15-AESSR-2304-484

Academy of
Education and Social Sciences Review
Vol. 3 No. 2 (May, 2023), 147-153
ISSN (p): 2790-8348 | ISSN (e): 2789-6781
DOI: https://doi.org/10.48112/aessr.v3i2.484

Original Article

Fundamental Considerations for a Writing Test Construction

Article history:

Received	March 30, 2023
Revised	April 18, 2023
Accepted	April 25, 2023
Published	May 31, 2023

Zahra Khan
Department of Humanities & Social Sciences, Bahria University Karachi Campus - Pakistan
zahrakhan.bukc@bahria.edu.pk
https://orcid.org/0009-0002-8519-6892

Jehanzeb Khan
Balochistan University of Engineering & Technology Khuzdar - Pakistan
jehanzebkhan@buetk.edu.pk
https://orcid.org/0000-0002-4879-3463

Pir Suhail Ahmed Sarhandi
Aror University of Art, Architecture, Design & Heritage, Sukkur - Pakistan
ssarhandi@yahoo.com
https://orcid.org/0000-0003-2145-0329

How to Cite:

Khan, Z., Khan, J., & Sarhandi, P. S. A. (2023). Fundamental considerations for a writing test construction. *Academy of Education and Social Sciences Review, 3*(2), 147-153. https://doi.org/10.48112/aessr.v3i2.484

Publisher's Note:
International Research and Publishing Academy (*i*RAPA) stands neutral with regard to jurisdictional claims in the published maps and institutional affiliations.

Abstract

Writing is challenging in second language and its test construction is rather problematic. The present study is aimed to offer an overview of the fundamental considerations in the test construction in English language teaching. The significance of test validity, test reliability, task interactiveness, test attentiveness, test impact, and test practicality have been employed to improve educational practices. The researcher has considered the validity and reliability as severe concerns for the issues in question and the review has been advanced to add to the understanding of the former's value in designing a writing test. Desktop reach formed the base of this research. The search confirms that the aforementioned characteristics are central to the design. This confirms with the research findings available in the literature of teaching and testing.

Keywords: ethical standards, test fairness, test reliability, test validity, writing test construction

INTRODUCTION

Writing tasks completion difficulty seems to have increased the challenges of a good writing test construction. For this reason, a quality writing test is necessary for the provision of assistance to the learners, their placement in learning courses, measurement of the test-takers' progress, problem analyses, suggestions, and evaluation of the course effectiveness. To do so, it is vital to predict a good relationship of the teachers to the design of a good writing test in/through teaching (Kunkun, 2015). The relevant literature addresses it in the context of test validity. There is a need for the prime consideration of professionalism in terms of constructing the standards of ethical conduct that are grounded on the usage of valid tests. Hence, without validity, a test is not considered ethical (Bachman, 2000).

Test validity has a considerable impact on making human judgments about writing tests (MacIver, et al., 2014). According to Lee and & Anderson (2007), the necessary feature is the identification of how and to what extent the writer interacts with the potential factors and their impact on measuring performance interactively or independently in terms of making such judgments of the tests. Additionally, the quality of the test items is also crucial for the validity. Major validity issues include: (a) test content relevance in terms of test purpose; (b) test quality including the phrasing of test items; and (c) test representativeness with respect to the appropriate usage of the target language (Riege, 2003).

For language teachers who want to develop the test items' writing techniques, it is important to ensure the test validity. The instructors should have knowledge of the test use and test construct in order to increase the validity considerations of a writing test (Rossi & Brunfaut, 2019). Crocker (2003) recommends teaching test validity considerations for fairness and moral actions. Furthermore, more attention is needed in the following four areas that include (a) improvement in the validation method involving the judgment of the test content; (b) the scope of fairness definition with respect to an achievement test; (c); and (d) the development of guiding principles of validation studies in terms of test use and its consequence and preparation of the instructors for assessment criteria and their validation. It is now well-established to gauge the evaluation of test qualities to maintain a balance of the test-specific situation (Kunkun, 2015).

The study of O'Neill, at al., (2005) may be used to focus the issue that the writing test influences the instructors' teaching style and content display. According to Wiggins (1993), good teaching tests have been used to improve learning and ensure test consideration standards for the validity and reliability. Moreover, previous research studies conclude that the instructors should modify their curriculum to help reflect the test form and content for the validity considerations. Test standardizations in addition, signify the assessment and its good testing impacts on teaching positively (Grant, 2000; Smith, 1991).

The research objective of the study was to consider how the fundamentals in a writing test construction are documented. The main reason to select this area was to find a way to deepening the understanding of the factors that impact on the validation of writing tests. The present study may be considered as an attempt to fill in the theoretical gap that the research literature has guided to and provided the answer related to the issues of validity considerations in the test design.

