

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

Ghasem Keshavarz Gerami 

PhD student of Measurement and Assessment, Allameh Tabataba'i University, Tehran, Iran

Mohammad Reza Flasafi Nejad  *

Associat professor, Measurement and Assessment Dep, Allameh Tabataba'i University, Tehran, Iran

jalil Younesi 

Associat professor, Measurement and Assessment Dep, Allameh Tabataba'i University, Tehran, Iran

Ali Delavar 

professor, Measurement and Assessment Dep, Allameh Tabataba'i University, Tehran, Iran

Zahra Rezaei Ghahroodi 

Associat professor, Statistics Dep, Tehran University, Tehran, Iran

ORCID

Ghasem Keshavarz Gerami



<https://orcid.org/0000-0003-3757-3163/>

Mohammad Reza Flasafi Nejad



<https://orcid.org/0000-0002-3227-102X/>

jalil Younesi



<https://orcid.org/0000-0002-9619-1600/>

Ali Delavar



<https://orcid.org/0000-0002-8404-5980/>

Zahra Rezaei Ghahroodi



<https://orcid.org/0000-0002-9138-8535/>

* Corresponding Author: falsafinejad@yahoo.co.uk

How to Cite: Keshavarz Gerami, Gh., Falsafi Nejad, M. H., Younesi, J., Delavar, A., & Rezaei Ghahroodi, Z. (2025). Application of Response Time Models in Measuring Ability of the Children's Quantitative Reasoning. *Educational Measurement*,

Abstract

There is a long tradition in the ability testing literature, that response speed is as much an indicator of ability as is the correctness of responses to items of increasing difficulty. Therefore, this study seeks to investigate the application of response time model in measuring cognitive ability of the children's quantitative reasoning. To this end, a sample of 502 children from different areas of Tehran was recruited through Multistage cluster random sampling and evaluated with a quantitative reasoning test. To analyze the collected data, modeling response time with the hierarchical model was used. Item and person parameters were also estimated and interpreted. The modeling results revealed that in addition to the response correctness parameters including discrimination coefficient, difficulty coefficient and theta, response Time parameters including Time discrimination coefficient, Time difficulty coefficient and gamma also been estimated. Moreover, the estimated parameters of the response time model compared with the 2-pl IRT model in order to identify their main similarities and differences. Findings revealed that there are some differences in discrimination coefficients and person's thetas, which indicates the power of the response time model. It is worth mentioning that the lowest differences were observed in the difficulty coefficients. Finally, based on the number of parameters, increasing the variance of theta scores, and research background support, the study and application of response time models in measuring children's cognitive abilities were suggested.

Keywords: Response Time models, Measuring cognitive ability, Quantitative reasoning, Item response theory

Extended Abstract**1. 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 better 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.

Research Question(s)

2. 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 and colleagues (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.

3. 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 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 upheld, and inclusion criteria ensured participants had no

neurological disorders or medication use that could influence 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 test demonstrated strong reliability (internal consistency = 0.816, marginal reliability = 0.856). For data analysis, the study applied the hierarchical model of response-time (van der Linden, 2007) using the LNIRT package in R 3.5, alongside SPSS 25 and BILOG 3 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.

4. 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 the highest 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 a more normally distributed estimation of test-takers' abilities and allowed for more sensitive classification, resulting in a much higher number of ability strata (502 compared to 35 in the traditional model).

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 precision in distinguishing among test-takers and emphasized the importance of incorporating response time in test analysis.

5. 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—emerged as an important issue, which the analyses confirmed. A suggested solution is explicit modeling of the relationship between primary and auxiliary variables to reduce bias and improve accuracy. The ability estimates also showed more dispersion in the response-time model, 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.

6. 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 considering response time as a latent variable, more precise information about mental effort and test-taker motivation can be obtained. 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.



کاربرد مدل های زمان-پاسخ در سنجش توانایی استدلال کمی کودکان

دانشجوی دکتری رشته سنجش و اندازه گیری، دانشگاه علامه طباطبائی، تهران، ایران

قاسم کشاورز گرامی 

دانشیار گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی، تهران، ایران

محمدرضا فلسفی نژاد 

دانشیار گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی، تهران، ایران

جلیل یونسی 

استاد ممتاز گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی، تهران، ایران

علی دلاور 

دانشیار گروه آمار دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران

زهره رضایی قهرودی 

چکیده

سنتی قدیمی در ادبیات سنجش توانایی وجود دارد که سرعت پاسخگویی به سؤال همان اندازه نشان دهنده توانایی است که درستی پاسخ به سؤالات، است. بنابراین، این مطالعه به دنبال بررسی کاربرد مدل زمان-پاسخ در سنجش توانایی شناختی استدلال کمی 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 کمی مورد ارزیابی قرار گرفتند. داده‌های جمع‌آوری شده با استفاده از مدل سلسله‌مراتبی زمان-پاسخ تحلیل شده است. همچنین، پارامترهای سؤال و توانایی فرد نیز برآورد و تفسیر شده است. مدل‌بندی 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

مقاله حاضر برگرفته از رساله دکتری رشته سنجش و اندازه گیری دانشگاه علامه طباطبائی است.

* نویسنده مسئول: falsafinejad@yahoo.co.uk

شناسایی شود. یافته‌ها نشان داد که تفاوت‌هایی در ضرایب تمیز و توانایی افراد وجود دارد که نشان‌دهنده قدرت مدل زمان-پاسخ است. شایان‌ذکر است که کمترین تفاوت‌ها در ضریب دشواری مشاهده شد. در نهایت 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 کمی کودکان، مدل سؤال-پاسخ

مقدمه

تعریف و مدل‌بندی زمان در اندازه‌گیری‌های روان‌شناختی به‌عنوان یک پارامتر آگاهی دهنده، جدیدترین استقراض مهم روان‌سنجی از فیزیک است. از سال ۱۹۴۶ که استیونز^۱ مقابل هیات فیزیکدانان انگلیسی فرگوسن^۲ 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^۳ 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 مورد توجه روان‌شناسان بوده است. گاهی زمان پاسخگویی در سنجش توانایی به شکل ضمنی وارد می‌شد، مانند آزمون‌های محدودیت زمانی، گاهی به طور صریح در آزمون‌های به‌اصطلاح سرعتی و گاهی به‌عنوان متغیر اصلی و خود توانایی، مانند زمان واکنش^۴ یا آزمون‌های پردازش اطلاعات^۵، مورد توجه قرار می‌گرفت. می‌توان گفت پیش‌گامان سنجش و اندازه‌گیری توانایی‌ها، گالتون و اسپیرمن^۶ 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 غلط بودن پاسخ است و دومین نکته، مدت زمانی 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. به‌عنوان مثال، پاسخ‌کند یک آزمودنی ممکن است منعکس‌کننده سرعت پردازش کند او یا دقت کم او باشد. آزمودنی

-
1. Stanley Smith Stevens
 2. the British Ferguson Committee
 3. Response Time Models
 4. Reaction Time
 5. processing speed
 6. Galton & Spearman

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 زمان کافی برای پردازش، یا سردرگم شدن هنگام پاسخگویی و اتمام آزمون می‌توانند منجر به پاسخ اشتباه شود. اگر چه به‌طور کلی بین سرعت و دقت در پاسخگویی تعادل وجود دارد، اما این تعادل بین سؤالات یا افراد ثابت نیست. ممکن است پاسخ آزمودنی در شرایط تحت فشار 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^۱ متوجه شدند که سرعت پاسخ افراد نسبت به محرک‌های مختلف، نشانه‌ای از نوعی برتری در عملکرد مغز و توانایی‌های فردی است. در نتیجه، 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 وقت گذاشتن برای کنش‌های پیچیده بیهوده است (Anastasi & Urbina, 1997).

