

توسعه فردی و تحول سازمانی

تعیین توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۹

شیوه استناددهی: محمدی زائر، سارا، انجم شعاع، زهرا، سلاجقه، سنجر، نظری، اکبر، و پور رشیدی، رستم. (۱۴۰۴). تعیین توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۹. توسعه فردی و تحول سازمانی، ۳(۱)، ۱-۱۲.

فاطمه آزادی^۱، اصغر مینایی^۲، مسعود گرامی پور^۳

۱. دانشجوی دکتری سنجش و اندازه‌گیری، گروه روانشناسی، دانشکده انسانی، واحد ساوه، دانشگاه آزاد اسلامی، ساوه، ایران
۲. دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
۳. دانشیار، گروه آموزشی برنامه‌ریزی درسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران

* ایمیل نویسنده مسئول: asghar.minaei@yahoo.com

چکیده

تاریخ چاپ: ۱۵ فروردین ۱۴۰۴

تاریخ پذیرش: ۰۴ فروردین ۱۴۰۴

تاریخ بازنگری: ۲۱ اسفند ۱۴۰۳

تاریخ ارسال: ۱۸ بهمن ۱۴۰۳

هدف این پژوهش تعیین توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۹ بود. جامعه آماری پژوهش حاضر شامل کلیه شرکت‌کنندگان در آزمون علوم پایه هشتم تیمز در سال ۲۰۱۹ است. نمونه این پژوهش شامل کلیه شرکت‌کنندگان ایرانی در آزمون علوم پایه هشتم تیمز بوده است. در این مطالعه، سبک‌های مختلف پاسخ‌دهی، از جمله افراطی، میانی و تصدیقی و تأثیر آن‌ها بر توانایی افراد بررسی شد. داده‌ها با استفاده از مدل‌های مختلف IRT تحلیل شدند. نتایج نشان داد در سال ۲۰۱۹، سبک تصدیقی (ARS) تأثیر قابل‌توجهی بر نتایج آزمون‌ها داشت و به افزایش واریانس نمرات منجر شد. همچنین، تفاوت‌های جنسیتی نشان داد که پسران تمایل بیشتری به سبک‌های افراطی و دختران به سبک‌های میانی داشتند. یافته‌های پژوهش نشان می‌دهد که سبک‌های پاسخ‌دهی می‌توانند بر توانایی افراد تأثیرگذار باشند. این نتایج می‌تواند به بهبود طراحی آزمون‌های استاندارد، کاهش سوگیری‌های فرهنگی و ارائه راهکارهایی برای ارزیابی دقیق‌تر در سطوح ملی و بین‌المللی کمک کند.

کلیدواژگان: سبک‌های پاسخ، آزمون‌های تیمز ۲۰۱۹، توانایی افراد.

این مقاله متعلق به نویسنده است. انتشار این مقاله به صورت

دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0)

صورت گرفته است.



Personal Development and Organizational Transformation

Determining the Ability of Iranian Students Based on Their Response Style in the 2019 Eighth Grade TIMSS Science Assessment

Fatemeh Azadi¹, Asghar Minaee^{2*}, Masoud Geramipour³

1. Ph.D Candidate in psychology, Sav.C., Islamic Azad University, Saveh, Iran

2. Associate Professor, Department of Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabatabaei University, Tehran, Iran

3. Associate Professor, Department of Curriculum Planning Education, Faculty of Education and Psychology, Kharazmi University, Tehran, Iran

*Corresponding Author's Email: asghar.minaei@yahoo.com

How to cite: Azadi, F., Minaee, A., & Geramipour, M. (2025). Determining the Ability of Iranian Students Based on Their Response Style in the 2019 Eighth Grade TIMSS Science Assessment. *Personal Development and Organizational Transformation*, 3(1), 1-12.

Abstract

The aim of this study was to determine the ability of Iranian students based on their response style in the 2019 Eighth Grade TIMSS Science Assessment. The statistical population of this study included all participants in the 2019 TIMSS Eighth Grade Science Assessment. The sample of this study consisted of all Iranian participants in the TIMSS Eighth Grade Science Assessment. In this study, different response styles, including extreme, middle, and acquiescent styles, and their effects on individuals' abilities were examined. The data were analyzed using various Item Response Theory (IRT) models. The results showed that in 2019, the acquiescent response style (ARS) had a significant effect on test results and led to an increase in the variance of scores. Furthermore, gender differences indicated that boys tended to adopt extreme styles more frequently, while girls were more inclined toward middle styles. The findings of the study suggest that response styles can influence individuals' abilities. These results may contribute to improving the design of standardized tests, reducing cultural biases, and providing solutions for more accurate assessment at both national and international levels.

Keywords: response styles, 2019 TIMSS assessments, individual ability.