METHODOLOGY

The researchers have frequently used the following procedures for literature survey, data evaluation, data analysis, and research finding presentation:

- Comprehensive search of the published literature to examine the relevant studies
- Titles and abstracts of all articles scanned and potentially relevant studies included
- Studies excluded if the focus was found beyond the objectives
- Relevant studies analyzed and included in the review

WRITING TEST CONSIDERATIONS

The purpose of this review article is to provide an overview of various scientific methods or techniques in terms of validity considerations in a writing test design. Theoretical constructs or foundations and practicalities of test item design have not been extensively researched though (Rossi & Brunfaut, 2019). The construct of test validity remains challenged in the educational domain for assessment purposes; therefore, theoretical foundations are required for writing test design (Weideman, 2019). For this, a design accuracy is crucial for the test validity.

Test Validity

Test validity is defined as the quality having an impact on the writing assessment (Kunkun, 2015). Hughes (2003) explains the validity if "It measures accurately what is intended to measure'' (p.26). In order to evaluate the research quality, it is important that the utilization of findings is to be with great care. The validity includes accuracy in terms of precisely reflecting the findings from the data (Noble & Smith, 2015). Moreover, different validity aspects including content and criterion are the provision supporting the test interpretations. It is important to understand the need for the new perspectives of the validity named user validity that emphasizes on interpretation of validity. This validity is made by test users depending upon the availability of information as it also provides result appropriateness of test outputs. Hence, the user validity construct is an important provision for test users and designers to research the practicality of test use. Unfortunately, the information about structural format of the validity and reliability in the field of research and education is inadequate. Another dimension of the validity includes content and face validity for subject experts. In designing a writing test, construct validity is achieved by means of item analysis, item discrimination, and key check (Considine, et al. 2005).

According to Kunkun (2015), the validity leading to a fair method of test writing consists of numerous elements. Each item provides different views to gather and interpret data. Its face validity in writing tests appears properly designed for a good visual reception. Moreover, content validity draws attention to the topics that test designers have discussed after the needs analysis at the learner's level and a check upon the test content area. As far as construct validity is concerned, it measures the quality of test tasks. If a test has a low construct validity, the test is reportedly found less reliable. Another aspect of the validity is the criterion-related validity, also known as external validity, in which the test takers' scores are compared with some external criterion measures. Alderson, et al., (1995) explain that other types of the validity including response validity (gathering data in order to identify the interest of the test takers /learners in terms of evaluating the behavior and opinion on test taking), concurrent validity (students marks co-relate with the instructor marking and also the learners' scores on test co-relates with other tests), and predictive validity (learners scores on test co-relates with other tests that have been taken later).

A research study conducted by Weideman (2019) has defined the validity to enable language testers claiming a test yields as effective. Moreover, it is important to note that a language test designer seeks to measure the language ability consistently or use claims or hypotheses of the test when it is validated. Thus, without a defendable and comprehensible construct validity, the interpretations of the scores of a language test lacks meaningfulness. A study's conclusion becomes more significant when the validity and reliability are taken together.

Test Reliability

Writing task assessment is considered as reliable if the test consistently measures the same learners writing tests at different time slots and assesses the same test by varied raters. Moreover, if the test scores are inconsistent, they lack competence of assessment measure. However, it is also realized that the removal of the inconsistencies caused by numerous factors of time, text type, learners' instructions, and previous knowledge of the learners is almost impossible. Test reliability is maximized by means of the construct of reliability factors as mentioned above (Kunkun, 2015). Test reliability, test validity, and test impact have been investigated to assess test design performance.

Test reliability demonstrates the operationalized measures of the research inquest that other researchers have repeated with similar and consistent findings of the research. This definition is similar to Hughes (2003) and Noble and Smith (2015), who state that test reliability requires consistency of analytical procedures and test scores that are employed for research. Moreover, it has also been reported that establishing a higher level of reliability and validity shows the significant and appropriate data collection methods and use of results in terms of decision-making or conclusions (Noble & Smith, 2015). Moreover, other reliability considerations in the writing test involve external evidence, calculations of coefficient and correlation reliability, test stability, consistency of test items and scores, and inter and intra-rater reliability (Alderson, et al., 1995; Hughes, 2003; Lissitz & Samuelsen, 2007; Kunkun, 2015). Table 1 outlines the establishment of the marking criteria of reliability.

Table 1

Reliability in Writing Test

Establishing the criteria of the reliability in a writing test	
Intra-rater reliability	Refers to sharing the same scripts in terms of having the same marks in different instances. If some marks are different, it is also reliable in a sense. It is measured through analysis of variance or through correlation coefficient.
Inter-rater reliability	Refers to the different examiners in terms of marking similarity without any influence with other examiners or an agreement among two raters. Measuring criteria is also measured by means of analysis of variance and correlation coefficient.
Routine double-marking	Each script is marked by different examiners and marked independently. The candidate receives the means of marks given by two different instructors or examiners.
Normality Test	This test is used to check the data in control and if experimental group is normally distributed.
Homogeneity Test	It is used to check the homogeneity of the data.
Hypothetical Test	This test is used to improve the significance level of the treatment group.

