AMERICAN JOURNAL OF EDUCATION AND LEARNING
ISSN: 2996-5128 (online) | ResearchBib (IF) = 10.91 IMPACT FACTOR
Volume-4| Issue-6| 2026 Published: |30-06-2026|

METHODOLOGICAL APPROACHES TO EVALUATING THE VALIDITY AND RELIABILITY OF PEDAGOGICAL ASSESSMENT MATERIALS

https://doi.org/10.5281/zenodo.20521247

Bazarbayeva Aynura Raximnazarovna
1st year of Master's degree
Nukus state pedagogical institute named after Ajiniyaz
(Republic of Karakalpakstan, Nukus)

Abstract

This article examines the methodological foundations for assessing the reliability and validity of pedagogical tests as the main criteria for measurement quality. It provides a comparative analysis of tools from classical (CTT) and modern (IRT) test theories, particularly the calculation of Cronbach's alpha coefficient and correlational methods. The article also describes the structure of content, criterion, and construct validity and their influence on the objectivity of educational outcomes. This work systematizes new psychometric approaches aimed at the verification of assessment tools to improve the accuracy of pedagogical diagnostics.

Keywords

Pedagogical testing, reliability, validity, classical test theory, IRT, Cronbach's alpha coefficient, psychometric analysis, qualimetry, assessment objectivity.

The issue of objectivity and accuracy in pedagogical measurements is one of the fundamental problems in modern educational qualimetry, as test results form the basis for making administrative decisions, certifying learners, and evaluating the effectiveness of educational programs. Within the scientific discourse, a pedagogical test is considered not merely a collection of assessment tasks, but a complex measurement system, qualitatively defined by two fundamental psychometric properties: reliability and validity. The scientific substantiation of these parameters requires the application of a rigorous mathematical-statistical apparatus and a methodological consistency that neutralizes subjectivity in the interpretation of data.

The reliability of a pedagogical test indicates the degree of precision and stability of measurement results obtained when the instrument is used repeatedly with the same group of test-takers. In the context of classical test theory, reliability is based on the postulate that a test-taker's observed score consists of two

components: the true score and the random error of measurement. Therefore, the reliability coefficient is defined as the proportion of the true score variance to the total observed score variance. In modern pedagogy, several primary strategies are used to determine reliability empirically. The test-retest method involves administering the same instrument twice with a certain time interval, where the Pearson correlation coefficient between the results of the two sessions serves as the measure of reliability. However, this method is associated with the risk of data distortion due to the learners' practice and memory effects. As an alternative, there is the parallel forms method, which requires creating two equivalent versions of a test that conform to the same specification. A more economical and frequently used method is the assessment of internal consistency, specifically, the calculation of Cronbach's Alpha coefficient. This indicator makes it possible to assess how homogeneous the test items are and whether they measure a single, unified construct. Mathematically, this coefficient compares the number of items and the sum of their individual variances to the total variance of the entire test. For dichotomous items, the Kuder-Richardson formula, which is a special case of the alpha coefficient, is traditionally used. When analyzing reliability, it is also crucial to consider the length of the test because, according to the Spearman-Brown prophecy formula, an increase in the number of quality items leads to a systematic increase in the overall reliability indicator. Nevertheless, high reliability is considered a necessary but not sufficient condition for test quality, as it only guarantees the precision of the measurement but does not guarantee its alignment with the intended objectives.

The validity of a pedagogical test is a more complex, multifaceted characteristic that determines the degree of its appropriateness for its functional purpose—that is, it indicates whether the test truly measures the knowledge and skills it was designed to measure. Unlike reliability, validity is not a single, predetermined numerical coefficient but rather a collection of evidence that substantiates the interpretation of the results. In modern classification, content, criterion-related, and construct validity are distinguished. Content validity is established during the test design phase through expert analysis. A group of subject matter experts compares the content of the items with the curriculum, the didactic units being assessed, and the learning objectives. Evaluating content validity often involves calculating specific coefficients, a process in which experts assess the relevance of each item to the domain being measured. Criterion-related validity, in turn, relies on the statistical comparison of test scores with an external, independent criterion, such as academic performance indicators or the results of other standardized exams. It is further divided into concurrent and predictive

validity. Construct validity is the most profound level of validation, aimed at confirming the theoretical model underlying the test. Factor analysis is employed to confirm it, which allows for identifying the latent structure of the test and determining how much of the results' variance is explained by the target construct and how much by extraneous factors. Concurrently, convergent and divergent validity methods are also used: in the former, the presence of a high correlation with tests measuring similar constructs is verified, while in the latter, the absence of a correlation with theoretically unrelated abilities is checked.

