Ethical Issues in Language Testing from a Validity Perspective

Shuling Gao, Jiayi Liu

School of Foreign Languages, Northwest University, Xi'an, China
Email: Shulinggao1998@163.com

Abstract

Ethical issues in language testing have become one of the major research topics and challenges in the field of language testing in the 21st century. Ethics refers to the normative rules governing relationships and behaviors among individuals in human society, as well as the behavior order among people, society, and nations. Any group or professional behavior that affects society as a whole carries inherent ethical requirements. The exploration of ethical issues in language testing began to gain prominence in the field after Messick expanded the concept of validity in language testing. With the development of language testing, researchers have increasingly focused on the study of testing ethics. In 2000, the International Test Commission (ITC) issued a set of "Ethical Guidelines", where ethical issues were reflected to varying degrees in various principles. In 2007, based on the specific implementation of the "Ethical Guidelines", the ITC issued corresponding "Code of Conduct" to define the responsibilities and obligations of language testers and test takers. In essence, ethical issues in language testing fall within the scope of validity in language testing. This paper, starting from the current status of research on ethical issues in language testing in China, which is still in its infancy in both theory and practice, uses literature review, descriptive research, and qualitative analysis. Within the framework of validity, it discusses the factors involved in ethical issues in language testing: development of teaching and examination syllabus, test item production, exam preparation and administration, organization of scoring and score interpretation, result statistical analysis, feedback, and summarization. Based on the "Ethical Guidelines" and the corresponding "Code of Conduct" issued by the International Test Commission, it summarizes the causes of ethical issues and proposes corresponding solutions for each phase. Finally, some suggestions are made for the development of ethical issues in language testing.

Keywords

Validity, Language Testing, Ethics

1. Introduction

Language testing is the process of measuring a set of language behaviors that represent a sample of the abilities in question. The results of these tests are used to make inferences about the target abilities (Li, 2001). After more than half a century of development, language testing has evolved into an independent academic discipline. Following three distinct stages of development, language testing has now entered the phase of social application.

As a means for further education, employment, and advancement, testing profoundly influences the prospects and destinies of test-takers. Simultaneously, it has immeasurable effects on individuals, families, schools, and society as a whole. Consequently, an increasing number of people are becoming concerned about the fairness of testing. Ethical issues in testing encompass two major dimensions: the quality of test items designed by language testers and the consideration of factors influencing the use and outcomes of the test. Ethical concerns in language testing primarily revolve around test validity. When assessing test validity, two factors must be considered: the interpretation of test results and the use of test results. This article aims to study the ethical issues in language testing within the framework of validity.

2. Review of Ethical Issues in Language Testing Research

International Research Trends: Messick (1989) expanded the concept of validity, asserting that language testing should be understood within a broader societal context. This perspective led to ethical considerations in educational measurement, subsequently becoming a forefront issue in the field of language testing. Spolsky (1995) was among the first to discuss ethical concerns related to test administration methods and their consequences. He identified the third stage of language testing development, characterized as psycholinguistic-sociolinguistic, or postmodern language testing. Subsequently, scholars like Davies (1997) approached the social issues of the language test from a sociolinguistic perspective, emphasizing that the quality and value of language testing must ultimately be evaluated in a social context. Ethical concerns are a critical component of these social issues. In 1994, the Association of Language Testers in Europe (ALTE) introduced the "Code of Practice" and in 2000, the International Language Testing Association (ILTA) (2000) issued the "Code of Ethics", marking significant milestones in the history of ethical issues in language testing. In 2005, ILTA published the "Draft Code of Practice" as specific practical principles to complement the "Code of Ethics", defining the responsibilities and obligations of both language testers and test-takers. This draft, along with its appendices, was ratified at the 2007 Barcelona conference and underwent revisions in 2010. Davies (2008) elaborated on the ethical aspects of testing when exploring the professionalism of language testing. He emphasized that an ethical code of conduct was sufficiently necessary to guarantee the scientific and fair nature of language testing. Ethical codes and codes of conduct for testers helped to regulate and constrain