این نکته اهمیت دارد که بین زمان واکنش^۲ و زمان پاسخ^۳ 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^۴ در اندازه‌گیری توانایی‌ها مدل‌سازی می‌کند. تمایز سطح سرعت، ابعاد مختلف سرعت و سطوح آن در چارچوب توانایی‌های شناختی، تعادل سرعت-دقت، روش‌های تجزیه و تحلیل به‌ویژه نظریه سؤال-پاسخ و دیگر کاربردهای زمان، مباحث ویژه و کاربردی در اندازه‌گیری‌های شناختی هستند. از جمله پیشرفت‌های

1. Cattle & Galton
2. Reaction Time (RT)
3. Response Time (RT)
4. dimension

مدل‌سازی پردازش اطلاعات شناختی، مدل‌های زمان پاسخ، مدل انتشار^۱ 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؛ به عنوان مثال برنینگر و ولف^۲ (۲۰۱۶) در مدل خود برای تبیین نواقص شناختی کودکان دارای اختلال یادگیری 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 (Papanastasiou, 2017؛ Swanson et al., 2009) 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 اندوختن 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. در واقع تمرکز اصلی بیشتر بر درستی پاسخ معطوف بوده است، همان‌طور که گلدشتاین^۳ و همکاران (۲۰۱۵) می‌گویند: «هنوز همان سناریویی مطرح می‌شود که برای اولین بار در سال ۱۹۴۲

1. diffusion

2. Berninger & Wolf

3. Goldstein

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 نادیده گرفتن زمان پاسخ این است که بیشتر مدل‌های «نظریه سؤال-پاسخ»^۱ برای آزمون‌های قدرت^۲ محض تهیه شده‌اند که متفاوت با آزمون‌های سرعت^۳ است. با این وجود، 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. قابل توجه است که این حقیقت ساده، مدت‌های طولانی توسط بیشتر روان‌سنج‌ها نادیده گرفته شده است (Goldstein et al., 2015).

برای وارد کردن 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^۴ (۱۹۹۷) و ورهلست^۵ 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 به توجه به نوع استفاده از زمان پاسخ در برآورد صفت مکنون مطرح شده است. این سه رویکرد شامل در نظر گرفتن زمان پاسخ به عنوان 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 رابطه زیر به دست می‌آید:

$$P_i(\theta_j) = [1 + \exp(\theta_j - \ln \tau_j - b_i)]^{-\pi_i}$$

1. Item-Response Theory (IRT)

2. pure power test

3. speed tests

4. Roskam

5. Verhelst

در رابطه بالا b_i پارامتر ضریب دشوار سوال i ام، θ_j پارامتر توانایی فرد j ام، τ_j پارامتر سرعت فرد j ام و π_i پارامتر شکل وابسته به سوال 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، به عنوان یک رگرسیون وارد مدل می شود.

$$P_i(\theta_j) = [1 + \exp(\theta_j + \ln t_{ij} - b_i)]^{-1}$$

پیشینه پژوهش

دی بوک و جئون^۱ (۲۰۱۹) با مروری بر مدل های زمان-پاسخ در فرایندهای آزمون های شناختی، کاربرد مدل های زمان-پاسخ را برای بررسی فرایندهای شناختی معرفی کردند. آنها، مدل های زمان-پاسخ را به چهار دسته متمایز تقسیم و معرفی کردند. کیلون و زو^۲ (۲۰۱۶) استفاده از زمان پاسخگویی را برای سنجش توانایی شناختی، با مرور برخی از ادبیات در مورد زمان پاسخ معرفی کردند و نشان دادند که این مدل ها 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 توانایی های شناختی، مبادله دقت و صحت، رویکردهای پرداختن به مبادله سرعت-دقت، روش های تحلیل، به ویژه نظریه سؤال-پاسخ، مدل های زمان-پاسخ از روان شناسی شناختی (عملکرد گاوسی، و مدل انتشار) و سایر کاربردهای زمان پاسخ در آزمون علاوه بر اندازه گیری توانایی بحث کرده اند. پارتچف^۳ و همکاران (۲۰۱۳) با پرداختن به این موضوع که «در آزمون چقدر قدرت و چقدر سرعت اندازه گیری می شود؟» این مسئله قدیمی در ارزیابی روان شناسی را بازنگری و شفاف سازی کرده اند. اگر چه در ادبیات و پیشینه پژوهش، اشاره شده است که استفاده از مدل های زمان-پاسخ می تواند آگاهی دهنده باشد اما همچنان پیشرفت هر دو علم روان سنجی و علوم شناختی در تئوری و کاربرد نیازمند پژوهش و پاسخگویی به سؤالات در زمینه کاربرد مدل های زمان-پاسخ است.

1. De Boeck & Jeon
2. Kyllonen & Zu
3. Partchev

روش

به منظور 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 قابل استفاده بودن پارامترهایی که در نتیجه 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 پیشنهادات مطرح شده، داده‌ها از حجم نمونه ۵۰۲ نفر جمع‌آوری و تحلیل شد. در این راستا ۱۹ منطقه آموزش 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. در مجموع، 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. در نهایت 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 کمی که مبتنی بر سنجش کامپیوتری است، استفاده شد. این آزمون توانایی‌های شناختی استدلال کمی و محاسباتی فرد را مورد سنجش 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^۱، پیل^۲، کن‌تب^۳ و وینا^۴ وجود دارد. در این آزمون فرد باید عدد بزرگ‌تر را انتخاب کند. در صورتی که دو عدد مساوی باشند، بر روی دکمه مساوی کلیک می‌کند. در مراحل بالاتر به جای 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^۵ ۰/۸۵۶ برآورد شده است.

برای تجزیه و تحلیل مدل سلسله‌مراتبی زمان-پاسخ (van der Linden, 2007) از بسته 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 مبنای 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. از آنجا که 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 مدل‌بندی زمان-پاسخ و درستی پاسخ وجود دارد: (۱) اضافه کردن پارامتر زمان به مدل‌های نظریه سؤال-پاسخ (Roskam, 1997)؛ (۲) مدل‌بندی مجزای زمان-پاسخ از درستی پاسخ (Maris, 1993) و (۳) مدل‌بندی زمان-پاسخ و درستی پاسخ به صورت سلسله‌مراتبی (Van der Linden, 2007).

در مدل سلسله‌مراتبی در اولین مرتبه به گونه‌ای 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

-
1. Lumosity
 2. PEBL
 3. CANTAB
 4. VIENNA
 5. marginal

به سؤال‌ها است. در اولین سطح از مدل‌بندی، 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 مدل‌بندی سلسله‌مراتبی می‌شود به‌گونه‌ای که ارتباط بین سرعت و درستی در سطح بالاتری از مدل‌بندی تعریف می‌شود.

چون کران پایین مقادیر زمان پاسخ صفر است، بنابراین از توزیع لگ‌نرمال برای مدل‌بندی زمان پاسخ‌ها استفاده می‌شود. در واقع 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 لیدن^۱ (۲۰۰۷) که فاکس^۲ و همکاران (۲۰۰۷) مدل‌بندی توأم در بسته LNIRT در نرم‌افزار R استفاده کرده‌اند، معرفی می‌گردد. همان‌گونه که قبلاً اشاره شد، مدل‌بندی در دو سطح انجام می‌شود. سطح اول «مرتب‌بندی» است که در ادامه معرفی می‌شود.