Submit Date: 07 February 2025

Revise Date: 12 March 2025

Accept Date: 24 March 2025

Publish Date: 08 April 2025



© 2025 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

مقدمه

میزان پیشرفت تحصیلی دانش آموزان همواره از شاخص‌های موفقیت در فعالیت‌های علمی و رشد و توسعه به شمار می‌رود و از اهداف اصلی هر نظام آموزشی است، از این رو نظام‌های آموزشی به‌عنوان بازوی اجرایی در تلاش برای ارائه رویکردها و روش‌هایی برای رسیدن به اهداف نظام آموزش هستند تا جامعه هدف که همان دانش آموزان می‌باشند، به حد مطلوبی از پیشرفت تحصیلی در زمینه‌های مختلف علوم رایج دست یابند (Sahibzada et al., 2022). از مهمترین ارزشیابی‌هایی که کیفیت یادگیری دانش آموزان را در نظام‌های آموزشی جهان می‌سنجد، می‌توان به آزمون بین‌المللی پیشرفت تحصیلی تیمز اشاره کرد (Kalagarastani et al., 2018). از دستاوردهای دیگر آزمون تیمز این است که میزان دانش دانش آموزان را در ریاضیات، علوم و مهارت‌ها در زمینه‌های مختلف زندگی نشان می‌دهد (Ismail et al., 2018). ایران تاکنون در پنج مطالعه بین‌المللی تیمز در پایه چهارم ابتدایی در طول سال‌های ۱۹۹۵، ۲۰۰۳، ۲۰۰۷، ۲۰۱۱ و ۲۰۱۵ شرکت کرده است (Bairamipour et al., 2020). در آزمون تیمز ۲۰۱۹، ایران با شرکت ۴۳۹۶ دانش‌آموز پایه هشتم در حوزه علوم رتبه ۳۹ از ۴۳ کشور را کسب کرد. در این آزمون میانگین نمره دانش آموزان ایرانی در درس علوم پایه هشتم ۴۴۹ گزارش شد که ۱۵۹ نمره پایین‌تر از کشور برتر (سنگاپور با نمره ۶۰۸) و پایین‌تر از سطح متوسط بین‌المللی (۵۰۰ نمره) قرار دارد (Zia Nejad Shirazi et al., 2022). این نمره نشان‌دهنده افت ۷ نمره‌ای نسبت به سال ۲۰۱۵ (۴۵۶) و اختلاف ۲۶ نمره‌ای با بالاترین نمره ایران در سال ۲۰۱۱ (۴۷۴) است (Aghaei et al., 2020). حدود ۲۸٪ دانش‌آموزان ایرانی حتی به حداقل نمره (۴۰۰) نرسیدند، درحالی‌که میانگین جهانی این شاخص ۱۵٪ بود. در بین ۳۹ کشور شرکت‌کننده، ایران در علوم پایه هشتم رتبه ۳۲ را کسب کرد و از کشورهایی مانند قطر، امارات، بحرین و ارمنستان عقب ماند (Zia Nejad Shirazi & Qaltash, 2016). کشورهایی مانند امارات، قطر و ارمنستان با سرمایه‌گذاری در آموزش و استخدام کارشناسان بین‌المللی، رشد قابل توجهی داشتند، درحالی‌که ایران با روند کند بهبودی مواجه شد؛ بنابراین مدل‌های آموزشی جدیدی مطابق با الگوهای در حال تغییر در صنعت، تکنولوژی و آموزش مورد نیاز است (Aleksankov, 2017). در این زمینه پژوهش یافتیان و عباسی (۱۴۰۳) با مقایسه مسائل کتاب‌های درسی ریاضی ایران و ژاپن دریافتند عملکرد دانش‌آموزان ایرانی در برخی مسائل به نسبت هم‌تایان ژاپنی پایین‌تر بود (Yaftian & Abbasi, 2024).

ایران در میان ۱۶ کشور منطقه، رتبه دهم را در علوم پایه هشتم به دست آورد که نشان‌دهنده وضعیت نامناسب در مقایسه با همسایگان است. سیستم آموزشی ایران بر حفظیات و حل مسائل ساده تمرکز دارد و دانش‌آموزان در درک معنادار مفاهیم علمی ضعف نشان می‌دهند (Minaei, 2013). برنامه‌ریزان آموزشی در سراسر جهان تمام تلاش خود را به کار می‌گیرند تا برنامه‌های آموزشی نسل بعدی دانش‌آموزان به حدی از سواد تکنولوژیکی برسند که در مواجهه با رقابت‌های فزاینده اقتصادی به موضوعاتی مانند علوم، فناوری، مهندسی و ریاضیات مسلط شوند (Tsupros et al., 2009). جذب معلمان دوره ابتدایی بدون آموزش تخصصی (مانند انتقال معلمان متوسطه به ابتدایی) و نبود انگیزه در معلمان به دلیل مشکلات معیشتی، کیفیت تدریس را کاهش داده است. مدارس شهری و خاص (مانند نمونه دولتی) امکانات بیشتری دریافت می‌کنند، درحالی‌که ۷۰٪ دانش‌آموزان مدارس روستایی به حداقل استانداردهای یادگیری دست نمی‌یابند. تعطیلات طولانی و آلودگی هوا باعث کاهش ساعات مفید آموزشی شده است. برخلاف بهبود نسبی در ریاضیات، عملکرد علوم پایه هشتم از سال ۲۰۱۱ تا ۲۰۱۹ روندی نزولی داشته است (Piroozi & Kazemi, 2022). این موضوع نشان‌دهنده بی‌توجهی به آموزش مفهومی و آزمایش محور در علوم است. اگرچه داده‌های جنسیتی در گزارش‌ها کمتر برجسته شده، اما در مدارس روستایی پسرانه، ۷۰٪ دانش‌آموزان به حداقل سطح یادگیری نمی‌رسند

(Khalilnejadi et al., 2024). وان دیگر و آدامز (۲۰۱۰) دریافت که رابطه بین نگرش و موفقیت متغیر است، بدین معنی که دانش‌آموزان علیرغم پیشرفت بالا در دروس، نگرش منفی نسبت به درس خواندن و مطالعه دارند؛ به عبارت دیگر ممکن است رابطه‌ی مثبت بین نگرش و موفقیت در یک فرهنگ و کشور وجود داشته باشد، درحالی‌که ممکن است در فرهنگ دیگر رابطه‌ی منفی بین نگرش و موفقیت وجود داشته باشد (Jeon & De Boeck, 2016). از این رو نتایج متفاوت^۱ در سطوح فردی و گروهی اعتبار نتایج به دست آمده از مطالعات را کاهش می‌دهد. یکی از دلایل این تفاوت‌ها، سبک‌های پاسخ^۲ است که عبارت است از گرایش سیستماتیک به پاسخ‌دهی به سؤال‌های یک پرسشنامه بدون توجه به محتوای آن‌ها (Khorramdel et al., 2019).

سبک پاسخ‌دهی به عنوان یک عامل غیرشناختی، به دنبال شناسایی الگوهای رفتاری دانش‌آموزان در مواجهه با سؤالات آزمون است. سبک پاسخ‌دهی شامل رفتارهایی مانند حدس زدن، حذف سؤال^۳ یا تخصیص زمان ناکافی است که ممکن است نتایج آزمون را تحت تأثیر قرار دهد (Wise & DeMars, 2005). در داده‌های تیمز ۲۰۱۹، دانش‌آموزان ایرانی نرخ بالاتری از حذف سؤالات را نسبت به میانگین جهانی نشان دادند (Mullis et al., 2020). نتایج پژوهش‌های پیشین نشان می‌دهد که سبک‌های پاسخ‌دهی افراطی^۴ به عنوان یکی از الگوهای رفتاری رایج در آزمون‌های روان‌سنجی، می‌تواند اعتبار نتایج آزمون را تحت تأثیر قرار دهد. هنگامی که پاسخ‌دهندگان به جای انتخاب گزینه‌های میانی، تمایل به انتخاب گزینه‌های انتهایی مقیاس (مثبت یا منفی) داشته باشند، این رفتار نه تنها برآورد واریانس را افزایش می‌دهد، بلکه ممکن است منجر به سوگیری سیستماتیک در داده‌ها شود. به عنوان مثال، تمایل به انتخاب گزینه‌های انتهایی مثبت (مانند «کاملاً موافقم») می‌تواند نشان‌دهنده سوگیری خوش‌بینانه باشد، درحالی‌که گرایش به گزینه‌های انتهایی منفی («کاملاً مخالفم») ممکن است سوگیری بدبینانه را القا کند (Bolt & Newton, 2011). از آنجاکه همبستگی بین متغیرها و واریانس نمرات به یکدیگر وابسته هستند، وجود سبک پاسخ نهایی می‌تواند روابط آماری را مخدوش سازد. به ویژه، افزایش واریانس ناشی از سبک پاسخ نهایی باعث کاهش مقادیر ضریب همبستگی بین متغیرها می‌شود، چرا که پراکندگی غیرواقعی داده‌ها، قدرت ارتباط خطی را تضعیف می‌کند (Wetzel et al., 2013). این مسئله در روش‌های آماری مبتنی بر ماتریس همبستگی، مانند تحلیل عاملی اکتشافی یا مدل‌سازی معادلات ساختاری، چالش‌های جدی ایجاد می‌کند، زیرا سبک پاسخ نهایی می‌تواند ساختار عاملی و برازش مدل‌ها را تغییر دهد (Dibek & Cikrikci, 2021).

