Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference.
May 13-14, 2006. Shizuoka, Japan: Tokai University College of Marine Science. (p. 48 - 73).

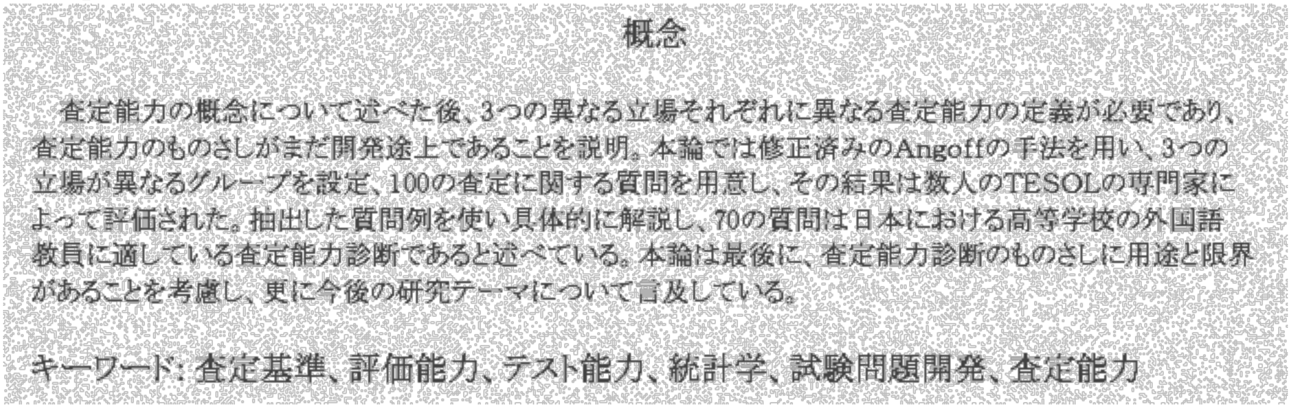
Teacher development and assessment literacy
外国語教員教育と査定能力

by Tim Newfields (Toyo University)

Abstract

After defining the concept of assessment literacy and possible operationalizations of this concept for three different populations, the rationale for developing an assessment literacy scale is explained. Using a modified Angoff procedure, the suitability of 100 possible assessment literacy items for three target populations was evaluated by a small panel of experts. Sample items are described and 70 items concerning assessment related issues that may be appropriate for high school foreign language teachers are outlined. This paper concludes by considering possible uses and limitations of the *Assessment Literacy for High School Foreign Language Teachers Inventory* and a call for further research on assessment literacy.

Keywords: assessment standards, evaluation skills, test competence, statistical literacy, test development



概念

査定能力の概念について述べた後、3つの異なる立場それぞれに異なる査定能力の定義が必要であり、査定能力のものさしがまだ開発途上であることを説明。本論では修正済みのAngoffの手法を用い、3つの立場が異なるグループを設定、100の査定に関する質問を用意し、その結果は数人のTESOLの専門家によって評価された。抽出した質問例を使い具体的に解説し、70の質問は日本における高等学校の外国語教員に適している査定能力診断であると述べている。本論は最後に、査定能力診断のものさしに用途と限界があることを考慮し、更に今後の研究テーマについて言及している。

キーワード: 査定基準、評価能力、テスト能力、統計学、試験問題開発、査定能力

The notion of "literacy" has been interpreted in many ways (Wagner and Kozma, 2003). Originally denoting a familiarity with literature or condition of being cultured, it later came to be associated with a capacity to read, write, and count (Openjuru, 2003). More recently this concept has been described in economic terms as a way of enhancing "human capital" (Machlup, 1984; Vittal, 2002). Since the skills reputedly involved in each form of literacy reflect the prevailing norms of a culture at a given point in time, shifts are inevitable. A generation ago, for example, few people considered computer skills as indispensable for assessment literacy (Séror, 2005). However, as post-classical statistical methods have become more sophisticated, familiarity with at least one statistical software program appears to be essential to evaluate many types of tests (Donoho, 2000).