مدل‌بندی درستی پاسخ RA: ماتریس پاسخ‌ها یک ماتریس $N \times K$ است که N تعداد افراد آزمون‌دهنده و k تعداد سؤال‌ها است. الگوهای پاسخ توسط آزمون‌دهنده و سؤال‌ها مشخص می‌شوند. در یک مدل IRT دو پارامتری، احتمال پاسخ صحیح با معلوم بودن پارامترهای فرد و سؤال تعریف می‌شود (Lord & Novick, 1968). احتمال پاسخ صحیح آزمون‌دهنده λ_k 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 (θ_i)، توانایی آزمون‌دهنده λ_k و a_k و b_k به‌ترتیب پارامتر ضریب تمیز و ضریب دشواری سؤال k ام هستند. Φ تابع توزیع تجمعی نرمال است. به‌جای تابع توزیع تجمعی نرمال می‌توان از تابع توزیع تجمعی 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:

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

1. Van der Linden

2. Fox

در فرمول فوق η_i و $\tilde{b}_k$ بر اساس مقیاس مشابه تعریف می‌شوند. در بسته LNIRT هر دو روش پارامترسازی قابل اجرا است. لازم به ذکر است که پارامتر ضریب دشواری سؤال در دو معادله فوق به روش مستقیم قابل مقایسه 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 به صورت زیر است:

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

مدل‌بندی زمان-پاسخ RT: متغیر پاسخ لگاریتم زمان-پاسخ را در یک ماتریس $K \times N$ با نام RT در نظر بگیرید. زمانی که سرعت کار ثابت فرض شود، در واقع هر آزمون‌دهنده با سرعت ثابتی کار می‌کند که آن را با نماد ζ_i نمایش می‌دهیم. زمان مورد نیاز برای پاسخ‌دهی به یک سؤال نیز به پارامترهای مشخص هر سؤال بستگی دارد که آنها را با نماد λ_k و Φ_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 به صورت زیر تعریف می‌شود:

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

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

به‌طور کلی پارامتر شدت زمان λ_k معرف متوسط زمان مورد نیاز برای کامل کردن سؤال (در مقیاس لگاریتم)، پارامتر سرعت ζ_i ، معرف سرعت کار کردن آزمون‌دهنده i ام و پارامتر تشخیص زمان، Φ_k ، معرف اثر ویژه-سؤال سرعت کارکردن روی زمان-پاسخ است. افزایش پارامتر شدت زمان λ_k منجر به جابجایی مثبت مکان توزیع زمان سؤال k ام می‌شود. به همین ترتیب، افزایش پارامتر سرعت ζ_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:

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

در فرمول فوق پارامتر تشخیص زمان Φ_k بر حسب $(\tilde{\lambda}_k - \zeta_i)$ عمل می کند. در این مدل، پارامتر Φ_k به عنوان پارامتر شیب سرعت، حساسیت سؤال برای سطوح مختلف سرعت آزمون شوندگان را مدل بندی می کند. عبارت خطای ϵ_{ik} ، تغییرات زمان-پاسخ به دلیل رفتار تصادفی یک آزمون دهنده را مدل بندی می کند.

سطح دوم «مرتبه دوم و سوم: مدل های ساختاری پارامتر فرد و سؤال» است که در ادامه معرفی می شود:

پارامترهای افراد: توزیع نرمال دومتغیره برای پارامتر سرعت و توانایی به صورت زیر مدل بندی می شود:

$$(\eta_i, \zeta_i) \sim N_2(\mu_p - \Sigma_p)$$

که در آن μ_p و Σ_p برابر است با:

$$\mu_p = (\mu_\eta - \mu_\zeta)$$

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

کوواریانس بین پارامترهای فرد با ρ معرفی می شود. مدل دوسطحی سرعت و توانایی، برای معرفی توزیع جامعه آزمون دهندگان معرفی می شود. با توجه به اینکه از روش بیزی استفاده می شود، توزیع پیشین برای ماتریس واریانس-کوواریانس Σ_p توزیع ویشارت معکوس با درجه آزادی η_p و V_p در نظر گرفته می شود:

$$\Sigma_p \sim \text{Iw}(\eta_p, V_p)$$

پارامترهای سؤال: برای پارامترهای سؤال، توزیع نرمال چندمتغیره به صورت زیر در نظر گرفته می شود:

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

در آن Σ_I برابر است با:

$$\Sigma_I = \begin{pmatrix} \Sigma_{a,b} & \Sigma_{(a,b)(\eta,\lambda)} \\ \Sigma_{(\eta,\lambda)(a,b)} & \Sigma_{\eta,\lambda} \end{pmatrix}$$

$$= \begin{bmatrix} \sigma_a^2 & \sigma_{a,b} & \sigma_{a,\Phi} & \sigma_{a,\lambda} \\ \sigma_{b,a} & \sigma_b^2 & \sigma_{b,\Phi} & \sigma_{b,\lambda} \\ \sigma_{\Phi,a} & \sigma_{\Phi,b} & \sigma_{\Phi}^2 & \sigma_{\Phi,\lambda} \\ \sigma_{\lambda,a} & \sigma_{\lambda,b} & \sigma_{\lambda,\Phi} & \sigma_{\lambda}^2 \end{bmatrix}$$

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

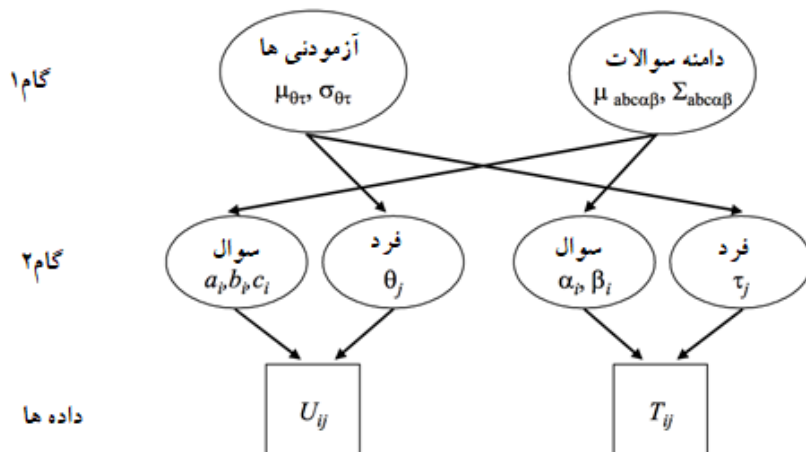
$$\Sigma_I \sim IW_{\mathbb{I}_I}(V_I^{-1})$$

و

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

$\mathbb{I}_I$ و V_I درجه‌های آزادی ماتریس توزیع ویشارت معکوس، μ_0 میانگین توزیع پیشین، k تعداد اندازه‌گیری‌های پیشین که مقدار پیش فرض آن ۱ است. از توزیع پیشین بتا برای پارامتر حدس C_k استفاده می‌شود که مقایر ابر پارامتری‌های آن ۲۰ و ۸۰ است. این انتخاب پارامترها منجر به نسبت حدس پیشین $\frac{1}{5}$ با انحراف استاندارد ۰/۰۴ می‌شود.

