



Research Paper

Factorial Invariance of the PAL-2 Mathematics Assessment Across Genders in Elementary School Students

Fereshteh Ezzati¹ , Masoud Geramipour^{1*} , Hamid Khanipour²

1. Department of Assessment and Measurement, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran.

2. Department of Psychological Studies, Educational, Institute of Psychological and Social Research, Kharazmi University, Tehran, Iran.

Article info:

Received: 16.05.2026

Revised: 31.05.2026

Accepted: 19.06.2026

Keywords:

Factorial invariance, Multi group confirmatory factor analysis (MG-CFA), PAL-2 test, Psychometrics

Abstract

Background: This study examined the factorial invariance of the PAL-2 mathematics test among male and female elementary school students in Tehran. As an instrument for assessing specific mathematical skills and related cognitive processes, PAL-2 is important for identifying learning processes underlying mathematical performance. Establishing factorial invariance across gender groups is essential to ensure fairness and unbiased interpretation of the results.

Method: The research method was descriptive analytical, employing multi group confirmatory factor analysis (MG-CFA). The statistical population consisted of elementary students, and a sample of 404 participants (210 girls and 194 boys) was selected through multistage cluster sampling. Data was analyzed using AMOS version 26, and factorial invariance was tested at four levels: configural, metric, scalar, and strict.

Results: Findings indicated that the two-factor model of the test (specific mathematical skills and cognitive mathematical skills) demonstrated a good fit across both groups. The results of the multi-group analysis supported configural, metric, and scalar invariance, indicating that the same factorial structure, factor loadings, and item intercepts were equivalent across boys and girls. However, strict invariance was not confirmed, suggesting that the equality of residual variances across gender groups did not hold.

Conclusion: Based on these results, the PAL-2 mathematics test can be considered a valid tool for comparing the conceptual structure of abilities across gender groups. However, because strict invariance was not supported, direct comparisons of observed scores should be interpreted with caution. By highlighting the importance of factorial invariance testing, this study contributes to the standardization of assessment tools and the promotion of educational equity in the Iranian context.



Publisher: University of Zanjan



Citation: Ezzati, F., Geramipour, M. & Khanipour, H. (2026). Factorial Invariance of the PAL-2 Mathematics Assessment Across Genders in Elementary School Students. *Iranian Journal of Psychoeducational Assessment*, 2 (1), 153-165.

<https://doi.org/10.30470/ijpa.2026.2088813.1031>

***Corresponding Author:** Masoud Geramipour

Address: Department of Assessment and Measurement, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran

Email: mgramipour@khu.ac.ir

Extended Abstract

Introduction

The assessment of students' learning processes and cognitive abilities represents a cornerstone of contemporary educational psychometrics. Such evaluations are instrumental in identifying the nuanced cognitive profiles of learners, facilitating targeted interventions that foster academic trajectory and success. Within this framework, the Process Assessment of the Learner (PAL-2) has emerged as a sophisticated test designed to capture the complexity of cognitive and educational functioning. The PAL-2 includes separate domains related to reading, writing, and mathematics (Pals & Houghton, 2004); however, the present study focuses specifically on the mathematics assessment component of the instrument. However, the transition of such high stakes assessment tools across diverse demographic and cultural contexts necessitates rigorous empirical validation. A critical psychometric prerequisite in this regard is Factorial Invariance. This principle posits that the underlying factor structure must remain stable across different sub populations, such as those defined by gender, socio economic status, or linguistic background. Without establishing measurement invariance, any observed differences in scores might reflect systematic measurement bias rather than true variations in the latent ability, thereby compromising the fairness and interpretability of the assessment results (Vandenberg & Lance, 2000).

Despite the widespread application of the PAL-2 in various international settings, there is a continuous need to scrutinize its structural stability within specific educational ecosystems. Accordingly, the present study investigates the factorial invariance of the PAL-2 Mathematics Assessment across gender among elementary school students in Tehran. By employing Advanced Structural Equation Modeling (SEM) techniques, this research aims to ensure that the mathematics assessment, specific mathematics skills and mathematics-related cognitive skills, functions equivalently for all students, providing a robust foundation for evidence-based decision making in educational diagnostics and psychometric practice.

Methods

The present study employed a descriptive analytical research design using multi group confirmatory factor analysis. The statistical population consisted of male and female elementary school students in Tehran. A total of 404 students, including 210 girls and 194 boys, were selected through multistage cluster sampling. The study used the PAL-2 mathematics assessment, derived from the Process Assessment of the Learner, Second Edition (PAL-2). In the present study, only the mathematics-related component of the PAL-2 was examined, including indicators related to specific mathematics skills and mathematics-related cognitive abilities. Therefore, the analyses and interpretations are limited to the mathematics assessment component rather than the full PAL-2 battery. Data was analyzed using AMOS software, version 26. To examine the psychometric properties of the instrument, factorial invariance was tested across four levels: configural, metric, scalar, and strict invariance. Additionally, the reliability of PAL-2 test was calculated using Cronbach's alpha coefficient. Given that some PAL-2 subtests are performance-based or include a small number of indicators, Cronbach's alpha was interpreted cautiously.

Results

The findings of the multi group confirmatory factor analysis indicated that the PAL-2 mathematics assessment demonstrated satisfactory factorial invariance across gender groups up to the scalar level. Fit indices for successive models showed that the configural model provided a good fit to the data ($\chi^2/df = 1.70$, CFI = 0.97, RMSEA = 0.04). Constraining factor loadings to equality (metric invariance) resulted in a model with acceptable fit ($\chi^2/df = 1.65$, CFI = 0.97, RMSEA = 0.04; $\Delta CFI < 0.01$, $\Delta RMSEA < 0.01$). Similarly, adding equality constraints on indicator intercepts (scalar invariance) did not significantly degrade model fit ($\chi^2/df = 0.96$, CFI = 0.97, RMSEA = 0.04; $\Delta CFI < 0.01$, $\Delta RMSEA < 0.01$). In the next step, equality constraints were imposed on residual variances to test strict invariance. Although the strict invariance model retained an acceptable level of absolute fit ($\chi^2/df = 2.26$, CFI = 0.95, RMSEA = 0.05), the deterioration in fit relative to the scalar model indicated that strict invariance was not supported ($\Delta CFI = -0.02$, $\Delta RMSEA = 0.01$). Nested model comparisons further supported this pattern: $\Delta\chi^2 (\Delta df = 12) = 13.23$, $p = 0.35$ between the

configural and metric models; $\Delta\chi^2$ ($\Delta df = 15$) = 14.45, $p = 0.49$ between the metric and scalar models; and $\Delta\chi^2$ ($\Delta df = 40$) = 90.47, $p < 0.01$ between the scalar and strict models. Hence, configural, metric, and scalar invariance were supported, whereas strict invariance was not confirmed. These results indicate that the same two-factor structure, factor loadings, and indicator intercepts were equivalent across boys and girls. Therefore, the PAL-2 Mathematics Assessment measures the same mathematics-related latent constructs comparably across gender groups, and the establishment of scalar invariance provides a psychometric basis for comparing latent mean scores between boys and girls. However, because strict invariance was not supported, direct comparisons of observed scores should be interpreted with caution, as residual variances were not equivalent across gender groups.

In addition, the instrument demonstrated acceptable internal consistency across its multi-item subscales. Cronbach's alpha coefficients were 0.93 for Facts Retrieval and Writing (FRL), 0.70 for Verbal Computations (COVE), 0.90 for Place Value (PV), 0.64 for Spatial Working Memory (SWM), and 0.76 for Rapid Automatic Naming (RAND). These values indicate very high reliability for FRL and PV, acceptable reliability for COVE and RAND, and relatively lower reliability for SWM. However, given the small number of indicators and the performance-based nature of some PAL-2 Mathematics Assessment subtests, Cronbach's alpha coefficients should be interpreted cautiously, particularly for SWM. Single-item subtests (COSA, COSP, FB, MSPS, QWM, RASW, FW) were exempted from internal consistency evaluation.