This paper examines the notion of assessment literacy and some of its possible components. After mentioning why assessment literacy is important for teachers, let's briefly conceptualize this term, then attempt to operationalize it, and finally examine some screening items that might actually begin to express what this notion represents for three different groups. Those who are hoping to find a single, cogent definition of "assessment literacy" that works for all groups will be disappointed because I believe the construct represents a wide matrix of skills which vary significantly from population to population. What might be called "assessment literacy" from the viewpoint of a university student, a high school teacher, and a professional test developer probably involve vastly different skills.

Three reasons for teacher assessment literacy

Why is assessment literacy important for teachers? There are three compelling reasons. First, assessment is a widespread (if not intrinsic) feature of most educational systems. Teachers are estimated to spend from 10% - 50% of their work time on assessment-related activities (MacBeath & Galton, 2004, p. 31; Gunn, cited in Brindley, 1997) In many schools, a good portion of the budget also goes into formal testing. With so much time and money devoted to assessment, it's worth critically understanding how assessment decisions are made.

A second reason assessment literacy is essential because it's necessary to understand much of the educational literature. A familiarity with basic statistical terms is needed not only to critically read specialized journals, but even many general articles in academic publications. Without grasping basic statistical concepts it's often difficult to weigh the evidence for or against any point described in an article. And when this happens, research moves further away from the realm of science and closer towards unfounded sophistry.

A final reason assessment literacy is needed is that it allows teachers to communicate their own classroom results with others. As Hopkins (1985, p. 58-60) suggests, teachers should share their research with peers and develop a community that fosters learning. To make classroom research comprehensible to a wide audience, mastering the conventions of qualitative and quantitative inquiry are essential. Assessment literate researchers should be committed to an ongoing self-critique their own research and sharing the results in ways that are technically convincing. All too often articles with interesting insights lack sufficient analysis and/or evidence to allow readers to critically interpret the ideas.

Assessment literacy: A relativistic conceptualization

Interestingly, the term assessment literacy isn't listed in the *Dictionary of Language Testing* (1999) or *Japan Language Testing Association's Bilingual List of Language Testing Terms* (2006). Nor does it appear in the ALTE *Multilingual Glossary of Language Testing Terms* (1998) or Mousavi's *Encyclopedic Dictionary of Language Testing* (2002). Though each of these works devote ample space to the concept of assessment, the issue of how people actually become competent at assessing isn't mentioned. Instead of conceptualizing assessment literacy solely as a set of given skills, we also need to focus on the conditions needed to foster such skills. How do people actually learn about concepts often expressed in involved in assessment? That issue is worth exploring.

"*Instead of conceptualizing assessment literacy solely as a set of given skills, perhaps we should also focus on the conditions needed to foster such skills*."

Rather than describe assessment literacy as a single phenomena with some sort of unitary meaning, this paper adopts a grounded theory perspective by suggesting it means different things to different populations. For students, it largely means knowing how to perform well on exams. For teachers, it is associated with the ability to grade students ethically and accurately. And for professional test developers, every facet of their work hinges assessment literacy. This paper explores what assessment literacy might mean from three contrasting views: the perspectives of professional test developers, high school foreign language instructors, and incoming university undergraduates. An operationalization of assessment literacy for each of these three populations follows. However, those conceptualizations should be regarded as preliminary since not enough structured interviews with respondents in each of these three populations have been completed.

Assessment literacy: Three preliminary operationalizations

What does assessment literacy likely mean from the perspective of a first year university undergraduate? It might seem that at least two competencies are likely required: (1) an ability to effectively interpret test scores/grades and (2) some understanding of what does/doesn't constitute ethical testing/grading practices. For the first task, what is often referred to as *statistical literacy* (Wallman, 1992, p. 1) is needed. In simplest terms, that means a capacity to make sense of the quantitative data encountered in everyday life.

When I ask university students in Japan about assessment literacy in their native language, very few are able to articulate anything. Is this because they lack the meta-language needed to describe the concept? Or perhaps is this because the term *satei nouryoku* (perhaps the best translation of "assessment literacy") is not a pervasive word in Japanese? Such questions are fascinating, but beyond the scope of this paper.

