

# Language Testing

<http://ltj.sagepub.com>

---

## **Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience**

Richard Johnstone

*Language Testing* 2000; 17; 123

DOI: 10.1177/026553220001700202

The online version of this article can be found at:  
<http://ltj.sagepub.com/cgi/content/abstract/17/2/123>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Language Testing* can be found at:**

**Email Alerts:** <http://ltj.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ltj.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.co.uk/journalsPermissions.nav>

**Citations** <http://ltj.sagepub.com/cgi/content/refs/17/2/123>

# Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience

Richard Johnstone *University of Stirling*

Although much more limited in scope and intensity than second language (L2) immersion, MLPS (Modern Languages at Primary School; known elsewhere as FLES, Foreign Languages at Elementary School) is being extensively implemented across the European Union as a reflection of national and European Commission (EC) policies. Despite an increase of research interest in the area, issues of MLPS assessment have not been widely addressed. There are good reasons for developing such assessments:

- gauging the return on a massive public investment;
- feedback to parents and the public;
- informing national policy development;
- self-evaluation at the level of institutions, departments, teachers and learners.

There are, however, major problems at present in doing so across schools:

- variability of context;
- embeddedness of children's language in a flow of events;
- their relative lack of cultural knowledge;
- unfamiliarity of teachers with concepts of L2 testing;
- lack of consensus concerning what MLPS proficiency might mean.

Examples are given from Scotland of how some of these problems were addressed in the pilot phase of the national MLPS initiative of the early to mid-1990s, leading to the development of more comprehensive and standardized assessments in the national Assessment of Achievement Programme (AAP) intended for the late 1990s and early years of the twenty-first century.

## I Introduction

The teaching of modern languages at primary school (MLPS; known elsewhere as FLES, Foreign Languages at Elementary School) has

---

Address for correspondence: Richard Johnstone, Head, Institute of Education, University of Stirling, Stirling, FK9 4LA, UK; email: R.M.Johnstone@stir.ac.uk

had an erratic career in the UK. The large-scale introduction of French at primary school in the late 1960s and early 1970s in England and Wales met a generally negative evaluation (Burstall, 1974) and went into sharp decline. A parallel innovation in Scotland met a similar fate (HM Inspectorate, 1969). However, in Scotland a national initiative was relaunched in the late 1980s, prompted mainly by a realization that the European Single Market was imminent and, by an assumption explicitly stated by some right-wing Scottish politicians at the time, that MLPS would help deliver a higher level of foreign language competence in secondary schools, and thereby ultimately make Scottish business more competitive in the new Europe.

MLPS refers here to a provision that is generally one to three hours per week, hence much less than partial or total immersion and with objectives that are much more limited, though even within this there is enormous variation across the European Union, since the starting age ranges from 5 or even younger to 12. Unlike immersion programmes, MLPS teachers are normally not native or highly fluent speakers of the other language which normally may be described as 'foreign'. On the other hand, MLPS massively exceeds immersion programmes in the numbers of schools and pupils involved and in its centrality to national and EC policy priorities.

The reappearance of MLPS in Scotland has been part of a wider movement within the European Union where two different models have been adopted: one geared to 'second language (L2) learning' (*apprentissage*) which seeks to develop competence in one foreign language, the other geared to 'sensitization' (*sensibilisation*) which seeks to introduce children to language more generally through a language awareness approach as a precursor to the subsequent learning of one or more foreign languages. Because of its largely instrumental rationale, the Scottish initiative, according to the then specialist HM Inspector of Schools, was consciously within the 'L2 learning' camp (Giovanazzi, 1993). The assessment of foreign-language attainments was therefore of central importance in order to ascertain whether or not pupils in the MLPS pilot – beginning at age 10 or in a smaller number of cases at age 8 – did in fact gain an advantage over those who began their foreign language learning at secondary level – beginning at age 12. As a result, an independent evaluation of the Scottish pilot was commissioned (Low *et al.*, 1993; 1995; Johnstone *et al.*, 1996), for which a variety of assessment instruments were developed that are discussed later in this article.

### *1 Recent MLPS research that involves assessment*

It is recognized that MLPS (FLES) is an international phenomenon, not restricted to member states of the European Union, but the present

article is intended to highlight a range of key assessment issues that have emerged from the attempts to introduce MLPS across several countries within the European Union. MLPS research that in one way or another involves assessment has focused on four main aspects.

