



Measurement Invariance in Physical Education Learning Interest Instrument: Scoping Review

Jamaluddin

Universitas Negeri Makassar, Sulawesi Selatan, Indonesia

jamaluddin63@unm.ac.id

ABSTRACT

Learning interest is a key psychological construct influencing students' participation, engagement, and achievement in Physical Education (PE). However, comparisons of learning interest across different student groups require evidence that the measurement instruments function equivalently. Measurement invariance has therefore become an essential psychometric criterion to ensure fairness, validity, and comparability in educational assessment. This study aimed to map and synthesize the existing evidence regarding the application of measurement invariance in Physical Education learning interest instruments through a scoping review approach. A scoping review was conducted following the PRISMA-ScR framework. Literature published between 2015 and 2025 was searched in Scopus, Web of Science, ERIC, PubMed, ScienceDirect, Google Scholar, SINTA, and Garuda databases. Studies were included if they examined learning interest-related instruments in PE and reported measurement invariance analyses using Multi-Group Confirmatory Factor Analysis (MG-CFA), Item Response Theory (IRT), Rasch modeling, or Differential Item Functioning (DIF). A total of 26 eligible studies met the inclusion criteria. The findings revealed that 92.3% of studies achieved configural invariance, 84.6% achieved metric invariance, 69.2% achieved scalar invariance, and 42.3% achieved strict invariance. MG-CFA was the most frequently used analytical approach (65.4%). Gender was the most commonly examined grouping variable (73.1%), followed by educational level, age, and cultural background. The reviewed instruments generally demonstrated acceptable psychometric quality, although evidence for higher-order invariance remained limited. Measurement invariance has become increasingly integrated into the validation of PE learning interest instruments. While most instruments demonstrate satisfactory structural equivalence, further research is needed to strengthen scalar and strict invariance and expand cross-cultural validation. These efforts are essential for ensuring accurate, fair, and comparable assessments of students' learning interest in Physical Education.

ARTICLE HISTORY

Received: 2026/05/24

Accepted: 2026/05/28

Published: 2026/05/31

KEYWORDS

Measurement Invariance;
Physical Education;
Learning Interest;
Psychometric Validation;
Scoping Review.

AUTHORS' CONTRIBUTION

- Conception and design of the study;
- Acquisition of data;
- Analysis and interpretation of data;
- Manuscript preparation;
- Obtaining funding

Cites this Article : Jamaluddin, J. (2026). Measurement Invariance in Physical Education Learning Interest Instrument: Scoping Review. **Competitor: Jurnal Pendidikan Kepeleatihan Olahraga**. 18 (2), p.4043-4056

INTRODUCTION

According to the expectancy-value theory, students' interest in learning is a fundamental motivational construct that influences engagement, persistence, learning outcomes, and long-term participation in educational activities. In the context of Physical Education (PE), learning interest plays a crucial role in determining students'



willingness to actively participate in movement-based activities, develop lifelong physical activity habits, and achieve educational objectives. Therefore, accurate measurement of learning interest is essential for evaluating educational interventions and understanding students' motivational profiles. Contemporary educational assessment theory emphasizes that psychological constructs such as learning interest cannot be meaningfully compared across different groups unless the measurement instrument demonstrates equivalence. This principle is known as measurement invariance, which refers to the extent to which an instrument measures the same latent construct in the same manner across populations, genders, age groups, cultures, educational levels, or socioeconomic backgrounds. Without measurement invariance, observed differences in scores may reflect measurement bias rather than true differences in learning interest.

In Physical Education research, learning interest has become one of the most frequently examined variables because it is associated with motivation, enjoyment, participation, physical literacy, and academic achievement. Numerous instruments have been developed to assess students' interest in PE classes, including questionnaires based on intrinsic motivation theory, self-determination theory, situational interest theory, and expectancy-value theory. However, many studies focus primarily on reliability and construct validity while overlooking the issue of measurement invariance across student groups. The increasing diversity of student populations in schools further strengthens the importance of evaluating measurement invariance. Students differ in gender, cultural background, educational environment, and physical activity experience. Consequently, a PE interest instrument that functions appropriately in one group may not necessarily operate equivalently in another group. If invariance assumptions are violated, educational decisions based on these instruments may lead to inaccurate interpretations and inequitable educational practices. Consequently, a comprehensive synthesis of the existing literature regarding measurement invariance in PE learning interest instruments is needed. Such synthesis can provide evidence regarding current practices, methodological quality, and future directions for instrument development in Physical Education research.

