Educational Measurement
https://jem.atu.ac.ir/

ALLAMEH TABATABA'I UNIVERSITY PRESS

17(65), 2026, 37-69
eISSN: 2470-6240; ISSN: 2252-004x

Article Type: Research Article
...Received in revised form March 16, 2025
....Published Online September 28, 2026
Received December 29, 2024
Accepted July 15, 2025

# Application of Response Time Models in Measuring Ability of the Children's Quantitative Reasoning

**Ghasem Keshavarz Gerami** [1]✉, **Mohammad Reza Flasafi Nejad** [2], **Jalil Younesi** [3], **Ali Delavar** [4], and **Zahra Rezaei Ghahroodi** [5]

1. Department of Measurement and Assessment, Allameh Tabataba'i University, Tehran, Iran. E-mail: gh.kz.59@gmail.com
2. *Corresponding Author*, Department of Measurement and Assessment, Allameh Tabataba'i University, Tehran, Iran .E-mail: falsafinejad@yahoo.co.uk
3. Department of Measurement and Assessment, Allameh Tabataba'i University, Tehran, Iran. E-mail: jalilyounesi@gmail.com
4. Department of Measurement and Assessment, Allameh Tabataba'i University, Tehran, Iran. E-mail: delavarali@yahoo.com
5. Department of Statistic, Tehran University, Tehran, Iran. E-mail: gh.kz.59@yahoo.com

## ABSTRACT

There is a long tradition in the ability testing literature that response speed is as much an indicator of ability as the correctness of responses to items. Therefore, this study seeks to investigate the application of response-time models in measuring children's quantitative reasoning ability. To this end, a sample of 502 children from different areas of Tehran was selected through multistage cluster random sampling and evaluated with a quantitative reasoning test. To analyze the collected data, a hierarchical response-time model was used. Item and person parameters were also estimated and interpreted. The modeling results revealed that in addition to the response-accuracy parameters, including the discrimination coefficient, difficulty coefficient, and theta, response-time parameters, including the time discrimination coefficient, time difficulty coefficient, and gamma, were also estimated. Moreover, the estimated parameters of the response-time model were compared with those of the 2PL IRT model in order to identify their main similarities and differences. Findings revealed that there are differences in discrimination coefficients and individuals' theta estimates, which indicate the power of the response-time model. The smallest differences were observed in the difficulty coefficients. Finally, based on the number of parameters, greater dispersion in ability scores, and support from previous research, the study and application of response-time models in measuring children's cognitive abilities were recommended.

**Keywords:** response-time models, measuring cognitive ability, children's quantitative reasoning, item response theory

**Cite this Article:** Keshavarz Gerami, Gh., Falsafi Nejad, M. H., Younesi, J., Delavar, A., & Rezaei Ghahroodi, Z. (2026). Application of Response Time Models in Measuring Ability of the Children's Quantitative Reasoning. *Educational Measurement*, *17*(65), 37-69. https://doi.org/10.22054/jem.2024.72326.3444

© **2026** by The author(s) retain the copyright
**Publisher:** Allameh Tabataba'i University Press
**DOI:** https://doi.org/10.22054/jem.2024.72326.3444

# EXTENDED ABSTRACT

## Introduction

In recent years, time has emerged as a key parameter in modern psychological measurement models, leading to the development of response-time models. These models integrate both speed and accuracy to estimate cognitive abilities by incorporating response time as informative data. Historically, psychologists such as Galton and Cattell emphasized the importance of reaction time in assessing individual differences; however, classical test theory focused solely on response correctness. With advancements in cognitive science, the distinction between reaction time and response time has gained prominence, prompting the adoption of advanced statistical models—such as generalized logistic models and item response theory (IRT) extensions—that explicitly consider both speed and accuracy. These developments, alongside research highlighting the critical role of cognitive abilities in academic success, underscore the significance of response time as an essential element in cognitive assessment and analysis.

## Literature Review

Recent studies highlight the significant role of response-time models in analyzing cognitive processes. De Boeck and Jeon (2019) categorized these models into four distinct types and introduced their application in examining cognitive mechanisms. Kyllonen and Zu (2016), through a comprehensive review of the literature, emphasized the utility of response time in assessing cognitive abilities, discussing dimensions of speed and accuracy, the trade-off between them, and analytical approaches such as item response theory (IRT), as well as cognitive psychology-based models like the Gaussian performance model and diffusion model. Partchev et al. (2013) revisited the longstanding question in psychological assessment regarding the relative contribution of power and speed in test performance. Although the literature acknowledges the informative value of response-time models, further research is still needed to advance both the theoretical and practical applications of these models within psychometrics and cognitive science.

## Methodology

This study employed a descriptive-correlational research design, with an emphasis on psychometric modeling. The statistical population included 11- and 12-year-old boys and girls enrolled in fifth and sixth grades from central

Homepage: https://jem.atu.ac.ir/

districts of Tehran during the 2021 academic year. A multistage cluster random sampling approach was used to select a final sample of 502 students (250 boys and 252 girls). Ethical standards were observed, and the inclusion criteria required participants to have no history of brain injury and no use of analgesics or other medications affecting cognitive processing speed.

Data were collected using a computerized quantitative reasoning test, assessing cognitive abilities through arithmetic comparison tasks involving the four basic operations. The reliability of the test was estimated (internal consistency = 0.816, marginal reliability = 0.856). For data analysis, the study applied the hierarchical response-time model (van der Linden, 2007) using the LNIRT package in R 3.5 and SPSS 25; BILOG 3 was used for IRT-based modeling. This hierarchical framework allows simultaneous modeling of both accuracy and response-time data, facilitating a more nuanced understanding of individual differences in cognitive speed and ability.

## Results

In this study, descriptive statistics of the test items were first reported. The findings showed that some items (including items 23, 35, 37, 44, and 46 to 48) had the lowest average correct response rates and relatively high mean response times. Next, the unidimensionality of the instrument was examined using the multidimensional item response theory (IRT), and the fit indices (RMSEA, CFI, and TLI) supported the unidimensional structure.

Subsequently, a hierarchical response-time modeling approach was applied, and its estimated parameters were compared with those from the conventional item response theory that does not consider response time. The findings revealed discrepancies in item functioning across the two models. While some items showed strong discrimination in the IRT model, they had low discrimination in the response-time model, and vice versa.

Correlation analyses between the difficulty and discrimination parameters of the two models showed moderate to weak associations in some cases. The results indicated that the two modeling approaches produced significantly different estimates of both item difficulty and discrimination. The response-time model provided estimates of test-takers' abilities that were closer to a normal distribution and allowed for more sensitive classification, resulting in a much higher number of ability strata (502 compared to 35 in the traditional model).

Homepage://jem.atu.ac.ir/

Finally, the comparison of ability ranks across the two models revealed considerable differences. In some cases, test-takers were classified in drastically different ability groups depending on whether response time was considered. For example, one participant ranked 30th in the IRT model but dropped to 478th in the response-time model. Overall, the response-time model demonstrated greater sensitivity in distinguishing among test-takers and emphasized the importance of incorporating response time in test analysis.

## Discussion

The main objective of this study was to investigate the application of hierarchical response-time models in measuring children's cognitive abilities. The results indicated that there were correlations and distinctions between parameters estimated by the response-time model and those from the traditional item-response model. Specifically, some items showed stronger discrimination indices in the model without response time, and a significant relationship was observed between response time and item difficulty. Moreover, spurious correlations between responses and response times—both at the item and respondent levels—may arise. A suggested solution is explicit modeling of the relationship between observed and auxiliary variables. The ability estimates also differed in their distribution and dispersion across the two models, which in some cases led to notable differences in participant ranking. These findings underscore that even simple reasoning processes may be influenced by hidden and complex factors such as processing time, numerical structure, and the type of cognitive strategies employed.

Homepage: https://jem.atu.ac.ir/

## Conclusion

The findings of this study revealed that response-time models can serve as powerful tools for improving the accuracy of cognitive ability assessments. By incorporating response time as a latent parameter in the model, information about mental effort and cognitive activity can be obtained indirectly. Additionally, various applications of these models have been proposed, ranging from evaluating problem-solving strategies and motivation in low-stakes tests to detecting cheating and designing adaptive tests. In conclusion, due to their high capacity in delivering richer and more precise data, the use of response-time models is recommended for cognitive assessment—particularly in children.

### *Conflict of Interest*

The authors declare no conflict of interest.

انتشارات دانشگاه علامه طباطبائی

اندازه‌گیری تربیتی
https://jem.atu.ac.ir/

دوره ۱۷، شماره ۶۵، ۱۴۰۵، ۳۷-۶۹
eISSN: 2470-6240; ISSN: 2252-004x

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۱۰/۰۹؛ تاریخ بازنگری: ۱۴۰۳/۱۲/۲۶
تاریخ پذیرش: ۱۴۰۴/۰۴/۲۷؛ تاریخ انتشار: ۱۴۰۵/۰۷/۰۶

# کاربرد مدل‌های زمان-پاسخ در سنجش توانایی استدلال کمّی کودکان

**قاسم کشاورز گرامی [۱]، محمدرضا فلسفی‌نژاد [۲]✉، جلیل یونسی [۳]، علی دلاور [۴]، و زهرا رضایی قهرودی [۵]**

۱. گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: gh.z.59@gmail.com

۲. *نویسنده مسئول*، گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: falsafinejad@yahoo.co.uk

۳. گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: jalilyounesi@gmail.com

۴. گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: delavarali@yahoo.com

۳. گروه آمار دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران. رایانامه: gh.kz.59@yahoo.com

## چکیده

سنتی قدیمی در ادبیات سنجش توانایی وجود دارد که سرعت پاسخگویی به سؤال همان اندازه نشان‌دهنده توانایی است که درستی پاسخ به سؤالات، است؛ بنابراین، این مطالعه به دنبال بررسی کاربرد مدل زمان-پاسخ در سنجش توانایی شناختی استدلال کمّی کودکان است. برای دستیابی به این هدف نمونه‌ای شامل ۵۰۲ کودک از مناطق مختلف شهر تهران به روش نمونه‌گیری تصادفی خوشه‌ای انتخاب شدند و با آزمون استدلال کمّی مورد ارزیابی قرار گرفتند. داده‌های جمع‌آوری‌شده با استفاده از مدل سلسله‌مراتبی زمان-پاسخ تحلیل شده است. همچنین، پارامترهای سؤال و توانایی فرد نیز برآورد و تفسیر شده است. مدل بندی نشان داد، علاوه بر پارامترهای صحت پاسخ شامل ضریب تمیز، ضریب دشواری و تتا، پارامترهای زمان-پاسخ شامل ضریب تمیز زمان، ضریب دشواری زمان و گاما نیز در مدل برآورد می‌شود. در این مقاله، پارامترهای برآورد شده مدل زمان-پاسخ با مدل سؤال-پاسخ دو پارامتری مقایسه شده است تا شباهت‌ها و تفاوت‌های اصلی آنها شناسایی شود. یافته‌ها نشان داد که تفاوت‌هایی در ضرایب تمیز و توانایی افراد وجود دارد که نشان‌دهندهٔ قدرت مدل زمان-پاسخ است. شایان‌ذکر است که کمترین تفاوت‌ها در ضریب دشواری مشاهده شد. درنهایت بر اساس تعداد پارامترها، افزایش پراکندگی نمرات توانایی و حمایت پیشینه پژوهش، مطالعه و کاربرد مدل‌های زمان-پاسخ در سنجش توانایی‌های شناختی کودکان پیشنهاد شد.

**کلیدواژه‌ها:** مدل‌های زمان-پاسخ، سنجش توانایی شناختی، استدلال کمّی کودکان، مدل سؤال

**استناد به این مقاله:** کشاورز گرامی، قاسم، فلسفی نژاد، محمدرضا، یونسی، جلیل، دلاور، علی، و رضایی قهرودی،زهرا. (۱۴۰۵). کاربرد مدل‌های زمان-پاسخ در سنجش توانایی استدلال کمّی کودکان. *اندازه‌گیری تربیتی*، ۱۷(۶۵)، ۳۷-۶۹. https://doi.org/10.22054/jem.2024.72326.3444

© ۲۰۲۶ نویسنده(گان)

**ناشر:** دانشگاه علامه طباطبائی

## مقدمه

تعریف و مدل بندی زمان در اندازه‌گیری‌های روان‌شناختی به‌عنوان یک پارامتر آگاهی‌دهنده، جدیدترین استقراض مهم روان‌سنجی از فیزیک است. از سال ۱۹۴۶ که استیونز مقابل هیئت فیزیکدانان انگلیسی فرگوسن قرار گرفت و عنوان کرد «اندازه‌گیری در علوم انسانی یعنی اختصاص اعداد به اشیاء و رویدادها بر اساس قوانینی مشخص» (McGrane, 2015)، روان‌سنجی پیشرفت‌های مهم و چشمگیری داشته است که یکی از آنها ورود پارامتر زمان در مدل‌های جدید اندازه‌گیری و شکل‌گیری مدل‌های زمان–پاسخ است.

