African Journal of Social and Behavioural Sciences (AJSBS)
Volume 16, Number 5 (2026) ISSN: 2141-209X

RETHINKING MEASUREMENT IN EDUCATION: A CRITICAL EXAMINATION OF PROPERTIES IN EDUCATIONAL SCALES

Kingsley Oku

Department of Educational Foundations, University of Delta, Agbor, Delta State, Nigeria

kingsley.oku@unidel.edu.ng

ABSTRACT: Educational measurement is essential to teaching, learning, assessment, curriculum evaluation, and educational decision-making. As educational systems increasingly rely on quantitative evidence to guide policy and practice, concerns have emerged regarding the quality and psychometric adequacy of educational scales used to measure latent constructs such as achievement, depression, locus of control, motivation, self-efficacy, engagement, attitude, and learning outcomes. Traditional approaches to measurement often emphasize score generation while overlooking critical scale properties that determine the trustworthiness and interpretability of results. This position paper advocates a rethinking of educational measurement with a stronger emphasis on validity, reliability, factor structure, measurement invariance, fairness, and technological responsiveness. The paper argues that educational scales should be evaluated not merely by their ability to produce scores but by the extent to which those scores accurately, consistently, and equitably represent the intended constructs. Implications for educational research, assessment practices, and curriculum evaluation are discussed.

Keywords: Educational Measurement, Educational Scales, Validity, Reliability, Factor Structure, Measurement Invariance, Psychometrics

INTRODUCTION

Education has increasingly become data-driven, making measurement one of the most essential components of teaching, learning, and educational decision-making. Educational measurement provides the scientific basis for collecting, analyzing, and interpreting evidence about learners' knowledge, skills, attitudes, abilities, and other educational outcomes. Through standardized tests, classroom assessments, rating scales, questionnaires, and observational instruments, educational measurement informs curriculum development, instructional improvement, student placement, certification, and educational policy. As educational systems continue to evolve in response to technological advancement, competency-based education, and digital learning environments, there is an increasing need to critically examine whether the scales commonly used in educational research possess the necessary measurement properties to support valid and reliable interpretations of scores (Yu et al., 2026).

Traditionally, educational researchers have relied heavily on measurement scales such as Likert-type scales, semantic differential scales, and rating scales to quantify abstract constructs including academic motivation, depression, locus of control, self-efficacy, teacher effectiveness, student engagement, school climate, and learning satisfaction. These scales are generally assumed to

represent interval-level measurements, thereby justifying the application of sophisticated statistical analyses such as regression, factor analysis, multilevel modeling, and structural equation modeling (SEM). However, contemporary psychometric research increasingly questions these assumptions, arguing that many educational scales merely generate ordinal data rather than true interval measurements. Consequently, treating ordinal responses as interval data may introduce bias, distort parameter estimates, and weaken the validity of research conclusions (Feuerstahler, 2023).

The debate surrounding the properties of educational scales has built up over the past decade due to renewed interest in modern measurement theories, particularly Item Response Theory (IRT), Rasch Measurement Theory (RMT), and Evidence-Centered Design (ECD). Unlike Classical Test Theory (CTT), which primarily focuses on observed scores and reliability estimates, modern psychometric approaches emphasize construct validity, invariance, item functioning, measurement precision, and the transformation of ordinal observations into interval measures where appropriate. These developments have shifted attention from simply developing reliable instruments toward ensuring that measurement instruments genuinely reflect the underlying constructs they are intended to measure (Cook & Pitoniak, 2026; Oku, 2018).

Another important development influencing educational measurement is the rapid integration of digital technologies, artificial intelligence (AI), adaptive testing, learning analytics, and computer-based assessments. These innovations have expanded the sources of educational evidence beyond traditional test scores to include process data, interaction logs, behavioural indicators, and multimodal learning traces. Consequently, educational scales must now satisfy more complex measurement requirements involving fairness, transparency, interpretability, and validity across diverse educational contexts. Recent systematic reviews indicate that technology-enhanced educational measurement demands stronger theoretical foundations regarding how evidence is generated, interpreted, and linked to educational constructs (Yu et al., 2026; Xiong & Li, 2026).

