Journal of New Approach to Children's Education (JNACE) **Original Article**

Summer-2024, Volume 6, No 2 **DOI: 10.22034/NAES.2024.464023.1456**

ISSN: 2716-9766

GrossMark

The Study and Application of Cognitive Diagnostic Models in Estimating the Psychometric Properties of Published 8th Grade Mathematics Questions of Tims 2019

Hamideh Khalilpour [1], Asghar Minaei [*2], Abdolazim Karimi [3]

[1] Department of Management, Faculty of Humanities, Saveh Branch, Islamic Azad University, Saveh, Iran

[2] Department of Psychology, Faculty of Measurement and Measurement, Allameh Tabatabai University, Tehran, Iran

[3] Associate Professor of Education Research Institute

*** Corresponding author:** asgharminaeie@gmail.com

Received: 2024-04-20 **Accepted:** 2024-06-21

Abstract

Purpose: The purpose of the current research was to study and use cognitive diagnostic models in estimating the psychometric properties of the published 8th grade mathematics questions of Tims 2019.

Research method: The research method of this study is quantitative-qualitative, which was carried out with the approach of retrofitting existing tests. The statistical population of this study was all Iranian, Korean and American eighth grade students participating in the 2019 TEAMS study. The size of Iranian students included 1,095,026 students from 23,895 schools, the size of South Korean students included 465,626 students from 3,006 schools, and the size of American students included 4,059,757 students from 48,557 schools. The sample of Iranian, Korean, and American eighth grade students participating in the TEAMS 2019 study is that through a two-stage cluster sampling method, 5980 Iranian eighth grade students from 220 schools with an average age of 14.1, 3861 Korean eighth grade students from 168 schools with an average age of 14.5 and 8698 American eighth grade students from 273 schools with a mean age of 14.2 were reported.

Findings: Using the qualitative method, 16 skills in four areas of content, knowledge, application and reasoning were identified and the Q matrix was compiled. Then, in the quantitative section, the fit of the three Dina, Dino, and G-Dina models was examined with the data obtained from the students' answers to the questions of the math section in blocks 1, 2, 3, and 5, and the G-Dina model had a somewhat better fit.

Keywords: Cognitive diagnostic models, TEAMS 2019, Mathematics, Cross-cultural, Skill profile

© 2019 Journal of New Approach to Children's Education (JNACE)

This work is published under CC BY-NC 4.0 license. **© 2022 The Authors.**

How to Cite This Article: Minaei, A & et al. (2024). The Study and Application of Cognitive Diagnostic Models in Estimating the Psychometric Properties of Published 8th Grade Mathematics Questions of Tims 2019. *JNACE*, 6(2): 239-258.

CrossMark Open Access

فصلنامه رویکردی نو بر آموزش کودکان | مقاله پژوهشی

تابستان ۱۴۰۳، دوره ۶، شماره ۲

DOI: 10.22034/NAES.2024.464023.1456

ISSN: 2716-9766

GrossMark

مطالعه و کاربرد مدل‌های تشخیصی شناختی در برآورد ویژگی‌های روانسنجی سؤال‌های انتشار یافته ریاضی پایه هشتم تیمز ۲۰۱۹

حمیده خلیل پور[۱]، اصغر مینایی[*۲]، عبدالعظیم کریمی[۳]

[۱] گروه مدیریت، دانشکده علوم انسانی، واحد ساوه، دانشگاه آزاد اسلامی، ساوه،ایران

[۲] گروه روانشناسی، دانشکده سنجش و اندازه گیری، دانشگاه علامه طباطبائی، تهران، ایران

[۳] دانشیار پژوهشگاه مطالعات تعلیم و تربیت، تهران، ایران

* نویسنده مسئول: asgharminaeie@gmail.com

تاریخ دریافت مقاله: ۱۴۰۳/۰۲/۰۱ **تاریخ پذیرش مقاله:** ۱۴۰۳/۰۴/۰۱

چکیده

هدف: هدف از پژوهش حاضر مطالعه و کاربرد مدل‌های تشخیصی شناختی در برآورد ویژگی‌های روانسنجی سؤال‌های انتشار یافته ریاضی پایه هشتم تیمز ۲۰۱۹ بود.

روش پژوهش: روش تحقیق این مطالعه، کمی-کیفی است که با رویکرد ریتروفیت آزمون‌های موجود انجام شد. جامعه آماری این مطالعه، کلیه دانش‌آموزان ایرانی، کره‌ای و امریکایی پایه هشتم شرکت‌کننده در مطالعه تیمز ۲۰۱۹ بود. حجم جامعه دانش‌آموزان ایرانی شامل ۱۰۹۵۰۲۶ دانش‌آموز از ۲۳۸۹۵ مدرسه، حجم جامعه دانش‌آموزان کره جنوبی شامل ۴۶۵۶۲۶ دانش‌آموز از ۳۰۰۶ مدرسه، و حجم جامعه آمریکایی شامل ۴۰۵۹۷۵۷ دانش‌آموز از ۴۸۵۵۷ مدرسه بود. نمونه دانش‌آموزان پایه هشتم ایرانی، کره‌ای و آمریکایی شرکت‌کننده در مطالعه تیمز ۲۰۱۹ است که از طریق روش نمونه‌گیری خوشه‌ای دو مرحله‌ای، ۵۹۸۰ دانش‌آموز پایه هشتم ایرانی از ۲۲۰ مدرسه با میانگین سنی ۱۴/۱، ۳۸۶۱ دانش‌آموز پایه هشتم کره‌ای از ۱۶۸ مدرسه با میانگین سنی ۱۴/۵ و ۸۶۹۸ دانش‌آموز پایه هشتم آمریکایی از ۲۷۳ مدرسه با میانگین سنی ۱۴/۲ گزارش شده‌است.

یافته‌ها: با استفاده از روش کیفی ۱۶ مهارت در چهار زمینه حیطه محتوایی، دانش، به کاربستن و استدلال کردن شناسایی و ماتریس کیو تدوین شد. سپس در بخش کمی برازش سه مدل دینا، دینو و جی دینا با داده‌های حاصل از پاسخ دانش‌آموزان به سؤالات بخش ریاضی در بلوک‌های ۱، ۲، ۳ و ۵ بررسی شد که مدل جی-دینا تا حدودی دارای برازش بهتری بود. هم‌چنین، در بلوک ۱ براساس روش برآورد بیشینه درستنمایی در همه مهارت‌ها به‌جز مهارت (R5 تعمیم دادن) بیش از ۵۰ درصد از آزمودنی‌ها به‌حد تسلط رسیده‌اند.

واژگان کلیدی: مدل‌های تشخیصی شناختی، تیمز ۲۰۱۹، ریاضی، بین فرهنگی، نیمرخ مهارت

تمامی حقوق نشر برای فصلنامه رویکردی نو بر آموزش کودکان محفوظ است.

شیوه استناد به این مقاله: خلیل پور، حمیده؛ مینایی، اصغر و کریمی، عبدالعظیم (۱۴۰۳) مطالعه و کاربرد مدل‌های تشخیصی شناختی در برآورد ویژگی‌های روانسنجی سؤال‌های انتشار یافته ریاضی پایه هشتم تیمز ۲۰۱۹. *فصلنامه رویکردی نو بر آموزش کودکان* ، ۶(۲): ۲۵۸-۲۳۹.

مقدمه

ریاضیات در زندگی روزمره زیاد مورد استفاده قرار می‌گیرد و می‌توان ریاضیات را زبانی دانست که الگوهای طبیعی و الگوهای یافت شده در ذهن انسان را توصیف می‌کند. این الگوها می‌توانند واقعی یا خیالی، دیداری یا ذهنی، ساکن یا پویا، کمی یا کیفی، صرفاً منفعت طلبانه یا فراتر از علائق تفریحی باشند. آن‌ها می‌توانند از دنیای اطراف ما، از اعماق فضا و زمان یا از عملکردهای درونی ذهن انسان پدیدار شوند. هرکس متناسب با ظرفیت خود فعالیت‌های ریاضی انجام می‌دهد. به عنوان مثال یک جنجگو برای شمارش، کشاورزان با مشاهده تغییرات و مشکلاتی که وجود دارد، تولید خود را افزایش می‌دهد، کسی می‌خواهد با انتخاب راه صحیح به‌موقع به مقصد برسد. با در نظر گرفتن همه امکانات برای رسیدن به مقصد، فعالیت‌های تفکر اعم از محاسبه اعداد یا فکر کردن در مورد تصمیم‌گیری بخشی از تفکر و ریاضی است (Aripin et al, 2019) زندگی عادی در بسیاری از مناطق جهان در قرن بیستم بدون استفاده از ریاضیات بسیار دشوار و غیرممکن است. بنابراین مهارت و توانایی‌های ریاضیات مزایای بسیار مه‌ی برای زندگی بشر هستند (Shadiq, 2007). آموزش ریاضیات، یکی از مهم‌ترین عناصر سازنده آموزش مقدماتی، وظیفه‌ای فراتر از آموزش و ارائه مهارت‌های محاسبه را برعهده دارد و از لحاظ ارتباط با وقایع، استدلال و حل مسئله پشتیبانی مهمی را فراهم می‌کند. تعیین اینکه مهارت‌ها و توانایی‌ها و پیشرفت دانش آموزان از نظر دانش ریاضی به چه میزان است، به ارزیابی دانش آموزان با روش‌های صحیح بستگی دارد (Mullis et al, 2011). در دهه‌های گذشته، متخصصان برای نشان‌دادن موفقیت و برآورد دستاوردهای حوزه‌های شناختی و ارزیابی دانش‌آموزان، از رویکردهای سنتی سنجش مثل نظریه کلاسیک[1] و نظریه سؤال- پاسخ[2] استفاده کرده‌اند.

ظریه کلاسیک آزمون در اوایل قرن بیستم شکل گرفت. این نظریه که به آن «نظریه نمره واقعی» نیز گفته‌می‌شود، بر اساس این فرض عمل می‌کند که تفاوت بین پاسخ‌های آزمون‌ها سیستماتیک است. آنها تحت‌تأثیر توانایی آزمودنی‌ها قرار می‌گیرند. این نظریه توجه خود را فقط بر روی توانایی متمرکز می‌کند، و یکی از بزرگ‌ترین فرضیاتی که اغلب بررسی نتایج را به خود جلب می‌کند این است که سایر منابع تغییر، از جمله عوامل خارجی شرایط پیرامونی یا جسمی و روحی آزمودنی‌ها ثابت است در طول روش تکرار استاندارد، یا فقط یک اتفاق تصادفی و غیر سیستماتیک است. مدل اصلی تئوری، سه مفهوم هستند: نمرات مشاهده‌شده (TO)، که نتیجه نمره واقعی (T) و نمره خطا (E) است (Vincent & Shanmugam, 2020). در پی مهم‌ترین نقص آن یعنی عدم تفکیک‌پذیری ویژگی‌های آزمودنی‌ها و ویژگی‌های سؤال از یکدیگر رویکرد جدیدی در روانسنجی به نام نظریه سؤال پاسخ ایجاد شد.

نظریه سؤال پاسخ شامل مدل‌های دقیق آماری است که برای تجزیه و تحلیل آزمون‌های روانشناختی و پیشرفت تحصیلی کاربرد دارد. در این روش امکان پیش‌بینی عملکرد آزمودنی براساس مجموعه‌ای از عوامل که صفت مکنون یا توانایی نامیده می‌شود، میسر می‌گردد. با این حال این نظریه تنها یک نمره کلی را به‌عنوان نمره توانایی ارائه می‌دهد و قادر به بررسی مهارت‌ها و عملکرد آزمودنی‌ها در جریان آزمون نیست این نظریه مشخص می‌کند که آزمودنی چه چیزی را یاد گرفته است، اما مشخص نمی‌کند که او چه ظرفیتی برای یادگیری دارد (Embertson, 2019).