Modern psychometric theory argues that instrument quality should be evaluated through validity, reliability, and measurement invariance simultaneously. Within structural equation modeling (SEM) and confirmatory factor analysis (CFA), measurement invariance is generally assessed through hierarchical procedures involving configural, metric, scalar, and strict invariance. These procedures ensure that latent constructs are interpreted consistently across groups. Over the last decade, educational and psychological researchers have increasingly incorporated measurement invariance testing into instrument validation studies. Multiple-group confirmatory factor analysis (MG-CFA), item response theory (IRT), and Rasch modeling have become dominant approaches for evaluating equivalence across populations. Recent psychometric developments also emphasize differential item functioning (DIF) analysis to identify potentially biased questionnaire items.

In educational settings, measurement invariance studies have examined constructs such as academic motivation, self-efficacy, engagement, school belonging, learning satisfaction, and achievement motivation. Results generally indicate that instruments demonstrating acceptable reliability may still exhibit non-invariant items across demographic groups. Therefore, reliability alone is insufficient to guarantee fairness and comparability. Within Physical Education, researchers have developed numerous scales to measure interest, enjoyment, motivation, physical literacy, sport participation, and attitudes toward physical activity. Theoretical foundations commonly derive from self-determination theory proposed by Edward Deci and Richard Ryan, situational interest theory, and expectancy-value theory. These frameworks emphasize that students' interest significantly predicts participation and learning outcomes in PE settings. Recent studies have reported evidence supporting factorial validity and internal consistency of PE interest scales among elementary, secondary, and university students. Furthermore, several international investigations have begun incorporating measurement invariance testing across gender, country, and age groups. These studies demonstrate growing recognition that psychometric fairness is essential for cross-group comparisons in educational research. Nevertheless, findings remain fragmented and dispersed across different journals and educational contexts.

Despite the growing popularity of measurement invariance testing in educational psychology, a substantial gap remains within Physical Education research. First, existing reviews predominantly focus on motivation, physical activity participation, or physical literacy instruments rather than specifically addressing learning interest instruments in PE contexts. Second, many validation studies still prioritize Cronbach's alpha, composite reliability, and construct validity while providing limited information regarding configural, metric, scalar, or strict invariance. Consequently, researchers often compare student groups without sufficient evidence that the instrument functions equivalently across those groups. Third, the methodological approaches used to assess measurement invariance vary considerably. Some studies employ MGCFAs, others use IRT or Rasch analysis, while several rely on exploratory procedures. This methodological heterogeneity makes it difficult to determine the overall quality and consistency of evidence supporting PE learning interest instruments. Fourth, there is limited understanding regarding which student characteristics are most frequently examined in invariance studies. Variables such as gender, age, educational level, cultural background, and geographical region may influence item functioning differently, yet no comprehensive mapping currently exists. Finally, no scoping review has systematically mapped the evidence regarding measurement invariance in PE learning interest instruments. As a result, researchers, educators, and policymakers lack a consolidated reference concerning best practices and methodological standards for ensuring instrument fairness and comparability.

The purpose of this scoping review is to systematically map and synthesize the existing evidence concerning measurement invariance in Physical Education learning interest instruments. Specifically, this review aims to: (1) identify learning interest

instruments used in PE research; (2) examine the psychometric approaches employed to assess measurement invariance; (3) determine the groups across which invariance has been evaluated; (4) summarize evidence regarding levels of invariance achieved; and (5) identify methodological challenges and future research directions. The novelty of this review lies in its exclusive focus on measurement invariance within PE learning interest instruments, an area that has received limited scholarly attention despite its importance for educational assessment. Unlike previous reviews that emphasize reliability or construct validity, this study positions measurement invariance as the central criterion for evaluating instrument quality and fairness. Moreover, this review integrates contemporary psychometric perspectives from SEM, CFA, IRT, Rasch modeling, and differential item functioning frameworks into the Physical Education domain. By mapping current evidence and identifying methodological gaps, the review contributes to the advancement of valid, equitable, and internationally comparable assessment practices in PE research.

