

تعیین تفاوت توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۱، ۲۰۱۵، ۲۰۱۹

۱. فاطمه آزادی: دانشجوی دکتری سنجش و اندازه‌گیری، گروه روانشناسی، دانشکده انسانی، واحد ساوه، دانشگاه آزاد اسلامی، ساوه، ایران
۲. اصغر مینایی*: دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. (ایمیل: asghar.minaei@yahoo.com)
۳. مسعود گرامی پور: دانشیار، گروه آموزشی برنامه‌ریزی درسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران

چکیده

پژوهش حاضر با هدف تعیین تفاوت توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۱، ۲۰۱۵، ۲۰۱۹ انجام شد. در این مطالعه، سبک‌های مختلف پاسخ‌دهی، از جمله افراطی، میانی و تصدیقی، و تأثیر آن‌ها بر پارامترهای روان‌سنجی سؤال‌ها و توانایی افراد بررسی شد. داده‌ها با استفاده از مدل‌های مختلف IRT تحلیل شدند. نتایج نشان داد که در آزمون علوم تیمز پایه هشتم ۲۰۱۱ افراد دارای سبک پاسخ معمولی توانایی بالاتری دارند در رده بعدی افراد دارای سبک پاسخ معمولی و افراد سبک پاسخ میانی قرار دارند و کمترین توانایی نیز مربوط به افراد با سبک پاسخ رضایت است. در آزمون علوم تیمز پایه هشتم ۲۰۱۵، میانگین توانایی افراد با سبک پاسخ افراطی بالاترین میانگین در بین گروه‌ها بوده است، می‌توان ادعا کرد که افراد این گروه به لحاظ توانایی از سایر افراد برتر هستند. همچنین در آزمون علوم تیمز پایه هشتم ۲۰۱۹، 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 ($P < 0.05$). بین توانایی افراد سبک پاسخ معمولی با توانایی افراد سبک پاسخ رضایت نیز تفاوت معنادار آماری وجود دارد. این نتایج می‌تواند به بهبود طراحی آزمون‌های استاندارد، کاهش سوگیری‌های فرهنگی و ارائه راهکارهایی برای ارزیابی دقیق‌تر در سطوح ملی و بین‌المللی کمک کند.

واژگان کلیدی: سبک‌های پاسخ، آزمون‌های تیمز، پارامتر سؤال، توانایی افراد

مقدمه

یکی از مهم‌ترین مسائل دنیای امروز، پرورش و توسعه درک علم برای دانش‌آموزان است. در واقع 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 (Chen, 2022). موضوعات علوم نباید به‌طور جداگانه تدریس شوند، از این‌رو در رویکرد بین‌رشته‌ای آموزش مفاهیم آکادمیک ارتباط نزدیکی با تجربیات دنیای واقعی دارند و دانش‌آموزان، علوم، فناوری، مهندسی و ریاضیات را به‌گونه‌ای آموزش می‌بینند که ارتباط بین مدرسه، جامعه، کار و تجارت برقرار شود و به دنبال آن امکان توسعه سواد و توانایی رقابت در اقتصاد فراهم شود (Saphira, 2024; Tsupros et al., 2009). از این‌رو محققان برآنند تا ابزارهای مناسب جهت سنجش و ارزیابی میزان پیشرفت علوم، تدوین و اجرا نمایند (Meadows, 2023; Saphira, 2024; Xu & Ouyang, 2022). یکی از این آزمون‌ها مجموعه آزمون‌های تیمز است که داده‌های قابل‌اعتماد 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 به‌منظور هماهنگی برنامه‌های درسی و سیستم آموزشی ریاضیات و علوم به‌طور گسترده در کشورهای شرکت‌کننده ایجاد شده است (Martin et al., 2020).