با در نظر گرفتن ضعف نظریه سؤال پاسخ یک راه‌حل پیشنهادی که توسط Snow & Lohman(1989) ارائه شد، ترکیب اصول روانشناختی با شیوه‌های اندازه‌گیری بود که منجر به پیدایش رویکرد سنجش شناختی تشخیصی[3] شد. سنجش شناختی تشخیصی نوعی از سنجش آموزشی است که برخلاف مدل‌های سنتی سنجش، این توانایی را دارد که ساختارهای دانش و فرآیندهای مهارت را در آزمودنی‌ها اندازه بگیرد و اطلاعاتی راجع به نقاط ضعف و قدرت آنان ارائه دهد (Leighton & Gierl, 2007).

برای ارزیابی نقاط ضعف و قدرت دانش‌آموزان، تعیین مهارت تسلط آنها و طراحی دستورالعمل‌ها و مداخلات در فرایند یادگیری، رویکرد شناختی تشخیصی می‌تواند مفید باشد. این رویکرد یک رویکرد جایگزین مناسب برای سؤال پاسخ است و اطلاعات بیشتری را با استفاده از مهارت‌های عالی چندگانه در فرآیند حل مسئله فراهم می‌کند. رویکرد شناختی تشخیصی یک مدل متغیر پنهان است که در آن مهارت‌ها به‌عنوان ویژگی تعریف می‌شوند و توسط یک مسیر دو دویی α برای ارزیابی تسلط دانش‌آموزان و عدم مهارت آنها نشان داده‌می‌شود (Evran, 2019).

در سنجش تشخیصی شناختی، با توجه به تسلط یا عدم تسلط فرد بر مهارت‌ها و صفاتی که برای پاسخ‌گویی درست به سؤال مورد نیاز هستند احتمال پاسخ فرد به سؤال برآورد می‌شود. در این مدل می‌توان براساس وضعیت تسلط افراد در هر یک از مهارت‌ها و صفات زیربنایی مورد نیاز سؤال‌ها، آنها را به دو گروه مسلط و غیرمسلط و در صورت لزوم به سه گروه مسلط، غیرمسلط و نامعین تقسیم‌بندی نمود و از این طریق مشخص کرد که افراد یا دانش‌آموزان در کدام یک از مهارت‌ها و صفات زیربنایی دارای مشکل هستند. با اجرای این سنجش و شناسایی

افراد ضعیف می‌توان برای آنها برنامه‌های آموزشی جبرانی برگزار نمود و یا با انجام اصلاحات در کتاب‌های درسی و آموزشی، بر آن دسته از مفاهیم، مهارت‌ها و موضوعاتی که اکثریت دانش‌آموزان در آن‌ها دارای مشکل هستند تأکید بیش تری نمود (مینایی و همکاران، ۲۰۱۴).

از عوامل دیگری که باید در بررسی آزمون‌ها مورد توجه قرار داد بحث فرهنگ‌های مختلف آزمون شوندگان است. اینکه چرا بسیاری از دانش‌آموزان، از ریاضیات گریزان‌اند؟ چرا تصور ذهنی آنها از ریاضیات چندان لذت بخش نیست؟ و برعکس چرا دانش آموزانی علاقه‌مند به ریاضی هستند و توانایی‌هایی زیادی در ریاضیات دارند؟ از عواملی است که می‌تواند کمک کننده در برنامه‌ریزی‌های آموزشی باشد. از داده‌های مطالعات تیمز تاکنون گزارشات و مقالات علمی زیادی استخراج شده‌است که به روند بهبود عملکرد در دانش‌آموزان ایرانی کمک کننده بوده است. چنان‌چه تغییرات دو سال آخر برگزاری این آزمون نشان دهنده عملکرد بهتر دانش‌آموزان ایرانی بوده است، اما از آن‌جایی که درس ریاضی در اقتصاد، زندگی شخصی افراد جامعه، پیشرفت سیاسی و اجتماعی مؤثر است و از آنجایی که تنها توانایی‌های محاسباتی در عملکرد ریاضیات مؤثر نبوده و مدل‌های تشخیصی شناختی می‌تواند در این مورد کمک کننده باشند و مهارت‌های ادراکی مرتبط را به یادگیرندگان ارائه کنند، از سوی دیگر مقایسه بین کشورهای شرکت‌کننده با فرهنگ‌های مختلف مثل کره جنوبی و ایالات متحده آمریکا می‌تواند با الگو گرفتن از نقاط قوت آنها در جهت بهبود عملکرد و آموزش بهتر کارگشا باشد، بنابراین با توجه به همه نکات مطرح شده و با توجه به خلأ موجود در مقایسه بین کشورهای شرکت‌کننده، در این پژوهش مساله اصلی این‌گونه مطرح شده‌است که آیا مدل‌های تشخیصی شناختی ارائه شده در برآورد ویژگی‌های روانسنجی سؤال‌های انتشار یافته ریاضی پایه هشتم تیمز ۲۰۱۹ برازش دارد؟

۱. مهارت‌های مورد نیاز برای ارائه پاسخ صحیح به سؤالات ریاضی تیمز ۲۰۱۹ کدامند؟

۲. آیا مجموعه مهارت‌های شناختی سؤالات تیمز ۲۰۱۹ قابل تعمیم به عملکرد دانش‌آموزان پایه هشتم ایران است؟

۳. کدام یک از سه مدل DINA، DINO و G-DINA با داده‌ها برازش بهتری دارد؟

۴. نقاط قوت و ضعف دانش‌آموزان در مهارت‌های مورد نیاز برای پاسخ درست به سؤالات ریاضی کدام است؟

روش پژوهش

به‌منظور پاسخ به سؤالات این پژوهش در چارچوب مدل‌سازی تشخیصی شناختی هم از روش کیفی و هم از روش کمی استفاده شد. از روش‌های کیفی شامل مطالعه پیشینه و چارچوب سنجش تیمز، تحلیل تکلیف و بررسی و کدگذاری سؤالات توسط متخصص ریاضی برای تدوین ماتریس کیو به‌عنوان بخش اصلی مدل‌های تشخیصی شناختی استفاده شد. پس از تدوین ماتریس کیو و بررسی روایی آن، میزان برازش مدل به داده‌ها با استفاده از سه مدل پرکاربرد شامل دینا، جی دینا و دینو مورد تحلیل قرار گرفت. جامعه آماری این مطالعه، کلیه دانش آموزان ایرانی، کره‌ای و امریکایی پایه هشتم شرکت‌کننده در مطالعه تیمز ۲۰۱۹ بود. حجم جامعه دانش‌آموزان ایرانی شامل ۱۰۹۵۰۲۶ دانش‌آموز از ۲۳۸۹۵ مدرسه، حجم جامعه دانش آموزان کره جنوبی شامل ۴۶۵۶۲۶ دانش‌آموز از ۳۰۰۶ مدرسه، و حجم جامعه آمریکایی شامل ۴۰۵۹۷۵۷ دانش‌آموز از ۴۸۵۵۷ مدرسه بود. نمونه دانش‌آموزان پایه هشتم ایرانی، کره‌ای و آمریکایی شرکت‌کننده در مطالعه تیمز ۲۰۱۹ است که از طریق روش نمونه‌گیری خوشه‌ای دو مرحله‌ای، ۵۹۸۰ دانش‌آموز پایه هشتم ایرانی از ۲۲۰ مدرسه با میانگین سنی ۱۴/۱، ۳۸۶۱ دانش‌آموز پایه هشتم کره‌ای از ۱۶۸ مدرسه با میانگین سنی ۱۴/۵ و ۸۶۹۸ دانش‌آموز پایه هشتم آمریکایی از ۲۷۳ مدرسه با میانگین سنی ۱۴/۲ گزارش شده‌است در مطالعات تیمز و پرلز، برای انتخاب نمونه از طرح دو مرحله‌ای خوشه‌ای طبقه‌ای با احتمال متناسب با حجم نمونه[۴] استفاده شده‌است. در مرحله‌ای اول، چند مدرسه انتخاب می‌شود و در مرحله دوم از هر مدرسه منتخب، یک یا دو کلاس به‌صورت تصادفی برای اجرای آزمون برگزیده می‌شوند. در این روش، ابتدا همه مدارس براساس عواملی مثل نوع مدرسه، جنسیت و استان، طبقه‌بندی شده و از میان آنها تعدادی مدرسه انتخاب گردید. در مرحله بعد در هر مدرسه یک یا دو کلاس به‌صورت تصادفی انتخاب‌شده‌اند.

ابزار اندازه‌گیری و ویژگی‌های روانسنجی آن

ابزار مورد استفاده در این مطالعه، سؤالات قابل‌انتشار آزمون ریاضی پایه هشتم مطالعه تیمز ۲۰۱۹ بود. از این سؤالات قابل انتشار، ۶۶ سؤال که در هر سه نمونه ایران، کره جنوبی و آمریکا اجرا شده بود، مورد تحلیل قرار گرفت. ۱۶ سؤال از بلوک یک، ۲۰ سؤال از بلوک دو، ۱۵ سؤال از بلوک ۳ و ۱۵ سؤال از بلوک ۱۵ آزمون ریاضی انتخاب شده‌است. سؤالات مربوط به بلوک ۴ به این دلیل که در دو کشور کره جنوبی و آمریکا اجرا نشده است در مطالعه حاضر مورد بررسی قرار نگرفت. از آن‌جایی که سؤالات مطالعه تیمز توسط افراد متخصص و در سطح بین المللی طراحی شده‌است، روایی و پایایی این آزمون مناسب است.

جدول ۱: سؤالات بررسی شده آزمون ریاضی تیمز ۲۰۱۹

ردیف	سؤال	بلوک	حیطه محتوایی	حیطه موضوعی	حیطه شناختی	نوع سؤال
۱	MP52024	۱	اعداد	کسر و اعشار	دانستن	چندگزینه‌ای
۲	MP52058A	۱	اعداد	عدد صحیح	به کار بستن	پاسخ ساز
۳	MP52058B	۱	اعداد	نرخ، نسبت و درصد	به کاربستن	پاسخ ساز
۴	MP52125	۱	اعداد	اعداد صحیح	استدلال کردن	چندگزینه‌ای
۵	MP52229	۱	اعداد	کسر و اعشار	دانستن	پاسخ‌ساز
۶	MP52063	۱	جبر	عبارت، عملیات و معادله	به کار بستن	چندگزینه‌ای
۷	MP52072	۱	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۸	MP52146A	۱	جبر	رابطه و تابع	استدلال کردن	پاسخ‌ساز
۹	MP52146B	۱	جبر	رابطه و تابع	استدلال کردن	پاسخ‌ساز
۱۰	MP52092	۱	جبر	رابطه و تابع	به کاربستن	چندگزینه‌ای
۱۱	MP52046	۱	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	چندگزینه‌ای
۱۲	MP52083	۱	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	چند گزینه‌ای
۱۳	MP52082	۱	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	چندگزینه‌ای
۱۴	MP52161	۱	داده و احتمال	احتمال	به کار بستن	چندگزینه‌ای
۱۵	MP52418A	۱	داده و احتمال	داده	به کار بستن	چندگزینه‌ای
۱۶	MP52418B	۱	داده و احتمال	داده	به کار بستن	چندگزینه‌ای
۱	MP02-01A	۲	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۲	MP02-01B	۲	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۳	MP02-01C	۲	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۴	MP02-01D	۲	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۵	MP02-01E	۲	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۶	MP02-02	۲	اعداد	کسر و اعشار	به کاربستن	چند گزینه‌ای
۷	MP02-03	۲	اعداد	کسر و اعشار	استدلال کردن	پاسخ ساز
۸	MP02-04	۲	اعداد	نرخ، نسبت و اعشار	دانستن	پاسخ ساز
۹	MP02-05	۲	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۱۰	MP02-06	۲	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۱۱	MP02-07	۲	جبر	عبارت، عملیات و معادله	به کار بستن	پاسخ ساز
۱۲	MP02-08	۲	جبر	عبارت، عملیات و معادله	به کار بستن	پاسخ ساز
۱۳	MP02-09	۲	جبر	رابطه و تابع	به کار بستن	چندگزینه‌ای
۱۴	MP02-10	۲	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	پاسخ ساز
۱۵	MP02-11	۲	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	پاسخ ساز
۱۶	MP02-12A	۲	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	پاسخ ساز
۱۷	MP02-12B	۲	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	پاسخ ساز
۱۸	MP02-13	۲	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	پاسخ ساز
۱۹	MP02-14	۲	داده و احتمال	داده	دانستن	پاسخ ساز
۲۰	MP02-15	۲	داده و احتمال	داده	به کار بستن	پاسخ ساز
۱	MP03-01	۳	اعداد	عدد صحیح	دانستن	چند گزینه‌ای
۲	MP03-02	۳	اعداد	عدد صحیح	به کار بستن	پاسخ‌ساز
۳	MP03-03	۳	اعداد	کسر و اعشار	دانستن	چندگزینه‌ای
۴	MP03-04	۳	اعداد	نرخ، نسبت و درصد	به کار بستن	پاسخ‌ساز