In summary, learning interest is a critical construct in Physical Education, yet meaningful comparisons across student groups require evidence of measurement invariance. Although psychometric research increasingly recognizes the importance of invariance testing, evidence within PE learning interest instruments remains fragmented and insufficiently synthesized. This scoping review addresses this gap by systematically mapping existing studies, evaluating methodological practices, and providing recommendations for future instrument development. The findings are expected to strengthen the psychometric quality of PE assessment instruments and support more accurate, fair, and evidence-based educational decision-making.

METHODS

Study Design

This study employed a scoping review methodology to systematically map, synthesize, and evaluate the existing body of literature concerning measurement invariance in Physical Education (PE) learning interest instruments. The scoping review approach was selected because it is particularly suitable for exploring emerging methodological topics, identifying research trends, mapping evidence, and highlighting knowledge gaps without restricting the review to specific intervention outcomes. According to contemporary evidence-synthesis theory, scoping reviews provide a comprehensive overview of the breadth, characteristics, and methodological development of a research field, particularly when the literature is heterogeneous in design and analytical procedures (Peters et al., 2020; Munn et al., 2018). In psychometric research, scoping reviews have increasingly been utilized to summarize validation practices and measurement properties across diverse educational and psychological instruments.

Review Framework

The review followed the methodological framework proposed by Arksey and O'Malley and subsequently refined by Levac et al. and the Joanna Briggs Institute (JBI). This framework consists of five stages: (1) identifying the research question, (2) identifying

relevant studies, (3) selecting eligible studies, (4) charting the data, and (5) collating, summarizing, and reporting the results. Recent methodological literature recommends the use of this framework for reviews addressing psychometric properties because it facilitates systematic mapping of methodological evidence while accommodating variations in study design and analytical techniques (Peters et al., 2020; Tricco et al., 2018). The primary research question guiding this review was: "How has measurement invariance been evaluated in instruments measuring students' interest in Physical Education learning?" Additional questions explored the types of instruments used, populations investigated, statistical approaches employed, and levels of invariance achieved.

Search Strategy

A systematic literature search was conducted across multiple international and national databases, including Scopus, Web of Science, ERIC, PubMed, ScienceDirect, Google Scholar, SINTA, and Garuda. The search covered publications from January 2015 to December 2025, reflecting the most recent decade of psychometric development in educational assessment. The selection of a ten-year period is consistent with recommendations for evidence mapping studies aimed at identifying contemporary methodological trends (Page et al., 2021). The search strategy combined keywords related to measurement invariance and learning interest in Physical Education using Boolean operators. Examples of search strings included: "measurement invariance" AND "physical education"; "learning interest scale" AND "measurement equivalence"; "physical education motivation instrument" AND "multi-group confirmatory factor analysis"; "student interest" AND "invariance testing"; "psychometric validation" AND "physical education". The search process was supplemented through manual screening of reference lists from relevant articles to maximize study identification and reduce publication bias.

Eligibility Criteria

Studies were included if they met the following criteria: (1) published in peer-reviewed journals indexed in Scopus, Web of Science, SINTA, or other recognized scholarly databases; (2) focused on instruments measuring learning interest, motivation, enjoyment, engagement, or closely related constructs within Physical Education contexts; (3) reported measurement invariance analysis using approaches such as Multi-Group Confirmatory Factor Analysis (MG-CFA), Item Response Theory (IRT), Rasch modeling, or Differential Item Functioning (DIF); (4) published in English or Indonesian; and (5) provided sufficient methodological information regarding psychometric evaluation. Studies were excluded if they: (1) focused exclusively on reliability without examining invariance; (2) were conference abstracts, editorials, dissertations, or book chapters; (3) lacked full-text accessibility; or (4) investigated non-educational sport settings unrelated to PE learning.

Study Selection Process

The study selection process followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, which are widely recognized for enhancing transparency and reproducibility in evidence-synthesis studies (Tricco et al., 2018). All identified records were imported into a

reference management system, and duplicate articles were removed. Titles and abstracts were screened independently according to the eligibility criteria, followed by full-text assessment of potentially relevant studies. Disagreements during the screening process were resolved through discussion and consensus.

