشور k–ام است.

1 effect size
2 du Toit

جدول ۳. برآورد اثرات ثابت و تصادفی مدل خطی سه سطحی مقدار آماره‌ی برازش فرد H^T، شامل متغیرهای سطح دانش‌آموز و مدرسه

اثر ثابت					
	برآورد	خطای استاندارد	آماره‌ی آزمون (t)	درجه‌ی آزادی تقریبی	مقدار احتمال (p-value)
γ_{000}	۰/۳۴۶	۰/۰۲۴	۱۴/۴۷۶	۲	۰/۰۰۵
تأکید مدرسه بر موفقیت تحصیلی (سطح مدرسه (γ_{010}))	۰/۰۰۸	۰/۰۰۲	۴/۱۶۲	۵۸۲	<۰/۰۰۱
شرایط و منابع مدرسه (سطح مدرسه (γ_{020}))	۰/۰۰۳	۰/۰۰۲	۱/۳۲۵	۵۸۲	۰/۱۸۶
اطمینان در یادگیری ریاضیات (سطح دانش‌آموز (γ_{100}))	−۰/۰۰۶	۰/۰۰۳	−۲/۳۵۰	۱۷۲	۰/۰۲۰
برآورد توانایی دانش‌آموزان (سطح دانش‌آموز (γ_{200}))	۰/۰۵۱	۰/۰۲۸	۱/۸۳۳	۲	۰/۲۰۸
اثر متقابل برآورد توانایی دانش‌آموزان و تأکید مدرسه بر موفقیت تحصیلی (γ_{210})	۰/۰۰۲	۰/۰۰۳	۰/۷۶۸	۵۸۲	۰/۴۴۳
اثر متقابل برآورد توانایی دانش‌آموزان و شرایط و منابع مدرسه (γ_{220})	−۰/۰۰۲	۰/۰۰۴	−۰/۵۰۷	۵۸۲	۰/۶۱۳
اثر تصادفی					
	انحراف استاندارد	مؤلفه واریانس	آماره‌ی آزمون (x^2)	درجه‌ی آزادی	مقدار احتمال (p-value)
u_{00k}	۰/۰۴۱	۰/۰۰۱۶۷	۱۰۴/۲۰۶	۲	<۰/۰۰۱
u_{20k}	۰/۰۴۶	۰/۰۰۲۱۴	۵۰/۹۴۰	۲	<۰/۰۰۱
r_{0jk}	۰/۰۵۲	۰/۰۰۲۶۵	۶۸۵/۳۸۰	۴۵۵	<۰/۰۰۱
r_{2jk}	۰/۰۵۵	۰/۰۰۳۰۳	۵۹۲/۹۲۲	۴۵۵	<۰/۰۰۱
e_{ijk}	۰/۱۱۷	۰/۰۱۳۷۴			

در جدول ۳ برآورد پارامترهای مدل شامل متغیرهای سطوح دانش‌آموزان و مدارس و آزمون فرض آن پارامترها ارائه شده است. اندازه‌ی اثر این مدل در مقایسه با مدل اولیه بدون هیچ متغیر پیش‌بین، برابر ۲۲/۹۸ درصد است. گرچه این مدل در مقایسه با مدل قبلی، مدل شامل صرفاً متغیرهای سطح دانش‌آموز، درصد قابل توجهی از واریانس آماره‌ی برازش فرد H^T را تبیین نمی‌کند ولیکن با وارد کردن متغیر زمینه‌ای "تأکید مدرسه بر موفقیت تحصیلی" دارای ضریب معنی‌دار از مجموعه متغیرهای زمینه‌ای سطح مدرسه، ضریب متغیر زمینه‌ای "اطمینان دانش‌آموزان در یادگیری ریاضیات" که در مدل قبل مثبت و معنی‌دار بود در مدل اخیر منفی و معنی‌دار شده است. یعنی رابطه‌ی بین مقدار آماره‌ی برازش فرد H^T و متغیر زمینه‌ای "اطمینان دانش‌آموزان در یادگیری ریاضیات" معکوس شده است. به عبارتی در مدل اخیر هرچه دانش‌آموز اطمینان بیشتری در یادگیری ریاضیات داشته باشد، احتمال بیشتری دارد که الگوی پاسخ نامناسبی داشته باشد. مورد دیگری که قابل توجه است، این است که ضرایب متغیرهای زمینه‌ای همان‌طور که از جداول مشخص می‌شود ازلحاظ قدر مطلق مقادیر کوچکی هستند، این بدین دلیل است که متغیرهای زمینه‌ای با میانگین ۱۰ و انحراف استاندارد دو مقیاس‌سازی شده‌اند ولیکن مقدار آماره‌ی برازش فرد H^T می‌تواند بین منفی یک و یک باشد.

