



## Research Paper

# Measurement Invariance Across Gender of Wechsler Intelligence Scale for Children- Fourth Edition in Iranian Students

Noshiravan Mohammadi<sup>1\*</sup> 

1.PhD in Psychological Measurement, Department of Educational and Psychological Measurement, Faculty of Psychology and Education, Allameh Tabataba'i University, Tehran, Iran

### Article info:

Received: 06.05.2024

Revised: 08.04.2025

Accepted: 04.07.2025

### Keywords:

psychometrics,  
measurement invariance,  
cognitive ability, intelligence



Publisher: University of Zanjan

### Abstract

The Wechsler Intelligence Test, and specifically the fourth edition of the Wechsler Intelligence Scale for Children (WISC-IV), is one of the most widely used intelligence tests in the world. The aim of the present study was to examine gender measurement invariance in the Wechsler Intelligence Scale for Children (WISC-IV). The research methodology was applied in terms of its objective and was categorized as secondary analysis in terms of data collection. To achieve this, a random sample of 1,222 Iranian students (622 boys and 600 girls) was selected, and the Wechsler test was administered to them. The data collection tool was the fourth edition of the Wechsler Intelligence Scale for Children (WISC-IV). The data were analyzed using LISREL 8.7 software. For data analysis, multigroup confirmatory factor analysis was used. The results showed that the Wechsler Intelligence Scale exhibited all four types of invariances: configural invariance (RMSEA=0.037,  $\chi^2=321.67$ ), weak invariance (RMSEA=0.031,  $\chi^2=270$ ), strong invariance (RMSEA=0.034,  $\chi^2=326.76$ ), and strict invariance (RMSEA=0.032,  $\chi^2=326.76$ ). The invariance results indicate that the factor scores of the WISC-IV and its subscales have the same meaning for both boys and girls student groups. Therefore, the obtained results can be interpreted equally for both groups, allowing for meaningful comparisons between boys and girls.

Use your device to scan and  
read the article online



**Citation:** Mohammadi, N. (2025). Measurement invariance across gender of Wechsler Intelligence Scale for Children- fourth edition in Iranian students. *Iranian Journal of Psychoeducational Assessment*, 1 (1), 144-155. <https://doi.org/10.30470/ijpa.2025.726687>

**\*Corresponding Author:** Noshiravan Mohammadi

**Address:** Department of Assessment and Measurement, Faculty of Psychology and Education, Allameh Tabataba'i University, Tehran, Iran

**Email:** noshirwanm@gmail.com

## Extended Abstract

### Introduction

The Wechsler Intelligence Scale, particularly the fourth edition of the *Wechsler Intelligence Scale for Children (WISC-IV)*, is among the most widely used cognitive assessment tools worldwide. These tests are currently employed in approximately 20 countries to measure intelligence. According to exploratory and confirmatory factor analyses, this instrument consistently follows a four-factor structure across all these countries (Reverte et al., 2015). Given the high professional and ethical responsibilities associated with using such tests for diagnosis, intervention, and decision-making, it is imperative to assess their fairness prior to large-scale implementation (Rodríguez-Cancino & Concha-Salgado, 2023).

Effective measurement requires that these instruments reliably and validly assess a given characteristic or construct across diverse individuals, as well as within the same individual over time and under varying conditions. Therefore, measurement invariance is a critical prerequisite for the scientific and practical application of assessment tools. The question “*Does this instrument provide an invariant measure?*” remains largely unexplored in the behavioral sciences. Hypotheses regarding measurement invariance are seldom tested, and many researchers implicitly assume that the tests are invariant. As a result, the focus often shifts toward designing clear and unambiguous items, selecting optimal content, and presuming that these items are interpreted similarly across varied groups. Such assumptions lead to the belief that tests function equivalently across age groups and classifications. However, if this assumption proves invalid, the conclusions derived from these instruments may be misleading. It is only in recent years, with increased attention to the issue, that a handful of studies have begun to empirically test the hypothesis of invariance (Blankson & McArdle, 2013).

### Methods

This study utilizes data from the fourth edition of the Wechsler Intelligence Scale for Children (WISC-IV), originally collected by the Iranian Exceptional Education Organization. Accordingly, the analysis conducted here in qualifies as a secondary analysis. The statistical population consists of all elementary school students across the country. A total of 1,222 individuals participated in the study, comprising 622 boys and 600 girls, selected by a simple random sampling method.

The primary measurement instrument is the *Wechsler Intelligence Scale for Children, Fourth Edition*, designed for children aged 6 to 16. The test includes 15 subtests: 10 core and 5 supplemental. Each subtest is standardized with a mean of 10 and a standard deviation of 3. Additionally, the instrument provides a Full-Scale Intelligence Quotient (FSIQ), alongside four index scores: Verbal Comprehension Index (VCI), Conceptual Reasoning Index (CRI), Working Memory Index (WMI), and Processing Speed Index (PSI). The total score is standardized to a mean of 100 and a standard deviation of 15.

To investigate measurement invariance, the study employs Confirmatory Factor Analysis (CFA) using LISREL 8.7 software. This analytical approach enables the evaluation of whether test items function equivalently across different groups within the sampled population.

### Results

The analytical process was conducted in several structured stages. Initially, the Confirmatory Factor Analysis (CFA) model was tested independently within the two gender groups (boys and girls) to assess baseline model fit, with results yielding acceptable indices and confirming the integrity of the underlying factor structure in both groups. This was followed by a Multi-Group Confirmatory Factor Analysis (MGCFA) to evaluate the model simultaneously across the groups, allowing systematic investigation of measurement invariance through progressive equality constraints. Model parameters were estimated using the Maximum Likelihood (ML) method, a robust approach for multivariate normal data and commonly used in structural equation modeling. Invariance testing was performed at the subscale level using the hierarchical sequence

of configural, weak (metric), strong (scalar), and strict invariance. The WISC-IV demonstrated satisfactory model fit across all levels, Configural:  $\chi^2 = 321.67$ , RMSEA = 0.037; Metric:  $\chi^2 = 270.00$ , RMSEA = 0.031; Scalar:  $\chi^2 = 326.76$ , RMSEA = 0.034; Strict:  $\chi^2 = 326.76$ , RMSEA = 0.032, which supports the conclusion that the instrument maintains measurement invariance across gender groups. This confirmation adds rigor to interpretations of group differences and validates the use of WISC-IV for meaningful cross-group cognitive assessments.

### Conclusion

This study was conducted to examine the measurement invariance of the constructs assessed by the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) across boys and girls student groups. The results revealed that the four-factor model demonstrated acceptable fit indices in both boys' and girls' samples, supporting the gender invariance of the instrument and indicating that the WISC-IV measures cognitive abilities equivalently across genders. Model invariance was confirmed across all tested levels configural, metric, scalar, and strict demonstrating that the factor scores and subscales maintain consistent meaning between the two groups. These findings validate the use of WISC-IV for meaningful cross-gender comparisons in cognitive assessment. Nonetheless, the study's scope was limited to gender as the grouping variable, and future research is encouraged to extend invariance testing across other subpopulations, including distinct age brackets and clinical groups such as individuals diagnosed with ADHD. Such broader investigations would enhance the generalizability and applicability of the instrument and provide deeper insight into its psychometric properties across diverse populations.

