e-ISBN : 978-629-98755-7-4

SIG : e-Learning@CS
https://appspenang.uitm.edu.my/sigcs/
Publication Date : 26 – Sep - 2025

BEYOND CRONBACH'S ALPHA: RETHINKING RELIABILITY ANALYSIS IN THE AGE OF DIGITAL LEARNING

*Zuraira binti Libasin[1], Nurhafizah binti Ahmad[2] and Norazah binti Umar[3]
zuraira946@uitm.edu.my[1], nurha9129@uitm.edu.my[2], norazah191@uitm.edu.my[3]

[1,2,3]Department of Computer and Mathematical Sciences,
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

**Corresponding author*

ABSTRACT

Cronbach's alpha has been widely used to measure internal consistency in educational research due to its simplicity and ease of interpretation. However, its assumptions, such as tau equivalence, unidimensionality, and equal item contribution, are often violated in modern digital learning environments. This paper examines the limitations of Cronbach's alpha and presents alternative approaches that better suit technology enhanced assessments, namely McDonald's omega, Generalizability Theory, and Item Response Theory. A practical framework is proposed to guide the selection of reliability methods based on assessment characteristics. Examples from quizzes, reflective writing, and gamified simulations illustrate how these methods address specific psychometric challenges in digital contexts. The study shows that aligning reliability techniques with assessment design improves measurement accuracy, supports adaptive feedback, and enhances transparency in reporting. It concludes that adopting modern reliability approaches and investing in methodological training are essential for creating fair and valid digital assessment practices in education.

Keywords: Cronbach's alpha, McDonald's omega, Generalizability Theory, Item Response Theory, Digital learning assessment

Introduction

Reliability testing has long played a foundational role in educational research, underpinning the psychometric integrity of instruments used to assess learning outcomes, cognitive abilities, and affective constructs. Central to these efforts is the need to determine the consistency and precision of measurement tools, ensuring that observed scores reliably reflect the constructs they are intended to capture. Among the many indices developed to assess internal consistency, Cronbach's alpha (α) remains the most widely adopted. Since its introduction, α has become the de facto standard in education and psychology for estimating the reliability of scales composed of multiple items, particularly Likert-type survey instruments and standardized assessments.

Cronbach's alpha owes its popularity to its simplicity, ease of computation, and interpretability. However, over time, extensive theoretical and empirical scrutiny has revealed significant limitations in its assumptions and applicability. A presumes tau-equivalence, one-dimensionality, and uncorrelated errors across items conditions rarely met in real-world educational contexts (Zinbarg, Revelle, Yovel, & Li, 2005; Dunn, Baguley, & Brunsden, 2014). While researchers often rely on α due to convention or software defaults, its misuse can lead to misestimate reliability coefficients and misinformed conclusions about the quality of measurement instruments.

These concerns become even more pronounced in the context of digital learning environments. The rise of e-learning platforms, interactive simulations, gamified assessments, and complex data-driven learning analytics introduces new challenges for traditional reliability frameworks. Digital assessments often incorporate heterogeneous item types (e.g., video-based prompts, open-ended reflections, embedded quizzes), platform-dependent interactivity, and dynamic feedback mechanisms. Such features inherently violate key assumptions of α, calling into question its suitability for evaluating measurement precision in contemporary settings (Liu, Pek, & Maydeu-Olivares, 2024). As educational assessment evolves toward more diverse, adaptive, and data-rich formats, there is a pressing need to re-examine classical reliability approaches and explore alternatives better suited for the digital age.

This article explores the theoretical and practical limitations of Cronbach's alpha and evaluates more robust alternatives that better align with the evolving nature of digital learning environments. It aims to critically evaluate the continued reliance on Cronbach's alpha in educational research, particularly in the context of digital learning, and to propose more appropriate alternatives grounded in recent advancements in psychometric theory. Drawing upon recent literature, including comparative analyses of alpha and omega coefficients (Orcan, Celik, & Gungor, 2023), item response theory (Wang & Bao, 2010), and generalizability theory frameworks, this paper offers a structured synthesis of current reliability approaches and their applicability to various digital learning scenarios.

The article contributes by advancing a practical framework for selecting and applying reliability techniques tailored to specific digital data types, thereby enhancing the methodological robustness of assessment practices in modern educational contexts.