در مرحله‌ی بعد، متغیر "میانگین پیشرفت تحصیلی دانش‌آموزان کشورها" که توسط تیمز گزارش شده به مدل خطی سلسله‌مراتبی قبلی اضافه می‌شود.

سطح اول:

$$y_{ijk} = \pi_{0jk} + \pi_{1jk} * (BSBGSCM_{ijk}) + \pi_{2jk} * (ABILITY_{ijk}) + e_{ijk}$$

سطح دوم:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} * (BTBGEAS_{jk}) + \beta_{02k} * (BTBGSCR_{jk}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k} + \beta_{21k} * (BTBGEAS_{jk}) + \beta_{22k} * (BTBGSCR_{jk}) + r_{2jk}$$

سطح سوم:

$$\beta_{00k} = \gamma_{000} + \gamma_{001} * (MATHEMAT_k) + u_{00k}$$

$$\beta_{01K} = \gamma_{010} + \gamma_{011} * (MATHEMAT_k)$$

$$\beta_{02K} = \gamma_{020} + \gamma_{021} * (MATHEMAT_k)$$

$$\beta_{10k} = \gamma_{100} + \gamma_{101} * (MATHEMAT_k)$$

$$\beta_{20k} = \gamma_{200} + \gamma_{201} * (MATHEMAT_k) + u_{20k}$$

$$\beta_{21K} = \gamma_{210} + \gamma_{211} * (MATHEMAT_k)$$

$$\beta_{22K} = \gamma_{220} + \gamma_{221} * (MATHEMAT_k)$$

که در آن‌ها علاوه بر موارد قبلی

$MATHEMAT_k$: مقدار متغیر "میانگین پیشرفت تحصیلی" کشور k–ام است.

جدول ٤. برآورد اثرات ثابت و تصادفی مدل خطی سه سطحی مقدار آماره‌ی برازش فرد H^T، شامل متغیرهای سطح دانش‌آموز، مدرسه و کشور

اثر ثابت					
	برآورد	خطای استاندارد	آماره‌ی آزمون (t)	درجه‌ی آزادی تقریبی	مقدار احتمال (p-value)
γ_{000} میانگین پیشرفت تحصیلی (سطح کشور (γ_{001}))	۰/۳۵۰	۰/۰۰۸	۴۱/۹۲۷	۲	<۰/۰۰۱
تأکید مدرسه بر موفقیت تحصیلی (سطح مدرسه (γ_{010}))	۰/۰۰۱	۰/۰۰۰۱	۴/۸۷۰	۲	۰/۰۳۹
اثر متقابل میانگین پیشرفت تحصیلی کشور و تأکید مدرسه بر موفقیت تحصیلی (γ_{011})	۰/۰۰۷	۰/۰۰۲	۳/۸۱۳	۵۷۸	<۰/۰۰۱
اطمینان در یادگیری ریاضیات (سطح دانش‌آموز (γ_{100}))	−۰/۰۰۰۱	۰/۰۰۰۳	−۳/۵۴۱	۵۷۸	<۰/۰۰۱
برآورد توانایی دانش‌آموزان (سطح دانش‌آموز (γ_{200}))	−۰/۰۰۶	۰/۰۰۳	−۲/۳۸۹	۱۷۱	۰/۰۱۸
اثر متقابل میانگین پیشرفت تحصیلی کشور و برآورد توانایی دانش‌آموزان (γ_{201})	۰/۰۵۳	۰/۰۱۲	۴/۲۳۱	۲	۰/۰۴۱
	−۰/۰۰۰۶	۰/۰۰۰۲	−۳/۲۱۰	۲	۰/۰۸
اثر تصادفی					
	انحراف استاندارد	مؤلفه واریانس	آماره‌ی آزمون (x^2)	درجه‌ی آزادی	مقدار احتمال (p-value)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