علاوه بر سبک پاسخ نهایی، سبک پاسخ تصدیق یا مخالفت^۵ و سبک پاسخ نمره میانی^۶ و غیره نیز بر نتایج آزمون‌ها تأثیر می‌گذارند (Kupffer et al., 2024). سبک پاسخ تصدیق عبارت است از گرایش فرد به دادن پاسخ‌های مثبت یا موافق با سؤال‌ها بدون در نظر گرفتن محتوای آن‌ها است (Wang et al., 2022) که به آن «گرایش به موافقت» نیز گفته می‌شود. از یک سو، همان‌طور که گفته شد، سبک پاسخ نهایی گرایش به انتخاب گزینه‌های انتهایی مقوله‌های پاسخ مانند «کاملاً موافق/کاملاً مخالف» است (Dimock, 2019). به طور خاص، تفاوت‌ها در گرایش به پاسخ به گزینه‌های انتهایی مثبت یا منفی یعنی کاملاً موافق یا کاملاً مخالف نتایج ابزارهای گروهی را مخدوش می‌کند که بر میزان همبستگی نمرات سؤال‌ها با نمره‌ی کل آزمون تأثیر می‌گذارد یا اتکاپذیری را در رابطه با همسانی درونی افزایش یا کاهش می‌دهد. علاوه بر این، سبک پاسخ افراطی بر سطح همبستگی‌ها بین معیارها و در نتیجه نتایج به دست آمده از تحلیل‌های عاملی، تأثیر می‌گذارد. از سوی دیگر،

1 interchangeable

2 Response styles

3 Omission Rate

4 Extreme Response Style; ERS

5 Acquiescence Response Style (ARS) or disacquiescence(DRS)

6 middle response styles (MRS)

سبک پاسخ افراطی معمولاً باعث می‌شود که میانگین نمرات سؤال‌ها کمتر یا بیشتر از نمرات واقعی‌اش تخمین زده شود (Bolt & Newton, 2011) که منتج به ایجاد سوگیری در نتایج می‌شود. این نمره‌های دارای سوگیری ممکن است منتج به نتایج خطای اندازه‌گیری شوند (Khalilnejadi et al., 2024).

از آنجاکه نگرش‌ها و سبک پاسخ ممکن است از فرهنگی نسبت به فرهنگ دیگر متفاوت باشند، دستیابی به تفاوت میان این نگرش‌ها می‌تواند تأثیر تفاوت‌های فرهنگی بر نگرش و سبک‌های پاسخ را مشخص کند (Aghaei et al., 2020). مطالعات نشان می‌دهند که دانش‌آموزان در فرهنگ‌های مختلف، سبک‌های پاسخ متفاوتی دارند. به‌عنوان مثال، دانش‌آموزان آسیایی تمایل کمتری به حدس زدن دارند (Lee & Jia, 2014). از این جهت، برای نشان دادن نمرات واقعی آزمون‌ها، تأثیر سبک‌های پاسخ باید مورد بررسی قرار گیرد. در مقالات تحقیقاتی، هیچ روش واحد مورد اجماعی برای چاره‌جویی درباره اثرات منفی سبک پاسخ وجود ندارد، اما در این مورد اتفاق نظر وجود دارد که بر سنجش نگرش‌ها تأثیر منفی می‌گذارند. تفاوت در رویکردهای سبک پاسخ باعث مخدوش شدن نتایج و کنترل سیستماتیک سوگیری‌ها می‌گردد. پژوهش‌های اخیر تأکید می‌کنند که خودکارآمدی تحصیلی و اضطراب امتحان به‌طور غیرمستقیم از طریق سبک پاسخ‌دهی بر نمرات تأثیر می‌گذارند (Naeimi et al., 2023; Ulitzsch et al., 2020). سبک‌های پاسخ به‌عنوان یکی از عوامل مخدوش‌کننده در آزمون‌های روان‌سنجی، می‌توانند بر اعتبار نمرات تأثیر منفی بگذارند. دیبک و کیکریکی (۲۰۲۱) با بررسی سبک پاسخ افراطی تمایل به انتخاب گزینه‌های انتهایی مقیاس را تأکید کردند که این پدیده نه تنها واریانس نمرات را افزایش می‌دهد، بلکه می‌تواند منجر به سوگیری سیستماتیک در تخمین پارامترهای آزمون شود (Dibek & Cikrikci, 2021). این یافته با پژوهش کوپفر و همکاران (۲۰۲۴) همسوست که سبک پاسخ افراطی را عاملی کلیدی در کاهش دقت مدل‌های اندازه‌گیری معرفی کردند (Kupffer et al., 2024).