Furthermore, increasing emphasis on educational accountability, international large-scale assessments, competency-based curricula, and evidence-based policymaking has highlighted the importance of ensuring that measurement scales possess appropriate psychometric properties. Educational decisions involving student promotion, teacher evaluation, institutional accreditation, and resource allocation often depend on measurement instruments whose underlying scale properties are rarely scrutinized. Poorly constructed scales may produce misleading findings regarding learners' abilities or educational interventions, thereby affecting educational equity and decision quality. Contemporary educational measurement therefore advocates rigorous examination of construct representation, dimensionality, response processes, scale functioning, reliability, validity, measurement invariance, and fairness before educational data are used for high-stakes decision-making (Brookhart & Bonner, 2026).

Central to this discussion is the distinction between measurement and quantification. While many educational researchers assume that assigning numerical values automatically constitutes measurement, measurement theorists contend that true measurement requires evidence that numerical differences correspond meaningfully to differences in the underlying construct. This distinction challenges the widespread practice of interpreting Likert-scale scores as precise quantitative measures without adequate psychometric justification. Consequently, educational

researchers are increasingly encouraged to move beyond traditional assumptions about scale properties and adopt more rigorous approaches that establish meaningful measurement through empirical validation and theoretical justification (Feuerstahler, 2023). The concept of validity has also undergone significant transformation in contemporary educational measurement. Rather than viewing validity as a property of a test itself, current perspectives conceptualize validity as the degree to which evidence and theory support the intended interpretation and use of test scores. This unified view of validity integrates content evidence, response processes, internal structure, relationships with external variables, and consequences of testing into a comprehensive validity argument. Such perspectives require educational scales to be evaluated continuously rather than assumed to be valid based solely on previous applications or high reliability coefficients (Cook & Pitoniak, 2026).

Against this background, rethinking measurement in education has become both timely and necessary. Educational researchers, psychometricians, policymakers, and practitioners must critically examine whether existing educational scales adequately satisfy the properties required for meaningful measurement in increasingly complex educational environments. This involves questioning traditional assumptions regarding scale levels, investigating the appropriateness of statistical analyses applied to educational data, evaluating the extent to which scales represent latent constructs accurately, and considering how emerging psychometric theories and technological innovations can strengthen educational measurement. A critical examination of the properties of educational scales therefore provides an opportunity to improve the scientific quality of educational research, enhance evidence-based decision-making, and promote more accurate interpretations of educational outcomes.

The Need to Rethink Educational Measurement

Historically, educational measurement has focused primarily on quantifying individual differences through standardized testing, norm-referenced comparisons, and score interpretation. These traditional approaches have contributed significantly to educational evaluation, student placement, and accountability systems. However, contemporary educational contexts are considerably more complex than those in which conventional measurement theories were developed. Today's educational systems require instruments capable of measuring multidimensional psychological and educational constructs across culturally diverse populations, digitally mediated learning environments, and rapidly evolving instructional contexts (AERA, APA, & NCME, 2024; DeVellis & Thorpe, 2021).

The expansion of digital learning environments, artificial intelligence (AI), computer-adaptive testing, learning analytics, and technology-enhanced assessments has exposed important limitations in traditional measurement practices. Conventional psychometric approaches often assume static constructs, homogeneous populations, and paper-based testing conditions, assumptions that are increasingly inconsistent with modern educational realities. Consequently, educational scales must suit higher standards of psychometric quality to ensure valid interpretations, equitable decisions, and meaningful educational outcomes. Recent revisions of professional testing standards emphasize the growing importance of technological innovation, diversity, equity, inclusion, fairness,

transparency, and advanced psychometric methodologies in educational assessment (AERA et al., 2024; Mislevy, 2021; OECD, 2023).

Critical Properties of Educational Scales

Validity

Validity is widely regarded as the cornerstone of educational measurement because it determines whether evidence supports the intended interpretation and use of test scores. Contemporary psychometric theory no longer views validity as a fixed characteristic of an instrument; rather, it is conceptualized as the degree to which empirical evidence and theoretical rationale justify the interpretation of obtained scores for specific purposes (AERA et al., 2024; Kane, 2023). Thus, validity concerns the appropriateness of score interpretations rather than the test itself.