مدل‌های زمان–پاسخ شاخه‌ای از مدل‌های جدید اندازه‌گیری است که به شکل‌های متفاوت زمان پاسخگویی را در برآورد توانایی و عملکرد مغز مدل‌سازی می‌کند (Fox et al., 2007). از ابتدای اندازه‌گیری‌های روان‌شناختی توجه به زمان پاسخگویی در برآورد توانایی به شکل‌های متفاوت موردتوجه روان‌شناسان بوده است. گاهی زمان پاسخگویی در سنجش توانایی به شکل ضمنی وارد می‌شد، مانند آزمون‌های محدودیت زمانی، گاهی به‌طور صریح در آزمون‌های به‌اصطلاح سرعتی و گاهی به‌عنوان متغیر اصلی و خود توانایی، مانند زمان واکنش یا آزمون‌های پردازش اطلاعات، موردتوجه قرار می‌گرفت. می‌توان گفت پیش‌گامان سنجش و اندازه‌گیری توانایی‌ها، گالتون و اسپیرمن اندازه‌گیری زمان واکنش را به شکل ابتدایی اجرا کرده‌اند (Kyllonen & Zu, 2016). هر پاسخی که توسط آزمودنی ثبت می‌شود حداقل دو نکته مهم همراه خود دارد، اولین نکته، درست یا غلط‌بودن پاسخ است و دومین نکته، مدت‌زمانی است که آزمودنی صرف کرده تا پاسخ را ارائه دهد. نکته دوم در بسیاری از مدل‌های اندازه‌گیری نادیده گرفته می‌شود (van der Linden, 2011).

از مهم‌ترین چالش‌های اندازه‌گیری زمان پاسخ، وجود عوامل متعدد تأثیرگذار بر زمان پاسخ است که تقریباً کنترل تمامی آنها غیرممکن است. به‌عنوان‌مثال، پاسخ کند یک آزمودنی ممکن است منعکس‌کننده سرعت پردازش کند او یا دقت کم او باشد. آزمودنی می‌تواند با یک حدس خوش‌شانسانه سریع و درست جواب دهد، یا می‌تواند کند و درست پاسخ دهد. در برخی موارد وجود انگیزه لازم می‌تواند باعث پاسخ سریع و درست شود. همچنین ندانستن پاسخ، عدم اختصاص

Homepage: https://jem.atu.ac.ir/

زمان کافی برای پردازش، یا سردرگم شدن هنگام پاسخگویی و اتمام آزمون می‌توانند منجر به پاسخ اشتباه شود. اگرچه به‌طورکلی بین سرعت و دقت در پاسخگویی تعادل وجود دارد، اما این تعادل بین سؤالات یا افراد ثابت نیست. ممکن است پاسخ آزمودنی در شرایط تحت‌فشار بودن زمان، دقیق‌تر از زمان نامحدود باشد (Schnipke & Scrams, 2005).

از طرف دیگر زمان پاسخگویی نقش مهم و کلیدی در اندازه‌گیری‌های علوم‌شناختی دارد. در همان ابتدای اندازه‌گیری‌های روان‌شناسی که در آزمایشگاه‌های روان‌شناسی انجام می‌شد، دانشمندانی مانند کتل و گالتون متوجه شدند که سرعت پاسخ افراد نسبت به محرک‌های مختلف، نشانه‌ای از نوعی برتری در عملکرد مغز و توانایی‌های فردی است. درنتیجه، نقطه تلاقی روان‌شناسی تجربی و اندازه‌گیری کلاسیک در آزمایش‌های حسی و مطالعه تفاوت‌های فردی توسط این روان‌شناسان شکل گرفت. یکی از مفاهیمی که در همان سال‌ها در این زمینه شکل گرفت، زمان واکنش بود. گالتون اعتقاد داشت که با زمان واکنش می‌توان کنش‌های ذهنی را اندازه‌گیری کرد و کتل معتقد بود که با این کنش‌های ساده می‌توان با دقت کامل تفاوت‌های ذهنی افراد را اندازه‌گیری کرد و وقت‌گذاشتن برای کنش‌های پیچیده بیهوده است (Anastasi & Urbina, 1997).

این نکته اهمیت دارد که بین زمان واکنش و زمان پاسخ تفاوت قائل شویم. زمان واکنش یک متغیر پاسخ اصلی در مطالعه تفاوت‌های فردی است، ولی زمان پاسخ مفهومی است که زمان را به‌عنوان یک پارامتر یا بُعد در اندازه‌گیری توانایی‌ها مدل‌سازی می‌کند. تمایز سطح سرعت، ابعاد مختلف سرعت و سطوح آن در چارچوب توانایی‌های شناختی، تعادل سرعت -دقت، روش‌های تجزیه‌وتحلیل به‌ویژه نظریه سؤال-پاسخ و دیگر کاربردهای زمان، مباحث ویژه و کاربردی در اندازه‌گیری‌های شناختی هستند. ازجمله پیشرفت‌های مدل‌سازی پردازش اطلاعات شناختی، مدل‌های زمان پاسخ، مدل انتشار و تابع گاوسی هستند (Goldhammer, 2015).

از طرف دیگر پژوهش‌های متعدد بیانگر نقش کلیدی توانایی‌های شناختی در موفقیت تحصیلی کودکان است و در طول دهه اخیر به دلیل نقش بی‌بدیل آن‌ها در موفقیت تحصیلی، توجه فزاینده‌ای به آموزش توانایی‌های شناختی در فرایندهای شناختی، تحصیلی و یادگیری شده است؛ به‌عنوان‌مثال برنینگر و ولف (۲۰۱۶) در مدل خود برای تبیین نواقص شناختی کودکان دارای اختلال یادگیری

Homepage: https://jem.atu.ac.ir/

خاص که علی‌رغم برخورداری از هوش نرمال، دارای نواقص متعددی در حوزه تحصیلی هستند، توانایی‌های شناختی را مهم دانسته و بر آن تأکید کرده‌اند. تعداد زیادی از تحقیقات (Swanson et al., 2009؛ Papanastasiou, 2017) که در زمینه پردازش شناختی صورت گرفته است، بیانگر نواقصی در حوزه شناختی افراد دارای اختلال ریاضی است (Swanson, 2012). از عوامل مؤثر در آموزش و یادگیری، حافظه و میزان توجه کودکان به موضوع درس است. در تحقیقی که توسط دانمن و کارپنتر (cited in Swanson, 2012) انجام شد، افراد شرکت‌کننده در آزمون می‌بایست تعدادی از جملات را می‌خوانند و صحت و عدم صحت آنها را تأیید می‌کردند (ملزومات پردازشی) و هم‌زمان سعی کنند آخرین کلمه از هر جمله را به خاطر بسپارند (ملزومات مربوط به ذخیره‌سازی)؛ این تحقیقات هم‌راستا با هم به این نتیجه رسیدند که کودکان مبتلا به اختلال یادگیری خاص و به‌ویژه اختلال ریاضی نسبت به کسانی که در مهارت‌های تحصیلی، توانمند هستند از ضعف حافظه کاری رنج می‌برند و نمی‌توانند در تکالیفی که مستلزم تلفیق هم‌زمان پردازش و اندوزش است به‌خوبی عمل کنند که به‌نوعی مربوط به پردازش اجرایی مرکزی در زمینه به‌روزرسانی است (Miyake et al., 2000).

برخلاف روان‌شناسی شناختی که به زمان پاسخ توجه دارد، در نظریه قدیم اندازه‌گیری، توجهی به زمان نشده است و جایگاه ویژه‌ای برای زمان پاسخ وجود نداشته است. درواقع تمرکز اصلی بیشتر بر درستی پاسخ معطوف بوده است، همان‌طور که گلدشتاین و همکاران (۲۰۱۵) می‌گویند: «هنوز همان سناریویی مطرح می‌شود که برای اولین بار در سال ۱۹۴۲ عرضه شد: فردی با توانایی x با یک سؤال (آیتم) دشوار و متمایز مواجه می‌شود، و نتیجه آن، یک پاسخ دوتایی (یا شاید چند ارزشی) است.»

یک دلیل نادیده‌گرفتن زمان پاسخ این است که بیشتر مدل‌های «نظریه سؤال-پاسخ» برای آزمون‌های قدرت محض تهیه شده‌اند که متفاوت با آزمون‌های سرعت است. بااین‌وجود، همان‌طور که بسیاری از نویسندگان ذکر کرده‌اند، مفهوم آزمون قدرت محض، یک خطای توهمی در هر موقعیت امتحانی محسوب می‌شود، چون همیشه مقداری فشار زمانی در آن وجود دارد. قابل‌توجه است که این حقیقت ساده، مدت‌های طولانی توسط بیشتر روان‌سنج‌ها نادیده گرفته شده است (Goldstein et al., 2015).

Homepage: https://jem.atu.ac.ir/

برای واردکردن سرعت و دقت در برآورد سطح صفت، مدل‌های متفاوتی پیشنهاد شده است. بخشی از احتمال پاسخ درست، به زمان صرف شده برای پاسخ دادن به سؤال وابسته است. راسکام (۱۹۹۷) و ورهلست و همکاران (۱۹۹۷) مدل‌های مختلفی را مطرح می‌کنند که در آنها زمان صرف شده برای پاسخ دادن به سؤال در پیش‌بینی احتمال موفقیت در آن سؤال تعریف شده است. برای انواع دیگری از این مدل‌ها، متوسط یا مجموع زمان پاسخ‌دهی مهم است. ولی به شکل کلی سه دسته یا رویکرد برای مدل‌های زمان پاسخ با توجه به نوع استفاده از زمان پاسخ در برآورد صفت مکنون مطرح شده است. این سه رویکرد شامل در نظر گرفتن زمان پاسخ به‌عنوان اطلاعات جانبی، زمان پاسخ‌هایی که مشترک و هم‌زمان محاسبه می‌شوند و زمان پاسخی که سلسله‌مراتبی هستند (Tuerlinckx & De Boeck, 2001).

ورهلست و همکاران (۱۹۹۷) مدلی را برای آزمون‌های با محدودیت زمانی ارائه کردند. این مدل بر پایه فرض ارزش افراطی تعمیم‌یافته توزیع یک متغیر پاسخ مکنون استوار است که به زمان صرف شده برای سؤال و توزیع گاما برای آن زمان توجه می‌کند. با استفاده از این واقعیت که ترکیب این دو توزیع، از لوجستیک تعمیم‌یافته پیروی می‌کند (Dubey, 1969)، احتمال پاسخ صحیح به سؤال i توسط فرد j، از رابطهٔ زیر به دست می‌آید:

$$P_{(i)}(\theta_j) = \left[\!\left[ 1 + \exp(\theta_j - \ln \tau_j - b_i) \right]\!\right]^{(-\pi_i)}$$

در رابطه بالا $b_i$ پارامتر ضریب دشوار سؤال i ام، $\theta_j$ پارامتر توانایی فرد j ام، $\tau_j$ پارامتر سرعت فرد j ام و $\pi_i$ پارامتر شکل وابسته به سؤال i ام است. مدل مشابهی توسط راسکام (۱۹۹۷) ارائه شده است. این مدل مانند مدل راش است با این تفاوت که یک پارامتر اضافه که لگاریتم زمان است، به‌عنوان یک رگرسور وارد مدل می‌شود.

$$P_{(i)}(\theta_j) = \left[\!\left[ 1 + \exp(\theta_j + \ln t_{ij} - b_i) \right]\!\right]^{(-1)}$$

## پیشینه پژوهش

دی‌بوک و جئون (۲۰۱۹) با مروری بر مدل‌های زمان-پاسخ در فرایندهای آزمون‌های شناختی، کاربرد مدل‌های زمان-پاسخ را برای بررسی فرایندهای شناختی معرفی کردند. آن‌ها، مدل‌های زمان-پاسخ را به چهار دسته متمایز تقسیم و معرفی کردند. کیلونن و زو (۲۰۱۶) استفاده از زمان

Homepage: https://jem.atu.ac.ir/

پاسخگویی را برای سنجش توانایی شناختی، با مرور برخی از ادبیات در مورد زمان پاسخ معرفی کردند و نشان دادند که این مدل‌ها نقش مهمی در سنجش توانایی شناختی، ارائه یک دیدگاه تاریخی و همچنین پوشش تحقیقات عصر جدید دارند. در این راستا بر تمایز سطح سرعت، ابعاد سرعت و سطح دقت در چارچوب توانایی‌های شناختی، مبادله دقت و صحت، رویکردهای پرداختن به مبادله سرعت-دقت، روش-های تحلیل، به‌ویژه نظریه سؤال-پاسخ، مدل‌های زمان-پاسخ از روان‌شناسی شناختی (عملکرد گاوسی، و مدل انتشار) و سایر کاربردهای زمان پاسخ در آزمون علاوه بر اندازه‌گیری توانایی بحث کرده‌اند. پارتچف و همکاران (۲۰۱۳) با پرداختن به این موضوع که «در آزمون چقدر قدرت و چقدر سرعت اندازه‌گیری می‌شود؟» این مسئله قدیمی در ارزیابی روان‌شناسی را بازنگری و شفاف‌سازی کرده‌اند. اگرچه در ادبیات و پیشینه پژوهش، اشاره شده است که استفاده از مدل‌های زمان-پاسخ می‌تواند آگاهی‌دهنده باشد اما همچنان پیشرفت هر دو علم روان‌سنجی و علوم‌شناختی در تئوری و کاربرد نیازمند پژوهش و پاسخگویی به سؤالات در زمینه کاربرد مدل‌های زمان-پاسخ است.