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 (Ismail et al., 2018). 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 (Khalilpour, 2024). بوکلی (۲۰۰۹)، وان دگیر و آدامز (۲۰۱۰) دریافتند 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 (Bhutoria & Aljabri, 2022; Chen, 2022; Wu, 2009; Yoshino, 2012). از این‌رو نتایج متفاوت^۱ در سطوح فردی و گروهی اعتبار نتایج به‌دست‌آمده 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 پاسخ افراد مشکلات روان‌سنجی مختلفی در داده‌ها ایجاد کرده است. به بیان دیگر، از طریق تولید

1 interchangeable

2 Response styles

یک خطای سیستماتیک^۱ در نمرات افرادی که از سطح یکسانی از دانش^۲، نگرش^۳ یا خصوصیات شخصیتی^۴ برخوردارند، باعث کاهش اعتبار^۵ نمره‌های آزمون می‌شود (Kim & Bolt, 2021). نتایج تحقیقات پیشین نشان می‌دهد که سبک‌های پاسخ متفاوتی وجود دارند که هر کدام به‌نوبه‌ی خود تأثیر منفی بر اعتبار نمرات آزمون می‌گذارند. زمانی که سبک پاسخ به سبک پاسخی نهایی^۶ (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 پاسخی نهایی) بیشتر انتخاب شود، ممکن است سوگیری رخ 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 (Liu et al., 2017). از آنجاکه همبستگی و واریانس نمرات تا حدودی مرتبط با یکدیگر هستند، نتایج به‌دست‌آمده از همبستگی بین متغیرها از سبک‌های پاسخ تأثیر می‌پذیرند. مشخصاً، از آنجاکه سبک پاسخی 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 پاسخی نهایی بر نتایج به‌دست‌آمده از آن‌ها تأثیر خواهد گذاشت (Dibek & Cikrikci, 2021). علاوه بر سبک پاسخی نهایی، سبک پاسخ تصدیق یا مخالفت^۷ و سبک پاسخ نمره میانی^۸ و غیره نیز بر نتایج آزمون‌ها تأثیر می‌گذارند. 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 یک‌سو، همان‌طور که گفته شد، سبک پاسخی نهایی گرایش به انتخاب گزینه‌های انتهایی مقوله‌های پاسخ مانند «کاملاً موافق/کاملاً مخالف» است (Dibek & Cikrikci, 2021). 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 پاسخی نهایی بر سطح همبستگی‌ها بین معیارها و در نتیجه نتایج به‌دست‌آمده از تحلیل‌های عاملی، تأثیر می‌گذارد. از سوی دیگر، سبک پاسخی نهایی معمولاً باعث می‌شود که میانگین نمرات سؤال‌ها کمتر یا بیشتر از نمرات واقعی تخمین زده شود (Kim & Bolt, 2021) که منتج به ایجاد سوگیری در نتایج می‌شود. این نمره‌های دارای سوگیری ممکن است منتج به نتایج خطای اندازه‌گیری شوند (Ulitzsch et al., 2023; Wang et al., 2022). بنابراین، این واقعیت که سبک‌های پاسخ منتج به استنتاج‌های غلط و کاربردهای نادرست برای تصمیمات و خطامشی‌های آموزشی در سطح ملی

1 systematic error

2 knowledge

3 attitude

4 personality characteristics

5 reliability

6 extreme response style

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

8 middle response styles (MRS)

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 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 این سوال ایجاد می‌شود که آیا توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ متفاوت است؟

روش پژوهش

این پژوهش از نظر هدف کاربردی و از لحاظ روش از نوع همبستگی می‌باشد این پژوهش برای شناسایی سبک‌های پاسخ از داده‌های حاصل از اجرای ۳ دوره آزمون‌های تیمز (۲۰۱۱، ۲۰۱۵ و 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 است که طی سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ در این آزمون شرکت کرده‌اند. نمونه این پژوهش شامل کلیه شرکت‌کنندگان ایرانی در آزمون علوم پایه هشتم تیمز بوده است که طبق آمار دریافتی از مرکز مطالعات تیمز و پرلز کشور و نیز داده‌های موجود در سایت رسمی IEA تعداد کل شرکت‌کننده‌ها ۱۸۱۲۴ نفر بوده‌اند. با توجه به اینکه نمونه موردنظر بنا به درخواست پژوهشگر از بانک داده‌های آزمون تیمز تهیه شده و پژوهشگر هیچ دخالتی در امر انتخاب نداشته است، می‌توان روش نمونه‌گیری پژوهش را نمونه‌گیری ساده محسوب کرد. ابزار مورد استفاده در این پژوهش آزمون علوم تیمز در سال‌های ۲۰۱۱، ۲۰۱۵ و ۲۰۱۹ بوده است 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 است 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، در حالی که تیمز پیشرفته دانش‌آموزان را در سال آخر متوسطه در ریاضیات پیشرفته و فیزیک مورد ارزیابی قرار می‌دهد. 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 مساله 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 می‌توانند 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 با توجه به پیشینه پژوهش‌های انجام‌شده 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 به‌گزینه‌های 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 اعتمادبه‌نفس به‌صورت معکوس نمره‌گذاری شدند.