Conclusion

The results indicated that the PAL-2 Mathematics Assessment demonstrates a stable and unbiased two-factor structure across Iranian girls and boys. Configural, metric, and scalar invariance were supported (McDonald, 1999; Floyd & Widaman, 1995), indicating equivalence in factorial structure, factor loadings, and indicator intercepts across gender groups. These findings suggest that the constructs of specific mathematics skills and mathematics-related cognitive skills are measured comparably in boys and girls. However, strict invariance was not established (Marsh et al., 2014), indicating that residual variances were not fully equivalent across groups. Therefore, although the findings support measurement equivalence of the PAL-2 Mathematics Assessment up to the scalar level, direct comparisons of observed scores should be interpreted cautiously, as they may be influenced by unequal measurement error. Overall, the results are consistent with national and international research (Berry, 2003; Khanipour et al., 2022) and support the use of the PAL-2 Mathematics Assessment for cross-gender comparisons of latent constructs, while recommending caution in interpreting direct mean-score differences.

Ethical considerations

This study adhered to ethical standards by securing informed consent and safeguarding participant confidentiality. Participation was entirely voluntary, and all individuals were fully informed about the study's aims and procedures prior to their involvement.

Funding

The authors independently funded all aspects of this research and its publication, without external financial contributions.

Author contributions

This article is based on the first author's MSc thesis. The first author performed the research, including conceptualization, data collection, formal analysis, and drafting of the original manuscript. The second author provided supervision, contributed to the study design and methodology, and critically reviewed and revised the manuscript. The third author acted as an advisor, contributing to the conceptual framework and providing valuable feedback on the manuscript. All authors approved the final version of the manuscript.

Conflict of interest

The authors state that they have no conflicts of interest.



مقاله پژوهشی

تغییرناپذیری عاملی سنجش ریاضی آزمون سنجش فرایندی تشخیصی یادگیرنده (PAL-2) برحسب جنسیت در دانش‌آموزان ابتدایی

فرشته عزتی^۱، مسعود گرامی‌پور^{۱*}، حمید خانی‌پور^۲

۱. گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران.

۲. گروه مطالعات روان‌شناختی، مؤسسه تحقیقات تربیتی، روان‌شناختی و اجتماعی، دانشگاه خوارزمی، تهران، ایران.

چکیده

زمینه: پژوهش حاضر با هدف بررسی تغییرناپذیری عاملی سنجش ریاضی آزمون (PAL-2) سنجش فرایندی تشخیصی یادگیرنده) در دانش‌آموزان دختر و پسر پایه‌های ابتدایی شهر تهران انجام شد. آزمون ریاضی PAL-2 به‌عنوان ابزاری برای ارزیابی مهارت‌های خاص ریاضی و مهارت‌های شناختی مرتبط با ریاضی، نقش مهمی در شناسایی فرایندهای زیربنایی یادگیری مرتبط با عملکرد ریاضی ایفا می‌کند. یکی از الزامات اساسی استفاده از چنین ابزارهایی، اطمینان از تغییرناپذیری عاملی آن‌ها در گروه‌های جمعیت‌شناختی مختلف است تا نتایج به‌صورت عادلانه و بدون سوگیری تفسیر شوند.

روش: روش پژوهش، توصیفی - تحلیلی از نوع تحلیل عاملی تأییدی چندگروهی بود. جامعه آماری شامل دانش‌آموزان دختر و پسر مقطع ابتدایی در شهر تهران بود که ۴۰۴ نفر (۲۱۰ دختر و ۱۹۴ پسر) به‌صورت نمونه‌گیری خوشه‌ای چندمرحله‌ای انتخاب شدند. داده‌ها با استفاده از نرم‌افزار AMOS-26 تحلیل شدند و تغییرناپذیری عاملی آزمون در چهار سطح پیکربندی، متریک، اسکالر و سختگیرانه مورد بررسی قرار گرفت.

یافته‌ها: یافته‌ها نشان داد که مدل دو عاملی آزمون (شامل مهارت‌های خاص ریاضی و مهارت‌های شناختی ریاضی) در هر دو گروه برازش مناسبی دارد. نتایج تحلیل چندگروهی نشان‌دهنده آن بود که تغییرناپذیری پیکربندی، متریک و اسکالر برقرار است؛ اما تغییرناپذیری سختگیرانه مورد تأیید قرار نگرفت؛ به بیان دیگر، ساختار عاملی، بارهای عاملی و عرض از مبداهای سوالات در دختران و پسران معادل بود؛ اما برابری خطاهای اندازه‌گیری بین دختران و پسران برقرار نبود.

نتیجه‌گیری: بر اساس این نتایج، آزمون ریاضی PAL-2 را می‌توان ابزاری معتبر برای مقایسه ساختار مفهومی توانایی‌ها در گروه‌های جنسیتی دانست. با این حال، از آنجایی که تغییرناپذیری سختگیرانه تأیید نشد، مقایسه مستقیم نمرات مشاهده‌شده باید با احتیاط تفسیر شود؛ به‌علاوه، این پژوهش گامی در جهت استانداردسازی و ارتقای عدالت آموزشی در نظام سنجش ایران به‌شمار می‌رود.

اطلاعات مقاله:

تاریخ دریافت: ۱۴۰۵/۰۲/۲۶

تاریخ داوری: ۱۴۰۵/۰۳/۱۰

تاریخ پذیرش: ۱۴۰۵/۰۳/۲۹

واژه‌های کلیدی:

تغییرناپذیری عاملی، تحلیل عاملی
تأییدی چندگروهی، آزمون
PAL-2، روان‌سنجی.



ناشر: دانشگاه زنجان

استناد: عزتی، ف.، گرامی‌پور، م. و خانی‌پور، ح. (۱۴۰۵). تغییرناپذیری عاملی سنجش ریاضی آزمون سنجش فرایندی

تشخیصی یادگیرنده (PAL-2) برحسب جنسیت در دانش‌آموزان ابتدایی. *سنجش روانی تربیتی*، ۲(۱)، ۱۵۳-۱۶۵.

<https://doi.org/10.30470/ijpa.2026.2088813.1031>



از دستگاه خود برای اسکن و خواندن
مقاله به صورت آنلاین استفاده کنید

* نویسنده مسئول: مسعود گرامی‌پور

نشانی: گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران.

پست الکترونیکی: mgramipour@khu.ac.ir

مقدمه

ارزیابی فرایندهای یادگیری و توانایی‌های شناختی دانش‌آموزان از موضوعات اساسی در روان‌سنجی آموزشی است که نقش مهمی در شناسایی نقاط قوت و ضعف یادگیرندگان و هدایت آن‌ها به سوی پیشرفت تحصیلی ایفا می‌کند. استفاده از ابزارهای سنجش معتبر می‌تواند به درک دقیق‌تر از نحوه یادگیری دانش‌آموزان و عوامل مؤثر بر عملکرد تحصیلی آنان کمک کند. در این میان، آزمون PAL-2 به عنوان یکی از ابزارهای جامع برای سنجش فرایندهای شناختی و عملکرد آموزشی طراحی شده است. PAL-2 شامل حوزه‌های مرتبط با خواندن، نوشتن و ریاضی است. با این حال، پژوهش حاضر صرفاً بر بخش سنجش ریاضی این آزمون تمرکز دارد، نه بر کل مجموعه آزمون. بخش ریاضی PAL-2 با هدف ارزیابی مهارت‌های اختصاصی ریاضی و فرایندهای شناختی مرتبط با عملکرد ریاضی به کار می‌رود و می‌تواند در شناسایی برخی فرایندهای زیربنایی مؤثر بر یادگیری ریاضی، مانند پردازش عددی، حافظه کاری^۱ و حل مسئله، نقش داشته باشد (زیبالسکی و ویزل^۲، ۲۰۱۳؛ پترسون^۳ و همکاران، ۲۰۱۰).