## روش

به‌منظور بررسی اهداف پژوهش از روش توصیفی از نوع همبستگی با تمرکز بر محاسبات روان‌سنجی استفاده شده است. جامعه آماری پژوهش، کودکان دختر و پسر بخش مرکزی شهر تهران هستند که در سال تحصیلی ۱۴۰۰ دارای سن تقویمی ۱۱ و ۱۲ سال بوده‌اند و در پایه‌های پنجم و ششم ابتدایی مشغول به تحصیل بوده‌اند.

تعیین کمینه مقدار برای اندازه نمونه تجربی در مطالعات روان‌سنجی چالش مهمی است. مسلماً هر چه اندازه نمونه‌های تجربی بیشتر باشد، برآورد پارامترهای آزمون که از این نمونه‌ها حاصل می‌شود پایاتر خواهد بود. برای قابل‌استفاده‌بودن پارامترهایی که درنتیجه اجرای تجربی به دست می‌آیند، به نمونه‌هایی با کمینه حجم ۲۰۰ تا ۳۰۰ نفر برای هر گروه سنی نیاز است. از طرف دیگر برای محاسبات نظریه‌های جدید اندازه‌گیری و مدل‌های نظریه سؤال-پاسخ حجم نمونه بین ۳۵۰ تا ۷۵۰ آزمودنی پیشنهاد شده است (Reise & Yu, 1990). با جمع‌بندی نظری بر اساس پیشنهادهای مطرح شده، داده‌ها از حجم نمونه ۵۰۲ نفر جمع‌آوری و تحلیل شد. در این راستا ۱۹ منطقه آموزش

Homepage: https://jem.atu.ac.ir/

Homepage: https://jem.atu.ac.ir/

پرورش تهران به ۵ ناحیه شمالی، جنوبی، غربی، شرقی و مرکزی تقسیم شد، در ادامه از هر ناحیه یک منطقه انتخاب شد، سپس از هر منطقه یک مدرسه دخترانه و یک مدرسه پسرانه به روش نمونه‌گیری خوشه‌ای چندمرحله‌ای تصادفی انتخاب شد. درمجموع، ۱۰ مدرسه انتخاب و در ادامه با هماهنگی مدیران این مدارس هر تعداد کودک که شرایط ورود به پژوهش را داشتند آزمون شدند. شرایط ورود به گروه نمونه، سن تقویمی ۱۱ و ۱۲ سال، عدم آسیب مغزی، عدم مصرف داروهای مسکن و تأثیرگذار در پردازش و سرعت عملکرد مغز بود. همچنین کدهای اخلاقی در این فرایند رعایت شد. درنهایت گروه نمونه شامل ۲۵۰ پسر و ۲۵۲ دختر شد. در گروه پسران، ۱۳۲ نفر ۱۱ ساله و ۱۱۸ نفر ۱۲ ساله بودند، در گروه دختران ۱۱۹ نفر ۱۱ ساله و ۱۳۳ نفر ۱۲ ساله بودند.

برای جمع‌آوری داده‌ها از ابزار آزمون توانایی استدلال کمّی که مبتنی بر سنجش کامپیوتری است، استفاده شد. این آزمون توانایی‌های شناختی استدلال کمّی و محاسباتی فرد را موردسنجش قرار می‌دهد. در این روش، از فرد خواسته می‌شود که با سرعت و دقت به سؤالات محاسباتی و تشخیص بزرگ‌تر و کوچک‌تری اعداد و محاسبات در حد چهار عمل اصلی (جمع، ضرب، تقسیم و تفریق) پاسخ دهد. اعتبار این آزمون با روش‌های بازآزمایی بالای ۰/۹ گزارش شده است (Deloche et al., 1969). این آزمون در بسیاری از بسته‌های شناختی با نام‌های متفاوت و گرافیک جذاب مانند لموسیتی، پیل، کن‌تب و وِینا وجود دارد. در این آزمون فرد باید عدد بزرگ‌تر را انتخاب کند. درصورتی‌که دو عدد مساوی باشند، بر روی دکمه مساوی کلیک می‌کند. در مراحل بالاتر به‌جای اعداد، محاسبه‌ای ساده به فرد داده می‌شود که فرد باید آنها را حل کند. در این آزمون، فرد باید با حداکثر سرعت ممکن، پاسخ‌ها را ارائه و هر چه سریع‌تر پاسخ دهد، تا نمره بیشتری کسب کند. در سنجش برای این پژوهش از ۴۸ سؤال آزمون استفاده شده است. اعتبار این آزمون با روش همسانی درونی ۰/۸۱۶ و بر اساس برآورد حاشیه‌ای ۰/۸۵۶ برآورد شده است.

برای تجزیه‌وتحلیل مدل سلسله‌مراتبی زمان-پاسخ (van der Linden, 2007) از بستهٔ LNIRT در نرم‌افزار R 3.5 و همچنین از نرم‌افزار SPSS 25 استفاده شده است. برای محاسبات مدل نظریه سؤال-پاسخ نیز از نرم‌افزار BILOG 3 استفاده شده است.

در آزمون‌های رایانه مبنا، به‌دست‌آوردن صحت پاسخ و زمان‌های پاسخ برای هر سؤال آزمون، استاندارد شده است. مدل‌های نظریه سؤال-پاسخ برای اندازه‌گیری یک صفت مکنون (مانند توانایی

و هوش) با استفاده از مشاهدات صحت پاسخ استفاده می‌شوند. ازآنجاکه اطلاعات زمان پاسخ می‌تواند به بهبود عملیات معمول در آزمون (آموزشی) کمک کند و اطلاعاتی در مورد سرعت کار ارائه دهد، در کاربردهای مدرن، استفاده از مدل‌های هم‌زمان برای ادغام اطلاعات زمان-پاسخ در تحلیل آزمون ضروری است. سه روش کلی برای مدل بندی زمان-پاسخ و درستی پاسخ وجود دارد: ۱) اضافه کردن پارامتر زمان به مدل‌های نظریه سؤال-پاسخ (Roskam, 1997)؛ ۲) مدل بندی مجزای زمان-پاسخ از درستی پاسخ (Maris, 1993) و ۳) مدل بندی زمان-پاسخ و درستی پاسخ به‌صورت سلسله‌مراتبی (Van der Linden, 2007).

در مدل سلسله‌مراتبی در اولین مرتبه به‌گونه‌ای فرض می‌شود درستی پاسخ و زمان پاسخ از مدل‌های مجزایی پیروی می‌کنند و هریک از مدل‌ها از مجموعه متفاوتی از پارامترهای سؤال و فرد استفاده می‌کنند. پارامترهای فرد معرف سرعت و درستی (توانایی) پاسخ‌دهنده به سؤال‌ها است. در اولین سطح از مدل بندی، توزیع زمان پاسخ و درستی پاسخ به‌شرط پارامترهای سرعت و درستی از هم مستقل هستند؛ اما در دومین سطح این پارامترها به هم وابسته هستند. این رویکرد منجر به چارچوب مدل بندی سلسله‌مراتبی می‌شود به‌گونه‌ای که ارتباط بین سرعت و درستی در سطح بالاتری از مدل بندی تعریف می‌شود.

چون کران پایین مقادیر زمان پاسخ صفر است، بنابراین از توزیع لگ‌نرمال برای مدل بندی زمان پاسخ‌ها استفاده می‌شود. درواقع یا زمان-پاسخ از توزیع لگ‌نرمال پیروی می‌کند و یا لگاریتم زمان-پاسخ از توزیع نرمال پیروی می‌کند. انتخاب توزیع دوجمله‌ای برای درستی پاسخ و توزیع نرمال برای لگاریتم زمان-پاسخ، منجر به یک روش پارامترسازی تئوری سؤال-پاسخ سنتی می‌شود.

در ادامه مدل سلسله‌مراتبی ون در لیندن (۲۰۰۷) که فاکس و همکاران (۲۰۰۷) مدل بندی توأم در بستهٔ LNIRT در نرم‌افزار R استفاده کرده‌اند، معرفی می‌گردد. همان‌گونه که قبلاً اشاره شد، مدل بندی در دو سطح انجام می‌شود. سطح اول «مرتبه اول مدل بندی» است که در ادامه معرفی می شود.

**مدل بندی درستی پاسخ RA:** ماتریس پاسخ‌ها یک ماتریس N*K است که N تعداد افراد آزمون‌دهنده و k تعداد سؤال‌ها است. الگوهای پاسخ توسط آزمون‌دهنده و سؤال‌ها مشخص می شوند. در یک مدل IRT دو پارامتری، احتمال پاسخ صحیح با معلوم بودن پارامترهای فرد و سؤال

Homepage: https://jem.atu.ac.ir/

تعریف می‌شود (Lord & Novick, 1968). احتمال پاسخ صحیح آزمون‌دهنده i ام به سؤال k به‌صورت زیر تعریف می‌شود:

$$P(Y_{ik}=1|\theta_i,a_k,b_k)=\Phi(a_k\theta_i-b_k)$$

تتا ($\theta_i$)، توانایی آزمون‌دهنده iام، $a_k$ و $b_k$ به‌ترتیب پارامتر ضریب تمییز و ضریب دشواری سؤال kام هستند. $\Phi$ تابع توزیع تجمعی نرمال است. به‌جای تابع توزیع تجمعی نرمال می‌توان از تابع توزیع تجمعی توزیع لوجستیک استفاده کرد. زمانی که سختی سؤال در مقیاس یکسانی مانند مقیاس توانایی تعریف شود، احتمال پاسخ صحیح به سؤال k ام با فرمول زیر محاسبه می‌شود:

$$P(Y_{ik}=1|\theta_i,a_k,b_k)=\Phi(a_k(\theta_i-\tilde{b}_k))$$

در فرمول فوق $\theta_i$ و $\tilde{b}_k$ بر اساس مقیاس مشابه تعریف می‌شوند. در بسته LNIRT هر دو روش پارامترسازی قابل‌اجرا است. لازم به ذکر است که پارامتر ضریب دشواری سؤال در دو معادله فوق به روش مستقیم قابل‌مقایسه با هم نیستند و در مقیاس‌های مختلف تعریف شده‌اند.

