



Utilizing Item Response Theory (IRT) Models to Analyze and Compare the Response Styles of Eighth-Grade Students in the TIMSS Science Assessments of 2011, 2015, and 2019

Fatemeh Azadi¹ , Asghar Minaee² , Masoud Geramipour³

1. Ph.D Candidate in psychology, Department of psychology, Saveh Branch, Islamic Azad University, Saveh, Iran. E-mail: azadifatemehk@gmail.com

2. Associate Professor, Department of Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabatabaie University, Tehran, Iran. E-mail: asghar.minaei@yahoo.com

3. Associate Professor, Department of Curriculum and Measurement Studies, Faculty of Education and Psychology, Kharazmi University, Tehran, Iran. E-mail: mgramipour@yahoo.com

ARTICLE INFO

Article type:

Research Article

Article history:

Received 12 January 2026

Received in revised form 11 March 2026

Accepted 17 April 2026

Published Online 23 July 2026

Keywords:

TIMSS science test,
response styles,
Response Theory (IRT)

ABSTRACT

Background: International standardized tests such as TIMSS serve as rich sources for examining student academic performance. However, accurate interpretation of the results from these assessments necessitates an understanding of students' response patterns and styles. Despite the importance of analyzing response patterns in international tests like TIMSS for gaining a deeper understanding of student performance, few studies have comprehensively examined and compared the response styles of eighth-grade students in the TIMSS science assessment using Item Response Theory (IRT) models across multiple time cycles (2011, 2015, 2019).

Aims: The purpose of this study was to identify the response styles of Iranian students in the TIMSS Grade 8 Science assessments conducted in 2011, 2015, and 2019.

Methods: This research is applied in terms of its purpose and correlational in terms of its methodology. Additionally, it falls within the category of secondary analysis studies. The statistical population of this study included all participants in the TIMSS Grade 8 Science assessments during the years 2011, 2015, and 2019. According to official statistics, the total number of participants was 18,124. The instrument used in this study was the TIMSS Science test for the years 2011, 2015, and 2019. To determine the response styles and their stability, IRTree modeling based on a hierarchical approach was employed using the R/Rstudio software packages. The statistical significance level for this study was set at 0.05 ($\alpha = 0.05$).

Results: In the TIMSS 2011 assessment, more than half of Iranian students exhibited a typical response style. In the TIMSS 2015 assessment, approximately one-third of Iranian students showed an extreme response style, while another one-third demonstrated a typical response style. In the TIMSS 2019 assessment, around one-third of Iranian students exhibited both typical and extreme response styles.

Conclusion: Given the observed differences in students' response styles, it is necessary to identify their weaknesses and implement changes in the educational, technological, or cultural systems affecting students.

Citation: Azadi, F., Minaee, A., & Geramipour, M. (2026). Utilizing Item Response Theory (IRT) Models to analyze and compare the response styles of eighth-grade students in the TIMSS science assessments of 2011, 2015, and 2019. *Journal of Psychological Science*, 25(161), 1-14. [10.66224/jps.25.161.15](https://doi.org/10.66224/jps.25.161.15)

Journal of Psychological Science, Vol. 25, No. 161, 2026

© The Author(s). DOI: [10.66224/jps.25.161.15](https://doi.org/10.66224/jps.25.161.15)



✉ **Corresponding Author:** Asghar Minaee, Associate Professor, Department of Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabatabaie University, Tehran, Iran.

E-mail: asghar.minaei@yahoo.com, Tel: (+98) 9122599745

Extended Abstract

Introduction

The process of education in the technological era necessitates new models and appropriate assessment tools. Large-scale international assessments like TIMSS provide valuable data on student performance in science and mathematics and serve as a benchmark for comparing educational systems (Martin et al., 2020; Smail et al., 2018). However, comparative studies have shown inconsistent results regarding the relationship between attitude and performance in tests across different countries (Juen & Deboeck, 2016). A primary reason for these discrepancies and the reduced validity of results is the presence of response styles, which refer to systematic tendencies in responding to questions regardless of their content (Khorramdel et al., 2019).

Response styles, as systematic response patterns independent of content, introduce numerous psychometric issues and reduce test score validity by generating systematic error (Cronbach, 1964). Common styles include Extreme Response Style (ERS), a tendency to select the endpoints of a scale, and Midpoint/Ordinary Response Style (MRS), a propensity for choosing the middle options (Luo & Bolt, 2015; Kaufmann et al., 2023). Research indicates that ERS impairs the results of correlation-based statistical analyses by increasing score variance and introducing bias (positive or negative) (Liu et al., 2017; Deboeck & Kyriakides, 2021), and also leads to biased estimation of mean scores and measurement error (Kim & Bolt, 2021; Hooton, 2017). Given that attitudes and response styles may vary across cultures (Eid et al., 2003), investigating and correcting the impact of these styles is essential to obtain true test scores. Previous research has shown that modeling without considering response style can lead to serious bias in estimations (Tutz et al., 2018), and advanced IRT models such as Mixture IRT and IRTree are effective tools for identifying and controlling these styles (Buckenholt & Meiser, 2017; Deebak, 2019). Despite the importance of this topic, a review of the research literature indicated that no study has yet addressed the longitudinal identification of response styles in Iranian eighth-grade TIMSS science test data

across the 2011, 2015, and 2019 cycles. Therefore, the present study was conducted with the aim of filling this gap by identifying and comparing student response styles in these three time periods using Item Response Theory (IRT) models.

Method

The present study was applied in purpose and correlational in method, conducting a secondary analysis of data from the TIMSS eighth-grade science assessment in Iran for the years 2011, 2015, and 2019. The statistical population comprised all Iranian participants in this assessment during the mentioned years, and a sample size of 18,124 individuals was analyzed, with their data sourced from the TIMSS database; inclusion criteria included participation in the TIMSS eighth-grade science assessment during the specified years and being of Iranian nationality, while the sole exclusion criterion was the absence of complete response data for a participant. The research instrument was the TIMSS science assessment (2011, 2015, 2019), a standardized international assessment for measuring academic achievement in science. To identify and compare student response styles, Item Response Theory (IRT) models, specifically the IRTree model based on a hierarchical approach, were employed, and analyses were conducted using R/Rstudio software packages at a statistical significance level of 0.05 ($\alpha = 0.05$).

Results

The findings of this study are based on the analysis of data from 18,124 Iranian eighth-grade students who participated in the TIMSS science assessment in 2011, 2015, and 2019. Students' response styles to attitude questions about science were identified and categorized into four groups: "Extreme Response Style" (selecting options 1 or 4), "Midpoint Response Style" (selecting options 2 or 3), "Acquiescence Response Style" (selecting options 1 or 2), and "Ordinary Response Style" (no fixed pattern). Table 1 presents the frequency distribution of these styles in the TIMSS 2011 assessment, indicating that the Ordinary Response Style was the most prevalent (54.6%), while the Midpoint Response Style had the lowest frequency (8.5%) among Iranian students in that year.

Table1. Frequency and Percentage of Response Styles in the TIMSS 2011 Assessment

Response Style	Frequency	Percentage
Extreme Response Style	1540	25/6
Midpoint Response Style	511	8/5
Acquiescence Response Style	680	11/3
Ordinary Response Style	3283	54/6
Total	6104	100

Table 1 shows that in the TIMSS 2011 assessment, over half of the students exhibited the Ordinary Response Style (54.6%). Among the other three response styles, the Extreme Response Style had the

highest frequency (25.6%), followed by the Acquiescence Response Style (11.3%) and the Midpoint Response Style (8.5%), respectively.

