

Engendering assessment literacy: Narrowing the gap between teachers and testers

by Tim Newfields (Toyo University Faculty of Economics)

Abstract

How do high school foreign language teachers conceptualize the notion of "assessment"? How do their views tend to differ from ways that this term often appears in the language testing literature? This paper explores these questions and considers possible ways to bridge the chasm between classroom instructors and professional test developers. The need for teachers to become more "assessment literate" and for test developers to think instructionally while writing items and communicate test data more effectively with stakeholders is underscored.

Keywords: assessment literacy, teacher development, assessment models, teacher training

If we jog our memories, it should not be difficult for most of us to recall what it felt like going through various tests as students. Did those exams seem instructionally illuminating? Was the evaluation process empowering? Moreover, when we recall our first attempts to create classroom tests, what images come to mind? What prompted the decisions you made to learn more about testing?

To some degree, teachers need to be time travelers in order to develop a dialog with various parts of themselves so that they can have a better dialog with others. Candidly speaking, is there a part of you that hates tests? Is there another part that feels they are inevitable? Often it is helpful to get in touch with the conflicting feelings we have about tests. Few things about testing are simple, as many of the ongoing debates going about testing attest.

This paper presents some preliminary findings about how a small group of high school English teachers in Japan feel about testing and the sort of assessment practices they employ. I should begin with some provisos. First, this is a qualitative case study with merely seven informants, so the extent that any of the findings can be generalized to the population of about 30,000 high school English teachers in Japan (MEXT, 2006) is questionable. This study merely offers a limited sketch of the views of a handful of Japanese secondary school foreign language instructors – it is by no means a complete picture. However, just as a likeness of a face can emerge with a few lines, perhaps some of the comments made by these informants will begin to reveal a larger picture.

Second, in the tradition of qualitative research, this paper openly acknowledges its own biases. It is not a neutral report – in my view neutrality is difficult when it comes to questions about values and perceived educational purposes. At the same time, this study adopts some standards of thick description (Geertz, 1973, pp. 3-30) for rigor. In this paper you should notice the voices of the informants emerging. Before doing that, however, it is necessary to briefly concede my prejudices about language testing.

Some Acknowledged Biases

Working as an EFL instructor in Japan for over twenty years, it seems hard not to notice the deep-seated impact of testing on university students. The majority of students that I encounter appear to lack any sense of ownership in what they are learning: education is something "done" to them – not something they actually "do". I agree with McVeigh (2001) and Nevara (2003, pp. 1-10) that passivity is the norm.

Although Tanezawa (2007, n.p.) has pointed out how most students are often adept at memorizing massive amounts of material just before a test, it is astonishing how quickly they generally forget it afterwards. Moreover, when it comes to applying English in real life contexts, very few seem capable of performing authentic tasks in the time frames that would be expected of lifelike situations. Perhaps the kind of skills needed to perform real tasks are for the most part different from those needed for multiple-choice tests. Or since many formal tests emphasize formal correctness rather than communication, perhaps most students see little incentive to communicate in a foreign language. There are so many factors influencing language learning that we do not know about. There is still so much about language acquisition that we do not know.

I firmly believe that most so-called language tests are only scratching the surface when it comes to measuring language ability. As Douglas (2001) points out, since we still do not have a wholly satisfactory model of language or communication, it is hard to develop appropriate constructs, frameworks, and test specifications. Partly for that reason, I concur with Widdowson (2000) in asserting that language teaching (and testing) is many ways more of an art than a science. That does not mean that we should abandon empirical research, but we should be humble about the extent that is actually known. At this point in time, our questions seem bigger than our answers. And in some ways, that is what makes language teaching and testing an interesting field.

I do want to openly acknowledge my misgivings about applying large-scale testing practices to small-scale classrooms. Should, for example, a TOEIC® score determine what university a student can enter (Three Sisters Inc., 2007)? Or should a class supposedly devoted to "English listening" focus almost exclusively on multiple-choice tasks (cf. Chen, 2004, pp. 544 - 555)? These questions are not merely conjectural: they reflect actual cases of what I regard as test misuse and negative test washback.

I am not alone in voicing my concern about the ways that tests adversely impact learning on in expressing concerns about adopting large norm-reference tests for small classes. Popham (2001) has suggested that when testing drives the curriculum several disturbing things become more prevalent. First, the curricular focus tends to narrow. Second, the prevalence of skill-and-drill instructional activities increases. Third, students with low test scores become more subject to various forms of prejudice.

