

# ial

## Issues in Applied Linguistics

Volume 4 Number 2 • December 1993

### SPECIAL ISSUE

#### ETHICAL ISSUES FOR APPLYING LINGUISTICS

Jeff Connor-Linton and Carolyn Temple Adger, Guest Editors

#### MAIN ARTICLES

##### The Problem of Solutions: Two Cautionary Cases for Applying Conversation Analysis to Business

Jeff Connor-Linton

##### Ethical Considerations for Expert Witnesses in Forensic Linguistics

Edward Finegan

##### Ethical Issues for Applying Linguistics to Clinical Contexts: The Case of Speech-Language Pathology

Heidi E. Hamilton

##### Ethical Dilemmas for the Computational Linguist in the Business World

Heather McCallum-Bayliss

##### Ethics, Standards, and Professionalism in Language Testing

Charles W. Stansfield

##### Ethical Considerations in Language Awareness Programs

Walt Wolfram

##### Ethical Issues for Applying Linguistics: Afterword

Braj B. Kachru

#### EXCHANGE

##### Does Conceptualization Equal Explanation in SLA?

Cheryl Fantuzzi

### Ethics, Standards, and Professionalism in Language Testing<sup>1</sup>

Charles W. Stansfield  
Center for Applied Linguistics, Washington, DC

*This discussion addresses matters of ethics, standards, and professionalism in language testing. It begins with a definition of relevant terms, reviews current standards and codes that are relevant to language testing, and discusses their relationship to professionalism and the emergence of language testing as a profession. The author describes several personal experiences to illustrate the ethical choices that professional language testers often face. Sometimes these choices involve competing allegiances, which, in turn, create ethical dilemmas. The ethical issues, problems, choices, and dilemmas encountered by language testers provide evidence of the need for a code of ethics for language testing specialists who develop formal test instruments or work with formal language testing programs.*

#### PROLOGUE

After reading the language testing literature, I began teaching a graduate course in language testing at the University of Colorado in 1972. After ten years of teaching this course and consulting on occasional projects, I wanted to devote myself full time to the practice of language testing. So I left academia and entered the world of test makers and test publishers. I spent the next five years at Educational Testing Service, and have now spent seven years at the Center for Applied Linguistics. Throughout this time, there have been many occasions when I have been concerned about

Issues in Applied Linguistics  
© Regents of the University of California

ISSN 1050-4273  
Vol. 4 No. 2 1993 189-206

the issues of ethics, professionalism, and standards in the context of language testing. I have found that these matters do affect the decisions I must make and that ethical concerns can pose major dilemmas to language testers. Therefore, I agreed to accept the invitation to talk on the subject at this symposium on ethics in applied linguistics. I hope the following thoughts on the subject can be useful to others involved in language test development.

### DEFINITION OF TERMS: ETHICS AND STANDARDS

It is appropriate to begin a discussion of ethics by defining terms. In reviewing the definition of ethics and related words, naturally, I find that *ethics* has different meanings. As a branch of philosophy, it is concerned with the moral choices an individual must make in relationships with others. *Moral* seems to be a key word in understanding the meaning of ethics. Lexicographers define it as relating to the judgement of the goodness or badness of human action and character. So the concept of good and bad or good and evil is part of the philosopher's usage of the word *ethics*.

Ethics can also be applied to the professions, where it is sometimes called "applied ethics." When applied to a profession, *ethics* commonly refers to a set of principles governing the conduct of members of the profession. Here, *ethics* is used as the plural of *ethic*, which is a principle of correct or good conduct. So, in the context of language testing, *ethics* refers to a set of principles of good conduct in doing what language testers do. These principles would involve the regulation of professional conduct. A collection of ethics that has been formally adopted by an association is usually referred to as a *code of ethics*.

Closely related to ethics is *standards*. Indeed, the two terms are sometimes used synonymously. In this discussion however, I use the term *standards* to refer to technical criteria that are to be adhered to by professionals. This usage should not be confused with the psychometric meaning of standard, which is a basis for making comparisons in performance. Thus, in some sense, *ethics* is used here to refer to the moral conduct of language testers as people practicing their profession. *Standards* is used to refer to the procedures language testers follow in developing tests, operating test programs, or administering and interpreting tests and test

/ scores. Of course, standards and ethics invariably overlap since developing a test or operating a test program involves applying technical criteria within the context of moral and professional behavior.