Table2. Frequency and Percentage of Response Styles in the TIMSS 2015 Assessment

Response Style	Frequency	Percentage
Extreme Response Style	1982	32/3
Midpoint Response Style	494	8/1
Acquiescence Response Style	1688	27/5
Ordinary Response Style	1966	32/1
Total	6130	100

Table 2 indicates that in the TIMSS 2015 assessment, approximately one-third of the students demonstrated the Extreme Response Style (32.3%), and one-third

exhibited the Ordinary Response Style (32.1%). Comparing the four response styles, the Midpoint Response Style had the lowest frequency (8.1%).

Table3. Frequency and Percentage of Response Styles in the TIMSS 2019 Assessment

Response Style	Frequency	Percentage
Extreme Response Style	2609	34/6
Midpoint Response Style	360	6
Acquiescence Response Style	1402	23/4
Ordinary Response Style	2149	35/9
Total	5980	100

Table 3 reveals that in the TIMSS 2019 assessment, approximately one-third of the students showed the Ordinary Response Style (35.9%), and one-third exhibited the Extreme Response Style (34.6%). Among the four response styles, the Midpoint Response Style had the lowest frequency (6%).

Conclusion

This study aimed to determine and compare the response styles of Iranian eighth-grade students in the TIMSS science assessment in 2011, 2015, and 2019. The key findings of this study for each assessment cycle are discussed and explained below. In the TIMSS 2011 assessment, the Ordinary Response Style was dominant among Iranian students with over half of the frequency (54.6%), while the Extreme Response Style ranked second (25.6%), and the

Acquiescence (11.3%) and Midpoint (8.5%) styles were less common. The prevalence of the Ordinary style was linked to IRT concepts such as Differential Item Functioning (Majed & Abuharaz, 2019) and test-taker behavior theories (Papanastasiou & Michaelides, 2024), and was attributed to the traditional Iranian educational system's emphasis on rote learning and insufficient development of critical thinking (Ilanlou, 1398/2019).

In the TIMSS 2015 assessment, the distribution of styles became more balanced, with Extreme (32.3%) and Ordinary (32.1%) styles having nearly equal shares, while the Midpoint style still had the lowest frequency (8.1%). The Extreme style was associated with IRT parameters and confidence (Zhang & Wang, 2020; Sirota & Janczic, 2018). These findings, alongside related studies on aberrant response profiles (Henninger & Meiser, 2020) and item

difficulty (Saatcioglu & Sen, 2023) and comparisons with countries like Japan (Mullis & Martin, 2017; Isozaki, 2018), were interpreted as reflecting the combined influence of educational and cultural factors on decisive or cautious responding in Iran.

In the TIMSS 2019 assessment, the Ordinary (35.9%) and Extreme (34.6%) styles remained dominant, but the distribution became slightly more balanced compared to previous years, with the Midpoint style having the lowest frequency at 6%. This relative shift, which can be linked to developmental/generational change theories (Damas et al., 2020), was attributed to the potential impact of developments in Iranian educational policies and practices. However, the continued prevalence of the Ordinary and Extreme styles indicates persistent structural challenges within the educational and cultural system.

Despite providing valuable insights into the response styles of Iranian students on the TIMSS science assessment, the present study has limitations, many of which were beyond the direct control of the researcher. The primary limitation is the nature of secondary data analysis, meaning the researcher had no control over the initial data collection process, the selection of the original sample by the organization responsible for TIMSS in Iran, or the background variables collected. This restricted the possibility of a more in-depth investigation into factors influencing response styles (such as teacher or school characteristics, or family factors). Furthermore, the cross-sectional nature of the TIMSS data in each cycle does not allow for causal analysis or examination of individual student changes over time, with analyses conducted solely based on aggregated data for each year. Nevertheless, the findings of this study carry significant practical implications. Understanding the prevalent response styles among Iranian students can assist test designers in better calibrating items and minimizing the impact of confounding factors on the measurement of true ability. Additionally, teachers and curriculum planners can leverage this understanding to modify teaching methods, guiding students towards more critical thinking, expressing well-reasoned opinions, and reducing response behaviors driven by uncertainty or guessing.

Ethical Considerations

Compliance with ethical guidelines: This research was conducted as a secondary analysis of pre-collected data from the international TIMSS assessment. As the researcher had no direct contact with the participants and the data were obtained in a completely anonymized form from the TIMSS assessment database and the national TIMSS and PIRLS study center, the ethical considerations regarding obtaining direct informed consent from participants during the initial data collection process were addressed by the assessment administrators. In this study, the principle of participant data confidentiality was fully maintained, and the data used were devoid of any personally identifiable information. Access to the data was restricted to the research team, and all analyses were conducted solely for the scientific and research purposes defined in this study.

Funding: This research is part of a doctoral dissertation and was not externally funded.

Authors' contribution: This article is extracted from the first author's doctoral dissertation, supervised by the second author, and advised by the third author.

Conflict of interest: The authors declare no conflict of interest concerning the results presented in this study.

Acknowledgments: The authors would like to express their gratitude to the participants who took part in this study.



کاربرد مدل‌های نظریه سؤال پاسخ (IRT) در شناسایی و مقایسه سبک‌های پاسخ دانش‌آموزان پایه هشتم علوم

تیمز ۲۰۱۱، ۲۰۱۵، ۲۰۱۹

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. رایانامه: azadifatemeh@gmail.com

۲. دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: asghar.minaei@yahoo.com

۳. دانشیار، گروه مطالعات برنامه درسی و سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران. رایانامه: mgramipour@yahoo.com

چکیده

مشخصات مقاله

نوع مقاله:

پژوهشی

تاریخچه مقاله:

دریافت: ۱۴۰۴/۱۰/۲۲

بازنگری: ۱۴۰۴/۱۲/۲۰

پذیرش: ۱۴۰۵/۰۱/۲۸

انتشار برخط: ۱۴۰۵/۰۵/۰۱

کلیدواژه‌ها:

آزمون علوم تیمز،

سبک‌های پاسخ،

نظریه سؤال پاسخ.

زمینه: آزمون‌های استاندارد بین‌المللی نظیر تیمز منبع غنی برای بررسی عملکرد تحصیلی دانش‌آموزان هستند. با این حال، تفسیر دقیق نتایج این آزمون‌ها نیازمند درک الگوها و سبک‌های پاسخ‌دهی دانش‌آموزان است. با وجود اهمیت تحلیل الگوهای پاسخ‌دهی در آزمون‌های بین‌المللی نظیر تیمز برای درک عمیق‌تر عملکرد دانش‌آموزان، مطالعات اندکی به بررسی جامع و مقایسه‌ی سبک‌های پاسخ دانش‌آموزان پایه هشتم در آزمون علوم تیمز با استفاده از مدل‌های نظریه سؤال-پاسخ (IRT) در چرخه‌های زمانی متعدد (۲۰۱۱، ۲۰۱۵، ۲۰۱۹) پرداخته‌اند.

هدف: هدف از این پژوهش تعیین سبک‌های پاسخ دانش‌آموزان ایرانی در آزمون علوم تیمز پایه هشتم ۲۰۱۱، ۲۰۱۵، ۲۰۱۹ بود.

روش: این پژوهش از نظر هدف کاربردی و از لحاظ روش از نوع همبستگی بود از طرفی پژوهش حاضر در دسته تحقیقات تحلیل ثانویه قرار دارد. جامعه آماری پژوهش شامل تمامی شرکت‌کنندگان در آزمون علوم پایه هشتم تیمز بود که طی سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ در این آزمون شرکت کرده بودند. بر اساس آمار رسمی تعداد کل شرکت‌کننده‌ها (۱۸۱۲۴) نفر بوده‌اند. ابزار مورد استفاده در این پژوهش آزمون علوم تیمز در سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ بود. به منظور تعیین سبک‌های پاسخ و ثبات آن‌ها، از مدل‌سازی IRTree مبتنی بر رویکرد مدل سلسله‌مراتبی با استفاده از پکیج‌های نرم‌افزار R/Rstudio استفاده شد. سطح معناداری آماری در این پژوهش $(\alpha = 0.05)$ بود.