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



مطابق شکل ۱ و همان گونه که در بالا ذکر شد، در مدل سلسله مراتبی (Van der Linden, 2007) معرفی کرده است، فرایند برآورد پارامترها و توانایی به صورت سلسله مراتبی و در دو مرحله انجام می شود. در این مدل، τ_p سرعت p امین آزمون دهنده به $p = 1, \dots, P$ را نشان می دهد. فرض می شود پارامترهایی برای سرعت های زمانی وجود دارد که t_{pi} ، زمان پاسخ مشاهده شده p امین آزمون دهنده به i امین سؤال برای $i = 1, \dots, I$ است. به طور معادل، در این مدل فرض می شود زمان پاسخ مشاهده شده آزمون دهنده، ناشناخته و اثر سؤال مجزا است. از آنجایی که توزیع زمان پاسخ چوله به راست است، استفاده از لگاریتم های زمان به جای زمان، مرسوم است. از طرف دیگر ماهیت زمان پاسخ ها تصادفی است. در نتیجه $\ln t_{pi}$ از توزیع نرمال با میانگین $\tau_p - \beta_i$ و واریانس α_i^{-2} پیروی می کند. توزیع مشابه می تواند توسط تابع چگالی لگ نرمال باشد. این موضوع نشان می دهد به جای لگاریتم، متغیر توزیع شده به طور طبیعی، نمایشی است.

یافته ها

در این بخش، ابتدا به ارائه آمار توصیفی و سپس نتایج مدل سلسله مراتبی زمان-پاسخ و نظریه سؤال-پاسخ پرداخته می شود. در جدول ۱، شاخص های توصیفی سؤالات محاسبه و گزارش شده است. برای این منظور میانگین پاسخ صحیح به سؤال، واریانس سؤال، متوسط زمان پاسخ گویی به هر سؤال (بر حسب ثانیه)، کمترین و بیشترین زمان ثبت شده برای هر سؤال گزارش شده است. نتایج نشان می دهد 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 پاسخ گویی بالایی دارند.

جدول ۱. شاخص های توصیفی سؤالات استدلال کمی

شماره سؤال	میانگین پاسخ صحیح	واریانس پاسخ صحیح	میانگین زمان پاسخ گویی	واریانس زمان پاسخ گویی	حداقل زمان پاسخ گویی	حداکثر زمان پاسخ گویی
۱	۰/۸۸	۰/۱۰۸	۱/۹۳	۰/۷۴۰	۰/۳۲	۴/۰۵
۲	۰/۹۳	۰/۰۶۵	۱/۵۱	۰/۵۷۷	۰/۵۰	۴/۰۲
۳	۰/۹۵	۰/۰۴۷	۱/۳۲	۰/۴۴۳	۰/۲۰	۴/۰۲

نام خانوادگی نویسنده اول و دوم (بیش از دو نویسنده نام خانوادگی نویسنده اول و همکاران | ۲۱

شماره سؤال	میانگین پاسخ صحیح	واریانس پاسخ صحیح	میانگین زمان پاسخ‌گویی	واریانس زمان پاسخ‌گویی	حداقل زمان پاسخ‌گویی	حداکثر زمان پاسخ‌گویی
۴	۰/۹۶	۰/۰۳۶	۱/۲۴	۰/۳۷۳	۰/۴۰	۴/۰۱
۵	۰/۹۸	۰/۰۲۱	۱/۱۹	۰/۳۲۷	۰/۵۰	۴/۰۱
۶	۰/۸۷	۰/۱۱۱	۱/۵۱	۰/۴۴۳	۰/۱۰	۴/۰۱
۷	۰/۹۳	۰/۰۶۳	۱/۴۰	۰/۵۳۶	۰/۲۰	۶/۰۱
۸	۰/۶۳	۰/۲۳۵	۲/۳۶	۱/۹۲۳	۰/۵۰	۶/۰۱
۹	۰/۹۵	۰/۰۵۱	۱/۴۷	۰/۸۱۶	۰/۴۰	۶/۰۱
۱۰	۰/۹۵	۰/۰۵۱	۱/۸۴	۰/۸۷۱	۰/۲۰	۶/۰۱
۱۱	۰/۹۴	۰/۰۵۸	۱/۶۰	۰/۸۹۷	۰/۵۰	۶/۰۲
۱۲	۰/۹۱	۰/۰۸۳	۱/۸۴	۰/۸۶۸	۰/۱۰	۶/۰۲
۱۳	۰/۶۶	۰/۲۲۵	۳/۰۲	۳/۰۰۵	۰/۵۰	۶/۰۴
۱۴	۰/۹۵	۰/۰۴۶	۱/۵۹	۰/۷۴۵	۰/۱۰	۶/۰۱
۱۵	۰/۷۱	۰/۲۰۷	۲/۹۵	۲/۳۶۰	۰/۵۰	۶/۰۲
۱۶	۰/۸۵	۰/۱۲۵	۱/۶۱	۰/۶۲۵	۰/۱۰	۶/۰۱
۱۷	۰/۷۳	۰/۱۹۸	۲/۲۷	۱/۲۵۴	۰/۲۰	۶/۰۳
۱۸	۰/۸۴	۰/۱۳۷	۲/۱۸	۱/۲۵۵	۰/۲۰	۸/۰۱
۱۹	۰/۶۱	۰/۲۳۸	۳/۰۰	۲/۵۶۱	۰/۱۰	۸/۰۱
۲۰	۰/۴۸	۰/۲۵۰	۳/۲۲	۳/۳۰۳	۰/۱۰	۸/۰۲
۲۱	۰/۷۵	۰/۱۸۵	۲/۹۶	۳/۴۳۰	۰/۵۰	۸/۰۲
۲۲	۰/۸۱	۰/۱۵۱	۲/۶۳	۲/۴۰۱	۰/۲۰	۸/۰۱
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

شماره سؤال	میانگین پاسخ صحیح	واریانس پاسخ صحیح	میانگین زمان پاسخ گویی	واریانس زمان پاسخ گویی	حداقل زمان پاسخ گویی	حداکثر زمان پاسخ گویی
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

همان گونه که اشاره شد، این اطلاعات بر اساس ۵۰۲ الگوی پاسخ معتبر گزارش شده است. به منظور بررسی فرضیه تک بعدی بودن پرسشنامه که از الزامات نظریه های جدید اندازه گیری است، از تحلیل نظریه سؤال-پاسخ چند بعدی^۱ استفاده شده است. بر این اساس، مدل با فرض تک بعدی بودن، مورد تجزیه و تحلیل قرار گرفت تا با مقایسه مقدار برازش شده با ملاک های RMSEA، CFI و TLI، آزمون تک بعدی بودن پرسشنامه استنباط شود. نتایج این تحلیل در جدول ۲ گزارش شده است.

نام خانوادگی نویسنده اول و دوم (بیش از دو نویسنده نام خانوادگی نویسنده اول و همکاران | ۲۳

جدول ۲. شاخص‌های برازش مدل تک بعدی بر اساس محاسبه سؤال-پاسخ چند بعدی

شاخص	لگاریتم درست‌نمایی	درجه آزادی	معناداری	RMSEA	CFI	TLI
میزان مطلوب	-	-	$>0/05$	$<0/08$	$0/9-1$	$0/9-1$
مقدار محاسبه‌شده	-۱۲۰۴	۱۰۸۰	$0/001$	$0/053$	$0/92$	$0/91$

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

در ادامه، نتایج مدل‌بندی سلسله‌مراتبی زمان پاسخ ارائه شده است. به منظور 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 تمیز 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 تمیز منفی نشان می‌دهند سؤال عملکرد نامناسبی داشته است و ضرایب مثبت تأیید می‌کنند سؤال عملکرد مناسبی در سنجش صفت مورد نظر داشته است.