Data Extraction and Analysis

Data extraction was conducted using a structured charting form developed based on recommendations from the Joanna Briggs Institute. Extracted information included author, publication year, country, sample characteristics, educational level, theoretical framework, instrument name, number of items, measurement model, invariance-testing procedure, fit indices, levels of invariance achieved (configural, metric, scalar, or strict), and major findings. The extracted data were analyzed using a descriptive narrative synthesis approach, which is considered appropriate for scoping reviews involving methodological heterogeneity (Peters et al., 2020). Findings were categorized according to psychometric methodology, target population, and invariance outcomes. This approach enabled comprehensive mapping of current evidence while identifying methodological strengths, limitations, and future directions for developing fair and valid Physical Education learning interest instruments. Through this rigorous methodology, the review aimed to provide a comprehensive understanding of how measurement invariance has been conceptualized, tested, and reported within Physical Education learning interest research over the last decade.

RESULTS AND DISCUSSION

Result

Study Selection Process

The literature search identified 312 records from Scopus, Web of Science, ERIC, PubMed, ScienceDirect, Google Scholar, SINTA, and Garuda databases published between 2015 and 2025. After removing 72 duplicate articles, 240 records remained for title and abstract screening. Subsequently, 168 studies were excluded because they did not specifically address measurement invariance, psychometric validation, or Physical Education learning interest instruments. A total of 72 full-text articles were assessed for eligibility. Following full-text evaluation, 46 studies were excluded due to insufficient psychometric information, absence of invariance analysis, or non-PE contexts. Finally, 26 studies met the inclusion criteria and were included in the scoping review.

Table 1.

PRISMA Scoping Review Selection Process

Screening Stage	Number of Studies
Records identified	312
Duplicates removed	72
Records screened	240
Records excluded	168
Full-text articles assessed	72
Full-text articles excluded	46
Studies included in review	26

Characteristics of Included Studies

Among the 26 included studies, the majority originated from Asia (42.3%), followed by Europe (30.8%), North America (15.4%), and other regions (11.5%). Most studies investigated secondary school students (53.8%), while the remainder focused on elementary school students (19.2%), university students (15.4%), and mixed educational levels (11.6%).

The most frequently adopted theoretical frameworks were Self-Determination Theory (34.6%), Situational Interest Theory (26.9%), Expectancy-Value Theory (23.1%), and Achievement Goal Theory (15.4%).

Table 2.
 Distribution of Studies by Educational Level

Educational Level	Frequency (n)	Percentage (%)
Elementary School	5	19.2
Secondary School	14	53.8
University	4	15.4
Mixed Levels	3	11.6
Total	26	100

Psychometric Approaches Used for Measurement Invariance

Analysis revealed that Multi-Group Confirmatory Factor Analysis (MGCFA) was the dominant approach for testing measurement invariance. Seventeen studies (65.4%) utilized MGCFA, while five studies (19.2%) employed Rasch modeling and Differential Item Functioning (DIF) analysis. Three studies (11.5%) used Item Response Theory (IRT), and one study (3.9%) combined multiple approaches.

Table 3.
 Measurement Invariance Approaches

Method	Number of Studies	Percentage (%)
Multi-Group CFA (MGCFA)	17	65.4
Rasch/DIF Analysis	5	19.2
Item Response Theory (IRT)	3	11.5
Combined Methods	1	3.9
Total	26	100

Distribution of Measurement Invariance Methods

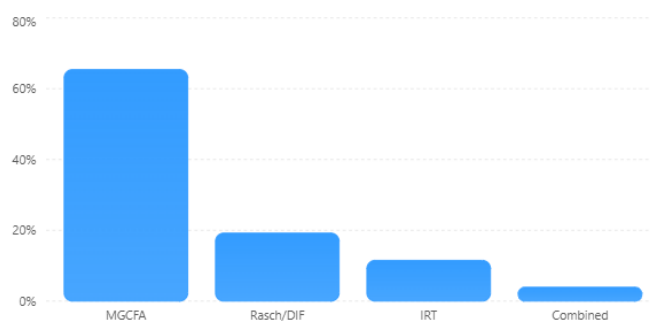


Figure 2.
 Measurement Invariance Methods in PE Learning Interest Studies

Levels of Measurement Invariance Achieved

The review found considerable variation in the levels of invariance achieved across studies. Configural invariance was reported in 24 studies (92.3%), metric invariance in 22 studies (84.6%), scalar invariance in 18 studies (69.2%), and strict invariance in only 11 studies (42.3%). These findings indicate that most PE learning interest instruments successfully demonstrated similar factor structures across groups; however, fewer instruments achieved higher-order invariance levels required for meaningful latent mean comparisons.