- First, the longer-term effects of MLPS on pupils' foreign-language attainments at secondary school (e.g. Low *et al.*, 1993; 1995; Genelot, 1996; Kahl and Knebler, 1996).
- Secondly, pupils' progress and attainments in a foreign language during their pre-primary or primary education (e.g. Balke, 1991; Helfrich, 1995; Low *et al.*, 1995; Kahl and Knebler, 1996; Luc, 1996; Ministry Group of Experts, 1996; Peltzer-Karpp, Hasiba and Zangl, 1996; and elsewhere in the world, e.g., Donato, Antoneck and Tucker, 1994; 1996).
- Thirdly, gauging the development of children's metalinguistic awareness in pre-primary or primary education, usually as a result of a 'sensitization' approach (Bailly and Luc, 1992; Charmeux, 1992; Luc, 1992; Pinto, 1993; Low *et al.*, 1993, 1995; Nagy, 1996; Genelot, 1996).
- Fourthly, measuring children's attitudes to learning a foreign language at primary (a large number of researchers but pre-eminently Djigunovich, 1995).

A review of MLPS research conducted within member states of the European Union and commenting on much of the above is available in Blondin *et al.* (1998).

## 2 Contexts for language testing

Despite the merits of the above research, it cannot be claimed that a substantial body of literature exists at present on how MLPS pupils may be assessed. Writing about language testing generally, Shohamy (1994) identifies three different contexts in which this has a role:

- first, 'an external context in which standardised tests are used for making decisions about individuals and programs regarding, for instance, certificates, diplomas, acceptance, rejection and placement . . .';
- secondly, 'the classroom context, where tests are used as part of the teaching and learning process'; and
- thirdly, 'the SLA [second language acquisition] research context, where language tests are used as tools for collecting language data in order to answer and test SLA research questions and hypotheses' (Shohamy, 1994: 133).

Most MLPS research falls into Shohamy's categories one and two.

Her third category is well represented by Pelzer-Karpf and Zangl's (1997) authoritative account, but this goes well beyond MLPS since it pertains to the Vienna Bilingual Schooling project (see Zangl, this issue).

If thus far relatively little research on or involving MLPS assessment has been undertaken, it is pertinent to ask why research on MLPS assessment is important. Four complementary reasons are suggested.

### *3 Importance of MLPS assessment research*

*a Gauging the return on public investment:* First, in most states of the European Union MLPS is moving well beyond being a merely interesting initiative adopted at the level of an individual school or locality. It is becoming a major aspect of national and, indeed, international policy for education. The White Paper 'Towards the learning society' of the European Commission (European Commission, 1995) specifically recommends that all pupils should receive MLPS during their primary school career and, indeed, that the process should begin at pre-primary level. Very large sums of public money are being spent on the implementation of this policy. Practising primary teachers in Scotland, for example, have the opportunity to receive a 160-hour training programme in one of the four MLPS languages, amounting to 27 days distributed across four terms (15 months). This has cost some £21 million for the first six annual cohorts of teachers. Moreover, in many countries of the EU the starting age is gradually being moved down the school, resulting in a larger investment of time, teachers and resources. It therefore seems appropriate that a range of measures should be adopted in order to ascertain what sort of return is being yielded for this substantial investment. Although, of course, many sorts of data will have an important role in enabling educational decision makers to gauge the return on the public investment they have authorized, it will be particularly important to develop valid and reliable assessments of pupils' performance.

*b Feedback to parents and the public:* Secondly, the interest of parents and other members of the public in MLPS is very high. A key question that they reasonably ask is: What are the children getting from this? Here, good assessments have a vital role to play, particularly as many parents have only a vague notion of what it is reasonable to expect their children to learn in a given period of time in school conditions. They may possibly wish to know not only how their particular child is progressing in the foreign language, but they may also have a more normative perspective and wish to know how

their child is faring in relation to others of the same age in their particular school or, indeed, nationally. They may have been seduced by the proposition that children of a young age are somehow magically equipped for the successful acquisition of a foreign language in school conditions, or they may have read commercial literature that claims Method X enables people to learn a foreign language in three months. Good-quality assessment data, and the sensitive but systematic feeding of this to parents and the public, have an important role to play in helping local and national communities develop realistic expectations concerning the outcomes of MLPS.

*c Informing national policy development:* Thirdly, good-quality assessment data is required in order to enable national policy makers to develop appropriate aims, attainment targets and curricular pathways, both within primary school and also connecting primary to secondary education. Indeed, ‘continuity’ (known in some countries as ‘articulation’) from primary to secondary education emerged as possibly the main unsolved problem in MLPS across the many research projects that contributed to the European review (Blondin *et al.*, 1998). Continuity is as pertinent in the Netherlands – with its years of MLPS experience, its concentration on one particular foreign language (English) and its high levels of exposure to that language within Dutch society – as to Scotland – with its much more limited experience, its four MLPS languages and its relative lack of societal exposure to any of these. The very scope and seriousness of the MLPS initiatives prompts national policy makers to attempt to establish national aims, guidelines and norms. It therefore becomes important to develop assessments that show what learners can do, so as to inform the development of such guidelines and/or to check their validity once implemented.