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

سؤال	مدل سؤال-پاسخ		مدل زمان-پاسخ			
	ضریب تمیز	ضریب دشواری	ضریب تمیز	ضریب دشواری	ضریب تمیز	ضریب دشواری
سؤال ۱	$0/37$	$-3/28$	$0/87$	$-1/11$	$0/201$	$0/245$
سؤال ۲	$0/62$	$-2/85$	$1/02$	$-1/43$	$0/298$	$0/134$
سؤال ۳	$0/51$	$-3/79$	$0/78$	$-1/52$	$0/319$	$0/080$
سؤال ۴	$0/76$	$-3/11$	$0/92$	$-1/65$	$0/365$	$0/051$
سؤال ۵	$0/59$	$-4/30$	$0/62$	$-1/74$	$0/237$	$0/037$
سؤال ۶	$0/43$	$-2/90$	$0/76$	$-1/08$	$0/297$	$0/142$

مدل زمان-پاسخ				مدل سؤال-پاسخ		سؤال
ضریب دشواری	ضریب تمییز	ضریب	ضریب	ضریب	ضریب	
زمان	زمان	دشواری	تمییز	دشواری	تمییز	
۰/۱۰۱	۰/۴۱۸	-۱/۴۵	۱/۰۲	-۳/۰۵	۰/۵۸	سؤال ۷
۰/۲۹۲	۰/۶۵۵	-۰/۳۷	۱/۸۱	-۰/۴۷	۰/۸۰	سؤال ۸
۰/۱۰۷	۰/۴۵۷	-۱/۵۲	۰/۸۷	-۳/۲۸	۰/۶۰	سؤال ۹
۰/۲۱۲	۰/۴۶۲	-۱/۵۱	۰/۸۳	-۲/۹۴	۰/۵۶	سؤال ۱۰
۰/۱۴۷	۰/۴۴۲	-۱/۵۴	۱/۲۲	-۲/۵۳	۰/۷۹	سؤال ۱۱
۰/۲۱۵	۰/۴۱۶	-۱/۴۱	۱/۴۹	-۱/۸۷	۰/۹۲	سؤال ۱۲
۰/۳۹۰	۱/۰۰۷	-۰/۳۹	۰/۵۸	۱/۳۶	۱/۲۷	سؤال ۱۳
۰/۱۵۱	۰/۴۹۴	-۱/۵۴	۰/۷۴	-۲/۹۵	۰/۴۵	سؤال ۱۴
۰/۳۹۵	۰/۸۸۲	-۰/۵۶	۱/۰۵	۰/۳۰	۰/۹۹	سؤال ۱۵
۰/۱۶۴	۰/۴۶۸	-۱/۰۱	۰/۶۹	۰/۹۵	۲/۴۷	سؤال ۱۶
۰/۳۰۲	۰/۶۰۶	-۰/۶۴	۱/۳۲	-۱/۳۲	۰/۵۰	سؤال ۱۷
۰/۲۸۳	۰/۶۶۷	-۱/۰۸	۱/۶۱	-۱/۶۸	۰/۷۰	سؤال ۱۸
۰/۳۹۹	۰/۹۴۲	-۰/۳۲	۱/۶۷	-۰/۲۱	۰/۷۸	سؤال ۱۹
۰/۴۱۶	۱/۲۰۸	۰/۰۵	۳/۸۳	۱/۳۳	۰/۴۷	سؤال ۲۰
۰/۳۶۹	۱/۱۹۳	-۰/۸۲	۱/۸۳	-۱/۰۱	۰/۸۵	سؤال ۲۱
۰/۳۴۰	۰/۸۷۸	-۰/۹۲	۱/۱۴	۰/۶۸	۲/۸۸	سؤال ۲۲
۰/۴۳۲	۱/۴۳۲	۰/۴۳	۲/۱۱	۰/۵۱	۱/۰۲	سؤال ۲۳
۰/۳۹۶	۱/۳۶۸	-۳۰	۱/۸۶	-۰/۰۸	۱/۰۱	سؤال ۲۴
۰/۴۵۱	۱/۶۶۰	۰/۰۲	۰/۵۷	۱/۴۹	۱/۹۴	سؤال ۲۵
۰/۳۹۰	۱/۵۲۶	-۰/۳۲	۱/۴۳	-۰/۰۵	۰/۷۹	سؤال ۲۶
۰/۴۰۵	۱/۵۵۲	-۰/۶۶	۱/۲۱	-۱/۳۰	۰/۵۱	سؤال ۲۷
۰/۴۱۴	۱/۵۷۳	-۰/۷۰	۱/۵۶	-۰/۰۹	۱/۲۳	سؤال ۲۸
۰/۴۱۶	۱/۷۵۳	-۰/۲۰	۱/۳۳	۰/۷۳	۲/۱۰	سؤال ۲۹
۰/۳۶۸	۱/۵۸۹	-۰/۳۲	۰/۹۳	-۰/۸۰	۰/۳۹	سؤال ۳۰
۰/۴۲۱	۱/۷۱۳	-۰/۳۰	۰/۷۷	۱/۱۸	۲/۰۶	سؤال ۳۱
۰/۴۶۰	۱/۸۸۹	۰/۸۳	۱/۸۵	۱/۱۰	۱/۴۹	سؤال ۳۲

نام خانوادگی نویسنده اول و دوم (بیش از دو نویسنده نام خانوادگی نویسنده اول و همکاران | ۲۵

مدل سؤال-پاسخ		مدل زمان-پاسخ				
ضرب	ضرب	ضرب	ضرب	ضرب	ضرب	سؤال
دشواری	تمییز	دشواری	تمییز	دشواری	تمییز	
زمان	زمان	زمان	زمان	زمان	زمان	
۰/۴۲۳	۱/۸۳۰	-۰/۳۲	۰/۷۵	۱/۰۷	۲/۱۲	سؤال ۳۳
۰/۴۲۶	۱/۸۹۰	۰/۵۷	۱/۷۲	۰/۸۲	۰/۸۳	سؤال ۳۴
۰/۳۸۱	۱/۸۶۷	۰/۳۷	۲/۰۴	۰/۵۴	۱/۲۳	سؤال ۳۵
۰/۴۵۸	۱/۹۴۶	-۰/۱۰	۱/۱۴	۰/۸۵	۱/۳۳	سؤال ۳۶
۰/۴۳۶	۱/۷۹۱	۰/۳۷	۱/۸۲	۰/۶۹	۱/۰۱	سؤال ۳۷
۰/۴۱۸	۱/۹۰۸	۰/۱۴	۰/۱۸	-۳/۵۰	-۰/۰۱	سؤال ۳۸
۰/۴۱۳	۱/۹۹۵	-۰/۰۷	۱/۳۳	۰/۷۰	۲/۱۶	سؤال ۳۹
۰/۴۰۵	۲/۰۲۰	۰/۱۲	۰/۴۷	۱/۶۱	۲/۴۹	سؤال ۴۰
۰/۳۹۵	۱/۸۹۴	-۰/۱۱	۱/۱۲	۰/۸۲	۱/۳۶	سؤال ۴۱
۰/۳۶۶	۱/۷۹۹	-۰/۴۲	۱/۱۱	۰/۸۱	۲/۳۳	سؤال ۴۲
۰/۳۶۶	۱/۹۳۶	-۰/۱۳	۰/۹۵	-۰/۳۲	۰/۳۵	سؤال ۴۳
۰/۳۹۰	۱/۹۵۰	۰/۴۷	۰/۸۱	۱/۶۷	۲/۱۵	سؤال ۴۴
۰/۳۴۵	۱/۷۶۵	-۰/۰۲	۱/۰۴	۰/۷۷	۰/۷۶	سؤال ۴۵
۰/۳۵۷	۱/۸۳۷	۰/۴۰	۰/۶۵	۱/۷۸	۰/۲۲	سؤال ۴۶
۰/۳۷۱	۱/۷۷۲	۰/۵۶	۰/۸۵	۱/۷۲	۰/۳۲	سؤال ۴۷
۰/۳۶۰	۱/۷۴۹	۰/۴۹	۰/۸۳	۱/۴۹	۰/۳۳	سؤال ۴۸