به‌منظور 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 R و Rstudio انجام شده است. در پژوهش‌های پیشین به‌منظور ویژگی‌های روانسنجی آزمون‌های تیمز از جمله پارامترهای سؤال، توانایی افراد از روش‌های مختلف مانند نظریه سؤال پاسخ به‌منظور برآورد پارامترهای سؤال و توانایی افراد پرداخته شده است، منحنی ویژگی سؤال، تابع آگاهی سؤال، تابع آگاهی آزمون و منحنی خطای استاندارد آزمون، پیشینه آگاهی سؤالات، پیشینه آگاهی آزمون به‌وسیله نرم‌افزارهای SPSS، WinBugs، BILOG-MG3، NOHARM و R استفاده شده است، همچنین شریفی یگانه (۱۳۹۱) به‌منظور کارکرد افتراقی آزمون تیمز از دو روش ماتل-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 توانایی افراد برحسب سبک‌های پاسخ در آزمون تیمز ۲۰۱۱ نشان داد میانگین توانایی سبک پاسخ افراطی از سایر گروه‌ها بهتر بوده است ($M=+0/151$) و میانگین توانایی سبک پاسخ رضایت ($M=-0/381$) از همه کمتر شده است. برای قضاوت درباره معناداری تفاوت‌ها نتایج آزمون تحلیل واریانس در جدول (۱) گزارش شده است.

جدول ۱. نتایج آزمون تحلیل واریانس جهت مقایسه توانایی افراد با سبک پاسخ متفاوت در آزمون تیمز ۲۰۱۱

| منابع تغییر | مجموع مجزورات | درجه آزادی | میانگین مجزورات | F | سطح معناداری |
|-------------|---------------|------------|-----------------|--------|--------------|
| بین گروهی | ۱۳/۲۷۹ | ۳ | ۴۵/۷۶ | ۴۹/۶۵۸ | ۰/۰۰۰۱ |
| درون گروهی | ۵۵۳۸/۱۵۹ | ۶۰۱۰ | ۰/۹۲۱ | | |
| مجموع | ۵۶۷۵/۴۳۸ | ۶۰۱۳ | | | |

بر اساس اطلاعات جدول (۱) نتیجه آزمون تحلیل واریانس به لحاظ آماری معنادار است ($P < 0/05$ و $F = 49/658$). 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 (۲) گزارش شده است.

جدول ۲. نتایج آزمون تعقیبی بونفرونی در آزمون تیمز ۲۰۱۱

| گروه ۱ | گروه ۲ | تفاوت میانگین | خطای استاندارد | سطح معناداری |
|-----------------|-----------------|---------------|----------------|--------------|
| سبک پاسخ معمولی | سبک پاسخ افراطی | -۰/۱۲۴ | ۰/۰۳۰ | ۰/۰۰۰۱ |
| | سبک پاسخ میانی | ۰/۰۷۹ | ۰/۰۴۶ | ۰/۴۹۲ |
| | سبک پاسخ رضایت | ۰/۴۰۷ | ۰/۰۴۰ | ۰/۰۰۰۱ |
| سبک پاسخ افراطی | سبک پاسخ میانی | ۰/۲۰۴ | ۰/۰۴۹ | ۰/۰۰۰۱ |
| | سبک پاسخ رضایت | ۰/۵۳۲ | ۰/۰۴۴ | ۰/۰۰۰۱ |
| سبک پاسخ میانی | سبک پاسخ رضایت | ۰/۳۲۸ | ۰/۰۵۶ | ۰/۰۰۰۱ |