u_{00k}	۰/۰۱۳	۰/۰۰۰۱۶	۱۳/۲۹۸	۱	<۰/۰۰۱
u_{20k}	۰/۰۱۷	۰/۰۰۰۲۹	۹/۵۷۳	۱	۰/۰۰۲
r_{0jk}	۰/۰۴۹	۰/۰۰۲۴۵	۶۷۰/۶۳۰	۴۵۵	<۰/۰۰۱
r_{2jk}	۰/۰۵۴	۰/۰۰۲۹۶	۵۹۰/۸۸۳	۴۵۵	<۰/۰۰۱
e_{ijk}	۰/۱۱۷	۰/۰۱۳۷۳			

در جدول ۴ برآورد پارامترهای مدل شامل متغیرهای سطوح دانش‌آموزان، مدارس و کشورها و آزمون فرض آن پارامترها ارائه شده است. اندازه‌ی اثر این مدل در مقایسه با مدل اولیه بدون هیچ متغیر پیش‌بین، برابر ۲۳/۰۴ درصد است. گرچه این مدل در مقایسه با مدل قبلی، مدل شامل صرفاً متغیرهای سطوح دانش‌آموز و مدرسه، درصد قابل توجهی از واریانس آماره‌ی برازش فرد H^T را تبیین نمی‌کند ولیکن ضریب متغیر زمینه‌ای "اطمینان دانش‌آموزان در یادگیری ریاضیات" که در مدل اولیه مثبت و معنی‌دار بود در مدل اخیر منفی و معنی‌دار باقی مانده است. یعنی رابطه‌ی بین مقدار آماره‌ی برازش فرد H^T و متغیر زمینه‌ای "اطمینان دانش‌آموزان در یادگیری ریاضیات" معکوس است. به عبارتی در مدل اخیر هرچه دانش‌آموز اطمینان بیشتری در یادگیری ریاضیات داشته باشد، احتمال بیشتری دارد که الگوی پاسخ نامناسبی داشته باشد. شایان ذکر است متغیرهای زمینه‌ای با میانگین ۱۰ و انحراف استاندارد دو مقیاس‌سازی شده‌اند و متغیر پیشرفت تحصیلی با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰ مقیاس‌سازی شده است.

بحث و نتیجه‌گیری

به‌منظور مدل‌سازی عوامل سطوح دانش‌آموزان، مدارس و کشورها که بر برازش فرد تأثیرگذار هستند، با توجه به ساختار آشیانه‌ای آن‌ها از مدل‌سازی خطی سلسله‌مراتبی استفاده شد. برای آماره‌ی برازش فرد H^T، از مدل‌سازی خطی سلسله‌مراتبی سه سطحی شامل سطوح دانش‌آموزان، مدارس و کشورها استفاده شد؛ به استناد ضریب همبستگی درون رده‌ای، مشخص گردید ۱۶/۳ درصد از تغییرات این آماره‌ی برازش فرد مؤثر از عوامل سطوح مدارس و کشورها است.