Moss (2003, cited in Fulcher & Davidson, 2006, pp. 193 - 202) provides a particularly insightful analysis of the ways that large-scale testing paradigms do not fit most in classroom contexts. Some of the dichotomies that she highlights appear in Table 1.

Table 1. *Some contrasts between humanistic and mechanistic testing paradigms*

Humanistic testing paradigm	Mechanistic testing paradigm
Often willing to consider complex questions with more than one "correct" answer.	Generally only one single correct answer and preference for simple "right-or-wrong" formats.
Preference for holistic information and multiple task types to enhance student interest.	Preference for isolated information which is ideally context-free and with just a few task-types.
Human factors and emotional responses are deemed a relevant part of learning.	Human factors and emotional responses are considered extraneous variables that are irrelevant.
Generally adaptive and variable.	Generally fixed and non-adaptive.
Social collaboration and team-play is a valued skill.	Usually only individual performance counts – "collaboration" is often regarded as cheating.
"Effort" is often considered significant.	Effort is irrelevant – only performance counts.
Feedback is often informal and verbose.	Feedback generally consists of formal grades or scores.
Validity through qualitative consensus.	Validity through quantitative statistical procedures and perceptions of beneficial washback and social impact of a test is seldom measured.

As Czajkowski and Montague (2005) posit, at many schools we are witnessing a clash of these two diverse paradigms. Though the paradigms above might appear to be mutually incompatible, efforts to achieve some sort of creative compromise are not hard to find. The interviews mentioned in this paper suggest some teachers make valiant attempts at humanistic teaching/testing despite the mechanistic paradigms they are compelled to work in. Particularly in primary education in Japan at least, a concern for humanistic teaching/testing seems prevalent. However, as students seek to enter competitive schools, teachers often feel compelled to switch their testing practices in order to

accommodate the mechanistic testing paradigm described in Table 1. When this happens, teachers could be aptly described as "teaching machines" rather than humane mentors.

What does test-driven education do to students? Forthcoming research by Pan suggests that they are savvy when it comes to taking exams, but cynical about the entire educational process.

Despite my misgiving about test misuse, I would be less than honest if I did not acknowledge that I am part of the problem. I have authored many test-driven educational materials over the years and made some lousy tests. Since one of the themes of this conference is bridging borders, we as teachers should also consider ways of bridging gaps between our beliefs and practice.

Research Questions

Seven basic questions are explored in this study:

1. What did the respondents formally learn about assessment prior to teaching?
2. What testing-related questions were in their teacher certification examinations?
3. What percentage of the respondents' work time is generally devoted to assessment?
4. How have the respondents' ideas about assessment changed since they began teaching?
5. What are the main criteria the respondents use for determining student grades?
6. How prevalent do the respondents believe text anxiety is among students?
7. What concerns (if any) do the respondents have about the English tests they encounter?

Methods

In 2007 I approached twelve teachers for interviews, but five refused because by saying they "did not know enough about testing" or they were simply too busy. Seven Japanese EFL instructors therefore constituted the respondents for this study. One thing the sampling process taught me is that many high school teachers are not comfortable talking in detail with outsiders about how they teach and test.

The fact that two of the informants preferred to conduct the interviews away from school is also noteworthy. Unlike university professors who generally have their own private rooms, most high school teachers in Japan have a small desk space and personal privacy at work is limited. Two informants said that they really couldn't say what they were thinking with so many ears around, so we went to a off-campus coffee shop. Information about the informants who did consent to be interviewed appears in Table 2.

Table 2. *A profile of the Japanese high school EFL teacher respondents used in this study.*

#	Gender	HS Teaching Experience	Language of Interview	Educational Background	Current posting
1	M	10+ yrs.	English	Ph.D. candidate	High-rank public HS in Tokyo
2	M	20 yrs.	Japanese	MA degree	High-rank private HS in Tokyo
3	F	12 yrs.	English	BA & some grad course	Mid-rank public HS in central Japan
4	M	25+ yrs.	English	BA & 3 MA degrees	Mid-rank public HS in central Japan
5	F	13 yrs.	English	MA degree	Mid-rank private HS near Tokyo
6	F	11 yrs.	Japanese	BA degree	Low-rank public HS in Tokyo
7	F	24 yrs.	Japanese	BA degree	Low-rank private HS in central Japan

This composition differs slightly from the profile of Japanese teachers suggested by Nohara (1997). Even though 60% of Japanese teachers nationwide are thought to be male, less than half the informants in my study were male. Nohara suggests most Japanese teachers are "about 40". I did not ask the age of my informants, but I suspect the median was closer to 45. Although the average Japanese teacher has about 15 years experience, the average for this sample was slightly higher.