از سوی دیگر، نادیده گرفتن سبک پاسخ در فرآیند مدل‌سازی می‌تواند به تعریف نادرست ساختار داده‌ها بینجامد. توتز و برگر (۲۰۱۶) در مطالعه‌ای نشان دادند که عدم کنترل سبک پاسخ افراطی در مدل‌های رگرسیونی، برآوردهای مربوط به سختی سؤالات^۱ و توانایی پاسخ‌دهندگان^۲ را به شدت تحت تأثیر قرار می‌دهد (Tutz & Berger, 2016). این مسئله به‌ویژه در مدل‌های IRT چندبعدی مشهود است، چرا که سبک‌های پاسخ به‌عنوان یک بعد پنهان اضافی، واریانس نامربوط را به مدل تزریق می‌کنند (Falk & Ju, 2020). در زمینه آموزش، دیبک (۲۰۲۰) با تحلیل داده‌های تیمز ۲۰۱۵ و استفاده از مدل IRTree، دریافت که دانش‌آموزان با نگرش مثبت به ریاضیات، تمایل بیشتری به انتخاب گزینه‌های میانی مقیاس لیکرت دارند (Dibek & Cikrikci, 2021). این الگو با ویژگی‌های فرهنگ‌های جمع‌گرا مرتبط است که در آن حفظ هارمونی اجتماعی از طریق اجتناب از پاسخ‌های افراطی تقویت می‌شود (Bonjeer & Vonkova, 2024)؛ بنابراین، این واقعیت که سبک‌های پاسخ منتج به استنتاج‌های غلط و کاربردهای نادرست برای تصمیمات و خط‌مشی‌های آموزشی در سطح ملی می‌شوند باعث می‌شود تا اصلاح اثرات این سبک‌های پاسخ بر روی امتیازات ساختارهای روانشناسی، همچون نگرش‌ها و آزمون‌های توانمندی^۳، مورد توجه قرار گیرد. از آنجایی که در پژوهش‌های بررسی‌شده تاکنون پژوهشی که به شناسایی سبک‌های پاسخ در تیمز ۲۰۱۹ پرداخته باشد، یافت نشد و مشخص نیست که کدامیک از سبک‌های پاسخ تأثیر بیشتری بر سؤالات آزمون دارد و همچنین با توجه به اینکه مقایسه آن‌ها در بین دانش‌آموزان ایرانی می‌تواند بر ویژگی‌های روان‌سنجی آزمون تیمز تأثیر بگذارد. از این رو به دلیل خلأ پژوهشی موجود در زمینه بررسی طولی

¹ Item Difficulty

² Ability Parameters

سبک‌های پاسخ در آزمون‌های تیمز، در این پژوهش به شناسایی سبک‌های پاسخ دانش آموزان پایه هشتم در تیمز ۲۰۱۹ با استفاده از مدل‌های نظریه سؤال پاسخ (IRT) و نیز اثربخشی سبک پاسخ در آزمون علوم تیمز پایه هشتم ۲۰۱۹ بر توانایی افراد پرداخته شده است.

روش‌شناسی

این پژوهش از نظر هدف کاربردی و ازلحاظ روش از نوع همبستگی می‌باشد این پژوهش برای شناسایی و مقایسه سبک‌های پاسخ از داده‌های حاصل از اجرای آزمون‌های تیمز (۲۰۱۹) استفاده کرده است بنابراین می‌توان گفت پژوهش حاضر در دسته تحقیقات تحلیل ثانویه قرار دارد. جامعه آماری پژوهش حاضر شامل کلیه شرکت‌کنندگان در آزمون علوم پایه هشتم تیمز ۲۰۱۹ است. نمونه این پژوهش شامل کلیه شرکت‌کنندگان ایرانی در آزمون علوم پایه هشتم تیمز بوده است. توجه به اینکه نمونه موردنظر بنا به درخواست پژوهشگر از بانک داده‌های آزمون تیمز تهیه شده و پژوهشگر هیچ دخالتی در امر انتخاب نداشته و به صورت تصادفی داده‌ها از بین کلیه داده‌های موجود انتخاب شده است، می‌توان روش نمونه‌گیری پژوهش را نمونه‌گیری ساده محسوب کرد. تعداد افراد پاسخ‌دهنده در آزمون علوم تیمز ۲۰۱۹ پایه هشتم دانش‌آموزان ایران (سؤال ۲۲، ۵۶۵۲؛ سؤال ۲۳، ۵۸۰۱؛ سؤال ۲۴، ۵۶۰۰؛ سؤال ۲۵، ۵۷۴۴).

ابزار پژوهش

ابزار مورد استفاده در این پژوهش آزمون علوم تیمز ۲۰۱۹ بوده است این آزمون یکی از پروژه‌های کلیدی انجمن بین‌المللی ارزیابی پیشرفت تحصیلی (IEA) است که با هدف مقایسه نظام‌مند عملکرد سیستم‌های آموزشی جهانی و تسهیل یادگیری سیاستی از طریق داده‌های تطبیقی طراحی شده است (Mullis & Martin, 2017). این مطالعه از سال ۱۹۹۵ به صورت چرخه‌های چهارساله اجرا شده و به کشورها امکان می‌دهد روند تغییرات آموزشی خود را در بلندمدت تحلیل کنند (Fishbein et al., 2021). سؤالات ریاضیات در این آزمون مربوط به سنجش مهارت‌ها در حوزه‌های جبر، هندسه، آمار و تحلیل داده‌ها؛ در زمینه علوم شامل ارزیابی دانش زیست‌شناسی، شیمی، فیزیک و علوم زمین؛ و تمرکز بر حل مسئله و تفکر تحلیلی به جای محفوظات است (Wiberg & Rolfsman, 2019). علاوه بر این تیمز داده‌های متنی را در برنامه تحصیلی، آموزشی و عوامل مرتبط با منابع فراهم می‌کند که می‌توانند بر فرآیند آموزش و یادگیری تأثیر بگذارند آزمون تیمز علاوه بر سنجش عملکرد تحصیلی، داده‌های زمینه‌ای را از چهار لایه (دانش‌آموزان، معلمان، مدیران مدارس، سیاست‌گذاران ملی) جمع‌آوری می‌کند (Pongsophon, 2023). این داده‌ها از طریق پرسشنامه‌های استاندارد شده جمع‌آوری می‌شوند و امکان شناسایی عوامل مؤثر بر یادگیری (مانند نابرابری جنسیتی، تأثیر فناوری آموزشی) را فراهم می‌کنند (Meinck & Brese, 2019). آزمون‌های تیمز شامل چندین دفترچه در حوزه‌های ریاضی و علوم است که یک دفترچه مخصوص دانش‌آموزان یک دفترچه معلم، دفترچه برنامه تحصیلی، دفترچه برنامه‌ریزی و علوم اختصاصی و علوم عمومی است، در سال ۲۰۱۹، تعداد سؤالات آزمون مدرسه ۲۱ سؤال، برنامه‌ریزی درسی عمومی ۱۵ و قسمت علوم ۷ سؤال، دفترچه دانش‌آموزان علوم عمومی ۲۷، دفترچه دانش‌آموزان علوم اختصاصی ۴۳ و دفترچه معلم ۲۲ سؤال بوده است. به دلیل نوع پاسخ‌های تیمز سبک پاسخ افراطی برای سؤالات بررسی شد و با توجه به پیشینه پژوهش‌های انجام شده از مدل IRTree استفاده شده است، در این پژوهش از مدل سلسله مراتبی با اثرهای ثابت و تصادفی استفاده شده است، مدل‌های سلسله‌مراتبی که به آن‌ها مدل‌های چند سطحی نیز گفته می‌شود، ابزارهای آماری قدرتمندی هستند که برای تحلیل داده‌هایی با ساختار آشیانه‌ای استفاده می‌شوند. در این مدل‌ها، داده‌ها در سطوح مختلف سازمان‌دهی شده‌اند و این ساختار سلسله‌مراتبی به محققان امکان می‌دهد تا تأثیرات متغیرها را در سطوح مختلف به صورت همزمان بررسی کنند.