در مدل خطی سلسله‌مراتبی شامل صرفاً متغیرهای سطح دانش‌آموز متغیر پیش‌بین "اطمینان در یادگیری ریاضیات دانش‌آموزان" دارای ضریب مثبت معنی‌دار در پیشگویی مقدار آماره‌ی برازش فرد H^T دانش‌آموزان دارد؛ این در حالی است که در مدل خطی سلسله‌مراتبی که متغیرهای پیش‌بین سطح مدرسه نیز به مدل قبلی اضافه می‌شود، ضریب

متغیر پیش‌بین "اطمینان در یادگیری ریاضیات دانش‌آموزان" از مقدار مثبت معنی‌دار به مقدار منفی معنی‌دار تغییر یافته است؛ یا به عبارتی رابطه‌ی این متغیر با مقدار آماره‌ی برازش فرد H^T دانش‌آموزان معکوس شده است و این بازنمایی مناسبی از واقعیت است. در همین راستا آلیورنینی، مانگانلی و وینسی[1] (۲۰۰۸) به نقل از نقش و مقدم (۱۳۹۱)، مدل خود را با داده‌های پرلز در دو حالت عدم توجه به ویژگی سلسله‌مراتبی داده‌ها و توجه به ویژگی سلسله‌مراتبی داده‌ها تحلیل کرده و با مقایسه نتایج دریافتند زمانی که به ساختار سلسله‌مراتبی داده‌ها توجه شده است، نتایج بازنمایی مناسبی از واقعیت را نشان داده‌اند.

با توجه به اندازه‌ی اثر مدل نهایی خطی سلسله‌مراتبی بین مقدار آماره‌ی برازش فرد H^T دانش‌آموزان و عوامل سه سطح دانش‌آموزان، مدارس و کشورها، ۲۳/۰۴ درصد از تغییرات آماره‌ی برازش فرد H^T توسط این مدل تبیین می‌شود. همچنین صرفاً متغیرهای میانگین پیشرفت تحصیلی کشورها، تأکید مدرسه بر موفقیت تحصیلی دانش‌آموزان، اطمینان در یادگیری ریاضیات دانش‌آموزان و برآورد توانایی دانش‌آموزان از عوامل با ضرایب معنی‌دار در مدل نهایی هستند. خاطر نشان می‌سازد از ضرایب معنی‌دار مذکور، ضرایب میانگین پیشرفت تحصیلی کشورها، تأکید مدرسه بر موفقیت تحصیلی دانش‌آموزان و برآورد توانایی دانش‌آموزان مثبت و رابطه‌ی مستقیمی با مقدار آماره‌ی برازش فرد H^T دارد، و می وان این‌طور تفسیر کرد که دانش‌آموزانی که در کشورهایی با میانگین پیشرفت تحصیلی بیشتر، دانش‌آموزانی که در مدرسه‌هایی تحصیل می‌کنند که تأکید بیشتری بر موفقیت تحصیلی دانش‌آموزان دارند و دانش‌آموزانی که توانایی بیشتری در یادگیری ریاضیات دارند، احتمال بیشتری دارد که الگوی پاسخ مناسبی داشته باشند. همچنین رابطه‌ی بین اطمینان در یادگیری ریاضیات دانش‌آموزان و مقدار آماره‌ی برازش فرد H^T، معکوس است یعنی هرچه اطمینان در یادگیری ریاضیات دانش‌آموزان افزایش می‌یابد، مقدار آماره‌ی برازش فرد H^T آن‌ها کاهش می‌یابد و به سمت داشتن الگوی پاسخی با برازش نامناسب میل خواهند کرد. این ممکن است برخلاف انتظار باشد که هر چه دانش‌آموز اطمینان بیشتری در یادگیری ریاضیات داشته باشد، احتمال داشتن الگوی پاسخ نامناسب او افزایش یابد؛ ولیکن این را می‌توان این‌طور تفسیر کرد که دانش‌آموزان با

1. Alivernini, Manganelli & Vinci

اطمینان بیشتر در یادگیری ریاضیات، در پاسخ به سؤال‌ها سریع‌تر عمل کرده و دقّت کمتری دارند.