برای مدل‌های سه پارامتری، پارامتر حدس که با نماد c_k معرفی می‌شود، و به معنای احتمال حدس درست سؤال k است، به مدل اضافه می‌شود و مدل احتمال موفقیت یا احتمال پاسخ درست به‌صورت زیر است:

$$P(Y_{ik}=1|\theta_i,a_k,b_k,c_k)=c_k+(1+c_k)\Phi(a_k\theta_i-b_k)$$

مدل بندی زمان-پاسخ RT: متغیر پاسخ لگاریتم زمان-پاسخ را در یک ماتریس N×K با نام RT در نظر بگیرید. زمانی که سرعت کار ثابت فرض شود، درواقع هر آزمون‌دهنده با سرعت ثابتی کار می‌کند که آن را با نماد ζ_i نمایش می‌دهیم. زمان موردنیاز برای پاسخ‌دهی به یک سؤال نیز به پارامترهای مشخص هر سؤال بستگی دارد که آنها را با نماد $\Phi_k$ و $\lambda_k$ که به ترتیب معرف پارامتر تشخیص زمان و پارامتر شدت زمان است، نمایش داده می‌شود. لگاریتم زمان-پاسخ با $RT_{ik}$ نمایش داده می‌شود و دارای توزیع نرمال است که به‌صورت زیر تعریف می‌شود:

$$RT_{ik}=\lambda_k-\Phi_k\zeta_i+\epsilon_{ik}$$

$$\epsilon_{ik}\sim N(0,\sigma_{ik}^2)$$

Homepage: https://jem.atu.ac.ir/

به‌طورکلی پارامتر شدت زمان $\lambda_k$ معرف متوسط زمان مورد نیاز برای کامل کردن سؤال (در مقیاس لگاریتم)، پارامتر سرعت $\zeta_i$، معرف سرعت کار کردن آزمون‌دهندهٔ iام و پارامتر تشخیص زمان، $\Phi_k$، معرف اثر ویژه-سؤال سرعت کارکردن روی زمان-پاسخ است. افزایش پارامتر شدت زمان $\lambda_k$ منجر به جابجایی مثبت مکان توزیع زمان سؤال k ام می‌شود. به همین ترتیب، افزایش پارامتر سرعت $\zeta_i$ منجر به جابجایی منفی مکان توزیع زمان سؤال k ام می‌شود. مشابه مدل سؤال-پاسخ در معادله دوم، پارامترهای شدت زمان می‌تواند در مقیاس مشابهی با پارامتر سرعت تعریف شود. در این صورت خواهیم داشت:

$$RT_{ik} = \Phi_k\left(\tilde{\lambda}_k - \zeta_i\right) + \epsilon_{ik}$$

در فرمول فوق پارامتر تشخیص زمان $\Phi_k$ بر حسب $\left(\tilde{\lambda}_k - \zeta_i\right)$ عمل می‌کند. در این مدل، پارامتر $\Phi_k$ به‌عنوان پارامتر شیب سرعت، حساسیت سؤال برای سطوح مختلف سرعت آزمون شوندگان را مدل‌بندی می‌کند. عبارت خطای $\epsilon_{ik}$، تغییرات زمان-پاسخ به دلیل رفتار تصادفی یک آزمون‌دهنده را مدل بندی می‌کند.

سطح دوم «مرتبه دوم و سوم: مدل‌های ساختاری پارامتر فرد و سؤال» است که در ادامه معرفی می‌شود:

پارامترهای افراد: توزیع نرمال دومتغیره برای پارامتر سرعت و توانایی به‌صورت زیر مدل بندی می‌شود:

$$(\Theta_i, \zeta_i) \sim N_2\left(\mu_p - \Sigma_p\right)$$

که در آن $\mu_p$ و $\Sigma_p$ برابر است با:

$$\mu_p = \left(\mu_\theta - \mu_\zeta\right)$$

$$\Sigma_p = \begin{pmatrix} \sigma_\theta^2 & \rho \\ \rho & \sigma_\zeta^2 \end{pmatrix}$$

کوواریانس بین پارامترهای فرد با $\rho$ معرفی می‌شود. مدل دوسطحی سرعت و توانایی، برای معرفی توزیع جامعه آزمون‌دهندگان معرفی می‌شود. با توجه به اینکه از روش بیزی استفاده می‌شود،

Homepage: https://jem.atu.ac.ir/

توزیع پیشین برای ماتریس واریانس-کوواریانس $\Sigma_p$ توزیع ویشارت معکوس با درجه آزادی $\upsilon_p$ و $V_p$ در نظر گرفته می‌شود:

$$\Sigma_p \sim Iw(\upsilon_p, V_p)$$

**پارامترهای سؤال:** برای پارامترهای سؤال، توزیع نرمال چندمتغیره به‌صورت زیر در نظر گرفته می شود:

$$(a_k, b_k, \Phi_k, \lambda_k) \sim N_4(\mu_I - \Sigma_I)$$

در آن $\Sigma_I$ برابر است با:

$$\Sigma_I = \begin{pmatrix} \Sigma_{a,b} & \Sigma_{(a,b)(\theta,\lambda)} \\ \Sigma_{(\theta,\lambda)(a,b)} & \Sigma_{\theta,\lambda} \end{pmatrix} = \begin{bmatrix} \sigma_a^2 & \sigma_{a,b} & \sigma_{a,\theta} & \sigma_{a,\lambda} \\ \sigma_{b,a} & \sigma_b^2 & \sigma_{b.\Phi} & \sigma_{b,\lambda} \\ \sigma_{\Phi,a} & \sigma_{\Phi,a} & \sigma_\Phi^2 & \sigma_{\Phi,\lambda} \\ \sigma_{\lambda,a} & \sigma_{\lambda,b} & \sigma_{\lambda\Phi} & \sigma_\lambda^2 \end{bmatrix}$$

$a_k$ و $\Phi_k$ مثبت هستند. توزیع ابر پیشین با $(\mu_I - \Sigma_I)$، توزیع نرمال-معکوس ویشارت به‌صورت زیر است:

$$\Sigma_I \sim IW_{\upsilon_I}(V_I^{-1})$$

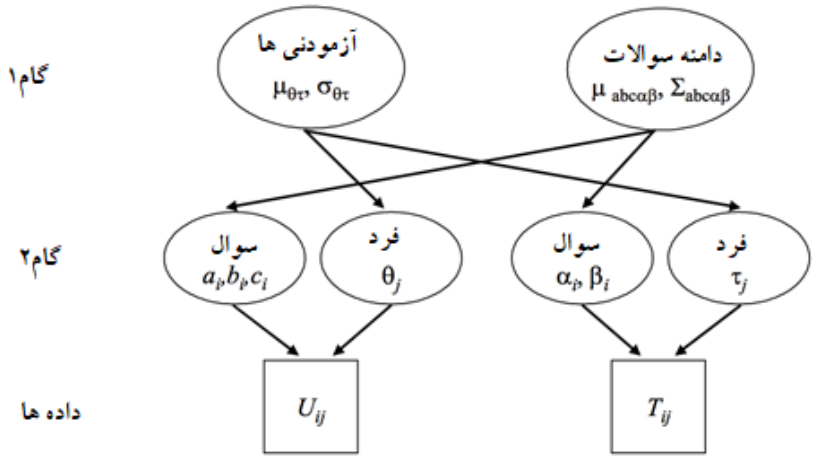
و

$$\mu_I | \Sigma_I \sim N(\mu_0, \Sigma_I / k)$$

$\upsilon_I$ و $V_I$ درجه‌های آزادی ماتریس توزیع ویشارت معکوس، $\mu_0$ میانگین توزیع پیشین، $k$ تعداد اندازه‌گیری‌های پیشین که مقدار پیش‌فرض آن ۱ است. از توزیع پیشین بتا برای پارامتر حدس $c_k$ استفاده می‌شود که مقادیر ابر پارامتری‌های آن ۲۰ و ۸۰ است. این انتخاب پارامترها منجر به نسبت حدس پیشین $\frac{۱}{۵}$ با انحراف استاندارد ۰/۰۴ می‌شود.

Homepage: https://jem.atu.ac.ir/

**شکل ۱.** سلسله‌مراتب برآورد پارامتر در مدل زمان–پاسخ سلسله‌مراتبی (Van der Linden, 2007)




مطابق شکل ۱ و همان‌گونه که در بالا ذکر شد، در مدل سلسله‌مراتبی (Van der Linden, 2007) معرفی کرده است، فرایند برآورد پارامترها و توانایی به‌صورت سلسله‌مراتبی و در دو مرحله انجام می‌شود. در این مدل، $\tau_p$ سرعت p امین آزمون‌دهنده به $p = 1, \dots, P$ را نشان می‌دهد. فرض می‌شود پارامترهایی برای سرعت‌های زمانی وجود دارد که $t_{pi}$، زمان پاسخ مشاهده‌شده pامین آزمون‌دهنده به iامین سؤال برای $i = 1, \dots, I$ است. به‌طور معادل، در این مدل فرض می‌شود زمان پاسخ مشاهده‌شدهٔ آزمون‌دهنده، ناشناخته و اثر سؤال مجزا است. ازآنجایی‌که که توزیع زمان پاسخ چوله به راست است، استفاده از لگاریتم‌های زمان به‌جای زمان، مرسوم است. از طرف دیگر ماهیت زمان پاسخ‌ها تصادفی است. درنتیجه $\ln t_{pi}$ از توزیع نرمال با میانگین $\beta_i - \tau_p$ و واریانس $\alpha_i^{-2}$ پیروی می‌کند. توزیع مشابه می‌تواند توسط تابع چگالی لگ‌نرمال باشد. این موضوع نشان می‌دهد به‌جای لگاریتم، متغیر توزیع‌شده به‌طور طبیعی، نمایی است.

## یافته‌ها

در این بخش، ابتدا به ارائه آمار توصیفی و سپس نتایج مدل سلسله‌مراتبی زمان–پاسخ و نظریه سؤال–پاسخ پرداخته می‌شود. در جدول ۱، شاخص‌های توصیفی سؤالات محاسبه و گزارش شده است.

Homepage: https://jem.atu.ac.ir/

برای این منظور میانگین پاسخ صحیح به سؤال، واریانس سؤال، متوسط زمان پاسخ‌گویی به هر سؤال (برحسب ثانیه)، کمترین و بیشترین زمان ثبت‌شده برای هر سؤال گزارش شده است. نتایج نشان می‌دهد سؤالات **۲۳، ۳۵، ۳۷، ۴۴** و **۴۶** تا **۴۸** کم‌ترین میانگین پاسخ صحیح را دارد (کمتر از ۴۰ درصد). لازم به ذکر است که این سؤالات دارای میانگین زمان پاسخ‌گویی بالایی دارند.

**جدول ۱.** شاخص‌های توصیفی سؤالات استدلال کمّی

| شماره سؤال | میانگین پاسخ صحیح | واریانس پاسخ صحیح | میانگین زمان پاسخ‌گویی | واریانس زمان پاسخ‌گویی | حداقل زمان پاسخ‌گویی | حداکثر زمان پاسخ‌گویی |
|---|---|---|---|---|---|---|
| ۱ | ۰/۸۸ | ۰/۱۰۸ | ۱/۹۳ | ۰/۷٤۰ | ۰/۳۲ | ٤/۰۵ |
| ۲ | ۰/۹۳ | ۰/۰٦۵ | ۱/۵۱ | ۰/۵۷۷ | ۰/۵۰ | ٤/۰۲ |
| ۳ | /۹۵ | ۰/۰٤۷ | ۱/۳۲ | ۰/٤٤۳ | ۰/۲۰ | ٤/۰۲ |
| ٤ | ۰/۹٦ | ۰/۰۳٦ | ۱/۲٤ | ۰/۳۷۳ | ۰/٤۰ | ٤/۰۱ |
| ۵ | ۰/۹۸ | ۰/۰۲۱ | ۱/۱۹ | ۰/۳۲۷ | ۰/۵۰ | ٤/۰۱ |
| ٦ | ۰/۸۷ | ۰/۱۱۱ | ۱/۵۱ | ۰/٤٤۳ | ۰/۱۰ | ٤/۰۱ |
| ۷ | ۰/۹۳ | ۰/۰٦۳ | ۱/٤۰ | ۰/۵۳٦ | ۰/۲۰ | ٦/۰۱ |
| ۸ | ۰/٦۳ | ۰/۲۳۵ | ۲/۳٦ | ۱/۹۲۳ | ۰/۵۰ | ٦/۰۱ |
| ۹ | ۰/۹۵ | ۰/۰۵۱ | ۱/٤۷ | ۰/۸۱٦ | ۰/٤۰ | ٦/۰۱ |
| ۱۰ | ۰/۹۵ | ۰/۰۵۱ | ۱/۸٤ | ۰/۸۷۱ | ۰/۲۰ | ٦/۰۱ |
| ۱۱ | ۰/۹٤ | ۰/۰۵۸ | ۱/٦۰ | ۰/۸۹۷ | ۰/۵۰ | ٦/۰۲ |
| ۱۲ | ۰/۹۱ | ۰/۰۸۳ | ۱/۸٤ | ۰/۸٦۸ | ۰/۱۰ | ٦/۰۲ |
| ۱۳ | ۰/٦٦ | ۰/۲۲۵ | ۳/۰۲ | ۳/۰۰۵ | ۰/۵۰ | ٦/۰٤ |
| ۱٤ | ۰/۹۵ | ۰/۰٤٦ | ۱/۵۹ | ۰/۷٤۵ | ۰/۱۰ | ٦/۰۱ |
| ۱۵ | ۰/۷۱ | ۰/۲۰۷ | ۲/۹۵ | ۲/۳٦۰ | ۰/۵۰ | ٦/۰۲ |
| ۱٦ | ۰/۸۵ | ۰/۱۲۵ | ۱/٦۱ | ۰/٦۲۵ | ۰/۱۰ | ٦/۰۱ |
| ۱۷ | ۰/۷۳ | ۰/۱۹۸ | ۲/۲۷ | ۱/۲۵٤ | ۰/۲۰ | ٦/۰۳ |
| ۱۸ | ۰/۸٤ | ۰/۱۳۷ | ۲/۱۸ | ۱/۲۵۵ | ۰/۲۰ | ۸/۰۱ |
| ۱۹ | ۰/٦۱ | ۰/۲۳۸ | ۳/۰۰ | ۲/۵٦۱ | ۰/۱۰ | ۸/۰۱ |
| ۲۰ | ۰/٤۸ | ۰/۲۵۰ | ۳/۲۲ | ۳/۳۰۳ | ۰/۱۰ | ۸/۰۲ |
| ۲۱ | ۰/۷۵ | ۰/۱۸۵ | ۲/۹٦ | ۳/٤۳۰ | ۰/۵۰ | ۸/۰۲ |
| ۲۲ | ۰/۸۱ | ۰/۱۵۱ | ۲/٦۳ | ۲/٤۰۱ | ۰/۲۰ | ۸/۰۱ |