یافته‌ها: نتایج نشان داد در آزمون تیمز ۲۰۱۱ بیش از نیمی از دانش‌آموزان ایرانی سبک پاسخ معمولی را داشته‌اند. در آزمون تیمز ۲۰۱۵ تقریباً یک‌سوم از دانش‌آموزان ایرانی سبک پاسخ افراطی و یک‌سوم نیز سبک پاسخ معمولی داشته‌اند. در آزمون تیمز ۲۰۱۹، تقریباً یک‌سوم از دانش‌آموزان ایرانی سبک پاسخ معمولی و سبک پاسخ افراطی داشته‌اند.

نتیجه‌گیری: با توجه به تفاوت‌های مشاهده‌شده در سبک‌های پاسخ‌دهی دانش‌آموزان بعد از شناسایی نقاط ضعف دانش‌آموزان می‌بایست تغییراتی در نظام آموزشی، فناوری یا فرهنگی دانش‌آموزان ایجاد کرد.

استاد: آزادی، فاطمه؛ مینایی، اصغر؛ و گرامی پور، مسعود (۱۴۰۵). کاربرد مدل‌های نظریه سؤال پاسخ (IRT) در شناسایی و مقایسه سبک‌های پاسخ دانش‌آموزان پایه هشتم آزمون

علوم تیمز ۲۰۱۱، ۲۰۱۵، ۲۰۱۹. مجله علوم روانشناختی، دوره ۲۵، شماره ۱، ۱۶۱-۱۴.

DOI: 10.66224/jps.25.161.15. ۱۴۰۵، شماره ۱، ۱۶۱. مجله علوم روانشناختی، دوره ۲۵، شماره ۱، ۱۶۱-۱۴.



نویسندگان.

✉ نویسنده مسئول: اصغر مینایی، دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه:

asghar.minaei@yahoo.com

تلفن: ۰۹۱۲۲۵۹۹۷۴۵

مقدمه

فرآیند آموزش در سراسر دنیا مسئله‌ای مهم و قابل تأمل است، زیرا تکنولوژی با سرعت بیشتری نسبت به تغییرات در سیستم آموزشی حرکت می‌کند؛ بنابراین مدل‌های آموزشی جدیدی مطابق با الگوهای در حال تغییر در صنعت، تکنولوژی و آموزش مورد نیاز است (الکسانکو، ۲۰۱۷). قرن بیست و یکم دوره منحصربه‌فردی از پیشرفت‌های تکنولوژی و گسترش فناوری در جهان بوده است که از رویدادهای ده‌های گذشته فراتر رفته است. 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 (سوپروس، کوهرلر و هالینن، ۲۰۰۹). یکی از مهمترین مسائل دنیای امروز، پرورش و توسعه درک علم برای دانش‌آموزان است. درواقع تلاش بر آن است که دانش‌آموزان امروز به شهروندانی در آینده تبدیل شوند که بتوانند آگاهانه برای خود و دنیایی که در آن زندگی می‌کنند، تصمیم بگیرند (چن، ۲۰۲۲). در کنار پیشرفت برنامه و الگوی درسی، مربیان و برنامه‌ریزان تمایل دارند که نسل جدید عملکردی خوب در جامعه آتی داشته باشد و مهارت‌های قرن بیست و یکم مجهز به خلاقیت، نوآوری و کارآفرینی شود. ازاین‌رو پژوهشگران برآنند تا ابزارهای مناسب جهت سنجش و ارزیابی میزان پیشرفت علوم، تدوین و اجرا نمایند. یکی از این آزمون‌ها مجموعه آزمون‌های تیمز^۱ است که داده‌های قابل اعتماد و ارزشمندی در مورد عملکرد دانش‌آموزان در ریاضیات و علوم در کشورهای مختلف ارائه می‌دهد. تیمز مجموعه‌ای از پروژه‌های ارزیابی و تحقیقاتی بین‌المللی است که برای اندازه‌گیری سطح پیشرفت تحصیلی دانش‌آموزان پایه چهارم و هشتم در سطح بین‌المللی طراحی شده و نیز به منظور هماهنگی برنامه‌های درسی و سیستم آموزشی ریاضیات و علوم به‌طور گسترده در کشورهای شرکت‌کننده ایجاد شده است (مارتین، وون داویر و مولیس، ۲۰۲۰).

از دستاوردهای دیگر آزمون تیمز این است که میزان دانش، دانش‌آموزان را در ریاضیات، علوم و مهارت‌ها در زمینه‌های مختلف زندگی نشان می‌دهد و معیاری است که کیفیت سیستم آموزشی کشورهای مختلف در سطح بین‌المللی را نشان می‌دهد (اسمیل و همکاران، ۲۰۱۸). نقش و اهمیت آزمون‌های تیمز در تشخیص و ارزیابی پیشرفت تحصیلی در علوم محققان را بر آن داشته که عوامل اثرگذار بر نتایج آزمون‌ها را بررسی نمایند، مطالعات تطبیقی بین‌المللی نتایج متناقضی را در خصوص رابطه بین نگرش و عملکرد افراد در آزمون نشان داده است. نتایج پژوهش‌ها نشان داد 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، درحالی که ممکن است در فرهنگ دیگر رابطه‌ی منفی بین نگرش و موفقیت وجود داشته باشد (جوئن و دیوئیک، ۲۰۱۶).

ازاین‌رو نتایج متفاوت^۲ در سطوح فردی و گروهی اعتبار نتایج به‌دست آمده از مطالعات را کاهش می‌دهد. یکی از دلایل این تفاوت‌ها، سبک‌های پاسخ^۳ است که گرایش سیستماتیک به پاسخ‌دهی به سؤال‌های یک پرسشنامه بدون توجه به محتوای آن‌ها است (خرمدل، ون داویر و پوکروپیک، ۲۰۱۹). سبک پاسخ افراد مشکلات روان‌سنجی مختلفی در داده‌ها ایجاد کرده است. به بیان دیگر، از طریق تولید یک خطای سیستماتیک^۴ در نمرات افرادی که از سطح یکسانی از دانش^۵، نگرش^۶ یا خصوصیات شخصیتی^۷ برخوردارند، باعث کاهش اعتبار^۸ نمره‌های آزمون می‌شود (کرنباخ، ۱۹۶۴).