Assessment literacy: What it means for university undergraduates

Ohata (2005) suggests that test anxiety is far from rare in Japan. Many Japanese have what Baker (2006) describes as a victim mentality when it comes to testing. At the same time, Japan can be described as an edumetric society in which test results significantly impact life outcomes.

One way of operationalizing assessment literacy in terms of what it might mean for university undergraduates majoring in education appears in Table 1.

Table 1. *A suggested operationalization of "assessment literacy" for one first year college student population*

1. An ability to interpret statistical raw data in terms of common measures of centrality (mean, mode, median) and deviation (SD, quartiles).
2. A basic understanding of the concept of measurement error and confidence intervals.
3. An ability to discern whether or not the difference between two or more data sets is significant.
4. A capacity to logically distinguish between correlation and causation.
5. An understanding of what constitutes ethical assessment – and what should be done if encountering unethical testing practices.

These points represent many of the competencies outlined in the American Association for the Advancement of Science's *Benchmarks for Science Literacy* (1993). In other words, they denote what some educators think that university students should know. Chances are, this is far from what most students themselves actually want to know. Often, the biggest challenge in promoting assessment literacy seems to be convincing end-users that the topic is actually worth learning: when many people encounter the arcane jargon and complex statistical formulas sometimes used in assessment, perplexity is a frequent response. There's so much to know and most learners (and perhaps teachers as well) are frankly unprepared. Generally speaking, little time is given in promoting assessment literacy skills in most curriculum in Japan: people take tests without considering why such tests are being conducted or whether they ethically assess a given skill.

What does assessment literacy likely mean for high school foreign language teachers? Perhaps in addition to the points in Table 1, items concerning test creation, grading, and communicating with stakeholders should be included. Table 2 lists some of the additional components that might be associated with assessment literacy for this specific population.

Table 2. *A suggested operationalization of six further criteria for "assessment literacy" for foreign language teachers working in high schools*

6. An ability to use a broad variety of assessment measures to assess students with minimal bias.
7. An ability to construct, administer, and score tests within a given field of expertise.
8. An ability to evaluate the reliability, item difficulty, item facility, and content validity of tests within ones field of teaching.
9. The ability to statistically determine where the cutoff point of a CRT examination should be.
10. The ability to intervene appropriately if students engage in unethical behavior during a test.
11. Skill in communicating assessment results effectively to parents, peers, and students.

These items are based in part on the *1990 Standards for Teacher Competence in the Educational Assessment of Students* published jointly by the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. Would these points also apply to Japanese contexts? There is no reason to think they wouldn't even though the "testing culture" (Sacks, 2000, p. 282) of both nations differs.

Professional test developers, who have the most authority in directly shaping testing outcomes, also need the highest level skills in order to promote fair testing practices. In addition to having all of the abilities mentioned in the two previous tables, professional test developers should have at least six additional skills as outlined in Table 3.

Table 3. A *suggested operationalization of six further criteria for "assessment literacy" among professional test developers*

10. The ability to provide clear evidence of what a specific test does and does not measure.
11. A commitment to indicating what the appropriate/inappropriate uses of a given test are.
12. A demonstrated concern for client confidentiality and test security.
13. Knowledge of how to detect poorly performing test items and how to factor out those items from the test scores.
14. An ability to detect various factors unrelated to a target skill are confounding examinee test performance.
15. An ongoing commitment to test improvement and cyclic validation.

Most of these points are based on the 2004 *Code of Fair Testing Practices in Education* adopted by the Joint Committee on Testing Practices of the American Psychological Association. What is disconcerting is how little they are actually implemented. For example, if a university in Asia wishes to use the Institutional TOEIC® as a placement test to stream its incoming students – a use for which that examination was not designed and probably should not be ethically applied – the testing agency will fully cooperate. When business imperatives collide with ethical concerns, it is often easy to predict the winner.

Method

Although this paper is influenced by Strauss and Corbin's (1990) work on grounded inquiry, methodologically owes more to Angoff (1971), who suggested a procedure to ascertain what might be regarded as a "minimal competence" in a given field based on input from a panel of expert informants. The opinions of the expert informants described here need to be further triangulated with data from the target populations and perhaps also with those who train or administer each of the specific groups.

This project has seven steps, as outlined in Figure 1.