سنجش ریاضی PAL-2 برای دانش‌آموزان پایه کودکستان تا ششم ابتدایی است و به جای بسندگی به نمره کل، به شناسایی فرایندهای شناختی و تحصیلی مرتبط با مشکلات یادگیری ریاضی تأکید دارد و در سطوح مختلف غربالگری، مداخله و تشخیص اختلالات یادگیری کاربرد دارد. بخش ریاضی آن شامل ۱۴ خرده‌آزمون برای سنجش محاسبه، درک عددی و حل مسئله است که در فرایند انطباق‌یابی در ایران، با برنامه درسی و بافت فرهنگی انطباق پیدا کرده است. پژوهش خانی‌پور و همکاران (۱۴۰۱) درباره ۴۰۴ دانش‌آموز تهرانی، ویژگی‌های روان‌سنجی مطلوب نسخه ایرانی را تأیید کرد. تغییرناپذیری اندازه‌گیری در این مطالعه بررسی نشد؛ اما تحلیل عاملی نشان داد که برخلاف ساختار دو عاملی نسخه اصلی، مدل سه عاملی برازش بهتری با جمعیت ایران دارد که نشان‌دهنده تناسب بیشتر این ابزار با کارکردهای عصبی - روان‌شناختی دانش‌آموزان ایرانی است.

با وجود اهمیت ابزارهای استاندارد در سنجش توانایی‌های شناختی، یکی از چالش‌های مهم در کاربرد آن‌ها، مسئله تغییرناپذیری عاملی^۴ است. تغییرناپذیری عاملی به این معنا است که ساختار عاملی آزمون باید در گروه‌های مختلف جمعیتی، از جمله گروه‌های مبتنی بر جنسیت، زبان یا قومیت، ثابت باقی بماند. در صورتی که این ویژگی در ابزار سنجشی برقرار نباشد، مقایسه نتایج بین گروه‌های مختلف ممکن است با سوگیری همراه شود و اعتبار و کارآمدی آزمون زیر سؤال برود (وندنبرگ و لنس^۵، ۲۰۰۰). از این رو، بررسی تغییرناپذیری عاملی از الزامات اساسی در اعتبارسنجی ابزارهای روان‌سنجی به‌شمار می‌رود. در کشوری مانند ایران که از تنوع فرهنگی، زبانی و اجتماعی قابل توجهی برخوردار است، بررسی تغییرناپذیری عاملی ابزارهای سنجشی اهمیت بیشتری پیدا می‌کند. تحلیل ساختار عاملی آزمون‌ها در چنین بافتی می‌تواند به بهبود عدالت سنجشی و افزایش اعتبار بین فرهنگی آن‌ها کمک کند. پژوهش‌های پیشین نشان داده‌اند که تحلیل عاملی تأییدی یکی از روش‌های مؤثر برای بررسی پایداری ساختار عاملی آزمون‌ها در گروه‌های مختلف است. در این رویکرد، سطوح مختلف تغییرناپذیری شامل تغییرناپذیری پیکربندی^۶، متریک^۷، اسکالر^۸ و سختگیرانه^۹ مورد بررسی قرار می‌گیرد تا میزان ثبات ساختار عاملی ابزار مشخص شود (چئونگ و رزنولد^{۱۰}، ۲۰۰۲).

بررسی تغییرناپذیری عاملی آزمون‌ها علاوه بر اهمیت نظری، پیامدهای عملی مهمی نیز برای نظام آموزشی دارد. نتایج چنین مطالعاتی می‌تواند به بهبود طراحی آزمون‌ها، کاهش سوگیری‌های فرهنگی^{۱۱} و افزایش دقت ابزارهای سنجش آموزشی کمک کند. همچنین شناخت دقیق ساختار عاملی آزمون‌ها می‌تواند سیاست‌گذاران و برنامه‌ریزان آموزشی را در درک بهتر فرایندهای یادگیری یاری دهد و زمینه تدوین برنامه‌های آموزشی و مداخلات هدفمندتر را فراهم سازد؛ امری که در نهایت به ارتقای کیفیت آموزش در نظام آموزشی کشور منجر خواهد شد.

از منظر روان‌سنجی، بررسی تغییرناپذیری عاملی به توسعه دانش مربوط به اعتبار سازه‌ای^{۱۲} ابزارهای اندازه‌گیری کمک می‌کند. مفهوم تغییرناپذیری عاملی در ادبیات بین‌المللی به خوبی شناخته شده است؛ اما بررسی آن در بسترهای فرهنگی خاص اهمیت ویژه‌ای دارد. تحلیل تغییرناپذیری عاملی آزمون‌هایی مانند سنجش ریاضی PAL-2 می‌تواند نشان دهد که آیا سازه‌های مربوط به مهارت‌های شناختی و ریاضی در گروه‌های مختلف به شکل مشابهی عمل می‌کنند یا خیر؟ همچنین به غنای مدل‌های نظری اندازه‌گیری در حوزه روان‌سنجی کمک می‌کند (مردیت^{۱۳}، ۱۹۹۳).

بر این اساس، پژوهش حاضر با هدف بررسی تغییرناپذیری عاملی سنجش ریاضی PAL-2 برحسب جنسیت در دانش‌آموزان پایه‌های اول تا ششم ابتدایی شهر تهران انجام شد. در این مطالعه با استفاده از تحلیل عاملی تأییدی چندگروهی^{۱۴}، پایداری ساختار عاملی آزمون در گروه‌های مختلف بررسی شد تا مشخص شود آیا این ابزار

1 Working Memory

2 Zibulsky & Veizel

3 Peterson

4 Factorial Invariance

5 Vandenberg & Lance

6 Configural

7 Metric

8 Scalar

9 Strict

10 Cheung & Rensvold

11 Cultural Biases

12 Construct Validity

13 Meredith

14 Multi-group Factor Analysis

از ساختار عاملی پایدار برخوردار است و می‌توان از آن برای مقایسه معتبر عملکرد دانش‌آموزان استفاده کرد یا خیر؟ نتایج این پژوهش می‌تواند به استانداردسازی بهتر آزمون‌های تشخیصی، بهبود عدالت سنجشی و ارتقای کیفیت ارزیابی آموزشی در کشور کمک کند (براون، ۲۰۱۵؛ میلز، ۲۰۱۱).

روش

پژوهش حاضر از نظر هدف، کاربردی و از نظر ماهیت و روش، توصیفی از نوع همبستگی - روان‌سنجی است. به‌طور خاص، این مطالعه از نوع پیمایشی - تحلیلی محسوب می‌شود؛ زیرا به بررسی ساختار عاملی و تغییرناپذیری اندازه‌گیری سنجش ریاضی PAL-2 در گروه‌های جنسیتی می‌پردازد.

برای جمع‌آوری داده‌ها، پس از اخذ مجوزهای لازم از مراجع ذی‌صلاح، شامل سازمان آموزش و پرورش شهر تهران و دانشگاه مربوط، هماهنگی‌های لازم با مدیران و معلمان مدارس منتخب انجام شد. در مرحله بعد، اهداف پژوهش و نحوه اجرای آن به‌طور کامل برای دانش‌آموزان والدین آنان توضیح داده شد و رضایت آگاهانه کتبی از والدین برای مشارکت فرزندانشان در پژوهش اخذ گردید. همچنین رضایت شفاهی دانش‌آموزان نیز دریافت شد. اجرای سنجش ریاضی آزمون PAL-2 توسط مجریان آموزش‌دیده و با رعایت کامل دستورالعمل‌ها و پروتکل‌های استاندارد آزمون انجام شد. بسته به نوع خرده‌آزمون‌ها، اجرا به‌صورت فردی یا گروهی و در محیطی آرام و استاندارد صورت گرفت. در تمام مراحل جمع‌آوری داده‌ها و تحلیل نتایج، ملاحظات اخلاقی شامل حفظ حریم خصوصی، محرمانه بودن اطلاعات و ناشناس ماندن شرکت‌کنندگان، رعایت شد. داده‌ها نیز در بازه زمانی تعیین‌شده در قلمروی زمانی پژوهش، یعنی سال ۱۴۰۳، گردآوری شدند.

داده‌های گردآوری‌شده با استفاده از نرم‌افزار SPSS (نسخه ۲۶) و AMOS (نسخه ۲۶) مورد تحلیل قرار گرفتند. از SPSS برای سازماندهی داده‌ها و انجام تحلیل‌های توصیفی اولیه و از AMOS برای انجام تحلیل‌های پیشرفته‌تر و آزمون مدل‌های اندازه‌گیری استفاده شد. در بخش آمار توصیفی از شاخص‌هایی مانند میانگین، انحراف معیار، فراوانی و درصد برای توصیف ویژگی‌های جمعیت‌شناختی نمونه و نیز نمرات خام خرده‌آزمون‌ها استفاده شد.