نتایج پژوهش‌های پیشین نشان می‌دهد که سبک‌های پاسخ متفاوتی وجود دارند که هرکدام به‌نوبه‌ی خود تأثیر منفی بر اعتبار نمرات آزمون می‌گذارند. زمانی که سبک پاسخ به سبک پاسخ نهایی^۹ (ERS) محدود شود، پاسخ را از مرکز (نقطه میانی) به بیرون می‌راند و در نتیجه برآورد واریانس را افزایش می‌دهد. علاوه بر این، وقتی یکی از نقاط انتهایی (طبقه‌بندی‌های پاسخ نهایی) بیشتر انتخاب شود، ممکن است سوگیری رخ

⁶ attitude⁷ personality characteristics⁸ Validity⁹ extreme response style¹ TIMSS Assessments² interchangeable³ Response styles⁴ systematic error⁵ knowledge

دهد. به عبارت دقیق‌تر، وقتی افراد، بیشتر گزینه‌ی انتهایی مثبت را انتخاب کرده باشند، یک سوگیری مثبت ممکن است رخ دهد. از سوی دیگر، اگر افراد، بیشتر مستعد انتخاب گزینه‌های انتهایی منفی، در مقایسه با گزینه‌های انتهایی مثبت باشند، سوگیری در جهت منفی خواهد بود (لیو و همکاران، ۲۰۱۷). از آنجاکه همبستگی و واریانس نمرات تا حدودی مرتبط با یکدیگر هستند، نتایج به‌دست آمده از همبستگی بین متغیرها از سبک‌های پاسخ تأثیر می‌پذیرند. مشخصاً، از آنجاکه سبک پاسخ نهایی موجب افزایش واریانس می‌شود، همبستگی بین متغیرهای موردنظر با افزایش انتخاب گزینه‌های انتهایی افراد کاهش می‌یابد. افزون بر این، به خاطر این واقعیت که چندین تکنیک آماری همچون تحلیل رگرسیون، تحلیل همبستگی بنیادی، تحلیل عاملی مبتنی بر همبستگی است، سبک پاسخ نهایی بر نتایج به‌دست آمده از آن‌ها تأثیر خواهد گذاشت (دیک و کیکریکی، ۲۰۲۱).

علاوه بر سبک پاسخ نهایی، سبک پاسخ تصدیق یا مخالفت^۱ و سبک پاسخ نمره میانی^۲ و غیره نیز بر نتایج آزمون‌ها تأثیر می‌گذارند (کافمن، بنی و استپ، ۲۰۲۳). سبک پاسخ تصدیق عبارت است از گرایش فرد به دادن پاسخ‌های مثبت یا موافق با سؤال‌ها بدون در نظر گرفتن محتوای آن‌ها است (وانگ و لویس، ۲۰۲۲) که به آن «گرایش به موافقت» نیز گفته می‌شود. از یک‌سو، همان‌طور که گفته شد، سبک پاسخ نهایی گرایش به انتخاب گزینه‌های انتهایی مقوله‌های پاسخ مانند «کاملاً موافق/کاملاً مخالف» است (دیمک، ۲۰۱۹). به‌طور 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 نهایی بر سطح همبستگی‌ها بین معیارها و در نتیجه نتایج به‌دست آمده از تحلیل‌های عاملی، تأثیر می‌گذارد. از سوی دیگر، سبک پاسخ نهایی معمولاً باعث می‌شود که میانگین نمرات سؤال‌ها کمتر یا بیشتر از نمرات واقعی‌اش تخمین زده شود (کیم و بولت، ۲۰۲۱) که منتج به ایجاد سوگیری در نتایج می‌شود. این نمره‌های دارای سوگیری ممکن است منتج به نتایج خطای اندازه‌گیری شوند (هوتن، ۲۰۱۷).

^۱ acquiescence (ARS) or disacquiescence (DRS)

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 (اید، لانگهین و دینر، ۲۰۰۳). از این جهت، برای نشان دادن نمرات واقعی آزمون‌ها، تأثیر سبک‌های پاسخ باید موردبررسی 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. دیک و کیکریکی (۲۰۲۱)، عنوان کردند که سبک‌های پاسخ مختلفی وجود دارند که تأثیر منفی بر نمرات آزمون می‌گذارند. آن‌ها به اثرات سبک پاسخ نهایی پرداختند که در آن گرایش پاسخ دادن به یکی از دو انتهای گزینه‌ها است (لو و بولت، ۲۰۱۵). توتر و همکاران (۲۰۱۸) نشان دادند که نادیده گرفتن سبک پاسخ در هنگام مدل‌سازی می‌تواند منجر به تعصب جدی در تخمین‌ها شود. مطالعات نشان داده‌اند که مدل‌ها اغلب به سمت طبقه‌بندی داده‌ها در دسته‌های افراطی (با ویژگی‌های بسیار شدید) یا متوسط (با ویژگی‌های معمولی) تمایل دارند. برای مثال، در یک سیستم طبقه‌بندی احساسات، ممکن است مدل همه نظرات را به‌عنوان بسیار مثبت یا بسیار منفی طبقه‌بندی کند و از در نظر گرفتن نظرات خنثی یا پیچیده غافل شود. همچنین دیمک (۲۰۱۹) در بررسی سبک پاسخ افراطی دانش آموزان با استفاده از IRTree در آزمون تیمز ۲۰۱۵، عنوان کردند که سبک پاسخ افراطی دانش آموزان بر نگرش آن‌ها بر ریاضیات اثرگذار بوده است؛ این نتیجه با نتایج پژوهش بوکنهلت و میسر (۲۰۱۷) نیز همسو است که در آن مدل‌های IRT مختلف (مدل‌های چندبعدی راش و مدل‌های درخت پاسخ) برای کنترل سبک‌های پاسخ در نمره دهی مقیاس‌ها استفاده می‌شدند. آن‌ها خاطر نشان کردند که سبک‌های پاسخ بر پاسخ دانش آموزان با توجه به ویژگی‌های شخصی تأثیر می‌گذارند و مدل‌های استفاده‌شده در مطالعه آن‌ها از نظر ارائه سبک‌های پاسخ به‌عنوان منابع چندبعدی واریانس‌های فردی متفاوت بود. علاوه بر این، دانش آموزانی که نگرش آن‌ها مثبت بود به انتخاب نمرات میانی گرایش داشتند. این نتیجه را می‌توان به ابعاد فرهنگی کشورهای منتخب ارتباط داد؛ به عبارت دیگر،

^۲ middle response styles (MRS)

۲۰۱۵ و ۲۰۱۹ شناسایی می‌کنند و این سبک‌ها در طول این چرخه‌های زمانی (۲۰۱۱، ۲۰۱۵، ۲۰۱۹) چه تفاوت‌ها و شباهت‌هایی با یکدیگر دارند؟

روش

الف) طرح پژوهش و شرکت‌کنندگان

این پژوهش از نظر هدف کاربردی و ازلحاظ روش از نوع همبستگی بود که در دسته تحقیقات تحلیل ثانویه قرار دارد. جامعه آماری پژوهش حاضر شامل تمامی شرکت‌کنندگان در آزمون علوم پایه هشتم تیمز بود که طی سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ در این آزمون شرکت کرده بودند. نمونه آماری این پژوهش، دانش‌آموزان پایه هشتم ایرانی شرکت‌کننده در آزمون علوم تیمز در سال‌های مذکور را در بر می‌گیرد. حجم نمونه مورد استفاده در این پژوهش برابر با ۱۸۱۲۴ نفر است که داده‌های آن‌ها از بانک داده‌های آزمون تیمز تهیه شده است. با توجه به ماهیت داده‌های ثانویه و عدم دخالت پژوهشگر در فرآیند جمع‌آوری اصلی، روش تهیه داده‌ها معادل نمونه‌گیری تصادفی ساده تلقی می‌شود. ملاک‌های ورود به این پژوهش شامل شرکت دانش‌آموزان پایه هشتم در آزمون علوم تیمز، ایرانی بودن شرکت‌کننده و حضور در یکی از چرخه‌های آزمون در سال‌های ۲۰۱۱، ۲۰۱۵ یا ۲۰۱۹ بود. در مقابل، تنها ملاک خروج از پژوهش، عدم وجود داده‌های پاسخ کامل برای شرکت‌کننده در آزمون در نظر گرفته شد.