Nonetheless, it is pedagogically useful to maintain the distinction, and a description of formally adopted standards and ethics that are relevant to language testers may contribute to an understanding of the minor differences between the two. I will focus on how they are developed and adopted, and the role they can play in the regulation of professional conduct.

### PREEXISTING CODES AND STANDARDS

The field of language testing is as much a part of the field of educational research and measurement as it is of applied linguistics. Since the field of educational measurement already has a set of standards and a code of ethics, it is worthwhile to consider their application to language testing.

#### The APA Standards

A well-defined set of standards for language testers is the *Standards for Educational and Psychological Tests*, developed jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1985). Commonly known as the *APA Standards* or simply the *Standards*, this is the "bible" of the test industry and the testing profession. First published in 1966, they were updated in 1974 and again in 1986.

The *APA Standards* consist of a set of rules that can serve as guidelines for the development of standardized tests and the operation of test programs. The standards are grouped in four sections and sixteen categories that begin with necessary background information authoritatively presented. Then, several standards are clearly stated in each category, and where necessary additional guidance, in the form of comments, follows. The *APA Standards* are organized as follows:

### **Technical Standards for Test Construction and Evaluation**

1. Validity (25 standards)
2. Reliability and Errors of Measurement (12 standards)
3. Test Development and Revision (25 standards)
4. Scaling, Norming, Score Comparability, and Equating (9 standards)
5. Test Publication: Technical Manuals and User's Guides (11 standards)

### **Professional Standards for Test Use**

6. General Principles of Test Use (13 standards)
7. Clinical Testing (6 standards)
8. Educational Testing and Psychological Testing in the Schools (12 standards)
9. Test Use in Counseling (9 standards)
10. Employment Testing (9 standards)
11. Professional and Occupational Licensure and Certification (5 standards)
12. Program Evaluation (8 standards)

### **Standards for Particular Applications**

13. Testing Linguistic Minorities (7 standards)
14. Testing People Who Have Handicapping Conditions (8 standards)

### **Standards for Administrative Procedures**

15. Test Administration, Scoring, and Reporting (11 standards)
16. Protecting the Rights of Test Takers (10 standards)

The *APA Standards* have had a profound effect on test publishing in the United States. The major test publishers use the *Standards* as a manual for carrying out their business. When publishing a test manual, for instance, they often refer to the *Standards* (see category 5 above) in an effort to comply with its maxims to the greatest degree possible. Although all test publishers are not equally guided by the *Standards*, the *Standards* have had a substantial impact on the practice of testing in the United States where they serve as a vehicle for voluntary quality control. This has

not been the case in other countries, however. The *Standards* are largely unknown outside the United States and to my knowledge no nationally accepted document comparable to the *APA Standards* exists in other countries.

### **The Code of Fair Testing Practices**

A few years ago, I was privileged to participate in the development of the *Code of Fair Testing Practices in Education* (National Council on Measurement in Education, 1988). The *Code* was developed under the sponsorship of several professional associations which are heavily involved in educational assessment: the American Educational Research Association (AERA), the American Psychological Association (APA), the National Council on Measurement in Education (NCME), the Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association.

The *Code* was the outcome of cooperation between commercial test publishers and representatives of relevant professional associations. This cooperation began in 1984 with a meeting of representatives of 23 test publishers and representatives of the AERA, NCME, and APA. At this meeting, the Joint Committee on Testing Practices was formed. This committee sanctioned the formation of two working groups, the *Code of Fair Testing Practices in Education* Work Group and the Test User Qualifications Work Group. Following additional meetings with more test publishers and relevant professional associations, each work group produced a document (Diamond & Fremer, 1989).