Literature Review

In the evolving landscape of digital learning, the reliability of assessment tools and measurement instruments has become a critical concern. The shift toward adaptive technologies, multimedia-rich content, and artificial intelligence in education has transformed how learning is delivered and assessed.

However, these innovations also challenge the adequacy of traditional reliability metrics, which were developed for static, unidimensional assessment formats. As digital learning environments become more complex, there is a growing need to reconsider the tools used to evaluate the consistency and dependability of educational measurements. This literature review critically examines both traditional and emerging approaches to reliability analysis, highlighting the limitations of legacy methods such as Cronbach's alpha, and exploring advanced alternatives better suited to the digital age.

Traditional Reliability Measures and Their Limitations

Cronbach's alpha remains one of the most widely used statistics for evaluating the internal consistency of measurement instruments, especially in educational, psychological, and biomedical research (Taber,

2018; Kotian et al., 2022). It provides an estimate of how closely related items in a scale are, serving as a proxy for reliability. Despite its widespread application, several researchers have highlighted critical limitations that question its continued relevance in increasingly complex and dynamic research contexts. A primary limitation of Cronbach's alpha is its reliance on the assumptions of tau-equivalence and unidimensionality conditions that are often violated in real-world applications (Sijtsma, 2009).

Furthermore, alpha is highly sensitive to the distribution of items, with skewed or non-normal data producing misleading estimates (Christmann, 2006). Its inability to account for the internal structure of multidimensional constructs also makes it inadequate for modern digital learning environments where assessments are diverse and interactive (Kumar,2024). Moreover, Cronbach's alpha is frequently misinterpreted. Researchers often over-rely on it as the sole indicator of reliability, overlooking more appropriate or nuanced alternatives. This reliance can compromise the validity of research findings, particularly in digital education, where traditional test characteristics may not apply. For instance, static assessments fail to capture the dynamic nature of sensor-based, AI-driven, or adaptive learning systems.

Additionally, traditional reliability measures such as test-retest or split-half methods do not adequately address the complexity of big data, machine learning, and real-time assessments in digital platforms. As Eagan (2020) and Rosli(2021) state, conventional coding reliability approaches in educational analysis can produce high Type I error rates, underscoring the need for more sophisticated and adaptive techniques.

Emerging and Alternative Approaches to Reliability

Given the pressing need for more robust reliability metrics, a range of emerging methods has been proposed to address the shortcomings of traditional approaches. These methods better align with the evolving demands of digital learning environments.

(1) McDonald's Omega (ω)

McDonald's Omega (ω) has gained traction as a superior alternative to Cronbach's alpha. Unlike alpha, Omega does not assume tau-equivalence and is derived from factor analytic models that allow for congeneric measures items that assess the same construct but with varying factor loadings and error variances (Hancock, 2020). This makes ω more flexible and applicable to a wider range of data sets.

Simulation studies have shown that while α tends to slightly underestimate reliability, ω provides a more accurate estimate, especially in large samples (Malkewitz,2023). Omega offers a more accurate estimate of the proportion of score variance attributable to the common factor, making it particularly useful in settings where item variances differ. ω generally performs better than α in handling missing data, providing more consistent reliability estimates (Malkewitz,2023). Its applicability in non-

uniform item structures, common in multimedia and gamified assessments that makes it ideal for digital learning environments (Hancock & An, 2020). ω has been successfully applied in diverse fields, from psychological assessments to educational measurements, demonstrating its versatility and robustness (Yupari,2023 ; Wang, 2024)

(2) Generalizability Theory (G-Theory)

Generalizability Theory provides a powerful framework for identifying and quantifying multiple sources of measurement error. G-theory assumes that any measurement situation has multiple sources of variation and error. Analysis of variance (ANOVA) methods is used to disentangle these sources, providing a more detailed understanding of measurement error compared to classical test theory (Vispoel, 2025). It extends classical test theory by considering facets such as raters, items, tasks, occasions, and even digital platforms. This is particularly relevant for digital assessments involving peer evaluations, interactive media, or multi-platform delivery.

