

Quality Control in Evaluation and Assessment

by J Charles Alderson

Paper presented to Seminar on Common Standards in Europe, Szeged, April 23rd, 2004.

J Charles Alderson

Contact details
Department of Linguistics and English Language,
Bowland College
Lancaster University
Lancaster,
LA1 4YT
UK

email: c.alderson@lancaster.ac.uk
fax: 00 44 1524 843085

Abstract

In this presentation, I argue that there are two main traditions of assessment in Europe. One is teacher-centred and school-based, and the other is centrally organised and externally administered. The virtues of the former are frequently proclaimed, for being close to the teaching and learning that has gone on in classrooms. The vices of the latter are equally frequently asserted, for being insensitive to the learning that has taken place in class and to development over time, since the assessment is a typically one-off event. However, the virtues of the former are rarely investigated and the advantages of the latter are increasingly recognised in an integrated Europe. I argue that whether the assessment is school-based or external, it has to conform to quality standards that are intended to ensure fairness and comparability.

Introduction

"Assessment is central to language learning, in order to establish where learners are at present, what level they have achieved, to give learners feedback on their learning, to diagnose their needs for further development, and to enable the planning of curricula, materials and activities."

But imagine a situation where a school-leaver sits for six hours, writing answers to questions that have been written by one untrained individual in a Ministry, without any test specifications or guidance on what the exam should contain, the difficulty of which is completely unknown, and whose responses are marked by teachers within that same school, without any instructions on how marks should be awarded, what criteria should be used, or whether other teachers should also mark the responses to

ensure a degree of consistency. Imagine that same student in front of a panel of four or more judges seated behind a table decorated with flowers, being required to recite a monologue in response to an essay-type question, whilst behind him other students scribble notes to prepare their answer to a similar question which they will have to deliver in the next fifteen minutes, where the external moderator of the panel understands no English and the sole mark is given by the student's own class teacher. Such a ceremony is an annual experience for thousands of school-leavers in many places in Europe, yet in what sense is this assessment "central to language learning"? In what sense does such a system establish what level a learner has reached, and in what sense does it enable curriculum planning or materials development? Sadly, in no sense, and the results of such examination rituals are generally ignored by employers and universities, because they cannot be compared with known standards, because results vary between schools and towns, and because there are available certificates that ARE recognised by the system which are developed and delivered by commercial organisations, for which students have to pay a substantial fee.

Imagine another situation, this time in the UK, where a test of spoken English for speakers of another language produced by a UK examinations board is being administered. It consists of a five-minute chat, delivered by an interlocutor who is a teacher in a further education college. He or she has not been trained to conduct interviews or oral tests of any sort, he or she has only one hour earlier actually seen the instructions, questions and pictures used in the "examination". He or she is expected to conduct the oral examination, ask questions, elicit the candidate's performance, and make judgements about that performance. He or she has not been trained in making reliable assessments, has had no formal induction into the meaning and use of the rating scales, yet he is the only person who actually interacts with the candidate and observes how well that person can perform. All in five minutes. The examination board that produces and profits from this examination claims, without any supporting evidence, that this five-minute chat provides valid and reliable evidence of learning, the college that administers the test claims that this exam is relevant to and a suitable measure of achievement on, their syllabus, and all assert that the exam covers levels A1, A2 and B1 on the Council of Europe Framework. What nonsense! Not surprisingly, the examination board refuses to provide any information on how reliable the assessments are, or on how valid the results are, or on what they mean. Enquirers are simply told that the information is confidential.

Believe it or not, such an examination exists in the UK in 2003, and it is officially recognized by a UK Government agency - the Learning and Skills Council. Colleges administering the exam can claim official recognition for any student who achieves a pass at levels 1 and 2 on this exam, and courses that lead to the award of the "certificate" are fully funded by the Government. The use of this examination is thus crucial for the existence of courses in English as a second language in this and many other colleges in the UK.

Clearly, delivering an "accredited exam", and offering courses that lead to accredited exams, that attract government money, means a lot of money for the colleges that run the courses, and for the examination boards that offer the exams. Language testing is big business.

However, language testing is not only big business. It can be a life-changing experience. Admission to university, graduation from university, the possibility of employment, the opportunity to become a citizen of a country, all these can depend upon passing a language test. Such tests are known as high-stakes exams, because they can have important consequences for individuals. Indeed, they can be literally life and death matters for individuals, as we see in Japan, for example, where numerous teenagers every year commit suicide because of the pressure of university entrance examinations, and the incredible importance in Japanese society of getting into a good university. There, frequent reference is made to “Examination Hell”. In such circumstances, it is surely crucial that Governments, the teaching and testing professions, and educational authorities make every effort to ensure that the examinations themselves are fair: that they measure relevant skills, that they are fit for their purpose, that the assessments that are made are made reliably by trained professionals, that the examinations follow accepted practices of test construction and validation, and that they measure ability at recognized and validated levels of language proficiency.

