The Role of a Language Testing Code Ethics in the Establishment of a Code of Practice

Randy Thrasher


To link to this article: http://dx.doi.org/10.1080/15434303.2004.9671782

Published online: 22 Jun 2011.

Article views: 175

View related articles

Citing articles: 1 View citing articles
The Role of a Language Testing Code Ethics in the Establishment of a Code of Practice

Randy Thrasher
Okinawa Christian Junior College

This article examines the relation between a language testing code of ethics and a code of practice by discussing the development of the Japan Language Testing Association (JLTA) draft Code of Practice. It claims that the relation between the two sorts of codes is not as straightforward as seems to be assumed by the authors of the International Language Testing Association (ILTA) Code of Ethics. But it argues that a code of ethics is useful in deciding to whom a code of practice should apply and in justifying the inclusion of the various elements in the code of practice. One reason JLTA had for demanding that our Code of Practice apply to both high- and low-stakes testing is rooted in the ILTA Code of Ethics—the demand that all test takers be treated as fairly as possible. It is also claimed that the purpose of a code of practice, in Japan at least, is not to discipline members of our professional organization but to hasten the day when testing practices that treat test takers unfairly are replaced by those a consensus of professional language testers believe are fairer.

In this article I would like to explore the relation of a code of ethics and a code of practice using the ILTA Code of Ethics (ILTA Code of Ethics, 2000) and the first draft of a code of practice (JLTA Code of Practice, 2001) being developed by the Japan Language Testing Association (JLTA). I believe that this relation is a crucially important issue for ILTA. As is well known, ILTA had tried for several years to draft a language testing code of practice but was unable to reach agreement. So the decision was made to draft a code of ethics and then later try to establish a code (or codes) of practice. At the Language Testing Research Colloquium (LTRC) 2000, in Vancouver, British Columbia, Canada, the draft ILTA Code of Ethics was presented and accepted by the membership. Soon after, at the direction of then

Requests for reprints should be sent to Randy Thrasher, Okinawa Christian Junior College, 777 Onaga, Nishihara-cho, Okinawa 903-0207, Japan. E-mail: thrasher@ocjc.ac.jp
ILTA president, Alan Davies, two committees were set up, one under the chairmanship of Lyle Bachman and the other with me as chair set up by the JLTA.

**RELATION BETWEEN A CODE OF ETHICS AND A CODE OF PRACTICE**

It may seem at first glance that the connection between a code of ethics and a code of practice is straightforward. The code of ethics states the broad ethical demands and the code of practice spells out in practical terms how those ethical demands are to be satisfied in the day-to-day work of language testers. This certainly was the view of the writers of the ILTA *Code of Ethics* (2000). They stated that

> The Code of Ethics is instantiated by the Code of Practice (currently under preparation by ILTA). Although the Code of Ethics focuses on the morals and ideals of the profession, the Code of Practice identifies the minimum requirements for practice in the profession and focuses on the clarification of professional misconduct and unprofessional conduct. (*ILTA Code of Ethics*, p. 15)

The authors clearly believed that the role of a code of practice was to state the practical implications of a code of ethics. And, I must admit that, for a few of the elements in the ILTA *Code of Ethics* (2000), it could be claimed that there is a need to spell out the demands of the code in practical terms in a code of practice. The first principle is one such element. It states that,

> Language testers shall have respect for the humanity and dignity of each of their test takers. They shall provide them with the best professional consideration and shall respect all persons needs, values, and cultures in the provision of their language testing services. (*ILTA Code of Ethics*, p. 15)

It could be argued that developing tests that are valid and reliable is necessary to "provide the best professional consideration." But it is much harder to see how Principle 5, “Language testers shall continue to develop their professional knowledge, sharing this knowledge with colleagues and other language professionals” (p. 19), could be spelled out in a code of practice. In fact, there seems to be no need to do so. And because the annotation to Principle 4, “Language testers shall not allow the misuse of their professional knowledge or skills, in so far as they are able”(p. 18), indicates that this principle is not concerned with the misuse of test results so much as prohibiting language testers from participating in unethical or immoral activities or projects, it is also difficult to see how this principle could be instantiated in a code of testing practice.
If it is the case, as I have tried to point out, that there is no simple relation between a code of ethics and a code of practice, we are faced with several possibilities. We can decide that there is no useful relation between the two sorts of codes and go ahead and try to build a code of practice without reference to the ILTA *Code of Ethics* (2000). Or we can assume that such a relation exists and begin to search for the ways the two types of codes are connected. It is this second stance that I adopted in trying to develop a code of practice that would fit the situation in Japan. In part this decision was based on my belief that there should be a connection and in part on the great need to justify the code of practice that we were trying to build.