ردیف	سؤال	بلوک	حیطه محتوایی	حیطه موضوعی	حیطه شناختی	نوع سؤال
۵	MP03-05	۳	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۶	MP03-06	۳	جبر	رابطه و تابع	دانستن	چندگزینه‌ای
۷	MP03-07	۳	جبر	عبارت، عملیات و معادله	به کار بستن	چندگزینه‌ای
۸	MP03-08	۳	جبر	رابطه و تابع	استدلال کردن	پاسخ‌ساز
۹	MP03-09	۳	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	چند گزینه‌ای
۱۰	MP03-10	۳	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	پاسخ‌ساز
۱۱	MP03-11	۳	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	چندگزینه‌ای
۱۲	MP03-12	۳	هندسه	اشکال هندسی و اندازه‌گیری	استدلال کردن	پاسخ ساز
۱۳	MP03-13	۳	داده و احتمال	داده	استدلال کردن	پاسخ‌ساز
۱۴	MP03-14A	۳	داده و احتمال	احتمال	دانستن	پاسخ‌ساز
۱۵	MP03-14B	۳	داده و احتمال	احتمال	به کار بستن	چندگزینه‌ای
۱	MP05-01	۵	اعداد	عدد صحیح	دانستن	چندگزینه‌ای
۲	MP05-02	۵	اعداد	عدد صحیح	دانستن	چندگزینه‌ای
۳	MP05-03	۵	اعداد	نرخ، نسبت و درصد	به کاربستن	چندگزینه‌ای
۴	MP05-04	۵	اعداد	کسر و اعشار	دانستن	چندگزینه‌ای
۵	MP05-5A	۵	اعداد	عدد صحیح	به کار بستن	بازپاسخ
۶	MP05-05B	۵	اعداد	عدد صحیح	به کار بستن	بازپاسخ
۷	MP05-06	۵	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۸	MP05-07	۵	جبر	عبارت، عملیات و معادله	دانستن	چندگزینه‌ای
۹	MP05-08	۵	جبر	عبارت، عملیات و معادله	دانستن	بازپاسخ
۱۰	MP05-09	۵	جبر	رابطه و تابع	به کار بستن	بازپاسخ
۱۱	MP05-10	۵	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	چندگزینه‌ای
۱۲	MP05-11	۵	هندسه	اشکال هندسی و اندازه‌گیری	به کار بستن	بازپاسخ
۱۳	MP05-12	۵	داده و احتمال	داده	به کار بستن	بازپاسخ
۱۴	MP05-13	۵	داده و احتمال	داده	به کاربستن	بازپاسخ
۱۵	MP05-14	۵	داده و احتمال	داده	دانستن	چندگزینه‌ای

با توجه به جدول ۱، ۶۶ سؤال بررسی شده در این مطالعه در چهار بلوک شامل بلوک‌های یک، دو، سه و پنج، پنج حیطه محتوایی شامل اعداد، جبر، هندسه، داده و احتمال، هشت حیطه موضوعی شامل کسر و اعشار، عدد صحیح، نرخ، نسبت و درصد، عبارت، عملیات و معادله، اشکال هندسی و اندازه‌گیری، داده، احتمال و رابطه و تابع، سه حیطه محتوایی شامل دانستن، به کار بستن و استدلال کردن، و دو نوع سؤال شامل چند گزینه‌ای و پاسخ‌ساز توزیع‌شده‌اند.

شیوه اجرا. از آن‌جایی که در پژوهش حاضر از داده‌های مطالعه تیمز ۲۰۱۹ استفاده شده‌است، جز تحلیل‌های ثانویه محسوب‌می‌شود که در آن نیاز به جمع‌آوری داده نیست. برای دسترسی به داده‌ها به سایت مراجعه شد و داده‌های مربوط به پاسخ دانش‌آموزان پایه هشتم به سؤالات ریاضی برای سه کشور ایران، کره جنوبی و آمریکا دانلود شد. علاوه به‌منظور گردآوری داده برای بخش کیفی نیز ضمن مطالعه پیشینه پژوهش و چارچوب سنجش مطالعه تیمز ۲۰۱۹، خلاصه نتایج این مطالعات در اختیار دو معلم ریاضی قرار داده شد و پس از چند جلسه بررسی مهارت‌های احتمالی مورد نیاز برای پاسخ‌گویی به این سؤالات شناسایی شد و از آن‌ها درخواست شد تا مشخص کنند که برای ارائه پاسخ صحیح به هر یک از سؤالات آزمون ریاضی به کدام یک از مهارت‌ها نیاز است. بدین ترتیب ماتریس کیو اولیه ساخته شد. سپس در بخش کمی میزان برازش مد با داده ها آزمون شد و به سؤالات این پژوهش پاسخ داده شد. در پژوهش حاضر برای بررسی اعتبار درونی درجه‌بندی کننده‌ها از آماره کاپا، برای پاسخ به سؤالات پژوهش از سه مدل تشخیصی شناختی شامل دینا، جی دینا و دینو، از بسته cdm نرم‌افزار RStudio و برای بررسی تفاوت‌ها بین فرهنگی نیمرخ مهارت ریاضی از تحلیل واریانس استفاده شد.

یافته‌ها

برای سؤالات پژوهش از مراحل تحلیل روانسنجی برای بلوک ۱، ۲، ۳ و ۵ اجرا و گزارش شده‌است تمامی تحلیل‌های روانسنجی ارائه شده در این پژوهش از طریق نرم‌افزار RStudio انجام شده‌است.

سؤال اول: مهارت‌های شناختی مورد نیاز برای ارائه پاسخ صحیح به سؤالات ریاضی تیمز ۲۰۱۹ کدامند؟

به‌منظور شناسایی مهارت‌های شناختی برای پاسخ‌گویی به سؤالات قابل‌انتشار ریاضی تیمز ۲۰۱۹ از روش‌های کیفی شامل بررسی چارچوب سنجش تیمز، تحلیل اهداف کتاب ریاضی، مطالعه پژوهش‌های قبلی و نظر متخصص استفاده شد. نتایج حاصل از بخش کیفی مطالعه حاضر، در جداول ۲ و ۳ ارائه شده است.

جدول ۲: مهارت‌های مورد نیاز برای ارئه پاسخ درست به سؤالات ریاضی تیمز ۲۰۱۹

مهارت‌ها با توجه به محتوایی	مهارت‌های دانشی	مهارت‌های به کاربستن	مهارت‌های استدلالی
اشکال هندسی	یادآوری کردن	تعیین کردن	تحلیل کردن
جبر	بازشناسی کردن	نشان‌دادن	ترکیب کردن
داده و احتمال	طبقه بندی کردن	انجام‌دادن	ارزیابی کردن
اعداد	محاسبه کردن		نتیجه گیری کردن
	بازیابی کردن		تعمیم دادن
	اندازه‌گیری کردن		توجیه کردن

همان طور که در جدول ۲ مشاهده می‌شود، ۱۹ مهارت شامل اشکال هندسی، جبر، داده و احتمال، اعداد، یادآوری کردن، بازشناسی کردن، طبقه‌بندی کردن، محاسبه کردن، بازیابی کردن، اندازه‌گیری کردن، تعیین کردن، نشان‌دادن، انجام‌دادن، تحلیل کردن، ترکیب کردن، ارزیابی کردن، نتیجه‌گیری کردن، تعمیم دادن، تایید کردن برای ۶۶ سؤال ریاضی مورد بررسی در این مطالعه شناسایی شده‌است. در جدول ۳ هم ماتریس کیو برای مشخص کردن مهارت‌های مورد نیاز برای پاسخ داده به سؤالات ریاضی ارائه شده‌است.

جدول ۳: ماتریس کیو آزمون ریاضی مطالعه تیمز ۲۰۱۹

کد سؤال	C1	C2	C3	C4	K1	K2	K3	K4	K5	K6	A1	A2	A3	R1	R2	R3	R4	R5	R6	تعدادمهارت
MP01_01	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	۳
MP01-02A	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP01-02B	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP01-03	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	۲
MP01-04	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	۲
MP01-05	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۴
MP01-06	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP01-07A	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	۵
MP01-07B	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	۳
MP01-08	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۴
MP01-09	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	۴
MP01-10	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	۴
MP01-11	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	۳
MP01-12	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۲
MP01-13A	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۲
MP01-13B	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۲

کد سؤال	C1	C2	C3	C4	K1	K2	K3	K4	K5	K6	A1	A2	A3	R1	R2	R3	R4	R5	R6	تعدادمهارت
MP02-01A	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	۲
MP02-01B	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	۲
MP02-01C	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	۲
MP02-01D	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۳
MP02-01E	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۲
MP02-02	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	۲
MP02-03	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	۳
MP02-04	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۳
MP02-05	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP02-06	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP02-07	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۴
MP02-08	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	۵
MP02-09	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	۵
MP02-10	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	۴
MP02-11	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	۴
MP02-12A	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	۳
MP02-12B	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	۳
MP02-13	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۲
MP02-14	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	۱
MP02-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	۳
MP03-01	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۳
MP03-02	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP03-03	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP03-04	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP03-05	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	۶
MP03-06	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	۶
MP03-07	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	۴
MP03-08	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	۵
MP03-09	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	۳
MP03-10	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP03-11	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	۳
MP03-12	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	۵
MP03-13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	۵
MP03-14A	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۱
MP03-14B	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۲
MP05-01	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۲
MP05-02	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	۴
MP05-03	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP05-04	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۳
MP05-5A	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳
MP05-05B	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۳

کد سؤال	C1	C2	C3	C4	K1	K2	K3	K4	K5	K6	A1	A2	A3	R1	R2	R3	R4	R5	R6	تعدادمهارت
MP05-06	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP05-07	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP05-08	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	۴
MP05-09	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	۳
MP05-10	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	۳
MP05-11	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	۲
MP05-12	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	۲
MP05-13	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	۱
MP05-14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	۱
تعداد سؤال	۱۳	۱۸	۱۸	۲۳	۱۳	۱۲	۲	15	۳	۱	۲۱	۱۲	۲۴	۱۱	۶	۳	۹	۴	۱	

سؤال دوم: آیا مجموعه مهارت‌ها و صفات شناختی سؤالات تیمز ۲۰۱۹ قابل تعمیم به عملکرد دانش آموزان پایه هشتم ایران است؟

برای پاسخ به این سؤال از دو دسته شواهد شامل شواهد درونی و بیرونی کمک گرفته شد. شواهد درونی شامل ملاک‌های ارائه شده توسط مدل‌های تشخیصی شناختی جهت ارزیابی مدل است، اما شواهد بیرونی شامل نتایجی است که از روش‌های دیگر به‌دست آمده و با نتایج حاصل از مدل‌های تشخیصی شناختی مورد مقایسه قرار گرفته و به‌نوعی ملاک بیرونی جهت ارزیابی مدل محسوب می‌شوند.

