

# Language Assessment Literacy in Theory and Practice

Daniel DUNKLEY

## Abstract

After examining several definitions of assessment literacy, or basic knowledge of testing, the testing awareness needs of several different groups are discussed. Firstly, what do teachers need to know about testing, and how is this knowledge best delivered to them? Then we observe the needs of stakeholders who are not directly involved in education, namely, policy makers, followed by those of university administrators. Finally the situation in university teaching is presented, and recommendations for improving assessment literacy are suggested.

The term literacy has broadened considerably in its scope over the last 100 years. Initially of course it means the ability to read and write, a skill which for many centuries was the exclusive preserve of social elites. In a fairly static society and before the wide availability of texts with the invention of printing, illiteracy was hardly a handicap for the mass of the population. However, with the increasing complexity of society and the resulting need for gaining knowledge by reading and writing, individuals were motivated to make the effort to become literate, and finally in advanced countries the state enforced literacy by providing free compulsory education. Naturally this worked two ways: people were able to advance socially through their knowledge, but the

state's power to influence citizens' thought and behavior was greatly increased by the rise of mass newspapers. However, the metaphor of literacy as a positive, beneficial skill has remained, and it has come to be a metaphor for the minimum level of ability required in a certain area for an individual to function competently in society. Thus the terms *technological literacy* means the ability to operate common domestic machines such as vacuum cleaners or telephones, and more recently *computer literacy* refers to a similar familiarity with personal computers. Equally the term has been used of more abstract skills such as cultural literacy or social literacy.

While literacy is thus a metaphor for minimum competence, it must be noted that literacy is not simply a binary concept. Just as with spoken language, each user of written language only has access to a subset of the complete repertoire of grammar and vocabulary. This becomes apparent when one glances at a specialist magazine, clearly written by people with the same common basic literacy as the general reader, but who have, for professional reasons, acquired a further specialist literacy. Thus literacy is best imagined as a continuum of skill, where the lower end is occupied by those who find it hard to function smoothly in society, the middle part is the mass of the population, while the top third possess extra literacy appropriate to their profession, which depending on the case may be largely irrelevant to those in different occupations.

When we transfer the term literacy to assessment or testing in education, an initial definition seems obvious: the minimum knowledge to be able to test effectively. But there are difficulties to add: Who needs this literacy? Should there be some system of mandatory enforcement? Is it just knowledge, or practical skill? Who decides who needs what degree of literacy? The more one considers assessment literacy the more difficult it is to devise a practically useful definition. As Inbar-Lourie (2013, 304) puts it: "Language assessment literacy is a multilayered entity, and ... defining it presents a major challenge."

As for a history of the term Assessment Literacy, its first coining is ascribed to Stiggins over twenty years ago (Stiggins, R. J. (1991)), referring to the training of language teachers. However, the scope of Assessment Literacy has broadened considerably, as it was realized that more and more categories of people were affected by testing, and as tests became more widespread in scope and more globalized.

It is not only teachers who are concerned with testing, but naturally students and their families, then those who use test results such as university admissions officers or employers. Finally, anyone in society who has the power to set minimum standards in any field will probably have to decide standards on the basis of tests, whether for example in the field of immigration or professions such as medicine or banking. Thus few members of society are immune from testing, and many must deal with tests as an ancillary skill to their main field of competence and responsibility.

As a result of this expansion of the number of people connected with testing, more recent definitions have tended to be more nuanced and necessarily wordy. Thus we read Fulcher's (2012, 125) definition: "The knowledge, skills and abilities required to design, develop maintain or evaluate large-scale standardized and/or classroom based tests, familiarity with test processes and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice." This seems to be, rather than a description of the average language teacher's competence, a taxonomy of skills possessed by few other than language testing researchers. If this is taken as a statement of basic competence, then we seem to be condemning the vast majority of people concerned with testing to illiteracy or incompetence.