After each respondent consented to be interviewed, I e-mailed them a list of the preliminary questions in [Appendix B](#). A short research statement was included along with those questions and all respondents (except for #4, who was a spontaneous snowball case) had several days to consider their responses. At the onset of the interview, permission to digitally record the conversation was obtained and standard procedures regarding confidentiality were described.

During the actual interviews, the preliminary questions were raised first. Depending on the answers and time available, additional exploratory questions were then asked. The interviews, all of which were conducted in the spring of 2007, averaged about 25 minutes, though the sessions with informants #3 and #4 were nearly an hour.

Results

Let us now explore the results of the preliminary interview questions.

(1) Pre-teaching knowledge of assessment

Four of the seven respondents did not recall taking any special courses on testing or statistics before becoming teachers. It does not seem that assessment or testing is a mandated course of study for teacher trainees in Japan. Two of the informants recalled going through some sort of statistics class while obtaining their MA degrees. Informant #1, who had the most formal training of all persons in this study, studied statistics both at the MA level and during his Ph.D. studies.

(2) Teacher certification exam questions

Only three of the informants recollected taking a teacher certification exam. None were able to recall any questions in their exams related to ways to systematically assess or test students. Informants #2, 3, 5 indicated that they obtained their certification papers without having to go through any formal exams because of their educational backgrounds. One informant said that her examination consisted of an oral interview and no questions about testing, just her general teaching philosophy, attitude towards discipline, and willingness to coach sports.

(3) Time on assessment-related activities

Only one informant was able to offer any sort of numerical estimate about how much time was spent on testing overall. Informant #6 drew a heavy breath and said, "Well, ahh, maybe 10%." Other informants indicated that the time devoted to testing varied considerably with the "season" (*jiki*) and noted how during mid-term and final exam periods a lot of time was spent on testing and test preparation. Informant #4 wryly indicated that he spent "as little time as possible" on testing because he felt that formal tests did not foster deep learning. "We have to do a certain amount of testing," he conceded, "but students generally do not like it and neither do I."

(4) Changes in attitudes about assessment

A point that Informant #3 emphasized is that there is a ranking system of high schools in each prefecture (*hensachi seido*) in Japan. In her view, what teachers can expect or do with students in high-rank schools is quite different from what is deemed possible in institutions with a low-rank. She indicated that in many public schools

teachers are transferred every 4-6 years, often to a school with a different rank. In her words, it takes a few semesters to get used to each school's conditions. "In top schools, you are often teaching kids smarter than you are," she added, "and they will learn complex things very quickly." She then pointed out how different the picture is in schools that students who do not do well on the admission tests go to. "Kids feel discouraged and discipline is often a problem there." She then stressed a truism many teachers have heard in various ways: what works well in one context with one group of students might not work well in a different context with others.

Informants #2, 6, and 7 mentioned that their personal attitudes towards assessment were not so important: a school-wide (or district-wide) system for assessing students was in place and uniformity was a highly valued feature of the system. For them, learning about "assessment" essentially meant becoming familiar with the rules of a particular school system. In a way quite like British Common Law, historical precedent seems to carry more weight than abstract theory. "Younger teachers follow the direction of older staff" Informant #6 added, ". . . and for the most part, history carries on".

Informant #5 was more shrewd in this regard and remarked, "We defer to authority . . . and apologize to profusely [to superiors] if there is ever any complaint . . . [but] pretty much do what we want [to do]." A man who has achieved a modicum of success in the school hierarchy, he suggested that the best way to maximize freedom is to appear outwardly as a conformist and "invisible" to a degree. He said that in some ways, teachers have to be *ninja* (stealthy warriors) and suggested that teachers in Japan also need to be adept at ritual play (*tatemaie*) while concealing their real intentions (*honne*). Successful teachers, in his view, must have a good capacity for *donkan-ryoku*. That term literally means "power of insensitivity" in English, but I suspect that the expression "thick skin" comes closest to conveying the nuance. In 2001 a book about the importance of "thick skin" by Watanabe Junichi was a best-seller in Japan, but the extent that this informant was influenced by that book is unknown.