ب) ابزار

ابزار مورد استفاده در این پژوهش آزمون علوم تیمز در سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ بود. تیمز یکی از مطالعات ایجادشده توسط IEA (انجمن بین‌المللی ارزیابی دستاورد آموزشی) است که هدف آن اجازه دادن به سیستم‌های آموزشی در سراسر جهان برای مقایسه دستاوردهای تحصیلی دانش‌آموزان با سیستم‌های آموزشی دیگر و یادگیری از تجربیات دیگران در طراحی سیاست آموزشی مؤثر است. این ارزیابی اولین بار در سال ۱۹۹۵ انجام شد و از آن پس هر چهار سال یک‌بار دایر می‌شود؛ بنابراین، برخی از سیستم‌های آموزشی شرکت‌کننده داده‌های روند را از سال ۱۹۹۵ تا ۲۰۱۹ به دست آورده‌اند. آزمون تیمز دانش‌آموزان کلاس چهارم و هشتم را ارزیابی می‌کند، درحالی‌که تیمز پیشرفته دانش‌آموزان را در سال آخر متوسطه در ریاضیات پیشرفته و فیزیک مورد ارزیابی قرار می‌دهد.

ساختار جوامع آن‌ها می‌تواند پاسخ‌هایشان را به آیتم‌های لیکرت شکل دهد. هوفستد (۲۰۰۱) تأکید می‌کند که در جوامع جمع‌گرا، افراد عموماً به‌عنوان اعضای گروه یا سازمان عمل می‌کنند. در چنین فرهنگ‌هایی، ارتباط متقابل بین افراد نقشی مهم در زندگی‌شان بازی می‌کند و وفاداری در این جوامع اهمیت دارد. افراد در جوامع جمع‌گرا به احتمال بیشتری، میل دارند برای حفظ هارمونی در جامعه، پاسخ‌های میانی را انتخاب کنند، دیمک (۲۰۱۹) به این نتیجه رسیده است که تفاوت در شیوه پاسخ دادن شرکت‌کنندگان به سؤالات آزمون، باعث تفاوت‌های مشاهده‌شده در پارامترهای افراد در کشورهای مختلف شده است. مطالعه حاضر نشان داد که وقتی تأثیر سبک‌های پاسخ در نظر گرفته نشده بود، تغییرپذیری پارامتر فردی در رابطه با ارزیابی در زمینه ریاضیات برای هر کشور افزایش می‌یافت. این یافته با مطالعه انجام‌شده توسط توتز و همکاران (۲۰۱۸) همخوانی دارد که تأثیر سبک‌های پاسخ بر پاسخ‌های افراد درباره آیتم‌ها در رابطه با سختی^۱ را موردبررسی قرار دادند. آن‌ها دریافتند که برآورد واریانس پارامترهای فردی وقتی آن‌ها تأثیر سبک پاسخ را در نظر گرفته‌اند کاهش می‌یابد. در این حالت، می‌توان بیان کرد که دلیل کاهش واریانس در میان افراد حذف تفاوت در سبک پاسخ است.

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، موردتوجه قرار گیرد. از آنجایی که در پژوهش‌های بررسی‌شده تاکنون پژوهشی که به شناسایی سبک‌های پاسخ در تیمز ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ پرداخته باشد، یافت نشد و مشخص نیست که کدام‌یک از سبک‌های پاسخ تأثیر بیشتری بر سؤالات آزمون دارد. از این رو به دلیل خلأ پژوهشی موجود در زمینه بررسی طولی سبک‌های پاسخ در آزمون‌های تیمز و ویژگی‌های روان‌سنجی آزمون، در این پژوهش به شناسایی سبک‌های پاسخ دانش‌آموزان پایه هشتم در تیمز ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ پرداخته شده است؛ بنابراین پژوهش حاضر با هدف پاسخگویی به این سوال انجام شد که مدل‌های نظریه سؤال-پاسخ (IRT) چه سبک‌های پاسخی را در پاسخ‌های دانش‌آموزان پایه هشتم به آزمون علوم تیمز در سال‌های ۲۰۱۱،

^۱ tenseness

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 گروهی 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 به صورت طیف لیکرت (کاملاً موافقم=۱ تا کاملاً مخالفم=۴) نمره‌گذاری گردید و فقط سؤالات اعتماد به نفس به صورت معکوس نمره‌گذاری شدند.

در پژوهش‌های پیشین به منظور ویژگی‌های روانسنجی آزمون‌های تیمز از جمله پارامترهای سؤال، توانایی افراد از روش‌های مختلف مانند نظریه سؤال پاسخ به منظور برآورد پارامترهای سؤال و 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 و به وسیله نرم‌افزارهای SPSS، WinBugs، BILOG-MG3، NOHARM و R استفاده شده است، همچنین شریفی یگانه (۱۳۹۱) به منظور کارکرد افتراقی آزمون تیمز از دو روش مانل-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؛ بنابراین به منظور تعیین سبک‌های پاسخ از مدل نظریه سؤال پاسخ برای تعیین ثبات سبک‌های پاسخ استفاده شد به عبارتی از مدل‌سازی IRTree بر اساس رویکرد مدل سلسله‌مراتبی که با استفاده از پکیج‌های نرم‌افزار Rstudio و R انجام شده است. در این پژوهش، سطح معناداری آماری ($\alpha = 0.05$) در نظر گرفته شد.

یافته‌ها

یافته‌های این پژوهش بر اساس تحلیل داده‌های پاسخ به آزمون علوم تیمز در نمونه‌ای متشکل از ۱۸۱۲۴ نفر از دانش‌آموزان پایه هشتم ایرانی شرکت‌کننده در این آزمون طی چرخه‌های سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ استوار است. این نمونه، کلیه شرکت‌کنندگان ایرانی در آزمون علوم پایه هشتم تیمز در چرخه‌های زمانی مورد بررسی را شامل می‌شود و تحلیل‌ها

ارزیابی‌های جامع تیمز، شامل داده‌های مربوط به عملکرد دانش‌آموز در حوزه‌های ریاضی و علوم مختلف (جبر، هندسه، زیست‌شناسی، شیمی و غیره) و عملکرد در حل چالش‌های حل مسئله در هر یک از این زمینه‌ها است. علاوه بر این، تیمز داده‌های متنی را در برنامه تحصیلی، آموزشی و عوامل مرتبط با منابع فراهم می‌کند که می‌تواند بر فرآیند آموزش و یادگیری تأثیر بگذارند. این داده‌ها با استفاده از دانش‌آموزان، معلم، مدرسه و برنامه درسی (ملی) جمع‌آوری شده‌اند که به ترتیب توسط دانش‌آموزان، معلمان، مدیران مدارس و هماهنگ‌کننده‌های تحقیقات ملی تکمیل شده است. 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 ۲۰۱۱، تعداد سؤالات آزمون مدرسه ۱۷ سؤال، برنامه‌ریزی درسی ۱۲ سؤال، دفترچه دانش‌آموزان ۳۳ و دفترچه معلم ۲۰ سؤال بوده است، در سال ۲۰۱۵، تعداد سؤالات آزمون مدرسه ۲۲ سؤال، برنامه‌ریزی درسی عمومی ۱۵ و قسمت علوم ۸ سؤال، دفترچه دانش‌آموزان علوم عمومی ۲۶، دفترچه دانش‌آموزان علوم اختصاصی ۳۹ و دفترچه معلم ۲۵ سؤال بوده است و 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 از مدل سلسله‌مراتبی با اثرهای ثابت و تصادفی استفاده شده است، مدل‌های سلسله‌مراتبی^۱ که به آن‌ها مدل‌های چند سطحی^۲ نیز گفته می‌شود، ابزارهای آماری قدرتمندی هستند که برای تحلیل داده‌هایی با ساختار آشیانه‌ای^۳ 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 به عنوان اثر تصادفی و سؤالات به عنوان اثر ثابت وارد تحلیل شدند و در مدل دوم علاوه بر آیت‌ها، گره هم به عنوان اثر ثابت گرفته شد، گره‌ها واحدهای تحلیلی هستند که در سطوح مختلف