در مدل اول تفاوت‌های فردی به‌عنوان اثر تصادفی و سؤالات به‌عنوان اثر ثابت وارد تحلیل شدند و در مدل دوم علاوه بر آیت‌ها، گره هم به‌عنوان اثر ثابت گرفته شد، گره‌ها واحدهای تحلیلی هستند که در سطوح مختلف سلسله‌مراتب قرار دارند، در این مدل ۴ گره تعریف شد که گره اول شامل گروه موافق بودند و در این گره هم دو گروه تعریف شدند، گروهی که شدیدتر و گروه بعدی گروه کمتر بوده است و گرهی دیگر افراد مخالف بودند که این گره هم به دو گره افراد شدیدتر و افرادی که به گزینه‌های پایین‌تر پاسخ داده بودند، قرار گرفتند. در مدل سوم در اثر تصادفی و اثر ثابت نیز گره در نظر گرفته شد. نمره‌گذاری سؤالات نگرشی به‌صورت طیف لیکرت (کاملاً موافقم=۱ تا کاملاً مخالفم=۴) نمره‌گذاری گردید و فقط سؤالات اعتماد به نفس به‌صورت معکوس نمره‌گذاری شدند.

در پژوهش‌های پیشین به‌منظور بررسی ویژگی‌های روان‌سنجی آزمون‌های تیمز از روش‌های مختلف مانند نظریه سؤال پاسخ به‌منظور برآورد توانایی افراد پرداخته شده است، منحنی ویژگی سؤال، تابع آگاهی سؤال، تابع آگاهی آزمون و منحنی خطای استاندارد آزمون، بیشینه آگاهی سؤالات، بیشینه آگاهی آزمون به‌وسیله نرم‌افزارهای SPSS، WinBugs، BILOG-MG3، NOHARM و R استفاده شده است، همچنین شریفی یگانه و همکاران (۱۳۹۷) به‌منظور کارکرد افتراقی آزمون تیمز از دو روش مانتل-هنزل و نظریه سؤال پاسخ استفاده کرده است. شناسایی سبک‌های پاسخ دانش‌آموزان پایه هشتم تیمز و برآورد آن‌ها بر پارامترهای سؤال‌های علوم و توانایی افراد با استفاده از نظریه سؤال پاسخ روشی است که در دانش‌آموزان ایرانی انجام نشده است.

یافته‌ها

اطلاعات توصیفی مربوط به توانایی دانش‌آموزان برحسب سبک پاسخ آن‌ها در آزمون تیمز ۲۰۱۹ در جدول (۱) گزارش شده است.

جدول ۱. اطلاعات توصیفی توانایی افراد برحسب سبک‌های پاسخ در آزمون تیمز ۲۰۱۹

گروه‌ها	تعداد	میانگین	انحراف استاندارد	کمترین	بیشترین
سبک پاسخ افراطی	۲۰۶۹	۰/۱۵۹	۰/۹۴۵	-۵/۲۶۳	۴/۰۹۳
سبک پاسخ میانی	۳۶۰	-۰/۰۵۲	۰/۹۴۱	-۴/۸۲۶	۴/۲۳۸
سبک پاسخ رضایت	۱۴۰۲	-۰/۱۷	۰/۸۸۵	-۰۴/۵۲۳	۳/۹۴۴
سبک پاسخ معمولی	۲۱۴۹	-۰/۰۱۰	۰/۸۶۹	-۳/۴۳۷	۳/۹۴۴
مجموع	۵۹۸۰	۰/۰۰۸	۰/۹۱۲	-۵/۲۶۳	۴/۲۳۸

با توجه به اطلاعات جدول (۱) میانگین توانایی سبک پاسخ افراطی از سایر گروه‌ها بهتر بوده است ($M=0/159$) و میانگین توانایی سبک پاسخ رضایت ($M=-0/17$) از همه کمتر شده است. برای قضاوت درباره معناداری تفاوت‌ها نتایج آزمون تحلیل واریانس در جدول (۲) گزارش شده است.

جدول ۲. نتایج آزمون تحلیل واریانس جهت مقایسه توانایی افراد با سبک پاسخ متفاوت در آزمون تیمز ۲۰۱۹

منابع تغییر	مجموع مجذورات	درجه آزادی	میانگین مجذورات	F	سطح معناداری
بین گروهی	۹۳/۸۱۲	۳	۳۱/۲۷۱	۳۸/۲۵۹	۰/۰۰۰۱
درون گروهی	۴۸۸۴/۴۱۵	۵۹۷۶	۰/۸۱۷		
مجموع	۴۹۷۸/۲۲۸	۵۹۷۹			

بر اساس اطلاعات جدول (۲) نتیجه آزمون تحلیل واریانس به لحاظ آماری معنادار است ($P < 0/05$ و $F = 38/259$)؛ بنابراین بین گروه‌ها با سبک پاسخ متفاوت به لحاظ توانایی‌های شناختی تفاوت وجود دارد. در ادامه برای مقایسه دوتایی بین گروه‌ها از آزمون تعقیبی بونفرونی استفاده شد که نتایج آن در جدول (۳) گزارش شده است.

جدول ۳. نتایج آزمون تعقیبی بونفرونی در آزمون تیمز ۲۰۱۹

گروه ۱	گروه ۲	تفاوت میانگین	خطای استاندارد	سطح معناداری
سبک پاسخ معمولی	سبک پاسخ افراطی	-۰/۱۶۹	۰/۰۲۸	۰/۰۰۰۱
	سبک پاسخ میانی	۰/۰۴۲	۰/۰۵۱	۰/۹۹۹
	سبک پاسخ رضایت	۰/۱۶۰	۰/۰۳۱	۰/۰۰۰۱
سبک پاسخ افراطی	سبک پاسخ میانی	۰/۲۱۱	۰/۰۵۲	۰/۰۰۰۱
	سبک پاسخ رضایت	۰/۳۲۹	۰/۰۳۱	۰/۰۰۰۱
سبک پاسخ میانی	سبک پاسخ رضایت	۰/۱۱۸	۰/۰۵۳	۰/۱۶۲

با توجه به اطلاعات جدول (۳) بین توانایی افراد سبک پاسخ معمولی و توانایی افراد سبک پاسخ میانی تفاوت معنادار وجود ندارد ($P > 0/05$). همچنین بین توانایی افراد سبک پاسخ میانی با افراد سبک پاسخ رضایت نیز تفاوت معنادار وجود ندارد ($P > 0/05$). بین توانایی افراد با سبک پاسخ افراطی با توانایی سایر گروه‌ها تفاوت معنادار آماری وجود دارد ($P < 0/05$). همچنین بین توانایی افراد سبک پاسخ معمولی با توانایی افراد سبک پاسخ رضایت نیز تفاوت معنادار آماری وجود دارد ($P < 0/05$).