در مدل نهایی خطی سلسله‌مراتبی سه سطحی در سطح سه (کشورها) یک متغیر (میانگین پیشرفت تحصیلی کشورها) در نظر گرفته شده بود که آن هم دارای ضریب معنی‌داری بود. در سطح دو (مدارس) از بین متغیرهای زمینه‌ای "ترکیب مدرسه ازلحاظ پیشینه دانش‌آموزان"، "شرایط کاری معلمان"، "تأکید مدرسه بر موفقیت تحصیلی دانش‌آموزان"، "رضایت شغلی معلمان"، "رویارویی معلمان با چالش‌ها"، "مشکلات انضباطی مدرسه"، "امنیت و انضباط مدرسه" و "محدودیت تدریس به دلیل نیازهای آموزشی دانش‌آموزان" صرفاً یکی از متغیرها (تأکید مدرسه بر موفقیت تحصیلی دانش‌آموزان) دارای ضریب معنی‌دار در مدل بود و سایر متغیرها ضرایب معنی‌دار نداشتند، این می‌تواند بدین دلیل باشد که تعداد نمونه‌های هر یک از مدارس حداکثر سه دانش‌آموز بوده است. این تعداد کم نمونه می‌تواند یکی از عوامل معنی‌دار نشدن ضرایب آن متغیرها باشد و نباید نتیجه گرفت که تحلیل سلسله‌مراتبی مفید واقع نمی‌شود.

در مدل‌های خطی این تحقیق قدرمطلق ضرایب مدل‌های خطی مقادیر کوچکی هستند، این بدین دلیل است که مقدار آماره‌ی برازش فرد H^T بین منفی یک و یک تغییر می‌کند ولیکن متغیرهای زمینه‌ای با میانگین ۱۰ و انحراف استاندارد دو؛ و متغیر پیشرفت تحصیلی دانش‌آموزان با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰ مقیاس‌سازی شده‌اند.

آماره‌های برازش فرد به دلیل پاسخگو نبودن این که برازش نامناسب چگونه رخ می‌دهد، همواره مورد انتقاد بوده است. با این حال ما معتقد هستیم که سنجش برازش فرد برای توسعه و اعتبارسنجی آزمون‌ها مفید است. آماره‌های برازش فرد را می‌توان در مطالعه‌ی مقدماتی آزمون‌ها مورد استفاده قرار داد. اگر در بررسی نتایج مطالعه‌ی مقدماتی آزمون مشخص شد که درصد الگوهای پاسخ با برازش نامناسب بیش‌ازحد مورد انتظار است، می‌بایست به بررسی بیشتر محتوای آزمون‌ها پرداخته شود. ازاین‌رو سنجش برازش فرد توان بالقوه‌ای برای کمک به بهبود روایی آزمون دارد و با شناسایی مشکلات احتمالی پیش روی آزمون و مرتفع کردن آن‌ها، می‌توان به آزمون‌هایی عادلانه‌تر و دقیق‌تر دست یافت. تجزیه‌وتحلیل تعقیبی برای بررسی دقیق‌تر دلایل ممکن برازش نامناسب، باید اجرا شود.