شواهد درونی

شاخص دقت[۵] و ثبات[۶] دو شاخصی هستند که یک مدل تشخیصی شناختی به‌عنوان خروجی ارائه کرده و به‌نوعی روایی و پایایی تشخیص آزمودنی‌ها در جایگاه مسلط یا غیرمسلط است. شاخص دقت نشان‌می‌دهد که آیا آزمودنی‌ها به‌درستی به‌عنوان مسلط / غیرمسلط در هر مهارت شناسایی می‌شوند. شاخص ثبات[۷] نیز نشان‌می‌دهد که آیا در اجرای دوباره یا اجراهای موازی آیا آزمودنی‌ها به‌طور باثباتی به‌عنوان مسلط / غیرمسلط در مهارت‌ها تشخیص داده می‌شوند. در جدول زیر شاخص‌های دقت و ثبات براساس دو روش برآورد بیشینه پسین[۸] (MAP) و بیشینه درست نمایی[۹] (MLE) گزارش شده‌است.

جدول ۴: شاخص‌های دقت و ثبات

بلوک	مهارت	MLE		MAP	
		دقت	ثبات	دقت	ثبات
بلوک ۱	C_1	۰/۷۱۱	۰/۷۰۸	۰/۶۹۹	۰/۷۸۱
	C_2	۰/۷۱۰	۰/۷۰۲	۰/۶۳۰	۰/۸۷۲
	C_3	۰/۶۷۲	۰/۶۸۹	۰/۷۰۵	۰/۷۶۳
	C_4	۰/۸۰۰	۰/۷۴۰	۰/۸۰۷	۰/۷۷۴
	K_1	۰/۶۹۶	۰/۶۹۳	۰/۷۱۱	۰/۷۷۲
	K_2	۰/۷۰۳	۰/۶۶۸	۰/۷۰۵	۰/۷۸۳
	K_4	۰/۶۶۶	۰/۷۴۴	۰/۶۷۶	۰/۸۰۵
	A_2	۰/۶۸۴	۰/۶۹۳	۰/۶۷۹	۰/۷۴۹
	A_3	۰/۷۱۸	۰/۷۲۹	۰/۷۷۹	۰/۷۴۴
	R_1	۰/۷۵۶	۰/۷۳۲	۰/۷۱۹	۰/۷۲۴
	R_2	۰/۶۴۱	۰/۸۵۳	۰/۶۴۰	۰/۸۵۴
	R_4	۰/۶۸۱	۰/۸۳۸	۰/۶۹۳	۰/۷۷۳
	R_5	۰/۷۴۵	۰/۹۲۱	۰/۷۳۷	۰/۸۱۳
بلوک ۲	C_1	۰/۷۲۸	۰/۷۵۸	۰/۷۴۲	۰/۷۸۴
	C_2	۰/۷۳۵	۰/۶۸۵	۰/۷۵۷	۰/۷۳۹

بلوک	مهارت	MLE		MAP	
		دقت	ثبات	دقت	ثبات
	C_3	۰/۷۴۳	۰/۷۷۸	۰/۷۵۲	۰/۷۹۷
	C_4	۰/۷۵۸	۰/۷۰۶	۰/۷۲۲	۰/۸۰۰
	K_1	۰/۷۶۹	۰/۷۲۸	۰/۷۹۸	۰/۷۹۸
	K_2	۰/۶۶۹	۰/۶۸۰	۰/۷۰۹	۰/۷۳۰
	K_4	۰/۷۲۵	۰/۷۱۶	۰/۷۱۹	۰/۸۲۴
	A_2	۰/۷۲۲	۰/۶۶۲	۰/۷۴۰	۰/۷۱۰
	A_3	۰/۷۳۲	۰/۷۱۹	۰/۷۵۷	۰/۷۵۵
	R_1	۰/۷۴۶	۰/۸۸۴	۰/۷۵۳	۰/۸۸۷
	R_2	۰/۶۳۵	۰/۸۱۸	۰/۶۸۷	۰/۸۵۲
	R_4	۰/۷۲۰	۰/۸۷۸	۰/۷۲۵	۰/۸۸۰
بلوک ۳	C_1	۰/۶۹۰	۰/۶۷۷	۰/۷۰۲	۰/۶۷۷
	C_2	۰/۷۲۹	۰/۷۰۱	۰/۷۴۷	۰/۷۳۱
	C_3	۰/۷۳۴	۰/۶۹۳	۰/۷۴۵	۰/۷۰۰
	C_4	۰/۷۰۰	۰/۶۵۳	۰/۷۱۱	۰/۶۹۲
	K_1	۰/۷۰۸	۰/۷۲۱	۰/۷۱۵	۰/۷۲۵
	K_2	۰/۷۴۲	۰/۷۳۶	۰/۷۴۵	۰/۷۳۹
	K_3	۰/۶۸۰	۰/۸۸۷	۰/۶۸۰	۰/۸۹۱
	K_4	۰/۶۹۲	۰/۶۶۸	۰/۶۹۹	۰/۶۸۳
	A_2	۰/۶۶۳	۰/۶۶۸	۰/۶۸۰	۰/۶۹۴
	A_3	۰/۷۶۵	۰/۷۱۸	۰/۷۲۹	۰/۷۸۰
	R_1	۰/۵۹۸	۰/۷۱۸	۰/۶۸۴	۰/۸۹۸
	R_2	۰/۵۹۴	۰/۷۵۸	۰/۶۲۶	۰/۶۵۲
	R_3	۰/۶۱۶	۰/۷۹۹	۰/۶۴۹	۰/۶۹۶
	R_4	۰/۶۱۶	۰/۷۹۹	۰/۶۴۹	۰/۶۹۶
بلوک ۵	C_1	۰/۶۷۱	۰/۶۵۱	۰/۷۰۸	۰/۷۱۰
	C_2	۰/۷۲۴	۰/۷۲۵	۰/۷۱۵	۰/۶۹۷
	C_3	۰/۸۸۳	۰/۸۶۱	۰/۸۳۵	۰/۸۳۳
	C_4	۰/۷۳۶	۰/۶۸۵	۰/۷۳۹	۰/۶۹۳
	K_1	۰/۶۸۷	۰/۷۰۰	۰/۷۱۲	۰/۷۰۸
	K_2	۰/۷۰۰	۰/۷۲۰	۰/۷۱۳	۰/۷۵۰
	K_4	۰/۷۴۴	۰/۷۵۸	۰/۷۴۳	۰/۸۱۱
	K_5	۰/۸۳۸	۰/۸۸۵	۰/۸۴۸	۰/۸۸۹
	K_6	۰/۶۱۰	۰/۷۷۳	۰/۶۳۹	۰/۷۶۳
	A_2	۰/۷۷۱	۰/۷۵۳	۰/۷۳۶	۰/۶۷۹
	A_3	۰/۷۹۹	۰/۷۴۶	۰/۷۹۶	۰/۷۴۹

نقطه برش معینی برای قضاوت درباره شاخص‌های دقت و ثبات وجود ندارد، ولی هرچه ارزش این شاخص‌ها به ۱ نزدیک‌تر باشد، نشان‌دهنده دقت و ثبات بالاتر در تشخیص آزمودنی‌های مسلط و غیرمسلط در مهارت‌ها است. در بررسی مهارت‌های بلوک ۱ براساس برآورد بیشینه درستنمایی، مشاهده می‌شود که بالاترین درجه دقت مربوط به مهارت C_4 (عدد) و بالاترین درجه ثبات مربوط به مهارت R_5 (تعمیم دادن) و کمترین درجه دقت و ثبات به‌ترتیب مربوط به مهارت‌ها R_2 (یکپارچه‌سازی / ترکیب

کردن) و K_2 (بازشناسی) است. همچنین براساس برآورد بیشینه پسین بالاترین دقت و ثبات به‌ترتیب متعلق به مهارت‌های C_4 (عدد) و C_2 (جبر) و کمترین آن مربوط به مهارت‌های C_2 (جبر) و R_1 (تحلیل کردن) است. در بررسی مهارت‌های بلوک ۲، براساس برآورد بیشینه درستنمایی، بالاترین درجه دقت مربوط به مهارت K_1 (یادآوری) و بالاترین درجه ثبات مربوط به مهارت R_1 (تحلیل کردن) و کمترین درجه دقت و ثبات به‌ترتیب مربوط به مهارت‌ها R_2 (یکپارچه‌سازی / ترکیب کردن) و K_2 (بازشناسی) است. همچنین براساس برآورد بیشینه پسین بالاترین دقت و ثبات به‌ترتیب متعلق به مهارت‌های A_3 (انجام دادن) و R_1 (تحلیل کردن) و کمترین آن مربوط به مهارت‌های R_2 (یکپارچه‌سازی / ترکیب کردن) و A_2 (نشان‌دادن / مدل) است. در بلوک ۳ براساس هر دو روش برآورد بالاترین ثبات مربوط به مهارت K_3 (طبقه‌بندی / مرتب کردن) است. همچنین مهارت R_2 (یکپارچه‌سازی / ترکیب کردن) در هر دو روش برآورد کمترین دقت را نشان داد، ضمن اینکه در برآورد با روش بیشینه پسین نیز این مهارت دارای کمترین ثبات است. در بررسی بلوک ۵ نیز مهارت K_5 (بازیابی) در برآورد با روش بیشینه پسین بالاترین دقت و ثبات را در بین سایر مهارت‌ها دارد و همچنین دارای بالاترین ثبات در روش بیشنه درستنمایی نیز است. در حالی که مهارت K_6 (اندازه‌گیری) در هر دو روش برآورد کمترین دقت را در بین مهارت‌های دیگر نشان داد.

شواهد بیرونی

نمره کل مشاهده‌شده و توانایی برآورد شده آزمودنی‌ها (از طریق مدل راش) به‌عنوان دو ملاک بیرونی در نظر گرفته شد و همبستگی این دو ملاک با تعداد مهارت‌هایی که آزمودنی در آن ها به تسلط رسیده است، محاسبه گردید. نتایج این همبستگی در جدول ۵ گزارش شده‌است.

جدول ۵: همبستگی بین نمرات

بلوک	نمره	۱	۲	۳
بلوک ۱	۱. تعداد مهارت‌های تسلط یافته	۱	۰/۸۱۱**	۰/۸۱۱**
	۲. نمره مشاهده‌شده	–	۱	۱**
	۳. توانایی	–	–	۱
بلوک ۲	۱. تعداد مهارت‌های تسلط یافته	۱	۰/۸۸۳**	۰/۸۷۹**
	۲. نمره مشاهده‌شده	–	۱	۰/۹۹۸**
	۳. توانایی	–	–	۱
بلوک ۳	۱. تعداد مهارت‌های تسلط یافته	۱	۰/۷۶۹**	۰/۷۶۹**
	۲. نمره مشاهده‌شده	–	۱	۱**
	۳. توانایی	–	–	۱
بلوک ۵	۱. تعداد مهارت‌های تسلط یافته	۱	۰/۸۸۶**	۰/۸۸۶**
	۲. نمره مشاهده‌شده	–	۱	۱**
	۳. توانایی	–	–	۱

همان طور که در ماتریس همبستگی بالا مشاهده می‌شود، در تمامی بلوک‌ها بین نمره کلی که از تعداد مهارت‌های تسلط یافته آزمودنی‌ها به‌دست آمده است با توانایی برآورد از مدل راش و نمره کل مشاهده‌شده افراد رابطه مستقیم معناداری وجود دارد ($p<۰/۰۵$). بنابراین نتیجه حاصل از مدل تشخیصی شناختی توسط ملاک‌های بیرونی تأیید می‌شود و نتایج حاصل از هر یک از روش‌ها با همدیگر همخوانی و تجانس دارند.

