

FAIRNESS AND ETHICAL STANDARDS IN LANGUAGE TESTING

*Students of DIEP:
Normurodova Nigora
Sayfiddionova Nodira
Eshqobilova Shamsiqamar*

ABSTRACT

The submitted paper outlines the impact of political decisions on language testing, predominantly on test scores of the tested population, test design and item-writing. Results from research referring to the comparison of test-takers' performances in the years 2011-2013 reveal a necessity to be sensitive and aware of ethics and fairness in language testing. The research is focused on the professional decision of test developers, item writers and administrators to avoid negative impact and maximize positive washback. The paper focuses on the construct which underlies the B2 English test to provide information about abilities that the test is designed to measure. Testing grammar and vocabulary was supported by using multiple linear regression analysis. The new conditions for test takers in 2012 are analyzed and evaluated in three areas the number of the students applying for tests in English B2, the total scores of students and their grouped distributions and the scores achieved in productive skills and language in use. We conclude by examining the ethical implications of the new conditions for test takers and by demonstrating the effect the ethical dimension had on item writers in their design of subsequent tests. Fairness plays an important yet hard to implement role in test design, and it is often up to the administrators and test designers to cope with external obstacles in promoting this.

KEY WORDS: Fairness, Ethical Standards, Validity, Reliability, Test Bias, Transparency, Accessibility, Confidentiality, Responsible Score Use, Professionalism

INTRODUCTION

Language testing has become an essential component of modern education, serving multiple purposes such as placement, progress monitoring, achievement measurement, and certification. In many educational systems, high-stakes language tests determine students' academic pathways, scholarship opportunities, employment prospects, and even migration eligibility. Because of these significant consequences, language assessments must be held to the highest standards of fairness and ethical practice. When tests lack fairness, they fail to measure true ability; when they lack ethical integrity, they risk violating students' rights and creating unequal opportunities. Therefore, educators, test developers, and institutions have a shared responsibility to design, administer, and interpret language tests in a manner that is transparent,

unbiased, and professionally accountable. This article provides an in-depth discussion of fairness and ethical standards in language testing by examining key principles such as validity, reliability, accessibility, transparency, responsible score use, and examiner professionalism. It highlights why these principles matter and how they contribute to trustworthy and meaningful assessment. Language tests are more than academic tools; they are instruments that influence real-life decisions. Therefore, maintaining fairness and ethical responsibility is not just desirable—it is necessary. Fair and ethical testing ensures that every learner is evaluated based on their actual language competence, not background factors such as culture, gender, or socioeconomic status.

1. Fairness in Language Testing

Fairness refers to the extent to which a test gives all individuals equal opportunities to perform well. A fair test reflects learners' abilities accurately and ensures that external factors do not influence performance.

a. Validity

Validity is the foundation of a fair assessment. A test must measure the intended construct—for example, a speaking test must assess speaking ability, not writing skills, personal background, or test anxiety.

There are several types of validity:

Content validity – the test content matches the skills students are expected to learn.

Construct validity – the test genuinely measures the theoretical construct (such as grammar knowledge or listening comprehension).

Face validity – the test appears fair and appropriate to teachers and learners.

Tests with weak validity can lead to unfair outcomes and incorrect decisions.

b. Reliability

Reliability ensures consistency. A reliable test gives similar results under similar conditions. This includes:

Consistent scoring across different examiners

Consistent difficulty across test forms

Consistent test conditions for all learners

If test results differ due to chance or inconsistency, the test becomes unfair.

c. Accessibility

Fair tests must be accessible to all learners, including:

Students with disabilities

Students with different learning needs

Students from diverse linguistic backgrounds

Providing accommodations such as extra time, large-print materials, or quiet testing rooms is essential. Accessibility ensures that learners are tested on their ability, not on their limitations.

d. Absence of Bias

Bias occurs when test items favor certain groups over others. Common types of bias include:

Cultural bias: requiring knowledge only some groups have

Linguistic bias: using unfamiliar vocabulary not related to the test skill

Gender bias: topics that benefit one gender over another Bias reduces test fairness and must be eliminated during test design.

Ethical Issues in Language Testing

Ethical considerations in language testing permeate various aspects of the testing process, including development of teaching and examination projects, test item development, exam administration, score reporting, statistical analysis of results, and feedback.