پاسخ‌های با برازش نامناسب که صرفاً بر اساس آماره‌های برازش فرد شناسایی شده‌اند را نباید هرگز به‌عنوان شواهد کافی برای نامعتبر شناختن نتایج آزمون دانشجویان لحاظ کرد. تجزیه‌وتحلیل تعقیبی برای بررسی دقیق‌تر دلایل ممکن برازش نامناسب باید اجرا شود. اگر فرضیه‌ی خاصی در ارتباط با نوع رفتار با برازش نامناسب دانش‌آموزان وجود داشته باشد، روش‌هایی از قبیل اندازه‌گیری مناسب بهینه برای آزمون آماری آن فرضیه در دسترس است (لوین و دراسگو، ۱۹۸۸؛ لامپریانو، ۲۰۱۰). مزیت چنین آزمون فرضیه‌هایی این است که نتایج آن‌ها در مورد نوع برازش نامناسب آگاهی بخش‌تر است. با این حال، برای درک واقعی این که چرا برازش نامناسب رخ داده است، اطلاعاتی درباره‌ی پاسخ دانش‌آموزان از قبیل مصاحبه با دانش‌آموزان و معلمان، گزارش‌های کلامی دانش‌آموزان، اطلاعات ردیابی چشمی و زمان واکنش ممکن است مورد نیاز باشد (انجمن تحقیقات آموزشی آمریکا[1]، شورای ملی اندازه‌گیری در آموزش‌وپرورش[2] و انجمن روان‌شناسی آمریکا[3]، ۱۹۹۹). این نوع از اطلاعات تصویر نسبتاً مشروحی از چگونگی پاسخ به سؤال‌های آزمون‌ها فراهم می‌کند و می‌تواند به شناسایی عوامل مرتبط با برازش نامناسب کمک کند. بدین ترتیب نتایج حاصل از آماره‌های برازش فرد به‌طور قابل‌ملاحظه‌ای قابل تفسیر و ارزشمند هستند و مهم‌تر از همه این‌که با استفاده از آماره‌های برازش فرد می‌توان روایی نمره‌های آزمون را بهبود بخشید. نمونه‌ای از این دسته مطالعات، تحقیق میجر و همکاران (۲۰۰۸) است که در آن داده‌های حاصل از پرسشنامه خوداظهاری نیم‌رخ خودانگاره هارتر[4] (۱۹۸۵) کودکان مورد تجزیه‌وتحلیل قرار گرفته است. آماره‌های برازش فرد به همراه اطلاعات حاصل از مشاهده و مصاحبه با کودکان، در شناسایی و درک پاسخ کودکانی که پاسخشان با مدل IRT مورد استفاده ناسازگار بود، کمک شایانی کرد. در حوزه‌ی تحقیقات برازش فرد، مطالعات بیشتری که اطلاعات حاصل از منابع مختلف را برای درک بهتر دلایل پاسخ با برازش نامناسب ترکیب کند، مورد نیاز است.

علاوه بر این اغلب متغیرهای سطوح دانش‌آموزان و مدرسه، توسط پرسشنامه‌های خوداظهاری توسط دانش‌آموزان، معلمان و مدیران مدارس گردآوری شده است. گرچه

1. American Educational Research Association
2. National Council for Measurement in Education
3. American Psychological Association
4. Harter

اطلاعاتی که با استفاده از پرسشنامه‌های خوداظهاری حاصل می‌شوند، در تحقیقات رایج هستند و به دست آوردن آن اطلاعات به‌طور نسبی آسان هستند و اغلب تنها راه عملی برای اندازه‌گیری برخی از سازه‌ها هستند، ولیکن روایی این قبیل پرسشنامه‌ها مورد سؤال هستند. هم تئوری و هم تحقیقات بیان‌گر آن هستند که عوامل روان‌شناختی، جامعه‌شناختی، زبان‌شناختی، تجربی و متنی می‌توانند به اریبی پاسخ خوداظهاری منجر شوند (به‌طور مثال، هاریسون، مک‌لاولین و کولتر[1]، ۱۹۹۶؛ لانیون و گودشتاین[2]، ۱۹۹۷). مثلاً افراد ممکن است در پاسخ به سؤال‌ها تمایل به این داشته باشند که به نحوی به سؤال‌ها پاسخ دهند که ازلحاظ اجتماعی، مطلوب جلوه داده شوند؛ و این مورد می‌تواند به‌طور بالقوه در پاسخ‌دهی دانش‌آموزان، معلمان و مدیران مدارس به پرسشنامه‌های زمینه‌ای تأثیرگذار باشد. ازاین‌رو با توجه به توضیحات فوق و این که متغیرهای زمینه‌ای این تحقیق اغلب توسط پرسشنامه‌های خوداظهاری گردآوری شده‌اند، لذا نتایج این تحقیق لازم است با احتیاط بیشتری تفسیر شوند.

تشکر و قدردانی

در اینجا لازم است از دکتر مسعود کبیری، مدیر فنی و داده‌پردازی مرکز ملی مطالعات تیمز و پرلز پژوهشگاه مطالعات آموزش‌وپرورش، به خاطر کمک‌های متعددشان در طول تحقیق، تشکر و قدردانی شود.

منابع

بودون، ریمون (۱۳۷۳). *روش‌های جامعه‌شناسی*، ترجمه عبدالحسین نیک گهر، شرکت انتشارات علمی و فرهنگی، چاپ دوم.