Adapted from Alderson, et al., (1995).

Task Interactiveness

The interactiveness entails that the test takers use more personal resources. Task interactiveness is described as how the test task involves the test takers linguistic knowledge and meta-cognitive strategies (Kunkun, 2015). Moreover, higher degree of interactive tasks necessitates test takers to demonstrate the linguistic knowledge as well as strategic knowledge. It is important to ensure that writing tasks must determine the learners' needs and their level. Test tasks also develop learner's motivation of/about interactivity and authenticity (Weigle, 2002). It is also important to establish the criteria of the validity and practicality of a test. In addition, there is a need to acknowledge the data collected in a valid and an effective manner (Long & Johnson, 2000). Hence, developing suitable assessment criteria or methods is the utmost need for writing as there is a strong relationship between assessment of test and learning (Baartman, et al., 2006).

Test Authenticity

Test of English as a Foreign Language (TOEFL) is projected for academic purposes. The tasks in this test are aimed to check the significant features of academic English and are considered to be authentic. However, Weigle (2002) has opined that the authenticity of TOEFL tasks is limited in a sense as it lacks prewriting of reading opportunities or discussion based on an assigned topic. That is, when authenticity is higher in a writing test, it is helpful for assessment purposes in terms of bridging the knowledge gap between what the learner's challenges in this world are and how they are tested. For instance, TOEFL tests are authentic because of their association with the real world and global practices (Kunkun, 2015).

Writing includes a plethora of set skills and demands different text types, purposes, and audiences that

writing tests consider. It also requires the authenticity and practicality in designing rubrics for different text types. Aligning the grading quality criteria helps improve the assessment criteria (Humphry & Heldsinger, 2014). According to Moss (1994), the researchers who have conducted research on validity, stress the importance of authenticity in assessing performance. Charney's (1984) study also reports that the instructors, researchers, administrator, and testing agencies need to have test validity and test reliability in method and authenticity for the assessment of writing performances. In fact, research literature has revealed the demand for the authenticity and set a criterion for writing tests that engage the learners with problems or questions to think in order to work effectively.

Test Impact

According to Bachman (1990), in terms of the impact, the tests have an intended purpose as per the education system or the demands of the society. Moreover, it affects potentially the test-takers' test perceptions and performance (Bachman, 2000). When a quality test is designed, it has an impact. This should lie with the testing procedures and clarity for human cooperation. Importantly, the assessors while designing a writing test must think about its impact on the learners and instructors. Additionally, in order to see the test tasks impact, the assessors analyze how the learners are engaged and then make comparison of different tests on the particular subject. The ethics of respect is mandatory to see the connection of technical practices regarding the question's merit and its probable negative impact on the learners' performance, level of motivation, and trust. With respect to large scale testing, if the test items lack quality and impact, they are rejected during the review process or revised several times before the test is approved and imlemented. It is important to provide training to the assessors in order to achieve quality impact of the test tasks (Rossi & Brunfaut, 2019).

Test Practicality

Another significant aspect of validity consideration in writing a test construction is test practicality, requiring the available resources to develop, administer, and score or mark the test. In terms of having a finite resource for test purposes, it is needed to optimize the resource allocation to meet the desired level of writing considerations and test usefulness. For instance, if the inter-rater reliability is not high, one should explore several options including an increase in the writing task numbers and raters for the assessment, training raters, and devoting more time to bring refinement in the rating scale (Weigle, 2002). Importantly, the practicality of a good test is also confirmed in terms of evidence that substantiates the result. Fowler's (2008) study reflects that in this age of test propagation, the focus is on test validity and practicality. The educators need to be concerned with the practicality issues when they select the measurement tools. For instance, even if a test has a high degree of test reliability, it may fail to indicate that a test is practical and meaningful. There is a need that scores must be tied with other scoring variables occurring in the network data. In this way, it benefits professional education and reputation.

CONCLUSION & RECOMMENDATIONS

This study sets out to provide an overview of the issues related to test validity construction for writing tests. The literature review research has shown that a proper consideration of test validity, test reliability, task interactiveness, test authenticity, test impact, and test practicality are essential for designing meaningful tests. The result of this review also shows that validity and reliability are the key factors for test consideration. Moreover, the classroom discourse lends opportunities to learn and ensures the scoring criteria of writing. In addition, test authenticity and test practicality are also important for rubrics design. Furthermore, research has also shown that the assessment of writing performances is considered central to a teacher's job. Enabling the teachers for good test takers and test designers is helpful to make teaching and learning effective. The findings of this study suggest that test tasks must include a higher degree of interactive skills for demonstration of linguistics and strategic knowledge. Moreover, test tasks must follow needs analysis and level of the learners that develop their motivation. The theoretical findings also highlight the importance of face validity, content validity, construct validity, response validity, criterion-related validity, predictive validity, simply internal validity, and external validity. Therefore, accuracy is an important consideration for a writing test design.