Homepage: https://jem.atu.ac.ir/

| شماره سؤال | میانگین پاسخ صحیح | واریانس پاسخ صحیح | میانگین زمان پاسخ‌گویی | واریانس زمان پاسخ‌گویی | حداقل زمان پاسخ‌گویی | حداکثر زمان پاسخ‌گویی |
|---|---|---|---|---|---|---|
| ۲۳ | ۰/۳۶ | ۰/۲۳۲ | ۳/۴۹ | ۴/۷۴۴ | ۰/۵۰ | ۸/۰۲ |
| ۲۴ | ۰/۵۹ | ۰/۲۴۲ | ۳/۱۵ | ۳/۵۰۰ | ۰/۵۰ | ۸/۰۳ |
| ۲۵ | ۰/۴۹ | ۰/۲۵۰ | ۳/۷۵ | ۵/۳۷۸ | ۰/۲۰ | ۸/۴۰ |
| ۲۶ | ۰/۶۱ | ۰/۲۳۸ | ۳/۲۴ | ۴/۳۲۸ | ۰/۵۰ | ۸/۰۲ |
| ۲۷ | ۰/۷۳ | ۰/۱۹۷ | ۳/۳۹ | ۴/۸۲۱ | ۰/۲۰ | ۱۰/۰۱ |
| ۲۸ | ۰/۷۳ | ۰/۱۹۹ | ۳/۵۰ | ۵/۵۲۹ | ۰/۵۰ | ۱۰/۰۱ |
| ۲۹ | ۰/۵۷ | ۰/۲۴۶ | ۳/۶۵ | ۶/۹۵۲ | ۰/۵۰ | ۱۰/۰۹ |
| ۳۰ | ۰/۶۲ | ۰/۲۳۶ | ۳/۲۴ | ۶/۲۰۷ | ۰/۴۰ | ۱۰/۰۲ |
| ۳۱ | ۰/۶۱ | ۰/۲۳۸ | ۳/۶۵ | ۶/۳۴۷ | ۰/۵۰ | ۱۰/۰۲ |
| ۳۲ | ۰/۲۴ | ۰/۱۸۱ | ۴/۰۹ | ۸/۳۰۱ | ۰/۲۰ | ۱۰/۰۲ |
| ۳۳ | ۰/۶۲ | ۰/۲۳۶ | ۳/۷۳ | ۶/۸۶۸ | ۰/۳۰ | ۱۰/۰۱ |
| ۳۴ | ۰/۳۱ | ۰/۲۱۴ | ۳/۸۶ | ۷/۹۱۶ | ۰/۵۰ | ۱۰/۰۲ |
| ۳۵ | ۰/۳۷ | ۰/۲۳۵ | ۳/۵۰ | ۷/۰۸۶ | ۰/۵۰ | ۱۰/۰۲ |
| ۳۶ | ۰/۵۳ | ۰/۲۵۰ | ۴/۲۷ | ۱۰/۳۱۶ | ۰/۳۰ | ۱۲/۰۱ |
| ۳۷ | ۰/۳۷ | ۰/۲۳۴ | ۴/۰۶ | ۱۰/۰۴۶ | ۰/۱۰ | ۱۲/۰۱ |
| ۳۸ | ۰/۴۴ | ۰/۲۴۷ | ۳/۹۳ | ۹/۲۹۷ | ۰/۱۰ | ۱۲/۰۲ |
| ۳۹ | ۰/۵۲ | ۰/۲۵۰ | ۳/۹۸ | ۱۰/۵۲۱ | ۰/۴۰ | ۱۲/۰۱ |
| ۴۰ | ۰/۴۵ | ۰/۲۴۸ | ۴/۱۱ | ۱۳/۳۲۵ | ۰/۱۰ | ۱۲/۰۳ |
| ۴۱ | ۰/۵۴ | ۰/۲۴۹ | ۳/۹۶ | ۱۲/۱۹۵ | ۰/۱۰ | ۱۲/۰۲ |
| ۴۲ | ۰/۶۵ | ۰/۲۲۸ | ۳/۴۴ | ۷/۷۲۱ | ۰/۴۰ | ۱۲/۰۱ |
| ۴۳ | ۰/۵۵ | ۰/۲۴۸ | ۳/۶۹ | ۱۰/۴۴۸ | ۰/۵۰ | ۱۲/۰۲ |
| ۴۴ | ۰/۳۲ | ۰/۲۱۹ | ۳/۹۷ | ۱۲/۹۰۸ | ۰/۱۰ | ۱۲/۰۲ |
| ۴۵ | ۰/۵۱ | ۰/۲۵۰ | ۳/۶۳ | ۱۲/۸۰۵ | ۰/۱۰ | ۱۴/۰۰ |
| ۴۶ | ۰/۳۵ | ۰/۲۲۷ | ۳/۸۳ | ۱۴/۱۹۰ | ۰/۲۰ | ۱۴/۰۰ |
| ۴۷ | ۰/۲۹ | ۰/۲۰۷ | ۳/۸۲ | ۱۲/۲۳۸ | ۰/۱۵ | ۱۴/۰۰ |
| ۴۸ | ۰/۳۱ | ۰/۲۱۶ | ۳/۷۸ | ۱۳/۴۵۷ | ۰/۲۰ | ۱۴/۰۰ |

Homepage: https://jem.atu.ac.ir/

همان‌گونه که اشاره شد، این اطلاعات بر اساس ۵۰۲ الگوی پاسخ معتبر گزارش شده است. به‌منظور بررسی فرضیهٔ تک‌بعدی بودن پرسشنامه که از الزامات نظریه‌های جدید اندازه‌گیری است، از تحلیل نظریه سؤال-پاسخ چندبعدی[1] استفاده شده است. بر این اساس، مدل با فرض تک‌بعدی بودن، مورد تجزیه‌وتحلیل قرار گرفت تا با مقایسه مقدار برازش‌شده با ملاک‌های RMSEA، CFI و TLI، آزمون تک‌بعدی بودن پرسشنامه استنباط شود. نتایج این تحلیل در جدول ۲ گزارش شده است.

**جدول ۲.** شاخص‌های برازش مدل تک‌بعدی بر اساس محاسبه سؤال-پاسخ چندبعدی

| شاخص | لگاریتم درست‌نمایی | درجه آزادی | معناداری | RMSEA | CFI | TLI |
|---|---|---|---|---|---|---|
| میزان مطلوب | - | - | >۰/۰۵ | <۰/۰۸ | ۱-۰/۹ | ۱-۰/۹ |
| مقدار محاسبه‌شده | -۱۲۰٤ | ۱۰۸۰ | ۰/۰۰۱ | ۰/۰۵۳ | ۰/۹۲ | ۰/۹۱ |

نتایج جدول ۲ نشان می‌دهد شاخص‌های برازش داده‌شده در دامنه مناسبی قرار دارند و داده‌های تحلیل شده از تک‌بعدی بودن آزمون حمایت می‌کند.

در ادامه، نتایج مدل بندی سلسله‌مراتبی زمان پاسخ ارائه شده است. به‌منظور مقایسه، پارامترهای برآورد شده از محاسبات نظریه سؤال-پاسخ که زمان در آن لحاظ نشده است، نیز گزارش شده است. با توجه به برآوردهای مربوط به مدل سلسله‌مراتبی برای هر سؤال، دو ضریب تمییز و دو ضریب دشواری که یکی مربوط به سؤال و دیگری مربوط به زمان پاسخ‌گویی به آن سؤال است، برآورد شده است. ضرایب دشواری منفی (تا ۴-) نشان می‌دهند که سؤال آسان برآورد شده است و ضرایب دشواری مثبت (تا ۴) نشان می‌دهند سؤال دشوار برآورد شده است. ضرایب تمییز منفی نشان می‌دهند سؤال عملکرد نامناسبی داشته است و ضرایب مثبت تأیید می‌کنند سؤال عملکرد مناسبی در سنجش صفت موردنظر داشته است.

1. Multidimensional Item Response Theory

Homepage: https://jem.atu.ac.ir/

**جدول ۳.** پارامترهای برآورد شده بر اساس مدل سلسله‌مراتبی زمان پاسخ

| سؤال | مدل سؤال-پاسخ | | مدل زمان-پاسخ | | | |
|---|---|---|---|---|---|---|
| | ضریب تمییز | ضریب دشواری | ضریب تمییز | ضریب دشواری | ضریب تمییز زمان | ضریب دشواری زمان |
| سؤال ۱ | ۰/۳۷ | -۳/۲۸ | ۰/۸۷ | -۱/۱۱ | ۰/۲۰۱ | ۰/۲٤٥ |
| سؤال ۲ | ۰/٦۲ | -۲/۸٥ | ۱/۰۲ | -۱/٤۳ | ۰/۲۹۸ | ۰/۱۳٤ |
| سؤال ۳ | ۰/٥۱ | -۳/۷۹ | ۰/۷۸ | -۱/٥۲ | ۰/۳۱۹ | ۰/۰۸۰ |
| سؤال ٤ | ۰/۷٦ | -۳/۱۱ | ۰/۹۲ | -۱/٦٥ | ۰/۳٦٥ | ۰/۰٥۱ |
| سؤال ٥ | ۰/٥۹ | -٤/۳۰ | ۰/٦۲ | -۱/۷٤ | ۰/۲۳۷ | ۰/۰۳۷ |
| سؤال ٦ | ۰/٤۳ | -۲/۹۰ | ۰/۷٦ | -۱/۰۸ | ۰/۲۹۷ | ۰/۱٤۲ |
| سؤال ۷ | ۰/٥۸ | -۳/۰٥ | ۱/۰۲ | -۱/٤٥ | ۰/٤۱۸ | ۰/۱۰۱ |
| سؤال ۸ | ۰/۸۰ | -۰/٤۷ | ۱/۸۱ | -۰/۳۷ | ۰/٦٥٥ | ۰/۲۹۲ |
| سؤال ۹ | ۰/٦۰ | -۳/۲۸ | ۰/۸۷ | -۱/٥۲ | ۰/٤٥۷ | ۰/۱۰۷ |
| سؤال ۱۰ | ۰/٥٦ | -۲/۹٤ | ۰/۸۳ | -۱/٥۱ | ۰/٤٦۲ | ۰/۲۱۲ |
| سؤال ۱۱ | ۰/۷۹ | -۲/٥۳ | ۱/۲۲ | -۱/٥٤ | ۰/٤٤۲ | ۰/۱٤۷ |
| سؤال ۱۲ | ۰/۹۲ | -۱/۸۷ | ۱/٤۹ | -۱/٤۱ | ۰/٤۱٦ | ۰/۲۱٥ |
| سؤال ۱۳ | ۱/۲۷ | ۱/۳٦ | ۰/٥۸ | -۰/۳۹ | ۱/۰۰۷ | ۰/۳۹۰ |
| سؤال ۱٤ | ۰/٤٥ | -۲/۹٥ | ۰/۷٤ | -۱/٥٤ | ۰/٤۹٤ | ۰/۱٥۱ |
| سؤال ۱٥ | ۰/۹۹ | ۰/۳۰ | ۱/۰٥ | -۰/٥٦ | ۰/۸۸۲ | ۰/۳۹٥ |
| سؤال ۱٦ | ۲/٤۷ | ۰/۹٥ | ۰/٦۹ | -۱/۰۱ | ۰/٤٦۸ | ۰/۱٦٤ |
| سؤال ۱۷ | ۰/٥۰ | -۱/۳۲ | ۱/۳۲ | -۰/٦٤ | ۰/٦۰٦ | ۰/۳۰۲ |
| سؤال ۱۸ | ۰/۷۰ | -۱/٦۸ | ۱/٦۱ | -۱/۰۸ | ۰/٦٦۷ | ۰/۲۸۳ |
| سؤال ۱۹ | ۰/۷۸ | -۰/۲۱ | ۱/٦۷ | -۰/۳۲ | ۰/۹٤۲ | ۰/۳۹۹ |
| سؤال ۲۰ | ۰/٤۷ | ۱/۳۳ | ۳/۸۳ | ۰/۰٥ | ۱/۲۰۸ | ۰/٤۱٦ |
| سؤال ۲۱ | ۰/۸٥ | -۱/۰۱ | ۱/۸۳ | -۰/۸۲ | ۱/۱۹۳ | ۰/۳٦۹ |
| سؤال ۲۲ | ۲/۸۸ | ۰/٦۸ | ۱/۱٤ | -۰/۹۲ | ۰/۸۷۸ | ۰/۳٤۰ |
| سؤال ۲۳ | ۱/۰۲ | ۰/٥۱ | ۲/۱۱ | ۰/٤۳ | ۱/٤۳۲ | ۰/٤۳۲ |
| سؤال ۲٤ | ۱/۰۱ | -۰/۰۸ | ۱/۸٦ | -/۳۰ | ۱/۳٦۸ | ۰/۳۹٦ |
| سؤال ۲٥ | ۱/۹٤ | ۱/٤۹ | ۰/٥۷ | ۰/۰۲ | ۱/٦٦۰ | ۰/٤٥۱ |
| سؤال ۲٦ | ۰/۷۹ | -۰/۰٥ | ۱/٤۳ | -۰/۳۲ | ۱/٥۲٦ | ۰/۳۹۰ |
| سؤال ۲۷ | ۰/٥۱ | -۱/۳۰ | ۱/۲۱ | -۰/٦٦ | ۱/٥٥۲ | ۰/٤۰٥ |
| سؤال ۲۸ | ۱/۲۳ | -۰/۰۹ | ۱/٥٦ | -۰/۷۰ | ۱/٥۷۳ | ۰/٤۱٤ |