سرمد، زهره؛ بازرگان، عباس؛ و حجازی، الهه (۱۳۸۴). *روش‌های تحقیق در علوم رفتاری*. تهران. نشر آگاه.

کبیری، مسعود؛ کریمی، عبدالعظیم؛ و بخشعلی زاده، شهرناز (۱۳۹۵). *یافته‌های ملی تیمز ۲۰۱۵، روند ۲۰ ساله آموزش علوم و ریاضیات ایران در چشم‌انداز بین‌المللی*. پژوهشگاه مطالعات آموزش‌وپرورش. انتشارات مدرسه.

1. Harrison, McLaughlin & Coalter
2. Lanyon & Goodstein

نقش، زهرا؛ و مقدم، اعظم (۱۳۹۱). کاربرد تکنیک‌های مدل‌یابی چندسطحی در تحلیل داده‌های تیمز ۲۰۰۷ و مقایسه آن با تحلیل یک‌سطحی. *فصلنامه اندازه‌گیری تربیتی*. دوره دوم. شماره هشتم.

Alivernini,F., Manganelli, S., & Vinci, E. (2008). *Multilevel analysis of PIRLS 2006 data for Italy*. Paper presented at the 3rd IEA International Research Conference, Taipei, and Chinese Taipei.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Armstrong, R. D., & Shi, M. (2009). An IRT-based cumulative sum statistic for person fit. *Applied Psychological Measurement, 33*(5), 391-410.

Childs, R. A., & Jaciw, A. P. (2003). Matrix sampling of items in large-scale assessments. *Practical Assessment, Research & Evaluation, 8*(16).

Coleman, J. S., Compbell, E. Q., Hobson, C. J., Mcpartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of Educational Opportunity*. Washington DC: Department of Helth, Education & Welfare Office of Education.

Conijn, J. M., Emons, W. H. M., & Sijtsma, K. (2014). Statistics l_z-based person-fit methods for noncognitive multiscale measures. *Applied Psychological Measurement, 38*(2), 122-136.

Conijn, J. M., Sijtsma, K., & Emons, W. H. M. (2016). Identifying person-fit latent classes, and explanation of categorical and continuous person misfit. *Applied Psychological Measurement, 40*(2), 128-141.

Cui, Y., & Li, J. (2015). Evaluating Person fit for cognitive diagnostic assessment. *Applied Psychological Measurement, 39*(3), 223-238.

Cui, Y., & Mousavi, A. (2015). Explore the usefulness of person-fit analysis on large-scale assessment. *International Journal of Testing, 15*(1), 23-49.

De la Torre, J., & Deng, W. (2008). Improving person-fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement, 45*(2), 159–177.

De Leeuw, J. and Kreft, I. G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational and Behavioral Statistics, 11*(1), 57-85.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association, 76*, 341-353.

Dodeen, H., & Darabi, M. (2009). Person-fit: relationship with four personality tests in mathematics. *Research Papers in Education, 24*(1), 115–126.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology, 38*(1), 67–86.

Du Toit, M. (2002). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Scientific Software International: Lincolnwood, IL, USA.

Emons, W. H. M., Sijtsma, K., Meijer, R. R. (2005). Global, local, and graphical person-fit analysis using person-response functions. *Psychological Methods, 10*(1), 101-119.

Ferrando, P. J. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. *Personality and Individual Differences, 52*(6), 718-722.

Finkelman, M., & Kim, W. (2007). *Using person fit in a body of work standard setting*. Paper presented at the American Educational Research Association, Chicago, IL, USA.

Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd ed. London: Hodder Arnold.

Guo, J., & Drasgow, F. (2010). Identifying cheating on unprotected internet tests: The Z-test and the likelihood ratio test. *International Journal of Selection and Assessment, 18*(4), 351–364.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*, 139–150.

Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement, 18*(3), 133–146.

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes, 68*(3), 246–261.

Harter, S. (1985). *Manual for the self-perception profile for children*. Denver, CO: University of Denver.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics, *Applied Measurement in Education, 16*(4), 277-298.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963-974.