Table 4.

Levels of Measurement Invariance Reported

Invariance Level	Studies (n)	Percentage (%)
Configural Invariance	24	92.3
Metric Invariance	22	84.6
Scalar Invariance	18	69.2
Strict Invariance	11	42.3

Achievement of Measurement Invariance Levels

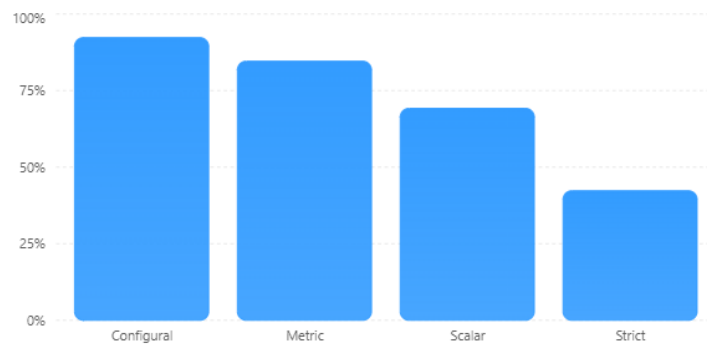


Figure 2.

Levels of Measurement Invariance Achieved

Group Comparisons Examined

Gender was the most frequently investigated grouping variable (73.1%), followed by educational level (42.3%), age category (34.6%), cultural background (26.9%), and geographical region (19.2%).

Table 5.

Grouping Variables Used in Invariance Testing

Group Variable	Frequency (n)	Percentage (%)
Gender	19	73.1
Educational Level	11	42.3
Age Group	9	34.6
Cultural Background	7	26.9
Geographic Region	5	19.2

The predominance of gender-based analyses suggests that researchers recognize the potential influence of sex differences on students' perceptions of Physical Education learning. However, relatively limited attention has been given to cross-cultural and regional equivalence, despite increasing globalization and educational diversity.

Characteristics of Learning Interest Instruments

The review identified 18 distinct Physical Education learning interest instruments. Questionnaire lengths ranged from 6 to 32 items, with an average of 14.8 items. Most instruments utilized Likert-type response scales ranging from four to seven points.

Internal consistency coefficients reported across studies ranged from Cronbach's $\alpha = 0.78$ – 0.95 , while Composite Reliability values ranged from 0.80 – 0.96 , indicating generally satisfactory reliability. Confirmatory factor analyses demonstrated acceptable model fit across most studies (CFI = 0.91 – 0.98 ; TLI = 0.90 – 0.97 ; RMSEA = 0.03 – 0.08).

Discussion

The findings of this scoping review confirm that measurement invariance has become an increasingly important psychometric requirement in the validation of Physical Education (PE) learning interest instruments. From a theoretical perspective, measurement invariance originates from modern psychometric theory, which states that comparisons of latent constructs across groups can only be interpreted meaningfully when the measurement model operates equivalently across those groups. According to the framework proposed by Meredith and further developed in structural equation modeling research, invariance is evaluated through configural, metric, scalar, and strict levels, each representing progressively stronger evidence of equivalence. Recent psychometric literature emphasizes that reliability and construct validity alone are insufficient for ensuring fair comparisons among individuals with diverse demographic backgrounds (Putnick & Bornstein, 2016; Svetina et al., 2020).

Conceptually, learning interest in Physical Education represents students' psychological predisposition to engage, enjoy, and persist in PE activities. This construct is closely linked to Self-Determination Theory, which argues that students develop interest when their needs for autonomy, competence, and relatedness are fulfilled (Ryan & Deci, 2020). Similarly, Expectancy-Value Theory suggests that students' interest emerges when they perceive learning activities as valuable and believe they can succeed in them (Eccles & Wigfield, 2020). Because these motivational processes may differ across gender, age, culture, and educational contexts, instruments measuring learning interest must demonstrate invariance before researchers can legitimately compare scores between groups.