با توجه به اطلاعات جدول (۲) تنها بین توانایی افراد سبک پاسخ معمولی و توانایی افراد سبک پاسخ میانی تفاوت معنادار وجود ندارد ($P > 0/05$)، سایر گروه‌ها با همدیگر دارای تفاوت معنادار آماری هستند ($P < 0/05$)؛ یعنی افراد سایر گروه‌ها به لحاظ توانایی تفاوت دارند و با توجه به جدول اطلاعات توصیفی، افراد دارای سبک پاسخ معمولی توانایی بالاتری دارند در رده بعدی افراد دارای سبک پاسخ معمولی و افراد سبک پاسخ میانی قرار دارند و کمترین توانایی نیز مربوط به افراد با سبک پاسخ رضایت است.

نتایج توصیفی توانایی افراد برحسب سبک‌های پاسخ در آزمون تیمز ۲۰۱۵ نشان داد میانگین توانایی سبک پاسخ افراطی از سایر گروه‌ها بهتر بوده است ($M=+0/125$) و میانگین توانایی سبک پاسخ میانی ($M=-0/099$) از همه کمتر شده است. برای قضاوت درباره معناداری تفاوت‌ها نتایج آزمون تحلیل واریانس در جدول (۳) گزارش شده است.

جدول ۳. نتایج آزمون تحلیل واریانس جهت مقایسه توانایی افراد با سبک پاسخ متفاوت در آزمون تیمز ۲۰۱۵

| منابع تغییر | مجموع مجذورات | درجه آزادی | میانگین مجذورات | F | سطح معناداری |
|-------------|---------------|------------|-----------------|--------|--------------|
| بین گروهی | ۴۳/۳۶۶ | ۳ | ۱۴/۴۵۵ | ۱۷/۴۵۵ | ۰/۰۰۰۱ |
| درون گروهی | ۵۱۳۶/۰۳۴ | ۶۱۲۶ | ۰/۸۳۸ | | |
| مجموع | ۵۱۷۹/۴ | ۶۱۲۹ | | | |

بر اساس اطلاعات جدول (۳) نتیجه آزمون تحلیل واریانس به لحاظ آماری معنادار است ($F=17/455$ و $P<0/05$). 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 (۴) گزارش شده است.

جدول ۴. نتایج آزمون تعقیبی بونفرونی در آزمون تیمز ۲۰۱۵

| گروه ۱ | گروه ۲ | تفاوت میانگین | خطای استاندارد | سطح معناداری |
|-----------------|-----------------|---------------|----------------|--------------|
| سبک پاسخ معمولی | سبک پاسخ افراطی | -۰/۱۵۵ | ۰/۰۲۹ | ۰/۰۰۰۱ |
| | سبک پاسخ میانی | ۰/۰۶۸ | ۰/۰۴۶ | ۰/۸۳۵ |
| | سبک پاسخ رضایت | ۰/۰۳۰ | ۰/۰۳ | ۰/۹۹۹ |
| سبک پاسخ افراطی | سبک پاسخ میانی | ۰/۲۲۳ | ۰/۰۴۶ | ۰/۰۰۰۱ |
| | سبک پاسخ رضایت | ۰/۱۸۵ | ۰/۰۳ | ۰/۰۰۰۱ |
| سبک پاسخ میانی | سبک پاسخ رضایت | ۰/۰۳۸ | ۰/۰۵ | ۰/۹۹۹ |

با توجه به اطلاعات جدول (۴) تنها بین توانایی افراد با سبک پاسخ افراطی با توانایی سایر سبک‌ها تفاوت معنادار آماری وجود دارد ($P<0/05$). با توجه به اینکه میانگین توانایی افراد با سبک پاسخ افراطی بالاترین میانگین در بین گروه‌ها بوده است، می‌توان ادعا کرد که افراد این گروه به لحاظ توانایی از سایر افراد برتر هستند.