برای بررسی فرضیه‌های پژوهش درباره تغییرناپذیری عاملی سنجش ریاضی PAL-2 بین گروه‌های جنسیتی، از تحلیل عاملی تأییدی چندگروهی استفاده شد. این تحلیل به‌صورت سلسله‌مراتبی و در چهار مرحله انجام گرفت: اول، مدل تغییرناپذیری پیکربندی برای بررسی یکسانی الگوی بارهای عاملی در دو گروه؛ دوم، مدل تغییرناپذیری متریک برای آزمون برابری بارهای عاملی؛ سوم، مدل تغییرناپذیری اسکالر برای آزمون برابری عرض از مبدأهای نشانگرها در دو گروه و چهارم، مدل تغییرناپذیری سختگیرانه برای آزمون برابری واریانس خطاهای اندازه‌گیری در دو گروه دختران و پسران.

برآزش هر یک از مدل‌ها با استفاده از شاخص‌های استاندارد برازش شامل نسبت مجذور خی به درجه آزادی (χ^2/df)، ریشه دوم میانگین مجذور خطای برآورد^۱، شاخص برازش نسبی^۲، شاخص تاکر-لوئیس^۳ و ریشه میانگین مجذور باقی‌مانده^۴ ارزیابی شد. ملاک‌های پذیرش برازش مدل‌ها شامل $RMSEA < 0.08$ ، $CFI > 0.90$ و $RMR/ SRMR < 0.05$ بود؛ افزون بر این، تغییرات شاخص‌های CFI و RMSEA بین مدل‌های متوالی نیز برای ارزیابی افت معنی‌دار در برازش مدل مورد بررسی قرار گرفت.

مشارکت‌کنندگان

مشارکت‌کنندگان این پژوهش را دانش‌آموزان پسر و دختر پایه‌های اول تا ششم ابتدایی شهر تهران در سال تحصیلی ۱۴۰۲-۱۴۰۳ تشکیل دادند. این جامعه آماری به دلیل برخورداری از تنوع فرهنگی، اجتماعی و اقتصادی، بستر مناسبی برای بررسی ویژگی‌های روان‌سنجی و ارزیابی ساختار عاملی ابزارهای اندازه‌گیری در گروه‌های مختلف جمعیتی فراهم می‌کند. نمونه پژوهش با استفاده از روش نمونه‌گیری خوشه‌ای چندمرحله‌ای^۵ انتخاب شد؛ به‌گونه‌ای که در مرحله نخست، مناطق مختلف آموزش و پرورش شهر تهران شامل مناطق شمالی، جنوبی، شرقی و غربی به‌عنوان خوشه‌های اولیه در نظر گرفته شدند. سپس از هر منطقه، چند مدرسه به‌صورت تصادفی انتخاب و در مرحله نهایی، کلاس‌ها و دانش‌آموزان هر مدرسه به‌طور تصادفی برای مشارکت در مطالعه گزینش شدند. حجم نمونه (بین ۲۰۰ تا ۴۰۰ نفر) با هدف تأمین کفایت آماری لازم برای انجام تحلیل عاملی تأییدی چندگروهی و دستیابی به قدرت آزمون مناسب، بر اساس توصیه جکسون^۶ (۲۰۰۱) مد نظر قرار گرفته شد. بر این اساس، ۴۰۴ نفر در این مطالعه مشارکت داشتند که ۲۱۰ نفر از آن‌ها را دختران (۵۲٪) و ۱۹۴ نفر از آن‌ها را پسران (۴۸٪) تشکیل دادند. ۶۰ نفر (۱۴/۹٪) در مقطع پیش‌دبستان، ۵۳ نفر (۱۳/۱٪) در پایه اول، ۵۲ نفر (۱۲/۹٪) در پایه دوم، ۵۵ نفر (۱۳/۶٪) در پایه سوم، ۵۸ نفر (۱۴/۴٪) در پایه چهارم، ۶۱ نفر (۱۵/۱٪) در پایه پنجم و ۶۵ نفر (۱۶/۱٪) در پایه ششم ابتدایی مشغول به تحصیل بودند. ۲۹۵ نفر از مشارکت‌کنندگان پژوهش (۷۳٪) در مدارس دولتی، ۹۱ نفر (۲۲/۵٪) در مدارس نیمه‌دولتی و ۱۸ نفر (۴/۵٪) در مدارس غیرانتفاعی حضور داشتند. به‌علاوه، ۱۸ نفر (۴/۵٪) در منطقه ۱، ۷۸ نفر (۱۹/۳٪) در منطقه ۲، ۱۱۴ نفر (۲۸/۲٪) در منطقه ۳، ۴۳ نفر (۱۰/۶٪) در منطقه ۵، ۲۰ نفر (۵٪) در منطقه ۹، ۷۵ نفر (۱۸/۶٪) در منطقه ۱۰، ۱۲ نفر (۳٪) در منطقه ۱۵ و نهایتاً ۴۴ نفر (۱۰/۹٪) در منطقه ۲۱ بودند. همچنین پدر ۳۵ تن از مشارکت‌کنندگان (۸/۷٪) به کارگری، ۱۳۳ نفر (۳۲/۹٪) به کارمندی، ۱۲۸ نفر (۳۱/۲٪) به شغل آزاد، ۲۹ نفر (۷/۲٪) در مشاغل مربوط به دانشگاه، پزشکی و مهندسی، ۲۱

¹ Brown

² Millsap

³ Root Mean Squared Error of Approximation (RMSEA)

⁴ Comparative Fit Index (CFI)

⁵ Tucker-Lewis Index (TLI)

⁶ Root Mean Square Residual (RMR/SRMR)

⁷ Multistage Cluster Sampling

⁸ Jackson

نفر (۵/۲٪) در بازار، ۳۰ نفر (۷/۴٪) در خدمات عمومی و ۱۸ نفر (۴/۵٪) در ارتش مشغول به کار بودند. ۲۹۷ نفر از مشارکت کنندگان (۷۳/۵٪) تک‌زبانه و ۱۰۷ نفر (۲۶/۵٪) دوزبانه بودند. همچنین ۱۴۲ نفر از مشارکت کنندگان (۳۵/۱٪) خواهر/ برادر نداشتند. ۲۱۰ نفر (۵۲٪) ۱ خواهر/ برادر، ۴۵ نفر (۱۱/۱٪) ۲ خواهر/ برادر، ۵ نفر (۱/۲٪) ۳ خواهر/ برادر، ۱ نفر (۰/۲٪) ۴ خواهر/ برادر و ۱ نفر (۰/۲٪) نیز ۵ خواهر/ برادر داشتند.

ابزار

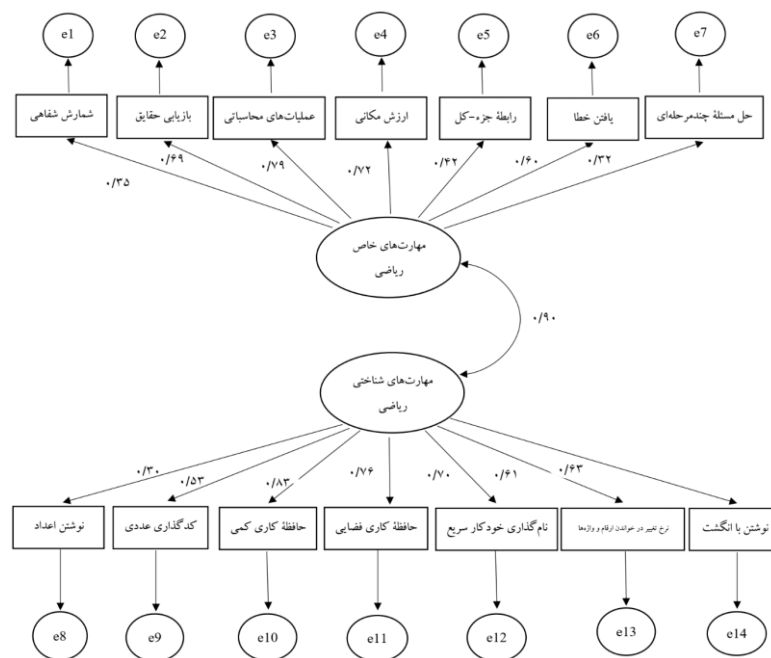
ابزار اندازه‌گیری این پژوهش، بخش سنجش ریاضی آزمون ارزیابی فرایندی تشخیصی یادگیرنده^۱ (PAL-2) است. این آزمون ابزاری استاندارد و جامع روان‌سنجی است که توسط پالس و هوتن^۲ (۲۰۰۴) برای ارزیابی مهارت‌های شناختی و فرایندهای یادگیری دانش‌آموزان طراحی شده است. در پژوهش حاضر، کل مجموعه آزمون PAL-2 بررسی نشده، بلکه صرفاً خرده‌آزمون‌های مرتبط با سنجش ریاضی آن مورد استفاده قرار گرفته است. این آزمون توسط زیبالسکی و ویزل (۲۰۱۴) برای ارزیابی فرایندهای شناختی و تحصیلی مرتبط با یادگیری طراحی شده است. PAL-2 در نسخه اصلی شامل دو بخش اصلی خواندن/ نوشتن و بخش ریاضی آن با هدف سنجش مهارت‌های اختصاصی ریاضی و فرایندهای شناختی مرتبط با عملکرد ریاضی طراحی شده است.