*d Self-evaluation:* Fourthly, in a number of states within the EU the notion of ‘self-evaluation’ (also known as ‘self-assessment’) is being promoted as central to national policy for education. This is intended to apply at a number of levels: e.g., the self-evaluating local authority, school, department, teacher and learner. In Scotland, all state schools are required to take account of the national policy document ‘How good is our school?’ (SOEID, 1997) which sets out performance indicators, each at a number of levels, covering a wide range of aspects of school life. Visits by HM Inspectorate check the extent to which a school is adequately engaged in evaluating itself against these national indicators. The purpose of this approach is intended to be educational, i.e., promoting self-awareness and self-development, rather than instrumental, i.e., promoting external accountability (for

which nationally there are other procedures). Integral to this process of self-awareness must be not only teachers' assessments of their pupils, but also pupils' and teachers' assessments of their own learning, teaching and development. Assessment – including assessment in MLPS – then becomes a major element by which a school, and those in it, are enabled to understand more about themselves.

If the above arguments have established the importance of MLPS assessment, in the next section we turn to the main problems which confront those who at present wish to develop MLPS assessments for implementation across schools.

## II Current problems in MLPS assessment

### *I Variability in context*

The European research review of MLPS (Blondin *et al.*, 1998) confirms the huge degree of contextual variability across and within member states. It has already been stated that there may be wide differences in aim and approach (e.g., aiming to 'sensitize' children to language by possibly introducing them to more than one new language as opposed to aiming to help them make progress in 'learning' one particular language), but there is also enormous variability in relation to factors such as:

- starting age;
- amount of time available per week;
- number of teachers available;
- teachers' foreign-language proficiency;
- teachers' knowledge of how to teach another language;
- degree of support available to teachers;
- extent to which national or local aims for MLPS are specified, understood and agreed;
- extent to which MLPS is taught as a separate primary school subject or is embedded in the wider curriculum that children experience;
- extent to which school ethos supports MLPS or, in fact, perceives it as a problem imposed from the outside; and
- extent to which staff in secondary schools are supportive and seek to build on what children bring with them from their MLPS experience.

This massive variability makes it difficult for teachers to collaborate across schools and poses problems for researchers in identifying what is common from one school to another. If commonality across schools on at least some key dimensions of language teaching cannot be established, then testing across schools becomes much more problematical.



Another aspect of contextual variability to take into account is the extent to which MLPS is a new phenomenon or has a supportive history. The Netherlands, for example, (Edelenbos and Suhre, 1996) has some 25 years' experience of attempting MLPS, and almost exclusively for one language (English), whereas other countries are either starting or are making a restart after a bad experience in the 1960s, and attempting a number of foreign languages in the process (e.g., in Scotland the national initiative covers French, German, Spanish or Italian). In some countries, therefore, there are publicly shared understandings about MLPS, whereas in most parts of the European Union this is not the case.

## *2 Embeddedness in a flow of events*

Empirical evidence from classroom observations (Low *et al.*, 1993; 1995) shows that in Scotland many if not all classroom activities at primary school are interlinked. Foreign-language activities therefore do not exist in isolation but are intended to fit into a wider pattern of pupils' learning activity. For example, a first-language activity in mathematics may lead to foreign-language work on numbers in relation to (say) a geography task. The concepts underlying these numbers will have been established through the first language, and the foreign-language activity will have the double purpose of not only enabling the pupils to learn these numbers in the foreign language but also of strengthening their understanding of the underlying concepts by extending their application to a new and different domain. The pupils' command of these foreign-language numbers may then be further exploited some time later in various other areas of the primary school curriculum, e.g., within a project that draws on history, art and drama that is undertaken mainly in the first language but that allows for various foreign-language activities to be situated within it, such as acting out a historical scene or designing an imaginary village. This natural flow of events in which the foreign language pops in and out of relevant classroom activity reflects a view of the primary school curriculum in which the universe of children's knowledge is not divided into discrete areas called 'subjects' but is organized more holistically into broader areas that allow children to integrate a variety of different experiences. It is only when they move to secondary school at the age of 12 that recognizably separate subjects appear.

A consequence of this deliberate 'embeddedness in a flow of events' is that children rarely begin a foreign-language activity 'cold' or unprepared, or solely in relation to previous foreign-language activities. On the contrary, they have a great deal of conceptual



knowledge that is currently being developed within their curriculum as a whole through their first language that they can bring with them into their foreign-language activities. However, this very embeddedness constitutes a problem when it comes to testing, since in a test there may not be time in which to embed the assessment activity in a series of prior activities. In the case of primary school children an assessment task is unlikely to be valid unless it represents a type of activity with which they have some familiarity; however, in addition, if they are asked to make a 'cold start' in an assessment task, when they are accustomed every day to being 'warmed up' for it cognitively as well as linguistically, then questions must arise about the validity of the process.