Homepage: https://jem.atu.ac.ir/

| سؤال | مدل سؤال-پاسخ | | مدل زمان-پاسخ | | | |
|---|---|---|---|---|---|---|
| | ضریب تمییز | ضریب دشواری | ضریب تمییز | ضریب دشواری | ضریب تمییز زمان | ضریب دشواری زمان |
| سؤال ۲۹ | ۲/۱۰ | ۰/۷۳ | ۱/۳۳ | -۰/۲۰ | ۱/۷۵۳ | ۰/٤۱٦ |
| سؤال ۳۰ | ۰/۳۹ | -۰/۸۰ | ۰/۹۳ | -۰/۳۲ | ۱/٥۸۹ | ۰/۳٦۸ |
| سؤال ۳۱ | ۲/۰٦ | ۱/۱۸ | ۰/۷۷ | -۰/۳۰ | ۱/۷۱۳ | ۰/٤۲۱ |
| سؤال ۳۲ | ۱/٤۹ | ۱/۱۰ | ۱/۸٥ | ۰/۸۳ | ۱/۸۸۹ | ۰/٤٦۰ |
| سؤال ۳۳ | ۲/۱۲ | ۱/۰۷ | ۰/۷٥ | -۰/۳۲ | ۱/۸۳۰ | ۰/٤۲۳ |
| سؤال ۳٤ | ۰/۸۳ | ۰/۸۲ | ۱/۷۲ | ۰/٥۷ | ۱/۸۹۰ | ۰/٤۲٦ |
| سؤال ۳٥ | ۱/۲۳ | ۰/٥٤ | ۲/۰٤ | ۰/۳۷ | ۱/۸٦۷ | ۰/۳۸۱ |
| سؤال ۳٦ | ۱/۳۳ | ۰/۸٥ | ۱/۱٤ | -۰/۱۰ | ۱/۹٤٦ | ۰/٤٥۸ |
| سؤال ۳۷ | ۱/۰۱ | ۰/٦۹ | ۱/۸۲ | ۰/۳۷ | ۱/۷۹۱ | ۰/٤۳٦ |
| سؤال ۳۸ | -۰/۰۱ | -۳/٥۰ | ۰/۱۸ | ۰/۱٤ | ۱/۹۰۸ | ۰/٤۱۸ |
| سؤال ۳۹ | ۲/۱٦ | ۰/۷۰ | ۱/۳۳ | -۰/۰۷ | ۱/۹۹٥ | ۰/٤۱۳ |
| سؤال ٤۰ | ۲/٤۹ | ۱/٦۱ | ۰/٤۷ | ۰/۱۲ | ۲/۰۲۰ | ۰/٤۰٥ |
| سؤال ٤۱ | ۱/۳٦ | ۰/۸۲ | ۱/۱۲ | -۰/۱۱ | ۱/۸۹٤ | ۰/۳۹٥ |
| سؤال ٤۲ | ۲/۳۳ | ۰/۸۱ | ۱/۱۱ | -۰/٤۲ | ۱/۷۹۹ | ۰/۳٦٦ |
| سؤال ٤۳ | ۰/۳٥ | -۰/۳۲ | ۰/۹٥ | -۰/۱۳ | ۱/۹۳٦ | ۰/۳٦٦ |
| سؤال ٤٤ | ۲/۱٥ | ۱/٦۷ | ۰/۸۱ | ۰/٤۷ | ۱/۹٥۰ | ۰/۳۹۰ |
| سؤال ٤٥ | ۰/۷٦ | ۰/۷۷ | ۱/۰٤ | -۰/۰۲ | ۱/۷٦٥ | ۰/۳٤٥ |
| سؤال ٤٦ | ۰/۲۲ | ۱/۷۸ | ۰/٦٥ | ۰/٤۰ | ۱/۸۳۷ | ۰/۳٥۷ |
| سؤال ٤۷ | ۰/۳۲ | ۱/۷۲ | ۰/۸٥ | ۰/٥٦ | ۱/۷۷۲ | ۰/۳۷۱ |
| سؤال ٤۸ | ۰/۳۳ | ۱/٤۹ | ۰/۸۳ | ۰/٤۹ | ۱/۷٤۹ | ۰/۳٦۰ |

بر اساس نتایج جدول ۳، مشابهت‌ها و تفاوت‌هایی در برآورد پارامترها و دامنه تغییرات آنها وجود دارد. به‌عنوان‌مثال برای سؤال ۴۰، ضریب تمییز در نظریه سؤال-پاسخ بالا (۲/۴۹) ولی در نظریه زمان-پاسخ، ضریب کوچکی (۰/۴۷) دارد. هرچند که ضریب تمییز زمان در نظریه زمان-پاسخ این سؤال، بالا (۲/۰۲) است. در برخی سؤالات مانند سؤال ۴۷ نیز مشاهده می‌شود سؤالات در یک پارامتر بهترین در مدل خود هستند. سؤال ۴۷ در هر دو مدل ضریب دشواری بالایی دارد. به‌منظور بررسی دقیق‌تر رابطه بین پارامترهای برآوردشده از مدل زمان-پاسخ با مدل سؤال-پاسخ،

Homepage: https://jem.atu.ac.ir/

ضرایب همبستگی بین پارامترهای مدل زمان-پاسخ و پارامترهای مدل سؤال-پاسخ، در جدول ۴ گزارش شده است.

**جدول ٤.** ضرایب همبستگی بین پارامترهای مدل زمان-پاسخ با پارامترهای مدل سؤال-پاسخ

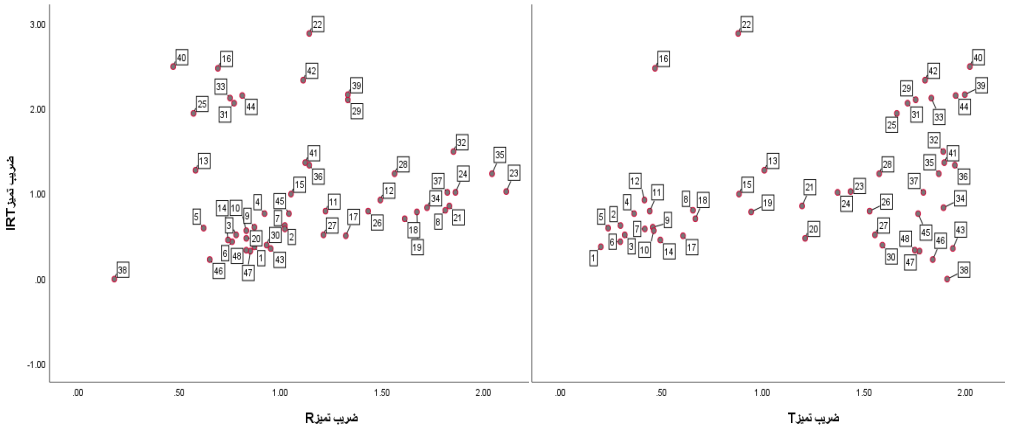
| پارامترها | ضریب تمییز R | ضریب دشواری R | ضریب تمییز T | ضریب دشواری T | ضریب تمییز IRT | ضریب دشواری IRT |
|---|---|---|---|---|---|---|
| ضریب تمییز R | ۱ | | | | | |
| ضریب دشواری R | ۰/۲۱ | ۱ | | | | |
| ضریب تمییز T | ۰/۱۰ | **۰/۸۷ | ۱ | | | |
| ضریب دشواری T | ۰/۲۷ | **۰/۸۵ | **۰/۸۵ | ۱ | | |
| ضریب تمییز IRT | ۰/۰۲ | ۰/۱۸ | *۰/۳۲ | *۰/۳۲ | ۱ | |
| ضریب دشواری IRT | ۰/۱۵ | **۰/۸۱ | **۰/۷۵ | **۰/۷۹ | **۰/٥٤ | ۱ |

**$p \leq ۰/۰۱$

بر اساس نتایج جدول ۴، بیشترین رابطه بین پارامترهای دو مدل، مربوط به ضریب دشواری پاسخ در مدل زمان-پاسخ و ضریب دشواری نظریه سؤال-پاسخ با مقدار ۰/۸۱ است. ضریب تمییز زمان در نظریه زمان-پاسخ نیز همبستگی ۰/۳۲ با ضریب تمییز در مدل سؤال-پاسخ دارد. همچنین مشاهده می‌شود بین ضریب تمییز زمان با ضریب دشواری زمان با ضریب تمییز سؤال-پاسخ رابطه معنی‌دار شده است. در شکل‌های ۲ و ۳، نمودار پراکندگی هم‌تغییری بین پارامترهای مدل سلسله‌مراتبی زمان-پاسخ با مدل سؤال-پاسخ گزارش شده است.

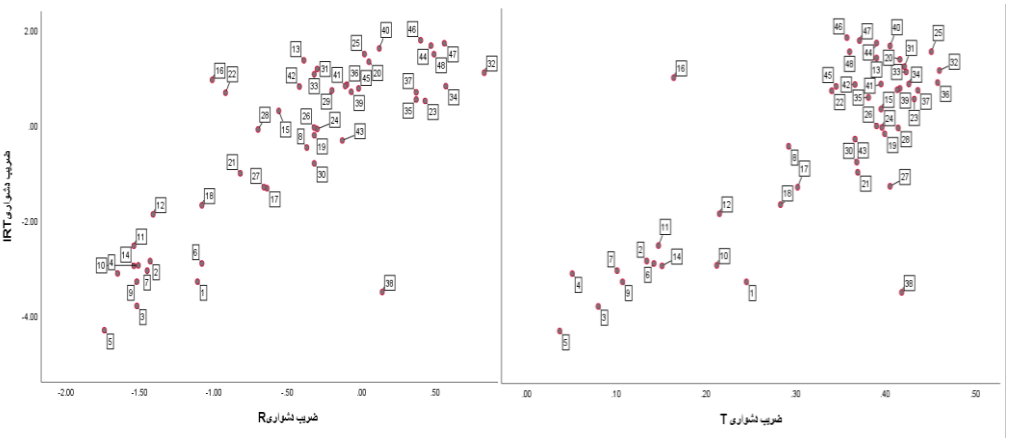
Homepage: https://jem.atu.ac.ir/

**شکل ۲.** نمودار رابطه بین ضریب تمییز زمان و زمان پاسخ با ضریب تمییز مدل سؤال-پاسخ




نتایج شکل ۲ نشان می‌دهد بین ضرایب تمییز حاصل از پاسخ مدل زمان-پاسخ و ضرایب تمییز سؤال-پاسخ ارتباط و هم‌تغییری ضعیفی وجود دارد. در دامنه ۰/۵۰ تا ۱/۵ مدل زمان-پاسخ، تفاوت مقادیر ضریب تمییز مدل‌ها بیشتر است. به‌عبارت‌دیگر این نمودار نشان می‌دهد در مواردی که سؤالات ازنظر دشواری متوسط هستند، در نظر گرفتن مدل زمان پاسخ ضرایب را تعدیل کرده است.