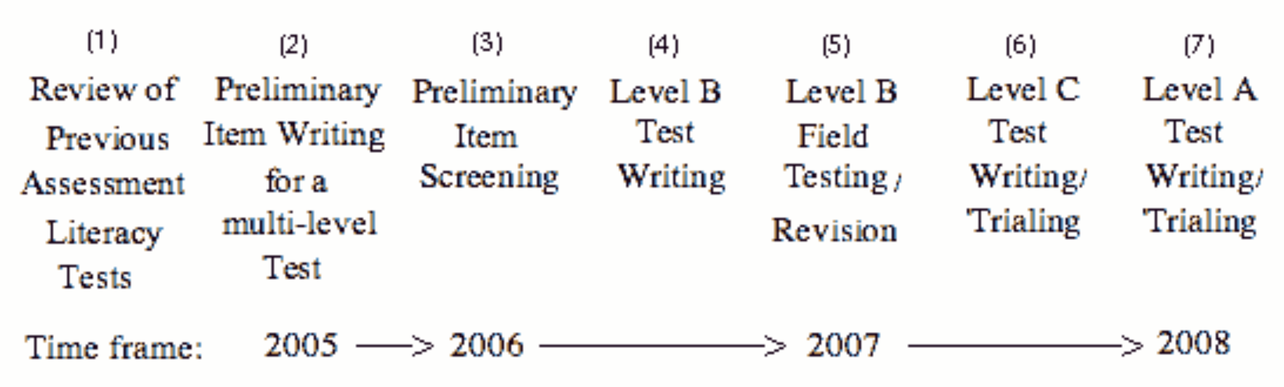

(1) Review of Previous Assessment Literacy Tests
(2) Preliminary Item Writing for a multi-level Test
(3) Preliminary Item Screening
(4) Level B Test Writing
(5) Level B Field Testing / Revision
(6) Level C Test Writing/ Trialing
(7) Level A Test Writing/ Trialing
Time frame: 2005 —> 2006 —> 2007 —> 2008

Figure 1. Procedure adopted in this assessment literacy research

The first four steps of this procedure are outlined in this paper; subsequent steps will be detailed in future publications.

(1) *Review of previous assessment literacy tests*

Two types of works informed this study: (1) theoretical descriptions of what assessment literacy supposedly entails, and (2) examples of actual tests, surveys, or encyclopedic entries pertaining to testing and assessment.

In the first category, the APA *Report of the Task Force on Test User Qualification*s (2000) provided a good overview of what well-informed test users purportedly should know. In addition to addressing ethical issues about testing, it also outlines basic statistical concepts. Another helpful source of information was the *Standards for Teacher Competence in Educational Assessment of Students* (1990) which highlights some of the duties of teachers with respect to assessment. The *Code of Fair Testing Practices in Education* (2004) also provides broad ethical guidelines about how tests should be conducted.

In the second category, Mertler's (2002) *Classroom Assessment Literacy Inventory* - a collection of 35 questions classroom assessment for K-12 teachers - was helpful in framing some questions. One other resource which informed this study was the *Online Educational Assessment Tutorial* by Yu Chong-ho. Finally, I relied on encyclopedic references about assessment literacy. Mousavi's *Encyclopedic Dictionary of Language Testing* (3rd Edition) and Alkin's *Encyclopedia of Educational Research* (1992) were both particularly helpful.

(2) *Preliminary item writing for a multi-level test*

After reviewing the literature mentioned above and gathering a list of many potential questions touching on assessment issues, the next step was to consider how to organize the vast amount of information in a way that seemed representative. Influenced by Bloom's (1956) taxonomy of learning domains and a suggestion by Randy Thrasher to include ethical issues in the pre-screening test and subsequent test for teachers, I opted to focus on these four content areas for the scales in this study: (1) terminology, (2) procedures, (3) test interpretation, and (4) assessment ethics. Table 4 summarizes the characteristics of the 100 preliminary items for the *Language Assessment Literacy Test - Preliminary Item Screening* which appears in Appendix A.