Lamprianou, I. (2010). The practical application of optimal appropriateness measurement of empirical data using Rash models. *Journal of Applied Measurement, 11*(4), 409–423.

Lamprianou, I., & Boyle, B. (2004). Accuracy of measurement in the context of mathematics national curriculum tests in England for ethnic minority pupils and pupils who speak English as an additional language. *Journal of Educational Measurement, 41*(3), 239–259.

Lanyon, R. I., & Goodstein, L. D. (1997). *Personality assessment* (3rd ed.). New York, NY: Wiley.

Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika, 53*(2), 161–176.

Liu, M. T., & Yu, P. T. (2011). Aberrant learning achievement detection based on person-fit statistics in personalized e-learning systems learning systems. *Educational Technology & Society, 14*(1), 107-120.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74*(4), 817-827.

Longford, N. T. (1993). *Random Coefficient Models*. New York: Oxford University Press.

Martin, M. O., Mullis, I. V.S., & Hooper, M. (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Meijer, R. R. (1997). Person fit and criterion-related validity: An extension of the Schmitt, Cortina, and Whitney study. *Applied Psychological Measurement, 21*(2), 99 -113.

Meijer, R. R., Egberink, J. H. L., Emons, W. H. M., & Sijtsma, K. (2008). Detection and validation of unscalable item score patterns using item response theory: An illustration with Harter's self-perception profile for children. *Journal of Personality Assessment, 90*(3), 227–238.

Meijer, R. R., & Sijtsma, K. (1995). Detection of aberrant item score patterns: A review of recent developments. *Applied Measurement in Education, 8*(3), 261–272.

Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*(2), 107-135.

Mousavi, S. A. (2015). *The effect of person misfit on item parameter estimation: A simulation study*. Doctoral dissertation, University of Alberta.

Mousavi, A., Tendeiro, J. N., & Younesi, J. (2016). Person fit assessment using the PerFit package in R. *The Quantitative Methods for Psychology. 12*(3), 232-242.

Olson, J. F., Martin, M. O., & Mullis, I. V.S. (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Petridou, A., & Williams, J. (2007). Accounting for aberrant test response patterns using multilevel models. *Journal of Educational Measurement, 44*(3), 227–247.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Thousand Oaks, CA: Sage.

Rudner, L. M., Bracey, G., & Skaggs, G. (1996). The use of a person-fit statistic with one high quality achievement test. *Applied Measurement in Education, 9*(1), 91–109.

Rupp, A. A. (2013). A systematic review of the methodology for person fit research in Item Response Theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling, 55*(1), 3-38.

Schmitt, N. S., Cortina, J. M., & Whitney, D. J. (1993). Appropriateness fit and criterion-related validity. *Applied Psychological Measurement, 17*(2), 143-150.

Sijtsma, K. (1986). A coefficient of deviance of response patterns. *Kwantitatieve Methoden, 7*(22), 131–145.

Sijtsma, K., & Meijer, R. R. (1992). A method for investigating the intersection of item response function in Mokken's nonparametric IRT model. *Applied Psychological Measurement, 16*(2), 149-157.

Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and Psychological Measurement, 45*(3), 433–444.

Snijders, T. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika, 66*(3), 331-342.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (1st ed.): Thousand Oaks: Sage Publications.

Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement, 20*(3), 221–230.

Tendeiro, J. N., Meijer, R. R., Schakel, L., & Maij-de Meij, A. M. (2013). Using cumulative sum statistics to detect inconsistencies in unproctored Internet testing. *Educational and Psychological Measurement, 73*(1), 143-161.

Trabin, T. E., & Weiss, D. J. (1983). The person response curve: fit of individuals to item response theory models. In D. J. Weiss (Ed.), *new horizons in testing*. New York: Academic Press.

Van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology, 13*(3), 267–298.

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: an application of evaluating person fit with two-level logistic regression. *Psychological Assessment, 20*(2), 159-168.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design. Rasch measurement*. Chicago: Mesa Press.