**شکل ۳.** نمودار پراکندگی نقاط بین ضریب دشواری زمان و زمان پاسخ با ضریب دشواری مدل سؤال-پاسخ




Homepage: https://jem.atu.ac.ir/

نتایج شکل ۳ نشان می‌دهد که در پارامتر ضریب دشواری، شباهت‌ها و ترتیب‌های یکسانی بین نتایج دو مدل بیشتر است. در چند سؤال، پارامتر دشواری در مدل‌ها، نتایج متفاوتی ارائه می‌دهند. این سؤالات مربوط به مواردی است که افراد قوی که پاسخ درست داده‌اند، زمان زیادی برای پاسخگویی صرف کرده‌اند و با در نظر داشتن زمان، این موضوع مشخص شده است. ولی در مدل سؤال-پاسخ که زمان لحاظ نشده است، پاسخ‌های تصادفی که درست بوده، ضریب دشواری کمتری را برآورد کرده است.

**جدول ۵.** نتایج مقایسه پارامترهای برآورد شده مدل‌های زمان-پاسخ و سؤال-پاسخ

| متغیر | منابع تغییرات | SS | df | MS | F | P |
|---|---|---|---|---|---|---|
| پارامتر | مدل | ٦٩/٤٠ | ١ | ٦٩/٤٠ | ٦٩/٩٦ | ٠/٠٠٠ |
| ضریب تمییز | خطا | ٤٦/٦٣ | ٤٧ | ٠/٩٩ | | |
| پارامتر | مدل | ١٥/١٥ | ١ | ١٥/١٥ | ١٠/١٠ | ٠/٠٠٣ |
| ضریب دشواری | خطا | ٧٠/٤٨ | ٤٧ | ١/٥٠ | | |

جدول ۵، نتایج مقایسه وجود یا عدم وجود تفاوت پارامترهای دو مدل با زمان پاسخ و بدون زمان پاسخ به تفکیک ضریب تمییز و ضریب دشواری را نشان می‌دهد. نتایج نشان می‌دهد تفاوت پارامتر ضریب تمییز و پارامتر ضریب دشواری معنی‌دار است. به‌عبارت‌دیگر بین دو روش تفاوت معنی‌داری در برآورد پارامترها وجود دارد. این معنی‌داری در کنار طبقات بیشتری که مدل زمان-پاسخ ایجاد کرده است قابل تفسیر خواهد بود. در جدول ۶، شاخص‌های مرکزی و پراکندگی توانایی‌های برآورد شده بر اساس مدل زمان-پاسخ گزارش شده است.

**جدول ٦.** توصیف توانایی‌های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ

| شاخص‌ها | تتا IRT | تتا RT | گاما RT |
|---|---|---|---|
| میانگین | ٠/٢١٨٧ | ٠/٠٠ | ٠/٠٠٠ |
| واریانس | ٠/١٣٥ | ٠/١٣٤ | ٠/٠٣٤ |
| کمینه | -١/١٨ | -١/١٢ | ٠/٧٦ |
| بیشینه | ١/٥٦ | ١/٠٦ | -٠/٣٢ |
| تعداد نفرات | ٥٠٢ | ٥٠٢ | ٠/٤٤ |

Homepage: https://jem.atu.ac.ir/

نتایج جدول ۶ نشان می‌دهد، در مدل زمان-پاسخ، برآورد میانگین تتا صفر و واریانس ۰/۱۳۴ است ولی برای مدل سؤال-پاسخ میانگین تتا ۰/۲۱۸ برآورد شده است. تتای صفر نشان می‌دهد میانگین پارامتر توانایی بر اساس زمان پاسخ شکل نسبتاً قرینه‌ای دارد و به توزیع نرمال نزدیک‌تر است. کمترین توانایی ۱/۱۲- و بیشترین توانایی ۱/۰۶ برآورد شده است؛ اما در گاما (مبتنی بر زمان) کمترین مقدار ۰/۷۷- و بیشترین توانایی ۰/۴۴ برآورد شده است. به‌نوعی در توانایی که بر اساس اطلاعات زمان تقویت شده است افراد در مقیاس نسبت به میانگین با حساسیت بیشتری طبقه‌بندی شده‌اند و طبقات بیشتری تشکیل شده است، همچنین توزیع به نرمال نزدیک‌تر است.

**جدول ۷.** مقایسه توانایی‌های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ

| متغیر | منابع تغیر | SS | df | MS | F | P |
|---|---|---|---|---|---|---|
| توانایی | مدل | ۱۲/۰۱ | ۱/۰۰ | ۱۲/۰۱ | ۱۱۳/۸۳ | ۰/۰۰۰ |
| | خطا | ۵۲/۸۶ | ۵۰۱/۰۰ | ۰/۱۱ | | |

نتایج جدول ۷ نشان می‌دهد، تفاوت معنی‌داری ($p<۰/۰۰۱$) بین توانایی‌های برآورد شده در دو مدل برای هر فرد وجود دارد. برای مشخص‌تر شدن چگونگی این تفاوت‌ها رتبه‌های اختصاص‌یافته به هر فرد بر اساس مدل با زمان پاسخ و مدل بدون زمان پاسخ محاسبه و اختلاف بین این رتبه‌ها بررسی شد. نکته قابل‌توجه دیگر این است که طبقات تولیدشده بر اساس توانایی‌ها در نظریه سؤال پاسخ ۳۵ طبقه و در مدل زمان پاسخ ۵۰۲ طبقه بوده شد. در جدول ۸، خلاصه‌ای از نتایج این رتبه‌ها گزارش شده است.

**جدول ۸.** مقایسه توانایی‌های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ

| کد | رتبه در IRT | رتبه در RT | تفاوت | کد | رتبه در IRT | رتبه در RT | تفاوت |
|---|---|---|---|---|---|---|---|
| **۱۵۰** | ۳۰ | ٤٧٨ | ٤٤٨ | **٦٨** | ٤٩٤ | ۲۲ | -٤٧۲ |
| **۱۹۸** | ۳٤ | ٤٧۱ | ٤۳۷ | **۳٤۹** | ٤۸۰ | ۱٥ | -٤٦٥ |
| **۲٤۲** | ۲۰ | ٤۲۳ | ٤۰۳ | **۱۱٦** | ٤۹۸ | ۱٥ | -٤٦۳ |
| **۷۲** | ٥٦ | ٤٥٤ | ۳۹۸ | **۱۰۷** | ٤٦۷ | ۱۲ | -٤٥٥ |
| **٦٤** | ٥٥ | ٤٥۱ | ۳۹٦ | **۳۳۷** | ٤٥۹ | ۱۱ | -٤٤۸ |
| **۲۰٥** | ۲٤ | ٤۱۹ | ۳۹٥ | **۳۸۳** | ٤۷۲ | ٦۸ | -٤۰٤ |

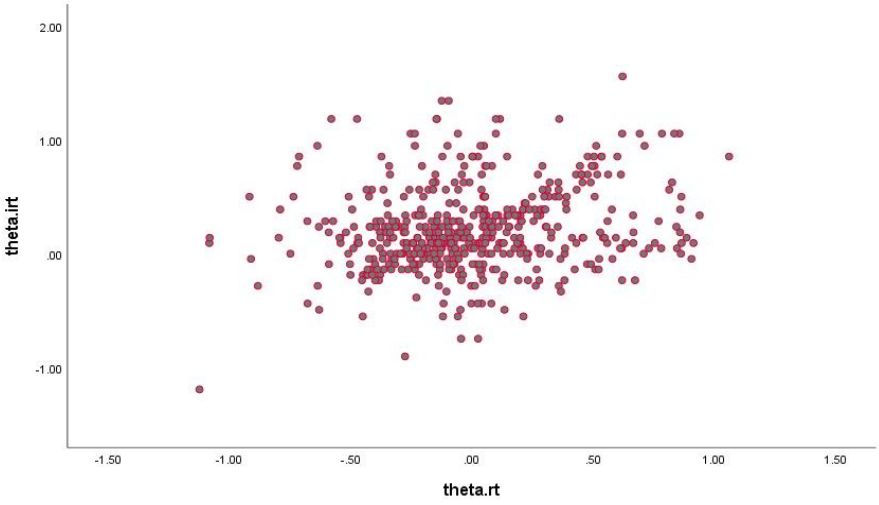
Homepage: https://jem.atu.ac.ir/

| کد | رتبه در IRT | رتبه در RT | تفاوت | کد | رتبه در IRT | رتبه در RT | تفاوت |
|---|---|---|---|---|---|---|---|
| **۲۷۳** | ۱۰٤ | ٤۹۹ | ۳۹٥ | **۳۲٥** | ٤۰٥ | ٤ | -٤۰۱ |
| **۲۰٦** | ۳٥ | ٤۲٤ | ۳۸۹ | **۳۲۸** | ٤۰٦ | ۱۰ | -۳۹٦ |
| **۱۸۱** | ۷ | ۳۷۹ | ۳۷۲ | **۳٦۰** | ٤۲۲ | ٤۳ | -۳۷۹ |
| **٤٤** | ۹٥ | ٤٦٥ | ۳۷۰ | **۳۲۰** | ٤٥۸ | ۸۰ | -۳۷۸ |
| **۱۲٦** | ۷٥ | ٤٤٥ | ۳۷۰ | **۳۸٦** | ٤۰۷ | ۲۹ | -۳۷۸ |
| **۷٥** | ۲۳ | ۳۹۱ | ۳٦۸ | **۳۲٦** | ۳۸۱ | ۸ | -۳۷۳ |
| **۱٦** | ۷٤ | ٤٤۱ | ۳٦۷ | **٦۹** | ٤۸٦ | ۱۱٤ | -۳۷۲ |
| **۲٤۰** | ٦٥ | ٤۳۱ | ۳٦٦ | **۸۰** | ٤۸۷ | ۱۲۱ | -۳٦٦ |
| **۱۸۰** | ۳۳ | ۳۹٦ | ۳٦۳ | **۱۱۷** | ٤۱۷ | ٥۲ | -۳٦٥ |
| **۲۸٥** | ۱۳٥ | ٤۹٥ | ۳٦۰ | **۳۱۹** | ٤٤۲ | ۸۱ | -۳٦۱ |
| **۱۷٥** | ۱۰۰ | ٤٥٥ | ۳٥٥ | **۳۹٥** | ٤۸۲ | ۱۲٤ | -۳٥۸ |
| **۱٥۷** | ۳۱ | ۳۸٥ | ۳٥٤ | **۲۱۷** | ٤۰۱ | ٤۷ | -۳٥٤ |
| **۲۳۲** | ۱۲۸ | ٤۸۰ | ۳٥۲ | **۳٥۰** | ۳۸۲ | ۳۲ | -۳٥۰ |
| **۲۰۱** | ۱۰ | ۳٥۱ | ۳٤۱ | **٦۷** | ٤۱٥ | ۷۱ | -۳٤٤ |

در جدول ۸ از بین ۵۰۲ آزمودنی، ۴۰ اختلاف که قابل‌توجه بودند و تفاوت بیشتری داشتند گزارش شده‌اند. همان‌گونه که در جدول مشاهده می‌شود آزمودنی با کد ۱۵۰ بر اساس محاسبات زمان-پاسخ جایگاه ۴۷۸ ام را در بین ۵۰۲ نفر کسب کرده است، اما همین فرد در مدل بدون زمان جایگاه ۳۰ ام را کسب کرده است که نشان‌دهنده اختلاف ۴۴۸ رتبه‌ای در دو مدل است. بر اساس مدل با زمان، فرد در گروه قوی جای گرفته است ولی در مدل با مدل بندی کردن زمان-پاسخ، فرد در گروه ضعیف جای گرفته است. در بین این ۵۰۲ نفر، تنها ۱۸ نفر بودند که در دو مدل رتبه‌های یکسانی داشتند و باقیمانده دارای رتبه اختلاف بودند.

Homepage: https://jem.atu.ac.ir/

**شکل ٤.** رابطه توانایی‌های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ




در شکل ۴ روابط بین نمرات تتا با زمان پاسخ و بدون زمان-پاسخ قابل‌مشاهده است همان‌گونه که در نمودار نیز مشاهده می‌شود در نمرات با زمان پاسخ طبقات بیشتری تولید کرده است و افراد از یکدیگر متمایزتر هستند.

## بحث و نتیجه‌گیری

هدف از انجام این پژوهش، کاربرد مدل‌های سلسله‌مراتبی زمان-پاسخ در اندازه‌گیری‌های شناختی کودکان است. برای دستیابی به این هدف، از روش استدلال از نوع کمّی به‌عنوان یکی از شناخته‌شده‌ترین آزمون‌های شناختی انتخاب و اجرا شده است. بر این اساس، پارامترهای سؤال و توانایی بر اساس مدل سلسله‌مراتبی زمان-پاسخ برآورد شده است. سپس به‌منظور بررسی کاربرد این مدل زمان-پاسخ، نتایج این مدل با نتایج مدل سؤال-پاسخ که در آن زمان مدل‌سازی نشده است، گزارش و مقایسه شده است.