G-Theory enables researchers to design assessments that minimize error and optimize reliability across conditions. According to Vispoel (2025), G-theory supports both univariate and multivariate designs, allowing researchers to assess score consistency and measurement error at different levels of score aggregation. Its application across different fields demonstrates its versatility and effectiveness in enhancing the reliability and validity of measurement (Clayson,2021).

(3) Item Response Theory (IRT) – Based Reliability

Item Response Theory (IRT)-based reliability is a modern, model-based approach to measuring the precision and consistency of a test or scale, especially when items vary in difficulty, discrimination, and format. It provides item-level metrics such as difficulty and discrimination, allowing researchers to estimate reliability across different levels of ability or score ranges (Embretson & Reise, 2000). This is especially useful in adaptive digital testing, where each learner may encounter a different subset of items.

IRT-based reliability goes beyond static estimates, capturing measurement precision tailored to individual learner profiles. IRT provides more precise reliability estimates by considering the properties of individual items and their interaction with the latent trait being measured (Milanzi,2015). IRT helps in developing and validating instruments for assessing psychological construct and health-related quality of life (Cui,2025)

Rethinking Reliability Assessment in the Digital

The emergence of advanced technologies such as machine learning and artificial intelligence has redefined the methodologies used in reliability assessment. These innovations have enabled more

dynamic, adaptive, and responsive forms of measurement, allowing reliability to be evaluated in real time with high precision and minimal human intervention (Teixeira, 2024). Digital learning platforms are now capable of automatically adapting to shifting learner behaviours, data patterns, and system parameters—features essential for personalized and adaptive education. Moreover, adaptive learning systems employ intelligent algorithms to tailor learning content and strategies according to individual student behaviours and characteristics (Cai, 2024).

Modern digital environments also support sophisticated tools for calculating, simulating, and visualizing advanced reliability coefficients beyond traditional static indices. Learning management systems like Moodle and Blackboard offer integrated analytic capabilities that not only streamline data collection but also improve the efficiency and transparency of reliability estimation, contributing to scalable and sustainable educational practices (Gavrus, 2025).

Despite these advancements, traditional reliability approaches such as test-retest and alternate-form methods still hold relevance, particularly in evaluating the temporal stability of assessments. However, their application must be reconsidered within the context of modern, technology-enhanced learning. For instance, when assessing stability in an adaptive learning system, test-retest procedures must account for fluctuating item exposure, personalized content sequencing, and diverse learner interaction pathways. As suggested by Wyse (2021), a retest interval of just over three weeks strikes a balance between preserving reliability and accommodating natural learner development, making it a useful guideline even in digital contexts.

Challenges in Digital Learning Contexts

The evolution of digital learning environments has redefined the structure and delivery of educational assessments. As education shifts further into virtual and hybrid spaces, the challenges associated with measuring reliability through traditional psychometric tools, such as Cronbach's alpha, become increasingly evident. The digital learning context introduces a series of complexities that directly challenge the assumptions and limitations of conventional reliability analysis.

Variety of Assessment Types

Unlike traditional classroom-based assessments, digital learning integrates a wide spectrum of assessment formats including e-quizzes, simulations, reflective discussion forums, and gamified tasks. These modalities are often designed to assess a range of competencies cognitive, metacognitive, affective, and even social engagement using diverse approaches.

This heterogeneity in format disrupts the uniformity typically required by classical test theory (CTT), where reliability measures such as Cronbach's alpha assume homogeneity in item structure and function. When items vary significantly in form and cognitive demand, interpreting internal consistency

becomes problematic, leading to misleading or oversimplified reliability estimates. Green (2015) stated that, when items are multidimensional, measures like Cronbach's alpha can yield high but misleading reliability coefficients

Use of Non-Uniform Items

In digital learning, items are not limited to traditional multiple-choice or Likert-type scales. Assessments frequently involve multimedia elements (e.g., video or audio prompts), drag-and-drop interfaces, or open-ended written reflections. These non-uniform item types carry different response structures, scoring schemes, and interaction modalities. As a result, assumptions such as equal item contribution and linearity, which are central to Cronbach's alpha, are violated.