سؤال سوم: کدام مدل تشخیصی شناختی (مقایسه بین مدل DINA، DINO و G-DINA) با داده‌های پژوهش برازش بهتری دارد؟

برای پاسخ به این سؤال سه دسته اطلاعات جمع‌آوری و گزارش شده‌است. ابتدا برازش سؤالات هر بلوک با هر کدام از این ۳ مدل بررسی شد. در گام دوم شاخص‌های برازش مطلق هر مدل مورد بررسی قرار گرفت و در گام سوم سه مدل مدنظر از طریق شاخص‌های برازش نسبی مورد مقایسه قرار گرفتند. گزارش مبسوط مربوط به هر یک از این مراحل در ادامه آمده است.

گام اول: برازش سؤالات با مدل‌ها

به‌منظور بررسی برازش سؤال با مدل از شاخص RMSEA یا ریشه میانگین مجذور خطای تقریب[10] استفاده شد. ارزش‌های مربوط به این شاخص در جدول ۶ گزارش شده‌است.

جدول ۵: شاخص‌های برازش سؤال

بلوک	سؤال	DINA	G-DINA	DINO
بلوک ۱	MP52024	۰/۱۳۷	۰/۱۲۷	۰/۱۴۸
	MP52058A	۰/۰۵۷	۰/۱۴۰	۰/۱۰۲
	MP52058B	۰/۰۷۳	۰/۰۷۹	۰/۱۳۵
	MP52125	۰/۰۳۳	۰/۰۶۳	۰/۰۴۲
	MP52229	۰/۰۴۴	۰/۱۰۴	۰/۰۸۴
	MP52063	۰/۱۵۹	۰/۱۹۵	۰/۱۳۳
	MP52072	۰/۰۵۹	۰/۱۴۳	۰/۰۰۰۱
	MP52146A	۰/۱۲۳	۰/۱۳۷	۰/۰۵۲
	MP52146B	۰/۰۳۱	۰/۰۸۱	۰/۰۱۶
	MP52092	۰/۱۱۷	۰/۱۰۳	۰/۱۲۱
	MP52046	۰/۰۴۶	۰/۱۰۵	۰/۰۹۶
	MP52083	۰/۱۲۲	۰/۱۲۰	۰/۱۱۳
	MP52082	۰/۰۶۵	۰/۰۸۰	۰/۰۲۴
	MP52161	۰/۱۲۲	۰/۱۴۱	۰/۰۹۹
	MP52418A	۰/۰۷۵	۰/۱۵۷	۰/۱۳۷
	MP52418B	۰/۱۲۳	۰/۱۷۲	۰/۱۱۷
بلوک ۲	MP72007A	۰/۰۶۸	۰/۱۹۸	۰/۰۵۷
	MP72007B	۰/۱۱۷	۰/۱۲۰	۰/۱۱۱
	MP72007C	۰/۱۲۲	۰/۱۸۰	۰/۱۱۵
	MP72007D	۰/۰۹۶	۰/۱۷۵	۰/۰۶۸
	MP72007E	۰/۱۰۶	۰/۲۲۴	۰/۱۰۲
	MP72025	۰/۰۸۶	۰/۲۰۷	۰/۰۷۴
	MP72017	۰/۰۴۷	۰/۱۶۰	۰/۱۰۵
	MP72190	۰/۱۲۶	۰/۱۷۴	۰/۱۳۳
	MP72068	۰/۰۸۵	۰/۱۹۰	۰/۰۶۲
	MP72076	۰/۰۸۲	۰/۱۶۹	۰/۰۹۴
	MP72056	۰/۰۷۰	۰/۱۴۹	۰/۰۴۲
	MP72098	۰/۰۳۹	۰/۱۱۷	۰/۱۹۶
	MP72103	۰/۰۸۰	۰/۱۳۶	۰/۱۰۲
	MP72121	۰/۰۷۹	۰/۱۶۵	۰/۰۴۴
	MP72180	۰/۱۳۱	۰/۱۴۹	۰/۱۰۱
	MP72198A	۰/۰۵۰	۰/۰۷۲	۰/۰۵۴
	MP72198B	۰/۰۴۰	۰/۱۰۵	۰/۰۳۴
	MP72227	۰/۰۳۴	۰/۱۲۸	۰/۱۰۹
	MP72170	۰/۰۸۹	۰/۱۷۹	۰/۰۹۷
	MP72209	۰/۰۴۳	۰/۰۹۹	۰/۰۷۹

بلوک	سؤال	DINA	G-DINA	DINO
بلوک ۳	MP62005	۰/۰۹۸	۰/۱۱۷	۰/۰۶۲
	MP62139	۰/۰۸۲	۰/۱۷۱	۰/۰۷۴
	MP62164	۰/۰۶۱	۰/۰۹۴	۰/۰۴۳
	MP62142	۰/۰۷۸	۰/۲۰۹	۰/۰۵۹
	MP62084	۰/۰۵۴	۰/۱۰۰	۰/۰۵۶
	MP62351	۰/۰۲۳	۰/۱۳۳	۰/۰۶۲
	MP62223	۰/۰۵۹	۰/۱۵۱	۰/۰۲۸
	MP62027	۰/۰۴۰	۰/۱۲۸	۰/۰۴۶
	MP62174	۰/۰۸۹	۰/۰۶۷	۰/۰۶۵
	MP62244	۰/۰۵۹	۰/۱۸۸	۰/۰۳۱
	MP62261	۰/۰۷۲	۰/۱۱۶	۰/۰۵۹
	MP62300	۰/۰۴۹	۰/۰۹۱	۰/۰۸۰
	MP62254	۰/۰۵۸	۰/۱۱۴	۰/۰۹۵
	MP62132A	۰/۰۸۷	۰/۱۰۳	۰/۰۳۳
	MP62132B	۰/۰۴۹	۰/۱۹۱	۰/۰۳۶
بلوک ۵	MP52413	۰/۰۹۶	۰/۱۴۴	۰/۰۷۷
	MP52134	۰/۰۷۲	۰/۰۶۴	۰/۰۵۱
	MP52078	۰/۰۹۸	۰/۱۶۲	۰/۰۸۶
	MP52034	۰/۰۹۴	۰/۱۳۱	۰/۰۹۱
	MP52174A	۰/۰۸۵	۰/۱۴۸	۰/۰۷۰
	MP52174B	۰/۰۴۲	۰/۰۵۲	۰/۰۸۶
	MP52130	۰/۰۳۹	۰/۰۸۸	۰/۰۳۸
	MP52073	۰/۰۲۷	۰/۰۹۹	۰/۰۳۵
	MP52110	۰/۰۵۱	۰/۰۷۵	۰/۱۳۷
	MP52105	۰/۰۲۸	۰/۰۳۵	۰/۰۳۴
	MP52407	۰/۰۷۲	۰/۰۷۹	۰/۰۸۰
	MP52036	۰/۰۵۹	۰/۱۱۱	۰/۰۳۳
	MP52502	۰/۰۳۸	۰/۰۸۷	۰/۰۲۷
	MP52117	۰/۰۶۵	۰/۰۹۹	۰/۰۷۸
	MP52426	۰/۰۰۸	۰/۰۷۴	۰/۰۳۱

تفسیر شاخص RMSEA به این صورت است که سؤال‌هایی با میزان شاخص RMSEA بیشتر از ۰/۱ برازش ضعیفی با مدل‌دارند، سؤال‌هایی با میزان شاخص RMSEA بین ۰/۱ تا ۰/۰۵ برازش متوسط و سؤال‌های با میزان شاخص RMSEA کمتر از ۰/۰۵ برازش خوبی با مدل دارند (Kunina-Habenicht, 2009). در بلوک ۱ بیشتر سؤالات برازش مطلوبی با مدل‌ها نشان ندادند، البته وضعیت برازش سؤالات در مدل G-DINA نسبت به دو مدل دیگر نامناسب‌تر بود. در بلوک ۲ نیز برازش سؤالات با مدل G-DINA بسیار بد بود و تنها شاخص RMSEA دو سؤال از ۰/۱ کوچک‌تر شده‌است. اما در بررسی برازش سؤالات این بلوک مشاهده می‌شود که تنها شاخص RMSEA ۴ سؤال در مدل DINA از ۰/۱ بزرگ‌تر شده‌است. در بلوک ۳ مشاهده می‌شود که شاخص RMSEA تمامی سؤالات در هر دو مدل DINA و DINO کوچک‌تر از ۰/۱ است؛ در حالی که در مدل G-DINA تنها ۳ سؤال با شاخص RMSEA کوچک‌تر از ۰/۱ مشاهده می‌شود. در بررسی بلوک ۵ مشاهده می‌شود که در مدل DINA هیچ سؤالی دارای RMSEA بزرگ‌تر از ۰/۱ نیست؛ یعنی هیچ‌یک از سؤالات این

بلوک با مدل DINA برازش ضعیف ندارند. در بررسی مدل G-DINA ۵ سؤال و در مدل DINO یک سؤال برازش ضعیفی با مدل نشان دادند.

گام دوم: برازش مطلق مدل‌ها

شاخص‌های برازش مطلق، صرفاً به بررسی برازش مدل با داده‌ها پرداخته و به مقایسه مدل با سایر مدل‌ها نمی‌پردازد. در این پژوهش از سه شاخص برازش مطلق شامل MADcor، SRMSR و $maxx^2$ که در جدول ۶ گزارش‌شده، استفاده شده‌است.