نتایج بررسی‌ها نشان داد که بین پارامترهای برآوردشده از مدل زمان-پاسخ، همبستگی وجود دارد و همچنین بین پارامترهای مدل زمان-پاسخ با مدل بدون زمان پاسخ روابط معنی‌داری وجود دارد. در نتایج مشاهده شد برخی از سؤالاتی که در مدل زمان پاسخ، پارامتر ضریب تمیز متوسطی

Homepage: https://jem.atu.ac.ir/

داشتند، در مدل بدون زمان پاسخ با ضرایب تمییز قوی‌تر برآورد شده‌اند. از طرف دیگر همبستگی بالایی بین ضرایب تمییز و میانگین زمان پاسخگویی به سؤالات وجود داشت. به‌گونه‌ای که سؤالاتی که دشوارتر بودند و نیاز به صرف زمان بیشتری داشتند، دشوارتر قلمداد شدند. اگر پاسخ‌های صحیح و کل زمان پاسخ‌ها را جمع کنیم و دو نمره کل محاسبه کنیم، بسته به علامت همبستگی بین این متغیرها، ممکن است همبستگی‌های مثبت و منفی کاذب بین پاسخ‌ها و زمان پاسخ‌ها مشاهده شود. همین همبستگی کاذب در مورد جمع میان آزمون‌دهندگانی که در صحت و سرعت متفاوت هستند صدق می‌کند. اگر نمره کل سؤالات و نمره کل آزمون‌دهندگان را محاسبه کنیم، همبستگی‌های بین متغیرهای آنها با هم تقابل می‌کند، و بسته به علائم آنها و توان نسبی، پیش‌بینی نتایج بدون هرگونه تحلیل رسمی دشوارتر می‌شوند (Van Der Linden & Ren, 2015). یک راه‌حل مؤثر برای حل مشکل همبستگی کاذب، مدل‌سازی صریح رابطه بین متغیرهای مشاهده‌شده و تمام متغیرهای کمکی است. به گفتهٔ ون در لیندن و گلس[1] (۲۰۱۰)، اگر مدل با داده‌های تجربی به‌طور رضایت بخشی برازش داشته باشد، قادر خواهیم بود اثرات متغیرهای کمکی را از طریق مطلوب‌سازی آن‌ها خنثی کنیم. این اصل، که در زمینه تمام صفات مکنون نهفته است، کاربردهایی در زمینه‌های مختلف مانند فیزیک و روان‌شناسی پیدا کرده است.

در ضرایب دشواری برآورد شده برای سؤالات، شباهت‌های قابل‌توجهی وجود دارد. از طرف دیگر نکته جالب‌توجه، رابطه معنی‌دار ضریب دشواری زمان پاسخ با ضریب تمییز بدون زمان پاسخ است. به‌عبارت‌دیگر در مواردی که زمان پاسخگویی نسبت به میانگین بیشتر است، احتمال پاسخگویی افراد قوی‌تر و تتای بالا در استدلال زیاد شده است و پاسخ درست اتفاق افتاده است، اما برای گروه ضعیف با زمان کم یا زیاد نسبت به میانگین، پاسخ صحیحی رخ نداده است.

در برآورد توانایی، به‌طورکلی هم‌تغییری وجود داشت، ولی تفاوت‌های قابل‌توجهی در توزیع و پراکندگی تتاها مشاهده شد. در برخی از آزمودنی‌ها (آزمودنی ۳۷۶ ام) رتبه در مدل زمان-پاسخ توانایی بالاتر را نشان داد و در برخی از آزمودنی‌ها (آزمودنی ۱۷۷ ام) رتبه در مدل زمان-پاسخ توانایی کمتر را نشان داد. این عدم انطباق دلایل متفاوتی دارد. یکی از شناخته‌شده‌ترین دلایل اریبی[2]

1. Van Der Linden & & Glas
2. bias

Homepage: https://jem.atu.ac.ir/

در برآورد توانایی در روش بیشینه درست‌نمایی[1] که در مدل‌های سؤال-پاسخ مطرح است، به آزمودنی‌های کرانه‌ای[2] مربوط است (Ranger & Kuhn, 2015).

درنهایت، عملکرد شناختی نمی‌تواند به‌طور مستقیم و خالص شناسایی شود. حتی برای دو استدلال جمع ساده، اگرچه طول زمانی دو مرحله استدلال‌کردن می‌تواند به‌وضوح عاملی باشد که بر میزان تلاشی که آنها نیاز دارند تأثیرگذار باشد، یکسان فرض‌کردن این دو فرایند کاملاً غلط است؛ چون عوامل دیگری مانند ترکیباتی از اعداد که باید در هر مرحله اضافه شوند، می‌تواند تأثیرگذار باشد؛ با افزایش یک مرحله به استدلال‌ها متغیرهای دیگری نیز می‌توانند روی استدلال تأثیرگذار شوند (De Boeck & Jeon, 2019).

در این پژوهش تلاش شد تا مزیت‌های کاربرد زمان پاسخ در اندازه‌گیری توانایی شناختی، معرفی شود. مدل‌سازی زمان پاسخ کاربردهای قابل‌توجه دیگری مانند ارزیابی سطوح انگیزه آزمون‌شوندگان، به‌ویژه برای آزمون‌های کم‌ریسک (Finn, 2015)، ارزیابی کاربرد استراتژی، با تفاوت در زمان پاسخ برای زیرگروه‌هایی که از استراتژی‌های حل مسئله متفاوت استفاده می‌کنند (DiTrapani et al., 2016)، امتیاز در آزمون (Molenaar et al., 2016)، تشخیص تقلب (Van Der Linden, 2008)، طراحی فرم‌های موازی (Van Der Linden, 2005) و انتخاب آیتم در تست تطبیقی (Van Der Linden, 2008) و درنهایت مدل‌سازی زمان پاسخ بر روی شخصیت و ارزیابی نگرش (Ranger, 2013؛ Ferrando & Lorenzo, 2007) دارد.

در جمع‌بندی تمام موارد می‌توان گفت اگرچه عملکرد شناختی نمی‌تواند به‌طور مستقیم اندازه‌گیری شود، ولی ملاحظه شد که زمان پاسخ اطلاعاتی را به‌عنوان یک پارامتر نهفته در هر سؤال وارد مدل می‌کند. لازم به ذکر است که این پارامتر از طریق اثر خود روی زمان صرف شده در مورد سؤال، میزان تلاش و فعالیت ذهنی شخص را به شکل غیرمستقیم نشان می‌دهد. بر اساس قوانین مدل‌سازی، هرچه اطلاعات واردشده به مدل بیشتر و ناهمبسته باشند، توان و آگاهی‌دهندگی مدل افزایش می‌یابد (Witelski & Bowen, 2015). بر این اساس، استفاده و کاربرد مدل‌های زمان-پاسخ اندازه‌گیری در سنجش توانایی شناختی پیشنهاد می‌شود.

1. maximum likelihood
2. extreme

Homepage: https://jem.atu.ac.ir/

## تعارض منافع

نویسندگان مقاله هیچ‌گونه تعارض منافعی ندارند.

## ORCID


Ghasem Keshavarz Gerami https://orcid.org/0000-0003-3757-3163
Mohammad Reza Flasafi Nejad http://orcid.org/0000-0002-3227-102X
Jalil Younesi http://orcid.org/0000-0002-9619-1600
Ali Delavar http://orcid.org/0000-0002-8404-5980
Zahra Rezaei Ghahroodi http://orcid.org/0000-0002-9138-8535


Homepage: https://jem.atu.ac.ir/

## References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Berninger, V. W., & Wolf, B. J. (2016). *Dyslexia, dysgraphia, OWL LD, and dyscalculia*. Paul H. Brookes Publishing Co.

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology, 10*, Article 102. https://doi.org/10.3389/fpsyg.2019.00102

Deloche, G., Seron, X., Larroque, C., Magnien, C., Metz-Lutz, M. N., Noel, M. N., ... & Pradat-Diehl, P. (1994). Calculation and number processing: Assessment battery; role of demographic factors. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 195–208. https://doi.org/10.1080/01688639408402631

DiTrapani, J., Jeon, M., De Boeck, P., & Partchev, I. (2016). Attempting to differentiate fast and slow intelligence: Using generalized item response trees to examine the role of speed on intelligence tests. *Intelligence, 56*, 82–92. https://doi.org/10.1016/j.intell.2016.02.012

Dubey, S. D. (1969). A new derivation of the logistic distribution. *Naval Research Logistics Quarterly, 16*(1), 37–40. https://doi.org/10.1002/nav.3800160103

Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement*, *31*(6), 525–543. https://doi.org/10.1177/0146621606295197

Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series, 2015*(2), 1–17. https://doi.org/10.1002/ets2.12067

Fox, J. P., Entink, R. K., & van der Linden, W. (2007). Modeling of responses and response times with the package cirt. *Journal of Statistical Software, 20*, 1–14. https://doi.org/10.18637/jss.v020.i07

Goldhammer, F. (2015). Measuring ability, speed, or both? Challenges, psychometric solutions, and what can be gained from experimental control. *Measurement: Interdisciplinary Research and Perspectives, 13*(3-4), 133–164. https://doi.org/10.1080/15366367.2015.1100020

Goldstein, S., Princiotta, D., & Naglieri, J. A. (Eds.). (2015). *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*. Springer. https://doi.org/10.1007/978-1-4939-1562-0

Kyllonen, P. C., & Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence, 4*(4), Article 14. https://doi.org/10.3390/jintelligence4040014

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. Information Age Publishing.

Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika, 58*, 445–469. https://doi.org/10.1007/BF02294651

Homepage: https://jem.atu.ac.ir/

McGrane, J. A. (2015). Stevens' forgotten crossroads: The divergent measurement traditions in the physical and psychological sciences from the mid-twentieth century. *Frontiers in Psychology, 6*, Article 431. https://doi.org/10.3389/fpsyg.2015.00431

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Molenaar, D., Bolsinova, M., Rozsa, S., & De Boeck, P. (2016). Response mixture modeling of intraindividual differences in responses and response times to the Hungarian WISC-IV Block Design test. *Journal of Intelligence, 4*(3), Article 10. https://doi.org/10.3390/jintelligence4030010

Papanastasiou, N. (2017). How does scale mean? A critical approach to scale in the study of policy. *Critical Policy Studies, 11*(1), 39–56. https://doi.org/10.1080/19460171.2015.1119052

Partchev, I., De Boeck, P., & Steyer, R. (2013). How much power and speed is measured in this test? *Assessment, 20*(2), 242–252. https://doi.org/10.1177/1073191111411658

Ranger, J. (2013). Modeling responses and response times in personality tests with rating scales. *Psychological Test and Assessment Modeling, 55*(4), 361–387.

Ranger, J., & Kuhn, J. T. (2015). Modeling information accumulation in psychological tests using item response times. *Journal of Educational and Behavioral Statistics, 40*(3), 274–306.

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27*(2), 133–144. https://doi.org/10.1111/j.1745-3984.1990.tb00738.x

Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 187–208). Springer. https://doi.org/10.1007/978-1-4757-2691-6_11

Schnipke, D. L., & Scrams, D. J. (2005). Exploring issues of examinee behavior: Insights gained from response-time analyses. In C. G. Mills, M. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 237–266). Routledge.

Swanson, H. L. (2012). Working memory and reading disabilities: Both phonological and executive processing deficits are important. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 59–88). Psychology Press.

Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities, 42*(3), 260–287. https://doi.org/10.1177/0022219409331958

Homepage: https://jem.atu.ac.ir/

Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods, 6*(2), 181–195. https://doi.org/10.1037/1082-989X.6.2.181

Van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 149–164). American Psychological Association. https://doi.org/10.1037/12343-007

Van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement, 42*(3), 283–302. https://doi.org/10.1111/j.1745-3984.2005.00015.x

Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*(3), 287–308. https://doi.org/10.1007/s11336-006-1478-z

Van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics, 33*(1), 5–20. https://doi.org/10.3102/1076998607302626

Van der Linden, W. J., & Glas, C. A. (2010). Statistical tests of conditional independence between responses and/or response times on test items. *Psychometrika, 75*(1), 120–139. https://doi.org/10.1007/s11336-009-9129-9

Van der Linden, W. J., & Ren, H. (2015). Optimal Bayesian adaptive design for test-item calibration. *Psychometrika, 80*(2), 263–288. https://doi.org/10.1007/s11336-013-9391-8

Verhelst, N. D., Verstralen, H. H., & Jansen, M. G. H. (1997). A logistic model for time-limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 169–185). Springer.

Witelski, T., & Bowen, M. (2015). *Methods of mathematical modelling*. Springer. https://doi.org/10.1007/978-3-319-23042-9

Homepage: https://jem.atu.ac.ir/