On the other hand, in contrast to this rather comprehensive and demanding definition, it has been pointed out that different people need different degrees of literacy. For example Taylor (2009, 25), an exam administrator and researcher,

writes “an appropriate level of assessment literacy needs to be nurtured not just among technicians ... or even among language teachers ... but much more broadly in the public domain if a better understanding of the function and values of assessment tools and their outcomes are to be realized across society.” This concept of different levels for various needs has been systematized in relation to scientific literacy by Bybee (1997), who suggests a scale of five stages, from *illiteracy*, through the lower stage of *nominal literacy*, the middle stages of *functional literacy* and *conceptual literacy* to the highest level of *multidimensional literacy*. Moving from theory to practice in assessment literacy, what these levels might mean for the various stakeholders in language testing will emerge from an analysis of several research projects and practical implementations.

### **Assessment literacy for teachers**

Regardless of the problems of arriving at a universal working definition of assessment literacy, many projects have been undertaken to implement a greater understanding of tests in several different countries and educational fields. We will examine one example originating in the UK, and one in the USA.

The first example comes from the University of Bedfordshire, U.K. where there is a relatively new institute named CRELLA (Center for Research into English Language Learning and Assessment). Here Professor Anthony Green and his team are developing material for teacher training courses in Russia. The reason for working on assessment literacy is that they detect a lack of communication between teachers and testers, which has led to a weakness in teachers’ testing skills. Teachers could find out much more about what their students have learned and what they are not learning if they used more effective

tests. Green remarks: “I think we’re seeing a realization that teachers can actually benefit a lot from understanding how assessment works, and they can get a better understanding of what tools they can use to really find what their students are able or aren’t able to do.” (Dunkley, 2014, 15)

In addition to giving training teachers a grasp of the principles of assessment, and skill in choosing the correct test format, there is the question of washback. This is the influence of the test on classroom teaching and on students’ attitude; generally speaking if the features of the test are known, then students and teachers will pay more attention. To give but one familiar case, the introduction of a listening component in the Japanese Center examination for university entrance has led teachers and students to pay more attention to speaking and listening to English than previously, when the exam was on grammar, vocabulary and reading only. Green explains: ‘How can we harness the power of the test in ways that helps people to get something useful out of that process [of study for the test], rather than just trying to “trick the test”.’ (Dunkley, 2014, 15) In this way the test can be a stimulus to more interesting and effective teaching.

CRELLA’s project in Russia is ambitious. In co-operation with the Russian Ministry of Education and the national association of teachers of English they are realising a project known as *Promoting sustainable excellence in testing and assessment of English*. This aims to train not just new teachers entering the profession, but every secondary school teacher of languages, in the basic concepts of language testing and assessment. This policy should slowly contribute to the improvement of language teaching in Russia.

### **Assessment literacy online for teachers**

While initial teacher training needs to include an assessment literacy course, as in Green's project above, the needs of currently serving teachers should not be neglected. These days in-service training of teachers is increasingly migrating to the internet, in a wide range of disciplines and course lengths, from short courses to diploma and Master's level offerings. It is through the internet that CAL, The Center for Applied Linguistics in Washington D.C. chose to improve assessment literacy among teachers, sponsored by the U.S. Department of Education, in a project lasting from 2005 to 2009. The result is an online tutorial entitled *Understanding Assessment: A Guide for Foreign Language Educators* ([www.cal.org/flad/tutorial](http://www.cal.org/flad/tutorial)).

The need for a minimum level of assessment literacy was detected in several studies in which teachers' testing behavior was found to be unhelpful to their students. For example, a study of EFL teachers in Colombia (Lopez Medoza and Bernal Arandia (2009)) found two main problems resulting from a lack of testing knowledge: assessment was mainly summative (evaluating past learning) rather than formative (motivating future learning), and test scores were reported to students in term of degrees of failure rather than degrees of success, or one global grade which did not separate skills such as reading, writing, grammar and speaking, in other words ways which do not "facilitate the learning process" (p.57). To remedy these issues the authors suggest that teachers need to receive training in the making of tests and interpreting test results.