### Ethical Considerations

This study utilized secondary data obtained from the Iranian Exceptional Education Organization; therefore, there was no direct contact with the participants. Nevertheless, confidentiality of the received information was maintained.

### Acknowledgments

The researcher sincerely acknowledges the valuable collaboration of the Iranian Exceptional Education Organization, which generously provided the necessary data for this study.

### Authors' contributions

All activities related to this research from ideation, data collection, analysis, and conclusion to manuscript writing were carried out by Noshiravan Mohammadi.

### Conflicts of interest

The current study has no financial support or conflicts of interest.



## مقاله پژوهشی

تغییرناپذیری اندازه‌گیری جنسیت در آزمون هوش و کسلر کودکان-نسخه چهارم در  
بین دانش‌آموزان دختر و پسر ایرانینوشیروان محمدی<sup>۱\*</sup>

۱. دکترای سنجش و اندازه‌گیری، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران

## چکیده

آزمون هوش و کسلر و به‌طور خاص نسخه‌ی چهارم آزمون هوش و کسلر برای کودکان، از آزمون‌هایی است که بیشترین استفاده را در جهان دارد. هدف از پژوهش حاضر بررسی نامتغیر بودن اندازه‌گیری جنسیت در آزمون هوش و کسلر کودکان نسخه چهارم بود. روش پژوهش از لحاظ هدف کاربردی و از نظر شیوه جمع‌آوری داده‌ها از نوع تحلیل‌های ثانویه است. به همین منظور از جامعه دانش‌آموزان ایرانی ۱۲۲۲ (۶۲۲ نفر پسر، ۶۰۰ نفر دختر) نفر به صورت تصادفی انتخاب شدند و آزمون و کسلر بر روی آن‌ها اجرا گردید. ابزار گردآوری داده‌ها، آزمون هوش و کسلر کودکان نسخه چهارم بود. داده‌ها با استفاده از نرم افزار لیزرل ۸.۷ تحلیل شدند. برای تحلیل داده‌های به دست آمده از روش تحلیل عامل تأییدی چندگروهی استفاده شد. نتایج نشان داد که آزمون هوش و کسلر از هر چهار نوع نامتغیر بودن یعنی نامتغیر بودن شکل ( $RMSEA=0/037$ ),  $x^2=321/67$ ، ضعیف ( $x^2=321/67$ ),  $RMSEA=0/031$ ، قوی ( $x^2=270$ ),  $RMSEA=0/034$ ، و دقیق ( $x^2=326/76$ ) و  $RMSEA=0/032$ )، نتایج نامتغیر بودن نشان می‌دهد که نمره‌های عامل‌های مربوط به آزمون هوش و کسلر کودکان نسخه چهارم و خرده مقیاس‌های مربوط به آن معنای مشابهی برای دو گروه دانش‌آموزان پسر و دختر دارد. بنابراین می‌توان نتایج به دست آمده برای این دو گروه را به صورت یکسان تفسیر کرد و مقایسه‌های معنادار بین دو گروه پسر و دختر انجام داد.

## اطلاعات مقاله:

تاریخ دریافت: ۱۴۰۳/۰۲/۱۷

تاریخ داوری: ۱۴۰۴/۰۱/۱۹

تاریخ پذیرش: ۱۴۰۴/۰۴/۱۳

## واژه‌های کلیدی:

روانسنجی، نامتغیر بودن، اندازه‌گیری،

توانایی‌های شناختی، هوش



ناشر: دانشگاه زنجان

\* استناد: محمدی، ن. (۱۴۰۴). تغییرناپذیری اندازه‌گیری جنسیت در آزمون هوش و کسلر کودکان-نسخه چهارم در بین

دانش‌آموزان دختر و پسر ایرانی. *سنجش روانی تربیتی*، (۱): ۱۴۴-۱۵۵.<https://doi.org/10.30470/ijpa.2025.726687>از دستگاه خود برای اسکن و خواندن  
مقاله به صورت آنلاین استفاده کنید.

\* نویسنده مسئول: نوشیروان محمدی

نشانی: گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران  
پست الکترونیکی: noshirwanm@gmail.com

در میان آزمون‌های روان‌شناسی، آزمون‌های استاندارد هوش از جمله ابزارهایی هستند که بیشترین استفاده را توسط روان‌شناسان دارند. روان‌شناسان مدرسه به‌طور خاص اغلب از آزمون‌های استاندارد هوش به‌عنوان یکی از اجزای اصلی ارزشیابی روان‌شناختی دانش‌آموزان برای تعیین شایستگی افراد برای آموزش‌های خاص استفاده می‌کنند (پناه<sup>۱</sup> و همکاران، ۲۰۲۵). در میان آزمون‌های استاندارد هوش موجود، آزمون وکسلر کودکان-نسخه چهارم<sup>۲</sup> دارای بیشترین استفاده است (ریچرسون<sup>۳</sup> و همکاران، ۲۰۱۴). آزمون هوش وکسلر و به‌طور خاص آزمون هوش وکسلر برای کودکان-نسخه چهارم، از آزمون‌هایی است که بیشترین استفاده را در جهان دارد. آزمون‌های هوش وکسلر تقریباً در ۲۰ کشور جهان برای اندازه‌گیری هوش استفاده می‌شوند. با توجه به تحلیل‌های عاملی اکتشافی و تأییدی این آزمون از یک ساختار چهار عاملی در همه این کشورها تبعیت می‌کند (ریورت<sup>۴</sup> و همکاران، ۲۰۱۵ و ابوباکر-عبداللطیف<sup>۵</sup>، ۲۰۲۵).