Open-ended responses, for instance, may be scored subjectively or via AI-based rubrics, which introduces another layer of variability. The inclusion of multimedia in assessments can influence response accuracy and perceived difficulty, a phenomenon known as the Multimedia Effect in Testing (Arts et al., 2024). This diversity in item structure calls for more nuanced reliability approaches that can handle multidimensionality and heterogeneity within digital instruments.

Dependence on Platforms, Algorithms, and Digital Data Analytics

Digital assessments rely heavily on underlying platforms and algorithms that govern content delivery, data logging, adaptive testing, and scoring. While these systems enhance personalization and interactivity, they also introduce new sources of error variance. For example, a platform's recommendation algorithm may expose learners to different item sets based on prior responses, creating inconsistencies in test experience across participants.

Similarly, platform-based analytics used for formative feedback may confound the measurement process if the feedback loop affects subsequent item responses. These platform-dependent variances are not accounted for by traditional reliability coefficients, which assume a static and uniform assessment experience.

Difficulty Maintaining Assumptions Like Equal Item Contribution

Cronbach's alpha relies on the assumption that all items in a scale contribute equally to the latent construct being measured. In digital learning, however, this assumption is difficult to uphold. Digital learning environments often involve a variety of activities and assessments, such as synchronous and asynchronous sessions, interactive activities, and digital resources, which may not equally contribute to the overall construct being measured (Fuster,2025). For example, in a gamified task, certain levels or scenarios may have greater impact on learner engagement or performance than others.

Similarly, in a reflective discussion forum, some prompts may elicit richer responses than others, depending on the learner's context or prior knowledge. This unequal item contribution undermines the conceptual foundation of internal consistency and calls for alternative reliability indices that can accommodate hierarchical or weighted item structures.

Alternative Reliability Analysis Methods

Considering the limitations associated with Cronbach's alpha, particularly in settings that violate assumptions of tau-equivalence and one-dimensionality, several alternative methods have emerged as more theoretically sound and empirically appropriate for assessing reliability in modern digital learning environments. These alternatives namely McDonald's omega, Generalizability Theory (G-Theory), and Item Response Theory (IRT) provide greater flexibility in handling multidimensional constructs, heterogeneous item formats, and complex data sources, which are increasingly prevalent in digital educational assessments.

Omega Coefficient

McDonald's omega (ω) has gained prominence as a more robust measure of internal consistency, especially in scales where items exhibit varying factor loadings or multidimensional structure. Unlike α, which assumes equal item contributions (tau-equivalence), ω accounts for the actual loadings of each item on a latent factor, thereby providing a more accurate estimate of true score variance (Zinbarg et al., 2005; Dunn et al., 2014). This makes ω particularly suitable for digital learning instruments, such as multimedia-based surveys or performance tasks, where item characteristics often differ in complexity and cognitive demand. Simulation studies by Orcan, Celik, and Gungor (2023) further demonstrate that ω yields more stable reliability estimates under conditions of low item homogeneity or limited sample size, both of which are common in digital pilot evaluations. Nevertheless, ω still relies on a factor analytic model and assumes correct model specification; as such, its performance is sensitive to mis-specified factor structures, which may occur in exploratory assessments with minimal theoretical underpinning.

Generalizability Theory (G-Theory)

Generalizability Theory extends classical test theory by modelling multiple sources of measurement error simultaneously. Unlike α or ω, which yield a single reliability index, G-Theory decomposes variance into multiple facets such as items, raters, tasks, or occasions enabling researchers to estimate generalizability coefficients across complex designs. This approach is particularly beneficial in digital contexts where assessments often involve multifaceted interactions, such as combinations of auto-graded quizzes, peer-assessed discussions, and reflective journal entries. For example, Liu, Pek, and

Maydeu-Olivares (2024) highlighted the ability of G-Theory to capture contextual variance arising from different modes of item delivery, such as synchronous versus asynchronous digital formats. While G-Theory provides a highly nuanced picture of score dependability, it requires larger sample sizes and carefully structured study designs to estimate variance components accurately, which may pose practical constraints in small-scale classroom settings.