In response to various calls for improved assessment literacy among teachers, CAL developed an online tutorial, to be freely available to anyone with an internet connection. The ideal was to make it possible for all teachers to assess competently, based on the premise that "Strong properly implemented

assessment provides teachers, students and all testing stakeholders with important information about student performance and about the extent to which learning objectives have been attained in the classroom.” (Malone, 2013, 330) However, deciding on the content was a complex problem. What exactly do teachers need to know, and to what depth?

CAL’s strategy was initially to create a first version of the tutorial, then to elicit opinions and make a second version in the light of these criticisms and finally, after subjecting it to evaluation again, to make the final version. The tutorial was divided into seven sections, entitled: *introduction*, *reliability*, *validity*, *practicality*, *impact*, *putting it together*, and *resources*. There were two distinct groups of critics: language teachers and language testing experts, and the principal method of data collection was group interviews with written feedback. 44 teachers, the majority from K-12 schools and the minority from universities, covering seven different foreign languages, took part. The language testing experts, recruited from US government agencies and academe, were divided into two groups, one (17 people) taking part in a group interview and one (14 people) completing an online survey. The results showed a wide divergence between the priorities of teachers and experts: teachers were more concerned with user-friendliness, or “presentation and delivery” (*ibid*, 338) while the experts focused on testing concepts. In short, teachers wanted a quick and easily understandable tutorial, while experts favored the inclusion of detail and theory, in other words “clarity and conciseness” as opposed to “accuracy and detail.” (*ibid*, 340). In the light of these results the authors modified the tutorial to reconcile these two almost diametrically opposed points of view, but concluded with the question: “How can resource developers combine fidelity of definition with succinctness, particularly given the often nuanced and technical nature of language testing definitions?” (*ibid*, 342) These considerations are relevant not just to this specific online tutorial project, but to teacher training

courses in general, beyond the field of assessment literacy.

### **Assessment literacy for policy makers**

As we noted above, a vast number of people are affected by testing and test results in their lifetime. These are often called the stakeholders in testing. In general tests, especially those on which life-changing decisions about college entrance or professional qualification depend, often arouse fear and anxiety in candidates, but one must remember that before tests in many fields were introduced, the alternative was the cause of great uncertainty, insecurity and inefficiency. Take for example the civil service: for centuries posts were allocated on the basis of acquaintanceship or kinship in an enveloping web of patronage, and finally only from the mid-19th century did examinations ensure that entry was based on merit rather than connections.

Far removed both from teachers and candidates are the administrators who make use of test scores to decide admission to the professions. They have crucial power, which if misused deliberately or accidentally, can have serious consequences. They are people “who use language test scores as the basis for decisions ... [and] may make assumptions about tests ... that are at odds with what is intended ... by the language testing community”.(Pill and Harding, 2013, 382) Naturally they are laymen compared to language test developers or teachers, but what knowledge should be considered essential for them? Where do they need to be along the assessment literacy continuum?

One way to answer this question is to assess what administrators know and what gaps appear to exist in their knowledge. Rather than attempt this by means of a survey or a test, two researchers in Australia took an original approach: they studied a verbatim record of a parliamentary committee concerned with



professional accreditation. The committee's task was to recommend ways by which the process of registration of foreign doctors might be simplified and accelerated. Currently two tests are used to assess candidates' English skills: the IELTS test and the *Occupational English Test*, an Australia-specific test in English for medicine.

The author found that there were widespread misconceptions about the tests. The first problematic area was knowledge of tests and responsibilities. Frequently the two tests were confused, whereas they are distinct - one general, the latter occupational. Another problem was that no-one was aware which organ is responsible for setting pass marks. In fact the provider (IELTS or OET) produces results for each candidate, while the Medical Board of Australia determines what is a satisfactory result. A second area of difficulty was erroneous notions of tests and testing procedure. In one instance a committee member assumed that the tests being discussed are written only, whereas in fact they both include a speaking section. Additionally, the person who said "This community is being denied a doctor because probably her grammar is not very good" (*ibid*, 391) assumed that a written test is a grammar test. Furthermore an ignorance of test delivery practices was evident, with some participants indicating that they thought that if you take the same test twice you are looking at the same questions again, which is not true. In sum, many participants in this committee's activities showed a serious lack of knowledge about the relevant language tests.