با توجه به مسئولیت حرفه‌ای و اخلاقی بالای استفاده از آزمون‌ها برای تشخیص، مداخله و تصمیم‌گیری، بررسی منصفانه بودن آزمون‌ها قبل از استفاده آنها در مقیاس بزرگ لازم و ضروری است (رودریگوئز-کانسینو و کونچا-سالگادو<sup>۶</sup>، ۲۰۲۳). استفاده از آزمون‌ها برای اندازه‌گیری نیازمند این است که این آزمون‌ها به اندازه‌گیری یک ویژگی یا سازه در میان افراد مختلف، یا یک فرد در طول زمان‌های مختلف و یا در شرایط متفاوت بپردازند. بنابراین نامتغیر بودن اندازه‌گیری<sup>۷</sup> در استفاده کاربردی و علمی از ابزارهای اندازه‌گیری یک شرط اساسی و مهم است (ابوباکر-عبداللطیف، ۲۰۲۵). هنوز هم این سؤال که آیا این ابزار اندازه‌گیری، اندازه‌گیری نامتغیر را فراهم می‌آورد؟ به‌دردت در علوم رفتاری تحقیق می‌شود و فرضیه‌های مربوط به نامتغیر بودن اندازه‌گیری خیلی کم آزمون می‌شوند. اغلب محققان فرض می‌کنند که آزمون نامتغیر است. بیشتر توجه و تمرکز در این حوزه بر روی نوشتن سؤال‌های واضح و روشن است و تحلیل‌هایی که بهترین سؤال‌های ممکن را انتخاب کند، پس از آن، فرض می‌کنند که این سؤال‌ها برای افراد مختلف معانی و تفسیرهای مشابهی دارد، پس می‌توان گفت که آزمون برای مقایسه گروه‌های سنی متفاوت و طبقه‌بندی‌های مختلف نامتغیر است. حال اگر این پیش فرض نامتغیر بودن نادرست باشد، پس احتمالاً نتایجی هم که بر اساس این ابزارهای اندازه‌گیری به‌دست می‌آیند، نادرست خواهند بود. تنها اخیراً با توجه به اهمیتی که به این موضوع داده شده است، تعداد کمی پژوهش فرضیه نامتغیر بودن را آزمون می‌کنند (بلنکسون و مک‌آردل<sup>۸</sup>، ۲۰۱۳).

نامتغیر بودن در همه پژوهش‌هایی که در آن ویژگی‌ها، سازه یا خصیصه‌های اندازه‌گیری، و رابطه میان آن خصیصه‌ها در میان گروه‌های مختلف، و یا افراد یکسانی که در شرایط مختلف (زمان و مکان) اندازه‌گیری شده‌اند، باید ارزیابی شود. نامتغیر بودن باید برای هر فردی که به‌وسیله یک ابزار اندازه‌گیری می‌شود، بررسی شود و فرضیه نامتغیر بودن باید برای گروه‌های مختلفی از افراد آزمون شود. سه طبقه‌بندی که در آن‌ها ممکن است نامتغیر بودن برای آزمون‌های شناختی وجود نداشته باشد عبارت‌اند از: قومیت<sup>۹</sup>، جنسیت و زمان (ابوباکر-عبداللطیف، ۲۰۲۵).

نامتغیر بودن به عنوان یکی از ویژگی‌های اساسی و مهم هر ابزار بالینی است که برای مقایسه افراد در گروه‌های مختلف استفاده می‌شود. به‌عبارت دیگر حتی اگر یک مقیاس در گروه‌های مختلف استفاده شود، باید یک ساختار مشابه داشته باشد (جلیک<sup>۱۰</sup> و همکاران، ۲۰۲۰). به‌طور کلی، نامتغیر بودن اندازه‌گیری به این اشاره دارد که دو فرد با یک نتایج<sup>۱۱</sup> (توانایی مکنون) یکسان باید نمره‌ی برابری در آزمون یا سؤال‌ها بگیرند، صرف‌نظر از گروهی که در آن عضویت دارند. اگر سؤال‌ها به‌صورت دوازده‌گانه<sup>۱۲</sup> نمره‌گذاری شوند، طبق مفهوم نامتغیر بودن اندازه‌گیری، افراد دو گروه باید شانس برابر برای دادن پاسخ درست داشته باشند. مِلنبرگ<sup>۱۳</sup> (۱۹۸۹) به این صورت نامتغیر بودن را تعریف می‌کند که نمره مشاهده‌شده  $X$  زمانی نسبت به گروه‌بندی نامتغیر است که:

$$P(X|\theta, \text{group}) = P(X|\theta).$$

بر اساس این تعریف، توزیع نمرات مشاهده‌شده  $X$ ، که وابسته به توانایی شناختی مکنون (نتا) است، وابسته به متغیرهای گروهی که فرد در آن عضویت دارد، نیست. اگر این معادله برقرار نباشد، آزمون نامتغیر نیست و تفاوت‌های گروهی در نمرات  $X$  را نمی‌توان به‌تنهایی به تفاوت‌های گروهی در توانایی‌های شناختی مکنون (نتا) نسبت داد. عدم رعایت نامتغیر بودن نشان‌دهنده این است که نمرات مشاهده‌شده  $X$  به‌عنوان تابعی از ویژگی‌های گروهی در کنار نتا هستند (ویچرت<sup>۱۴</sup>، ۲۰۱۶). بنابراین برای مقایسه معنادار نمرات آزمون‌ها در گروه‌های مختلف، یک آزمون باید نسبت به عضویت گروهی افراد نامتغیر باشد. وقتی که یک آزمون دارای نامتغیری اندازه‌گیری است، تفاوت نمرات افراد در گروه‌های مختلف را می‌توانیم به سازه‌ای که قصد اندازه‌گیری آن را داریم، نسبت دهیم (جک<sup>۱۵</sup>، ۲۰۱۴).

1 Panah  
2 Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV)  
3 Richerson  
4 Reverte  
5 Abubakar-Abdullateef  
6 Rodríguez-Cancino & Concha-Salgado  
7 Measurement invariance

8 Blankson & McArdle  
9 Ethnicity  
10 Celik  
11 Theta  
12 Dichotomously  
13 Mellenbergh  
14 Wicherts  
15 Jak

مردیت (۱۹۹۳)، نامتغیر بودن اندازه‌گیری را به‌عنوان پارامترهایی از مدل که وابسته به عضویت گروهی افراد است، تعریف می‌کند. برای مثال شما دو گروه در اختیار دارید که ممکن است متفاوت باشند (مردان در مقابل زنان، فرانسوی در مقابل استرالیایی) و می‌خواهید بدانید آزمون‌هایی که در اختیار دارید را می‌توانید به‌صورت برابر و مساوی در این گروه‌ها استفاده کنید و گروه‌ها را با آن مقایسه کنید، و آن‌ها را در یک متغیر مکنون مقایسه کنید یا خیر. در همه این مقایسه‌ها نامتغیر بودن اندازه‌گیری به ما می‌گوید که آیا ساختار عاملی مشابهی در دو گروه وجود دارد، و اینکه آیا بارهای عاملی<sup>۲</sup> و میانگین‌ها و توزیع خطا<sup>۳</sup> در دو گروه یکسان است. در واقع با آزمون نامتغیر بودن اندازه‌گیری به دنبال بررسی مدل معادلات ساختاری یکسانی در گروه‌ها هستیم. وقتی به مقایسه مدل‌ها می‌پردازیم، یک روش معمول مقایسه یک مدل با یک نسخه محدودشده<sup>۴</sup> از همان مدل است. وقتی که به آزمون نامتغیر بودن اندازه‌گیری می‌پردازیم، به برآزش یک مدل با همه داده‌ها زمانی که پارامترهای مشخصی را ثابت<sup>۵</sup> کرده‌ایم تا در گروه‌ها برابر باشند، با مدلی می‌پردازیم که در آن همه پارامترهای آزاد<sup>۶</sup> می‌باشند. اگر مدلی که دارای پارامترهای آزاد است برآزش بهتری نسبت به مدلی که دارای پارامترهای ثابت است، داشته باشد، این پارامترها در بین گروه‌ها نامتغیر نیستند. باید یادآوری شود که برای این کار باید از ماتریس کوواریانس نه ماتریس همبستگی استفاده کرد (هورتینسیوس<sup>۷</sup>، ۲۰۱۲).