(5) Student grading criteria

All of the respondents indicated clearly that decisions about grading criteria were set collectively by either the prefectural board of education or, in the case of private institutions, the school itself. In both cases, there are suggested guidelines by the central Ministry of Education, Culture, Sports, Science and Technology. The prefectural boards of education are expected to adhere to the general parameters of the central government, but they also have some leeway in interpreting policy. As a case in point, Informant #1 indicated that 25% of the grade in his oral communication class was determined by "student performance", 25% by performance on the mid-term test, and 50% by performance on the final grade. Four of the other interviewees also considered "student performance" as an important factor in grading, but when it came to operationalizing this concept and pinning down exactly how student performance was measured, the responses were vague. "You just have to know the students," Informant #7 said.

Other teachers mentioned slightly different formula, but in all cases but for one private school in Tokyo paper-and-pencil tests provided a significant portion of student grades. That private school had special "oral English classes. For those classes, an oral interview testing procedure was used and recently they adopted the GTEC test

developed by Benesse Inc. and Berlitz International to serve as an external benchmark for the school's internal grading system. Reputed to be a measure of "genuine English proficiency for today's business environment" (Benesse, n.d.), I find it alarming and fascinating to see the way that this test is being marketed. No reliability studies for this test are available to the public – but there is a testimonial from Tully's Coffee Inc. on their website.

(6) Test anxiety among students

Informant #3 remarked that the "bottom 10%" of the students she deals with are generally not anxious about tests. By the time they reach high school, most have already given up on traditional academic success. Many in fact consider themselves slow learners who are less gifted academically than the general population. Informant #5 remarked that "success" with such students simply entailed helping them obtain enough graduation credits so that they could begin the next phase of their lives. Informant #2 remarked that test anxiety was seldom a problem with the students at the top because ". . . they know that they are smart". In his view anxiety seems to be highest among students in the middle. They are no so bright as to be able to ace their tests, yet not so indifferent as not care whether they pass or fail. Informant #5 indicated that her students were apt to be more anxious about speaking and listening exams than reading or writing exams. In short, responses to this question were considerably mixed.

(7) Current concerns about testing

Informants #3, 5, and 6 felt that the emphasis on the Japanese National University Entrance Exam was more detrimental than beneficial for students. In light of the validity questions about that oral section of exam raised by Sage and Tanaka (2006, pp. 74-98), their concerns are worth noting. Informant #7 raised the question whether or not Japan's language testing practices were still relevant in light of the nation's dwindling population. "It is getting easier for almost anyone to go to college now," she added, ". . . but many students are sort of tired of school . . . [and] do not want to do anything that involves too much effort." That informant went on to describe how tests might motivate some students to study, but that many students are essentially indifferent to their test scores. She was frankly at a loss when it came to the question of how to meaningfully motivate students to study.

Discussion

Perhaps the best way to think of this research project is as a preliminary case study which touches upon various areas of interest, but due to methodological limitations it would not be wise to make too many generalizations about the population of high school teachers as a whole. In particular, these questions are worth reflecting on:

1. How reliable are self-reported accounts? Do they reflect what teachers actually do? As Yu (2007) and Brown (2004) have suggested, self-reports are by no means an accurate picture of what teachers are actually doing or

even thinking. Many self-reports are subject to a Hawthorne effect in which interviewer expectations subtly influence what respondents say. Skilled interviewers might be able to conceal some of their biases, but some interviewees are sharp enough to pick up on non-verbal gestures which are difficult to mask.

2. How consistent are teacher behaviors? Even if teachers actually do some of the things they report, their behavior may change over time and with teaching circumstances. This paper suggests that some informants adapt some of their testing behaviors each time they are transferred to a new school. Whereas intensive testing might be a feature of many high schools with a sizable number of college applicants, this is seldom the case at schools at the other end of the spectrum.