### *3 Children's relative lack of prior cultural knowledge*

Various researchers, e.g., Ausubel (1964: 421), have argued that one of the reasons why children tend to be less effective than older students in learning a foreign language in school conditions, is that they possess only a limited repertoire of concepts about the world. Older learners have generally acquired a richer network of concepts and can map their foreign language learning on to this. In creating assessments for MLPS pupils across schools, therefore, care has to be taken in ensuring that children are being tested for their language and not primarily for their general cognitive capacity or their knowledge about the world. Thus far, this point has been argued in relation to differences between younger and older learners, but there is also the dimension of social disadvantage within the category of younger learners. In certain parts of Scotland, as in all other countries, there are children living in conditions of poverty, exclusion and poor health. Such children may not have encountered certain concepts about the world that may be taken for granted in others.

To offer one specific example, in the development of a reading test for young learners some passages were found on the internet written by French youngsters on the topic of the film *Titanic* which, in the late 1990s, had gained international prominence. The language of the internet texts was authentic and interesting and seemed to lend itself very well to a test of reading for children in late primary or early secondary. However, when considering information such as this, one has to ask oneself what would happen in the case of those immature or socially disadvantaged young children who had not seen the film and/or who knew nothing about the real, original Titanic disaster as distinct from the film. Most children would have heard about the film and some would have seen it, but to what extent is it valid to build

in cultural content such as this when developing foreign-language assessments for children across schools?

#### *4 Relative unfamiliarity of primary school teachers with L2 testing*

At present there is little evidence to suggest that primary school teachers, not only in Scotland but also in many other countries of the European Union, possess a good working knowledge of how pupils' emerging competence in a modern language might be assessed. Very different evidence from three counties may be taken to illustrate this one point. In Scotland practising teachers received 160 hours of intensive training, in order to equip them for teaching MLPS, mainly geared to developing their proficiency in a particular foreign language, and to a lesser extent to helping them develop appropriate teaching methods. However, there is nothing built into the national training programme that focuses on children's language development, e.g., theories about error, interlanguage, age-factor, proficiency, communicative competence, language description or the possible relations (e.g., positive and negative transfer) between the first language and the second language. There is therefore nothing on how language development might be understood, described, measured or assessed. In France a recent major national initiative (e.g., Luc, 1996; Ministry Group of Experts, 1996) concentrated on making video material available to pupils from age 7 onwards, for 15 minutes per day, with the teacher playing a mediating role rather than providing the language input himself or herself, on the grounds that most teachers did not possess an appropriate command of the language. This raises complex and unanswered questions as to how children's progress might be assessed when they are expected to comprehend a foreign language in which video input rather than the teacher plays a major role. Until recently in Germany there was a tradition (Kubanek-German, 1996) of not assessing pupils' foreign-language progress at primary school. The prevailing view was that MLPS was a valuable educational process that would only become measurable later at secondary school. On the other hand, with a move towards beginning MLPS at age 6 as opposed to 8, there has been renewed interest in developing assessment tasks in order to gauge pupils' progress during their primary school period; see, for example, Kahl and Knebler's (1996) excellent study which developed useful task-types, criteria and procedures for different age groups and which yielded valuable empirical data on pupils' performance in relation to these.

The above examples suggest strongly that, for primary school teachers, thus far the priority has resided in developing their competence

and confidence in the foreign language and in finding ways of introducing it effectively in their classrooms, with little or no support available to them for developing the knowledge and skills that would enable them to assess their MLPS pupils. It follows then that, if valid and reliable assessments are to be developed, 'experts' from outside primary schools are likely to be required, desirably working in collaboration with primary school teachers.

However, it is precisely at this point that anxieties may begin to surface in teachers' minds. They begin to ask what and who these assessments are for; for example:

- 'Are they testing my pupils, or are they testing me as a teacher with my limited competence in the language?'
- 'Will I be expected to perform either in class or in a special test that these outsiders have contrived?'
- 'Is it fair to be attempting any assessment at all, given that we are all still learning how to do the job properly and need more time and resources?'

Before long, an experience for which (in Scotland at least) primary school teachers had volunteered and from which they were deriving professional satisfaction can assume a more threatening aspect. Sensitivities of this sort are well documented in Low (1997).