اطلاعات توصیفی توانایی افراد برحسب سبک‌های پاسخ در آزمون تیمز ۲۰۱۹ نشان داد میانگین توانایی سبک پاسخ افراطی از سایر گروه‌ها بهتر بوده است ($M=0/159$) و میانگین توانایی سبک پاسخ رضایت ($M=-0/117$) از همه کمتر شده است. برای قضاوت درباره معناداری تفاوت‌ها نتایج آزمون تحلیل واریانس در جدول (۵) گزارش شده است.

جدول ۵. 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 (۵) 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 (۶) گزارش شده است.

جدول ۶. 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 (۶) 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

این پژوهش با هدف تعیین تفاوت توانایی دانش‌آموزان ایرانی بر حسب سبک پاسخ آن‌ها در آزمون علوم تیمز پایه هشتم ۲۰۱۱، ۲۰۱۵، ۲۰۱۹ انجام شد. نتایج تحلیل واریانس یک‌طرفه نشان داد که بین توانایی شناختی دانش‌آموزان پایه هشتم در آزمون علوم تیمز ۲۰۱۱ بر اساس سبک پاسخ آن‌ها تفاوت معناداری وجود دارد. این یافته تأیید می‌کند که سبک پاسخ دانش‌آموزان می‌تواند بر توانایی‌های شناختی آنان در آزمون تأثیر بگذارد. بر اساس تحلیل توصیفی، میانگین توانایی شناختی گروه دارای سبک پاسخ افراطی از سایر گروه‌ها بالاتر بود، در حالی که گروه دارای سبک پاسخ رضایت کمترین میانگین توانایی را داشتند. همچنین نتایج آزمون تعقیبی بونفرونی نشان داد که به جز تفاوت میان گروه‌های دارای سبک پاسخ معمولی و میانی، سایر تفاوت‌ها بین گروه‌ها از لحاظ آماری معنادار بودند.

دانش‌آموزانی که سبک پاسخ افراطی دارند، ممکن است به دلیل گرایش به انتخاب گزینه‌های انتهایی (کاملاً موافق یا کاملاً مخالف)، دقت بیشتری در خواندن سؤالات و ارزیابی گزینه‌ها داشته باشند. این سبک پاسخ می‌تواند ناشی از اعتمادبه‌نفس بالا یا تمایل به ابراز قطعی تر عقاید باشد که در این پژوهش با توانایی شناختی بالاتر مرتبط شده است. از سوی دیگر، این سبک می‌تواند بیانگر خودکارآمدی بالاتر در آزمون‌های استاندارد باشد که به‌عنوان عامل مهمی در موفقیت تحصیلی شناسایی شده است. سبک پاسخ رضایت

معمولاً با گرایش به انتخاب گزینه‌های متوسط یا میانه مرتبط است. این الگو ممکن است ناشی از عدم اطمینان، اضطراب آزمون، یا کمبود انگیزه باشد که به عملکرد ضعیف‌تر در آزمون منجر می‌شود.

در آزمون علوم تیمز ۲۰۱۱ نشان داد تفاوت معناداری بین توانایی شناختی دانش‌آموزان دارای سبک پاسخ معمولی و میانی وجود ندارد. این امر می‌تواند به مشابهت در نحوه پردازش اطلاعات و استراتژی‌های حل مسئله در این دو گروه اشاره داشته باشد، بنابراین نتایج این پژوهش به‌طور واضح نشان داد که سبک پاسخ دانش‌آموزان می‌تواند تأثیر معناداری بر توانایی شناختی آنان در آزمون‌های استاندارد داشته باشد. این یافته‌ها، علاوه بر تأیید اهمیت سبک‌های پاسخ، بر لزوم توجه به راهبردهای آموزشی متناسب با نیازهای شناختی و فرهنگی تأکید دارند.