^۱ Hierarchical Models

^۲ Multilevel Models

^۳ Nested Data

بر روی این مجموعه داده‌های ثانویه انجام پذیرفته است. به منظور شناسایی سبک‌های پاسخ در آزمون علوم تمیز پایه هشتم ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ سبک پاسخ هر دانش آموز برحسب پاسخی که به سؤالات نگارش نسبت به درس علوم داده بود، تعیین گردید. افراد برحسب پاسخی که به سؤالات نگارش به علوم داده بودند به ۴ گروه «سبک پاسخ افراطی، سبک پاسخ میانی، سبک پاسخ رضایت و سبک پاسخ معمولی» تقسیم شدند. سبک پاسخ افراطی یعنی کسانی که گزینه‌های ابتدای «۱» و «۴» را انتخاب

کرده‌اند. سبک پاسخ میانی یعنی کسانی که گزینه‌های وسط «۲» و «۳» را انتخاب کرده‌اند. سبک پاسخ افراطی یعنی کسانی که گزینه‌های کاملاً موافقم «۱» و موافقم «۲» را انتخاب کرده‌اند. سبک پاسخ معمولی یعنی کسانی که در پاسخ به سؤالات الگوی خاصی نداشته و بر اساس شرایطی که دارند پاسخ مناسب را انتخاب کرده‌اند. در جدول ۱ شاخص‌های پرازش مدل برای هر متغیر مقایسه شده است.

جدول ۱. فراوانی و درصد فراوانی سبک‌های پاسخ در آزمون تیمز ۲۰۱۱

متغیر	فراوانی	درصد
سبک پاسخ افراطی	۱۵۴۰	۲۵/۶
سبک پاسخ میانی	۵۱۱	۸/۵
سبک پاسخ رضایت	۶۸۰	۱۱/۳
سبک پاسخ معمولی	۳۲۸۳	۵۴/۶
مجموع	۶۱۰۴	۱۰۰

در جدول (۱) مشخص شد در آزمون تیمز ۲۰۱۱ بیش از نیمی از دانش‌آموزان سبک پاسخ معمولی (۵۴/۶ درصد) داشته‌اند. در بین ۳ سبک پاسخ دیگر، بیشترین فراوانی مربوط به سبک پاسخ افراطی (۲۵/۶ درصد)

است و سبک پاسخ رضایت (۱۱/۳ درصد) و سبک پاسخ میانی (۸/۵ درصد) به ترتیب در رده‌های بعدی قرار دارند.

جدول ۲. فراوانی و درصد فراوانی سبک‌های پاسخ در آزمون تیمز ۲۰۱۵

متغیر	فراوانی	درصد
سبک پاسخ افراطی	۱۹۸۲	۳۲/۳
سبک پاسخ میانی	۴۹۴	۸/۱
سبک پاسخ رضایت	۱۶۸۸	۲۷/۵
سبک پاسخ معمولی	۱۹۶۶	۳۲/۱
مجموع	۶۱۳۰	۱۰۰

در جدول (۲) مشخص شد در آزمون تیمز ۲۰۱۵ تقریباً یک‌سوم از دانش‌آموزان سبک پاسخ افراطی (۳۲/۳ درصد) و یک‌سوم سبک پاسخ

معمولی (۳۲/۱ درصد) داشته‌اند. در مقایسه بین ۴ سبک پاسخ کمترین فراوانی مربوط به سبک پاسخ میانی (۸/۱ درصد) است.

جدول ۳. فراوانی و درصد فراوانی سبک‌های پاسخ در آزمون تیمز ۲۰۱۹

متغیر	فراوانی	درصد
سبک پاسخ افراطی	۲۰۶۹	۳۴/۶
سبک پاسخ میانی	۳۶۰	۶
سبک پاسخ رضایت	۱۴۰۲	۲۳/۴
سبک پاسخ معمولی	۲۱۴۹	۳۵/۹
مجموع	۵۹۸۰	۱۰۰

در جدول (۳) مشخص شد در آزمون تیمز ۲۰۱۹ تقریباً یک‌سوم از دانش‌آموزان سبک پاسخ معمولی (۳۵/۹ درصد) و یک‌سوم سبک پاسخ افراطی (۳۴/۶ درصد) داشته‌اند. در مقایسه بین ۴ سبک پاسخ کمترین فراوانی مربوط به سبک پاسخ میانی (۶ درصد) است.

بحث و نتیجه‌گیری

این پژوهش با هدف تعیین و مقایسه سبک‌های پاسخ دانش‌آموزان ایرانی در آزمون علوم تیمز پایه هشتم در سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ انجام شد. یافته‌های کلیدی این مطالعه در هر چرخه زمانی به شرح زیر مورد بحث و تبیین قرار می‌گیرد. بررسی داده‌های آزمون تیمز ۲۰۱۱ نشان داد که بیش از نیمی از دانش‌آموزان ایرانی سبک پاسخ معمولی را اتخاذ کرده بودند. این سبک حاکی از رویکردی متوازن و بدون تمایل افراطی به گزینه‌های پاسخ است و می‌تواند نشان‌دهنده گرایش به انتخاب گزینه‌های میانی یا عدم قطعیت قوی در پاسخ‌دهی باشد. از منظر نظریه سؤال-پاسخ (IRT)، سبک پاسخ معمولی می‌تواند با مفهوم عملکرد متفاوت آیتم یا عملکرد متفاوت گزینه مرتبط باشد، جایی که عوامل خارج از توانایی (مانند استراتژی پاسخ‌دهی یا عدم قطعیت) بر انتخاب پاسخ تأثیر می‌گذارند (ماجد و ابوحراز، ۲۰۱۹). همچنین، این سبک می‌تواند بازتاب‌دهنده تمایل به ریسک‌پذیری پایین در موقعیت‌های آزمون باشد که در نظریه‌های رفتار آزمون‌دهنده مورد بحث قرار می‌گیرد (پاپاناستاسیو و مایکلیدس، ۲۰۲۴). تبیین این یافته در زمینه آموزشی ایران می‌تواند به تأکید سنتی نظام آموزشی بر یادگیری رسمی و حفظی بازگردد. در چنین سیستمی، دانش‌آموزان ممکن است در مواجهه با سؤالاتی که نیازمند ابراز نظر یا قضاوت قاطعانه است، رویکرد محتاطانه‌ای اتخاذ کرده و گزینه‌های میانی را ترجیح دهند. این موضوع می‌تواند نشانگر عدم پرورش کافی تفکر انتقادی و مهارت‌های تصمیم‌گیری قاطع در مواجهه با مسائل علمی باشد که توسط برخی مطالعات در زمینه آموزش علوم در ایران نیز مطرح شده است (ایلانلو، ۱۳۹۸).