بخش سنجش ریاضی PAL-2 مجموعه‌ای از خرده‌آزمون‌ها را شامل می‌شود که به ارزیابی مهارت‌هایی نظیر نوشتن اعداد، شمارش شفاهی، کدگذاری عددی، بازیابی حقایق، عملیات محاسباتی، ارزش مکانی، رابطه جزء-کل، یافتن خطا، حل مسئله چندمرحله‌ای، حافظه کاری کمی، حافظه کاری فضایی، نام‌گذاری خودکار سریع، نرخ تغییر در خواندن ریاضی و واژه‌ها و نوشتن یا انگشت می‌پردازد. این آزمون اطلاعات ارزشمندی درباره نقاط ضعف و قوت یادگیرندگان در حوزه ریاضی ارائه می‌دهد و می‌تواند برای شناسایی چالش‌های یادگیری ریاضی و طراحی مداخلات آموزشی متناسب مورد استفاده قرار گیرد. از آنجایی که برخی خرده‌آزمون‌های PAL-2 ماهیت عملکردی دارند و برخی نیز از تعداد سؤال محدودی تشکیل شده‌اند، تفسیر شاخص پایایی^۳ آلفای کرونباخ باید با احتیاط انجام شود. این آزمون در نسخه اصلی خود از پایایی و روایی^۴ مطلوبی برخوردار است که در مطالعات پیشین به اثبات رسیده است.

یافته‌ها

نتایج تحقیق حاضر در سه بخش اصلی شامل ساختار عاملی، تغییرناپذیری اندازه‌گیری^۵ و پایایی، سازمان‌دهی و در ادامه ارائه شده است.

ساختار عاملی: به منظور بررسی ساختار عاملی آزمون PAL-2 در دو گروه، از مدل دو عاملی استفاده شد که دیاگرام آن در شکل ۱ قابل مشاهده است.



شکل ۱

نمودار مدل تحلیل عاملی تأییدی

1 Process Assessment of the Learner, Second Edition

2 Pals & Houghton

3 Reliability

4 Validity

5 Measurement Invariance

همچنین ضرایب برآورد شده در مدل دو عاملی پژوهش حاضر در جدول ۱ فراهم گشته است.

جدول ۱

پارامترهای برآورد شده در مدل دو عاملی

سازه	شاخص	بار عاملی	بار عاملی استاندارد شده	خطای استاندارد	مقدار بحرانی	سطح معنی داری
مهارت‌های خاص ریاضی	یافتن خطا	۱	۰/۶۰	-	-	-
مهارت‌های خاص ریاضی	رابطه جزء - کل	۲/۷۲	۰/۴۲	۰/۴۷	۵/۷۲	<۰/۰۰۱
مهارت‌های خاص ریاضی	ارزش مکانی	۴/۴۳	۰/۷۲	۰/۵۵	۷/۹۶	<۰/۰۰۱
مهارت‌های خاص ریاضی	عملیات محاسباتی	۵/۴۶	۰/۷۶	۰/۶۷	۸/۳۲	<۰/۰۰۱
مهارت‌های خاص ریاضی	بازیابی حقایق	۶/۶۷	۰/۶۹	۰/۸۷	۷/۶۲	<۰/۰۰۱
مهارت‌های خاص ریاضی	شمارش شفاهی	۳/۰۷	۰/۳۵	۰/۶۸	۴/۴۹	<۰/۰۰۱
مهارت‌های خاص ریاضی	حل مسئله چندمرحله‌ای	۰/۷۹	۰/۳۲	۰/۱۶	۴/۹۲	<۰/۰۰۱
مهارت‌های شناختی ریاضی	نوشتن اعداد	۱	۰/۳۰	-	-	-
مهارت‌های شناختی ریاضی	کدگذاری عددی	۰/۳۰	۰/۵۳	۰/۰۶	۴/۷۳	<۰/۰۰۱
مهارت‌های شناختی ریاضی	حافظه کاری کمی	۱/۳۵	۰/۸۳	۰/۳۲	۴/۱۶	<۰/۰۰۱
مهارت‌های شناختی ریاضی	حافظه کاری فضایی	۰/۹۴	۰/۷۶	۰/۲۳	۴/۱۱	<۰/۰۰۱
مهارت‌های شناختی ریاضی	نام‌گذاری خودکار سریع	۲/۰۴	۰/۷۰	۰/۵۰	۴/۰۶	<۰/۰۰۱
مهارت‌های شناختی ریاضی	نرخ تغییر در خواندن ارقام و واژه‌ها	۵/۹۵	۰/۶۱	۱/۵۱	۳/۹۴	<۰/۰۰۱
مهارت‌های شناختی ریاضی	نوشتن با انگشت	۰/۴۵	۰/۶۳	۰/۱۱	۳/۹۷	<۰/۰۰۱

همان‌طور که در جدول ۱ نشان داده شده است، تمامی بارهای عاملی سؤالات با عوامل مربوط در سطح ۰/۰۱ معنی‌دار بودند.

تغییرناپذیری اندازه‌گیری: در این بخش، چهار مرحله تغییرناپذیری پیکربندی، متریک، اسکالر و سختگیرانه، برای بررسی تغییرناپذیری اندازه‌گیری ابزار مورد نظر بررسی شد. جدول ۲ نشان‌دهنده شاخص‌های برازش هر یک از چهار مدل مذکور است.

جدول ۲

مقایسه شاخص‌های آماری چهار مدل

مدل	χ^2/df (مقدار مطلوب: کمتر از ۲)	RMSEA (مقدار مطلوب: کمتر از ۰/۱)	CFI (مقدار مطلوب: بیشتر از ۰/۹۰)
پیکربندی	۱/۷۰	۰/۰۴	۰/۹۷
متریک	۱/۶۵	۰/۰۴	۰/۹۷
اسکالر	۰/۹۶	۰/۰۴	۰/۹۷
سختگیرانه	۲/۲۶	۰/۰۵	۰/۹۵

همان‌طور که در جدول ۲ قابل مشاهده است، تمام شاخص‌های برازش مدل در سطح تغییرناپذیری پیکربندی، متریک، اسکالر و سختگیرانه، نشان‌دهنده برازش مناسب مدل هستند؛ افزون بر این، نتایج مقایسه مدل‌ها بر اساس آزمون تفاوت‌خی‌دو، تفاوت CFA و تفاوت RMSEA در جدول ۳ نشان داده شده است.