### *5 Lack of consensus concerning proficiency and its development in respect of MLPS pupils*

Finally, a shared view does not yet exist concerning what is meant by proficiency in the case of MLPS pupils, e.g., what its components are, how they relate to each other and how they develop. This is not surprising, partly because the aims of MLPS are not clear, partly because the context for MLPS is so variable, partly because teachers have had very little training in how to assess MLPS pupils, and partly because relatively little research has been undertaken.

On the other hand multilevel proficiency frameworks, constructed on the basis of professional expertise and experience, already do exist that may offer some help. However, even as distinguished a scheme as the Common European Framework of Reference (Council of Europe, 1996) must be treated with some caution if applied to MLPS. For example, its most basic level of Spoken Production states: 'I can use simple phrases and sentences to describe where I live and people I know.' This may be fine as far as it goes but it does not reflect the songs, poems, games and aspects of mathematics, science, history, geography and drama that MLPS beginners soon experience through their foreign language, and which enables them to pull chunks of

language from their long-term memory store that can go well beyond 'simple phrases or sentences'. What then does a basic level of MLPS proficiency consist of? It is at present not clear, for example, what the relationship at that level is between production of short phrases based on internalized rules and the production of larger memorized chunks.

A danger of frameworks that are constructed on the basis of professional experience rather than research is, therefore, that they may reflect an idealized rather than a real notion of what proficiency is and how it develops. The national guidelines for early secondary school in both Scotland and England, for example, envisage grammatical progression in terms of increasing accuracy as well as increasing range of structure; yet, research in those two countries (Low *et al.*, 1993; 1995; Mitchell and Martin, 1997) suggests that progression to a higher level of proficiency may in fact for a while be accompanied by reduced rather than enhanced grammatical accuracy, an insight supported in a very different context by Pelzer-Karpf and Zangl (1997) who use the term *Systemturbulenz* ('system turbulence') to describe what happens to a learner's developing language system when it first attempts to cope with a higher level of functioning. It follows therefore that frameworks which set out hypothesized levels of proficiency should ideally be informed by specially devised prior research, but if this is not possible, then such frameworks must at all costs be evaluated and modified in the light of second language acquisition (SLA) research and assessments specially devised for this purpose.

### **III Assessments within the independent evaluation of the national MLPS pilot projects**

The detailed findings of the independent evaluation in Scotland have already been published (Low *et al.*, 1993; 1995; Johnstone *et al.*, 1996). The present discussion is not intended to cover matters such as research design and sample nor to present the findings, but simply to illustrate how pupils were assessed. A major aim of the evaluation was to compare the linguistic attainments of pupils who had received MLPS with those who had not. In fact, three groups were compared:

- Group A: those who had received MLPS for one year and who were now at the end of their first year of secondary education;
- Group B: those who had not received MLPS but who had begun a foreign language at secondary and were at the end of their first year;
- Group C: those who also had begun at secondary but who were at the end of their second year.

Twelve secondary schools were involved, drawing on a total of 72 associated primary schools. A secondary aim, to be addressed in subsequent years, was to track the extent to which MLPS pupils were making progress in their foreign language from Primary 6 (P6) through to Primary 7 (P7) (ages 5–7) and from Secondary 1 (S1) to the end of Secondary 2 (S2) (ages 12–14).

The research team's general approach to the evaluation is well described by Shohamy's comment on language testing: 'language ability is a broad and complex construct that cannot be fully measured by tests' (Shohamy, 1994: 135) and 'language testers could begin to expand their repertoire of procedures for collecting language data beyond the traditional test approach . . . [these could include] . . . judgement tests, observation of natural language use, documents and self-assessments' (Shohamy, 1994: 140); that is, a range of testing and more general observational procedures were adopted.

Each of the five problems in MLPS assessment outlined in the preceding section are discussed in turn below, so that an indication may be given of how the evaluation attempted to cope with them through this combination of testing and observation.

### *1 Coping with variability in context*

In the early years of the national pilots this problem was acute. Aims and teaching approach were not prescribed at the start but were intended to emerge in the light of good practice that classroom teachers gradually found to work. There were therefore major differences across schools in the language content that was taught, the extent to which the language was embedded in the wider curriculum, and whether or not reading and writing should be introduced. In order to cope with this variability, the research team devised three data-collection procedures that could be applied across all participating schools:

- a systematic classroom observation instrument;
- a paired-speaking task; and
- a vocabulary retrieval task.

All three were relatively content-free, i.e., they were vessels into which pupils could pour whatever language they were able to. Each will be briefly discussed.