When organizations, institutions, and individual professionals adopt the *Code*, they commit to safeguarding the rights of test takers by following the principles prescribed in it. The *Code* is similar to certain parts of the *Standards for Educational and Psychological Testing*, but it differs in audience, purpose, and focus. The *Code* is meant to be understood by the general public; its scope is limited to educational tests; and it focuses on those issues that affect the proper use of tests. The *Code* does not create new principles over and above those in the *Standards*. Rather, it presents a selected portion of the *Standards* in a way that is meaningful to test takers and their parents or guardians.<sup>2</sup>

Although the *Code of Fair Testing Practices in Education* is, strictly speaking, not a code of ethics, it does have some of the

characteristics of a code of ethics in that it focuses on professional behavior with the goal of reducing the frequency of behaviors that could negatively impact examinees. It states the obligations to test takers of test makers and of professionals who use educational tests. While the *Code* is meant to apply broadly to the use of tests in education (admission, educational assessment, educational diagnosis, and student placement), it does not cover employment, licensure, or certification testing. Although the *Code* is relevant to many types of educational tests, it is directed primarily at professionally developed tests rather than tests made by individual teachers for use in their own classrooms.

The *Code* addresses the roles of test developer and test users separately. Test users are those who select tests, commission test development services, or make decisions based on test scores. The *Code* focuses on desirable behaviors in four main areas:

- A. Developing/Selecting Tests
- B. Interpreting Scores
- C. Striving for Fairness
- D. Informing Test Takers

The discussion of each area begins with general behavioral principles for each group, followed by specific behaviors that relate to the general principles for that area. In all, 21 behaviors are delineated side by side for each group. For example, the two general principles for the first area (Developing/Selecting Tests) are as follows.

Test developers should provide the information that test users need to select appropriate tests.

Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test-taking populations.

Since its publication, the *Code of Fair Testing Practices in Education* has been endorsed by over 100 professional associations and test publishers who have agreed to implement its tenets. This is an example of the contribution that professional cooperation through associations can make to the regulation of the conduct of individual and corporate practitioners.

The Test User Qualifications Work Group produced a series of papers specifying the qualifications that test purchasers should have. They produced sample forms for publishers that identify desirable knowledge and skills and require the purchaser to affirm that the test will be used responsibly. Both this document and the *Code* are examples of professionals working together to both clarify and specify ethical practices and standards within their profession.

## PROFESSIONALISM

A discussion of standards and ethics in the context of language testing raises the issue of professionalism. Professionalism is important only if one believes language testing to be a profession, and central to this belief is the question of domain. In whose domain does language testing lie and what ethical ramifications are traceable to this question? One could argue that competency in language testing should be achieved by all language teachers.<sup>3</sup> However, it is not realistic to expect all teachers to have the knowledge, skills, and abilities that professional language testers must acquire, and which are crucial to the field.

Whether one views language testing as a profession depends in part on one's concept of what it means to be a profession. The traditional professions of law, medicine, and divinity are sometimes considered examples of what it means to be a profession. These fields require extensive training, an apprenticeship (except in the case of law), certification of one's background, skills, and knowledge, approval by professional boards, and even licensure (in the case of law and medicine). If such requirements were applied uniformly to all fields, it is doubtful that language testing would ever be viewed as a profession. However, a broader definition of *profession* is an occupation that requires considerable education or training. Defined as such, language testing clearly meets the criteria of a profession. Since language educators need not possess the special understanding of general testing concerns that language testers possess, this is further evidence that language testing could claim to be a profession of its own.

In recent years some events have moved language testing in the direction of having the status of a profession. Clearly, an important event was the commencement in 1984 of a journal

(*Language Testing*) devoted to the subject. A more recent event was the formal creation of a professional association, the International Language Testing Association (ILTA), in February 1992, with a constitution, officers and a board of directors. The founding of ILTA was preceded by a year of formal and informal discussions among language testing specialists concerning the need for such an association. These discussions took place at international conferences devoted to language testing.<sup>4</sup> Because I moderated most of them, I know that many testing specialists are concerned about the degree to which language testers are accorded appropriate respect as professionals. This was, in fact, one perceived benefit of forming an association; that is, one characteristic of a professional is that he or she belongs to an association that represents the professional concerns of its members.

The concern for recognition of professional status motivates some ILTA members' support for other initiatives. Among these are the creation of a code of ethics for language testers, standards for language tests, formal certification of testers, and even licensure by the association. The latter two possibilities were considered when the ILTA Constitution was being drafted, since certification and licensure could have implications for one's status as a member of the association. Ultimately, we decided that membership in the association should be open to anyone interested in language testing. Still, this does not preclude ILTA from establishing a certification or licensure program at some point in the future.