Development of Teaching and Examination Syllabus

An English syllabus is a document that sets out the objectives, content, methods and assessment criteria for English teaching. It aims to ensure the effectiveness and quality of education and teaching and is the central document of the English language education system in schools. An English syllabus usually involves teaching objectives based on improving students' language proficiency and practical needs, including skills in listening, speaking, reading and writing. It should also focus on developing students' intercultural communication and thinking skills. In addition, the syllabus specifies the teaching contents as well as the allocation of hours for lectures and internships, experiments and assignments. According to the teaching plan, the document in the form of an outline specifies the teaching content of a course. In short, English syllabus is an important document to guide English teaching, which provides teachers with clear teaching objectives and directions and helps to improve the quality of English teaching. And the English examination syllabus is a document which specifies the objectives, contents, methods and assessment criteria of the English examination. It is formulated on the basis of the English syllabus and other relevant standards, and is designed to ensure the scientific and fair nature of the English language examination. It specifies the scope to be covered by the examination, the format of the examination, the difficulty of the examination, the duration of the examination, the types of questions and the marks. The English test syllabus is the basis for proposing and evaluating questions, and it is also an important reference document for candidates to prepare for the test. There is a close connection and mutual influence between the English teaching syllabus and the English examination syllabus, and they should be adapted and coordinated with each other to ensure the effectiveness and quality of English teaching.

The makers of English teaching syllabus and examination syllabus are usually the education departments and relevant academic institutions. In China, education departments and related academic institutions, such as the Ministry of Education, the

Higher Education Press, and the Foreign Language Teaching and Research Press, will formulate corresponding English syllabi and examination syllabi according to the national education policy and the characteristics of the disciplines, and in the light of the needs of different levels and types of English teaching.

However, due to the different regions, environments and modes of thinking of the syllabus developers, their understanding of the syllabus and examination syllabus will be more or less individual, and the preparation of the syllabus will inevitably be affected by the subjective judgment of the developers to a certain extent, which will lead to a series of ethical problems in testing. In view of this, the author suggests focusing on the diversified construction of the test-making body in the process of formulating the teaching and examination syllabus, which can include the authority of language testing, educationalists and excellent teachers with rich teaching experience, so that they can speak freely in the seminar to form a hundred schools of thought, so as to ensure the objectivity and fairness of teaching and examination syllabus. In addition, we can also focus on the backwash effect of language testing, constantly test the scientific use of the teaching and examination syllabi in daily teaching practice, and use this as the basis for constant revision and improvement of the teaching and examination syllabus.

Test Item Development

The first stage of testing involves test item development, where test creators design questions based on syllabi and test specifications. Guided by the first principle of the “Code of Ethics” which emphasizes that language testers should approach test item creation from a professional perspective and respect the needs of each test-taker (International Language Testing Association, 2000), the process of test item development involves ethical considerations related to material selection and item construction.

Material Selection

The ethical aspects of language material selection primarily revolve around the choice of test content. Test creators should maintain fairness in their thinking and not impose their own biases on test-takers. When selecting material, attention should first be given to whether the chosen content might elicit different responses from test-takers of various cultural and religious backgrounds (Xu, 2001). Additionally, the distribution of genres in test items should be considered carefully. A lack of variety in genres or inadequate coverage can lead to unfairness, as some test-takers may excel in certain genres while others may struggle. To avoid ethical issues in test materials, item creators should prioritize fairness and familiarity in material selection and ensure a balanced distribution of genres and topics in questions.

Question Confidentiality

In recent years, incidents of exam question leaks have become increasingly prevalent in the media. These leaks have exposed vulnerabilities in the process of maintaining question confidentiality. It is widely recognized that question confidentiality directly impacts the interests of test-takers. Question leaks seriously violate the rights of numerous candidates and undermine the validity of exams. The “Code of Practice” stipulates that test papers should be securely stored to ensure that no test-taker gains an advantage (International Language Testing Association, 2007).

Fairness during the Exam

Fairness during the exam implies that all test-takers participate under equal and consistent conditions (Zhang, 2013). Conditions consistency refers to external factors affecting the exam, such as exam timing, location, and equipment. For example, in a foreign language listening test, the placement of the recording equipment or speakers in the test room—whether central, left, or right—can have a significant impact on the scores of test-takers situated differently. Currently, we cannot precisely calculate the influence of these external differences on exam results. Therefore, during the exam administration phase, every effort should be made to consider all possibilities and minimize the impact of external factors on exam results.

Fairness in Proctoring

Proctors must strictly follow the exam design requirements, distributing and collecting test papers at the specified times. Overall, all proctors must be well-versed in the proctoring guidelines to avoid being too lenient or too strict. Organizing Scoring and Scoring

Scoring is a crucial phase in determining the reliability of an examination. If the scoring process is poorly executed, the reliability of the examination is compromised. Scoring involves both subjective and objective components, with large-scale examinations in China currently relying heavily on machine-based reading and scoring for objective questions. Therefore, our primary focus here is on the ethical issues related to subjective question scoring.