their duties and behaviors, thus enhancing the scientific and professional standards of language testing, which in turn enabled the test to have a positive and active impact on the social dimension. Bachman was committed to construct a theoretical model for examining the impacts caused by testing. In 1996, Bachman and Palmer included the "impact" of test in the framework of test usefulness and proposed that when considering the use of test, it was necessary to consider its possible impact on the three dimensions of the individual, the education system and the society. It can be seen that foreign research on test ethics is mainly based on the validity framework, which elaborated on the possible impact of test ethics on the various levels of the social system, and was less involved in the localized construction of the test ethics system, which provides a broad space for subsequent research on language test ethics.

Domestic Research Trends: In China, ethical research in language testing has not received sufficient attention from scholars. According to search results from the China National Knowledge Infrastructure (CNKI), only a few individuals, such as Yang Huizhong & Gui Shichun (2007), Mei Ye (2008), Mei Ye and Nie Jianzhong (2009), Xu Shihong (2011), Chen Xiaokou & Li Shaoshan (2013), Gan Lin & Xia Jimei (2016) and Zheng Yan (2018) have explored ethical issues in language testing.

In summary, both domestic and international research on ethical issues in testing are still in their nascent stages, and the research on test ethics from the perspective of validity is even more rare. Ethical research in testing, whether in theory or practice, warrants further exploration.

In view of this objective reality, this paper adopts the method of literature analysis, descriptive statistics and qualitative analysis, and discusses the elements involved in the ethics of language testing under the theoretical framework of validity, including six aspects: teaching, syllabus development, test question production, examination preparation and examination implementation, organization of marking and scoring, result statistical analyses, feedback and summarisation. Based on the Code of Ethics and the corresponding Code of Conduct issued by the International Test Association (ITA), the causes of ethical problems are summarized and corresponding solutions are proposed. Finally, some constructive suggestions are made for the development of ethical issues in language testing.

3. 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.

3.1. 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.

3.2. 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.

3.2.1. 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.

3.2.2. Item Construction

Item construction in test item development is also intertwined with ethical considerations. Test questions consist of stems and options, and clarity of instructions, the absence of ambiguity in stems, and readability are essential. Ambiguous instructions or unclear stems can confuse test-takers and make it challenging for testers to assess responses accurately. To improve the scientific and ethical aspects of item construction, two steps are recommended: rigorous training of item developers and the recognition of the importance of professional ethics during the item creation process.

3.3. Exam Administration

The "Code of Practice" explicitly outlines the obligations of institutions involved in the preparation and implementation of high-stakes exams in its first section, third section (International Language Testing Association, 2007), addressing both the pre-exam and during-exam phases. Ethical issues in exam administration encompass the fairness of preserving question confidentiality, ensuring fairness during the examination, and maintaining fairness in proctoring.

3.3.1. 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).

3.3.2. 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.

3.3.3. 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.

3.4. 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) intro-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.

Ongoing Monitoring and Feedback: Continuous monitoring and feedback are essential to detect scoring issues and address them promptly. Specific issues should be handled on a case-by-case basis.

Finally, during the official scoring phase, the ethical conduct of the raters plays a significant role in subjective question grading. This ethical responsibility is essential to the candidates' fate and constitutes a sacred duty for the raters.

3.5. 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.

3.5.1. 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.

3.5.2. 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.

3.6. Feedback and Conclusion

Feedback and conclusion represent the final steps of an examination. Information obtained from the examination should be relayed to three parties: examination users, candidates (students, schools, and teachers), and test designers and

producers. Furthermore, examinations should have a positive and negative societal impact.

3.6.1. 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).

3.6.2. Test Societal Impact

The societal aspect of language testing cannot be ignored in modern language testing. Generally, the larger the scale of an exam, the greater its societal significance. Therefore, a good test should consider its societal impact. The ninth item of the Code of Ethics mentions the societal impact of testing: "Language testers have a responsibility to assess the moral consequences of their work for the projects for which they are responsible." While they may not foresee all potential consequences, they should comprehensively assess foreseeable outcomes. A test with significant societal influence, if mishandled, could have dire consequences. Therefore, testers must uphold professional ethics, maintain their stance, and promote positive societal effects through testing.