با توجه به اهمیت نامتغیر بودن، به جای بررسی و ارزیابی آن در پژوهش‌های رفتاری، اغلب نادیده گرفته شده است. اکنون می‌دانیم که نامتغیر بودن ممکن است رعایت نشود، پس نمی‌توان به‌راحتی آن را فرض کرد و باید آن را به‌صورت یک فرضیه آزمون کرد. حداقل چهار نوع نامتغیر بودن باید درباره یک آزمون بررسی شود: شکلی<sup>۸</sup>، ضعیف<sup>۹</sup>، قوی<sup>۱۰</sup> و دقیق<sup>۱۱</sup>. در مورد اول، باید نشانگرهای یکسان بر روی عامل‌های یکسان در گروه‌های مختلف بار عاملی داشته باشند. در نامتغیر بودن ضعیف یا مقیاسی<sup>۱۲</sup>، بارهای عاملی باید در گروه‌های مختلف یکسان باشند. دو نوع آخر با واریانس و کوواریانس مشاهده سروکار دارند. نامتغیر بودن قوی (اسکالر<sup>۱۳</sup>) باید بارهای عاملی و عرض از مبدأ<sup>۱۴</sup> متغیرهای نشانگر نامتغیر باشد. در نهایت در نامتغیر بودن دقیق، همه بارهای عاملی، عرض از مبدأ<sup>۱۴</sup> متغیرهای نشانگر، و باقی‌مانده‌های متغیرهای نشانگر باید در گروه‌های مختلف یکسان باشد (بلنکسون و مک‌آردل، ۲۰۱۳).

آزمون نامتغیر بودن اندازه‌گیری با استفاده از تحلیل عامل تأییدی که به نامتغیر بودن عاملی محدود شناخته می‌شود، در دهه‌های اخیر محبوبیت زیادی را به‌عنوان ابزاری برای مطالعه منصفانه بودن آزمون‌های توانایی‌های شناختی مانند وکسلر کسب کرده است. وقتی که نامتغیر بودن برقرار نباشد، توانایی شناختی مکنونی که آزمون یا سوالات می‌خواهند آن را اندازه‌گیری کنند، نمی‌تواند همه تفاوت‌های گروهی را تبیین کند. عواملی که باعث تورش در نمرات آزمون می‌شوند گسترده و متنوع هستند، بعضی از آن‌ها به مشکلاتی در زمینه‌ی اجرای آزمون مربوط می‌شود، موضوعات فرهنگی در رابطه با محتوای سوالات، تفاوت‌های آموزشی، تفاوت‌های زبانی و عوامل شناختی و عاطفی از عوامل دیگر هستند (ویچرت، ۲۰۱۶).

تحقیقات زیادی به بررسی نامتغیر بودن در آزمون‌های شناختی پرداخته‌اند. ویچرت و دولان<sup>۱۵</sup> (۲۰۱۰) در پژوهشی که به بررسی نامتغیر بودن در آزمون هوش کودکان هلند در اقلیت‌های قومی پرداخته بودند، به این نتیجه رسیدند که میانگین نمره در قسمت کلامی برای اقلیت‌های قومی پایین بود و این به خاطر تفاوت دانشی درباره لغات در زبان هلندی بود. نوتبلی<sup>۱۶</sup> و همکاران (۲۰۱۰) به این نتیجه رسیدند که کشورهای آفریقایی نمرات پایین‌تری نسبت به کشورهای غربی در نسخه‌های مختلف آزمون هوش ریون به دست می‌آورند. همچنین نتایج ده مطالعه درباره ماهیت عاملی آزمون ریون در میان آفریقایی‌ها نشان داد آزمون چندین عامل زیر بنایی را به‌جای یک عامل استدلال سیالی اندازه می‌گیرد. چن<sup>۱۷</sup> و همکاران (۲۰۱۵) به بررسی نامتغیر بودن آزمون هوش وکسلر نسخه چهار برای متغیر جنسیت در امریکا پرداختند و به این نتیجه رسیدند که آزمون وکسلر نسبت به جنسیت نامتغیر است. چن و ژو<sup>۱۸</sup> (۲۰۱۲) به بررسی نامتغیر بودن آزمون هوش وکسلر نسخه چهار در نمونه نرمال و بالینی در امریکا پرداختند و به این نتیجه رسیدند آزمون دارای نامتغیر بودن جزئی<sup>۱۹</sup> است. دولان (۲۰۰۰) به نامتغیر بودن آزمون نسخه تجدیدنظر شده وکسلر در جامعه سیاهان و سفیدپوستان امریکا پرداخت و به این نتیجه رسید که آزمون نامتغیر است. نیلکسلا<sup>۲۰</sup> و همکاران (۲۰۱۳) به بررسی نامتغیر بودن آزمون هوش وکسلر نسخه چهار از نظر متغیر سن در امریکا پرداخته بودند و به این نتیجه رسیدند که آزمون دارای نامتغیر بودن جزئی است. توماسی<sup>۲۱</sup> و همکاران (۲۰۱۵) نامتغیر بودن آزمون هوش تجدیدنظر شده وکسلر را از نظر متغیر سطح تحصیلات در ایتالیا بررسی کرد و به این نتیجه دست‌یافتند که آزمون نامتغیر است. رینولدز<sup>۲۲</sup> و همکاران (۲۰۱۳) نامتغیر بودن آزمون هوش وکسلر نسخه چهار را در افراد نرمال و افراد دارای ناتوانی‌های شناختی بررسی کردند و مشخص شد که آزمون دارای نامتغیر بودن

1 Merdith

2 Factor loadings

3 Error distribution

4 Restricted version

5 Fixed

6 Free

7 Hortensius

8 Configural

9 Weak

10 Strong

11 Strict

12 Metric

13 Scalar

14 Intercepts

15 Dolan

16 Notably

17 Chen

18 Chen &amp; Zhu

19 Partial

20 Niileksela

21 Tommasi

22 Reynolds

جزئی است. ریورت و همکاران (۲۰۱۵) به بررسی نامتغیر بودن چند گروهی ساختار عاملی آزمون هوش و کسلر چهار در دو کشور فرانسه و سوئیس پرداختند. نتایج نشان داد که آزمون نامتغیر بودن جزئی دارد، آزمون دارای نامتغیر بودن شکلی و ضعیف بود اما نامتغیر بودن قوی را نداشت. والتر<sup>۱</sup> و همکاران (۲۰۲۱) نامتغیر بودن جنسیت در آزمون و کسلر پیش دبستان در آلمان را بررسی کردند. نتایج نشان داد آزمون و کسلر پیش دبستان نسبت به جنسیت نامتغیر است و این آزمون توانایی‌های شناختی را برای دختران و پسران به یک صورت اندازه‌گیری می‌کند. در نهایت، بکر<sup>۲</sup> و همکاران (۲۰۲۴) به بررسی نامتغیر بودن آزمون هوش و کسلر در نمونه بالینی از کودکان و بزرگسالان با بیش‌فعالی و گروه کنترل با روش تحلیل عامل تاییدی چندگروهی پرداختند. نتایج نشان داد آزمون در دو گروه به صورت یکسان عمل می‌کند.

