CSCanada
Studies in Literature and Language
Vol. 10, No. 5, 2015, pp. 19-21
DOI: 10.3968/6957

ISSN 1923-1555[Print]
ISSN 1923-1563[Online]
www.cscanada.net
www.cscanada.org

Validity Considerations in Designing a Writing Test

LUO Kunkun[a],*

[a]Associate Professor, School of Foreign Languages, Henan University of Technology, Zhengzhou, China.
*Corresponding author.

Received 24 January 2015; accepted 20 April 2015
Published online 26 May 2015

Abstract
It is essential to give proper consideration to reliability, validity, authenticity, interactiveness, impact and practicality in designing meaningful tests. This article attempts to provide an overview of issues relevant to validity considerations in designing a writing test. Clearly reliability and validity are both extremely important test considerations. Performance assessment is a crucial aspect of the writing teacher's job. Designing a writing test enables teachers to be good test users.

Key words: Reliability; Validity; Test considerations

Luo, K. K. (2015). Validity Considerations in Designing a Writing Test. *Studies in Literature and Language, 10*(5), 19-21. Available from: http://www.cscanada.net/index.php/sll/article/view/6957
DOI: http://dx.doi.org/10.3968/6957

INTRODUCTION
Learning how to write is one of the most challenging aspects of second language learning. Considering how to design a writing test is one of the crucial aspects of teaching. Test validity is a critical consideration when designing a test. The challenge is determining the domain of skills and competencies that should be tested.

A good writing test can provide data which could be used to provide assistance to students, place students in appropriate courses, certify proficiency, measure test taker progress, find problems, suggest solutions, and evaluate course effectiveness. The desirable qualities of tests are proposed by many language testing researchers. The one adopted in this project is from Bachman & Palmer. Bachman & Palmer (1996, pp.17-18) propose a "model of test usefulness" based on six qualities of tests: reliability, construct validity, authenticity, interactiveness, impact and practicality. It is impossible to maximize overall test usefulness. It is, however, probably to evaluate combined test qualities and strike a balance between the qualities from each specific testing situation.

1. TEST CONSIDERATIONS

1.1 Reliability
A writing assessment task is believed reliable if it measures consistently both in terms of the same student on two or more separate occasions and the same task by different raters. Hughes notes, "The more similar the scores would have been, the more reliable the test is said to be" (1989, p.29). If test scores are relatively inconsistent, they provide no information about the competence we want to measure. We realize that it is impossible to eliminate inconsistencies entirely because many factors such as instructions given to students, conditions under which tests are taken, genre, time of day, previous experience and so on can influence a writer's performance and test reliability. Reliability can be maximized, however, by standardizing as many of these factors above as possible. Hughes argues, "reliability of performance can be achieved by taking sufficient samples, restricting the candidate's choice of topics and genres, giving clear task directions, and ensuring students are familiar with the assessment format" (1989, p.28).

Another element of reliability is the consistency with student writing which is rated, for writing assessments involves subjective judgments. All assessors, it is required, should agree on the rating of the same student performance. Meanwhile, the same performance should be assessed in the similar way on different occasions.

One of the methods of estimating reliability involves calculating a reliability coefficient—a correlation coefficient. "Reliability coefficients can range between zero and one, with a higher coefficient indicating greater reliability" (Hughes,1989, p.31). Hughes suggests "if appropriate steps are taken, speaking and writing tests can have reliability coefficients as high as 0.9 " (1989, p.87).

1.2 Validity

The quality that most affects the value of a writing assessment is validity. In defining validity, Hughes states, "a test is said to be valid if it measures accurately what it is intended to measure" (1989, p.22). Validity is considered to be the central quality for meaningful and fair writing test, which means that a writing assessment task must judge what it claims to assess and what has been taught. For example, it is invalid to give a writing test that does not ask students to write, enables students to write in a genre that has not been studied, or requires professional knowledge that has not been had. There are several types of validity, each of which offers a slightly different point of view on gathering and interpreting data.

Face validity is when a writing test appears valid by test takers and other untrained observers. It is also concerned with whether or not a writing test looks like a proper test in the eyes of the teachers and the students. This indicates that an assessment ought to surround an actual writing sample and have a connection with students' writing needs.

Content validity draws on evidence of the topics that writers are supposed to discuss in the target domain based on a thorough needs analysis and concerns whether the test adequately and representatively measures the content area. This is closely linked to the issue of direct versus indirect testing. The writing task setting, task demands, and the test setting and administration should be taken into account to achieve content validity (Weir, 2005, pp.56-84). One of the effective writing assessments based on face and content validities should offer opportunities for writing that are as much like the real competence as they can by manifesting the authenticity of target contexts. For instance, since new TOEFL IBT is introduced, the gist of the changes in writing: Read, listen, and then write in response to a question verifies that the new TOEFL is much more reflective of how English is really used and more associated with real life with high authenticity.