بحث و نتیجه‌گیری

نتایج تحلیل واریانس یک‌طرفه برای داده‌های آزمون تیمز پایه هشتم ۲۰۱۹ نشان داد که توانایی شناختی دانش‌آموزان بر حسب سبک پاسخ آنان تفاوت معناداری دارد. این نتایج بار دیگر بر تأثیر سبک پاسخ بر عملکرد دانش‌آموزان تأکید می‌کند. رایت (۱۹۶۸) سنجش نامتغیر بودن پارامتر توانایی را از طریق دو یا بیشتر از دو مجموعه سؤال که به‌طور فاحشی از نظر دشواری باهم تفاوت دارند تعریف کرده است. مجموعه سؤال‌ها از بانک سؤالاتی که بر اساس آن توانایی سنجیده می‌شود ساخته می‌شوند (Minaei, 2013; Minaei & Falsafi Nejad, 2010). دانش‌آموزان دارای سبک پاسخ افراطی بیشترین میانگین توانایی را داشتند. دانش‌آموزان دارای سبک پاسخ رضایت کمترین میانگین توانایی را نشان دادند. میانگین توانایی دانش‌آموزان با سبک پاسخ میانی و سبک پاسخ معمولی نزدیک به هم و در حد متوسط بودند. آزمون تعقیبی بونفرونی نشان داد، سبک پاسخ افراطی در مقایسه با سایر سبک‌ها به‌طور معناداری توانایی بیشتری دارد. بین دانش‌آموزان دارای سبک پاسخ معمولی و میانی تفاوت معناداری وجود ندارد. میان دانش‌آموزان دارای سبک پاسخ میانی و رضایت نیز تفاوت معناداری مشاهده نشد. تفاوت معناداری میان سبک پاسخ معمولی و رضایت وجود داشت. این سبک پاسخ به دلیل انتخاب‌های قطعی‌تر (کاملاً موافق یا کاملاً مخالف)، ممکن است نشان‌دهنده درک بهتر، اعتماد به نفس بیشتر و توانایی پردازش سریع‌تر اطلاعات باشد. در شرایط آزمون استاندارد، دانش‌آموزانی که سریع‌تر تصمیم می‌گیرند و از تردید اجتناب می‌کنند، ممکن است عملکرد بهتری داشته باشند. سبک پاسخ رضایت که معمولاً با گرایش به گزینه‌های میانی همراه است، می‌تواند ناشی از بی‌انگیزگی، اضطراب آزمون، یا تمایل به اجتناب از ریسک باشد. این رفتارها، در آزمون‌های استاندارد، معمولاً به عملکرد ضعیف‌تر منجر می‌شوند. کمترین میانگین توانایی در این گروه نشان می‌دهد که این دانش‌آموزان احتمالاً استراتژی‌های شناختی عمیق یا انگیزه کافی برای موفقیت در آزمون را ندارند. عدم تفاوت معنادار بین سبک پاسخ میانی و معمولی می‌تواند

بیانگر استراتژی‌های مشابه این دو گروه در پردازش اطلاعات باشد. اگرچه میانگین توانایی سبک پاسخ معمولی کمی بالاتر از سبک میانی است، اما این تفاوت از نظر آماری معنادار نیست که نشان‌دهنده همپوشانی در نحوه تصمیم‌گیری آن‌هاست. تحلیل داده‌های آزمون تیمز باعث درک فرآیندهای آموزشی و شناسایی موضوعات جدید که منجر به اصلاح سیستم نظام آموزشی می‌شود و می‌توان از اطلاعات آن برای تصمیم‌گیری و برنامه‌ریزی به‌منظور بهبود نظام آموزشی اقدام کرد (Wang et al., 2022). بر اساس نتایج پژوهش به مسئولین پیشنهاد می‌شود با توجه به تفاوت‌های مشاهده‌شده در سبک‌های پاسخ‌دهی دانش‌آموزان، می‌توان برنامه‌های آموزشی شخصی‌سازی‌شده طراحی کرد که به تقویت نقاط ضعف پاسخ‌دهندگان کمک کند. تحلیل سبک‌های پاسخ‌دهی در بازه‌های زمانی مختلف (۲۰۱۱، ۲۰۱۵ و ۲۰۱۹) می‌تواند به شناسایی تأثیر تغییرات نظام آموزشی، فناوری یا فرهنگی بر عملکرد دانش‌آموزان کمک کند. این نتایج می‌تواند به طراحی استراتژی‌های آموزشی بهتر در دوره‌های زمانی آینده منجر شود. از یافته‌های پژوهش می‌توان برای بهبود طراحی آزمون‌های استاندارد داخلی، مانند امتحانات نهایی یا کنکور، استفاده کرد تا این آزمون‌ها عادلانه‌تر و دقیق‌تر شوند. از طرفی محدودیت پژوهش یکی از چالش‌های اصلی، دسترسی محدود به داده‌های آزمون تیمز (۲۰۱۹) بوده است. برخی اطلاعات ممکن است ناقص یا غیرقابل استفاده باشند که تحلیل نتایج را تحت تأثیر قرار داده است. سبک‌های پاسخ‌دهی دانش‌آموزان ممکن است تحت تأثیر عواملی خارج از کنترل پژوهش (مانند استرس امتحان، شرایط اجتماعی) قرار گرفته باشد که در این پژوهش قابل اندازه‌گیری نبوده‌اند.

تشکر و قدردانی

از تمامی کسانی که در انجام این پژوهش ما را همراهی کردند تشکر و قدردانی به عمل می‌آید.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافی وجود ندارد.

موازین اخلاقی

در تمامی مراحل پژوهش حاضر اصول اخلاقی مرتبط با نشر و انجام پژوهش رعایت گردیده است.