Scoring of subjective questions involves three aspects: 1) inter-rater consistency, 2) intra-rater consistency, and 3) ethical conduct of the raters (Li, 2001). If these three aspects are maintained consistently, the examination’s reliability can be ensured. Conversely, if they are not consistent, the results cannot accurately represent the actual abilities of the candidates. This irresponsibility is detrimental to the candidates, as well as to educational institutions and employers who rely on these results. To ensure these three aspects of consistency, the following steps should be taken: Establishing Unified Scoring Standards: It is imperative to develop unified scoring standards, as in the case of essay assessments, whether comprehensive or analytic scoring should be used. Rater Training: Raters should undergo rigorous training to ensure they adhere to uniform

scoring principles. They should be well-versed in the testing concepts, question types, and scoring guidelines.

Result Analysis

The Code of Practice explicitly stipulates that organizations involved in testing must take necessary steps to ensure accurate calculation of each candidate's scores and incorporate these results into data-based evaluations. In this process, continuous review should ensure that the scoring process progresses as planned. Ethical issues in the statistical analysis of results mainly manifest in the interpretation and utilization of scores. Score Interpretation. Ethical considerations during the scoring process also extend to the interpretation of scores. Presently, many large-scale exams in China do not provide assessment criteria for individual score points or descriptions of language proficiency corresponding to each score point. Instead, score interpretation is left entirely to the discretion of the scorers or employing institutions, resulting in a lack of transparency and credibility. In future exams, testing authorities should establish clear scoring standards and provide guidelines for score interpretation. This transparency is essential to uphold the ethical standards of testing.

Score Utilization

The ethical dimension of testing is ultimately reflected in the utilization of scores. Testing purposes vary and can include diagnostic testing, placement testing, selection testing, recruitment testing, admission testing, preparation testing, and research testing. Different tests have different implications for score utilization. For instance, some tests aim to assess candidates' abilities, while others determine different decisions based on scores. Administrative decision-makers often lack expertise in language testing theory, which can lead to scientifically or ethically unsound decisions. To prevent such situations, test creators must possess professional knowledge and assist score users in making informed and fair decisions based on test scores.

Test Impact on Teaching

The prevalence of test-oriented education has led to the utilitarianization of language testing. Language testing has been employed to serve various societal needs, such as graduation, advancement, certification, and employment. With testing becoming increasingly intertwined with society, exams have become the guiding principle of education. This shift has caused a negative feedback loop in teaching: instead of improving students' overall abilities, teaching has come to focus solely on test-taking skills. Testing has inadvertently undermined the quality of education. Despite efforts to minimize the negative impact, it remains challenging to eliminate entirely. Measures to mitigate the negative impact include: 1) balancing reliability and validity according to the test's purpose, 2) analyzing data and summarizing experiences based on test results, and 3) improving test methods and question types to reduce negative effects (Chen, 2007).

CONCLUSION

Ethical issues in language testing are among the major challenges and concerns in the field of language testing in the 21st century. Currently, international research on testing ethics primarily addresses ethical standards and behavioral norms. These standards are not obligatory, as indicated by the use of the modal verb “shall” in the Code of Ethics. In China, research on testing ethics remains in its infancy, and ethical considerations have not garnered enough attention from language testers. However, as language testing continues to evolve, ethical concerns are expected to gain increasing prominence.

References

1. International Language Testing Association. (2000) ILTA Code of Ethics in English.
https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/ILTA_2018_CodeOfEthics_Engli.pdf
2. Davies, A. (1997). Introduction: The Limits of Ethics in Language Testing. *Language Testing*, 14, 235-241. [Google Scholar] [CrossRef]
3. Davies, A. (2008). Ethics, Professionalism, Rights and Codes. *Encyclopedia of Language and Education*, 7, 429-443. [Google Scholar] [CrossRef]
4. Chorieva Iroda Kurbonboevna. (2024). Aphorisms with Medical Terms in the Works of Uzbek Thinkers: Avicenna, Alisher Navoi and Abu Rayhan Al-Biruni. *Miasto Przyszłości*, 49, 1368–1374.
5. Zhao, H. M. (2001). Reflections on the Modernization of Language Testing. *Journal of Chongqing Institute of Technology*, No. 4, 95-97
6. Spolsky, B. (1995). *Measured Words: The Development of Objective Language Testing*. Oxford University Press.