جدول ۳

نتایج مقایسه مدل‌ها

مدل‌های مقایسه شده	χ^2	درجه آزادی	مقدار احتمال	ΔCFI	$\Delta RMSEA$	نتیجه
متریک در مقابل پیکربندی	۱۳/۲۳	۱۲	۰/۳۵	<۰/۰۱	<۰/۰۱	تأیید
اسکالر در مقابل متریک	۱۴/۴۵	۱۵	۰/۴۹	<۰/۰۱	<۰/۰۱	تأیید
سختگیرانه در مقابل اسکالر	۹۰/۴۷	۴۰	۰/۰۰	-۰/۰۲	۰/۰۱	رد

اطلاعات جدول ۳ نشان می‌دهد که تفاوت معنی‌داری بین مدل متریک و مدل پیکربندی وجود ندارد؛ به عبارت دیگر، بارهای عاملی بین گروه‌ها یکسان هستند و افراد در گروه‌ها، سازه مورد نظر را به شکل مشابه درک می‌کنند؛ علاوه بر این، تفاوت معنی‌داری بین مدل اسکالر و مدل متریک وجود ندارد؛ به بیان دیگر، عرض از مبدأ نشانگرها در گروه‌ها یکسان است. این یافته نشان می‌دهد که افزون بر برابری بارهای عاملی، سطح پایه یا نقطه شروع

نشانه‌ها نیز در دو گروه دختران و پسران معادل است. بر این اساس، تغییرناپذیری اسکالر تأیید می‌شود و مقایسه میانگین‌های نهفته بین گروه‌های جنسیتی از نظر روان‌سنجی قابل دفاع است.

در نهایت، برخلاف دو مورد پیشین، تفاوت معنی‌داری بین مدل سختگیرانه و مدل اسکالر وجود دارد؛ به عبارت دیگر، برابری خطاهای اندازه‌گیری بین گروه‌ها برقرار نیست و اعمال این محدودیت باعث افت برازش مدل شده است؛ بنابراین، می‌توان مدل را تا سطح اسکالر به صورت تغییرناپذیر در نظر گرفت؛ اما مقایسه مستقیم نمرات مشاهده‌شده بین گروه‌ها ممکن است با محدودیت همراه باشد؛ چون مدل سختگیرانه رد شده است.

در مجموع، نتایج تحلیل عاملی تأییدی چندگروهی نشان داد که سنجش ریاضی آزمون PAL-2 از نظر ساختار عاملی (تغییرناپذیری پیکربندی)، وزن‌های عاملی (تغییرناپذیری متریک) و عرض از مبدأهای نشانگرها (تغییرناپذیری اسکالر) در دو گروه مختلف از دانش‌آموزان دارای تغییرناپذیری است؛ اما به دلیل تفاوت معنی‌دار در واریانس خطاهای اندازه‌گیری، تغییرناپذیری سختگیرانه در این آزمون برقرار نیست؛ بنابراین، می‌توان از این آزمون برای مقایسه ساختار مفهومی توانایی‌ها بین گروه‌ها استفاده کرد؛ اما مقایسه نمرات مشاهده‌شده باید با احتیاط تفسیر شود.

پایایی: به منظور اطمینان از دقت و ثبات اندازه‌گیری ابزار مورد استفاده در پژوهش حاضر، پایایی خرده‌آزمون‌های چندسؤالی سنجش ریاضی PAL-2 از طریق محاسبه ضریب آلفای کرونباخ^۱ بررسی شد. آلفای کرونباخ یکی از شاخص‌های رایج برای برآورد همسانی درونی^۲ مجموعه‌ای از سؤالات است. با این حال، تفسیر آن در خرده‌آزمون‌های عملکردی، کوتاه یا دارای تعداد سؤالات محدود باید با احتیاط انجام شود. بر اساس دیدگاه گرامی‌پور (۱۳۹۳) و معیارهای متداول در روان‌سنجی، مقدار آلفای بالاتر از ۰/۷۰ نشان‌دهنده پایایی قابل قبول، بالاتر از ۰/۸۰ نشان‌دهنده پایایی مطلوب و بالاتر از ۰/۹۰ نشان‌دهنده پایایی بسیار عالی است. پایایی هریک از زیرمقیاس‌های آزمون و همچنین پایایی کلی ابزار تحقیق حاضر در جدول ۴ آورده شده است.

جدول ۴			
مقادیر ضریب آلفای کرونباخ			
زیرمقیاس	تعداد سؤالات	آلفای کرونباخ	سطح پایایی
(بازیابی حقایق و نوشتن) FRL	۱۲	۰/۹۳	بسیار مطلوب
(محاسبات کلامی) COVE	۴	۰/۷	قابل قبول
(سازمان‌دهی فضایی) COSA	۱	-	تک‌سؤالی
(حل مسئله محاسباتی) COSP	۱	-	تک‌سؤالی
(ارزش مکانی) PV	۳	۰/۹	بسیار مطلوب
(یافتن خطا) FB	۱	-	تک‌سؤالی
(حل مسئله چندمرحله‌ای) MSPS	۱	-	تک‌سؤالی
(حافظه کاری کمی) QWM	۱	-	تک‌سؤالی
(حافظه کاری فضایی) SWM	۲	۰/۶۴	نسبتاً پایین
(نام‌گذاری خودکار سریع) RAND	۴	۰/۷۶	قابل قبول
(نرخ تغییر در خواندن و نوشتن) RASW	۱	-	تک‌سؤالی
(نوشتن با انگشت) FW	۱	-	تک‌سؤالی

با توجه به اطلاعات مندرج در جدول ۴، نتایج به دست آمده از تحلیل پایایی مؤید آن است که ساختار خرده‌آزمون‌های چندسؤالی سنجش ریاضی PAL-2 از لحاظ همسانی درونی سؤالات، کیفیت مطلوبی دارد و در مجموع، ابزار انتخاب‌شده برای هدف این پژوهش (بررسی تغییرناپذیری عاملی) از اعتبار روان‌سنجی کافی برخوردار است. ضرایب به دست آمده برای FRL و PV نشان‌دهنده همسانی درونی بسیار مطلوب و ضرایب COVE و RAND نشان‌دهنده همسانی درونی قابل قبول بودند. ضریب آلفای SWM نسبتاً پایین‌تر از سایر خرده‌آزمون‌ها بود. با این حال، از آنجایی که این خرده‌آزمون تنها از دو سؤال تشکیل شده است، مقدار آلفای آن باید با احتیاط تفسیر شود. همچنین برای خرده‌آزمون‌های تک‌سؤالی، آلفای کرونباخ گزارش نشد که این امر به معنای ضعف پایایی آن‌ها نیست، بلکه ناشی از نامناسب بودن ضریب آلفا برای خرده‌آزمون‌های تک‌سؤالی است.

بحث و نتیجه‌گیری

هدف اصلی پژوهش حاضر، بررسی تغییرناپذیری عاملی آزمون ریاضی PAL-2 در دانش‌آموزان دختر و پسر بود تا مشخص شود که آیا این ابزار، سازه‌های مورد نظر را در دو گروه جنسیتی به صورت یکسان اندازه‌گیری می‌کند یا خیر؟ یافته‌های حاصل از تحلیل عاملی تأییدی چندگروهی نشان

¹ Chronbach's Alpha

² Internal Consistency

داد که مدل دو عاملی شامل «مهارت‌های شناختی ریاضی» و «مهارت‌های خاص ریاضی» از برازش مطلوبی برخوردار است و در سطوح پیکربندی، متریک و اسکالر، تغییرناپذیری لازم را حفظ می‌کند. این نتایج نشان دهنده آن است که ساختار مفهومی آزمون در هر دو گروه مشابه است و دختران و پسران مؤلفه‌های آزمون را به شیوه‌ای نسبتاً یکسان درک می‌کنند. تأیید تغییرناپذیری پیکربندی نشان می‌دهد که الگوی کلی روابط میان سؤالات و عوامل نهفته در دو گروه یکسان است؛ موضوعی که با دیدگاه‌های مک‌دونالد^۱ (۱۹۹۹) و همچنین رویکردهای بین فرهنگی ون دو ویور و لیونگ^۲ (۱۹۹۷) و بری^۳ (۲۰۰۳) همسو است. این پژوهشگران ثبات ساختار عاملی را شرط اساسی برای انجام مقایسه‌های معتبر بین گروهی می‌دانند. همچنین همخوانی یافته‌های حاضر با نتایج خانی‌پور و همکاران (۱۴۰۱) نشان می‌دهد که انطباق فرهنگی آزمون ریاضی PAL-2 در بافت ایرانی توانسته است ساختار بنیادین این ابزار را حفظ کند؛ با این تفاوت که پژوهش حاضر به‌طور خاص بر آزمون تغییرناپذیری اندازه‌گیری برحسب جنسیت متمرکز بوده است.