Extended Abstract

Introduction

Academic achievement has long been recognized as a fundamental indicator of success in scientific and social progress, making it a main goal of educational systems (Sahibzada et al., 2022). Among the most important international assessments that evaluate students' learning quality worldwide is the TIMSS assessment (Kalagarastani et al., 2018). TIMSS measures students' knowledge in mathematics, science, and life-related skills (Ismail et al., 2018). Iran has participated in five TIMSS studies at the fourth-grade level between 1995 and 2015 (Bairamipour et al., 2020). In the 2019 TIMSS assessment, Iran ranked 39th out of 43 countries in eighth-grade science, with a mean score of 449, significantly lower than Singapore's 608 and below the international average of 500 (Zia Nejad Shirazi et al., 2022). Around 28% of Iranian students failed to meet even the basic proficiency level of 400, compared to the international average of 15% (Zia Nejad Shirazi & Qaltash, 2016). Countries like Qatar, the UAE, and Armenia have improved significantly through educational investments and international collaboration, while Iran has lagged (Aleksankov, 2017).

The Iranian educational system has been criticized for focusing excessively on memorization at the expense of meaningful understanding (Minaei, 2013). Internationally, policymakers are working to ensure students achieve technological literacy to compete globally (Tsupros et al., 2009). Factors such as inadequate teacher preparation, unequal resource distribution, and excessive school closures due to pollution have further impeded learning (Khalilnejadi et al., 2024; Piroozi & Kazemi, 2022). Response styles, including extreme and acquiescent patterns, have been shown to significantly affect test results, potentially inflating or deflating estimates of ability (Bolt & Newton, 2011; Wetzel et al., 2013). Such biases disrupt statistical analyses like factor analysis and SEM (Dibek & Cikrikci, 2021), and cultural factors influence these styles (Khorramdel et al., 2019; Wang et al., 2022). Despite extensive work internationally, studies focusing on Iranian students' response styles in TIMSS 2019 remain limited.

Methods and Materials

This study was an applied, correlational research based on secondary analysis of TIMSS 2019 eighth-grade science assessment data. The statistical population included all Iranian students who participated in the TIMSS assessment. Data sampling was simple random, extracted from the TIMSS database without researcher intervention. Responses from 5980 students were analyzed.

The assessment instrument was the official TIMSS 2019 Science Test, developed by the IEA. It evaluated content areas such as biology, chemistry, physics, and earth sciences and emphasized analytical skills over rote learning. Various hierarchical models, including IRTree models, were applied. Three models were tested: Model 1 included random effects for students and fixed effects for items, Model 2 added node effects, and Model 3 incorporated random node effects. Scoring for Likert-scale items ranged from 1 (Strongly Agree) to 4 (Strongly Disagree), with reverse scoring applied to self-confidence items.

Findings

Descriptive statistics revealed that students with an extreme response style had the highest mean ability (0.159), while students with an acquiescent response style had the lowest mean ability (-0.17). Students with middle and regular response styles had mean abilities close to each other.

Analysis of variance (ANOVA) showed a statistically significant difference in cognitive ability among the different response style groups ($F=38.259$, $p<0.0001$). Post-hoc Bonferroni tests indicated that students with extreme response styles had significantly higher abilities than those with middle, regular, or acquiescent styles. No significant difference was found between students with middle and regular styles or between middle and acquiescent styles. However, regular responders scored significantly higher than acquiescent responders.

Discussion and Conclusion

The findings of this study highlight the substantial influence of response style on the cognitive ability scores of Iranian eighth-grade students in the TIMSS 2019 science assessment. Students who tended toward extreme responses performed better cognitively, possibly reflecting greater decisiveness, confidence, or more efficient information processing. In contrast, students with an acquiescent response style demonstrated lower cognitive ability, which may indicate lower engagement, motivation, or test-taking skills.

The similarity between students with middle and regular response styles suggests that hesitation or moderation does not drastically differentiate performance levels compared to more polarized answering behaviors. These findings underscore the importance of accounting for response style in educational assessments to achieve more accurate and fair evaluations of student competencies.

From a practical perspective, recognizing the role of response styles can inform the design of educational interventions and assessment tools. Strategies to mitigate the impact of biased response styles could lead to more valid and equitable evaluations. Moreover, longitudinal analyses of response style trends across different TIMSS cycles could offer valuable insights into the impact of educational reforms and social changes on student learning behaviors.

Despite its contributions, the study faces limitations, including reliance on available secondary data and the inability to control for all external factors affecting student responses. Future research should consider mixed-method approaches, integrating qualitative insights into students' attitudes and test-taking behaviors to provide a more comprehensive understanding of the phenomenon. Tailored interventions that help students engage more critically and confidently with assessment tasks could further enhance the validity and reliability of large-scale educational evaluations.

References

- Aghaei, A., Madani, M. M., & Abdollahi, S. (2020). Examination of the Dimensions of the TIMSS Test and the Performance of Iranian Students in This Test. Third International Conference on Psychology, Educational Sciences, Social Sciences, and Humanities, <https://en.civilica.com/doc/1163106/>
- Aleksankov, A. M. (2017). The Fourth industrial revolution and modernization of education: international experience. *Strategic Priorities*, 1(13), 53-69.
- Bairamipour, A., Samsari, Z., & Hashemi, S. E. (2020). The Relationship Between Background Factors and Iranian Students' Mathematics Performance in TIMSS 2015. *Educational Innovations*, 19(3), 33-62. https://noavaryedu.oerp.ir/article_114613.html
- Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and psychological measurement*, 71(5), 814-833. <https://doi.org/10.1177/0013164410388411>
- Bonjeer, T., & Vonkova, H. (2024). Relationships between response styles and the Hofstede and GLOBE dimensions of culture in a sample of adolescents from 33 countries. *Cross-Cultural Research*, 58(2-3), 180-207. <https://doi.org/10.1177/10693971231203759>
- Dibek, M. I., & Cikrikci, R. N. (2021). Modelling of the Attitude-Achievement Paradox in TIMSS 2015 with Respect to the Extreme Response Style Using Multidimensional Item Response Theory. *International Journal of Progressive Education*, 17(2), 194-209. <https://doi.org/10.29329/ijpe.2021.332.12>
- Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center*, 17(1), 1-7. <https://www.pewresearch.org/short-reads/2019/01/17/where-millennials-end-and-generation-z-begins/>
- Falk, C. F., & Ju, U. (2020). Estimation of response styles using the multidimensional nominal response model: A tutorial and comparison with sum scores. *Frontiers in psychology*, 11, 72. <https://doi.org/10.3389/fpsyg.2020.00072>
- Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database. <https://timssandpirls.bc.edu/timss2019/international-database>
- Ismail, M. E., Samsudin, M. A., Amin, N. F. M., Kamarudin, N., Daud, K. A. M., & Halim, L. (2018). Contributing factors to science achievement in TIMSS Malaysia: Direct model and indirect model. *International Journal of Engineering & Technology*, 7(4.30), 423-428. <https://doi.org/10.14419/ijet.v7i4.30.22354>
- Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods*, 48, 1070-1085. <https://doi.org/10.3758/s13428-015-0631-y>
- Kalagarastani, A., Gorji, M., Gorji Daroonkalai, M., & Mousazadeh, S. M. R. (2018). The International TIMSS Test and the Performance of Iranian Students in Mathematics. Third National Conference on New Approaches in Education and Research, Mahmoudabad, <https://en.civilica.com/doc/829932/>
- Khalilnejadi, A., Mohsenpour, M., & Moghaddam, A. (2024). Differential Functioning of Fourth-Grade Mathematics Test Items in the TIMSS 2019 Study by Gender. *Educational Measurement and Evaluation Studies*, 14(45), 7-26. https://jresearch.sanjesh.org/article_713306.html
- Khorramdel, L., von Davier, M., & Pokropek, A. (2019). Combining mixture distribution and multidimensional IRTree models for the measurement of extreme response styles. *British Journal of Mathematical and Statistical Psychology*, 72(3), 538-559. <https://doi.org/10.1111/bmsp.12179>
- Kupffer, R., Frick, S., & Wetzel, E. (2024). Detecting careless responding in multidimensional forced-choice questionnaires. *Educational and psychological measurement*, 84(5), 887-926. <https://doi.org/10.1177/00131644231222420>