4. Suggestions

The ethical study of language testing is still in its infancy in China, and there is much room for development. In the future, the development of language testing can be started from four aspects: the professionalization of testing, the duties of testers, the legalization of testing, and the diversification of testing methods.

4.1. Professionalization of Testing

4.1.1. Current Situation of Testing

As teaching and testing are inextricably linked, teachers inevitably come into contact with testing in teaching. However, there are very few professional testers or those with some professional knowledge of testing in the existing teaching force (Zhao, 2001). However, some of them will be involved in the preparation, marking and scoring of the test questions, and in some national language tests, the designers involved in the test questions are partly appointed on an ad hoc basis, and the test papers they design will determine the future of the candidates,

which is the current situation of language testing in China. It is not known whether these papers are scientific and whether they could measure the language ability of the candidates, furthermore, in the marking of the tests, many markers do not have a certain knowledge of the tests, which may result in unscientific marking. And lastly, after the tests are over, there are few people to analyze and evaluate the papers in order to improve the design of the questions. All these will inevitably result in biased testing and even seriously affect the validity of the test.

4.1.2. Test Prospects

China's examination bodies are administrative organizations, and administrators find it easy for teachers to go about ordering one or two sets of examination papers per year, while external examiners struggle to enforce a strict post-examination confidentiality system for the examination papers. This has resulted in many public examination questions not being reused, while at the same time the professional labour of many of the teachers who order the questions is not valued. This violation of science is not only detrimental to the recycling of test questions and the quality of test questions, but even makes it difficult to conduct examinations in the long term. Therefore, we suggest establishing an independent English testing organization with high social credibility, realizing the separation of the politics and English test and taking the road of professional English testing. This is the general trend of the development of the testing discipline in the 21st century.

4.2. Duties of Testers

Professional Duties

Language testers bear certain social responsibilities, especially for some national exams such as the college entrance examination and the fourth and sixth grades, the responsibility of language testers is undoubtedly significant. Therefore, language testers must do their job well. According to Article 5 of the Code of Ethics, "It is the tester's duty to continuously enrich and strengthen his/her professionalism and to communicate with each other and with his/her peers or other relevant researchers in language" (International Language Testing Association, 2000). Firstly, continuous learning and enhancing one's knowledge is the basis for maintaining professional skills, otherwise it will be detrimental to the testers. In addition, testers should exchange knowledge with their peers through professional journal publications or conferences, and they should also provide career development planning for testers who are undergoing training and teaching guidance for students majored in language. Only by constantly improving their own professionalism, and at the same time developing a pool of language testers will the language testing system become more complete and better serve the candidates or the disorganization concerned.

4.3. Social Responsibilities

Language testers should fulfill some social duties in addition to improving their professionalism. Article 7 of the Code of Ethics sets out the social duties of lan-

guage testers. These include endeavoring to improve the quality of testing, assessment and teaching, promoting the rational distribution of these services, facilitating language learning and improving language proficiency.

Legalization of Testing

In recent years, the language testing community has focused on the development of ethical standards and codes of conduct for the profession (Mei & Nie, 2009). The International Association for Language Testing has successively issued the Code of Ethics and the Draft Code of Practice, which stipulate the responsibilities and obligations of language testers and test takers. China's research on testing ethics is still at an early stage, and we should draw on some international theories and practical experience to formulate language testing regulations that suit China's national conditions. The duties and obligations of language testers are clearly stipulated, so that language testers have laws to follow. The core of ruling the test according to law lies in applying legal norms to constrain the behaviour of language testers and the management authority of the organizers, so as to safeguard the fairness and authority of the test, and at the same time protect the legitimate rights and interests of the candidates. On the basis of establishing a sound and localized code of ethics for testing that suits China's national conditions and the current situation of testing, and in accordance with national laws and regulations, we need to build a scientific and strict legal system for testing, an enforcement system and a supervision system, and only in this way can the ethical value of China's large-scale English language testing be truly safeguarded.