Table 4. *Characteristics of the Language Assessment Literacy Test - Preliminary Item Screening*

		Part I: Terminology	
question #	response format(s)	sample task(s)	sample topic(s)
Q1 - Q15	matching	match testing terms with appropriate symbols	sample variance, null hypothesis
Q16 - Q29	multiple choice	select the correct term for a concept described	exam types, variable types
Q30 - Q35	open response	explain or contrast various statistical terms	explain the central limit theorem
		Part II: Procedures	
Q36- Q40	short completion	specify the M and SD for 5 types of scores	
Q41- Q45	short completion	calculate & interpret basic statistics	calculate M, SD for a test
Q46- Q50	short completion	select & employ advanced statistics	determine effect size for two groups
Q51 - Q55	short completion	calculate five correlation statistics	determine the Pearson cor. index
Q56 - Q59	mostly open response	interpret pretest/posttest results	decide what learning occurred
		Part III: Test Interpretation	
Q51 - Q74	mostly open response	interpret published research	construct validity, accommodation
		Part IV: Assessment Ethics	
Q75 - Q100	multiple choice	select the most appropriate response for each problem	grading procedures, score reporting, handling ethical violations

(3) *Preliminary item screening*

Adopting a method suggested by Angoff (1971), a panel of six testing experts were asked to give their opinions about the appropriateness of the 100 possible assessment-related questions in Appendix A for the following populations:

1. professional test validators (Level A),
2. high school foreign language teachers (Level B), and
3. first year university education majors (Level C).

The panel of experts were all university professors or associate professors teaching EFL and/or testing. Four of the experts had PhDs and the other two were PhD candidates. All of the candidates were male and four were American, one was Japanese, and one was Dutch. An electronic version of the test was emailed to four of the informants and two of them received printed copies. Two of the expert informants did not complete the task. The raw data from the four informants that did appears in Appendix B.

The task that the experts were asked to perform was to simply decide which population, if any, they felt each item represented a "minimal competency" for. In other words, if they felt that a test developer should be able to solve a given item, but not a teacher or student they would check the response box for "A" to the right of that item – but leave other boxes unchecked. Let us illustrate this with a concrete example. The first question in the *Assessment Ethics* section (Part IV) of the preliminary item screening was –

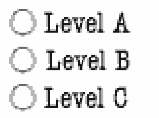
76. Which of the following is NOT an ethical means of helping students prepare for a school-wide test?

(A) Outlining the test format and rubrics in detail.

(B) Giving students a parallel form of the test.

(C) Pointing out one or two of the more difficult problems in that test.

(D) Mentioning basic statistical information about this test


○ Level A
○ Level B
○ Level C

Notice the three check boxes next to this question: they represent the three different population groups this might be an appropriate question for. An unchecked box could indicate two things: the expert informant did not feel it was within the "minimal required knowledge" for that population, or perhaps the informant was becoming tired and simply left the box blank. Concerning our specific case, two of the expert informants felt that test developers should be able to answer Q76. All four felt that teachers should be able to answer that question, and none felt that students needed to know the answer. This suggests that Q76 might be appropriate for an assessment literacy test for teachers, but not for students, and perhaps not for test developers either.

In addition to indicating which population they felt a question might be appropriate for, some informants also offered feedback on the tool itself. Two sorts of comments came out. First, a number of ambiguities about the test came to light. For example, in the original version Q73- Q75 asked respondents to mention any problematic points about a data table. The prompt did not specify how many problems needed to be mentioned, nor how the situation should be rectified. Based on informant feedback, the prompt was reconstructed with more detail.

A second type of feedback focused on validity issues. For example, the original version of Q58 asked respondents to conduct a pretest/posttest ANCOVA without mentioning what the continuous explanatory variable was or the factor under investigation. In other words, not enough information was provided to answer the question, so it was subsequently revised.

(4) *Level B test writing*

Tallying up the raw data in Appendix B provided a good idea about which items from Appendix A might work for an assessment literacy test geared for high school foreign language teachers. Items which received three or more recommendations were automatically included in the test in Appendix C; those receiving less than two recommendations were rejected. Those which received two recommendations were considered on a case-by-case basis. Table 5 summarizes the how the informants "voted" on the items in Appendix A for inclusion in an assessment literacy test for teachers.