These points are mentioned because readers have good reasons to be skeptical about the extent that any statements made by respondents about actual their actual beliefs or practices. Despite these limitations, it might be possible to cautiously venture the following three tentative generalizations:

1. At the all of the high schools where the interviewees worked, grading and testing procedures seem to be largely informed by past precedent rather than theory. Historical reverse engineering is the main approach to test construction at Japanese high schools: teachers simply look at what has existed before and build a new test based on previous material without much critical reflection on what the model is actually measuring. Though periodic lip service is devoted to discussions of reform, for the most part stability and tradition are highly esteemed, a point Cooper (1997) highlights in her overview of life in a Japanese school.
2. On the whole, it is probably safe to say that there is little, if any, attempt by high school teachers to conduct detailed any statistical analyses of their tests. Most interviewees seem to have neither the time, ability, nor inclination to go through rigorous validation procedures for their tests. When it comes to school entrance exams, four of the seven informants indicated that the opinions of noted cram schools (*juku*) about those tests seem to be more important than quantitative validation procedures. There is seldom, if ever, any detailed item analysis of a test and only one informant knew about Rasch analysis.
3. There is sometimes an interesting gap between official policy and actual testing practice. For example, Informant #1 commented that teachers will often allow students to pass if they "make effort". Officially, passing grades are based on student performance. Culturally, however, many Japanese seem to believe that "effort" (*yaruki*) is a facet that should be graded (Wu, 1999). In fact, many teachers in Japan I have encountered believe that grading itself is not solely an intellectual process – it is a holistic and often intuitive assessment of how given student compares with others in the class. This sort of view of assessment is admittedly fuzzy and at odds with the quantitative test paradigms.

In this regard, Informant #5 suggested that Japanese high school teachers are subject to two competing pressures: (1) a need to conform and test/teach in ways that bureaucrats want them to, and (2) a desire test/teach according to their core beliefs. Some teachers appear to align their core beliefs to the system effectively and experience minimal conflict. The comment by Informant #4 that teachers sometimes need to be like "stealth warriors" seems to be recognized by many teachers in Japan. Older teachers in particular seem to know it is fruitless to outwardly fight the system or exhibit too much overt insubordination, so seek subtle ways to work within the system and still bring some of their core beliefs into practice.

Conclusion

One theme this paper has emphasizes is the significant gap between the way assessment is described in the language testing literature and the way the limited number of informants in this study conceive of that term. As more and more sub-fields within testing field arise and the responsibilities of high school teachers in Japan which are unrelated to academic assessment increase, the gap seems likely widen. Two questions worth addressing at this point are "What factors seem to promote assessment literacy among educators?" and "What factors seem to inhibit it?" Before doing so, however, let us briefly define what is meant by assessment literacy. In Newfields (2006, p. 49) I described how this term has quite different meanings for students, teachers, and professional test developers. However, this description by Stiggins (1999) of what so-called "assessment literate" teachers should be able to is illuminating. Stiggins claims that assessment literate teachers should be able to:

- Understand what assessment methods to use, and when to use them in order to gather dependable information about student achievement;
- Communicate assessment results effectively to all intended users – including principals, other teachers, parents and students – whether using report card grades, test scores, portfolios, or conferences; and
- Understand how to use assessment to maximize student motivation and learning by involving students as full partners in assessment, record-keeping, and communication.

Let's explore both questions in detail.

Factors promoting assessment literacy

The one informant in this survey who knew the most about assessment was pursuing his doctorate degree in education. In his view, formal study does expose teachers to new ideas that they otherwise would not encounter. The recent Japanese Ministry of Education, Culture, Sports, Science and Technology directive requiring teachers have to take additional courses in order to become recertified every ten years might stimulate more interest in assessment (Japan Echo, 2007).

There are dozens, if not hundreds, of minor academic journals for teachers in Japan and associations for English teachers. Publications such as *Eigo Kyouiku* [English Education] and *Shin-Eigo Kyouiku* [The New English Classroom] attract a devoted number of readers. The teachers who write up their classroom research for those journals or make presentations about it are often seem become the most assessment literate. If we accept a constructionist theory of learning (Brown, et al., 1989) with respect to assessment literacy, this makes sense. Assessment literacy is not something one can simply acquire from reading one or two books – it is probably a lifelong process involving an ongoing and evolving dialog between empirical research and theory. Actively writing about classroom research often encourages or forces a writer to become more assessment literate. Whereas prestigious journals can afford to flatly reject articles with poor statistics or sloppy designs, with some smaller journals a mentoring process is more likely to take place. Though standards vary widely from journal to journal, one way some teachers learn about testing is through a process of mentoring when editorial advisor work over their manuscripts. This conjecture certainly requires further corroborating research.