بر اساس نتایج جدول ۳، مشابهت ها و تفاوت هایی در برآورد پارامترها و دامنه تغییرات آنها وجود دارد. به عنوان مثال 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 می شود سؤالات در یک پارامتر بهترین در مدل خود هستند. سؤال ۴۷ در هر دو مدل ضریب دشواری بالایی دارد. به منظور بررسی دقیق تر رابطه بین پارامترهای برآورد شده از مدل زمان-پاسخ با مدل سؤال-پاسخ، ضرایب همبستگی بین پارامترهای مدل زمان-پاسخ و پارامترهای مدل سؤال-پاسخ، در جدول ۴ گزارش شده است.

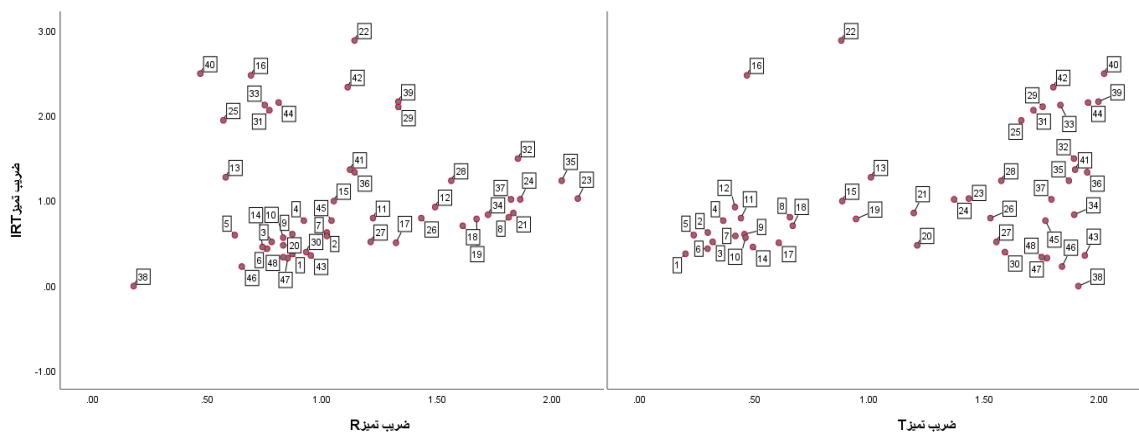
جدول ۴. ضرایب همبستگی بین پارامترهای مدل زمان-پاسخ با پارامترهای مدل سؤال-پاسخ

پارامترها	ضریب تمیز	ضریب دشواری	ضریب تمیز	ضریب دشواری	ضریب تمیز	ضریب دشواری
	R	R	T	T	IRT	IRT
ضریب تمیز R	۱					
ضریب دشواری R	۰/۲۱	۱				
ضریب تمیز T	۰/۱۰	۰/۸۷**	۱			
ضریب دشواری T	۰/۲۷	۰/۸۵**	۰/۸۵**	۱		
ضریب تمیز IRT	۰/۰۲	۰/۱۸	۰/۳۲*	۰/۳۲*	۱	
ضریب دشواری IRT	۰/۱۵	۰/۸۱**	۰/۷۵**	۰/۷۹**	۰/۵۴**	۱

** $p \leq 0/01$

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 تمیز زمان در نظریه زمان-پاسخ نیز همبستگی ۰/۳۲ با ضریب تمیز در مدل سؤال-پاسخ دارد. همچنین مشاهده می شود بین ضریب تمیز زمان با ضریب دشواری زمان با ضریب تمیز سؤال-پاسخ رابطه معنی دار شده است. در شکل های ۲ و ۳، نمودار پراکندگی هم تغییری بین پارامترهای مدل سلسله مراتبی زمان-پاسخ با مدل سؤال-پاسخ گزارش شده است.

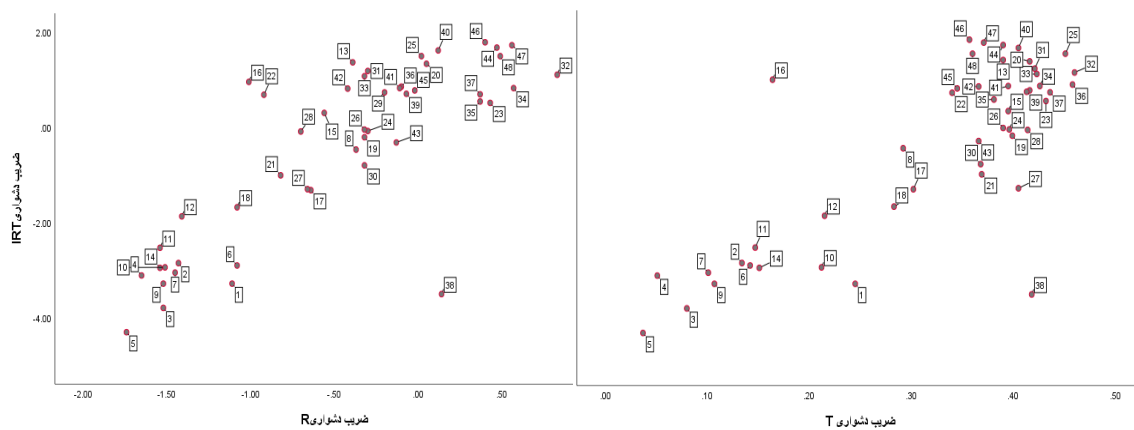
شکل ۲. نمودار رابطه بین ضریب تمیز زمان و زمان پاسخ با ضریب تمیز مدل سؤال-پاسخ



نتایج شکل ۲ نشان می‌دهد بین ضرایب تمیز حاصل از پاسخ مدل زمان-پاسخ و ضرایب تمیز سؤال-پاسخ ارتباط و هم‌تغیری ضعیفی وجود دارد. در دامنه ۰/۵۰ تا ۱/۵ مدل زمان-پاسخ، تفاوت مقادیر ضریب تمیز مدل‌ها بیشتر است. به عبارت دیگر این نمودار نشان می‌دهد در مواردی که سؤالات از نظر 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



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	۶۹/۴۰	۱	۶۹/۴۰	۶۹/۹۶	۰/۰۰۰
ضریب تمیز	خطا	۴۶/۶۳	۴۷	۰/۹۹		

متغیر	منابع تغییرات	SS	df	MS	F	P
پارامتر	مدل	۱۵/۱۵	۱	۱۵/۱۵	۱۰/۱۰	۰/۰۰۳
ضریب دشواری	خطا	۷۰/۴۸	۴۷	۱/۵۰		

جدول ۵، نتایج مقایسه وجود یا عدم وجود تفاوت پارامترهای دو مدل با زمان پاسخ و بدون زمان پاسخ به تفکیک ضریب تمیز و ضریب دشواری را نشان می دهد. نتایج نشان می دهد تفاوت پارامتر ضریب تمیز و پارامتر ضریب دشواری معنی دار است. به عبارت دیگر بین دو روش تفاوت معنی داری در برآورد پارامترها وجود دارد. این معنی داری در کنار طبقات بیشتری که مدل زمان-پاسخ ایجاد کرده است قابل تفسیر خواهد بود. در جدول ۶، شاخص های مرکزی و پراکندگی توانایی های برآورد شده بر اساس مدل زمان-پاسخ گزارش شده است.