Item Response Theory (IRT)

Item Response Theory offers an entirely different paradigm by modelling the probability of item responses as a function of underlying latent traits. Its flexibility in addressing item characteristics such as discrimination, difficulty, and guessing makes it particularly well-suited for adaptive digital assessments and log-based evaluations. In adaptive environments, IRT models especially the 2PL and 3PL models enable real-time tailoring of items based on learner performance, enhancing both engagement and measurement precision (Wang & Bao, 2010). Additionally, IRT has been employed to model user interaction patterns via clickstream data, capturing response times, navigation paths, and keystroke dynamics as behavioural indicators of cognitive processes. However, IRT-based reliability coefficients require sophisticated estimation procedures and large item pools with pre-calibrated parameters, which may not always be feasible in emerging educational technology applications or low-resource environments.

Collectively, these methods represent a methodological shift toward more context-sensitive and theoretically grounded approaches to reliability assessment in education. Each offers distinct advantages depending on the nature of the instrument and digital data environment. Omega is suitable for multidimensional constructs with variable item loadings, G-Theory excels in multifaceted assessment designs, and IRT provides precision and adaptability in dynamic testing contexts. However, none are universally optimal, and each requires specific assumptions, data structures, and technical expertise. Accordingly, the selection of reliability methods must be informed not only by statistical considerations but also by the epistemological and practical demands of digital education.

Proposed Practical Framework

This article proposes a practical framework to guide the selection of appropriate reliability analysis methods for digital learning assessments. As assessment formats continue to diversify, including objective quizzes, reflective writing, and gamified tasks, relying solely on Cronbach's Alpha is no longer sufficient. Aligning the psychometric method with the specific characteristics of the assessment enhances the validity, precision, and interpretability of results. A simplified visual (see Figure 1) supports this approach by mapping common assessment types to their recommended reliability methods (Orçan, 2023; Njeri, Rop, & Too, 2023).

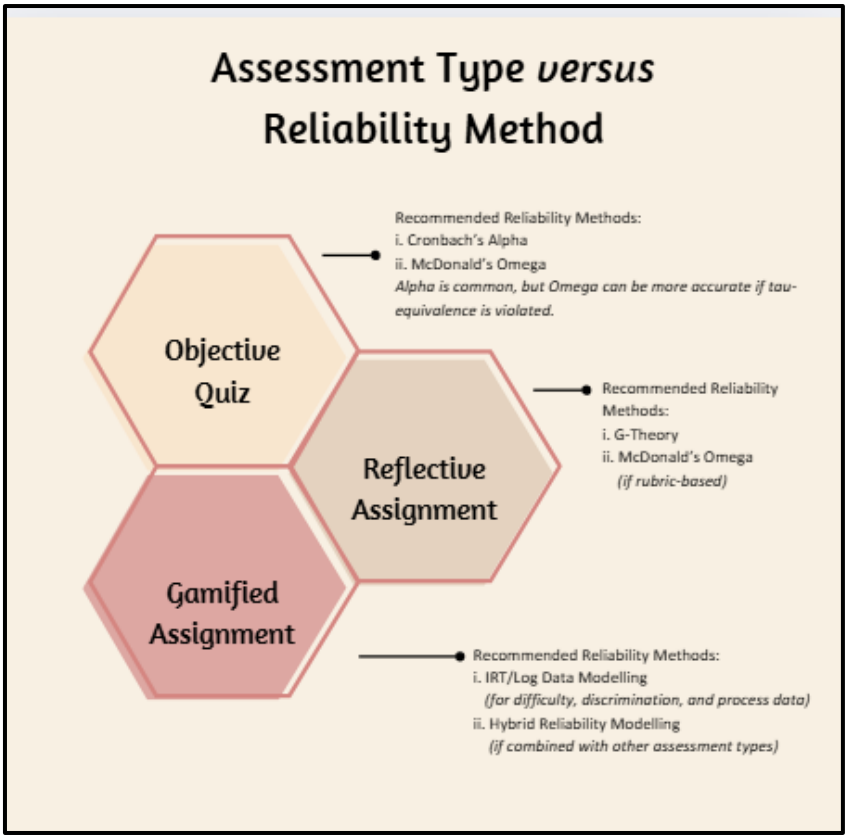

Figure 1. Proposed Framework for Selecting Reliability Methods in Digital Learning Assessments