پس می‌توان گفت نامتغیر بودن ویژگی اصلی هر ابزاری است که برای مقایسه افراد در گروه‌های مختلف استفاده می‌شود. آزمون و کسلر معمولاً در ارزیابی‌های روان‌شناختی استفاده می‌شود. با توجه به کاربرد این آزمون فرض بر این است که مقیاس‌های آزمون هوش و کسلر معنای مشابهی برای افراد در گروه‌های مختلف دارد. بنابراین بررسی نامتغیر بودن اندازه‌گیری این آزمون از اهمیت بالایی برخوردار است. در میان طبقه‌بندی‌های گروهی مختلف، نامتغیر بودن از نظر جنسیت به‌عنوان یک ویژگی بنیادی برای اندازه‌گیری در حوزه‌های مختلف در نظر گرفته می‌شود (چن و همکاران، ۲۰۱۵). بنابراین با توجه به اهمیت موضوع اندازه‌گیری یکسان در آزمون‌ها، هدف از پژوهش حاضر بررسی نامتغیر بودن اندازه‌گیری آزمون هوش و کسلر کودکان نسخه چهار بر اساس جنسیت در دانش‌آموزان ایرانی است.

## روش پژوهش

روش این پژوهش از نظر هدف کاربردی است. همچنین از آنجایی که در این تحقیق از داده‌های آزمون هوش و کسلر نسخه چهار که در سازمان آموزش و پرورش استثنائی گردآوری شده بود، استفاده شده است؛ بنابراین از نظر جمع‌آوری داده‌ها از نوع تحلیل‌های ثانویه است.

## مشارکت‌کنندگان

جامعه آماری این پژوهش از کلیه دانش‌آموزان ابتدایی سراسر کشور تشکیل می‌شود. نمونه آماری این پژوهش برابر با ۱۲۲۲ نفر متشکل از ۶۲۲ نفر پسر و ۶۰۰ نفر دختر است. این افراد با استفاده از روش نمونه‌گیری تصادفی ساده انتخاب شده‌اند. دامنه سنی شرکت‌کنندگان در این پژوهش ۷ تا ۱۲ سال و میانگین سنی آنها برابر ۱۰/۳۷ بوده است.

## ابزار پژوهش

ابزار اندازه‌گیری در این پژوهش آزمون و کسلر کودکان-نسخه چهارم بود. این آزمون برای دامنه سنی ۶ تا ۱۶ استفاده می‌شود. از ۱۵ خرده‌مقیاس تشکیل شده است، ۱۰ مقیاس اصلی و ۵ مقیاس مکمل، هر مقیاس دارای میانگین ۱۰ و انحراف استاندارد ۳ است. این آزمون دارای یک نمره هوش بهر کلی<sup>۳</sup> (FSIQ) برای کل مقیاس و چهار شاخص نمره‌گذاری: درک کلامی<sup>۴</sup> (VCI)، استدلال مفهومی<sup>۵</sup> (PRI)، حافظه فعال<sup>۶</sup> (WMI)، سرعت پردازش<sup>۷</sup> (PSI) است. میانگین نمره کل برابر ۱۰۰ و انحراف استاندارد آن ۱۵ است.

**روایی:** شواهد زیادی در مورد روایی آزمون هوش و کسلر وجود دارد. نتایج تحلیل‌های مختلفی را برای این آزمون نشان می‌دهد. همچنین آزمون هوش و کسلر همبستگی بالایی با سایر آزمون‌های هوش را اندازه‌گیری می‌کنند دارد که این نشان از روایی واگرایی این آزمون است. برای مثال همبستگی آن با آزمون هوش سیال-متبلور کافمن برای کودکان<sup>۸</sup> برابر ۰/۸۹ و با آزمون پردازش ذهنی کافمن برای کودکان<sup>۹</sup> برابر ۰/۸۸ و همچنین با مقیاس توانایی‌های افتراقی<sup>۱۱</sup> برابر ۰/۸۴ است که همبستگی‌های بالایی محسوب می‌شود و نشان از روایی بالای آن دارد (والستروم<sup>۱۲</sup> و همکاران، ۲۰۱۲).

**پایایی:**<sup>۱۳</sup> مقدار ضریب پایایی به‌دست آمده برای تمامی خرده‌مقیاس‌ها به‌جز دو خرده‌مقیاس برابر ۰/۸ تا ۰/۹ است. برای خرده‌مقیاس‌های تکمیل خط‌زنی و نمادبایی برابر ۰/۷۹ است. همچنین ضریب بازآزمایی برای نمره کل آزمون برابر ۰/۹۲ برای چهار شاخص اصلی برابر ۰/۸۴ تا ۰/۹۳ و برای خرده‌مقیاس‌ها برابر دامنه‌ای از ۰/۷۴ تا ۰/۹۲ است (والستروم و همکاران، ۲۰۱۲).

در ایران هم روایی و پایایی این آزمون بررسی شده است. صادقی و همکاران (۱۳۹۰) به بررسی روایی و پایایی آزمون در کودکان ۶ تا ۱۶ سال استان چهارمحال و بختیاری پرداختند. برای بررسی روایی آزمون از آزمون هوش و کسلر مقیاس تجدید نظر شده کودکان (۱۹۷۴) و آزمون ماتریس‌های