Table 5. *Expert informant responses on the appropriateness of the items from Appendix A for an assessment literacy test for high school foreign language teachers*

4 recommendations in favor :
Q3, Q5, Q11, Q13-14, Q20, Q22-24, Q32, Q36, Q41-42, Q56, Q71, Q76-82, Q84-86, Q88-90, Q92, Q94-100 (36 items total)
3 recommendations in favor:
Q16, Q28, Q32, Q37, Q43, Q72-74, Q75, Q83, Q87, Q91 (12 items total)
2 recommendations in favor:
Q17, Q39-40, Q45, Q47, Q50 (6 items total)
1 recommendation in favor:
Q1, Q4, Q6, Q8, Q15, Q21, Q26, Q38, Q44, Q57-59, Q66 (13 items total)
No recommendations in favor:
Q2, Q7, Q9-10, Q12, Q18-19, Q25, Q27, Q29-31, Q33-35, Q46, Q48-49, Q51-55, Q60-65, Q67-70, Q93 (34 items total)

Based on this procedure, 48 items from Appendix A were adopted into the first version of the *Assessment Literacy Test for High School Foreign Language Teachers* in Appendix C and 47 were rejected. The remaining six items that had two votes were considered on a case-by-case basis.

Since the *Assessment Literacy Test for High School Foreign Language Teachers* has four sections and a test with too few items will tend to have a low reliability, it was necessary to add more items to the test. For the first part of the test, five new items were included. For the next part, seven new items were incorporated. The Interpretative Section had three new items and the final section did not require any new items. The subsequent test had 70 items in total. The format of that test is summarized in Table 6.

Table 6. *Characteristics of the Assessment Literacy Test for High School Foreign Language Teachers*

		Part I: Terminology	
question #	response format(s)	sample task(s)	sample topic(s)
Q1 - Q9	matching	match testing terms with appropriate symbols	sample variance, null hypothesis
Q10 - Q16	multiple choice	select the correct term for a concept described	exam types, variable types, cutoffs
Q17 - Q20	open response	explain or contrast various statistical terms	masters vs. non-masters
		Part II: Procedures	
Q21- Q25	short completion	calculate basic & interpret statistics	calculate mean & S.D. for a test
Q26- Q29	open response	interpret pre-test/posttest gains	assess whether "learning" occurred
Q30 - Q33	short completion	calculate three descriptive statistics	describe a box-plot and bell curve
Q34 - Q36	open response	think of three ways to increase validity	validity & reliability issues
		Part III: Test Interpretation	
Q37 - Q44	mostly open response	interpret tests and research	interpreting error of measurement
		Part IV: Assessment Ethics	
Q45 - Q56	multiple choice	select the most appropriate response for each question	grading procedures, reporting scores handling ethical violations
Q57 - Q70	mostly open response	identify an ethical problem and/or suggest a solution	grading procedures, confidentiality, dealing with test anxiety

Discussion

In the process of developing and pre-screening this test, several points became clear. First, I came to realize that assessment literacy test categories are not mutually exclusive. For example, Q29 in Appendix C asks respondents to decide on a cut-off point for a post-test. That involves issues of judgment and as well as possible statistical calculations. In short, many aspects of assessment are inter-related: ethics often impinge upon interpretation and statistical procedure use.

> " . . . *many aspects of assessment are inter-related: ethics often impinge upon interpretation and statistical procedure use*."

Another point clear from the inventory in Appendix C1 is that many items tend to focus on those aspects of assessment literacy which are easily-measurable. As a result, the Assessment Literacy Test for High School Foreign Language Teachers Inventory has a strong quantitative orientation and perhaps too many questions about statistics. These aspects can be measured in vitro through writing, but perhaps the most important forms of classroom assessment happen in vivo and informally. Moreover, if we look at the tentative operationalization of assessment literacy for teachers suggested in Table 2, it is clear that some areas are under-represented in the test in Appendix C. Specifically, items #6 and #11 are not sufficiently covered. This suggests that the test needs to be augmented in some areas (quite likely), or the operationalization of the concept needs to be worked out more (also likely), or both.