در سطح تغییرناپذیری متریک، ثابت ماندن بارهای عاملی نشان داد که روابط بین شاخص‌ها و سازه‌های نهفته در دو گروه مشابه است؛ بنابراین، مقایسه روابط ساختاری میان متغیرها در دختران و پسران امکان‌پذیر خواهد بود. این یافته با دیدگاه‌های فلویید و ویدامان^۴ (۱۹۹۵) درباره ضرورت ثبات بارهای عاملی برای تفسیر صحیح روابط بین متغیرها مطابقت دارد؛ افزون بر این، یافته‌های پژوهش حاضر با مطالعات داخلی نظیر قاسمی و نجفی (۱۳۹۹) همسو است که بر اهمیت تغییرناپذیری متریک در تحلیل‌های چندگروهی تأکید کرده‌اند. از لحاظ نظری نیز پالس و هوتن (۲۰۰۴) و اسمیت^۵ و همکاران (۲۰۲۰) معتقدند فرایندهای شناختی پایه، مانند حافظه کاری، پردازش عددی و استدلال ریاضی^۶، معمولاً در گروه‌های جنسیتی مختلف از ثبات نسبی برخوردار هستند؛ یافته‌ای که در پژوهش حاضر نیز تأیید شد. بنا بر گفته‌های پیشین، یافته‌های این مطالعه نشان می‌دهد که نشانگرهای ریاضی و شناختی PAL-2 در دو گروه دختران و پسران به شیوه‌ای مشابه با عامل‌های زیربنایی خود مرتبط هستند.

در سطح تغییرناپذیری اسکالر، نتایج نشان داد که عرض از مبدأهای نشانگرها در دو گروه دختران و پسران تفاوت معناداری ندارند. این یافته یکی از مهم‌ترین نتایج پژوهش حاضر است؛ زیرا تغییرناپذیری اسکالر پیش‌شرط اصلی برای مقایسه میانگین‌های نهفته در گروه‌های مختلف محسوب می‌شود؛ به عبارت دیگر، تأیید این سطح از تغییرناپذیری نشان می‌دهد که تفاوت‌های احتمالی میان دختران و پسران در عامل‌های نهفته را می‌توان با اطمینان بیشتری به تفاوت در سازه‌های زیربنایی نسبت داد، نه به سوگیری نظام‌مند در نقطه شروع نشانگرها (میلسپ، ۲۰۱۱)؛ بنابراین، سنجش ریاضی PAL-2 از نظر روان‌سنجی مبنای لازم برای مقایسه میانگین‌های نهفته مهارت‌های اختصاصی ریاضی و مهارت‌های شناختی مرتبط با ریاضی را در دو گروه جنسیتی فراهم می‌کند. این نتیجه با ادبیات کلاسیک تغییرناپذیری اندازه‌گیری همسو است؛ زیرا در ادبیات تأکید شده است که پس از تأیید تغییرناپذیری پیکربندی و متریک، برقراری تغییرناپذیری اسکالر شرط ضروری برای تفسیر معتبر تفاوت‌های میانگین نهفته میان گروه‌ها است (مردیت، ۱۹۹۳؛ وندبرگ و لنس، ۲۰۰۰). از این منظر، یافته‌های پژوهش حاضر نشان می‌دهد که سنجش ریاضی PAL-2 نه تنها از نظر ساختار عاملی و بارهای عاملی در دو گروه جنسیتی ثبات دارد، بلکه سطح پایه نشانگرهای آن نیز در دختران و پسران معادل است.

با وجود این، تغییرناپذیری سختگیرانه به‌طور کامل تأیید نشد و تفاوت‌هایی در خطاهای اندازه‌گیری مشاهده گردید. این یافته نشان می‌دهد ساختار کلی آزمون و روابط بین سازه‌ها در دو گروه پایدار است؛ اما ممکن است برخی سؤالات در دختران و پسران با دقت یا حساسیت متفاوتی پاسخ داده شوند. چنین الگویی با یافته‌های مارش و همکاران (۲۰۱۴) همسو است که بیان می‌کنند دستیابی به تغییرناپذیری کامل در سطح خطای اندازه‌گیری در مطالعات آموزشی و روان‌شناختی دشوار است و محقق نشدن آن لزوماً به معنای بی‌اعتباری ابزار نیست. براون و گرین^۷ (۲۰۱۸) نیز اشاره کرده‌اند که تفاوت‌های جزئی در خطاهای اندازه‌گیری می‌تواند ناشی از تفاوت در سبک‌های پاسخ‌دهی^۸، تجارب آموزشی یا نحوه تفسیر سؤالات باشد؛ بنابراین، یافته‌های پژوهش حاضر نشان می‌دهد که آزمون PAL-2 برای مقایسه ساختارهای شناختی و روابط بین سازه‌ها در دو جنسیت مناسب است؛ اما در تفسیر و مقایسه مستقیم میانگین نمرات مشاهده‌شده باید با احتیاط عمل کرد.

به‌علاوه، یافته‌های مربوط به پایایی ابزار نشان داد که خرده‌آزمون‌های چندسؤالی سنجش ریاضی PAL-2 از همسانی درونی برخوردار هستند. با این حال، تفسیر آلفای کرونباخ در این پژوهش باید با احتیاط انجام شود؛ زیرا برخی خرده‌آزمون‌های PAL-2 ماهیت عملکردی دارند، برخی تک‌سؤالی هستند و برخی دیگر از تعداد سؤالات محدودی تشکیل شده‌اند. در چنین شرایطی، آلفای کرونباخ همیشه شاخص کاملی برای برآورد پایایی نیست؛ برای مثال، مقدار پایین‌تر آلفا در خرده‌آزمون‌هایی مانند حافظه کاری فضایی می‌تواند تا حدی ناشی از تعداد اندک سؤالات باشد، نه الزاماً ضعف جدی در کیفیت اندازه‌گیری؛ بنابراین، نتایج پایایی در این مطالعه باید به‌عنوان شاخصی توصیفی از همسانی درونی خرده‌آزمون‌های چندسؤالی تفسیر شود. متعاقباً

1 McDonald

2 Van de Ve Vijver & Leung

3 Berry

4 Floyd & Widaman

5 Smith

6 Mathematical Reasoning

7 Green

8 Response Styles

پیشنهاد می‌شود که در پژوهش‌های بعدی، در کنار آلفای کروناخ، از شاخص‌هایی مانند پایایی بازآزمایی، توافق بین ارزیابان، پایایی مرکب یا امگا^۱ نیز استفاده شود تا تصویر دقیق‌تری از ثبات اندازه‌گیری این ابزار به دست آید.

در مجموع، نتایج پژوهش حاضر مؤید آن است که سنجش ریاضی PAL-2 از ثبات مفهومی و ساختاری مطلوبی تا سطح اسکالر در بافت آموزشی ایران برخوردار است و می‌تواند به‌عنوان ابزاری معتبر برای ارزیابی مهارت‌های ریاضی و شناختی دانش‌آموزان مورد استفاده قرار گیرد. این یافته با چارچوب نظری سنجش فرایندی در آموزش ریاضی (کوهن^۲ و همکاران، ۲۰۱۷؛ اسمیت و همکاران، ۲۰۲۰) هم‌راستا است و نشان می‌دهد که این آزمون توانسته است سازه‌های مورد نظر را تا حد زیادی مستقل از سوگیری جنسیتی اندازه‌گیری کند. با این حال، چون تغییرناپذیری سختگیرانه تأیید نشد، مقایسه نمرات خام یا مشاهده‌شده باید با احتیاط تفسیر شود. از نظر کاربردی، نتایج این پژوهش می‌تواند در طراحی برنامه‌های غربالگری آموزشی، شناسایی دانش‌آموزان دارای مشکلات ریاضی و تدوین مداخلات آموزشی مبتنی بر شواهد، مورد استفاده قرار گیرد. همچنین آگاهی معلمان، مشاوران و متخصصان سنجش از ویژگی‌های روان‌سنجی سنجش ریاضی PAL-2 می‌تواند به تفسیر دقیق‌تر عملکرد دانش‌آموزان و طراحی برنامه‌های تقویتی متناسب کمک کند.