Based on these indications, the general level of assessment literacy of those involved in the discussions ranged from *illiterate* to *nominal literacy*. The authors conclude "There is a need for a fuller understanding among policy makers, test users and the general public of the scope of language tests and the claims that can be made based on their results".(*ibid*, 401) To achieve this they recommend a PR offensive, or *language assessment communication* effort to

raise awareness.

### **Assessment literacy for college administrators**

Many of the major public examinations around the world are taken by young people for the purpose of entering higher education. As a result, the three main stakeholders in this context are the candidates, the exam makers and the college administrators. These administrators can be in several different categories: admissions, marketing, academic and English language departments. Among these functions, admissions decisions are the most crucial, where an understanding of the tests on which admission criteria are based plays a vital role in ensuring fairness.

Recently university admissions have come to be based on a series of well-known standardized tests, especially where the admission of foreign students to English-speaking universities is concerned. In North America TOEFL is the standard, while in the UK and Australia IELTS is preferred. In certain circumstances several factors relating to each candidate, other than test scores, are considered when admitting foreign students, especially for higher degrees (Banerjee, 2003), but in many institutions, especially at the undergraduate level, only the test score is taken into account. This has led to criticism of the reliance on test scores alone. Spolsky (2008, 300) regretted that the public has “an oversimplified view of the ease of production of meaningful measurement”, and pointed out that numerical scales are blunt tools in measuring something as complex as language ability: “... proficiency is dynamic, ... a single measure at one time does not permit predicting what it will be later.”

In a recent series of studies O’Loughlin (2011, 2103) has investigated the use of proficiency test scores by admissions administrators in Australian

universities, where the usual test is IELTS. He found that the minimum score was determined more by tradition than as a result of consideration of the language needs of future students. Furthermore, there was no system for checking whether a high IELTS score on entry resulted in a more successful college career. In conclusion, it appears that there were “no beneficial educational consequences flowing from use of the test” (O’Loughlin, 2103, 366.)

In his most recent study O’Loughlin (2013) set out to analyze the assessment literacy needs of university admission staff, using an online survey and interviews. In general it was found that staff are mainly interested in basic information about IELTS, especially the minimum score needed, rather than the meaning of or reporting of the scores. The respondents were mainly satisfied with the information provided by IELTS on its website and in the printed IELTS guide (2009). However, there was a general opinion that more could be done to educate staff about the test. The author concludes that there is over-reliance on a single score and that IELTS’ own advice urging admissions decisions to be based on a broad range of criteria such as “age and motivation, educational and cultural background, first language and language learning history” (IELTS Handbook, 2009, 5) is being ignored. To improve the situation, the author suggests that web-based information should be available, to be accessed by staff according to their needs: in other words “adaptive and interactive online tutorials” (O’Loughlin, 2103, 377.)

### **Conclusion: the assessment literacy continuum in practice**

We have seen that different stakeholders in the field of testing have widely differing assessment literacy needs, from a knowledge of concepts such as

validity and reliability in the case of teachers, to a much less specialized range of awareness for policy makers. As a result, a blanket definition such as that of Fulcher (2012) is perhaps too restrictive. A more practical definition is given by Inbar Lourie (2008, 389), who states that assessment literacy is “having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results.”

An extension of this idea appears in Taylors’s (2013) suggestion that each stakeholder group has a different assessment skill profile. To graphically illustrate her concept of differentiated assessment literacy she presents three concentric circles. The inner circle is the one closest to the tests, occupied by researchers and test makers, where theory and statistical knowledge are important. The next circle, numerically larger, is occupied by teachers, who need less theory and more practical skill. Finally the outer circle, the peripheral group, is where the public and the policy makers are to be found, needing a simplified but accurate understanding of testing. This scheme helpfully complements Pill and Harding’s (2013) notion of the assessment literacy continuum.