Whenever the learners' writing performance is assessed, the assessors' operational framework for the assessment needs to be clear for a true measurement. The test must be aligned with the course contents, objectives, and different activities so that the construct validity in writing the test is easily achievable. Another significant finding emerging from this study is the reliability considerations of inter-rater reliability, intra-rater reliability, normality test, homogeneity test, hypothetical test, and routine double-marking in the test construct. The reliability also requires test scores consistency in terms of analytical procedures. Taken together, these

findings suggest that the validity and reliability are the key factors for writing tests. To add to the point, the assessment performance is also important in the teaching domain in order to enable the instructors to design a writing test effectively by following ethical considerations of the validity. Hence, the present study seems to have been pushed to enhance the understanding of validity consideration in designing or constructing a writing test. Considerably, more research will need to be done to review other validity considerations of writing tests. These inquiries may help the instructors teach and test effectively.

Competing Interest

The authors have declared no competing interest.

References

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge University Press.

Baartman, L. K., Bastiaens, T. J., Kirschner, P. A., & Vander Vleuten, C. P. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation, 32*(2), 153-170. **https://doi.org/10.1016/j.stueduc.2006.04.006**

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford University Press.

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17*(1), 1-42. **https://doi.org/10.1191/026553200675041464**

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English,* 65-81.

Considine, J., Botti, M., & Thomas, S. (2005). Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian, 12*(1), 19-24. **https://doi.org/10.1016/s1322-7696(08)60478-3**

Crocker, L. (2003). Teaching for the test: Validity, fairness, and moral action. *Educational Measurement: Issues and Practice, 22*(3), 5-11. **https://doi.org/10.1111/j.1745-3992.2003.tb00132.x**

Fowler, R. C. (2008). Validity, test. *Encyclopedia of Special Education,* 2087-2089.

Grant, S. G. (2000). Teachers and tests changes in the New York state testing program. *Education Policy Analysis Archives, 8,* 14-14. **https://doi.org/10.14507/epaa.v8n14.2000**

Hughes, A. (2003). *Testing for language teachers.* Cambridge University Press.

Humphry, S. M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher, 43*(5), 253-263. **https://doi.org/10.3102/0013189x14542154**

Kunkun, L. U. O. (2015). Validity considerations in designing a writing test. *Studies in Literature and Language, 10*(5), 19-21.

Lee, H. K., & Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing, 24*(3), 307-330. **https://doi.org/10.1177/0265532207077200**

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*(8), 437-448. **https://doi.org/10.3102/0013189x07311286**

Long, T., & Johnson, M. (2000). Rigour, reliability, and validity in qualitative research. *Clinical Effectiveness in Nursing, 4*(1), 30-37. **https://doi.org/10.1054/cein.2000.0106**

MacIver, R., Anderson, N., Costa, A. C., & Evers, A. (2014). Validity of Interpretation: A user validity perspective beyond the test score. *International Journal of Selection and Assessment, 22*(2), 149-164.

https://doi.org/10.1111/ijsa.12065

Moss, P. A. (1994). Validity in high stakes writing assessment: Problems and possibilities. *Assessing Writing, 1*(1), 109-128.

https://doi.org/10.1016/1075-2935(94)90007-8

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-based Nursing, 18*(2), 34-35.

https://doi.org/10.1136/eb-2015-102054

O'Neill, P., Murphy, S., Huot, B., & Williamson, M. (2005). What teachers say about different kinds of mandated state writing tests. *Journal of Writing Assessment, 2*(2), 81-108.

Riege, A. M. (2003). Validity and reliability tests in case study research: a literature review with "hands-on" applications for each research phase. *Qualitative Market Research: An International Journal, 6*(2),75-86.

https://doi.org/10.1108/13522750310470055

Rossi, O., & Brunfaut, T. (2019). Test item writers. *The TESOL Encyclopedia of English Language Teaching,* 1–7.

https://doi.org/10.1002/9781118784235.eelt0981

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

https://doi.org/10.3102/0013189x020005008

Weideman, A. (2019). Degrees of adequacy: The disclosure of levels of validity in language assessment. *Koers, 84*(1), 1-15.

https://doi.org/10.19108/koers.84.1.2451

Weigle, S.C. (2002). *Assessing writing.* Cambridge, UK: Cambridge University Press.

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* Jossey-Bass.