در آزمون تیمز ۲۰۱۵، توزیع سبک‌های پاسخ در میان دانش‌آموزان ایرانی نشان داد که تقریباً یک‌سوم سبک پاسخ افراطی و یک‌سوم نیز سبک پاسخ معمولی داشته‌اند که بیانگر پراکندگی قابل توجه پاسخ‌ها بین این دو قطب است. سبک پاسخ افراطی معمولاً نشان‌دهنده پاسخ‌دهی قاطعانه و تمایل

مشخص به گزینه‌های موافق یا مخالف است. سبک پاسخ افراطی می‌تواند با پارامترهای شیب و دشواری در مدل‌های IRT مرتبط باشد؛ به این معنا که دانش‌آموزان با توانایی‌های متفاوت، تمایل مشخصی به انتخاب گزینه‌های دور از میانه دارند (ژانگ و وانگ، ۲۰۲۰). همچنین، این سبک ممکن است به اعتمادبه‌نفس (بالا یا پایین) در پاسخ‌دهی مرتبط باشد که در چارچوب نظریه‌های شناختی پاسخ‌دهی به آزمون قابل بررسی است (سیروتا و جانچیچ، ۲۰۱۸). این یافته می‌تواند بازتاب‌دهنده تأثیر رویکردهای آموزشی باشد که بر تحمیل نظر یا پاسخ‌های درست مشخص به جای تشویق به تفکر مستقل تأکید دارند. مطالعه‌ای با استفاده از پرسشنامه دانش‌آموزان پایه هشتم تیمز ۲۰۱۵ نشان داد که دانش‌آموزان پروفایل‌های پاسخ‌دهی تصادفی متفاوتی از خود نشان می‌دهند، از جمله پاسخ‌دهندگان غیر تصادفی ثابت، پاسخ‌دهندگان تصادفی متناوب متوسط، پاسخ‌دهندگان تصادفی مکرر و دانش‌آموزانی که فقط در مقیاس‌های خاصی به صورت تصادفی پاسخ داده‌اند. این مطالعه از رویکرد ترکیب پنهان برای ارزیابی وضعیت پاسخ‌دهنده تصادفی دانش‌آموزان استفاده کرد و دریافت که این پروفایل‌ها در کشورهای مختلف قابل تعمیم هستند (هینگر و میسر، ۲۰۲۰). مطالعه‌ای که داده‌های تیمز ۲۰۱۵ از سنگاپور، ترکیه و آفریقای جنوبی را تحلیل کرد، دریافت که پارامترهای دشواری آیتم به طور معناداری در کلاس‌های پنهان در هر کشور متفاوت است. در سنگاپور، آیتم‌ها برای هر دو کلاس آسان بودند، در حالی که در ترکیه و آفریقای جنوبی، آیتم‌ها برای یک کلاس آسان و برای کلاس دیگر دشوار بودند (ساتچی اوغلو و سن، ۲۰۲۳). در مقایسه با کشورهای نظیر ژاپن که در گزارش‌های (مولیز و مارتین، ۲۰۱۷) تمایل بیشتری به سبک‌های پاسخ مثبت یا فعال نشان می‌دهند، این تفاوت را برجسته می‌سازد. در ژاپن، تمرکز بر حل مسئله و یادگیری تجربی ممکن است دانش‌آموزان را به اتخاذ رویکردهای پاسخ‌دهی فعال‌تر و انتقادی‌تر سوق دهد (ایسوزاکی، ۲۰۱۸)، در حالی که در ایران، ترکیبی از عوامل فرهنگی و آموزشی می‌تواند به شیوع سبک‌های افراطی (ناشی از قاطعیت یا عدم انعطاف‌پذیری) و معمولی (ناشی از عدم قطعیت یا احتیاط) منجر شود.

در آزمون تیمز ۲۰۱۹ نیز، سبک‌های پاسخ معمولی و افراطی همچنان بالاترین درصد فراوانی را در میان دانش‌آموزان ایرانی به خود اختصاص دادند. با این حال، توزیع سبک‌ها در این سال نسبت به سال‌های گذشته

دانش‌آموزان را به سمت تفکر انتقادی‌تر، ابراز عقیده مستدل و کاهش رفتارهای پاسخ‌دهی مبتنی بر عدم قطعیت یا حدس و گمان سوق دهند.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش: این پژوهش به صورت تحلیل ثانویه بر روی داده‌های از پیش جمع‌آوری شده آزمون بین‌المللی تیمز انجام گرفته است. از آنجایی که پژوهشگر به صورت مستقیم با شرکت‌کنندگان در ارتباط نبوده و داده‌ها به صورت کاملاً بی‌نام از بانک داده‌های آزمون تیمز و مرکز مطالعات تیمز و پرلز کشور دریافت شده‌اند، ملاحظات مربوط به کسب رضایت آگاهانه مستقیم از شرکت‌کنندگان در فرآیند جمع‌آوری داده‌های اولیه توسط متولیان آزمون رعایت شده است. در این پژوهش، اصل محرمانگی اطلاعات شرکت‌کنندگان به طور کامل حفظ گردید و داده‌های مورد استفاده فاقد هرگونه شناسه قابل انتساب به فرد بودند. دسترسی به داده‌ها محدود به تیم پژوهشی بود و کلیه تحلیل‌ها تنها برای اهداف علمی و پژوهشی تعریف شده در این مطالعه انجام پذیرفت.

حامی مالی: این پژوهش در قالب رساله دکتری و بدون حمایت مالی می‌باشد.

نقش هر یک از نویسندگان: این مقاله از رساله دکتری نویسنده اول و به راهنمایی نویسنده دوم و مشاوره نویسنده سوم استخراج شده است.

تضاد منافع: نویسندگان همچنین اعلام می‌دارند که در نتایج این پژوهش هیچ‌گونه تضاد منافی وجود ندارد.

تشکر و قدردانی: بدین وسیله از شرکت‌کنندگانی که در این پژوهش شرکت کردند، تشکر و قدردانی می‌گردد.

منابع

ایلاتلو حسین (۱۳۹۸). بررسی تفکر انتقادی و آموزش آن در آموزش و پرورش. *رویش روان شناسی*، ۸ (۵): ۱۰۶-۱۰۱. URL: <http://frooyesh.ir/article-147-fa.html>

(به‌ویژه ۲۰۱۵) تا حدی به سمت تعادل بیشتر میل کرده است. تغییرات در توزیع سبک‌های پاسخ در طول زمان می‌تواند با نظریه‌های تغییرات رشدی یا تأثیرات نسلی در حوزه روان‌سنجی یا روان‌شناسی تربیتی مرتبط باشد. این نظریه‌ها بیان می‌کنند که عوامل محیطی و تجربیات مشترک یک گروه سنی در طول زمان می‌توانند بر ویژگی‌های شناختی و رفتاری، از جمله نحوه پاسخ‌دهی در آزمون‌ها، تأثیر بگذارند (داماس و همکاران، ۲۰۲۰). این تغییرات نسبی در توزیع سبک‌ها می‌تواند به‌طور غیرمستقیم به تحولات در سیاست‌ها و شیوه‌های آموزشی در ایران در طول سال‌های اخیر اشاره داشته باشد. هرچند تغییرات ریشه‌ای ممکن است زمان‌بر باشد، اما تأکید بیشتر بر روش‌های تدریس فعال‌تر، استفاده از فناوری در آموزش، یا تغییر در محتوای درسی علوم می‌تواند به تدریج بر رویکرد دانش‌آموزان در مواجهه با سؤالات آزمون و در نتیجه سبک‌های پاسخ آن‌ها تأثیرگذار باشد. با این حال، تداوم شیوع سبک‌های افراطی و معمولی نشان می‌دهد که چالش‌های ساختاری در نظام آموزشی و فرهنگی همچنان نیازمند توجه جدی هستند.