A final point which came out concerns the nature of relativism and universality in this test. Originally my objective was to write a assessment literacy test that could be used for any academic field. What I soon discovered was that it was necessary to narrow down the focus to language in general, and later foreign language in particular. Why? Practically speaking, how could a single person answer so many questions about dissimilar disciplines? How could I get in touch with expert informants from many diverse fields? As a result, the test in Appendix C is highly contextualized: it represents only the sort of questions high school foreign language teachers might have to contend with, but not so much the ones that math teachers, for example, would find important.

Conclusion

This paper has begun to explore the concept of assessment literacy. However, it leaves many questions unanswered and the instrument in Appendix C needs extensive trialing and refinement. The current version of the *Assessment Literacy Test for High School Foreign Language Teachers* is rough and more work is needed before it is widely adopted. When we reflect on how much most teachers know about testing and assessment, the need for some kind of scale like this to promote more familiarity with assessment concepts seems evident. In that sense, this paper represents a useful first step.

One major question pertaining to this study needs to be answered. If this research is going to have any impact, the issue of incentive must be addressed. What incentive is there for anyone to spend the hours needed to complete the instrument in Appendix C? Unless there is some kind of structured incentive is offer to complete a test, who would make the effort? Although graduate school students taking an assessment or research methods course might be inspired to go through the *Assessment Literacy Test for High School Foreign Language Teachers*, why should high school teachers who are already busy with other work go through this? Subsequent versions of the instrument must address the issue of rewards and incentives.

If we take the optimistic view that given the time and resources, most teachers will be motivated to improve their own assessment literacy skills, several suggestions are in order. Table 7 lists some ways that ordinary teachers can to become more literate about assessment.

Table 7. *Some suggested ways teachers can enhance their own "assessment literacy"*

1. When encountering an unfamiliar word about assessment, learn it and the concept behind it.
2. When constructing a test, try to beta-test it and revise it after examining examinee responses.
3. When grading a test, pay special attention to your cut-off points and questions that over 90% of the students answered correctly or didn't answer correctly.
4. When mentioning test scores to students, be sure to explain the key descriptive statistics in a way that the audience can understand.
5. When deciding how to grade a course, make sure it is both educationally valid and sufficiently clear to all stakeholders.
6. Don't be overly obsessed with formal testing: regularly seek to improve your micro-assessment and daily feedback skills.
7. When encountering an unethical assessment practice, consult with peers and try to think of the best way to rectify the problem.

This paper has suggested that assessment literacy is an important aspect of overall teacher development. All teachers wishing to develop professionally should also learn more about assessment.

Acknowledgement:

I am grateful to Kristie Sage and Peter Ross for their feedback on this article.
The limitations of this paper, however, are my responsibility.

References

Alkin, M.C. (Ed.). (1992). *Encyclopedia of educational research* (6th ed). New York: Macmillan Publishing Company.

ALTE. (1998). Multilingual glossary of language testing terms. *Studies in Language Testing 6*. Cambridge University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. Accessed on October 12, 2006 from http://www.project2061.org/publications/bsl/online/bolintro.htm

American Federation of Teachers, National Council on Measurement in Education, National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Accessed on August 12, 2006 from http://www.lib.muohio.edu/edpsych/stevens_stand.pdf

American Psychological Association Practice and Science Directorates (ed.). (2000). *Report of the Task Force on Test User Qualifications*. Accessed November 12, 2006 from http://www.apa.org/science/tuq.pdf.

Angoff, W. H. (1971). Scale norms and equivalent scores. In R. L. Thorndike (Ed.) *Educational Measurement, 2nd Edition*. (pp. 508-600) Washington D.C.: American Council of Education.

Baker, B. (n.d.) *Victim mentality*. Accessed December 12, 2006 from http://www.selfgrowth.com/articles/Baker4.html

Bloom B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay Company Inc. (republished in 1984 by Pearson Education).