*a Systematic observation instrument:* First, the systematic observation of classroom teaching in P6 and P7 and in S1 and S2 generated a picture of the main types of teacher-talk, pupil-talk, listening, speaking, reading, writing and nonverbal activity, tasks and levels of

initiation and response. It proved possible to develop a language profile of MLPS and non-MLPS classes, to compare P6 classes with those in P7, S1 and S2, and different categories of pupil. For this latter purpose, 6 pupils per class were tracked systematically through their lessons: high-attaining, middle-attaining and low-attaining boys and girls respectively, thereby detailing the participation of pupils in lessons according to gender and language attainment (as perceived by the normal class teacher) and whether or not they had received MLPS.

*b Paired-speaking task:* Second, the paired-speaking tasks were the research team's main testing procedure. Pupils were asked to come in pairs rather than individually because talking in a foreign language with an unknown adult in a private room could easily be anxiety inducing, and the paired format allowed for different types of talk to come into play – e.g. researcher–pupils, pupil–researcher, pupil–pupil – and corresponded to the patterns of talk with which classroom observations had shown pupils to be familiar. Initially, paired-speaking was used for the MLPS vs. non-MLPS comparative study and in later years was used to chart pupils' progress in the language from P6 to S2 within the MLPS cohort; this was mostly on a cross-sectional basis although there was also a small longitudinal element. Pairs were in one of 6 possible categories – which, for convenience, are termed here: high-attaining, medium-attaining and low-attaining boys and girls respectively – to enable judgements to be made about how, across the range of schools, MLPS and non-MLPS pupils were performing according to gender and perceived attainment (as judged by the normal class teacher).

*c Vocabulary retrieval task:* Third, in the vocabulary retrieval task pupils were asked to say whatever words or phrases came into their head in relation to topics with which they were familiar. The researcher entered these into a computer and the pupils could see the results of their free associations. This stimulated them to add to their lists. The rationale for this task rested on the assumption that there was a difference between the language resource that was 'available' to pupils in their long-term memory and the language they would produce when under the constraint of engaging in meaningful real-time interactions. This corresponds to the distinction between *disponibilité* ('availability') and *fréquence* ('frequency') of the famous 'le français fondamental' study of the 1950s. The paired-speaking task picked up the language that pupils used, whereas the vocabulary retrieval task conveyed some sense of what elements of language pupils were able to summon up when under no pressure to use language for communication.

The systematically applied classroom observation study revealed much more commonality across schools in relation to processes of classroom interaction than to language content. Indeed, language content was highly variable from one school to another. It was therefore considered inappropriate to develop common tests of listening or reading comprehension, since common content would have been essential for these. Instead, schools were invited to provide their own evidence of tests of reading, listening and writing that they administered, and the researchers (having assured the schools of complete anonymity) were thus enabled to comment on these in their reports, e.g., making statements as to what areas of content appeared to be covered and what sorts of reading appeared to have been developed.

With regard to spoken interaction, on the other hand, the research team had the three relatively content-free instruments described above that enabled a large amount classroom and test data to be collected for purposes of making comparisons across schools, despite the high degree of variability in context and language content.

## *2 Coping with 'embeddedness in events'*

The problem of testing pupils 'cold' when they were accustomed to being 'warmed up' for language activities that were 'embedded in a flow of events' was alleviated, though not solved, by viewing the paired-speaking task as a set of related activities rather than as one single activity. It catered for socializing talk; for conversations in which pupils had the opportunity to talk about topics of their own choice with the researcher, 'oiling the wheels' of the exchange and occasionally probing further; and it gave the pupils an opportunity to elicit information from the researcher. In this latter regard, the researcher was able to talk more fluently, at a higher level and at greater length in response to the pupils' questions, thereby introducing an element of listening comprehension into the overall interview. There were also opportunities for pupils to provide extended narrative talk, drawing on this aspect of their primary school MLPS experience. This pattern of one activity flowing naturally into the next, in an overall foreign-language interaction that lasted some 20 minutes per pair, but with each activity capable of being separately assessed, was systematically applied across all of the age groups and classes in the study.

## *3 Coping with pupils' lack of cultural knowledge*

This problem was greatly reduced by not including common tests of listening or reading comprehension, for these would certainly have



contained a range of cultural concepts as well as items of language with which some pupils in some schools were not familiar. The strategy of concentrating on schools' own assessments of pupils' reading (see above) meant that children were being assessed there on cultural concepts and language items that their teachers considered appropriate, and that probably reflected teaching they had received. In the paired-speaking interviews the agenda for discussion was deliberately designed so as to be that of the pupils themselves, and they were informed of this in advance. That is, they were able to bring with them photos of their pet animals, family and favourite sports-persons, and to talk about the sorts of topics that they had encountered in class and with which they felt comfortable and familiar. It became the task of the adult interlocutor to follow this agenda rather than to set their own one, and to help them develop these topics through the patterns of talk with which the classroom observations had shown them to be familiar.