Yet all too often, the quality of such important examinations is not monitored, there is no obligation on those delivering the examinations to prove that their exams are relevant, fair, unbiased, and reliable, and that they do indeed measure relevant skills fairly. All too often, it is simply accepted that because somebody has a degree in a foreign language, that person is qualified to examine another person's language competence, regardless of the fact that it is highly unlikely that during the degree course, that individual will have been trained in language testing, and indeed may not have been trained in any aspect of applied linguistics or language pedagogy. Even worse, in many circumstances it is assumed that simply by virtue of being a native speaker, one is qualified to assess the language competence of another person. In a survey I once conducted for the Australian Government I discovered that UK policemen - any policeman, not an official trained in assessment - are authorized to hold an interview with individuals wishing to become a citizen, and if they are satisfied that the person knows English, he or she qualifies, at least on linguistic grounds.

In contrast, internationally, there has been an increasing professionalisation of language testing. Ten years or so ago the International Language Testing Association (ILTA) was formed, it has already developed a Code of Ethics which sets out the moral standards by which professional language testers should abide, and it is currently developing a Code of Practice which is intended to set out the professional standards and procedures according to which tests, especially high stakes tests, should be developed. ALTE - the Association of Language Testers in Europe - has developed a similar Code of Practice, which describes how professionally developed language tests are constructed and validated, and ALTE is already developing self-assessment procedures, according to which European examination boards will assess whether they adhere to such standards or not. One day, presumably, there will be external inspection and monitoring of these standards, to ensure that exam boards do indeed adhere to the agreed standards. Also in Europe, a new association has recently been formed - EALTA, the European Association for Language Testing and Assessment - www.ealta.eu.org - which is an independent professional association for individual language testers in Europe. EALTA's interests are independent of those of any other organization. The purpose of EALTA is to promote the understanding of theoretical

principles of language testing and assessment, and the improvement and sharing of testing and assessment practices throughout Europe.

In general, the professional language testing community throughout the world is increasingly concerned to ensure that its tests have adequate quality control mechanisms, that they follow professional ethical standards of design, development and validation, and that the tests also have a positive impact on candidates: research into the washback of tests on teachers and learners is increasing and is addressing questions like: how do tests impact on classrooms, how do teachers prepare their students for tests, what influence do tests have on curricula and on materials? Moreover, language testers have been discussing and publishing articles for about five years now on the ethics of their profession, on the grounds that to be a true profession, like doctors or lawyers, language testers need codes of conduct and of professional ethics.

However, in Europe, as elsewhere, there are two traditions in assessment, and language assessment is no exception. The first and older tradition is one of teacher-centred, school- and university-based assessment. In such a tradition, the teacher of the student or pupil being examined has the right to develop the questions that the student will be asked, the teacher's opinion of the student's performance is the one that counts, sometimes moderated by some sort of examination panel. It is very common for teacher-examiners not to have explicit criteria according to which they grade students: rather it is assumed that by virtue of being a teacher, and by virtue of having taught the student being examined, the teacher-examiner will make reliable and valid judgements. In such a tradition, the authority, professionalism, reliability and validity of the teacher is rarely questioned. Students can challenge judgements, usually by inspection of exam papers, but it is rare for students to be failed. For example, in the current school-leaving examination in Hungary, less than 1% of the school-leaving population fails the exam. As a result, the exam is largely regarded by universities and employers as worthless. Instead, school-leavers have to pay to take either a Hungarian certificate or a foreign certificate like Cambridge, if they want to have a certificate which has any value at all. The principle that free education for all should be completed by a freely available and valid/ reliable certificate of achievement is flagrantly violated in Hungary, to the financial benefit of private or quasi-private, commercial examination boards.

The second tradition is relatively new, and that is the development and administration of external measures of competence and ability. In this tradition, national or regional or quasi-governmental agencies are responsible for developing professionally respectable measures, which generally follow accepted standards, along the lines of the ILTA/ ALTE recommendations. Such tests are centrally constructed, piloted and revised, their difficulty levels are empirically determined, and procedures are established for the assessment of performances on tasks by externally trained assessors. They are empirically equated to known standards or levels of proficiency.

The first tradition is prevalent in many countries, especially in Central Europe, but as Central Europe changes, many countries are trying to adjust their examination traditions to different ways of doing things, and are moving towards the second tradition I outlined, one which is closer to what I have called a professional approach.