Language testing is a highly professional job, and due to the increasing social weight of the test, the language testers' own rights and interests and their intellectual property rights are not protected. Thus, it is imperative to develop a set of regulations for language testing. Only in this way can the long-term development of the language testing discipline be ensured, the chaotic domestic data market be cleared up, and the intellectual property rights of testers be protected.

4.4. Diversification of Testing Methods

There are a large number of high-stake social exams in China (such as the college entrance examination), and these high-stake exams typically have the characteristic of determining a person's fate in one test. Many candidates will experience anxiety before and during the large-scale exams that can affect their normal performance. To a certain extent, this kind of one-time test cannot test the real level of students, and the failure of the test will have far-reaching effects on the candidates themselves, their families and the society, thus involving the fairness and ethical issues of language testing. In view of this, the author suggests that testing methods should diversify and develop in multiple ways. On the basis of traditional language testing, we should conduct comprehensive evaluations of students' abilities by incorporating methods such as process evaluation in daily teaching practices, establishing student learning portfolios, paying attention to

students' classroom participation and performance as well as conducting holistic assessment of students' ability from multiple angles and at multiple levels by means of daily teaching observation.

However, as China is a large country with a large population, the traditional language test still plays a pivotal role, and there is still a long way to go to establish a comprehensive test that can be adapted to China's national conditions in the future.

5. 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.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Chen, X. K., & Li, S. S. (2013). Ethics—An Indispensable Dimension in Language Test Validity Research. *Foreign Language Testing and Teaching, No. 3*, 1-7+47.

Chen, Y. F. (2007). Achieving Positive Washback Effects of Language Testing on Language Teaching. *College English (Academic Edition), No. 1*, 226-229.

Davies, A. (1997). Introduction: The Limits of Ethics in Language Testing. *Language Testing, 14*, 235-241. https://doi.org/10.1177/026553229701400301

Davies, A. (2008). Ethics, Professionalism, Rights and Codes. *Encyclopedia of Language and Education, 7*, 429-443. https://doi.org/10.1007/978-0-387-30424-3_191

Gan, L., & Xia, J. M. (2016). Ethics in Language Testing: Review and Implication. *Journal of Guangdong University of Foreign Studies, No. 2*, 58-64.

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

International Language Testing Association. (2007) ILTA Guidelines for Practice in English. https://www.iltaonline.com/resource/resmgr/docs/guidelines_for_practice/2020_revised/ilta_guidelines_for_practice.pdf

Li, X. J. (2001). *The Science and Art of Language Testing*. Hunan Education Press.

Mei, Y. (2008). Ethical Considerations in Language Testing. *Journal of Taiyuan City Vocational and Technical College, No. 2*, 136-137.

Mei, Y., & Nie, J. Z. (2009). A Review of Research on Language Testing Ethics. *Foreign Language World, No. 4*, 91-96.

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher, 18*, 5-11. https://doi.org/10.2307/1175249

Spolsky, B. (1995). *Measured Words: The Development of Objective Language Testing*. Oxford University Press.

Xu, S. H. (2011). Ethical Issues in Large-Scale Language Testing. *Journal of Nantong University (Social Science Edition), No. 4*, 97-103.

Yang, H. Z., & Gui, S. C. (2007). Social Considerations of Language Testing. *Modern Foreign Languages, No. 4*, 368-374+437.

Zhang, Z. (2013). A Review of University English Teaching Practices and the Current Situation and Prospects of Language Testing. *Examination Weekly, No. 32*, 3-4.

Zhao, H. M. (2001). Reflections on the Modernization of Language Testing. *Journal of Chongqing Institute of Technology, No. 4*, 95-97.

Zheng, Y. (2018). Washback and Test Ethics Based on a Unified View of Validity—Reconstruction of the Participants Relation in English Test. *Journal of North Minzu University (Philosophy and Social Science Edition), No. 4*, 156-160.