1 Walter

2 Becker

3 Full Scale IQ

4 Verbal comprehension index

5 Perceptual reasoning index

6 Working memory index

7 Processing speed index

8 Validity

9 Kaufman Assessment Battery for Children- Fluid-Crystalized index

10 Kaufman Assessment Battery for Children- mental processing index

11 Differential Ability Scales- General conceptual ability

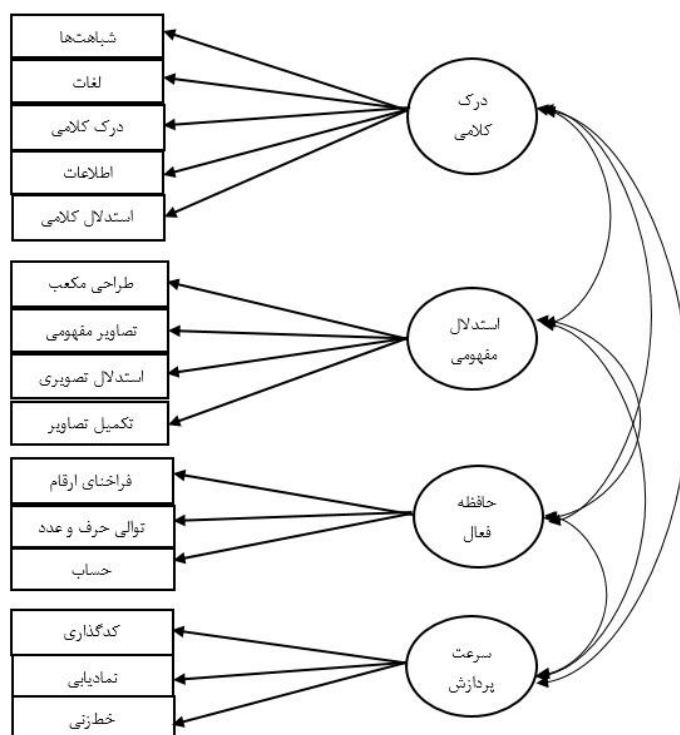
12 Wahlstrom

13 Reliability

پیشرونده ریون (۱۹۳۸) استفاده شد. نتایج نشان داد که بین ویرایش چهارم آزمون وکسلر و مقیاس تجدید نظر شده هوش وکسلر و آزمون ریون همبستگی معنادار وجود دارد. برای بررسی اعتبار آزمون از روش‌های بازآزمایی و دونیمه‌کردن استفاده شد. ضریب پایایی به‌دست آمده با استفاده از روش بازآزمایی برای نمره کلی هوش برابر ۰/۹۱ و برای شاخص‌ها از ۰/۸۸ تا ۰/۸۱ بود که مقادیرهای بالایی حساب می‌شوند. برای خرده مقیاس‌ها نیز بالاترین اعتبار مربوط به خرده مقیاس واژگان با ۰/۹۴ و پایین‌ترین آن مربوط به مفاهیم تصویری با ۰/۶۵ بود. در روش دو نیمه‌سازی نیز پایایی نمره کل هوش برابر ۰/۹۵ و برای شاخص‌ها از ۰/۹۱ تا ۰/۸۳ بود. در خرده مقیاس‌ها بالاترین پایایی برای واژگان و استدلال تصویری ۰/۸۶ و کمترین پایایی مربوط به فراختای ارقام با ۰/۷۱ می‌باشد.

### یافته‌های پژوهش

در این پژوهش برای بررسی نامتغیر بودن سؤالات از روش تحلیل عامل تأییدی و از نرم افزار Lisrel-8.7 استفاده شده است. فرایند تحلیل به این صورت بود که ابتدا مدل تحلیل عاملی تأییدی را که در تصویر شماره ۱ نشان داده شده است، به تفکیک در هر یک از دو گروه پسر و دختر اجرا شد. نتایج تحلیل عاملی تأییدی در دو گروه پسر و دختر نشان داد که مدل از برازش خوبی برخوردار است. سپس با استفاده از روش تحلیل عاملی تأییدی چندگروهی<sup>۱</sup> مدل همزمان در دو گروه اجرا شد. پارامترهای سؤال با استفاده از روش بیشینه درست‌نمایی برآورد شدند. همچنین تحلیل در سطح خرده مقیاس‌ها انجام شد. ابتدا نامتغیر بودن شکلی بررسی شد، در نامتغیر بودن شکلی همه پارامترهای مدل آزادانه برآورد می‌شوند (این مدل به عنوان مدل پایه شناخته می‌شود). در واقع در این روش شکل مدل تحلیل عاملی در دو گروه باید یکسان باشد. در مدل ضعیف، بارهای عاملی در دو گروه یکسان در نظر گرفته می‌شوند. در مدل قوی عرض از مبدأها در دو گروه یکسان در نظر گرفته می‌شود. در مدل دقیق، باقی مانده‌های متغیرهای نشانگر در دو گروه یکسان در نظر گرفته می‌شود.



تصویر ۱

مدل مفهومی آزمون هوش وکسلر کودکان ۴

جدول ۱ میانگین و انحراف استاندارد خرده مقیاس‌های آزمون هوش وکسلر چهار کودکان را بر حسب جنسیت نشان می‌دهد. نتایج به‌دست آمده بیانگر آن است که در خرده مقیاس‌های خطرتنی و کدگذاری تفاوت زیادی بین عملکرد گروه دختر و پسر در خرده مقیاس‌های آزمون هوش وکسلر چهار کودکان وجود ندارد.

<sup>1</sup> Multi-group confirmatory factor analysis

جدول ۱

میانگین و انحراف استاندارد خرده مقیاس‌های وکسلر بر اساس جنسیت				
جنسیت				
دختر		پسر		
انحراف استاندارد	میانگین	انحراف استاندارد	میانگین	
۱۰/۷۱	۱۹/۹۰	۱۰/۴۷	۱۹/۴۷	طراحی مکعب
۶/۳۵	۱۱/۸۴	۶/۳۵	۱۲/۶۲	شبهات‌ها
۳/۴۴	۱۲/۹۹	۳/۶۰	۱۳/۱۱	فراختای ارقام
۴	۱۲/۴۷	۳/۹۸	۱۲/۱۳	تصاویر مفهومی
۱۱/۵۹	۲۵/۸۸	۱۱/۰۷	۲۵/۲۷	لغات
۵/۵۱	۱۱/۰۹	۵/۷۱	۱۱/۴۲	توالی حرف-عدد
۵/۵۲	۱۴/۴۴	۵/۳۲	۱۴/۵۶	استدلال تصویری
۵/۸۵	۱۶/۵۴	۵/۹۱	۱۶/۷۴	درک کلامی
۶/۴۳	۱۵/۳۱	۵/۹۹	۱۵/۱۳	تکمیل تصاویر
۴/۵۵	۱۲/۱۹	۴/۴۴	۱۱/۹۷	اطلاعات
۵/۵۰	۱۷/۰۸	۵/۴۲	۱۶/۴۹	حساب
۳/۴۲	۹/۱۷	۳/۱۰	۹/۱۹	استدلال کلامی
۱۹/۳۳	۵۴/۱۱	۲۱/۵۵	۵۹/۶۰	خط زنی
۶/۷۸	۱۷/۶۹	۷/۳۱	۱۸/۳۳	نمادیابی
۱۰/۸۲	۳۴/۷۰	۱۱/۶۷	۳۷/۰۹	کدگذاری