جدول ۶: شاخص‌های برازش مطلق مدل‌ها

بلوک	مدل	MADcor	SRMSR	$\mathbf{maxx^2}$	p
بلوک ۱	DINA	۰/۰۳۴	۰/۰۴۴	۱۳/۵۷	۰/۰۲۸
	G-DINA	۰/۰۱۸	۰/۰۲۴	۵/۳۳۷	۰/۹۹۹
	DINO	۰/۰۳۸	۰/۰۴۹	۱۵/۲۷	۰/۰۱۱
بلوک ۲	DINA	۰/۰۳۰	۰/۰۴۱	۳۴/۸۲	۰/۰۰۰۱
	G-DINA	۰/۰۲۱	۰/۰۲۷	۵/۶۳	۰/۹۹۹
	DINO	۰/۰۴۲	۰/۰۵۶	۳۲/۷۱	۰/۰۰۰۱
بلوک ۳	DINA	۰/۰۲۶	۰/۰۳۴	۶/۷۱	۰/۹۹۹
	G-DINA	۰/۰۱۶	۰/۰۲۰	۲/۳۷۲	۰/۹۹۹
	DINO	۰/۰۲۹	۰/۰۳۷	۸/۵۹	۰/۳۵۵
بلوک ۵	DINA	۰/۰۲۹	۰/۰۳۹	۹/۰۶	۰/۲۷۴
	G-DINA	۰/۰۱۶	۰/۰۲۱	۲/۹۶۱	۰/۹۹۹
	DINO	۰/۰۴۱	۰/۰۵۲	۱۵/۱۷	۰/۰۱

در بررسی شاخص‌های مطلق برازش در بلوک ۱ شاخص‌های MADcor، SRMSR برای هر سه مدل کوچک‌تر از ۰/۰۵ است که نشان از برازش مطلوب مدل‌ها دارد. در بررسی بیشینه خی–دو نیز سطح معناداری هر سه مدل از ۰/۰۱ بزرگ‌تر شده است. در بررسی بلوک ۲ تنها شاخص SRMSR مربوط به مدل DINO بزرگ‌تر از ۰/۰۵ شده‌است. ولی براساس بیشینه خی–دو تنها مدل G-DINA برازش مطلوبی دارد، زیرا سطح معناداری آن از ۰/۰۵ بزرگ‌تر است. در بررسی بلوک ۳ شاخص‌های برازش مطلق برای هر سه مدل مطلوب بوده و برازش این مدل ها با داده‌ها تأیید می‌شود. در بلوک ۵، مشاهده می‌شود که شاخص‌های MADcor، SRMSR برای هر دو مدل DINA و G-DINA کوچک‌تر از ۰/۰۵ است که نشان از برازش مطلوب این دو مدل است؛ اما در مدل DINOشاخص SRMSR برابر با ۰/۰۵۲ است و چون میزان آن از ۰/۰۵ بزرگ‌تر است، نشان از برازش متوسط این مدل با داده‌ها دارد. این نتایج به‌وسیله معناداری شاخص بیشینه خی– دو نیز تأیید شده‌است، زیرا بیشینه خی– دو برای این دو مدل معنادار نیست و برای مدل DINO معنادار شده‌است؛ یعنی مدل DINO با داده‌ها متجانس نیست.

گام سوم: برازش نسبی مدل‌ها

در بررسی برازش نسبی مدل‌ها، راجع به هر مدل در قیاس با سایر مدل‌ها تصمیم‌گیری می‌شود. شاخص‌های برازش نسبی شامل لگاریتم درست نمایی، شاخص اطلاعات آکائیک[11] (AIC) و شاخص اطلاعات بیزی[12] (BIC) در جدول زیر گزارش شده است. علاوه بر این از آزمون LR که یک آزمون آماری است که برای مقایسه شاخص‌های برازش بین دو مدل که یکی از آن‌ها در دیگری آشیانه[13] می‌کند، استفاده شد. بنابراین براساس این شاخص تنها امکان مقایسه مدل G-DINA با دو مدل دیگر وجود دارد؛ زیرا دو مدل DINO و DINA در مدل G-DINA آشیانه می‌کنند.

جدول ۷: شاخص‌های برازش نسبی در مقایسه مدل‌ها

بلوک	مدل	-2Log like	AIC	BIC	Chisq	df	p
بلوک ۱	DINA	۱۵۱۱۸/۸۹	۳۱۶۳۶/۸۹	۷۰۷۷۲/۲۵	۳۹۰/۵۶	۷۹۴۳	۰/۹۹۹
	DINO	۱۵۲۴۳/۱۷	۳۱۶۸۹/۱۷	۷۰۸۲۴/۵۲	۴۴۲/۸۴	۷۹۴۳	۰/۹۹۹
	G-DINA	۱۴۸۰۰/۳۳	۱۵۳۶۰/۳۳	۱۶۶۹۲/۹۲	–	–	–
بلوک ۲	DINA	۱۷۱۲۴/۰۷	۲۵۳۹۴/۰۷	۴۵۰۰۵/۸۸	۴۷۸/۲۵	۳۸۲۸	۰/۹۹۹
	DINO	۱۷۲۱۸/۷۴	۲۵۴۸۸/۷۴	۴۵۱۰۰/۵۵	۵۷۲/۹۲	۳۸۲۸	۰/۹۹۹
	G-DINA	۱۶۶۴۵/۸۲	۱۷۲۵۹/۸۲	۱۸۷۱۵/۸۸	–	–	–
بلوک ۳	DINA	۱۲۹۵۰/۳۱	۴۵۷۷۶/۳۱	۱۲۳۵۸۲/۴۵	۴۱۹/۷۵	۱۶۰۳۹	۰/۹۹۹
	DINO	۱۳۰۱۶/۴۲	۴۵۸۴۲/۴۲	۱۲۳۶۴۸/۵۶	۴۸۵/۸۶	۱۶۰۳۹	۰/۹۹۹
	G-DINA	۱۲۵۳۰/۵۶	۱۳۲۷۸/۵۶	۱۵۰۵۱/۵۱	–	–	–
بلوک ۵	DINA	۱۲۵۲۶/۰۲	۱۶۶۸۰/۰۲	۲۶۵۴۸/۰۵	۳۴۱/۲۶	۱۷۱۰	۰/۹۹۹
	DINO	۱۲۶۳۱/۶۸	۱۶۷۸۵/۶۸	۲۶۶۵۳/۷۲	۴۴۶/۹۲	۱۷۱۰	۰/۹۹۹
	G-DINA	۱۲۱۸۴/۷۶	۱۲۹۱۸/۷۶	۱۴۶۶۲/۴۲	–	–	–

در بررسی تمامی بلوک‌ها مشاهده می‌شود که هر سه شاخص -2LL، AIC و BIC برای مدل G-DINA میزان کوچکتری را نشان می‌دهند که این نتیجه گویای برازش بهتر این مدل نسبت به دو مدل دیگر است. اما در مقایسه مدل G-DINA از طریق آزمون LR مشاهده می‌شود که این مدل تفاوت معناداری با دو مدل دیگر ندارد ($p>۰/۰۵$). لذا با درنظر گرفتن عدم معناداری تفاوت بین مدل G-DINA با سایر مدل‌ها و با توجه به برازش مطلوب‌تر سؤالات با مدل DINA برای تحلیل سؤالات از مدل DINA استفاده شد.

سؤال چهارم: نقاط قوت و ضعف دانش‌آموزان در مهارت‌های ریاضی پایه هشتم کدام است؟

نیمرخ مهارتی آزمودنی‌ها با دو روش بیشینه پسین (MAP) و بیشینه درستنمایی (MLE) برآورد گردید. روش MLE به‌دنبال یافتن یک سطح از توانایی (θ) است که الگوی پاسخ آزمودنی را به حداکثر برساند. در این روش کلیه اندازه‌های محتمل θ محاسبه و سپس اندازه‌ای که احتمال درست نمایی را بیشینه می کند را انتخاب می‌گردد. روش بیشینه پسین (MAP) که یک روش برآورد بیزی[۱۴] است از توزیع‌های پیشین در برآورد سطح توانایی آزمودنی استفاده می‌کند. در این روش اطلاعات پیشین درباره پارامترها با تابع لگاریتم درستنمایی براساس بیشینه کردن توزیع پسین برای برآورد سطح توانایی افراد استفاده می‌شود. توزیع پیشین یک توزیع احتمالی است که فرض می‌شود، آزمودنی‌های یک نمونه تصادفی از این توزیع هستند. نتایج این دو روش برآورد در جدول ۸ گزارش شده‌است.

جدول ۸: فراوانی و درصد فراوانی افراد مسلط و غیرمسلط براساس برآورد با روش بیشینه درستنمایی و بیشینه پسین

بلوک	مهارت	MLE		MAP	
		مسلط	غیرمسلط	مسلط	غیرمسلط
بلوک ۱	C_1	۶۷۶(۷۸٪)	۱۸۶(۲۲٪)	۷۵۳(٪۸۷)	۱۰۹(٪۱۳)
	C_2	۶۵۲(۷۶٪)	۲۱۰(۲۴٪)	۷۹۱(٪۹۲)	۷۱(٪۸)
	C_3	۴۷۵(۵۵٪)	۳۸۷(۴۵٪)	۵۴۴(٪۶۳)	۳۱۸(٪۳۷)
	C_4	۵۴۶(۶۳٪)	۳۱۶(۳۷٪)	۵۹۱(٪۶۹)	۲۷۱(٪۳۱)
	K_1	۵۴۲(۶۳٪)	۳۲۰(۳۷٪)	۵۳۲(٪۶۲)	۳۳۰(٪۳۸)
	K_2	۶۳۸(۷۴٪)	۲۲۴(۲۶٪)	۷۲۵(٪۸۴)	۱۳۷(٪۱۶)
	K_4	۵۷۳(۶۶٪/۵)	۲۸۹(۳۳٪/۵)	۵۱۸(٪۶۰)	۳۴۴(٪۴۰)
	A_2	۶۲۵(۷۲٪/۵)	۲۳۷(۲۷٪/۵)	۷۰۲(٪۸۱)	۱۶۰(٪۱۹)
	A_3	۶۵۶(۷۶٪)	۲۰۶(۲۴٪)	۵۸۵(٪۶۸)	۲۷۷(٪۳۲)
	R_1	۵۲۵(۶۱٪)	۳۳۷(۳۹٪)	۶۰۷(٪۷۰)	۲۵۵(٪۳۰)

