Evaluation and Assessment

Michael FLEMING

PRELIMINARY STUDY
Languages of Education

Evaluation and Assessment

Michael FLEMING, *University of Durham, United Kingdom*

Intergovernmental Conference
Languages of Schooling: towards a Framework for Europe

Strasbourg 16-18 October 2006

Contents

Abstract

Assessment often causes tension and disagreement because of the diverse and demanding functions that it is expected to fulfil. These differences can be compounded in the context of language as school subject (LS) because the aims and range of outcomes are particularly complex. It important that the values which underpin and inform a system of assessment are examined and made explicit, otherwise learners may unwittingly be affected in negative ways. One of the roles of a Framework for Language(s) of Education would be to make explicit the different functions of assessment and demonstrate ways in which these can be integrated in practice to raise standards.

Introduction

Few educators would disagree that assessment is of central importance. From the individual teacher in the classroom to those invested with the responsibility for formulating policy, everyone has an interest in knowing whether, and to what degree, teaching has been successful. The attention given to assessment in recent years by researchers has lead to a range of terminology to facilitate communication and bring clarity to the development of policy and practice. Technical terms abound and are widely familiar: e.g. norm-referencing; criterion-referencing; formative and summative; value added; competence based (proficiency assessment)[1]. Despite the shared terminology and dialogue, assessment is often the focus of disagreement in the education world, sometimes leading to entrenchment and disaffection. It is important therefore that a Framework for Language(s) of Education addresses issues related to assessment and the improvement of educational standards of quality. The purpose of this short paper is not to examine technical issues related to assessment in detail but to highlight in broad terms some of the central questions and sources of disagreement.

1. Purposes of assessment

In its simplest formulation, assessment provides information on whether teaching/learning has been successful. However the information it provides has a number of potential different audiences whose precise requirements may vary. Classroom teachers need regular information on how pupils' knowledge, skills and understanding are developing, both to inform how they should adjust their teaching and to determine what kind of feedback is needed to improve pupils' learning. On the other hand, school principals and policy makers need additional, broader information on the quality of education in a school or country. The sort of comparative data required for this purpose needs a high level of reliability and uniformity. In the case of language as school subject this requirement is challenging because it is difficult to create tests which are manageable but at the same time faithful to the aims of the subject. Employers and society at large also need reliable information which can help certify achievement and provide a basis for selection. Parents too require information which can help them understand their children's achievements and limitations. Learners themselves need to know how they are progressing and how to improve their performance but they may need to be protected from the potentially demotivating effects of negative assessment.

The concept of 'accountability' when used in relation to assessment usually refers to the imposition of systems of assessment external to the learning process as a form of 'policing' of standards to ensure that the education system is functioning effectively. But the term may be employed more broadly and more positively than this, referring to the different obligations that are relevant to all. Teachers have a responsibility to the learner but also to the needs of the wider society. Policy makers clearly have a duty to the public and need to ensure that the education system is delivering results but they also have responsibilities to the individual learners and need to consider consequences of policies in those terms. The concept of accountability interpreted in this way will take people outside of vested interests in order to see the larger context. Accountability needs to be linked with a process of sharing perceptions and fostering

---

[1] For a commented list of types of assessment see f. ex.: *Common European Framework of Reference for Languages*, Council of Europe, 2001, Cambridge University Press, p. 183-192

understanding. It is important therefore not to exaggerate differences between different potential 'stakeholders'; what all parties have in common is a fundamental concern that assessment should help raise achievement and improve learning. A starting point for resolving tensions related to matters of assessment is to develop understanding of other points of view. A Framework for Language(s) of Education would have an important role in making clear the different functions of assessment and how these in practice can be integrated.

