

Landmarks

Assessing and Testing Learners' Language Proficiency

Geetha Durairajan

As living social beings, we spend a large part of our lives either conversing with people, reading something written by others, or writing about something to someone. Every time we listen to what is said or read what has been written, we subconsciously evaluate what has been stated. This instinctive evaluation includes the language capability or proficiency of the person we are interacting with. We tell ourselves, at least in our minds: “good writer”, “speaks well”. We evaluate those we interact with, but such an evaluation is unconscious and instinctive; without it no conversation or correspondence between any two people would be possible. The evaluation of the language capability of a person is not mandatory, but the evaluation of what is said is essential.

As caregivers and parents, we also evaluate the language proficiency of our children, but without attaching a judgment to it; we do this to help them learn a language. This is the Language Acquisition Support System (LASS, see, Bruner, 1978). The interaction provided by parents, caregivers and more-abled peers enables language learning to happen. In the context of formal education however, we need to go beyond this informal evaluation; we need to test and assess the language proficiency of our learners/students.

In this short paper, I will focus only on the testing and assessment of English as a second language. I will first examine the differences between evaluation, testing and assessment, and then sketch a brief overview of the history of

language testing, and finally end with a discussion of current trends; this will include the testing and assessment of language in multilingual contexts. However, the paper will not provide a comprehensive overview of the latest research trends or findings in the area. Instead, it will focus primarily either on aspects of recent work done or emerging focuses which are likely to be of interest to a language teacher in the classroom.

Evaluating, Assessing and Testing Learners' Language

Evaluation is and can be done by all human beings. In the context of education, it happens nearly all the time informally, and whenever the system requires it, in a formal manner. We evaluate our students when they talk to us, or ask us questions or answer our questions. When they write something and submit it to us either as homework or as answer papers, we move away from evaluation and assess their responses. Whenever required, we also administer tests or give them assignments to test or assess their language ability/capability/proficiency.

A test is a “procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual” (Carroll, 1968, p. 46). Our language test paper is an instrument; students answer the paper and we make inferences based on these responses. If they get many answers correct, we infer that they have high proficiency or capability. We

need to remember, however, that our ‘inference’ has been based only on that one test, or rather that sample of performance. As such, we need to take care to ensure that our test sample is representative so that our inference is a valid one.

All of us know that these samples need not always be representative or reflective of the actual ability of the individual. Our students may not have done that particular test very well and therefore, although they are capable of more, they may have got bad grades/marks. This is where assessments become valuable. An assessment of language capability could be through assignments, projects or term papers. In an assessment, students can think about what they have to write and do, and have time to revise their responses. In assessments of student performance, unlike informal evaluation which can happen through observation, one assumes that there is some grading involved. However, in the history of language testing, we have moved from just evaluating (which continues even today) to experimenting with different types of testing practices (and staying with some) to advocating alternative assessment practices.

Language Testing History

The history of language testing and assessment as we know it is less than a hundred years old. Universities have of course existed for many more centuries and in these, written and oral examinations have been conducted. But these examinations were in different disciplines or subjects and evaluated mastery in that area, whether Philosophy, Science or History. Language as a skill was never evaluated in these examinations. The earliest official documentation on the testing and assessment of language as a skill where language proficiency or ability or capability was evaluated can be traced to the period of the First World War. The testing of language, (in this case,

English) as a skill, was divided into three phases. Spolsky (1975) identified these as the pre-scientific, the psychometric-structuralist, and the psycholinguistic-sociolinguistic (cited in Morrow, 1979, p. 144). In that seminal article, “Communicative language testing: revolution or evolution”, Morrow renamed these three phases as the Garden of Eden, the Vale of Tears and the Promised Land.