Objective quizzes are typically analysed using Cronbach's Alpha because of its accessibility and popularity. However, Cronbach's Alpha assumes tau-equivalence, meaning that all items contribute equally to the total score. In many real-world cases, this assumption is violated. McDonald's Omega provides a more accurate estimate of internal consistency in such situations by incorporating item loadings derived from factor analysis (Revelle & Condon, 2019; Stensen, Wendt, Wacker, & Esser, 2022). For reflective assessments such as journals and essays, variability in scoring is often introduced by differences among raters or scoring conditions. Generalizability Theory accounts for multiple sources of variance in these complex settings, making it a more appropriate method for estimating score reliability (Atılgan, 2019; Shavelson & Webb, 1991). Assessments that involve gamified or adaptive tasks generate interaction data that go beyond right or wrong responses. Item Response Theory is especially useful in modelling item characteristics such as difficulty and discrimination in these contexts. Furthermore, digital learning environments produce log data that can be analysed through behavioural modelling techniques like time-on-task and action-sequence analysis, offering additional insights into the consistency and reliability of learner behaviours (Bell, Ferrell, & Ward, 2024; Tempelaar, Rienties, & Giesbers, 2015). For assessments that combine multiple types of items or tasks, a hybrid approach is recommended. This might involve using McDonald's Omega for the objective

component, Generalizability Theory for reflective responses, and IRT or log-data modelling for interactive parts, followed by a composite reliability estimate that accounts for all components (Van der Linden, 2016; Bell et al., 2024).

To illustrate this framework, consider a digital engineering course in which students complete three types of assessments: a ten-item multiple-choice quiz, a reflective journal, and a simulation-based system design task. The multiple-choice quiz was analysed using Cronbach's Alpha, which yielded a reliability coefficient of 0.82. A confirmatory analysis using McDonald's Omega produced a comparable value, supporting the consistency of the instrument. Reflective journals scored by two raters were analysed using Generalizability Theory, resulting in a generalizability coefficient of 0.75 and showing that only 10 percent of the score variance was attributable to rater effects (Atılgan, 2019). The simulation assessment was evaluated using a two-parameter IRT model, which identified two items with low discrimination values, prompting revision. Additional log data were used to track time and interaction patterns, revealing meaningful trends that correlated with performance levels (Tempelaar et al., 2015). This case demonstrates the value of applying targeted reliability methods that suit the structure of each assessment type.

Discussion and Implications

Shifting from Cronbach's alpha to modern reliability methods has important implications for digital learning. While alpha remains common, it often underestimates reliability when assumptions like equal item contribution are violated. Omega provides a more accurate alternative by accounting for factor structure, making it more suitable for diverse and complex digital assessment formats (McNeish, 2018; Revelle & Zinbarg, 2009). This shift urges researchers and instructional designers to adopt accessible tools such as *psych*, *GeneralizIT*, and *mirt*, while platforms like OpenMx and easystats support more advanced modelling workflows (Makowski et al., 2022). Incorporating these methods not only improves assessment quality and personalized feedback but also promotes greater transparency in reporting measurement accuracy. Institutions that invest in methodological training and psychometric literacy will be better equipped to create fair, valid, and data-informed digital learning environments (Andersson et al., 2022).

Conclusion

In the era of digital learning, the continued reliance on Cronbach's alpha as the primary measure of reliability is increasingly untenable. While Cronbach's alpha has served as a valuable tool for decades, its restrictive assumptions, particularly tau equivalence, unidimensionality, and equal item contribution, are frequently violated in modern technology enhanced assessments. The complexity of digital learning

environments, encompassing diverse item formats, platform dependencies, adaptive algorithms, and multidimensional constructs, demands more flexible and context sensitive approaches.

McDonald's omega, Generalizability Theory, and Item Response Theory each offer distinct advantages for different assessment scenarios, enabling more accurate and nuanced measurement of score dependability. By aligning reliability analysis methods with the structural and functional characteristics of digital assessments, researchers can ensure higher psychometric precision, stronger validity, and more actionable insights. This shift requires both methodological awareness and institutional investment in psychometric literacy. Ultimately, adopting a tailored multi method reliability framework will support the creation of fairer, more transparent, and data driven educational environments that keep pace with the rapidly evolving digital landscape.