#### *4 Coping with teachers' relative unfamiliarity with language testing*

Since during the pilot phase most of the MLPS teaching was done by visiting teachers from local secondary schools who had received their full professional training as language teachers, this was less of a problem. Now that the pilot phase is over, however, and MLPS is being 'generalized' across all Scottish primary schools; teaching is being carried out by primary school teachers themselves, for whom MLPS is only a small part of their responsibility to their pupils, and as such the problem arises. This is briefly discussed in Section IV in the context of the Assessment of Achievement Programme (AAP) assessments that are due to take place within the 'generalization' phase.

#### *5 Lack of consensus concerning the nature of proficiency in the case of MLPS pupils*

The research team chose not to use any existing framework of levels of proficiency since none of these had been validated by research. For the paired-speaking assessments two approaches were adopted, each for a specific purpose. First, criteria (e.g., pronunciation-intonation, grammar control, range of structure) plus 3-point rating scales for each were found to be sufficient to discriminate between MLPS and non-MLPS pupils, i.e., Group A vs. Groups B and C as described previously. MLPS pupils were found to be ahead of their non-MLPS counterparts on all criteria except grammar control where

no differences were apparent. Secondly, a detailed linguistic description based on transcriptions of the full corpus of pupil utterances was used in order to compare pupils' performance in P6, P7, S1 and S2 in order to ascertain if the older groups were able to put a richer mixture of language into the common task. The analysis entailed a systematic count of the number and range of nouns, pronouns, verbs, structures, qualifiers, connectors and length of utterance (from 1 word to 6-plus words) that the pupils produced. The findings confirmed that the older pupils were indeed able to put *a greater amount* of language into a paired interview, but that little or no progression had taken place in *the range* of vocabulary, structure and discourse features. In addition, there was only limited evidence of pupils manipulating language, but substantial evidence of chunks of language that appeared to be learnt by heart.

#### **IV 'Generalizing' MLPS to all Scottish primary schools: current assessment issues**

The assessments conducted within the national MLPS pilot evaluation were constructed so as to be sensitive to the context in which MLPS was being developed: a fresh start after a traumatic experience 20 years earlier and in conditions that, though nationally supported, were highly variable. Since then the situation in Scotland has moved on. In roughly 80% of Scottish primary schools MLPS is now undertaken by primary teachers themselves (with the 160 hours of language training) rather than by visiting teachers from secondary. However, there are widespread concerns about the lack of further support for these teachers once they have received their training, and also about their continuing supply.

The next major step in assessing pupils at national level has taken place within the national Assessment of Achievement Programme (AAP). At present, pupils in P4, P7 and S2 are assessed in science, mathematics and English (one curricular area per year in a rolling programme). A nationally representative sample of Scottish schools is selected, as is a representative sample of pupils within these. Standardized tests are constructed by expert teams for implementation across the sample. No information on the performance of named pupils, classes, schools or local authorities is published. Instead, the purpose is to gauge the national norms for these three subject groups, to compare these for each particular subject every three years, to identify strengths and weaknesses in pupils' learning and to pinpoint any variation in performance according to categories such as gender and regional location.

A team, drawing on the MLPS researchers, was formed in order

to develop a first national AAP for modern languages. At the time of writing it is not possible to report on the findings of the first AAP for modern languages study, but the present brief account is intended to show how assessment is being tackled nationally now that the MLPS pilot phase is over. Since few schools have MLPS at P4, the AAP for Modern Languages focuses on P7 and S2. It constitutes a subsample within the large sample of schools used for AAP English, allowing comparisons to be made between the same pupils' performances in English and either French or German. It is intended that there should be a follow-up AAP for modern languages in 2000/01 or 2001/02 to allow those pupils assessed at P7 in the pilot AAP for modern languages in 1998/99 to be assessed again when they reach S2 two years later.

By the late 1990s the variability in context within Scotland has been somewhat reduced, since there are now clearer aims for MLPS and a consensus is building up on classroom approaches. For the AAP pilot in 1998 it proved possible, after extensive consultation with schools, to develop a common core of language content covered by the end of S2. This enabled AAP tests of reading and listening to be developed for application across schools, something that was inconceivable during the early years of the national MLPS pilot. In principle it therefore becomes possible to compare pupils at S2 who received MLPS with those at S2 who did not on these common tests of listening and reading. However, the common speaking and writing tests are relatively content free, as was the case with the paired-speaking test in the MLPS pilot, and so pupils who experienced MLPS have the opportunity to put a richer mixture of language into these content free tasks involving language production.