ILTA is in a unique position to lead in the development of such a code of ethics for language testing. Because ILTA is an international association, the code's impact would affect the language testing community on a world-wide basis. The current national basis of most professional codes of ethics limits their effectiveness, but a code adopted by ILTA would not be limited by national boundaries.<sup>5</sup>

It can be concluded that most professions have a code of ethics, and many people consider a code of ethics to be a *sine qua non* of a profession. Thus, having a code of ethics might increase the stature of language testing. A more compelling reason for establishing a code is that it would provide language testers with a set of guidelines to follow as they confront problems while going about their business. Such guidelines would not just be comforting; they would also result in higher quality language tests and test programs.<sup>6</sup>

## SOME ETHICAL ISSUES AND PROBLEMS IN LANGUAGE TESTING

A number of inter-related contemporary issues in language testing raise ethical concerns. Sometimes these issues are complex and sometimes they involve competing concerns of an ethical nature, producing ethical dilemmas. One such issue is preparing students to take standardized language tests, such as the TOEFL. Another broader ethical concern, encompassing many different kinds of behavior, is the applications of test scores by test users.

An important ethical dilemma with wide implications for language testing is the matter of cost versus quality of measurement, i.e., reliability and validity. In this respect, language testing is no different than health care, in that high quality health care is expensive and money spent on health care could instead be spent in other ways that would also improve the quality of life within a society. One could spend a large amount of money to produce a language test that is both highly valid and highly reliable. However, one has to choose between the gain in the quality of the measurement outcome and the cost of the gain to whoever is paying for the test, be it the examinee, or some institution, corporation, or agency. It is here that standards and a code of ethics can be of help to the test developer or test program manager. The adoption of standards and an ethical code ultimately means that one must either follow them to a very significant degree or consider withdrawing from involvement in the project. A professional who does neither would risk losing face among his or her colleagues.

### Test Preparation Courses

In the United States until the 1980s, the practice of preparing a student for certain kinds of standardized tests, such as aptitude tests like the SAT or language proficiency tests like the TOEFL, was considered unethical by many testing professionals. This was because such tests are supposed to measure some relatively stable trait that is not amenable to rapid improvement. Test publishers criticized the ethics of authors of test preparation books, the proprietors of test preparation schools, and the teachers of test preparation courses. Test publishers gave test candidates some

sample test questions and told them to have faith that the test would measure their abilities accurately.

While I do not condone misrepresenting or overstating the value of a test preparation course, to me such courses, when not misrepresented, do not raise serious ethical concerns. Each examinee has a true ability level in the construct being measured. A good test preparation course will familiarize the examinee with the way the construct is measured on the test. If the examinee thoroughly understands the test developer's approach, he or she will be able to demonstrate through the test his or her full ability level, at least to the degree that the test is valid and reliable. Indeed, on this basis, one could argue that teachers have an ethical obligation to familiarize students with the format of a standardized test, if the test will be used to make an important decision about the students.

Bachman and Palmer (1982), Shohamy (1984), and others have shown that the item types used to assess a construct can influence the scores of examinees. That is to say, examinees of equal ability will appear to differ in ability due to the differences in the format of each test. This phenomenon is known as the test method effect. By preparing students for a test, we can reduce or eliminate the test method effect.

In 1991, I led the development of an oral proficiency test that is used to certify Spanish, French, and bilingual education teachers in one of the larger states within the USA. With the cooperation of the state education agency, we developed a full-length sample test and an 80 page test preparation manual to accompany this test. The manual is written for prospective examinees. It thoroughly explains the format of the test, the nature of each generic test task, the scale by which raters evaluate performance, and the criteria raters consider when placing a performance on the scale. A cassette tape provides the test candidate with the opportunity to take the test. The reverse side of the tape provides the candidate with examples of examinee performance at each level on the test scale in response to items on the sample test. Candidates take the sample test and then compare their performance with the performances on the other side of the tape. Thus, they can give themselves a tentative rating and decide whether they feel ready to register for the test. We have received very positive feedback from examinees who clearly stated that they appreciated knowing what to expect when they entered the test center. They consistently comment that the test preparation kit

prepared them for what to expect with the result that they were able to perform as well as their second language skills would permit.<sup>7</sup>

### Cost versus Quality

Testing professionals who construct tests under contract with specific clients are often faced with the cost versus quality issue. Often, the issue is indeed an ethical dilemma. Typically, one takes the amount of available money into account in the design of the test. Thus, one constructs the best test that one can for the amount of money available for test development and for maintenance of the operational test program. However, one is often faced with turning down an interesting test development project because the client is not willing or able to provide enough funds to develop a high quality test. I have faced this situation so many times that it is now easy for me to reject such projects. In spite of this, no matter how inadequate the funding, there is always someone who claims they can develop a good test for the amount of money available.<sup>8</sup> A recent example may be illustrative.