جدول ۶. توصیف توانایی های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ

شاخص ها	تتا IRT	تتا RT	گاما RT
میانگین	۰/۲۱۸۷	۰/۰۰	۰/۰۰۰
واریانس	۰/۱۳۵	۰/۱۳۴	۰/۰۳۴
کمینه	-۱/۱۸	-۱/۱۲	۰/۷۶
پیشینه	۱/۵۶	۱/۰۶	-۰/۳۲
تعداد نفرات	۵۰۲	۵۰۲	۰/۴۴

نتایج جدول ۶ نشان می دهد، 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 می دهد میانگین پارامتر توانایی بر اساس زمان پاسخ شکل نسبتاً قرینه ای دارد و به توزیع نرمال نزدیک تر 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. به نوعی در توانایی که بر اساس اطلاعات زمان تقویت شده است افراد در مقیاس نسبت به میانگین با حساسیت بیشتری طبقه بندی شده اند و طبقات بیشتری تشکیل شده است، همچنین توزیع به نرمال نزدیک تر است.

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

متغیر	منابع تغییر	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 ($p < 0.001$) 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 تولید شده 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.

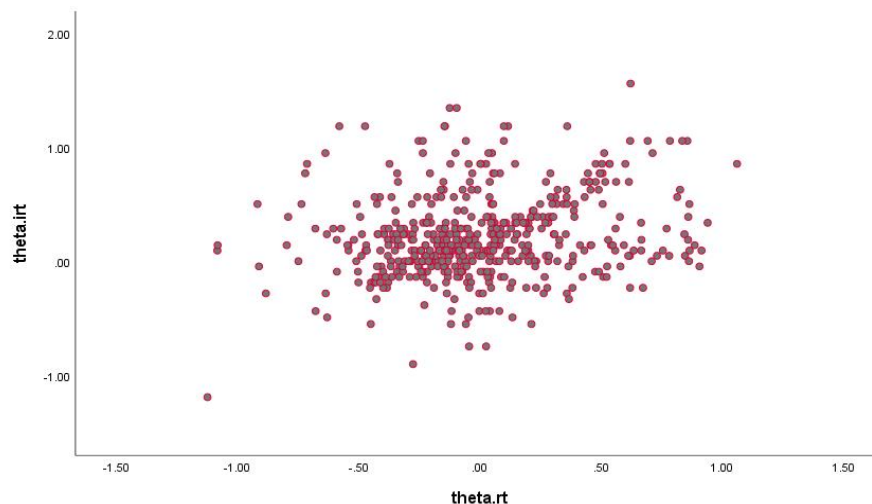
جدول ۸. مقایسه توانایی‌های برآورد شده در دو مدل زمان پاسخ و بدون زمان پاسخ

کد	رتبه در IRT	رتبه در RT	تفاوت	کد	رتبه در IRT	رتبه در RT	تفاوت
۱۵۰	۳۰	۴۷۸	۴۴۸	۶۸	۴۹۴	۲۲	-۴۷۲
۱۹۸	۳۴	۴۷۱	۴۳۷	۳۴۹	۴۸۰	۱۵	-۴۶۵
۲۴۲	۲۰	۴۲۳	۴۰۳	۱۱۶	۴۹۸	۱۵	-۴۶۳
۷۲	۵۶	۴۵۴	۳۹۸	۱۰۷	۴۶۷	۱۲	-۴۵۵
۶۴	۵۵	۴۵۱	۳۹۶	۳۳۷	۴۵۹	۱۱	-۴۴۸
۲۰۵	۲۴	۴۱۹	۳۹۵	۳۸۳	۴۷۲	۶۸	-۴۰۴
۲۷۳	۱۰۴	۴۹۹	۳۹۵	۳۲۵	۴۰۵	۴	-۴۰۱
۲۰۶	۳۵	۴۲۴	۳۸۹	۳۲۸	۴۰۶	۱۰	-۳۹۶
۱۸۱	۷	۳۷۹	۳۷۲	۳۶۰	۴۲۲	۴۳	-۳۷۹
۴۴	۹۵	۴۶۵	۳۷۰	۳۲۰	۴۵۸	۸۰	-۳۷۸
۱۲۶	۷۵	۴۴۵	۳۷۰	۳۸۶	۴۰۷	۲۹	-۳۷۸
۷۵	۲۳	۳۹۱	۳۶۸	۳۲۶	۳۸۱	۸	-۳۷۳
۱۶	۷۴	۴۴۱	۳۶۷	۶۹	۴۸۶	۱۱۴	-۳۷۲
۲۴۰	۶۵	۴۳۱	۳۶۶	۸۰	۴۸۷	۱۲۱	-۳۶۶
۱۸۰	۳۳	۳۹۶	۳۶۳	۱۱۷	۴۱۷	۵۲	-۳۶۵

کد	رتبه در IRT	رتبه در RT	تفاوت	کد	رتبه در IRT	رتبه در RT	تفاوت
۲۸۵	۱۳۵	۴۹۵	۳۶۰	۳۱۹	۴۴۲	۸۱	-۳۶۱
۱۷۵	۱۰۰	۴۵۵	۳۵۵	۳۹۵	۴۸۲	۱۲۴	-۳۵۸
۱۵۷	۳۱	۳۸۵	۳۵۴	۲۱۷	۴۰۱	۴۷	-۳۵۴
۲۳۲	۱۲۸	۴۸۰	۳۵۲	۳۵۰	۳۸۲	۳۲	-۳۵۰
۲۰۱	۱۰	۳۵۱	۳۴۱	۶۷	۴۱۵	۷۱	-۳۴۴

در جدول ۸ از بین ۵۰۲ آزمودنی، ۴۰ اختلاف که قابل توجه بودند و تفاوت بیشتری داشتند گزارش شده‌اند. همان‌گونه که در جدول مشاهده می‌شود آزمودنی با کد ۱۵۰ بر اساس محاسبات زمان-پاسخ جایگاه ۴۷۸ام را در بین ۵۰۲ نفر کسب کرده است، اما همین فرد در مدل بدون زمان جایگاه ۳۰ام 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 مدل‌بندی 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.

شکل ۴. 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 قابل مشاهده است همان گونه که در نمودار نیز مشاهده می شود در نمرات با زمان پاسخ طبقات بیشتری تولید کرده است و افراد از یکدیگر متمایزتر هستند.

بحث و نتیجه گیری

هدف از انجام این پژوهش، کاربرد مدل های سلسله مراتبی زمان-پاسخ در اندازه گیری های شناختی کودکان است. برای دستیابی به این هدف، از روش استدلال از نوع کمی به عنوان یکی از شناخته شده ترین آزمون های شناختی انتخاب و اجرا شده است. بر این اساس، پارامترهای سؤال و توانایی بر اساس مدل سلسله مراتبی زمان-پاسخ برآورد شده است. سپس به منظور بررسی کاربرد این مدل زمان-پاسخ، نتایج این مدل با نتایج مدل سؤال-پاسخ که در آن زمان مدل سازی نشده است، گزارش و مقایسه شده است.