It may also be helpful to distinguish between assessment for 'summative' and 'evaluative' purposes[2]. Summative assessment provides a summary judgement of the achievement of individual learners at a particular point in time, usually resulting in the designation of a mark or grade. Evaluative assessment however seeks to provide more general information in order to show average levels of proficiency achieved by groups in schools, education authorities or countries. The distinction is useful because it is possible to discharge the evaluative function of assessment by a process of sampling rather than by using summative data derived from an entire population. In the case of older pupils it is likely that summative data provided through national systems of assessment of the kind often administered at the end of compulsory schooling will be used for evaluative data. With younger pupils however it may be sufficient to derive evaluative data from sampling so that the summative and formative functions of assessment can merge. When assessment data is used to compare the progress of different cohorts of pupils, the use of a 'value-added' approach which takes account of the different base-lines from which the pupils are progressing is becoming more common.

Another useful distinction is between 'assessment' and 'monitoring'. (It should be noted that some writers use the term 'evaluation' for what here is being referred to as 'monitoring'[3].) If 'assessment' is often taken to refer to a more formal process of measurement or the designation of a mark or grade, 'monitoring' is a broader term which embraces the making of judgments in a less precise way. This distinction is particularly appropriate in the context of teaching language as school subject (LS) because the aims are inevitably varied and complex (see paper by Laila Aase). It is difficult, for example, to assign a formal grade to creative uses of language. Similarly, a classroom teacher may be able to make an informal judgement of a pupil's enthusiasm for reading but it is not easy to assign a specific grade against observable criteria for an aptitude of this kind. Effective monitoring means being alert to all the signs which may give some indication of a pupil's achievement or lack of success. It also carries connotations of a process which extends over time. Effective teaching requires awareness of all of the indicators which provide information about understanding, whether or not more formal assessments are taking place. In the language classroom this can include: contributions to group discussions; individual, informal conversations; reading aloud; drama presentations; drawings and diagrams. However not all of these activities lend themselves easily to formal assessment. An ongoing process of monitoring will inevitably feed into formative assessment.

---

[2] not to be confused with 'summative' versus 'formative' types of assessment (see footnote 1)

[3] Note however the following use of the term 'evaluation': Evaluation is a term which is (...) broader than assessment. All assessment is a form of evaluation, but in a language programme a number of things are evaluated other than learner proficiency. These may include the effectiveness of particular methods or materials, the kind and quality of discourse actually produced in the programme, learner/teacher satisfaction, teaching effectiveness, etc." (Common European Framework of Reference for Languages, Council of Europe, 2001, Cambridge University Press, p. 177

A Framework for Language(s) of Education would need to address how oral and verbal feedback to pupils can be best used to improve learning as part of the formative assessment process, drawing on the most up to date research findings. There may be dangers for example in comments to pupils which are excessively negative which serve to disillusion and demotivate them. However there may also be dangers for some pupils in feedback which is excessively and unrealistically positive which may serve to create complacency.

2. Approaches to assessment

Traditionally assessment of language as school subject (LS) often took a very simple form: pupils were given a narrow written task which was then awarded a grade or mark. This allowed them (and the teacher) to make a judgement of how they ranked in relation the rest of the group (normative assessment). However the absence of clear criteria meant that the information rarely gave an indication of how they could make progress in their learning. Also the test itself often embodied a very narrow conception of what competence in language entailed; it often centred on knowledge of language form and structures (syntax and grammar) and a narrow range of language uses (often only a written form of essay). It is important that any system of assessment takes account of the broad nature of language as school subject (LS) and the complexity of its aims.