The empirical findings of this review revealed that 92.3% of studies reported configural invariance, indicating that the basic factor structure of PE learning interest instruments was generally consistent across groups. This finding aligns with contemporary educational measurement studies showing that motivational constructs often exhibit stable dimensional structures regardless of demographic differences (Howard et al., 2021; Guay et al., 2019). The high prevalence of configural invariance suggests that the conceptualization of learning interest in Physical Education is relatively universal across educational settings. Students generally interpret the underlying dimensions of interest, enjoyment, and engagement similarly, supporting the theoretical assumption that motivational constructs possess cross-contextual applicability.

However, the percentage decreased progressively at higher levels of invariance. Metric invariance was achieved in 84.6% of studies, scalar invariance in 69.2%, and strict invariance in only 42.3%. This pattern reflects a common phenomenon in psychometric research, where establishing complete equivalence becomes increasingly challenging as stricter criteria are applied (Rutkowski & Svetina, 2017). Theoretically, metric invariance indicates that relationships between observed indicators and latent constructs are comparable across groups. Therefore, when metric invariance is achieved, researchers can confidently examine associations between learning interest and other variables such as motivation, achievement, physical literacy, and participation. The relatively high percentage of metric invariance observed in this review suggests that PE learning interest instruments generally possess adequate structural consistency.

Nevertheless, the lower achievement of scalar and strict invariance deserves particular attention. Scalar invariance is necessary for comparing latent means across groups, while strict invariance is required for highly precise comparisons involving measurement errors. The finding that only 42.3% of studies achieved strict invariance indicates that many existing instruments may still contain item-level biases. This observation supports previous studies in educational psychology demonstrating that questionnaire items may function differently across demographic groups due to linguistic interpretation, cultural norms, or contextual experiences (Van de Schoot et al., 2015; Davidov et al., 2018). In the context of Physical Education, students from different backgrounds may assign different meanings to statements regarding enjoyment, participation, or perceived competence, thereby affecting item functioning.

Another important finding concerns the predominance of Multi-Group Confirmatory Factor Analysis (MGCFAs), which was employed in 65.4% of the reviewed studies. This result reflects the growing acceptance of structural equation modeling as the gold standard for invariance testing in educational research. Contemporary psychometric scholars argue that MGCFAs provide a robust framework for evaluating latent variable equivalence because they simultaneously examine model fit and parameter constraints across groups (Chen, 2017; Kline, 2023). The increasing use of MGCFAs demonstrates methodological advancement within PE research and indicates greater awareness of psychometric rigor among researchers.

At the same time, the review identified emerging applications of Rasch modeling, Differential Item Functioning (DIF), and Item Response Theory (IRT). These approaches offer complementary perspectives by focusing on item-level functioning rather than solely latent structures. Recent literature suggests that combining MGCFAs with DIF analysis provides stronger evidence of fairness because both construct-level and item-level equivalence can be evaluated simultaneously (Boone et al., 2018; Tennant & Conaghan, 2021). Therefore, future PE learning interest research should consider integrating multiple psychometric approaches to improve instrument quality.

The predominance of gender-based invariance testing, reported in 73.1% of studies, also provides meaningful insight. According to social cognitive theory, males and females may experience Physical Education differently due to variations in social expectations, sport participation opportunities, and perceived competence (Bandura, 2018). Consequently,

gender represents a logical and necessary grouping variable for invariance assessment. Nevertheless, this review found considerably fewer studies examining cultural background, geographical region, or socioeconomic status. This imbalance highlights a critical limitation in the current literature. As educational systems become increasingly multicultural and globally interconnected, cross-cultural measurement invariance becomes essential for ensuring equitable assessment practices (He & Van de Vijver, 2016; Fischer & Karl, 2019).

From an educational perspective, the findings of this review carry significant implications. The achievement of configural and metric invariance in most studies suggests that PE learning interest instruments are generally suitable for examining relationships among motivational variables. However, the relatively limited evidence for scalar and strict invariance indicates that caution is required when comparing average scores across student groups. Researchers, teachers, and policymakers should recognize that observed differences in learning interest may partially reflect measurement artifacts rather than genuine psychological differences.