In 1990, a United States government agency issued a request for proposals to develop and validate speaking and reading proficiency tests for its employees in a large number of languages. It was expected that the work would be done in one year. Although the project seemed interesting, the number of languages involved and the time table indicated that the agency did not appreciate the complexity and, therefore, the cost of the task. Consequently, I decided not to bid for it. Ultimately, a contract was let to a vendor who had proposed to develop the tests for a fee that was within the amount the agency apparently expected the tests to cost. Later, I heard that the agency determined the tests were of such poor quality that they could not be used. Of course, the fault here lies with both the vendor, who said it could do the work just to get the money, and with the funding agency, which did not adequately investigate the cost of developing a high quality test. The vendor may not have had any idea how complex test development and validation is, since the company had almost no experience in test development. Thus, this outcome may have resulted from naive rather than unethical behavior on the part of the vendor. Yet, if one is aware that a reasonably good test cannot be developed for the amount of money available, it would seem to be at least somewhat unethical to accept the project

knowing that the outcome is unlikely to help the contractor. Yet contractors accept such work regularly.

Another more complex situation arises after one has created a high quality test and turned it over to the client, who has made it an operational program for some specific purpose. I have developed second language proficiency tests under contract with two state education agencies, which use them as part of the teacher certification process. Each time, after the agency has approved the type of test developed and indicated that it would implement it as discussed in our proposal and in our final report, due to competing budgetary demands, the cost of operating the program has eventually become a concern. When this happens, my support as the test developer is sought in order to legitimize some proposed cost-saving measures that would actually negatively impact the examinees. These measures typically involve reducing the length of the test (which reduces the cost of test administration), reducing the amount of training given to raters of productive skills tests (which reduces the cost of training), or scoring only a portion of the performance rather than the entire performance (which reduces the cost of scoring).

Each time this happens, I suggest any cost saving measures that can be taken without compromising test quality. However, each of the above proposals usually means a reduction in reliability or validity, and I explain why it cannot be done. One state has always accepted my response and continued to operate the program in a way that ensures quality assessment. In the other state, eventually my response was greeted with antipathy, even anger. Because the client funded me to develop the test, the client felt that I was obliged to endorse any new policies circumstances may dictate. When I refused to do so, the client ended our relationship. So, pressure from a client can be quite intimidating, particularly if language testing pays one's salary and the salary of one's employees.<sup>9</sup>

It should be pointed out, in fairness to the language tester's client, that the money spent to operate language testing programs could also be put to other educational uses. Money saved might be used to address other educational needs that appear more important at the moment.<sup>10</sup> Thus, a proposal that would diminish the quality of a test program poses ethical dilemmas on multiple planes. Nonetheless, I believe that cost saving proposals that will harm the quality of a test program or instrument are to be resisted, especially in the case of high stakes tests such as those used to determine

suitability for employment. Quite simply, if a high stakes test program cannot maintain an adequate degree of internal quality control, then it should probably be discontinued.<sup>11</sup>

Some test development companies have confronted this problem by developing a set of standards of their own, which they present to potential clients. Such standards can be quite useful in maintaining quality control, if they are followed by all test program managers. The *ETS Standards for Quality and Fairness* (1983) are an example. Nonetheless, program managers clearly wish to please the client, and may be more willing to compromise internal standards than to risk annoying or offending the client. Due to this possibility, some companies carry out an occasional internal audit of each test program in order to determine the degree to which it is in compliance.

It would be foolish for each individual test developer to develop his or her own set of standards. Each could argue that his or her standards are wholly appropriate and ensure quality control and responsibility, even if this were not the case. A set of standards for language tests in general could cause different test developers to follow similar practices which are widely held to be appropriate. Thus, a set of standards could contribute meaningfully to quality control across test projects. A code of ethics could also have a similar beneficial effect.