Where on the continuum do teachers really need to be? In a recent study of testing practice at a Japanese university, Dunkley (2004) analyzed the testing procedures of seven experienced native-speaker teachers of English “oral communication” for non-specialist freshmen. Most teachers set a written examination with a mixture of vocabulary, grammar and listening questions, while speaking was not tested directly because of the large classes (40–50 students). However, the crucial difference between these tests and large-scale high stakes tests is that the teacher has a wealth of data on the students’ progress as result of meeting them 15 times in the semester: thus additional factors such as attendance rates, attitude in class, pronunciation and homework performance

are added to the test scores to give a semester grade.

As for the teachers' attitudes to academic testing knowledge, most of them were glad that they had taken a testing module as part of their TESOL training courses, while the minority saw testing as practical activity where theory and statistics were hardly relevant. It seems therefore that for practicing teachers each person finds his or her own appropriate level of assessment literacy, depending on a number of factors. For example, the level of importance of the course plays a role: testing in a specialist subject is handled more rigorously than in a minor subject. Additionally, the students' stage in the university career affects the teachers' attitude to the test, thus for example senior students' examinations are very carefully checked for content validity, and marked with greater care, with the results reported to the students in greater detail.

As a result of these considerations, it is reasonable to conclude that assessment literacy for teachers is best furthered through its inclusion in initial teacher training courses. Subsequent refresher courses can be delivered online, to be used as and when teachers feel the need for them. For other stakeholders on the periphery, online resources and easily understandable media coverage, especially at times of year when examinations, especially university entrance examinations are a current topic (August in North America and the EU, February in Japan, for example), would be the best way to increase the level of assessment literacy. As with other types of literacy, it is only when people become aware of gaps in their knowledge that they are motivated to seek out information. Now, thanks to the new media, easily accessibility and well presented content, the motivated person can improve their level of knowledge quickly and thus achieve a suitable level of assessment literacy.

## References

- Banerjee, J. (2003) Interpreting and using proficiency test scores (unpublished PhD thesis) Lancaster University, Lancaster, U.K.
- Bybee, R. W (1997) *Achieving scientific literacy: From purpose to practices* Portsmouth, NH: Heinemann
- Dunkley, D. (2004) Current practice in the testing of oral communication: a case study in a Japanese university *Aichi Gakuin General Studies Journal*, 51, 3, 73–80
- Dunkley, D. (2014) Britain's new Language testing powerhouse: an interview with Profesor Tony Green *The Language Teacher*, 45(4), 13–16
- Fulcher, G. (2012) Assessment Literacy for the Language Classroom *Language Assessment Quarterly*, 9(2), 113–132
- Inbar-Lourie (2008) Constructing a language assessment knowledge base. A focus on language assessment courses. *Language Testing*, 25(3), 385–402
- Inbar-Lourie (2013) Guest editorial to the special issue on language assessment literacy *Language Testing*, 30(3), 301–307
- Lopez Medoza, A. A. and Bernal Arandia, R. (2009) Language Testing in Colombia: A call for more teacher education and teacher training in language assessment *PROFILE*, 11(2), 55–70
- Malone, M. (2013) The essentials of assessment literacy: Contrasts between testers and users *Language Testing*, 30(3), 329–344
- O'Loughlin, K. (2011) The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8(2), 146–160
- O'Loughlin, K. (2011) Developing the assessment literacy of university proficiency test users *Language Testing*, 30(3), 329–344
- Pill, J. and Harding, L. (2013) Defining the language assessment literacy gap: Evidence from a parliamentary inquiry *Language Testing*, 30(3), 381–402
- Spolsky, B. (2008) Introduction- Language Testing at 25: Maturity and Responsibility? *Language Testing*, 25(3), 297–305
- Stiggins, R. J. (1991) Assessment Literacy *Phi Delta Kappan*, 72(7), 534–539
- Taylor, L. (2009) Developing Assessment Literacy *Annual Review of Applied Linguistics*, 29, 21–36

### Notes

1. The author acknowledges the receipt of an Aichi Gakuin University Research Grant to gather information at the Center for Applied Linguistics in Washington D.C. in March 2014.
2. The author acknowledges with gratitude the assistance of his colleague at Aichi Gakuin University (Liberal Studies Department) Professor Geoffrey Blair in reading and commenting on this article.