The different purposes related to different potential audiences illustrate the challenge involved in devising assessment instruments. Two central tendencies emerge. One places emphasis on the assessment *of* learning where reliable, objective measures are a high priority. The focus here is on making summative judgements which in practice is likely to involve more formal examinations and tests with marks schemes to ensure that the process is sound. An alternative approach is to change the emphasis from assessment *of* learning to assessment *for* learning, implying a more formative approach where there is much more emphasis on feedback to improve performance. The approach here might be through course work and portfolio assessment in which diverse information can be gathered which reflects the true broad nature of the subject. A further advantage of this approach is that it can embody different forms of self-assessment which can be helpful ways of motivating learners and having them reflect on their progress. Self-assessment encourages pupils to take responsibility in the learning process although it is advisable for them to be trained in self-assessment techniques for this to work effectively. The broader the approach to assessment (incorporating the judgement of a range of different performances in different contexts) the more it can be said to constitute a meaningful assessment of performance in the subject . However tension emerges because it is sometimes difficult to compare, with any degree of accuracy, the results drawn from broad approaches to assessment. The quest for 'objective' and reliable methods of assessment driven by narrow ideas of accountability brings with it a number of dangers. So called 'teaching to the test' may not be a problem if the tests are sophisticated and wide-ranging but there may be practical difficulties in administering those that are too complex. If the tests are too narrow and simplistic then this may have an adverse effect on the teaching.

An ideal assessment system would reflect the full complexity of language as school subject (LS), and would motivate learners by giving useful feedback, while also providing other stake-holders (e.g. policy-makers and employers) with the information they need. An integrated approach to assessment would ensure that the different purposes and approaches are balanced so that no one priority has adverse and undue influence on the system as a whole.

A key concept is embodied in the notion of 'transparency', the view that those being assessed are aware of the criteria which are being used to make judgements about them and how those judgements are made. Knowledge of criteria can help performance and improve motivation but once again, in the context of language as subject, the issues are more complex than they first seem. A common assumption is that pupils learn best when they know what they are trying to achieve and why. While this view is largely true, there are exceptions. Because the development of language can in some ways be described as a 'natural' process learners do not always need to be fully focused on specific aspects of their performance in order to improve. In fact too much focal awareness on performance can make them too self-conscious: speakers can appear too groomed and artificial; the writer who has been told to strive for effect by using more adjectives may develop a highly artificial and awkward style. These insights do not negate the importance of transparency as a principle but highlight the fact that in pedagogical practice the principle needs to be interpreted and implemented with care.

The use of competences to describe outcomes which can be assessed has developed considerably in recent years. It is worth noting that the term 'competence' is used by writers in different ways which can be a source of confusion. Sometimes it is used in a very general way as a synonym for 'ability' or capability as in 'language competence'. Other writers use the term to describe broad language modes or domains such as reading, writing, speaking and listening. More commonly however 'competences' refers to the specific actions which a learner must perform and which in turn can be assessed to demonstrate achievement in a subject. Advocates of using competence statements for assessment purposes and syllabus design see their value largely in bringing clarity and transparency to the specification of learning outcomes. Critics of a competence approach take the view that performance statements are too narrow and specific, and do not reflect the range and subtlety of what is involved in language development. The balance of advantages and disadvantages needs to be considered.

3. Language as School Subject (LS)

One of the challenges posed by assessing language as school subject (LS) is that the content is so varied and complex (see paper by Florentina Sâmihaian) which is in turn a reflection of the complexity of the aims. Each mode of 'writing', 'reading', 'speaking 'and 'listening' can be broken down into further areas. A subject that is so multidimensional raises the question as to whether an assessment task in one area is representative of achievement in the subject as a whole. For example, it is fairly safe to assume that performance in speaking is not necessarily indicative of reading competence. On the other hand it is less clear whether it is necessary to assume that reading ability varies in relation to texts of different types (fiction, non-fiction, media). The assessment of reading can easily remain at a surface level only addressing recall or literal comprehension rather than deeper understanding. Multiple choice questions on a text are easy to mark and may yield high reliability (in the technical sense) but are less equipped to assess the learner's deep and individual response to a text. Even with very young children the reading process is more than simply decoding text and any system of assessment needs to reflect that fact.