Factors hindering assessment literacy

As Fulcher (1991) states, many teachers feel that assessment and testing is not particularly relevant to their small classroom practice. There is a lot of indifference, if not antipathy, towards testing among a good deal of teachers. Many frankly ask, "What is the pay off for taking all of the time needed to learn about testing and evaluation?" Many high school teachers in Japan I have spoken with openly wonder whether the effort to learn about testing and assessment is worth the reward. Until assessment literacy becomes a required course for teacher training, and until teachers who construct school entrance exams are required to be familiar with fundamental testing principles and ethics such as those embodied in *The JLTA Code of Good Testing Practice* (2007) many teachers will probably continue to feel this way about testing and assessment.

Specific recommendations

Let me now shift to some specific recommendations about ways that testing organizations and test developers – and then finally teachers themselves – can promote assessment literacy.

What testing organizations and test developers can do

1. Be more cautious about test claims. All too often, in-house testing literature seems like a thinly veiled advertising. Unit B6 in the most recent publication of Fulcher and Davidson (2006, pp. 230-248) is a good example of this. The teacher verification study by four researchers for ETS is basically as rough as this paper is – but instead of adequately acknowledging the limitations of their research, the authors gloss them over. As long as the vast majority of end users are not assessment literate, this trend will probably continue. In my

opinion, test developers need to be much more careful about stating what their tests actually do or do not measure.

2. Offer some interpretative scaffolding when describing data. In many academic journals detailed statistics about a test are offered without adequate interpretation. The test results might be clear to a few specialists in the field, but opaque to most readers. For example, a recent issue of the *Japan Language Testing Association Journal* mentions the Thurstone Index (Nakamura, 2002, p. 63). A golden opportunity to increase the assessment literacy of the readers was lost because that term was not explained. It is good to remember that nearly every publication has a range of readers: from novices in the field to illuminati. If no interpretative scaffolding for complex information is offered, it is all too likely that most readers will lose the point. Without lowering the academic standards of any publication, I believe academic journals can and should be more mindful of their readership. Many of us can think of teachers who have put down testing journals because they were unduly daunted by the language and/or statistical procedures.

2. On the topic of journals, I strongly believe that many testing journals need to become less elitist and more open. Many academic journals cost a good portion of a month's salary to receive for those of us in Asia. As a case in point, the journal *Language Testing* costs over US\$560 a year (Sage, 2007). That amounts to more than 500,000 Korean won – a fee which is frankly outrageous. Online access to that publication is just under US\$520 or approximately 440,000 Korean won. Is it any surprise that many teachers in Asia have no access to such journals? One can't help but wonder what the mission of such journals actually is.

What teachers can do

Experts who know far more than I do about testing have mentioned ways that teachers can become more assessment literate. JD Brown (2006, pp. 21-26) provides an extensive and recently updated list of testing resources about testing. The JALT Testing and Evaluation SIG webpage, online at jalt.org/test/pub.htm, also has about a hundred articles and dozens of links about testing and assessment. To this vast chorus of information, I offer just three simple suggestions for teachers to become more assessment literate:

1. Know – and communicate – your own strengths and weaknesses clearly to not only to your students, but also to other relevant stakeholders. When it comes to evaluating holistic student performance, I believe informal teacher assessment methods such as those described by Navarete, Wilde, Nelson, Martínez, and Hargett (1999) are generally far superior to narrow, analytic tests. The way that skilled teachers can adapt to student needs as they arise and integrate diverse teaching/testing modalities still exceeds even the best computer-adaptive tests created so far.

2. Seek transparency and communicate your evaluation criteria as clearly as possible to all relevant stakeholders. All of us can remember being in classes in which we had little idea what the teacher was saying, let alone how the grading was done. Others can remember being in classes in which the evaluation criteria seemed to change midway through the course. Neither of these should happen if the evaluation standards are sufficiently transparent and the curriculum fits the needs of students.
3. Give more power to students and, within appropriate parameters, have them assess themselves, their peers, and their instructors in ways suggested by Sloat (2006). Why should assessment be a one-way process that only teachers do to students (or school systems do to teachers)? We should think of assessment as at least a two-way communication process.