References:

A. Christmann, S. Van Aelst, Robust estimation of Cronbach's alpha, Journal of Multivariate Analysis, Volume 97, Issue 7, 2006, Pages 1660-1674, ISSN 0047-259X, https://doi.org/10.1016/j.jmva.2005.05.012.

Andersson, C., Gyllenpalm, J., & Lindberg, R. (2022). Assessing student learning in digital environments: The importance of reliable and valid data. Educational Assessment, 27(2), 95–112. https://doi.org/10.1080/10627197.2022.2030104

Arts, J., Emons, W., Dirkx, K., Joosten-ten Brinke, D., & Jarodzka, H. (2024, April). Exploring the multimedia effect in testing: the role of coherence and item-level analysis. In Frontiers in Education (Vol. 9, p. 1344012). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1344012

Atılgan, H. (2019). Reliability of essay ratings: A study using generalizability theory. Eurasian Journal of Educational Research, 82, 183–204. https://doi.org/10.14689/ejer.2019.82.10

Bell, A. R., Ferrell, J., & Ward, S. (2024). Behavioral indicators of learning in digital assessments: A framework for integrating log data and IRT. Journal of Educational Data Science, 6(1), 22–41. https://doi.org/10.1016/j.jeds.2024.03.005

Cai, Q., & Liu, Q. (2024, September). Exploration and practice of personalized education based on adaptive learning systems. In Proceedings of the 2024 International Symposium on Artificial Intelligence for Education (pp. 54-58).

Clayson, P. E., Carbine, K. A., Baldwin, S. A., Olsen, J. A., & Larson, M. J. (2021). Using generalizability theory and the ERP Reliability Analysis (ERA) Toolbox for assessing test-retest reliability of ERP scores Part 1: Algorithms, framework, and implementation. International Journal of Psychophysiology, 166, 174-187. https://doi.org/10.1016/j.ijpsycho.2021.01.006

Cui, J., Wang, J., Yue, A., Cao, J., Zhang, Z., & Shi, B. (2025). Psychometric evaluation of the Chinese version of the Patient Self-Advocacy Scale using classical test theory and item response theory. Scientific Reports, 15(1), 6871.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399–412. https://doi.org/10.1111/bjop.12046

Eagan, B., Brohinsky, J., Wang, J., & Shaffer, D. W. (2020, March). Testing the reliability of inter-rater reliability. In Proceedings of the tenth international conference on learning analytics & knowledge (pp. 454-461). https://doi.org/10.1145/3375462.3375508

Fuster-Guillén, D., Olivera Espinoza, J., Chumpen Elera, A. C., Peña Ricapa, I. L., & Hernández, R. M. (2025). Psychometric evidence of the questionnaire to evaluate the didactic sequence in digital environments in university students. Journal of Curriculum and Teaching, 14(1), 334–347. https://doi.org/10.5430/jct.v14n1p334

Gavrus, C., Petre, I. M., & Lupșa-Tătaru, D. A. (2025). The Role of e-Learning Platforms in a Sustainable Higher Education: A Cross-Continental Analysis of Impact and Utility. Sustainability (2071-1050), 17(7).

Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. Educational Measurement: Issues and Practice, 34(4), 14-20. DOI: 10.1111/emip.12100

Hancock, G. R., & An, J. (2020). A Closed-Form Alternative for Estimating ω Reliability under Unidimensionality. Measurement: Interdisciplinary Research & Perspective, 18(1), 1–14. https://doi.org/10.1080/15366367.2019.1656049

Kotian, H., Varghese, A. L., & Motappa, R. (2022). An R Function for Cronbach's Alpha Analysis: A Case-Based Approach. National Journal of Community Medicine, 13(08), 571–575. https://doi.org/10.55489/njcm.130820221149

Kumar, R. V. (2024). Cronbach's Alpha: Genesis, Issues and Alternatives. https://doi.org/10.1177/ijim.241234970

Liu, Y., Pek, J., & Maydeu-Olivares, A. (2024). On a general theoretical framework of reliability. arXiv preprint arXiv:2407.00716. https://doi.org/10.48550/arXiv.2407.00716

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2022). easystats: A suite of R packages to easily report results of statistical models. Journal of Open Source Software, 7(70), 3903. https://doi.org/10.21105/joss.03903