Construct validity concerns the qualities that the task measures. It simply demonstrates a suitable relationship between what we are testing and what we wish to assess. Namely, it shows what ability the task is attempting to measure and the domain of writing is seeking to test. Construct ability is seen to be superior to all other types of validity. Hughes states, "gross constructs such as reading ability and writing ability as well as more specific ones such as control of punctuation and sensitivity to demands on style" (1989, p.26). In second language writing classes, teachers manage to measure abstract constructs such as "writing ability" or "progress". Therefore, assessment tasks must produce writing that involves these abstract concepts. Construct validity does not cause a problem with direct testing, but severe difficulties may arise with indirect testing. The previous TOEFL writing test was multiple-choice question. It is clearly impossible for writing to be tested directly. This sort of writing theory reveals that numbers of sub-abilities are involved in writing such as accuracy of spelling, control of punctuation, grammatical accuracy and so on. In this way, the test would only have construct validity if the test was actually testing whether a taker could write well or not. While reliability is a necessary condition for validity (Bachman, 1990, p.160), Hughes (1989, p.42) and Bachman and Palmer (1996, p.23) find there is a tension between them. As a measure, a highly reliable writing test could lead to the use of a multiple choice error recognition test . The test, it seems, has less construct validity than a less reliable but more direct test of writing ability.

Criterion-related validity (Alderson et al., 1995, p.171) relates to how test takers' scores compare with an external criterion. "The first aspect of criterion-related validity is concurrent validity, which is the extent to which the results of the test in question agree with another independent, highly dependable second assessment method" (Hughes, 1989, p.23). The second aspect, predictive validity, considers how well a test is able to predict a future result (Ibid., p.25). It simply means whether the test results are closely related to match those from other writing tests.

1.3 Authenticity

The high degree of authenticity could be favorable in gaining the intended consequences of assessment by bridging the gap between what the students confront in the real world and the way they are tested. For instance, since new TOEFL IBT is introduced, the gist of the changes in writing: read, listen, and then write in response to a question verifies that the new TOEFL is much more reflective of how English is really used and more associated with real life in the high degree of authenticity.

1.4 Interactiveness

The higher degree of interactiveness requires more personal resources test takers use. The interaction between the test taker and the task can be described as how a test task engages the test-taker's language knowledge, metacognitive strategies. Highly interactive tasks require test takers not only to demonstrate their linguistic knowledge but also their strategic competence (Weigle, 2002, p.53).

1.5 Impact

Impact relates to the fact that tests "are virtually always intended to serve the needs of an educational system or of society at large" (Bachman, 1990, p.279). It potentially affects test takers' perceptions of the test and their

performance (Bachman, 2000). IELTS or TOEFL as a placement test, not a proficient test is used to assess the language level of students who do not have an appropriate qualification, which affects individuals, society and education system.

1.6 Practicality

Practicality relates to the implementation of the test and whether it will be developed and used. Bachman and Palmer (1996) classified the addressed resources into three types: human resources, material resources, and time. It is clearly pointless to design a well-principled writing test if resources to administer the writing test are lacking.

2. DESIGNING WRITING ASSESSMENT TASKS

It is essential to give proper consideration to reliability, validity, authenticity, interactiveness, impact and practicality in designing meaningful tests. In the classroom, giving students opportunities to display what they have mastered and ensuring that writing scoring is rated appropriately is of significance.

Further to the above research, designing writing assessment tasks is supposed to contain three basic elements. Rubric and prompt, and input material is crucial to successful writing assessment tasks.

Rubrics as a basis are performance-based assessments which assess student performance on any given task and use specific criteria for evaluating student performances. A good rubric describes levels of quality for each of the criteria. Writing rubrics are developed to assist teachers to conduct consistent and fairly formal assessments of learner writing proficiency. Good writing assessment tasks involve several key factors: Relating to what has been done in class; reflecting "authentic" communicative real world tasks that might be carried out in a particular environment, such as in the community, in school; having clear instructions for the learners; having clear performance conditions and indicators; providing the context in which the writer is to carry out the task in complete and logical details, such as the location of the task, the role of the writer. Writing rubrics may include the specification of the objective, the procedures for responding, the task format, the time for submission and the evaluation criteria (Douglas, 2000, p.50). Therefore, rubrics are supposed to make as comprehensive as possible, since the writer's performance in the task may be influenced considerably. The writing task must serve as a prompt to stimulate students' background knowledge and personal experiences. Kroll and Reid (1994) propose there are three basic types of writing prompts. A bare prompt is simple and direct. A framed prompt presents a set of circumstances. A reading-based prompt provides a text and the writer is asked to summarize, explain or interpret it. As a result, the higher degree of interactiveness will be demonstrated.

CONCLUSION

In this paper I have provided an overview of issues relevant to validity considerations in designing a writing test. Clearly reliability and validity are both extremely important test considerations. Performance assessment is a crucial aspect of the writing teacher's job. A writing test plays an important role in determining the success of the writing experience. Enabling teachers to be good test users and then good test designers so as to make teaching more effective and help students progress should not be neglected.

REFERENCES

Alderson, J. C., et al. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

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

Bachman, L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing,* (17).

Bachman, L., & Palmer, A. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats and cautions. *Journal of Second Language Writing, 13*(3).

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

Weir, C. J. (2005). *Language testing and validation—An evidence-based approach*. Basingstoke: Palgrave Macmillan.