These three phases echo and reflect the developments in two other disciplines—the assumptions about the nature of learning (psychology) and the nature of language (linguistics). When we look at the nature of learning we know that there were three paradigm shifts, from Behaviourism to Cognitivism and later to Social Constructivism. The mind was initially seen as a tabula rasa and systematic inputs along with positive and negative reinforcements were the only ways in which learning would happen. With the advent of Cognitivism, this changed and it was assumed that the mind was capable of perceiving patterns and would transform whatever was taught. Accommodation and assimilation of knowledge became important rather than the ‘empty’ output or repetition of transmitted knowledge by learners. However, by accepting the philosophy of social constructivism, we have gone far beyond the assimilation of knowledge. It is never just transmitted and not assimilated either. Today, all learners are perceived as co-constructors of knowledge. The focus therefore has shifted from understanding and comprehension of knowledge to its interpretation. These paradigm shifts can also be identified in the assumptions about the nature of language. Language was, at one point in time, seen as made up of only sentences and words. Later, the focus shifted to language as a cognitive tool. Today language as the primary tool of thinking is being foregrounded. The use of language itself may show high degrees of variability depending on the

varied understanding of users that is largely based on their backgrounds.

The three phases in English Language Testing, the pre-scientific, the psychometric-structuralist, and the psycholinguistic-sociolinguistic reflect these changes. To use the more cited Morrowian terms, the Garden of Eden was the period before the structuralist era. In this, the method of teaching English was mostly grammar translation and it was assumed that if students could write good essays, that capability could be equated with good language proficiency.

The problem with essay writing, however, was not what language learners and users could or could not do with the language but with the manner of evaluation. Long essays had to be evaluated subjectively and this meant that even if scoring criteria were specified, they could and would be interpreted differently by different evaluators. The reliability of such a marking system caused problems. When test creators, evaluators, and syllabus and curriculum designers (and other stake holders) realized that essays were good samples of capability but had problems with reliability, particularly scoring, the pendulum of testing swung to the other side.

This swing of the pendulum brought with it in-language testing, the Vale of Tears, or rather the psychometric-structuralist period. In this the method of teaching was largely audio-lingual, or structural, and testing was largely objective. Multiple choice tests of grammar, vocabulary, phonetic discrimination, and reading and listening comprehension were most common. The focus shifted from validity to reliability.

The third phase is what Morrow identified as the “Promised Land”. This was the era of communicative language testing. There was a significant attempt to test real life language use, and to use tasks where skills were integrated. With a deeper understanding of the nature of language proficiency (as multi-dimensional and

comprising many skills), the focus shifted to the validity of the test (authentic tasks that would also predict future performance) along with an attempt to evaluate these integrated ‘performances’ of students in a reliable manner. Thus, if the test creators and evaluators had to test tasks that required essays to be written in one section of the paper, they would balance it in another section with tasks that tested the aspects of grammar and vocabulary that such an essay demanded, through objective items. Different aspects of language proficiency were tested in a range of ways to ensure validity as well as reliability. An important justification for this shift to communicative language testing was positive washback.

Washback, or backwash refers to the influence of testing on teaching and learning (Alderson and Wall, 1993). Tests and examinations impact teaching and learning (Bachman and Palmer, 1996). For example, if tests and examinations focus on reproduction of knowledge then the teaching and learning of such knowledge is given importance in classrooms. The move from objective multiple choice items to task-based testing, it was hoped, would positively influence language teaching and learning and enable better proficiency. The testing practices did influence teaching, and communicative language teaching became very popular. Even today, as far as standardized proficiency tests are concerned, the practice of testing integrated skills is advocated for this purpose. Most coaching institutes are forced to go beyond the mere teaching of accuracy and focus on tasks that enable fluency as well.

In educational contexts, there was a realization, however, that the beneficial washback of testing on teaching and learning alone was not sufficient. All summative tests and examinations (achievement tests) asked students to write essays, or respond to tasks, but they required the student to deliver timed one-shot written responses (Wieggle, 2002). Students, as test

takers, when required to write essays in examinations, never had the time to go through the real-life writing cycle of thinking about what to write, jotting down ideas, planning and organizing, writing a draft version, revising what is written and then writing a final version of that essay. The decisions made about the language proficiency of students, based on such “instantaneous” writing would therefore lack predictive validity. Students may not do well on such examinations, but in reality, they were perhaps good writers.