If we fail to do our jobs communicating and empowering students, I am afraid that more McDonaldized testing which ignores many of the core principles in the ILTA Code of Practice (2006) is close at hand. That means more money for big testing conglomerates, but less dignity and power for most students and teachers.

Acknowledgement

*I am grateful to Kristie Sage for her comments on this article.
However, the responsibility of its short-comings is entirely mine.*

References

- Benesse Corporation. (n.d.). *GTEC - Online Test for Measurement of Practical English Proficiency*. Retrieved on Feb 6, 2008 <http://www.benesse.co.jp/gtec/english/about/>
- Brown, J. D. (2006). Resources available in language testing. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 10 (1) 21 - 26. Retrieved on July 29, 2007 http://jalt.org/test/bro_23.htm
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32 - 42. Retrieved on August 14, 2007 <http://www.exploratorium.edu/IFI/resources/museumeducation/situated.html>
- Brown, L. (2004). *Observational Field Research*. Retrieved August 15, 2007 from <http://www.socialresearchmethods.net/tutorial/Brown/lauratp.htm>
- Cheng, H. F. (2004). A comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals*, 37 (4) 544 - 555.
- Cooper, A. (1997). *Japanese lessons: A year in a Japanese school through the eyes of an American anthropologist and her children*. New York: New York University Press.
- Czajkowski, T. & Montague, M. Preparing for Change: Standardized Tests or Authentic Assessment? In A. L. Costa, , & B. Kallick, (Eds.). (1995). *Assessment in the Learning Organization: Shifting the Paradigm*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Douglas, D. (2001, July 1). *Language for specific purpose testing: The state of the art*. Presentation to the Japan Language Testing Association. Tokyo: Tokyo University of Economics.

- Fulcher, G. (1991). The role of assessment by teachers in schools. In T. Caudery (Ed.) *New Thinking in TEFL*. The Dolphin Series No. 21. Aarhus: Aarhus University Press, 138 - 158.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz, (Ed.) *The Interpretation of Cultures: Selected Essays*. New York: Basic Books, 3-30.
- ILTA. (2006). *Code of Practice [Draft Version]*. Retrieved on August 14, 2007 from <http://www.iltaonline.com/ILTA-COP-ver3-21Jun2006.pdf>
- Japan Echo Inc. (2007, April). *Controversy over Japanese education*. Retrieved on July 30, 2007 <http://www.japanecho.co.jp/sum/2007/340203.html>
- Japan Language Testing Association. (2007). *The JLTA Code of Good Testing Practice*. Retrieved on July 28, 2007 <http://www.avis.ne.jp/~youichi/COP.html>
- McVeigh, B. (2001, October). Higher education, apathy and post-meritocracy. *The Language Teacher*, 25 (10). Retrieved from <http://www.jalt-publications.org/tlt/articles/2001/10/mcveigh>
- Mellon, C. A. (1990). *Naturalistic inquiry for library science: methods and applications for research, evaluation, and teaching*. New York: Greenwood. Cited in Olson, H. (1995). Quantitative "versus" qualitative research: The wrong question. Retrieved on July 28, 2007 <http://www.ualberta.ca/dept/slis/cais/olson.htm>
- MEXT (2007). *Japan's education at a glance: 2006*. Retrieved on July 26, 2007 <http://www.mext.go.jp/english/statist/07070310.htm>
- Moss, P. (2003). Reconceptualizing validity for classroom assessment. Originally published in *Educational Measurement: Issues and Practice*. 22, 4, 13-25. [Abridged version in Fulcher, G. & Davidson, F. (eds.). *Language testing and assessment: An advanced resource book..* (2006). Abingdon, Oxfordshire & New York: Routledge. 193-202.]
- Nakamura, Y. (2002). Effectiveness of paired rating in the assessment of English compositions. *JLTA Journal*, 5 61-71.
- Navarete, C., Wilde, J., Nelson, C., Martínez, R., Hargett, G. (1990). *Informal assessment in educational evaluation: Implications for bilingual education programs*. NCBE Program Information Guide Series (3). Retrieved on August 16, 2007 <http://www.ncela.gwu.edu/pubs/pigs/pig3.htm>
- National Council on Ethics in Human Research. (2000). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans - Section 3*. Ottawa, Canada: National Council on Ethics in Human Research. Retrieved on March 29, 2007 http://www.ncehr-cnerh.org/english/code_2/sec03.html#1
- New English Teachers' Association. (Publisher). *Shin-Eigo Kyoiku: Subscription information*. Retrieved on July 30, 2007 at <http://www.shiramizu.org/~sineiken/activity/magazine.htm>
- Newfields, T. (2002). Challenging the notion of face validity. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 6 (3) 19. Retrieved on July 28, 2007 http://jalt.org/test/new_2.htm
- Newfields, T. (2006). Teacher development and assessment literacy. In T. Newfields, et al (Eds.) *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). Retrieved on July 30, 2007 <http://jalt.org/pansig/2006/HTML/Newfields.htm>
- Newfields, T. (2007) *Crassroom Voices: Poetry, Art, and Dialogs about Education*. Retrieved on July 26, 2007 <http://www.tnewfields.info/CrassroomVoices/>
- Nevara, J. (2003). Teaching English in Japan to Chinese Students. *Asian EFL Journal*. 5 (3) 1-10. www.asian-efl-journal.com/sept_03_sub1.pdf
- Nohara, D. (1997). *The training year: Teacher induction in Japan*. Retrieved on July 30, 2007 <http://www.ed.gov/pubs/APEC/ch4.html>