	R_2	۴۷۷(۵۵٪)	۳۸۵(۴۵٪)	۴۲۷(٪۴۹/۵)	۴۳۵(٪۵۰/۵)
	R_4	۶۰۶(۷۰٪)	۲۵۶(۳۰٪)	۵۹۷(٪۶۹)	۲۶۵(٪۳۱)
	R_5	۳۵۶(۴۱٪)	۵۰۶(۵۹٪)	۳۷۹(٪۴۴)	۴۸۳(٪۵۶)
بلوک ۲	C_1	۶۵۰(۷۷٪)	۱۹۸(۲۳٪)	۶۱۲(٪۷۲)	۲۳۶(٪۲۸)
	C_2	۵۰۷(۶۰٪)	۳۴۱(۴۰٪)	۵۱۸(٪۶۱)	۳۳۰(٪۳۹)
	C_3	۴۱۶(۴۹٪)	۴۳۲(۵۱٪)	۴۴۸(٪۵۳)	۴۰۰(٪۴۷)
	C_4	۵۷۱(۶۷٪)	۲۷۷(۳۳٪)	۶۸۴(٪۸۱)	۱۶۴(٪۱۹)
	K_1	۵۶۸(۶۷٪)	۲۸۰(۳۳٪)	۴۹۸(٪۵۹)	۳۵۰(٪۴۱)
	K_2	۴۵۸(۵۴٪)	۳۹۰(۴۶٪)	۵۰۹(٪۶۰)	۳۳۹(٪۴۰)
	K_4	۶۷۵(۸۰٪)	۱۷۳(۲۰٪)	۷۳۷(٪۸۷)	۱۱۱(٪۱۳)
	A_2	۵۴۸(۶۵٪)	۳۰۰(۳۵٪)	۵۶۷(٪۶۷)	۲۸۱(٪۳۳)
	A_3	۴۹۷(۵۹٪)	۳۵۱(۴۱٪)	۵۸۷(٪۶۹)	۲۶۱(٪۳۱)
	R_1	۴۹۲(۵۸٪)	۳۵۶(۴۲٪)	۵۰۱(٪۵۹)	۳۴۷(٪۴۱)
	R_2	۴۵۵(۵۴٪)	۳۹۳(۴۶٪)	۵۰۲(٪۵۹)	۳۴۶(٪۴۱)
	R_4	۵۳۹(۶۴٪)	۳۰۹(۳۶٪)	۵۱۰(٪۶۰)	۳۳۸(٪۴۰)
بلوک ۳	C_1	۴۹۴(۵۸٪)	۳۵۲(۴۲٪)	۵۳۴(٪۶۳)	۳۱۲(٪۳۷)
	C_2	۶۰۲(۷۱٪)	۲۴۴(۲۹٪)	۶۱۶(٪۷۳)	۲۳۰(٪۲۷)
	C_3	۴۸۲(۵۷٪)	۳۶۴(۴۳٪)	۵۷۴(٪۶۸)	۲۷۲(٪۳۲)
	C_4	۵۰۲(۵۹٪)	۳۴۴(۴۱٪)	۵۷۵(٪۶۸)	۲۷۱(٪۳۲)
	K_1	۶۱۵(۷۳٪)	۲۳۱(۲۷٪)	۶۱۲(٪۷۲)	۲۳۴(٪۲۷)
	K_2	۶۴۹(۷۷٪)	۱۹۷(۲۳٪)	۶۴۳(٪۷۶)	۲۰۳(٪۲۴)
	K_3	۴۳۳(۵۱٪)	۴۱۳(۴۹٪)	۴۱۹(٪۴۹/۵)	۴۲۷(٪۵۰/۵)
	K_4	۶۱۸(۷۳٪)	۲۲۸(۲۷٪)	۶۲۳(٪۷۴)	۲۲۳(٪۲۶)
	A_2	۵۸۰(۶۹٪)	۲۶۶(۳۱٪)	۵۹۱(٪۷۰)	۲۵۵(٪۳۰)
	A_3	۶۱۶(۷۳٪)	۲۳۰(۲۷٪)	۷۰۳(٪۸۳)	۱۴۳(٪۱۷)
	R_1	۴۶۴(۵۵٪)	۳۸۲(۴۵٪)	۷۹۸(٪۹۴)	۴۸(٪۶)
	R_2	۴۶۳(۵۵٪)	۳۸۳(۴۵٪)	۵۲۶(٪۶۲)	۳۲۰(٪۳۸)
	R_3	۴۷۶(۵۶٪)	۳۷۰(۴۴٪)	۷۳۰(٪۸۶)	۱۱۶(٪۱۴)
	R_4	۶۳۷(۶۳٪)	۳۰۹(۳۶٪)	۶۶۶(٪۷۹)	۱۸۰(٪۲۱)
بلوک ۵	C_1	۵۶۳(۶۶٪)	۲۹۲(۳۴٪)	۵۸۷(٪۶۹)	۲۶۸(٪۳۱)
	C_2	۵۷۲(٪۶۷)	۲۸۳(٪۳۳)	۶۰۶(٪۷۱)	۲۴۹(٪۲۹)
	C_3	۶۴۴(٪۷۵)	۲۱۱(٪۲۵)	۷۲۵(٪۸۵)	۱۳۰(٪۱۵)
	C_4	۴۹۱(٪۵۷)	۳۶۴(٪۴۳)	۵۰۹(٪۵۹/۵)	۳۴۶(٪۴۰/۵)
	K_1	۵۶۷(٪۶۶)	۲۸۸(٪۳۴)	۶۸۸(٪۸۸/۵)	۱۶۷(٪۱۹/۵)
	K_2	۴۸۲(٪۵۶)	۳۷۳(٪۴۴)	۴۶۸(٪۵۵)	۳۸۷(٪۴۵)
	K_4	۷۳۷(٪۸۶)	۱۱۸(٪۱۴)	۷۶۵(٪۸۹/۵)	۹۰(٪۱۰/۵)
	K_5	۶۲۴(٪۷۳)	۲۳۱(٪۲۷)	۶۱۸(٪۷۲)	۲۳۷(٪۲۸)
	K_6	۵۱۰(٪۶۰)	۳۴۵(٪۴۰)	۵۳۰(٪۶۲)	۳۲۵(٪۳۸)
	A_2	۳۷۱(٪۴۳)	۴۸۴(٪۵۷)	۴۹۶(٪۵۸)	۳۵۹(٪۴۲)
	A_3	۵۱۶(٪۶۰)	۳۳۹(٪۴۰)	۵۸۶(٪۶۸/۵)	۲۶۹(٪۳۱/۵)

در بلوک ۱ براساس روش برآورد بیشینه درستنمایی در همه مهارت‌ها به‌جز مهارت R_5 (تعمیم دادن) بیش از ۵۰ درصد از آزمودنی‌ها به‌حد تسلط رسیده‌اند. براساس روش بیشینه پسین ۹۲ درصد از آزمودنی‌ها در مهارت C_2 (جبر) به‌حد تسلط رسیده اند. در بلوک ۲ براساس هر دو روش برآورد بالاترین درصد تسلط متعلق به مهارت K_4 (محاسبه کردن) و کمترین درصد متعلق به مهارت C_3 (داده و احتمال) است. در بلوک ۳ براساس روش برآورد بیشینه درستنمایی در همه مهارت‌ها بیش از ۵۰ درصد افراد به‌حد تسلط رسیده‌اند. در روش بیشینه پسین نیز تنها مهارتی که کمتر از نیمی از افراد در آن تسطل دارند، مهارت K_3 (طبقه‌بندی / مرتب کردن) است (۴۹/۵ درصد). مهارت K_4 (محاسبه کردن) در بلوک ۵ براساس هر دو روش برآورد بیشترین درصد تسلط آزمودنی‌ها را کسب کرده است. در جدول ۹ احتمال تسلط بر هر مهارت در هر بلوک گزارش شده‌است.

جدول ۹: احتمال تسلط بر هر مهارت

مهارت	بلوک ۱	بلوک ۲	بلوک ۳	بلوک ۵
C_1	۰/۶۶	۰/۶۷	۰/۵۴	۰/۶۰
C_2	۰/۵۸	۰/۵۶	۰/۶۸	۰/۶۲
C_3	۰/۵۵	۰/۵۲	۰/۶۲	۰/۷۳
C_4	۰/۶۱	۰/۶۱	۰/۵۱	۰/۴۹
K_1	۰/۵۹	۰/۶۴	۰/۶۲	۰/۶۹
K_2	۰/۶۴	۰/۶۱	۰/۶۷	۰/۵۹
K_3	–	–	۰/۵۲	–
K_4	۰/۶۱	۰/۶۸	۰/۵۷	۰/۶۸
K_5	–	–	–	۰/۷۲
K_6	–	–	–	۰/۶۱
A_2	۰/۶۳	۰/۵۵	۰/۶۰	۰/۴۹
A_3	۰/۵۸	۰/۶۰	۰/۶۳	۰/۵۹
R_1	۰/۵۶	۰/۶۰	۰/۶۸	–
R_2	۰/۵۲	۰/۵۸	۰/۵۸	–
R_3	–	–	۰/۶۳	–
R_4	۰/۶۵	۰/۵۹	۰/۶۳	–
R_5	۰/۵۳	–	–	–

با مقایسه بین احتمال تسلط بر مهارت‌ها، مشاهده می‌شود که تنها در بلوک ۵ احتمال تسلط بر دو مهارت A_2 (نشان‌دادن / مدل) و C_4 (عدد) در بلوک ۵ کمتر از ۵۰ درصد است. همچنین مهارت با بیشترین احتمال تسلط مهارت C_3 (داده و احتمال) در بلوک ۵ است.

-بحث و نتیجه‌گیری

استفاده از مدل‌های تشخیصی شناختی به‌عنوان روشی جدید در سنجش آموزشی در تحلیل داده‌های کلان مقیاس توسط محققین مختلفی مورد تأکید قرار گرفته است. از آن‌جایی که در این مدل‌ها با داشتن رویکردی چندبعدی عملکرد آزمون شوندگان در تک تک مهارت‌های زیربنایی سؤالات آزمون بررسی می‌شود، به متخصصین حوزه آموزش و پروش، سیاست گذاران و معلمان در طراحی مناسب‌تر برنامه درسی، کتاب‌ها و روش‌های تدریس کمک شایانی می‌کند.

برای پاسخ به سؤال اول پژوهش، به‌منظور شناسایی مهارت‌های زیربنایی و طراحی ماتریس کیو با بررسی چارچوب سنجش تیمز ۱۹ (Mullis et al, 2012) مهارت برای ارائه پاسخ صحیح به سؤالات منتشر شده آزمون ریاضی تیمز ۲۰۱۹ در بلوک‌های ۱، ۲، ۳ و ۵ شناسایی شد و سپس ماتریس کیو بر اساس نظر متخصص تدوین شد. مهارت‌های زیربنایی برای ارائه پاسخ صحیح به سؤالات ریاضی عبارتند از:

- **اشکال هندسی**: درک اشکال هندسی و اندازه‌گیری
- **جبر**: درک عبارت، عملیات، معادله و رابطه و تابع

- **داده و احتمال**: خواندن و تفسیر داده از یک یا چند منبع و استفاده از آمار و درک احتمال
- **اعداد**: شناسایی اعداد صحیح، کسر و اعشار، نرخ، نسبت و درصد
- **یادآوری**: به یاد آوردن تعاریف، اصطلاحات، ویژگی‌های اعداد، واحدهای اندازه‌گیری، ویژگی‌های هندسی و نماد
- **بازشناسایی کردن**: شناختن اعداد، عبارات، مقادیر و اشکال
- **طبقه‌بندی کردن**: طبقه‌بندی کردن اعداد، عبارات، کمیت ها و اشکال را بر اساس ویژگی‌های رایج
- **محاسبه کردن**: انجام‌دادن رویه‌های الگوریتمی برای چهار عمل اصلی یا ترکیبی از آن‌ها با اعداد حسابی، کسرها، اعشار و اعداد صحیح و انجام‌دادن عملیات جبری ساده
- **بازیابی کردن**: بازیابی اطلاعات از نمودارها، جداول، متون یا منابع دیگر
- **اندازه‌گیری کردن**: استفاده از ابزار اندازه‌گیری و انتخاب واحدهای اندازه‌گیری مناسب
- **تعیین کردن**: مشخص کردن عملیات، راهبردها و ابزارهای کارآمد / مناسب برای حل مسائل
- **نشان‌دادن / مدل‌سازی**: نمایش داده‌ها در جدول یا نمودار؛ تدوین معادله، نامساوی، اشکال هندسی یا نمودار برای مدل‌سازی مساله و نمایش‌های معادل برای یک رابطه ریاضی معین
- **اجرا کردن**: اجرای راهبردها و عملیات برای حل مسائل مربوط به مفاهیم و رویه‌های ریاضی
- **تحلیل کردن**: تعیین و توصیف روابط بین اعداد، عبارات، کمیت‌ها و اشکال
- **ترکیب کردن**: مرتبط کردن عناصر مختلف دانش، بازنمایی ها و رویه‌های مناسب برای حل مسئله
- **ارزیابی کردن**: بررسی کردن راهبردها و راه‌حل‌های جایگزین حل مسئله
- **نتیجه‌گیری کردن**: استنباط کردن بر اساس اطلاعات و شواهد
- **تعمیم دادن**: بیان روابط با عبارات کلی‌تر و کاربردی‌تر
- **توجیه کردن**: ارائه استدلال‌های ریاضی برای حمایت از یک اراهبرد یا راه‌حل