Donoho, D. (2000, August 20). High-dimensional data analysis: The curses and blessings of dimensionality. Paper presented at the American Math Society 2000 Conference. Accessed on August 28, 2006 from http://www-stat.stanford.edu/~donoho/Lectures/AMS2000/Curses.pdf

Gunn, M. (1995). Criterion-based assessment: A classroom teacher's perspective. In G. Brindley (Ed.), *Language assessment in action*. Sydney: National Centre for English Language Teaching and Research, Macquarie University. Cited in Brindley, G. (1997, September). Assessment and the Language Teacher: Trends and Transitions. *The Language Teacher Online*. *21* (9). Accessed on January 28, 2007 from http://jalt-publications.org/tlt/files/97/sep/brindley.html

Hopkins, D. (1985). *A teacher's guide to classroom research*. Philadelphia: Open. University Press.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Accessed on November 2, 2006 from http://www.lib.muohio.edu/edpsych/stevens_stand.pdf

Machlup, F. (1984). *Knowledge: Its Creation, Distribution, and Economic Significance*. Princeton, NJ: Princeton University Press.

Nihon Tesuto Gakkai. (2006). *The JLTA list of bilingual testing terms / The JLTA code of good testing practice*. Nagano, Japan: Japan Language Testing Association.

MacBeath, F., & Galton, M. (2004, April). A life in secondary teaching: Finding time for learning. Accessed on October 29, 2006 from http://www.data.teachers.org.uk/resources/pdf/74626-MacBeath.pdf

Mertler, C. (2002). Classroom assessment literacy inventory. Accessed on May 21, 2006 from http://pareonline.net/htm/v8n22/cali.htm.

Mousavi, S. A. (2002). *Encyclopedic dictionary of language testing* (3rd Edition). Taipei: Tung Hua Book Company.

Ohata, K. (2005, December). Potential Sources of Anxiety for Japanese Learners of English: Preliminary Case Interviews with Five Japanese College Students in the U.S. *TESL-EJ, 9* (3) 1-21. Accessed on Dec 31, 2006 from http://www.kyoto-su.ac.jp/information/tesl-ej/ej35/a3.pdf

Openjuru, G. (2003). Adult literacy and development link: A perspective from a non-literate's literacy practices and environment. *Adult Education and Development,* 61. Accessed on April 31, 2006 from http://www.iiz-dvv.de/englisch/Publikationen/Ewb_ausgaben/61_2004/eng_georgeopenjuru.htm

Orrell, J. (2005). Assessment literacy: A precursor to improving the quality of assessment. Keynote presentation of the 2005 E*valuation and Assessment Conference*. University of Technology. Accessed on July 15, 2006 from http://www.iml.uts.edu.au/EAC2005/ speakers/OrrellKeynoteEAC2005.pdf

Parker, R. (1992). Process vs. product in language teacher education - Shifting the focus of course design. *ERIC Document ED369283*. Accessed on August 28, 2006 from http://www.eric.ed.gov/sitemap/html_0900000b80140c87.html

Sacks, P. (2000). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.

Séror, J. (2005). Computers and qualitative data analysis: Papers, pens, and highlighters vs. screen, mouse, and keyboard. *TESOL Quarterly, 39* (2), 321- 328.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications, Inc.

Sturgeon, J. (2006, August). A new kind of testing. *District Administration*. Accessed on October 22, 2006 from http://www2.districtadministration.com/viewarticle.aspx?articleid=573

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications, Inc.

Vittal, N. (2002). Enhancing human capital index: New approaches for the knowledge economy. Paper presented at the CII Seminar on March 14, 2002. New Delhi. Accessed on May 23, 2006 from http://cvc.nic.in/vscvc/cvcspeeches/sp15mar02.pdf

Wagner, D. A., & Kozma, R. (2003, Oct. 1). New technologies for literacy and adult education: A global perspective. University of Pennsylvania International Literacy Institute. Accessed on June 12, 2006 from http://ncal.literacy.upenn.edu/products/wagner_kozma.pdf

Wallman, K. (1993, March). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88* (421) 1- 8.

Yu, C. H. (n.d.). Educational Assessment. Accessed on August 12, 2006 from http://www.creative-wisdom.com/teaching/assessment/assessment.html.

Main Article	**Appendix A**	**Appendix B**	**Appendix C**

< Last Next >

http://www.jalt.org/pansig/2006/HTML/Newfields.htm (HTML)

http://www.jalt.org/pansig/PDF/Newfields.pdf (PDF)