**ROLE OF A CODE OF ETHICS**

From our experience (the JLTA) trying to develop a code of practice, I have found two areas in which the ILTA *Code of Ethics* (2000) has been important in our work. It has made us take a particular stance on both to whom the code should apply and how it might be structured. And it has provided us with a way of justifying the elements of the code that we have proposed. Let me deal with each of these in turn.

**To Whom a Code of Practice Should Apply**

A code of ethics should be universal—at least within the field for which it is designed to cover. But the creation of a universal code of practice may not be possible. Most of the discussion as to why such a universal code may or may not be possible has focused on cultural differences. This is certainly an issue in Japan. But there is another aspect of the universality question. This can be seen most clearly in the different approaches of the National Council on Measurement in Education (NCME) and the Joint Committee on Testing Practices of the American Psychological Association (APA). The NCME states, “The purpose of the *Code of Professional Responsibilities in Educational Measurement* (Schmeiser, Geisinger, Johnson-Lewis, Roeber, & Schafer, 1995) is to guide the conduct of NCME members who are involved in any type of assessment activities in education” (p. 1). However, the APA takes a different approach.

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in Education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed pri-
marily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The code is not intended to cover tests made by individual teachers for use in their own classrooms. (Joint Committee on Testing Practices, 1988, p. 9)

Clearly the APA is taking a more narrowly focused approach, and there are good practical reasons for this stance. But this approach raises an ethical issue. The APA stance (Joint Committee on Testing Practices, 1988) gives the impression that only test developers creating what the authors of the code call “professionally developed tests” need worry about building measurement devices that meet the ethical standards set by the association.

The first draft of the Code of Practice of the JLTA (Thrasher, 2001) attempts to avoid both the problem of appearing to claim that only certain test developers need to follow ethical standards and the practical difficulty of coming up with a code that fits all language testing situations by trying to state rules that apply to all language testing and then stating additional rules (or stricter versions of general rules) that are to be followed in commercial and other high-stakes testing situations.

There are two reasons for demanding that the code apply to every testing situation. The first is ethical. All test takers, whether facing a high- or low-stakes test should be treated as fairly as possible. The amount of evidence needed to demonstrate the validity of a classroom quiz will be less than that required of a commercially available examination used to make significant decisions in the lives of the test takers. The degree of rater reliability acceptable in the grading of an in-class essay does not need to be as high as that of the Test of Written English (TWE, published by the Educational Testing Service). But it does not follow from this that classroom teachers should not be concerned about validity and reliability.

The other reason for demanding that the code of practice cover all language testing is a practical one. If we are going to upgrade testing practice (in Japan, at least), we must get all test-writing teachers and the whole test-taking public to understand what good testing practice is. In Japan, it is usually the case that the classroom teachers write not only the tests for their classes but also those for high-stakes decisions such as deciding entrance to their school. Therefore, we do not have a chance of changing the quality of entrance examinations unless we first develop a better understanding of good testing practice among classroom teachers. We cannot expect them to wear an ethical testing hat when developing entrance exams if they have not already learned how to wear that hat in writing the tests they use in their classrooms.

The same argument can be made from the students’ point of view. We need to help them to understand what good testing practice is and have them begin to demand that the tests they must take reflect such practice. Yet, there is little that test takers can do about the high-stakes tests they must take. But they can speak up when classroom tests seem lacking in validity or reliability. The identity of en-
trance exam writers is a closely guarded secret, but the students know who to com-
plain to about poor classroom tests.

JLTA Draft Code of Practice

For these reasons I have proposed a six-part structure for the JLTA Code of Prac-
tice: (a) basic considerations for good testing practice in all situations, (b) respon-
sibilities of test designers and test writers, (c) obligations of institutions preparing 
or administering high-stakes exams, (d) obligations of those preparing and admin-
istering commercially available exams, (e) responsibilities of users of test results, 
and (f) special considerations. The first section contains elements such as the fol-
lowing basic considerations for good testing practice in all situations:

1. The test developer’s understanding of just what the test, and each subpart 
of it, is supposed to measure (its construct) must be clearly stated.
2. All tests, regardless of their purpose or use, must be valid and reliable to the 
degree necessary to allow the decisions based on their results to be fair to 
the test takers.
3. Test results must be reported in a way that allows the test takers and other 
stakeholders to interpret them in a manner that is consistent with their 
meaning and degree of accuracy.