برای بررسی سؤال دوم پژوهش یعنی قابلیت تعمیم مجموعه مهارت‌ها و صفات شناختی سؤالات تیمز ۲۰۱۹ به عملکرد دانش‌آموزان پایه هشتم ایران از شواهد درونی و بیرونی استفاده شد. شواهد درونی، ملاک‌های ارائه شده توسط مدل‌های تشخیصی شناختی جهت ارزیابی مدل است و شواهد بیرونی، نتایجی است که از روش‌های دیگر به‌دست آمده و مقایسه آن با نتایج حاصل از مدل‌های تشخیصی شناختی است. در شواهد درونی، شاخص‌های دقت و ثبات در تشخیص آزمودنی‌های مسلط و غیرمسلط براساس دو روش برآورد بیشینه پسین (MAP) و بیشینه درست نمایی (MLE) نشان داد که در بررسی مهارت‌های بلوک ۱ براساس برآورد بیشینه درستنمایی، بالاترین درجه دقت مربوط به مهارت C4 (عدد) و بالاترین درجه ثبات مربوط به مهارت R5 (تعمیم دادن) و کمترین درجه دقت و ثبات به‌ترتیب مربوط به مهارت‌ها R2 (یکپارچه‌سازی / ترکیب کردن) و K2 (بازشناسی) است. علاوه بر این برآورد بیشینه پسین بالاترین دقت و ثبات به‌ترتیب مربوط به مهارت‌های C4 (عدد) و C2 (جبر) و کمترین آن مربوط به مهارت‌های C2 (جبر) و R1 (تحلیل کردن) است. در مهارت‌های بلوک ۲، با توجه به برآورد بیشینه درستنمایی، بالاترین درجه دقت مربوط به مهارت K1 (یادآوری) و بالاترین درجه ثبات مربوط به مهارت R1 (تحلیل کردن) و کمترین درجه دقت و ثبات به‌ترتیب مربوط به مهارت‌ها R2 (یکپارچه‌سازی / ترکیب کردن) و K2 (بازشناسی) است. برآورد بیشینه پسین هم نشان داد که بالاترین دقت و ثبات به‌ترتیب متعلق به مهارت‌های A3 (انجام‌دادن) و R1 (تحلیل کردن) و کمترین آن مربوط به مهارت‌های R2 (یکپارچه‌سازی / ترکیب کردن) و A2 (نشان‌دادن / مدل) است. با توجه به هر دو روش برآورد در بلوک ۳، بیش‌ترین ثبات مربوط به مهارت K3 (طبقه‌بندی / مرتب کردن) و مهارت R2 (یکپارچه‌سازی / ترکیب کردن) در هر دو روش برآورد دارای کمترین دقت است. علاوه بر این، در روش برآورد بیشینه پسین نیز این مهارت دارای کمترین ثبات است. در بلوک ۵ نیز مهارت K5 (بازیابی) در برآورد با روش بیشینه پسین بیش‌ترین دقت و ثبات را دارد و دارای بالاترین ثبات در روش بیشنه درستنمایی نیز است ولی مهارت K6 (اندازه‌گیری) دارای کم‌ترین دقت در هر دو روش برآورد است. برای بررسی شواهد بیرونی نیز از نمره کل مشاهده‌شده و توانایی برآورد شده آزمودنی‌ها استفاده شد که همبستگی این دو ملاک با تعداد مهارت‌هایی که آزمودنی در آن ها به تسلط رسیده است، محاسبه شد. نتایج نشان داد که در همه بلوک‌ها بین نمره کلی که از تعداد مهارت‌های تسلط یافته آزمودنی‌ها به‌دست آمده است با توانایی برآورد از مدل راش و نمره کل مشاهده‌شده افراد همبستگی معناداری وجود دارد که نشان دهنده تایید نتیجه مدل تشخیصی شناختی توسط ملاک های بیرونی است.

برای پاسخ به سؤال دوم پژوهش شاخص‌های برازش سه مدل پر کاربرد DINA، DINO و G-DINA مقایسه شد. بدین

منظور سه دسته اطلاعات شامل برازش سؤالات هر بلوک با هر کدام از این ۳ مدل، شاخص‌های برازش مطلق و شاخص‌های برازش نسبی مورد مورد بررسی قرار گرفت. برازش سؤالات با مدل‌ها با استفاده از شاخص RMSEA نشان داد که در بلوک ۱ بیش‌تر سؤالات برازش مطلوبی با مدل‌ها خصوصاً در مدل G-DINA ندارند. در بلوک ۲ نیز برازش سؤالات با مدل G-DINA بسیار بد بود و فقط شاخص RMSEA ۴ سؤال در مدل DINA از ۰/۱ بیش‌تر شد. شاخص RMSEA در بلوک ۳ نشان داد که همه سؤالات در هر دو مدل DINA و DINO کم تر از ۰/۱ است اما در مدل G-DINA فقط ۳ سؤال با شاخص RMSEA دارای شاخص مناسبی هستند. نتایج حاصل از بررسی بلوک ۵ در مدل DINA نشان داد که هیچ سؤالی برازش ضعیف ندارد، در مدل G-DINA ۵ سؤال و در مدل DINO یک سؤال برازش ضعیفی با مدل داشتند. در گام دوم بررسی برازش مدل‌ها از سه شاخص برازش مطلق شامل MADcor، SRMSR و $maxx^2$ استفاده شد. در بلوک ۱ شاخص‌های MADcor، SRMSR برای هر سه مدل نشان دهنده برازش مطلوب مدل‌ها بود و بیشینه خی–دو نیز سطح معناداری هر سه مدل نیز حاکی از مطلوب بودن برازش بود. بررسی بلوک ۲ نشان داد که فقط شاخص SRMSR در مدل DINO مناسب است ولی براساس بیشینه خی–دو تنها مدل G-DINA برازش مطلوبی دارد. در بررسی بلوک ۳ شاخص‌های برازش مطلق برای هر سه مدل مطلوب بود. در بلوک ۵ هم شاخص‌های MADcor، SRMSR برای هر دو مدل DINA و G-DINA برازش مطلوبی را نشان داد. ولی در مدل DINO شاخص SRMSR دارای برازش متوسط این مدل با داده‌ها بود. معناداری شاخص بیشنیه خی– دو نیز مؤید این مطلب است، زیرا بیشینه خی– دو برای این دو مدل غیر معنادار و برای مدل DINO معنادار شده‌است. در گام سوم بررسی برازش، شاخص های برازش نسبی مورد بررسی قرار گرفت. نتایج بررسی برازش نسبی مدل‌ها نشان داد که فقط تنها امکان مقایسه مدل G-DINA با دو مدل دیگر وجود دارد. سه شاخص 2LL-، AIC و BIC برای مدل G-DINA در همه بلوک‌ها حاکی از برازش بهتر این مدل نسبت به دو مدل دیگر است. در حالی که مقایسه مدل G-DINA از طریق آزمون LR نشان داد که این مدل تفاوت معناداری با دو مدل دیگر ندارد. با توجه به نتیجه برازش مدل با داده‌ها در نهایت از مدل DINA برای تحلیل داده‌ها استفاده شد. مطالعات قبلی نیز نشان داده‌اند که در تحلیل‌های تشخیصی شناختی مدل‌های غیر جبرانی مثل مدل دینا محبوب تر هستند. از نظر De La Toure (2009)مدل DINA در مقایسه با سایر مدل‌های تشخیصی شناختی تفسیر پذیری ساده تری دارد.

به‌منظور تعیین نقاط قوت و ضعف دانش‌آموزان در مهارت‌های ریاضی پایه هشتم از نیمرخ مهارتی آزمودنی‌ها با دو روش بیشینه پسین (MAP) و بیشینه درستنمایی (MLE) استفاده شد. در بلوک ۱ با روش برآورد بیشینه درستنمایی در همه مهارت‌ها به‌جز مهارت R5 (تعمیم دادن) بیش از نیمی از آزمودنی‌ها به‌حد تسلط و با روش بیشینه پسین ۹۲ درصد از آزمودنی‌ها در مهارت C2 (جبر) به‌حد تسلط رسیده‌اند. در بلوک ۲ در هر دو روش برآورد بالاترین درصد تسلط مربوط به مهارت K4 (محاسبه کردن) و کمترین میزان تسلط متعلق به مهارت C3 (داده و احتمال) بود. در بلوک ۳ هم روش برآورد بیشینه درستنمایی نشان داد که در همه مهارت‌ها بیش از ۵۰ درصد آزمودنی‌ها به تسلط رسیده‌اند و در روش بیشینه پسین تنها مهارتی که کمتر از نیمی از افراد در آن مسلط شده‌اند، مهارت K3 (طبقه‌بندی / مرتب کردن) بود. در بلوک ۵ نیز در مهارت K4 (محاسبه کردن) با توجه به هر دو روش برآورد بیشترین درصد تسلط آزمودنی‌ها به‌دست آمده است. نتایج حاصل از بررسی این سؤال هم‌راستا با نتایج به‌دست آمده از مطالعه مینایی (۲۰۱۳) در زمینه کاربست مدل‌های تشخیصی شناختی در آزمون ریاضی تیمز ۲۰۰۷ است.

موازین اخلاقی

در این مطالعه اصول اخلاق در پژوهش شامل اخذ رضایت آگاهانه از شرکت کنندگان و حفظ اطلاعات محرمانه آنها رعایت گردیده است.

تشکر و قدردانی

پژوهشگران مراتب قدردانی و تشکر خود را از کلیه شرکت‌کنندگان این پژوهش که با استقبال و بردباری، در روند استخراج نتایج همکاری نمودند، اعلام می‌دارند.

تعارض منافع

نویسندگان این مطالعه هیچ گونه تعارض منافعی در انجام و نگارش آن ندارند.

واژه نامه	
۱. نظریه کلاسیک	1. Classical Test Theory
۲. نظریه سؤال- پاسخ	2. Item Response Theory
۳. سنجش شناختی تشخیصی	3. cognitive diagnostic assessment

۴. احتمال متناسب با حجم نمونه	4. probability proportional to their size (PPS)
۵. شاخص دقت	5. accuracy
۶. ثبات	6. consistency
۷. شاخص ثبات	7. consistency
۸. بیشینه پسین	8. Maximum A Posteriori
۹. بیشینه درست نمایی	9. Maximum Likelihood Estimate
۱۰. ریشه میانگین مجذور خطای تقریب	10. Root Mean Square Error Approximation
۱۱. شاخص اطلاعات آکائیک	11. Akaike Information Criterion
۱۲. شاخص اطلاعات بیزی	12. Bayesian Information Criterion
۱۳. آشیانه	13. nested
۱۴. بیزی	14. Bayesian
۱۵.	

فهرست منابع

Aripin, U., Setiawan, W., & Hendriana, H. (2019). Critical Thinking Profile of Mathematics in Integral Materials. Journal of Educational Experts (JEE), 2(2), 97-106.

Embertson, S. rice a. 2019. New psychometric theories for psychologists. Translated by Hasan Pashashreifi and colleagues. Tehran: Rushd Publications.

Evran, D. (2019). An application of cognitive diagnosis modeling in TIMSS: A Comparison of intuitive definitions of Q-Matrices. International Journal of Modern Education Studies, 3(1), 4-17.

Leighton, J., & Gierl, M. (Eds.). (2007). Cognitive diagnostic assessment for education: Theory and applications. Cambridge University Press.

Minaei, Asghar. Delaware, Ali. Philosopher Nezad, Mohammad Reza. Kiamanesh, Alireza and Mohajer, Yahya. (2013). Cognitive diagnostic modeling of TEAMS 2007 mathematics questions in Iran's eighth grade students using an integrated model with re-parameterization and comparing the mathematical skills of male and female students. Educational Measurement Quarterly. Year 4. Number 16.

Minaei, Asghar., Ghaffari, Zahra. (1394). The differential function of the eighth grade questions of the 2011 Tims math test among male and female students using the IRT question-answer theory approach, Educational Measurement Quarterly 6 (21).

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). TIMSS 2011 International Results in Mathematics. Chestnut Hill, MA: IEA.

Shadiq, F. (2007). Apa dan mengapa matematika begitu penting. Departemen Pendidikan Nasional Direktorat Jenderal Peningkatan Mutu Pendidik Dan Tenaga Kependidikan Pusat Pengembangan Dan Pemberdayaan Pendidik Dan Tenaga Kependidikan (PPPPTK) Matematika, 1-10.

Vincent, W., & Shanmugam, S. K. S. (2020). The Role of Classical Test Theory to Determine the Quality of Classroom Teaching Test Items. Pedagogia: Jurnal Pendidikan, 9(1), 5-34.