نتایج بررسی ها نشان داد که بین پارامترهای برآورد شده از مدل زمان-پاسخ، همبستگی وجود دارد و همچنین بین پارامترهای مدل زمان-پاسخ با مدل بدون زمان پاسخ روابط معنی داری وجود دارد. در نتایج مشاهده شد برخی از سؤالاتی که در مدل زمان پاسخ، پارامتر ضریب تمیز متوسطی داشتند، در مدل بدون زمان پاسخ با ضرایب تمیز قوی تر برآورد شده اند. از طرف دیگر همبستگی بالایی بین ضرایب تمیز و میانگین زمان پاسخگویی به سؤالات وجود داشت. به گونه ای که سؤالاتی که دشوارتر بودند و نیاز به صرف زمان بیشتری داشتند، دشوارتر قلمداد شدند. اگر پاسخ های صحیح و کل زمان پاسخ ها را جمع کنیم و دو نمره کل محاسبه کنیم، بسته به علامت همبستگی بین این متغیرها، ممکن است همبستگی های مثبت و منفی کاذب بین پاسخ ها و زمان پاسخ ها مشاهده شود. همین همبستگی کاذب در مورد جمع میان آزمون دهندگانی که در صحت و سرعت متفاوت هستند صدق می کند. اگر نمره کل سؤالات و نمره کل آزمون دهندگان را محاسبه کنیم، همبستگی های بین متغیرهای آنها با هم تقابل می کند، و بسته به علائم آنها و توان نسبی، پیش بینی نتایج بدون هرگونه تحلیل رسمی دشوارتر می شوند (Van Der Linden & Ren, 2015). یک راه حل مؤثر برای حل مشکل همبستگی کاذب، مدل سازی صریح رابطه بین متغیرهای مشاهده شده و تمام متغیرهای کمکی است. به گفته ون در لیندن و گلس^۱ (۲۰۱۰)، اگر مدل با داده های تجربی به طور رضایت بخشی برازش داشته باشد، قادر خواهیم بود اثرات متغیرهای کمکی

1. Van Der Linden & Glas

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 قابل توجهی وجود دارد. از طرف دیگر نکته جالب توجه، رابطه معنی‌دار ضریب دشواری زمان پاسخ با ضریب تمیز بدون زمان پاسخ است. به عبارت دیگر در مواردی که زمان پاسخگویی نسبت به میانگین بیشتر است، احتمال پاسخگویی افراد قوی‌تر و تنای 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، به‌طور کلی هم‌تغییری وجود داشت، ولی تفاوت‌های قابل توجهی در توزیع و پراکندگی نتایج مشاهده شد. در برخی از آزمودنی‌ها (آزمودنی ۳۷۶م) رتبه در مدل زمان-پاسخ توانایی بالاتر را نشان داد و در برخی از آزمودنی‌ها (آزمودنی ۱۷۷م) رتبه در مدل زمان-پاسخ توانایی کمتر را نشان داد. این عدم انطباق دلایل متفاوتی دارد. یکی از شناخته‌شده‌ترین دلایل ارببی^۱ در برآورد توانایی در روش بیشینه درست‌نمایی^۲ که در مدل‌های سؤال-پاسخ مطرح است، به آزمودنی‌های کرانه‌ای^۳ مربوط است (Ranger & Kuhn, 2015).

در نهایت، عملکرد شناختی نمی‌تواند به‌طور مستقیم و خالص شناسایی شود. حتی برای دو استدلال جمع ساده، اگرچه طول زمانی دو مرحله استدلال کردن می‌تواند به‌وضوح عاملی باشد که بر میزان تلاشی که آنها نیاز دارند تأثیرگذار باشد، یکسان فرض کردن 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 قابل توجه 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

1. bias

2. maximum likelihood

3. extreme

متفاوت استفاده می کنند (DiTrapani et al., 2016)، امتیاز در آزمون (Molenaar et al., 2016)، تشخیص تقلب (Van Der Linden, 2008)، طراحی فرم های موازی (Van Der Linden, 2005) و انتخاب آیتم در تست تطبیقی (Van Der Linden, 2008) و در نهایت مدل سازی زمان پاسخ بر روی شخصیت و ارزیابی نگرش (Ranger, 2013؛ Ferrando & Lorenzo, 2007) دارد.

در جمع بندی تمام موارد می توان گفت اگرچه عملکرد شناختی نمی تواند به طور مستقیم اندازه گیری شود، ولی ملاحظه شد که زمان پاسخ اطلاعاتی را به عنوان یک پارامتر نهفته در هر سؤال وارد مدل می کند. 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 می دهد. بر اساس قوانین مدل سازی، هرچه اطلاعات وارد شده به مدل بیشتر و ناهمبسته باشند، می توان و آگاهی دهندگی مدل افزایش می یابد (Witelski & Bowen, 2015). بر این اساس، استفاده و کاربرد مدل های زمان-پاسخ اندازه گیری در سنجش توانایی شناختی پیشنهاد می شود.

تعارض منافع

تعارض منافع ندارم.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice Hall/Pearson Education. <https://psycnet.apa.org/record/1998-07223-000>
- Berninger, V. W., & Wolf, B. J. (2016). *Dyslexia, dysgraphia, OWL LD, and dyscalculia*. Baltimore, MD: Paul H. Brookes Publishing Co. <https://products.brookespublishing.com/Dyslexia-Dysgraphia-OWL-LD-and-Dyscalculia-P892.aspx>
- De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology*. <http://dx.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://psycnet.apa.org/doi/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://psycnet.apa.org/doi/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.
<http://dx.doi.org/10.1177/0146621606295197>
- Finn, B. (2015). Measuring motivation in low- stakes assessments. *ETS Research Report Series*, 2015(2), 1-17.
<http://dx.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. <http://dx.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.
<http://dx.doi.org/10.1080/15366367.2015.1100020>
- Goldstein, S., Princiotta, D., & Naglieri, J. A. (2015). Handbook of intelligence. *Evolutionary theory, historical perspective, and current concepts*, 10, 978-1. <http://dx.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), 14.
<https://doi.org/10.3390/jintelligence4040014>
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.
- 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://psycnet.apa.org/doi/10.1007/BF02294651>
- 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, 431.
<http://dx.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), 10.
<http://dx.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.
<http://dx.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.
<http://dx.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. <http://dx.doi.org/10.1027/1614-2241/a000152>
- 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://psycnet.apa.org/doi/10.1111/j.1745-3984.1990.tb00738.x>
- Roskam, E. E. (1997). Models for speed and time-limit tests. In *Handbook of modern item response theory* (pp. 187-208). New York, NY: Springer New York. https://link.springer.com/chapter/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 *Computer-based testing* (pp. 237-266). Routledge.
https://www.researchgate.net/publication/340487658_Exploring_Issues_of_Examinee_Behavior_Insights_Gained_from_Response-Time_Analyses
- Swanson, H. L. (2012). Working memory and reading disabilities 1: Both phonological and executive processing deficits are important. *Working memory and neurodevelopmental disorders*, 59-88.
<https://awspntest.apa.org/record/2006-20457-003>
- 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.
<http://dx.doi.org/10.1177/0022219409331958>
- 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.
<http://dx.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.
<https://psycnet.apa.org/doi/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. <http://dx.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://psycnet.apa.org/doi/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. <http://dx.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.
<http://dx.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.
<http://dx.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 *Handbook of modern item response theory* (pp. 169-185). New York, NY: Springer New York.
<https://cito.nl/media/ahsp1tfg/1992-a-logistic-model-for-time-limit-tests.pdf>

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