The second section includes responsibilities of test designers and test writers:

1. A test designer must begin by deciding on the construct to be measured be-
fore deciding how that construct is to be operationalized.
2. Once the test tasks have been decided, their specifications should be 
spelled out in detail.
3. The work of the item writers needs to be edited before the items are pre-
tested. If pretesting is not possible, the items should be analyzed after the 
test has been administrated but before the results are reported. Malfunc-
tioning or misfitting items should not be included in the calculation of indi-
vidual test takers’ reported scores.

Although each element of the ILTA Code of Ethics (2000) begins with “lan-
guage testers” and thereby seems to be directed only to individual test writers and 
developers, the third section of the draft JLTA code is aimed specifically at institu-
tions. The reason for laying part of the responsibility for good testing practice on 
institutions stems from the experience of many JLTA members of being the only

---

1This article presents only examples from the JLTA Code of Practice. The complete draft of this code is available on the JLTA website at http://www.avis.ne.jp/~youichi/COP.html
testing professional on a committee set up to write a high-stakes test. We realize that in arguments in such committees we need to be able to say that good testing practice is the responsibility of not only the test writers but also the institutions who develop, administer, and use the test results for important decision making.

Obligations of Institutions Preparing or Administering
High-Stakes Exams

Responsibilities to test takers and related stakeholders:

**Before the test is administered.** The institution should provide all potential test takers with adequate information about the nature of the test, the construct (or constructs) the test is attempting to measure (Ideally this should include any evidence and arguments showing that the test tasks are in fact measuring what they are claimed to measure), the way the test will be graded, and how the results will be reported.

**At the time of administration.** The institution shall provide facilities for the administration of the test that do not disadvantage any test taker. The administration of the test should be uniform, assuring that all test takers receive the same instructions, time to do the test, and access to any permitted aids. If something occurs that calls into question the uniformity of the administration of the test, the problem should be identified, and any remedial action to be taken to offset the negative impact on the affected test takers should be promptly announced.

**At the time of scoring.** The institution shall take the steps necessary to see that each test taker’s exam paper is graded accurately and the result correctly placed in the database used in the assessment. There should be ongoing quality control checks to assure that the scoring process is working as intended.

**Other considerations.** If a decision must be made on candidates who did not all take the same test or the same form of a test, care must be taken to assure that the different measures used are in fact comparable. Equivalence must be demonstrated statistically. If more than one form of the test is used, interform reliability estimates should be published as soon as they are available.

The decision to include a section directly addressed to developers of commercially available tests reflects the need to have some influence on the huge language testing business in Japan. It is these tests that set the public’s understanding of what a test is, how tests should be administered, and test results reported. Unless the companies producing such tests can be convinced to use good testing practices, the
public will continue to believe that the practices these companies presently use are
good (or, at least, acceptable) testing procedures.

Obligations of Those Preparing and Administering
Commercially Available Exams

In addition to the obligations placed on any test designer and on those preparing
high-stakes examinations, developers and sellers of commercially available exam-
inations must:

1. Make a clear statement as to what groups the test is appropriate for and for
   which groups it is not appropriate.
2. Publish validity and reliability estimates for the test along with sufficient
   explanation to allow potential users to decide if the test is suitable in their
   situation.
3. Report the results in a form that will allow the test users to draw correct in-
   ferences from them and make them difficult to misinterpret.
4. Refrain from making any false or misleading claims about the test.

The fifth section of the proposed code deals with the use of test results. The rea-
son for its inclusion is to focus on the importance of using test results properly (As
Messick (1996) has pointed out, the crucially important element in test validity is
the inferences drawn on the basis of the test results) and to help test-using institu-
tions to both improve their test-based decision making and put pressure on test pro-
viders to give their clients the sort of results needed to make meaningful decisions.

Responsibilities of Users of Test Results

Persons who use test results for decision making must:

1. Use results from a test that is sufficiently reliable and valid to allow fair de-
   cisions to be made.
2. Make certain that the test construct is relevant to the decision to be made.
3. Clearly understand the limitations of the test results on which they will
   base their decision.
4. Take into consideration the standard error of the mean (SEM) of the device
   that provides the data for their decision.
5. Be prepared to explain and provide evidence of the fairness and accuracy of
   their decision-making process.
The last section of the code is an et cetera category that attempts to addresses issues that did not seem to fit nicely into the other sections.

**Special Considerations**

*In norm-referenced testing.* The characteristics of the population on which the test was normed must be reported so that test users can determine if this group is appropriate as a standard to which their test takers can be compared.

*In criterion-referenced testing.* (1) The appropriateness of the criterion must be confirmed by experts in the area being tested; (2) because correlation is not a suitable way of determining the reliability and validity of criterion-referenced tests, methods appropriate for such test data must be used.