- Lee, Y. H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education*, 2, 1-24. <https://doi.org/10.1186/s40536-014-0008-1>
- Meinck, S., & Brese, F. (2019). Trends in gender gaps: Using 20 years of evidence from TIMSS. *Large-scale Assessments in Education*, 7(1), 1-23. <https://doi.org/10.1186/s40536-019-0076-3>
- Minaei, A. (2013). Assessing Construct Comparability and Analyzing Differential Item Functioning (DIF) and Differential Test Functioning (DTF) in the Eighth-Grade TIMSS 2007 Science Test Between Iranian and American Students. *Educational Measurement Quarterly*, 4(11), 113-151. https://jem.atu.ac.ir/article_2682.html?lang=en
- Minaei, A., & Falsafi Nejad, M. R. (2010). Methods for Assessing the Unidimensionality of Items in Dichotomous IRT Models. *Educational Measurement Quarterly*, 3, 79-98. <https://www.sid.ir/paper/214519/en>
- Mullis, I. V., & Martin, M. O. (2017). TIMSS 2019 Assessment Frameworks. <https://timssandpirls.bc.edu/timss2019/frameworks/index.html>
- Mullis, I. V., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 international results in mathematics and science. <https://timssandpirls.bc.edu/timss2019/>
- Naeimi, S., Minaei, A., & Niyousha, B. (2023). Predicting Mathematical Self-Efficacy Through Self-Regulation in Students of Saveh City. *Razi Journal of Medical Sciences*, 29(12), 520-528. <https://www.magiran.com/paper/2758450/the-predicting-mathematic-self-efficacy-with-self-regulation-strategies-among-students-of-saveh-city?lang=en>
- Piroozi, A., & Kazemi, M. R. (2022). Evaluation of Eighth-Grade Students' Performance in Fars Province in the Cognitive and Content Domains of Mathematics Based on the Results of TIMSS 2019. *Mathematics and Society Journal*, 7(4), 43-77. https://math-sci.ui.ac.ir/article_27399.html?lang=en
- Pongsophon, P. (2023). Multilevel Analysis of Factors that Determine the Science Achievement of Fourth-Grade Students in TIMSS 2019. *Science Education International*, 34(2), 86-95. <https://doi.org/10.33828/sei.v34.i2.2>
- Sahibzada, U. F., Jianfeng, C., Latif, K. F., Shafait, Z., & Sahibzada, H. F. (2022). Interpreting the impact of knowledge management processes on organizational performance in Chinese higher education: mediating role of knowledge worker productivity. *Studies in Higher Education*, 47(4), 713-730. <https://doi.org/10.1080/03075079.2020.1793930>
- Tsupros, N., Kohler, R., & Hallinen, J. (2009). STEM education: A project to identify the missing components. Intermediate Unit,
- Tutz, G., & Berger, M. (2016). Response styles in rating scales: Simultaneous modeling of content-related effects and the tendency to middle or extreme categories. *Journal of Educational and Behavioral Statistics*, 41(3), 239-268. <https://doi.org/10.3102/1076998616636850>
- Ulitzsch, E., von Davier, M., & Pohl, S. (2020). Using response times for joint modeling of response and omission behavior. *Multivariate Behavioral Research*, 55(3), 425-453. <https://doi.org/10.1080/00273171.2019.1643699>
- Wang, L. H., Chen, B., Hwang, G. J., Guan, J. Q., & Wang, Y. Q. (2022). Effects of digital game-based STEM education on students' learning achievement: a meta-analysis. *International Journal of Stem Education*, 9(1), 1-13. <https://doi.org/10.1186/s40594-022-00344-0>
- Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of research in personality*, 47(2), 178-189. <https://doi.org/10.1016/j.jrp.2012.10.010>
- Wiberg, M., & Rolfman, E. (2019). The association between science achievement measures in schools and TIMSS science achievements in Sweden. *International Journal of Science Education*, 41(16), 2218-2232. <https://doi.org/10.1080/09500693.2019.1666217>
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17. https://doi.org/10.1207/s15326977ea1001_1
- Yaftian, N., & Abbasi, F. (2024). Alignment of Fourth-Grade Published Mathematics Problems from TIMSS 2019 with the Content and Problems of Iranian and Japanese Mathematics Textbooks. *Educational Innovations*, 23(2), 7-31. https://noavaryedu.oerp.ir/article_195789.html
- Zia Nejad Shirazi, A., Kouroshnia, M., Sohrabi, N., & Bogholi, H. (2022). Developing a Differential Equation of Individual Factors Affecting Performance in the TIMSS 2019 Fourth-Grade Test Among High- and Low-Performing Students. *Research in Educational Systems*, 16(58), 129-141. https://www.jiera.ir/article_167791.html
- Zia Nejad Shirazi, A., & Qaltash, A. (2016). The Role of Changes in the Educational System on the Results of International TIMSS Eighth-Grade Exams from the Beginning Until 2015. International Conference on New Horizons in Educational Sciences, Psychology, and Social Harms, Tehran, <https://en.civilica.com/doc/609977/>