### Rater Training

Because language testing is likely to involve the training of raters, language testers often encounter moral choices concerning the training of raters. On one occasion, while conducting a four day workshop on the oral proficiency interview after which the participating teachers were to receive a certificate, I was called aside by a high level administrator. He told me that he wanted his assistant certified also, but that the assistant was very busy and could only attend part of the workshop. I resolved this problem by requiring all participants to independently rate a series of tapes to a given standard of accuracy. Because the administrator's assistant did not meet the standard, no certificate was issued.

A similar situation occurred a week before the conference at which I delivered this paper. An administrator who works with a test I developed told me he needed to be approved to administer and score the test, but that he did not have time to go through the formal

training process. He suggested that he be allowed to develop his skills through informal self-study and on the job experience. Although it was a bit brazen of me to do so, my response was to phone this person's supervisor and suggest that he be required to go through the same training as everyone else. When it was explained to me that this was not possible, a compromise was reached. Under the compromise, the individual will go through a different training process involving self-study, yet the process will still be quite formal in nature. At the end, he will have to demonstrate the same proficiency in administering and scoring the test as all others who carry out this task.

A related dilemma involves the training of a large group of raters prior to the operational scoring of a productive skills test. Invariably, some trainees will fail a "calibration" test at the end of the training. Such tests are used to ensure that all trainees can use the rating scale with adequate accuracy. When someone fails he or she is given additional training and another calibration test. When someone fails the second calibration test, one is faced with bringing possible embarrassment to that individual by not allowing him or her to participate in the operational scoring. This may be an ethical issue also. My response is always to carefully explain to the individual that it is important that everyone be able to rate accurately so that no examinee will be treated unfairly. Thus far, each time I have explained this, the individual has volunteered not to participate in the scoring session.

In the situations I have described above, it seemed at the moment that I would be better off to acquiesce and to let the client, the client's representative, or some other relevant player in the process have his or her way. After all, all of us like to be seen as cooperative people who are willing to help solve problems in the simplest possible way. Ultimately, however, I believe that ethical conduct is one way that I am able to project my professionalism. The maintenance of this professionalism is in my long term best interest.

## CONCLUSION

This paper has reviewed some current standards and codes that are relevant to language testing and discussed their relationship to the acceptance of language testing as a profession. I have drawn on my own personal experience to describe some of the situations that language testers can encounter where faced with ethical choices.

As can be seen, professional language testers constantly confront problems that relate to professional ethics and standards. Sometimes, the problems involve conflicting perceptions of fairness. At other times, the problems relate to the degree of compromise that one can tolerate and still have a quality product or program. The degree of compromise is crucial, since it affects the test developer, the client, the examinees, score users, and others who could be impacted by the decisions made based on test scores. Ultimately, the quality of tests affects and is of concern to society at large. Guidance on how to deal with issues of ethics and standards can be provided by a code of professional ethics for language testers and by standards for developing language tests and operating language testing programs. Ultimately, the psychometric qualities and appropriateness of language tests and the ethical behavior of professional language testers affect the stature of language testers in particular and applied linguists in general.<sup>12</sup>

## NOTES

<sup>1</sup> This paper was distributed and summarized orally at the Colloquium on Ethics and Applied Linguistics held at the annual meeting of the American Association of Applied Linguistics in Atlanta, GA, April 17-19, 1993. Please direct comments and reactions to the author at 1118 22nd St. NW, Washington, DC 20037; email CAL@GUVAX. Helpful comments on an earlier draft were made by Dorry Kenyon, Meg Malone, and the editors of this special issue of IAL.

<sup>2</sup> At the 1990 business meeting, in Thessaloniki, Greece, of the AILA Scientific Commission on Language Tests and Testing, John de Jong, then chair of the Commission, proposed that the *Code of Fair Testing Practices*, with two additional statements added, become the set of standards for language testing. However, when introduced at the business meeting, the proposal was tabled following discussion. The Code and the additions were subsequently reprinted in *Language Testing Update* (deJong, 1991) with a request for comments. Other activities pertinent to standards are mentioned in a recently published article by Fred Davidson

(1993) that was brought to my attention at the AAAL conference at which this paper was presented.