Modern educational researchers are expected to accumulate multiple sources of validity evidence throughout the development and application of measurement instruments. These include:

- Content validity
- Construct validity
- Criterion-related validity
- Convergent validity
- Discriminant validity
- Consequential validity

Recent scholarship emphasizes that validity is an ongoing process of evidence collection rather than a one-time statistical verification. A scale lacking sufficient validity evidence may produce misleading results, regardless of how internally consistent or statistically reliable it appears (Kane, 2023; Zumbo, 2022). Consequently, comprehensive validation has become an indispensable component of educational research and scale development.

Reliability and Measurement Precision

Reliability refers to the consistency, stability, and reproducibility of measurement outcomes under comparable conditions. Reliable instruments minimize random measurement error, thereby increasing confidence in score interpretation (DeVellis & Thorpe, 2021). A single reliability estimate of a construct may result to making misleading research conclusions. On this note, introduction of person and item reliability in instrument standardization lends support to enhancing the quality of educational scales (Oku, 2018). Nevertheless, contemporary measurement theory increasingly advocates moving beyond simple reliability coefficients toward broader evaluations of measurement precision.

Researchers are encouraged to report multiple indicators of reliability, including:

- Internal consistency reliability (e.g., Cronbach's alpha and McDonald's omega)
- Composite reliability

- Test-retest reliability
- Inter-rater reliability
- Standard error of measurement (SEM)

The Standards for Educational and Psychological Testing (AERA et al., 2024) emphasize that reliability and measurement precision are essential prerequisites for meaningful educational decision-making. Likewise, recent psychometric studies contend that reporting several complementary reliability indices provides a more comprehensive evaluation of score consistency than reliance on a single coefficient alone (Dunn et al., 2024; Hayes & Coutts, 2020).

Factor Structure and Dimensionality

One of the most frequently overlooked aspects of educational measurement is the empirical evaluation of factor structure and dimensionality. Many educational constructs—including motivation, depression, locus of control, engagement, self-efficacy, grit, digital competence, and academic resilience—are inherently multidimensional. Consequently, researchers must establish evidence regarding the relationships among observed indicators and their underlying latent constructs before interpreting scale scores (Brown, 2024).

Advanced statistical approaches now provide robust procedures for evaluating dimensionality, including:

- Exploratory Factor Analysis (EFA)
- Confirmatory Factor Analysis (CFA)
- Exploratory Structural Equation Modeling (ESEM)
- Item Response Theory (IRT)

Recent psychometric research consistently demonstrates that validating factor structures is essential before applying educational scales across different contexts and populations. Studies examining academic engagement, educational resilience, task performance, and academic grit have shown that poorly specified factor structures can compromise the validity of educational inferences and policy decisions (Brown, 2024; Marsh et al., 2020; Morin et al., 2023).

Measurement Invariance

Measurement invariance refers to the extent to which an educational instrument measures the same construct in the same manner across different populations, such as gender, age, ethnicity, language groups, cultures, or educational levels. Without evidence of invariance, comparisons between groups may reflect measurement bias rather than genuine differences in the construct being assessed (Putnick & Bornstein, 2019).

Recent psychometric literature identifies measurement invariance as one of the most essential yet underutilized properties of educational measurement. Meaningful comparisons require evidence that factor loadings, intercepts, residual variances, and thresholds function equivalently across groups (Putnick & Bornstein, 2019).

Between 2023 and 2026, numerous educational scale validation studies have incorporated measurement invariance testing as a standard component of psychometric evaluation. This growing practice reflects increased recognition that educational assessments must support equitable comparisons across increasingly diverse educational populations (AERA et al., 2024; OECD, 2023).

Fairness and Equity

Fairness has surfaced as a foundational principle of contemporary educational measurement. A fair educational assessment provides all examinees with equal opportunities to demonstrate their knowledge, skills, and abilities regardless of cultural background, language, gender, disability status, or socioeconomic circumstances (AERA et al., 2024).