توانایی شناختی دانش‌آموزان ایرانی پایه هشتم در آزمون علوم تیمز ۲۰۱۵ بر اساس سبک پاسخ آنان تفاوت معناداری وجود دارد، این تفاوت نشان‌دهنده تأثیر سبک پاسخ بر عملکرد دانش‌آموزان در آزمون‌های استاندارد است. دانش‌آموزان دارای سبک پاسخ افراطی بیشترین میانگین توانایی شناختی را داشتند، دانش‌آموزان دارای سبک پاسخ میانی کمترین میانگین توانایی را نشان دادند و دانش‌آموزان با سبک‌های پاسخ رضایت و معمولی میانگین‌های تقریباً مشابه و نزدیک به صفر داشتند. نتایج آزمون تعقیبی بونفرونی نیز نشان داد که تنها میان دانش‌آموزان دارای سبک پاسخ افراطی با سایر گروه‌ها تفاوت معنادار وجود دارد، در حالی که تفاوت بین سایر گروه‌ها معنادار نبود. این یافته به وضوح نشان می‌دهد که سبک پاسخ افراطی نقشی مثبت و برجسته در عملکرد آزمونی دارد. دانش‌آموزانی که از سبک پاسخ افراطی استفاده می‌کنند، به انتخاب گزینه‌های قطعی (کاملاً موافق یا کاملاً مخالف) تمایل دارند. این رفتار ممکن است به اعتمادبه‌نفس بالاتر و تصمیم‌گیری سریع‌تر در شرایط آزمونی نسبت داده شود. این سبک پاسخ، همچنین می‌تواند بازتابی از خودکارآمدی یا استراتژی‌های یادگیری کارآمد باشد. دانش‌آموزانی که به‌طور افراطی پاسخ می‌دهند، ممکن است از توانایی‌های شناختی و مهارت‌های پردازش اطلاعات پیچیده بیشتری برخوردار باشند. سبک پاسخ میانی، که شامل انتخاب گزینه‌های میانه در طیف پاسخ‌هاست، ممکن است نشان‌دهنده عدم اطمینان یا تردید دانش‌آموزان در درک یا حل مسائل باشد. در واقع افرادی که به‌طور مداوم از گزینه‌های میانه استفاده می‌کنند، معمولاً در به‌کارگیری استراتژی‌های شناختی عمیق ضعیف‌تر عمل می‌کنند و ممکن است انگیزه کمتری برای شرکت فعالانه در آزمون داشته باشند. شباهت میانگین توانایی‌های این دو گروه ممکن است به ویژگی‌های مشترک آن‌ها در پردازش اطلاعات یا رویکرد مشابه در پاسخ‌دهی مرتبط باشد. با این حال، نتایج نشان می‌دهد که سبک پاسخ معمولی نسبت به سبک رضایت عملکرد کمی بهتر داشته است، که ممکن است ناشی از تصمیم‌گیری دقیق‌تر در انتخاب گزینه‌ها باشد.

در آزمون تیمز پایه هشتم ۲۰۱۹ توانایی شناختی دانش‌آموزان بر حسب سبک پاسخ آنان تفاوت معناداری دارد. این نتایج بار دیگر بر تأثیر سبک پاسخ بر عملکرد دانش‌آموزان تأکید می‌کند. 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. میانگین بالاتر توانایی دانش‌آموزان دارای سبک پاسخ افراطی، همانند نتایج تیمز ۲۰۱۱ و ۲۰۱۵، تأییدکننده نقش مثبت این سبک در آزمون‌های استاندارد است. این سبک پاسخ به دلیل انتخاب‌های قطعی‌تر (کاملاً موافق یا کاملاً مخالف)، ممکن است نشان‌دهنده درک بهتر، اعتمادبه‌نفس 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. همچنین نتایج این پژوهش در تیمز ۲۰۱۹ با الگوی مشاهده‌شده در سال‌های ۲۰۱۱ و ۲۰۱۵ مشابه است، جایی که سبک پاسخ افراطی عملکرد برتری نشان داده و سبک پاسخ رضایت کمترین توانایی را ثبت کرده است. این همسویی نشان می‌دهد که تأثیر سبک پاسخ بر عملکرد آزمون‌های دانش‌آموزان ایرانی در طول زمان پایدار است.