پژوهش حاضر، با وجود ارائه بینش‌های ارزشمند در زمینه سبک‌های پاسخ دانش‌آموزان ایرانی به آزمون علوم تیمز، دارای محدودیت‌هایی است که اغلب خارج از کنترل مستقیم پژوهشگر بوده‌اند. اصلی‌ترین محدودیت، ماهیت تحلیل ثانویه داده‌ها است؛ به این معنا که پژوهشگر کنترلی بر فرآیند جمع‌آوری اولیه داده‌ها، انتخاب نمونه اصلی توسط سازمان متولی تیمز در ایران، یا متغیرهای زمینه‌ای جمع‌آوری شده نداشته است. این موضوع، امکان بررسی عمیق‌تر عوامل مؤثر بر سبک‌های پاسخ (مانند ویژگی‌های معلم، مدرسه، یا عوامل خانوادگی) را محدود می‌سازد. همچنین، ماهیت مقطعی داده‌های تیمز در هر چرخه، اجازه تحلیل علی یا بررسی تغییرات فردی دانش‌آموزان در طول زمان را نمی‌دهد و تحلیل‌ها تنها بر اساس داده‌های تجمعی در هر سال صورت گرفته است. بر اساس یافته‌های این پژوهش، پیامدهای عملی متعددی قابل طرح است. شناخت سبک‌های پاسخ رایج در میان دانش‌آموزان ایرانی می‌تواند به طراحان آزمون در کالیبراسیون بهتر آزمون‌ها و کاهش تأثیر عوامل مزاحم بر اندازه‌گیری توانایی واقعی کمک کند. همچنین، معلمان و برنامه‌ریزان درسی می‌توانند با آگاهی از این سبک‌ها، روش‌های تدریس را به گونه‌ای اصلاح کنند که

References

- Aleksankov, A. M. (2017). The Fourth industrial revolution and modernization of education: international experience. *Strategic priorities*, 1(13), 53-69. DOI: [10.24411/2413-0324-2017-10013](https://doi.org/10.24411/2413-0324-2017-10013)
- Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British journal of mathematical and statistical psychology*, 70(1), 159-181. <https://doi.org/10.1111/bmsp.12086>
- Chen, X. (2022). The effects of individual-and class-level achievement on attitudes towards mathematics: An analysis of Hong Kong students using TIMSS 2019. *Studies in Educational Evaluation*, 72, 101113. DOI: [10.1016/j.stueduc.2022.101113](https://doi.org/10.1016/j.stueduc.2022.101113)
- 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. DOI: [10.18271/ijpe.87752](https://doi.org/10.18271/ijpe.87752)
- Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center*, 17(1), 1-7. DOI: [10.2139/ssrn.3419383](https://doi.org/10.2139/ssrn.3419383)
- Dumas, D., McNeish, D., & Greene, J. A. (2020). Dynamic measurement: A theoretical-psychometric paradigm for modern educational psychology. *Educational Psychologist*, 55(2), 88-105. DOI: [10.1080/00461520.2020.1744150](https://doi.org/10.1080/00461520.2020.1744150)
- Ehrlich, H. J. (1964). Instrument error and the study of prejudice. *Social Forces*, 43(2), 197-206. DOI: [10.2307/2574618](https://doi.org/10.2307/2574618)
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis: A primer. *Journal of cross-cultural psychology*, 34(2), 195-210. <https://doi.org/10.1177/0022022102250427>
- Henninger, M., & Meiser, T. (2020). Different approaches to modeling response styles in divide-by-total item response theory models (part 1): A model integration. *Psychological Methods*, 25(5), 560. DOI: [10.31234/osf.io/4m7vu](https://doi.org/10.31234/osf.io/4m7vu)
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture*, 2(1), 8. <https://doi.org/10.9707/2307-0919.1014>
- Hutton, A. C. (2017). *Assessing acquiescence in surveys using positively and negatively worded questions*. Virginia Commonwealth University. DOI: [10.25772/9499](https://doi.org/10.25772/9499)
- Ilanloo H. (2019). Examining critical thinking and teaching it in education. *Rooyesh*. 8(5), 101-106. <http://frooyesh.ir/article-1-947-fa.html> (In Persian).
- 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. DOI: [10.14419/ijet.v7i4.30.17713](https://doi.org/10.14419/ijet.v7i4.30.17713)
- Isozaki, T. (2018). Science teacher education in Japan: past, present, and future. *Asia-Pacific Science Education*, 4(1), 1-14. DOI: [10.1186/s41029-018-0027-2](https://doi.org/10.1186/s41029-018-0027-2)
- Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior research methods*, 48, 1070-1085. DOI: [10.3758/s13428-015-0647-7](https://doi.org/10.3758/s13428-015-0647-7)
- Kaufman, E. A., Beeney, J. E., & Stepp, S. D. (2023). Self-report response style bias and borderline personality features. *Current Psychology*, 42(24), 20443-20451. DOI: [10.1007/s12144-023-04431-z](https://doi.org/10.1007/s12144-023-04431-z)
- Khorramdel, L., von Davier, M., & Pokropek, A. (2019). Combining mixture distribution and multidimensional IRT models for the measurement of extreme response styles. *British Journal of Mathematical and Statistical Psychology*, 72(3), 538-559. DOI: [10.1111/bmsp.12178](https://doi.org/10.1111/bmsp.12178)
- Kim, N., & Bolt, D. M. (2021). A mixture IRT model for extreme response style: Accounting for response process uncertainty. *Educational and Psychological*

- Measurement, 81(1), 131-154. <https://doi.org/10.1177/0013164420913915>
- Liu, M., Harbaugh, A. G., Haring, J. R., & Hancock, G. R. (2017). The effect of extreme response and non-extreme response styles on testing measurement invariance. *Frontiers in psychology*, 8, 726. doi.org/10.3389/fpsyg.2017.00726
- Lu, Y., & Bolt, D. M. (2015). Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style. *Large-scale Assessments in Education*, 3, 1-18. <https://doi.org/10.1186/s40536-015-0012-0>
- Majeed, F. A., & Abuharaz, M. (2019, March). Multi DoF system modelling-Transfer Function and State variable approaches. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1-6). IEEE. DOI: [10.1109/ICASET.2019.8714448](https://doi.org/10.1109/ICASET.2019.8714448)
- Martin, M. O., von Davier, M., & Mullis, I. V. (2020). Methods and procedures: TIMSS 2019 Technical Report. *International Association for the Evaluation of Educational Achievement*. DOI: [10.37606/tmtr.54](https://doi.org/10.37606/tmtr.54)
- Papanastasiou, E. C., & Michaelides, M. P. (2024). Examining successful and unsuccessful time management through process data: two novel indicators of test-taking behaviors. *Large-scale Assessments in Education*, 12(1), 3. DOI: [10.1186/s40536-024-00193-z](https://doi.org/10.1186/s40536-024-00193-z)
- Saatcioglu, F. M., & Sen, S. (2023). The analysis of TIMSS 2015 data with confirmatory mixture item response theory: A multidimensional approach. *International Journal of Testing*, 23(4), 257-275. DOI: [10.17275/per.20.41.7.3](https://doi.org/10.17275/per.20.41.7.3)
- Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two-and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior research methods*, 50, 2511-2522. DOI: [10.3758/s13428-018-1029-4](https://doi.org/10.3758/s13428-018-1029-4)
- Tsupros, N. Kohler, R. & Hallinen, J. (2009). STEM education: A project to identify the missing components. *Intermediate Unit*, 1, 11-17. DOI: [10.4135/9781452226464.n6](https://doi.org/10.4135/9781452226464.n6)
- Tutz, G., Schauberger, G., & Berger, M. (2018). Response styles in the partial credit model. *Applied Psychological Measurement*, 42(6), 407-427. <https://doi.org/10.1177/0146621617748322>
- Wang, Y., & Lewis, S. E. (2022). Towards a theoretically sound measure of chemistry students' motivation; investigating rank-sort survey methodology to reduce response style bias. *Chemistry Education Research and Practice*, 23(1), 240-256. DOI: [10.1039/D1RP00206F](https://doi.org/10.1039/D1RP00206F)
- Zhang, Y., & Wang, Y. (2020). Validity of three IRT models for measuring and controlling extreme and midpoint response styles. *Frontiers in Psychology*, 11, 271. DOI: [10.3389/fpsyg.2020.00271](https://doi.org/10.3389/fpsyg.2020.00271)