The 2024 Standards for Educational and Psychological Testing position fairness alongside validity and reliability as a fundamental quality of educational assessment. Fairness encompasses every stage of assessment, including test design, administration, scoring, interpretation, and use of results. Educational assessments should therefore minimize sources of construct-irrelevant variance while promoting equitable educational opportunities (AERA et al., 2024).

Researchers are encouraged to routinely investigate:

- Differential Item Functioning (DIF)
- Cultural bias
- Linguistic bias
- Accessibility and accommodation issues
- Equity implications of score interpretation

Recent studies further emphasize that fairness extends beyond statistical analyses to include ethical considerations, inclusive assessment design, and responsible interpretation of educational outcomes in increasingly diverse learning environments (OECD, 2023; Soland, 2021).

Technological Responsiveness

Rapid advances in artificial intelligence, machine learning, educational data mining, and digital assessment platforms have fundamentally transformed educational measurement. Technology-enhanced assessments now support computer-adaptive testing, automated scoring, learning analytics, simulation-based assessment, and personalized feedback systems (OECD, 2023).

While these innovations provide unprecedented opportunities to improve assessment efficiency and diagnostic precision, they also introduce new psychometric and ethical challenges. Concerns regarding algorithmic bias, transparency, explainability, privacy, data security, validity, reliability, and fairness have become central issues in educational measurement research (AERA et al., 2024; UNESCO, 2023).

Recent scholarship contends that educational measurement must develop robust governance frameworks to ensure that AI-driven assessments remain transparent, auditable, valid, reliable, and educationally meaningful. Such frameworks should incorporate principles of responsible AI, human oversight, ethical accountability, and continuous psychometric evaluation (UNESCO, 2023; OECD, 2023).

Position Statement

This paper contends that educational measurement should be fundamentally reoriented from a narrow emphasis on score production toward a comprehensive evaluation of psychometric quality. Educational scales should not be considered adequate merely because they generate statistically significant findings or acceptable reliability coefficients. Rather, evidence supporting validity, reliability, dimensionality, measurement invariance, fairness, technological responsiveness, and ethical use should collectively determine the quality and appropriateness of educational instruments (AERA et al., 2024; Kane, 2023).

The increasing complexity of educational environments requires measurement systems capable of supporting accurate, equitable, transparent, and meaningful interpretations. Consequently, researchers, practitioners, policymakers, and educational institutions must adopt more rigorous standards for instrument development, validation, implementation, and continuous evaluation.

Implications for Educational Research

To enhance the quality and credibility of educational measurement, researchers should:

1. Adopt comprehensive validation frameworks throughout scale development and validation (AERA et al., 2024; Kane, 2023).
2. Conduct exploratory and confirmatory factor analyses before interpreting latent construct scores (Brown, 2024).
3. Test measurement invariance across relevant demographic, cultural, and educational groups to ensure valid comparisons (Putnick & Bornstein, 2019).
4. Report multiple forms of reliability, validity, and measurement precision rather than relying solely on Cronbach's alpha (DeVellis & Thorpe, 2021).
5. Assess fairness by investigating differential item functioning, accessibility, cultural bias, and equity considerations (AERA et al., 2024).
6. Integrate advanced psychometric methodologies such as Item Response Theory (IRT), Structural Equation Modeling (SEM), Exploratory Structural Equation Modeling (ESEM), and Bayesian approaches to improve measurement accuracy (Brown, 2024; Marsh et al., 2020; Oku, 2017).
7. Critically examine the implications of artificial intelligence, digital assessment technologies, learning analytics, and automated scoring systems to ensure that technological innovations enhance rather than compromise educational quality, fairness, and accountability (OECD, 2023; UNESCO, 2023).

Conclusion

The future of educational measurement relies on the quality of instruments used to measure educational phenomena. As educational systems become increasingly data-driven, reliance on poorly validated scales poses significant risks to research credibility and educational decision-making. Rethinking measurement in education requires a renewed commitment to psychometric rigor, fairness, and innovation. Critical properties such as validity, reliability, factor structure, measurement invariance, and equity must become central considerations in the development and evaluation of educational scales. Only through such an approach can educational measurement fulfill its intended purpose of providing trustworthy evidence for improving teaching, learning, and educational outcomes.