منابع

- Bhutoria, A., & Aljabri, N. (2022). Managerial Practices and School Efficiency: A Data Envelopment Analysis Across OECD and MENA Countries Using TIMSS 2019 Data. *Large-scale Assessments in Education*, 10(1). <https://doi.org/10.1186/s40536-022-00147-3>
- 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. <https://doi.org/10.1016/j.stueduc.2021.101113>
- 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. <https://doi.org/10.29329/ijpe.2021.332.12>
- 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>
- Khalilpour, H. (2024). Selecting an Appropriate Cognitive Diagnostic Model for Eighth Grade Math Questions: TIMSS 2019 Study. *Journal of Psychological Dynamics in Mood Disorders (PDMD)*, 3(2), 10-22. <https://doi.org/10.22034/pdmd.2024.462389.1118>
- 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>
- Kim, N., & Bolt, D. M. (2021). A Mixture IRTree 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. <https://doi.org/10.3389/fpsyg.2017.00726>

- 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.
- Meadows, J. (2023). Planning an Informal STEM Event? Try the Assets-Based Approach to Planning and Research for Informal STEM Events. *Journal of Research in Science Mathematics and Technology Education*, 1-16. <https://doi.org/10.31756/jrsmte.711>
- Saphira, H. V. (2024). Metaverse: A Paradigm Shift in STEM Education for Science Learning Beyond the Review. *E3s Web of Conferences*, 482, 04004. <https://doi.org/10.1051/e3sconf/202448204004>
- Tsupros, N., Kohler, R., & Hallinen, J. (2009). STEM education: A project to identify the missing components. *Intermediate Unit, 1*, 11-17.
- Ulitzsch, E., Lüdtke, O., & Robitzsch, A. (2023). The Role of Response Style Adjustments in Cross-Country Comparisons-A Case Study Using Data from the PISA 2015 Questionnaire. *Educational Measurement: Issues and Practice*. <https://doi.org/10.1111/emip.12552>
- 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>
- Wu, M. (2009). A Comparison of PISA and TIMSS 2003 Achievement Result in Mathematics. *Prospect*, 39(1), 33-46. <https://doi.org/10.1007/s11125-009-9109-y>
- Xu, W., & Ouyang, F. (2022). The Application of AI Technologies in STEM Education: A Systematic Review From 2011 to 2021. *International Journal of Stem Education*. <https://doi.org/10.1186/s40594-022-00377-5>
- Yoshino, A. (2012). The Relationship Between Self-Concept and Achievement in TIMSS 2007: A Comparison Between American and Japanese Students. *International Review of Education*, 58(2), 199-219. <https://doi.org/10.1007/s11159-012-9283-7>

Determining the differences in the ability of Iranian students based on their response styles in the TIMSS Grade 8 science test (2011, 2015, and 2019)

1. Fatemeh Azadi: PhD Student of Psychology, Saveh Branch, Islamic Azad University, Saveh, Iran
2. Asghar Minaei*: Associate Professor, Department of Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabatabaie University, Tehran, Iran. (Email: asghar.minaei@yahoo.com)
3. Masoud Geramipour: Associate Professor, Department of Curriculum Planning Education, Faculty of Education and Psychology, Kharazmi University, Tehran, Iran

Abstract

The present study aimed to identify the differences in the ability of Iranian students based on their response styles in the TIMSS (Trends in International Mathematics and Science Study) Grade 8 science test for the years 2011, 2015, and 2019. This study examined various response styles, including extreme, moderate, and acquiescent styles, as well as their effects on item psychometric parameters and individual abilities. Data were analyzed using various Item Response Theory (IRT) models. The results showed that in the 2011 TIMSS Grade 8 science test, individuals with a typical response style exhibited the highest ability, followed by those with a moderate response style. The lowest ability was observed among individuals with an acquiescent response style. In the 2015 TIMSS Grade 8 science test, individuals with an extreme response style had the highest average ability among all groups, suggesting their superior ability compared to other individuals. Furthermore, in the 2019 TIMSS Grade 8 science test, there was no significant difference in the abilities of individuals with a moderate response style and those with an acquiescent response style ($P > 0.05$). However, a statistically significant difference was found between the abilities of individuals with an extreme response style and those of other groups ($P < 0.05$). A statistically significant difference was also observed between the abilities of individuals with a typical response style and those with an acquiescent response style. These findings can contribute to the improvement of standardized test design, reduction of cultural biases, and provision of strategies for more accurate assessments at both national and international levels.

Keywords: Response styles, TIMSS exams, item parameters, individual abilities