جدول شماره ۲ نتایج مربوط به نامتغیر بودن اندازه‌گیری را نشان می‌دهد. شاخص‌های مختلفی برای بررسی نامتغیر بودن استفاده می‌شود که مهم‌ترین آنها در جدول ۲ آورده شده است. برای مقایسه مدل‌ها، می‌توان از شاخص‌های دو و شاخص برازش مقایسه‌ای (CFI) استفاده کرد. اما با توجه به اینکه تغییرات در  $\chi^2$  دو خیلی به اندازه نمونه وابسته است، شاخص‌های اصلی و مهم در بررسی نامتغیر بودن تغییرات در CFI و خطای میانگین مربعی ریشه‌ای تقریباً (RMSEA) هستند. با توجه به پیشنهاد چنگ و رنولد<sup>۳</sup> (۲۰۰۲) و چن<sup>۴</sup> (۲۰۰۷) و همچنین میلفونت و فیشر (۲۰۱۰)، مدل‌هایی در نهایت انتخاب می‌شوند که تغییرات در CFI و RMSEA آنها با توجه به مدل پیچیده‌تر بیشتر از ۰/۰۱ نباشد (لاوکس<sup>۵</sup> و همکاران، ۲۰۱۵). با توجه به این ملاک‌ها می‌توان گفت که آزمون هوش و کسلر نسخه چهار نسبت به جنسیت در بین دانش‌آموزان ایرانی نامتغیر است.

جدول ۲

نتایج تحلیل نامتغیر بودن آزمون هوش و کسلر کودکان					
مدل	مقدار $\chi^2$ دو	درجه آزادی	CFI	RMSEA	ECVI
مدل شکلی	۳۲۱/۶۷	۱۷۸	۰/۹۹	۰/۰۳۷ (۰/۰۳۰-۰/۰۴۳)	۰/۰۳۷ (۰/۰۳۳-۰/۰۴۲)
مدل ضعیف	۲۷۰	۱۷۰	۰/۹۹	۰/۰۳۱ (۰/۰۲۴-۰/۰۳۸)	۰/۰۲۹ (۰/۰۲۵-۰/۰۳۳)
مدل قوی	۳۲۶/۷۶	۱۹۴	۰/۹۹	۰/۰۳۴ (۰/۰۲۷-۰/۰۴)	۰/۰۳۴ (۰/۰۲۸-۰/۰۳۶)
مدل دقیق	۳۲۶/۷۶	۲۰۰	۰/۹۹	۰/۰۳۲ (۰/۰۲۶-۰/۰۳)	۰/۰۳۳ (۰/۰۲۷-۰/۰۳۵)

## بحث و نتیجه‌گیری

این مطالعه انجام شد تا نامتغیر بودن سازه‌های آزمون هوش و کسلر کودکان نسخه چهار را در بین دو گروه دانش‌آموزان پسر و دختر بررسی کنیم. اولین یافته این است که مدل چهار عاملی در دو گروه پسران و دختران به خوبی برازش پیدا کرد. همچنین نتایج نشان داد که نامتغیر بودن شکلی در دو

1 Comparative Fit Index

2 Root mean square error of approximation

3 Cheung & Rensvold

4 Chen

5 Laoux

گروه وجود دارد. نامتغیر بودن شکلی نشان می‌دهد که ساختار عاملی مشابهی در دو گروه پسر و دختر وجود دارد، بنابراین تعداد یکسانی متغیر مکنون، متغیرهای نشانگر و پارامترهای ثابت و آزاد شده در دو گروه وجود دارد. این تأیید می‌کند که آزمون هوش و کسلسر نسخه چهار سازه‌های یکسانی را در دو گروه اندازه‌گیری می‌کند. نامتغیر بودن ضعیف به این اشاره دارد که بارهای عاملی در دو گروه پسر و دختر برابر هستند، به این معنی که خرده مقیاس‌ها بر روی عامل‌های مربوط به خود با بارهای عاملی برابری در دو گروه بارگیری شده‌اند؛ در نتیجه نمره‌های مشاهده شده آزمون هوش و کسلسر نسخه چهار، عامل‌های مشابهی را در دو گروه اندازه‌گیری می‌کنند. نامتغیر بودن قوی به این اشاره دارد که عرض از مبدأها در دو گروه یکسان هستند، بنابراین وقتی مدل، نامتغیر بودن قوی دارد دل بر این است که میانگین متغیرهای مکنون در دو گروه یکسان است. در نامتغیر بودن دقیق باقی‌مانده‌های متغیرهای نشانگر در دو گروه یکسان در نظر گرفته می‌شوند، بنابراین وقتی مدل نامتغیر بودن دقیق دارد به این معنی است که متغیرهای مکنون در دو گروه دختر و پسر با دقت یکسانی اندازه‌گیری شده‌اند. نتایج به‌دست آمده در این پژوهش با پژوهش‌های چن و همکاران (۲۰۱۵)، دوالن و همکاران (۲۰۰۶) و بکر و همکاران (۲۰۲۴) همسو می‌باشد. این نتیجه همچنین همسو با نتایج به دست آمده از پژوهش والتر و همکاران (۲۰۲۱) است که نشان دادند آزمون هوش و کسلسر پیش دبستانی نسبت به جنسیت نامتغیر است و این آزمون می‌تواند توانایی‌های شناختی گروه دختران و پسران را به صورت یکسان اندازه‌گیری کند.

نامتغیر بودن اندازه‌گیری یکی از ویژگی‌های اساسی برای آزمون‌هایی است که به بررسی تفاوت‌های فردی در افراد می‌پردازد. در اندازه‌گیری روان‌شناختی، آزمونی که برای اندازه‌گیری یک ویژگی ساخته شده است، باید تفاوت‌های فردی افراد در آن ویژگی یا صفت را منعکس کند. اگر آزمون به این صورت عمل نکند، احتمالاً دارای ویژگی نامتغیر بودن نیست که معادل خطای سیستماتیک در اندازه‌گیری است. تحلیل‌های نامتغیر بودن اندازه‌گیری رابطه سؤال با سازه مکنون در افراد را نشان می‌دهد که وابسته به عضویت گروهی آنها نیست یا اینکه نسبت به گروه‌هایی که مقایسه می‌شوند، ثابت و نامتغیر است. عضویت گروهی افراد نباید رابطه بین سؤال و سازه مکنون را تحت تأثیر قرار دهد. در حوزه اندازه‌گیری هوش، بررسی نامتغیر بودن، یک شیوه معمول برای پاسخ به این پرسش است که آیا سؤال‌ها و خرده مقیاس‌هایی که توانایی‌های شناختی را اندازه‌گیری می‌کنند، در گروه‌ها و زیر مجموعه‌های گوناگون به صورت یکسان عمل می‌کنند یا خیر. در مورد نتایج مربوط به این پژوهش نتایج نامتغیر بودن نشان داد که نمره‌های عامل‌های مربوط به آزمون هوش و کسلسر کودکان نسخه چهار و خرده مقیاس‌های مربوط به آن معنای مشابهی برای دو گروه دانش‌آموزان پسر و دختر دارد. بنابراین می‌توان نتایج به دست آمده برای این دو گروه را به صورت یکسان تفسیر کرد و مقایسه‌های معنادار بین این دو گروه پسر و دختر انجام داد.