با وجود این، پژوهش حاضر با محدودیت‌هایی نیز همراه بود. نخست، جامعه آماری تنها به دانش‌آموزان شهر تهران محدود شد و این مسئله ممکن است تعمیم‌پذیری نتایج به سایر مناطق کشور، به‌ویژه مناطق روستایی یا دارای تفاوت‌های فرهنگی و اقتصادی را محدود کند. دوم، در این مطالعه فقط متغیر جنسیت به‌عنوان متغیر گروه‌بندی بررسی شد و عواملی مانند زبان مادری، وضعیت اقتصادی - اجتماعی، قومیت و سطح تحصیلات والدین، مورد تحلیل قرار نگرفتند، در حالی که این متغیرها می‌توانند بر عملکرد شناختی و پاسخ‌دهی به آزمون اثرگذار باشند. سوم، پژوهش حاضر از طراحی مقطعی^۳ استفاده کرد؛ پس امکان بررسی پایداری ساختار عاملی در طول زمان فراهم نبود؛ علاوه بر این، مطالعه حاضر فقط بر بخش سنجش ریاضی PAL-2 متمرکز بود؛ بنابراین، نتایج آن نباید به کل مجموعه آزمون PAL-2 تعمیم داده شود. همچنین با توجه به ماهیت عملکردی برخی خرده‌آزمون‌ها، استفاده از شاخص‌های پایایی متنوع‌تر می‌توانست تصویر کامل‌تری از ثبات اندازه‌گیری ابزار فراهم کند.

بر این اساس، پیشنهاد می‌شود پژوهش‌های آینده تغییرناپذیری عاملی سنجش ریاضی PAL-2 را در نمونه‌های گسترده‌تر، چندفرهنگی و چندمنطقه‌ای بررسی کنند تا میزان تعمیم‌پذیری یافته‌ها افزایش یابد. همچنین مطالعه نقش متغیرهایی مانند دوزبانگی، وضعیت اقتصادی - اجتماعی و تفاوت‌های آموزشی می‌تواند درک دقیق‌تری از عوامل مؤثر بر عملکرد دانش‌آموزان فراهم آورد. استفاده از طراحی‌های طولی^۴ نیز می‌تواند به بررسی پایداری ساختار آزمون در دوره‌های مختلف رشد شناختی کمک کند؛ علاوه بر این، بهره‌گیری از روش‌های آماری پیشرفته‌تر مانند مدل‌های چندسطحی^۵ یا بی‌زی^۶ می‌تواند دقت تحلیل‌ها را افزایش دهد. در سطح کاربردی نیز پیشنهاد می‌شود نسخه‌های بومی‌شده گسترده‌تر آزمون PAL-2 برای مقاطع تحصیلی بالاتر تدوین گردد و نتایج حاصل از این ابزار در سیاست‌گذاری‌های آموزشی، طراحی برنامه‌های مداخله‌ای و آموزش معلمان مورد استفاده قرار گیرد. در نهایت، یافته‌های این پژوهش نشان می‌دهد که سنجش ریاضی PAL-2 می‌تواند به‌عنوان ابزاری معتبر و کاربردی در ارزیابی توانایی‌های ریاضی دانش‌آموزان ایرانی مورد استفاده قرار گیرد؛ هرچند بررسی عمیق‌تر عوامل فرهنگی، اجتماعی و آموزشی در مطالعات آینده همچنان ضروری است.

سپاس‌گزاری

نویسندگان از همه شرکت‌کنندگان و والدین ایشان که در تکمیل آزمون‌ها مشارکت و همکاری کردند، تشکر می‌کنند.

ملاحظات اخلاقی

این پژوهش با رعایت موازین اخلاقی و با اخذ رضایت آگاهانه و حفظ محرمانگی اطلاعات مشارکت‌کنندگان انجام شد. شرکت در پژوهش کاملاً داوطلبانه بود و تمامی افراد پیش از ورود به مطالعه، از اهداف و روش‌های آن به‌طور کامل آگاهی یافتند.

حامی مالی

نویسندگان تمامی مراحل این پژوهش و انتشار آن را به‌طور مستقل و بدون دریافت هیچ‌گونه کمک مالی از منابع بیرونی، تأمین مالی کرده‌اند.

مشارکت نویسندگان

این مقاله برگرفته از پایان‌نامه کارشناسی ارشد نویسنده اول است. نویسنده اول، ایده‌پردازی، جمع‌آوری داده‌ها، تحلیل رسمی و نگارش پیش‌نویس مقاله را انجام داده است. نویسنده دوم نقش نظارت بر پژوهش، مشارکت در طراحی مطالعه و روش‌شناسی و نیز بازبینی انتقادی و ویرایش مقاله را بر عهده

¹ Omega

² Cohen

³ Cross-sectional

⁴ Longitudinal

⁵ Multilevel

⁶ Bayesian

داشته است. نویسنده سوم به عنوان مشاور در تدوین چارچوب مفهومی مشارکت کرده و بازخورد ارزشمندی درباره مقاله ارائه نموده است. همه نویسندگان نسخه نهایی مقاله را تأیید کرده‌اند.

تعارض منافع

نویسندگان اعلام می‌دارند که این پژوهش هیچ‌گونه تعارض منافی ندارد.

منابع

- خانی‌پور، ح.، رادفر، م.، پورطریق، م. و گرامی‌پور، م. (۱۴۰۱). انطباق‌یابی آموزشی و فرهنگی آزمون ریاضی پال ۲ (سنجش فرایندی تشخیصی ریاضی یادگیرنده) در ایران. *فصلنامه روان‌شناسی شناختی*. ۱۰(۲)، ۷۲-۸۶.
- قاسمی، آ و نجفی، م. ر. (۱۳۹۹). بررسی تغییرناپذیری عاملی آزمون‌های استاندارد سنجش مهارت‌های خواندن در دانش‌آموزان ایرانی. *پژوهش‌های آموزشی ایران*. ۱۶(۱)، ۴۴-۵۸.
- گرامی‌پور، م. (۱۳۹۳). *مبانی نظری و کاربرد نظریه‌های اندازه‌گیری در علوم رفتاری*. تهران: انتشارات تمدن علمی.

References

- Berry, J. W. (2003). Conceptual approaches to cross-cultural psychology: Issues and perspectives. In M. H. Bond (Ed.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 13-45). Allyn & Bacon.
- Brown, G., & Green, T. (2018). *Educational assessment: A practical approach*. Pearson.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. Routledge.
- Floyd, R. G., & Widaman, K. F. (1995). Reliability and construct validity of cognitive assessment instruments. *Psychological Assessment*, 7(2), 212-225.
- Geramipour, M. (2014). *Theoretical foundations and applications of measurement theories in behavioral sciences*. Tehran: Tamaddon-e- Elmi Publications. (Persian)
- Ghasemi, A., & Najafi, M. R. (2020). Examining the factorial invariance of standardized reading skills assessment tests among Iranian students. *Iranian Educational Research*, 16(1), 44-58. (Persian)
- Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 205-223
- Khanipour, H., Radfar, M., Pourtorogh, M., & Geramipour, M. (2022). Cultural adaptation of process assessment of the learner: Diagnostic assessment for math (Pall-II Math) in Iran. *Journal of Cognitive Psychology*, 10(2), 72-86. (Persian)
- Marsh, S. J., et al. (2014). Testing measurement invariance across gender in standardized achievement tests. *Educational Measurement: Issues and Practice*, 33(4), 39-46.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.
- Pals, J. L., & Houghton, D. A. (2004). Examining the reliability and validity of the PAL 2 test for cognitive abilities. *Journal of Educational Psychology*, 96(4), 784-798.

- Peterson, L. S., Martinez, A., & Turner, T. L. (2010). Test review: process assessment of the learner. *Journal of Psychoeducational Assessment*, 28(1), 80-86.
- Smith, R., Jones, M., & Taylor, F. (2020). Process-based diagnostic testing in mathematics education: The case of PAL-2. *Journal of Educational Psychology*, 112(2), 159-172.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Sage Publications.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-69.
- Zibulsky, J. and Veizel, K. (2014). Process Assessment of the Learner-Second Edition: Diagnostic Assessment for Math (PAL-II Math). In Encyclopedia of Special Education (eds C.R. Reynolds, K.J. Vannest and E. Fletcher-Janzen).