- Pan, Y. C. (forthcoming). Consequences of test use from students' perspectives. *Proceedings of the June 30 – July 11, 2008 LingFest 2008 Conference*. University of Sydney, Australia.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. New York: Da Capo Press (Perseus Books).
- Sage, K. & Tanaka, N. (2006). So what are we listening for? A comparison of the English listening constructs in the Japanese National Centre Test and TOEFL® iBT. In T. Newfields, et al (Eds.) *Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference*. May 13-14, 2006. Shizuoka, Japan: Tokai University College of Marine Science. (p. 74 - 98). Retrieved on July 30, 2007 <http://jalt.org/pansig/2006/HTML/SageTanaka.htm>
- Sage Publications. (2007). *2007 price list*. Retrieved on July 30, 2007 <http://www.sagepub.co.uk/subscAgents.nav>
- Sloat, J. M. (2006, Aug). *Balancing the power: An egalitarian model for course evaluation and assessment*. Paper presented at the annual meeting of the American Political Science Association. Abstract retrieved from http://www.allacademic.com/meta/p151019_index.html
- Stiggins, R. J. (1999). Learning teams can help educators develop their collective assessment literacy. *Journal of Staff Development*, 20 (3). Retrieved on August 14, 2007 <http://www.nsd.org/library/publications/jsd/stiggins203.cfm>
- Taishukan Shoten. (Publisher). *Eigo Kyoiku: Subscription information*. Retrieved on August 6, 2007 <http://www.taishukan.co.jp/magazine/magazine.html>
- Tanezawa, J. (2007, August 2). *Shibouko wo kaeta hou ga ii kamoshirenai yo*. [Isn't it high time to change dead schools?]. Retrieved on August 14, 2007 <http://www.juken-goukaku.jp/?gclid=CLGHge3B9I0CFRV6IgodVxPyMw>
- Three Sisters Inc. (2007). Daigaku Nyuushi de TOEIC wo katsuyo suru 140 no daigaku. [140 universities using TOEIC for entrance exams]. Retrieved on August 14, 2007 <http://www.toEICclub.net/univ.html>
- Yu, C. (2007). *Reliability of self-report data*. Retrieved on July 30, 2007 <http://www.creative-wisdom.com/teaching/WBI/memory.shtml>
- Watanabe, J. (2001). *Donkanryoku*. [The Power of Insensitivity]. Tokyo: Shuseisha.
- Widdowson, H. (2000). TESOL: Art and craft. *Journal of the Imagination in Language Learning*, 5. Retrieved on July 30, 2007 <http://www.njcu.edu/CILL/vol5/widdowson.html>
- Wu, A. (1999, January). *The Japanese education system: A case study summary and analysis*. Retrieved on August 17, 2007 <http://www.ed.gov/pubs/ResearchToday/98-3038.html>

This article may be viewed online at:

HTML: <http://tnewfields.info/Articles/testlit.htm> / **PDF:** <http://tnewfields.info/Articles/PDF/Newfields-Testlit.pdf>