Recommendations

1. Educational researchers should adopt comprehensive psychometric validation procedures when developing and adapting educational scales. Beyond reporting reliability coefficients, researchers should provide evidence of content validity, construct validity, criterion-related validity, and measurement invariance to ensure that scale scores accurately represent the intended constructs.
2. Educational institutions and research organizations should promote the use of advanced statistical techniques such as Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), Structural Equation Modeling (SEM), and Item Response Theory (IRT) in scale development and validation. These techniques enhance the accuracy, dimensionality assessment, and interpretability of educational measurement instruments.
3. Scale developers should routinely evaluate measurement invariance and fairness across demographic groups such as gender, age, culture, socioeconomic status, and educational level. This practice will help ensure that educational scales function equitably and support valid comparisons among diverse populations.
4. Teacher educators, assessment specialists, and educational practitioners should receive continuous professional development in modern psychometric practices to improve their capacity to select, develop, interpret, and utilize high-quality educational measurement instruments in research and classroom evaluation.
5. Policymakers and curriculum evaluation agencies should establish clear standards and guidelines for educational scale development and use**,** emphasizing validity, reliability, factor structure, fairness, and ethical considerations. Such standards will enhance the quality of educational research, strengthen evidence-based decision-making, and improve the credibility of educational assessments.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2024). *Standards for educational and psychological testing*. American Educational Research Association.

Brookhart, S. M., & Bonner, S. M. (2026). When is classroom assessment educational measurement? *Educational Measurement: Issues and Practice*, *45*(2), Article e70021.

Brown, T. A. (2024). *Confirmatory factor analysis for applied research* (3rd ed.). Guilford Press.

Cook, L. L., & Pitoniak, M. J. (Eds.). (2026). *Educational measurement* (5th ed.). Oxford University Press.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage.

Dunn, T. J., Baguley, T., & Brunsden, V. (2024). From alpha to omega: Practical solutions to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *115*(1), 1–20.

Feuerstahler, L. (2023). Scale type revisited: Some misconceptions, misinterpretations, and recommendations. *Psych*, *5*(2), 234–248. https://doi.org/10.3390/psych5020018

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Kane, M. T. (2023). *Validating the interpretations and uses of test scores*. Cambridge University Press.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2020). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *16*, 85–110.

Mislevy, R. J. (2021). Sociocognitive foundations of educational measurement. *Educational Measurement: Issues and Practice*, *40*(3), 5–17.

Morin, A. J. S., Myers, N. D., & Lee, S. (2023). Modern structural equation modeling approaches in educational and psychological research. *Educational Psychology Review*, *35*(2), 1–31.

OECD. (2023). *Digital education outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing.

Oku, K., & Iweka, F. (2018). Development, standardization and application of Chemistry Achievement Test using the one-parameter logistic model of item response theory. *American Journal of Educational Research*, *6*(3), 238–257.

Oku, K., & Orluwene, G. W. (2017). Verification of assumptions of the one-parameter logistic model of item response theory using chemistry achievement test. *African Social and Educational Journal (Nigerian Edition)*, *7*(1), 129–150.

Putnick, D. L., & Bornstein, M. H. (2019). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *53*, Article 100896. https://doi.org/10.1016/j.dr.2019.100896

Soland, J. (2021). Educational equity and fairness in assessment: Current issues and future directions. *Educational Assessment*, *26*(3), 177–193.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Xiong, J., & Li, F. (2026). A decade of reflection and thematic review on artificial intelligence's impact on educational measurement. *Educational Research Review*, *51*, Article 100789.

Yu, L., Wong, G. K. W., Zhang, B., & Wang, F. (2026). Educational measurement with emerging technologies: A systematic review through evidentiary lens on granularity and constructing measures theory. *Education Sciences*, *16*(4), 661. https://doi.org/10.3390/educsci16040661

Zumbo, B. D. (2022). Consequential validity and fairness in educational assessment: Contemporary perspectives. *Educational Measurement: Issues and Practice*, *41*(4), 32–41.