A similar question arises in relation to writing. Do pupils need to be assessed on a range of different tasks reflecting the fact that writing exists for different purposes and for different audiences? The assessment of writing appears to be more straightforward than that of reading and speaking because at least there is always a product which can be referred to after the event. But there is a major challenge in

determining what criteria should dominate in making a judgement (for example the accuracy of the writing as opposed to the impact of the content) and whether the criteria should change in relation to different types of writing. It also wrong to assume that the only response to pupils' writing comes when it is completed. Formative assessment in the form of a dialogue about the work in process is an important way of improving standards. Assessment is always a selection and therefore can unwittingly value some aspects of the subject more than others.

Speaking and listening is extremely difficult to assess because, even more than other aspects of language as subject, performance varies with the theme, context and level of motivation. Poor performance in oral work is often to do with the nature of the task which has been devised, the atmosphere of the classroom and the dynamics of the group rather than the competence of the pupils; to provide a valid assessment of speaking and listening, evidence needs to be drawn from a variety of situations. Some might argue that the assessment of speaking and listening is so complex and context specific that it should not be assessed formally. However because assessment so often determines the curriculum and the way it is taught there are arguments to suggest that speaking and listening should be assessed despite the difficulties; the ability to articulate a point of view orally and to argue a case are essential skills for meaningful participation in a democracy. A Framework for Language(s) of Education would have an important role in providing practical examples of the way all the language modes might be assessed.

4. Language Across the Curriculum (LAC)

What are the assessment implications for a policy that promotes language across the curriculum (LAC)? If aspects of language are taught in different subjects then two key questions emerge: (1) Does the method used in the assessment of the subject reflect the student's performance in language? (2) Do the subjects contribute to the assessment of language as school subject (LS)? In one sense the answer to question one must be yes; understanding of the subject and the ability to express it cannot easily be separated from competence in language use. However it is another matter whether what are sometimes referred to as 'surface' features of language (spelling, grammar, punctuation) should be taken into account when formally assessing pupils' achievement in a subject. The question might be phrased practically as follows: is a proportion of the marks allocated for accurate use of language? If that is the case learners may be penalised twice. Furthermore, the simple practical question conceals difficulties; how for example is 'accuracy' in language use determined? How much tolerance should there be for example of uses of dialect in written discourse? Recently it has become more common to speak of *appropriateness* of language use in specific contexts rather than accuracy. The practical problem still remains: not to recognise the importance of using language appropriately by ignoring 'mis-uses' may have adverse consequences on pupils' performance.

The second question refers to the possible contribution other subjects might make to the assessment of language. There are clearly practical difficulties involved here for it would not be a straightforward matter for teachers of history, for example, to contribute to the assessment of performance in language as subject. There is an argument to suggest that if language across the curriculum is to be taken seriously it must have some impact on how language is assessed. Underlying questions have to do with subject boundaries and the degree to which competence in language can be easily separated from the context in which it is used; learners of history could also in some sense be said to be learning the language of history.

5. Language(s) of Education in the context of plurilingualism

A consideration of language within a broader commitment to a policy of plurilingualism helps to highlight some of the issues that are pertinent to assessment.

Attention needs to be given to the wide range of different contexts in which language is used, both formal and informal, (for example written language takes different forms such as letters, emails, articles, reports and academic essays) and the different purposes language serves (for example, to persuade, inform, argue and report). The concept of plurilingualism which in its primary meaning refers to the speaking of several different languages can be extended to language as subject. Language learners have an entitlement to be exposed to, and to develop, a wide range of languages *uses* which can be seen as one aspect of a policy of plurilingualism.

Assessment is not a neutral process but inevitably embodies particular values; these need to be specifically examined because different approaches to assessment may have unintended consequences. Particular approaches embody different views of knowledge and conceptions of learning, and may promote surface rather than deep learning. Assessment may further exacerbate the disadvantage experienced by some learners by using tests that are culturally biased or may underestimate the competence of some pupils because their range is too narrow. A Framework for Languages of Education would be to make explicit the different functions of assessment, as well as some of the possible unintended consequences, and demonstrate ways in which these might be addressed in practical terms.