*In computer-adaptive testing.* (1) The sample sizes must be large enough to assure the stability of the Item Response Theory (IRT) estimates; (2) test takers and other stakeholders must be informed of the rationale of computer-adaptive testing and given advice on test-taking strategies for such tests.

**CODE OF ETHICS AS JUSTIFICATION OF POINTS IN A CODE OF PRACTICE**

The motives of the builders of a code of practice can be called into question just as Beyerstein, quoted in Boyd and Davies, claimed that “codes of ethics exist primarily to make professionals look moral” (2001, p. 10). But there are legitimate motives for trying to establish a code of practice, and these motives are rooted in the ILTA Code of Ethics (2000). It seems to me that we need such a code because the present practices treat the test takers unfairly. That is, they do not show “respect for the humanity and dignity of each … test taker” (Principle 1 in the ILTA Code of Ethics, 2000) and what is happening (in Japan, at least) does “advance purposes im- Mical to … test takers interests” (first annotation to Principle 4). The second annotation to Principle 6 indicates the societal responsibility of our profession to work for beneficial change. “Language testers develop and exercise norms on behalf of society” (p. 20). And the third annotation to Principle 7 states the goal of a code of practice—to ensure “that language testing test takers have available to them the best possible testing service” (p. 21). I believe that we can reach such a goal if we put the test taker at the center of our concerns. By demanding ethical treatment of test takers, we can not only justify the rules we put into a code of practice but also move toward the day of making the best possible testing service a reality.
BROADER ISSUES

In the various discussions of the development of a code of good testing practice, three major questions have emerged. One is the practical issue of what should be included in such a code. Another is how to enforce such a code, and a third concerns the possibility of a universal code of practice. This last issue is one that I feel strongly about but must admit I am not in a position to do more than state my belief that a universal code should be possible. However, I believe that my experience in helping develop a code of practice for the JLTA (Thrasher, 2001) allows me to say a bit more about the first two issues.

The issue of what to include in a code of good testing practice is often seen as the question, can language testing professionals agree on what should go into such a code? Because my experience in the field dictates a “no” answer to this question, I believe we have to settle for broad consensus rather than 100% agreement. And we must also give up the idea that a code can be produced that we will be able to set in stone. A code of practice will always be a work in progress. The code will have to be revised with the deepening of our knowledge of testing, developments in statistics, and our experience with the use of the code as originally stated. The authors of the ILTA Code of Ethics (2000) made the same point for their code, and a code of practice will probably need to be updated even more often than a code of ethics.

The issue of the enforcement of a code of practice is inevitably linked to what one sees as the purpose or function of such a code. Some seem to view a code of practice as similar to a legal code and expect that violators can be punished for unprofessional conduct. It is impossible to tell for sure, but it appears to me that this sort of thinking may be behind the words in the ILTA Code of Ethics: “While the Code of Ethics focuses on the morals and ideals of the profession, the Code of Practice identifies the minimum requirements for practice in the profession and focuses on the clarification of professional misconduct and unprofessional conduct” (2000, p. 15).

I am not convinced that this should be the proper role of a code of practice in our field, but even those who would like to have a code that could be used in this way must admit that it is not possible in our present situation. Language testers are not licensed and associations such as the ILTA and the JLTA, as they are presently constituted, have few if any procedures for disciplining members. And what is probably the most important reason for the impossibility of a code of practice as a means of disciplining test developers and users is that a large number of the developers and the bulk of the users are not members of any professional testing association. This leaves us with no alternative but to present the code of practice as the consensus view of professional language testers and urge test developers and users to follow it.
CONCLUSION

The attempt by the JLTA (Thrasher, 2001) to draft a code of good testing practice based on the ILTA Code of Ethics (2000) has raised important issues concerning the connection between such codes. I believe that it has shown that a code of practice is not simply the spelling out of each principle of the code of ethics. But there is, I have argued, a strong and necessary connection between the two sorts of codes. In the case of the JLTA Code of Practice (2001), the ILTA Code of Ethics has provided both the rationale for including the various points of the code and the moral ground to demand that the code of practice apply to all language testing, not just high-stakes exams. But I have also pointed out that the development of the JLTA Code of Practice and considerations of how it can be used in this society argue against the view reflected in two statements in the prologue to the ILTA Code of Ethics that indicate that failure to live up to the letter of such codes should result in what are called “serious penalties” (p. 14). Our experience in Japan has pushed us to take the position that the role of both the code of ethics and particularly the code of practice should be primarily educational and be an indication that testing practice has the seal of approval of language testing professionals. I believe that it would be wise for ILTA to monitor the level of acceptance of the JLTA Code of Practice to see if some statements in the prologue of the ILTA Code of Ethics might need to be modified.

REFERENCES