<sup>3</sup> This is true of the interface between testing and other areas of education as well. In most other disciplines teachers are even more lacking than second language teachers in the concepts of valid and reliable measurement. The growing importance of assessment in the educational process and the concern that teachers are lacking in their understanding of measurement, motivated the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (1990) to produce a document called *Standards for Teacher Competence in Educational Assessment of Students*.

<sup>4</sup> Major discussions were held at the 1990 Language Testing Research Colloquium in San Francisco, the 1990 Regional English Language Center conference in Singapore, at the 1990 AILA world conference in Thessaloniki Greece, and at the 1990 ACROLT meeting in Israel. Following these discussions, in 1991 a group of people met at the 1991 Language Testing Research Colloquium in Princeton, New Jersey to consider what had been learned during the discussion. At that meeting the author presented a formal written proposal to proceed with plans for the formation of a professional association. The proposal included the creation of a steering committee to draft a constitution for the organization. The work of the committee was presented and approved with minor revisions on February 28, 1992. Subsequently, officers and an executive board were elected.

<sup>5</sup> If at some point ILTA should begin to develop a code of ethics, some excellent and relevant codes that have already been developed should be considered. In addition to the *Code of Fair Testing Practices in Education*, I find the *Ethical Standards of the American Educational Research Association* (AERA, 1992) to be a helpful document in that it provides guidance on a number of issues that affect all researchers, including those in applied linguistics and language testing. Similarly, should ILTA ever wish to develop standards of competency in language testing for language teachers, the document mentioned in note 3 above would be helpful.

<sup>6</sup> Furthermore, the process of drafting, reviewing, revising or updating a set of standards or a code of ethics would be very valuable educational experience for all involved.

<sup>7</sup> Because the focus of this article is on ethical problems encountered by language test developers, we will not treat questionable test preparation practices of some language teachers.

<sup>8</sup> One could call this the P.T. Barnum effect.

<sup>9</sup> This example suggests that as test developers we must be responsible in the public statements we make about our language tests. It also suggests that we should not shirk responsibility for the consequences of our acts. Each of the two preceding sentences are close to being individual statements of ethics.

<sup>10</sup> On the other hand, a test that provides positive washback is a cost effective means of improving education.

<sup>11</sup> The situation professional language test developers sometimes confront is not unlike that described in the classic American play, *All My Sons*, by Arthur Miller. In this play, the father reduces the quality of airplane parts in order to earn more money which he uses to provide a better standard of living for his family. Eventually, the inferior parts cause planes to crash, with the result that the family suffers shame and, eventually, disaster.

<sup>12</sup> Many of the issues treated in this article (the need for a code of ethics, technical standards of quality, documents with international impact, and the short-term benefits versus the long term costs of acquiescing to a client's demands) may also apply to some degree to other subfields of applied linguistics whenever those professing to have expertise in these subfields "apply" their expertise outside of academia in a workplace, a school setting, a hospital, or in a court of justice.

## REFERENCES

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association. (1992). Ethical Standards of the American Educational Research Association. *Educational Researcher*, 21(7), 23-26.
- American Federation of Teachers, National Council on Measurement in Education, National Education Association. (1990). Standards for Teacher Competence in Educational Assessment of Students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Bachman, L. F. & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4), 449-65.
- Davidson, F. (1993). Testing English across cultures: Summary and comments. *World Englishes*, 12(1), 113-125.
- de Jong, J. (1991). Standards for language testing. *Language Testing Update*, 9, 7-11.
- Diamond, E. E. & Fremer, J. (1989). The Joint Committee on Testing Practices and the Code of Fair Testing Practices in Education. *Educational Measurement*, 8(1), 23-24.
- Educational Testing Service. (1983). *ETS Standards for Quality and Fairness*. Princeton, NJ: Author.
- Eyde, L. D., Moreland, K. L. & Robertson, G. J. (1989). Test user qualifications: A data-based approach to promoting good test use. Executive Summary. Washington, DC: American Psychological Association.
- Joint Committee on Testing Practices. (1988). *Code of fair testing practices in education*. Washington, DC: National Council on Measurement in Education.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147-70.

**Charles W. Stansfield** is director of the Division of Foreign Language Education and Testing as well as the ERIC Clearinghouse for Languages and Linguistics at the Center for Applied Linguistics in Washington, DC. He led the formation of the International Language Testing Association (ILTA) and served as the organization's first president. He has published extensively on language testing and has led the development of many formal second language proficiency tests.