This problematization of timed writing can be extended to other aspects of proficiency, namely, speaking, reading and listening. According to Cronbach (1970), under examination conditions that require maximum and not typical performances from a student, test takers would be tense and stressed and were not likely to perform to the best of their capability. Some form of tests and examinations are necessary for summative evaluation and certification and while these can be tweaked to provide positive washback they cannot be used as tools for teaching and learning. They can only inform or be the driving force behind them. Instruction driven by testing will become unavoidable in such situations.

Current Trends in Language Testing

A big step in the field of language testing in the twenty-first century was to separate large scale tests of language proficiency which are used for admission or stand alone certification from tests of language within educational contexts. Formative evaluation, along with alternative assessment practices are now given as much or even more importance than summative final examinations. This is in line with the idea of continuous and comprehensive evaluation (CCE) which is being seen as important in educational contexts. Tests and assessments are seen as pedagogic tools, integral to teaching and

learning. The shift is from system-oriented examinations to teacher-managed tests (Durairajan, 2015).

There is a parallel move from the assessment OF learning (whether students have successfully achieved their objectives) to assessment FOR learning, as a teaching and learning device. The focus has shifted to valuing student responses and the genuine pedagogic feedback provided by the teacher. Instead of worrying about reliability in marking in large scale public examinations, there is an attempt to see whether tests and examinations can genuinely be made learner-centred. Students are asked to reflect on their own capabilities through the use of ‘can- do’ descriptors. This has added a third preposition to the “of” and “for”, assessment AS learning.

Another major trend of the twenty-first century is to move away from the thinking that only prescriptive timed examinations can be used for summative evaluation and certification. Alternative modes of assessment, particularly portfolios, assignments and projects are a part of most evaluation practices in schools today. When working on these assignments, students have the time to think, plan, write, edit, revise and submit their work. When making presentations, they also have the time to plan, compose and work with what they would like to state before hand.

Thus far, the discussion about language proficiency has been made from a monolingual perspective. But in many countries, and this is particularly true of English, it is either a second or foreign language. The testing and assessment of learners’ language proficiency must take such bi/multilingual capabilities into account.

The Way Forward: Testing Proficiency Across Languages

In countries such as India, all language functions are never fulfilled through one language. More

importantly, English is rarely the language of thinking, organization and planning for the majority of the population. There is no zero level of English in India; our students' receptive capability (read and understand English) is much higher than their productive capability (speak and write English). The language(s) capability of students in grassroots multilingual contexts must be tested and assessed together. It should be possible to get our students to read complex texts in their more enabled language and respond orally in English, or read complex texts in English but write an essay in their more enabled language.

Languages are communicative and cognitive tools; they do not exist in separate compartments or corners without contact with each other inside our minds. Learners may be better enabled in the higher order skills of analyzing, applying theory to practice, synthesizing and creating in their more enabled language. Our testing and assessing practices need to learn how to access these capabilities and evaluate them across languages.

References

- Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Bachman, L. & Palmer A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bruner, J. (1978). Vygotsky: A historical and conceptual perspective. In J. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives*. (pp. 21-34). Cambridge: Cambridge University Press.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69). London: Oxford University Press.
- Cronbach, L. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper and Row.
- Durairajan, G. (2015). *Assessing learners: A pedagogic resource*. Delhi: Cambridge University Press.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution. In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp.143-157). Oxford: Oxford University Press.
- Spolsky, B. (1975). Language testing: Art or science? *Address to the Fourth AILA Congress* : Sumarized in Valette 1977.
- Wieggle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Geetha Durairajan heads the Department of Testing and Evaluation at EFL University, Hyderabad. Her primary areas of interest include second language learning and non-prescriptive language evaluation.
gdurairajan@gmail.com