Integrating reliability and validity assessment methods allows a researcher to form a comprehensive psychometric view of an instrument's quality. It is important to note the dialectical relationship between these parameters: a test with low validity can be highly reliable, but a low level of reliability inevitably lowers a test's validity, limiting its upper boundary to the square root of the reliability coefficient. In recent decades, classical test theory has been supplemented by methods from the modern theory of modeling and parameterizing pedagogical tests—Item Response Theory (IRT). Within this theory, the concept of reliability is re-examined through the concept of the test's information function, which indicates the precision of measurement at various proficiency levels of the test-takers. The Rasch model, as a one-parameter model, makes it possible to calculate the probability of a student answering correctly as an exponential function of the difference between their proficiency level and the difficulty level of a specific item. This approach ensures the invariance of the assessment of test-taker and item parameters, which is a significant step toward ensuring the objectivity of pedagogical measurements compared to traditional methods.

The validation process for a pedagogical test must be conducted continuously. It should include not only a priori analysis but also an a posteriori review of the results after each multifaceted application of the test. The psychometric analysis of distractors, calculation of the discrimination index, and analysis of the descriptive characteristics of items are considered standard procedures in academic practice. Particular attention must be paid to minimizing systematic errors that directly undermine construct validity, such as cultural bias or linguistic complexity unrelated to the subject of measurement. Thus, applying rigorous scientific methods to determine validity and reliability transforms pedagogical testing from empirical observation into a high-precision scientific instrument capable of providing relevant and objective information about the quality of the educational process. The study of these characteristics requires an interdisciplinary approach that combines pedagogy, psychology, and mathematical statistics. This allows for

the minimization of errors and ensures the fairness of assessment procedures within educational systems.

Conclusion: To summarize the foregoing, it can be noted that reliability and validity are fundamental determinants of the quality of pedagogical measurements, ensuring the scientific grounding and objectivity of test results. The analysis revealed that achieving a high level of reliability, manifested through internal consistency and stability of data, is a necessary condition for ensuring the validity of the measurement instrument. The application of mathematical-statistical methods, such as calculating the Cronbach's alpha coefficient, correlation and factor analysis, as well as models from modern test theory (IRT), not only minimizes the impact of random and systematic measurement errors but also allows for a deeper study of the latent structure of the educational achievements being assessed. The modern paradigm of pedagogical diagnostics demands a shift from the intuitive design of assessment materials to rigorous psychometric engineering. Here, validation is viewed as a continuous process of gathering empirical evidence of a test's conformity to its functional purpose. Consequently, the integration of classical and modern methods in assessing test quality guarantees the transparency of educational monitoring, serving to make measurement results a reliable basis for making informed pedagogical and managerial decisions within the modern education system.

REFERENCES:

1.Kroker L., Algina Dj. Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov. — M.: Logos, 2010.

2.Chelishkova M. B. Teoriya i praktika konstruirovaniya pedagogicheskix testov. — M.: Logos, 2002.

3.Cronbach L. J. Coefficient alpha and the internal structure of tests // Psychometrika. — 1951. — Vol. 16. — P. 297–334.

4.Lord F. M. Applications of Item Response Theory to Practical Testing Problems. — Mahwah, NJ: Lawrence Erlbaum, 1980.

5.Kengesbayevich, R. M. (2025). TRADITIONS OF RELIGIOUS PEDAGOGY IN THE STUDY OF SOCIAL EDUCATION. *AMERICAN JOURNAL OF EDUCATION AND LEARNING*, *3*(1), 28-32.

6. Kengesbayevich, R. M. (2025). PERSONAL VALUES IN THE STRUCTURE OF SPIRITUAL AND MORAL EDUCATION. *AMERICAN JOURNAL OF MULTIDISCIPLINARY BULLETIN*, *3*(1), 1-4.