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ωt and the greatest lower bound. Social Sciences & Humanities Open, 7(1), 100368. https://doi.org/10.1016/j.ssaho.2022.100368

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. Psychological Methods, 23(3), 412–433. https://doi.org/10.1037/met0000144

Milanzi, E., Molenberghs, G., Alonso, A., Verbeke, G., & De Boeck, P. (2015). Reliability measures in item response theory: Manifest versus latent correlation functions. British Journal of Mathematical and Statistical Psychology, 68(1), 43-64. DOI: 10.1111/bmsp.12033

Njeri, N. G., Rop, R. K., & Too, J. K. (2023). Comparing Cronbach's alpha and McDonald's omega in digital test environments. International Journal of Educational Measurement, 15(4), 144–159. https://doi.org/10.1177/ijem.2023.015041

Orcan, F., Celik, M., & Gungor, H. (2023). Comparison of Cronbach's alpha and McDonald's omega for ordinal data: A simulation study. Journal of Measurement and Evaluation in Education and Psychology, 14(3), 203–215. https://doi.org/10.21031/epod.1417464

Orçan, F. (2023). Revisiting internal consistency: When and why to use omega over alpha. Measurement and Evaluation in Counseling and Development, 56(1), 1–14. https://doi.org/10.1080/07481756.2023.2200147

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika, 74(1), 145–154. https://doi.org/10.1007/s11336-008-9102-z

Revelle, W., & Condon, D. M. (2019). Reliability from alpha to omega: A tutorial. Psychological Assessment, 31(12), 1395–1411. https://doi.org/10.1037/pas0000754

Shavelson, R. J., & Webb, N. M. (1991). Generalizability theory: A primer. SAGE Publications.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. Psychometrika, 74(1), 107–120. doi:10.1007/s11336-008-9101-0

Stensen, H. T., Wendt, L. P., Wacker, A., & Esser, G. (2022). The reliability debate revisited: Alpha, Omega, and the dimensionality of scales. Frontiers in Psychology, 13, 897055. https://doi.org/10.3389/fpsyg.2022.897055

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior, 47, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Van der Linden, W. J. (2016). Handbook of item response theory (Vol. 1). Chapman and Hall/CRC.

Villaseñor, Á. B., Paulis, J. C., Hernández-Mendo, A., Sánchez, C. R., & Usabiaga, O. (2014). Application of the Generalizability Theory in sport to study the validity, reliability and estimation of samples. Revista de Psicologia del Deporte, 23(1), 131-137.

Vispoel, W. P., Lee, H., & Chen, T. (2025). Structural Equation Modeling Approaches to Estimating Score Dependability Within Generalizability Theory-Based Univariate, Multivariate, and Bifactor Designs. Mathematics (2227-7390), 13(6). https://doi.org/10.3390/

Wang, J., & Bao, L. (2010). Analyzing Force Concept Inventory with Item Response Theory. American Journal of Physics, 78(10), 1064–1070. https://doi.org/10.1119/1.3467800

Wang, X., Ren, J., & Wang, H. (2024). Psychometric characteristics of the Chinese version of the Pregnancy Exercise Attitudes Scale (C-PEAS). BMC Pregnancy and Childbirth, 24(1), 679. https://doi.org/10.1186/s12884-024-06817-0

Wyse, A. E. (2021). How days between tests impacts alternate forms reliability in computerized adaptive tests. Educational and Psychological Measurement, 81(4), 644-667.

Yupari-Azabache, I. L., Díaz-Ortega, J. L., Bardales-Aguirre, L. B., Barros-Sevillano, S., & Paredes-Díaz, S. E. (2023). Validity and reliability of the knowledge, attitudes and practices Instrument regarding monkey pox in Peru. Risk Management and Healthcare Policy, 1509-1520. https://www.tandfonline.com/doi/pdf/10.2147/RMHP.S420330

Zhou, M., & Pardos, Z. A. (2020). Towards automatic skill discovery in educational games. Journal of Educational Data Mining, 12(2), 1–25. https://doi.org/10.5281/zenodo.3922961

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's alpha, Revelle's beta, and McDonald's omega: Their relations with each other and two alternative conceptualizations of reliability. Psychometrika, 70(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7