مهمترین محدودیت این پژوهش این بود که تنها به بررسی متغیر تغییرناپذیری جنسیت در اندازه‌گیری هوش پرداخته شد. به پژوهشگران بعدی پیشنهاد می‌شود که بررسی نامتغیر بودن اندازه‌گیری آزمون هوش و کسلسر در سایر گروه‌ها مانند سن یا گروه‌های بالینی دیگر مانند بیش‌فعالی پرداخته شود.

### سپاس‌گزاری

در این پژوهش صمیمانه از همکاری سازمان آموزش و پرورش استثنائی که داده‌های لازم را در اختیار پژوهشگر قرار دادند تشکر می‌شود.

### ملاحظات اخلاقی

در این پژوهش از داده‌های ثانویه به‌دست آمده از سازمان آموزش و پرورش استثنائی ایران استفاده شده بود و بنابراین مستقیماً با شرکت کنندگان ارتباط وجود نداشت. اما محرمانگی اطلاعات دریافت شده رعایت شد.

### حامی مالی

تمامی هزینه‌های مربوط به تحقیق و انتشار مقاله توسط نویسنده تقبل شده است و هیچ‌گونه حمایت مالی دریافت نشده است.

### مشارکت نویسندگان

همه فعالیت‌های مربوط به این پژوهش از ایده‌پردازی، گرفتن داده‌ها، تحلیل، نتیجه‌گیری و نگارش مقاله توسط نوشیروان محمدی انجام شده است.

### تعارض منافع

نویسنده تأیید می‌کند که هیچ‌گونه تعارض منافی در این پژوهش وجود ندارد.

### منابع

صادقی، ا.، ربیعی، م. و عابدی، م. ر. (۱۳۹۰). رواسازی و اعتباریابی چهارمین ویرایش مقیاس هوش و کسلسر. *مجله روانشناسی تحولی*، ۷ (۲۸)، ۳۸۶-۳۷۷.

### References

Abubakar-Abdullateef, A., Kusi-Mensah, K., Sun, L., Mohammed, A., & Sheikh, T. L. (2025). Assessing factor structure and measurement invariance of WISC-IV among almajiris attending

- quranic schools and pupils attending mainstream public primary schools in Northern Nigeria. *Applied Neuropsychology: Child*, 1-13.
- Becker, Angelika Beate Christiane, Jenny Maurer, Monika Daseking, and Franz Pauls. (2024). Measurement Invariance of the WISC-V across a Clinical Sample of Children and Adolescents with ADHD and a Matched Control Group. *Journal of Intelligence* 12, no. 1: 6.
- Blankson, A.N., & McArdle, J.J. (2013). Measurement invariance of cognitive abilities across ethnicity, gender, and time among older americans. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 70(3), 386-397.
- Celik C, Yigit İ, Guzey Yigit M, Erden G. (2020). Examining the factor structure of the WISC-IV in clinical and non-clinical samples: a multiple-group confirmatory factor analysis. Dusunen Adam. *The Journal of Psychiatry and Neurological Sciences*; 33:296-309.
- Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, 52, 161-166.
- Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children-Fifth Edition. *Personality and Individual Differences*, 86, 1-5.
- Dolan, c. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, 35, 21-50.
- Hortensius, L. M.(2012)."Project for Introduction to Multivariate Statistics: Measurement Invariance," [online] <http://www.tc.umn.edu/~horte005/docs/MeasurementInvariance.pdf>
- Jak, S (2014). Testing strong factorial invariance using three-level structural equation modeling. *Front. Psychol.* 5:745.
- Laloux, Y. R., Viladrich, C., Sousa, C., & Jannes, C. (2015). Exploring the factorial structure of the Sport Anxiety Scale-2: Invariance across language, gender, age and type of sport. *Psicothema*, 27(2), 174-181.
- Milfont, T.L., Fischer, R., (2010). Testing measurement invariance across groups: application in cross-cultural research. *International journal of psychological research*, 3(1), 111-121.
- Niileksela, c. r., Reynolds, M. r., & Kaufman, A. S. (2013). An alternative Cattell-Horn-Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70-90. *Psychological Assessment*, 25, 391-404.
- Panah, N. M. R., Kamkari, K., & Baghdasarians, A. (2025). Diagnostic Validity of the Fifth Edition of Wechsler Intelligence Scale for Children (WISC-V) in Students with ADHD. *Iranian Journal of Psychiatry & Behavioral Sciences/Progress in Psychiatry & Behavioral Sciences*, 19(1).
- Reverte, I., Golay, P., Favez, N., Rossier, J., & Lecerf, T. (2015). Testing for multigroup invariance of the WISC-IV structure across France and Switzerland: Standard and CHC models. *Learning and individual differences*, 40, 127-133.
- Reynolds, M. r., Ingram, P. B., Seeley, J. S., & Newby, K. D. (2013). Investigating the structure and invariance of the Wechsler adult intelligence scales, fourth edition in a sample of adults with intellectual disabilities. *Research in Developmental Disabilities*, 34, 3235-3245.
- Richerson, L. P., Watkins, M. W., & Beaujean, A. A. (2014). Longitudinal invariance of the Wechsler Intelligence Scale for Children-Fourth Edition in a referral sample. *Journal of Psychoeducational Assessment*, 32(7), 597-609.
- Rodríguez-Cancino, M., & Concha-Salgado, A. (2023). WISC-V measurement invariance according to sex and age: Advancing the understanding of intergroup differences in cognitive performance. *Journal of Intelligence*, 11(180).

- Sadeghi, A., Rabiei, M., & Abedi, M. R. (2011). Validation and reliability assessment of the fourth edition of the Wechsler Intelligence Scale for Children. *Developmental Psychology (Iranian Psychologists)*, 7(28), 377-386. [Persian]
- Tommasi, M., Pezzuti, I., Colom, R., Abad, F. J., Saggino, A., & Orsini, A. (2015). Increased educational level is related with higher IQ scores but lower g-variance: Evidence from the standardization of the WAIS-r for Italy. *Intelligence*, 50, 68-74.
- Wahlstrom, D., Breaux, K. C., Zhu, J., & Weiss, L. G. (2012). The Wechsler preschool and primary scale of intelligence—third edition, the Wechsler intelligence scale for children—fourth edition, and the Wechsler individual achievement test—. *Contemporary intellectual assessment: Theories, tests, and issues*, 224-248. Guilford Press.
- Walter, F., Daseking, M., & Pauls, F. (2021). Sex Differences in Intelligence in Children Aged 2:6-7:7: Analysis of the Factor Structure and Measurement Invariance of the German Wechsler Primary and Preschool Scale of Intelligence—Fourth Edition. *Journal of Psychoeducational Assessment*, 39(4), 395-421.
- Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, 30(7), 1006-1016.
- Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. (2010). Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn Effect. *Learning and Individual Differences*, 20(3), 135-151.