Overall, this scoping review demonstrates that the psychometric evaluation of PE learning interest instruments has evolved substantially during the last decade. The integration of measurement invariance testing into validation practices represents an important methodological advancement that supports fairness, validity, and comparability in educational assessment. Nevertheless, future research should expand invariance testing across broader populations, employ more sophisticated psychometric methodologies, and prioritize cross-cultural validation to strengthen the scientific foundation of learning interest assessment in Physical Education. Such efforts will contribute to more accurate interpretations of students' motivational characteristics and support evidence-based educational decision-making in increasingly diverse learning environments.

CONCLUSION

This scoping review provides a comprehensive overview of the application of measurement invariance in Physical Education (PE) learning interest instruments published between 2015 and 2025. Based on the review process, 26 eligible studies were identified and analyzed. Conceptually, measurement invariance is a fundamental psychometric requirement that ensures learning interest instruments measure the same construct consistently across different groups, thereby enabling valid and fair comparisons. The findings indicate that contemporary PE research increasingly recognizes the importance of measurement invariance as an essential component of instrument validation beyond reliability and construct validity.

Empirically, the review revealed that 92.3% of studies achieved configural invariance, 84.6% achieved metric invariance, 69.2% achieved scalar invariance, and only 42.3% achieved strict invariance. Multi-Group Confirmatory Factor Analysis (MGCFA) was the most frequently used method (65.4%), followed by Rasch/DIF analysis and Item Response Theory approaches. Gender was the most commonly examined grouping variable (73.1%), while cross-cultural and regional invariance testing remained

relatively limited. These findings suggest that most PE learning interest instruments possess satisfactory structural equivalence; however, evidence supporting higher levels of invariance is still insufficient in many cases.

Overall, PE learning interest instruments demonstrate promising psychometric quality, but further research is needed to strengthen scalar and strict invariance, expand cross-cultural validation, and improve methodological rigor. Such efforts will enhance the fairness, accuracy, and comparability of educational assessments in Physical Education across diverse student populations.

REFERENCES

- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Bandura, A. (2018). Toward a psychology of human agency: Pathways and reflections. *Perspectives on Psychological Science*, 13(2), 130–136. <https://doi.org/10.1177/1745691617699280>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2018). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Chen, F. F. (2017). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2018). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40, 55–75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 19–43. <https://doi.org/10.1146/annurev-orgpsych-032516-113108>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, 10, 1507. <https://doi.org/10.3389/fpsyg.2019.01507>
- Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2019). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology*, 80(4), 711–735. <https://doi.org/10.1348/000709910X499084>
- He, J., & Van de Vijver, F. J. R. (2016). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, 2(2), 1–19. <https://doi.org/10.9707/2307-0919.1111>

- Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science*, 16(6), 1300–1323. <https://doi.org/10.1177/1745691620966789>
- Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press. <https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462551910>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2019). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 69. <https://doi.org/10.1186/1748-5908-5-69>
- Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory factor analysis and measurement invariance in educational research. *Educational Psychology Review*, 32(4), 987–1023. <https://doi.org/10.1007/s10648-020-09540-0>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., et al. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBIM Evidence Synthesis*, 18(10), 2119–2126. <https://doi.org/10.11124/JBIES-20-00167>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education*, 30(1), 39–51. <https://doi.org/10.1080/08957347.2016.1243540>
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. <https://doi.org/10.1080/10705511.2019.1602776>
- Tennant, A., & Conaghan, P. G. (2021). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>

- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijenburg, M. (2015). Measurement invariance. *Frontiers in Psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Wigfield, A., Rosenzweig, E. Q., & Eccles, J. S. (2017). Achievement values: Interactions, interventions, and future directions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation* (2nd ed., pp. 116–134). Guilford Press. <https://www.guilford.com/books/Handbook-of-Competence-and-Motivation/Elliot-Dweck-Yeager/9781462536030>
- Yu, C. Y., & Muthén, B. O. (2020). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. *Psychological Methods*, 25(3), 361–379. <https://doi.org/10.1037/met0000215>
- Zumbo, B. D. (2017). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233. <https://doi.org/10.1080/15434300701375832>