The AAP assessments at P7 continue to pose the same five problems that were discussed earlier pertaining to MLPS. The test developed for the 1998 pilot mainly involved spoken interaction, although there was an optional reading task for those schools that claimed to have developed this aspect. The spoken interaction test engaged pupils in small groups in processes that were known to be common across all schools. This necessitated considerable discussion with each particular school in order to establish what types of language the pupils had experienced. The tests were administered by visiting teams of two teachers rather than by the primary school teachers themselves, in order to remove any notion that the teachers themselves would be under the microscope for their own language performance. This aspect has required careful and sensitive handling, precisely because the primary teachers are now solely responsible for their pupils' MLPS rather than supporting the visiting teacher from secondary school.

An additional feature of the AAP pilot for modern languages at both P7 and S2 was the inclusion of a brief test of pupils' metalinguistic awareness, using an instrument especially devised to suit the Scottish context. It therefore became possible to establish the extent to which the metalinguistic performance of pupils at P7 and S2 correlated with their AAP performance in French or German and in English. Moreover, instruments were designed to gauge pupils' evaluations of the assessment tasks they had attempted and their more general attitudes to learning a modern foreign language.

## **V Conclusions**

In conclusion, two thoughts are suggested. First, with reference to the 'L2 learning' model of MLPS as previously discussed, a number of conditions would have to be fulfilled before valid and reliable standardized MLPS assessments could be developed across schools in any one country, far less across different countries. Among these would be:

- 1) more consensus on the aims and intended outcomes of MLPS;
- 2) a substantial body of in-school assessment involving teachers and others (Shohamy's category two) to take account of diverse local and institutional curricula, cultures and circumstances; and
- 3) a range of good SLA studies (Shohamy's category three) that would suggest exactly what the components of proficiency are in the case of primary school children learning a foreign language under particular conditions and how these build up.

This latter aspect is vital in fostering an understanding of how particular characteristics of pupil performance that reveal themselves in assessments may be interpreted in relation to research on second language development, e.g., the understanding that 'system turbulence' is not necessarily a bad thing. Once a sufficient number of studies pertaining to (2) and (3) had been published covering a range of different local and national contexts, these would lead to a more informed view than what we have at present concerning what is common and what is specific to particular local circumstances. This would, in turn, inform thinking about how larger-scale, more standardized MLPS assessments might be constructed.

Secondly, although this article has focused on the 'L2 learning' model rather than the 'sensitization' model, our preliminary work on metalinguistic awareness within the Scottish AAP pilots suggests the importance of developing assessment instruments that will measure this aspect within the context within which it has been developed. It would be helpful, for example, to track the progress in a foreign language at secondary school of pupils who had followed a 'sensitization'

approach at primary school in which metalinguistic awareness had been explicitly taught. Indeed, how would primary school 'sensitization' and 'L2 learning' pupils fare when compared with each other for 'L2 learning' at secondary? In addition, although an 'L2 learning' approach does not necessarily entail a lot of explicit teaching of metalinguistic awareness, what sorts of incidental metalinguistic awareness are acquired by 'L2 learning' pupils at primary and, indeed, also at secondary? To help answer important questions such as these, it is important in the future to develop portfolios of assessments that cover a wide spectrum of pupil learning that include proficiency in the language, metalinguistic awareness, attitudes and the other forms of cognitive and cultural development that are activated through the particular ways in which MLPS is implemented.

## VI References

- Ausubel, D.** 1964: Adults versus children in second-language learning: psychological considerations. *Modern Language Journal* 48, 420–24.
- Bailly, D. and Luc, C.** 1992: *Approche d'une langue étrangère à l'école. Études psycholinguistiques et aspects didactiques* (Teaching a foreign language at school. Psycholinguistic studies and didactic considerations), Volume 2. Paris: INRP.
- Balke, G.** 1991: Multilevel factor analysis of proficiency in English as a foreign language. Paper presented at the symposium Multilevel factor analysis: Applications to Education: Annual meeting of American Educational Research Association, Chicago.
- Blondin, C., Candelier, M., Edelenbos, P., Johnstone, R., Kubanek-German, A. and Taeschner, T.** 1998: *Foreign languages in primary and pre-school education: context and outcomes. A review of recent research within the European Union*. London: CILT.
- Burstall, C.** 1974: *Primary French in the balance*. Slough: National Foundation for Educational Research.
- Charmeux, E.** 1992: Maîtrise du français et familiarisation avec d'autres langues (Mastery of French and familiarisation with other languages). *Repères*, 6, 155–72.
- Council of Europe** 1996: *Common European framework of reference for the learning and teaching of languages*. Strasbourg: Council of Europe Education Committee.
- Djigunovich, J.M.** 1995: Attitudes of young foreign-language learners: a follow-up study. In Vilke, M., editor, *Children and foreign languages*. Zagreb: University of Zagreb, Faculty of Philosophy, 16–33.
- Donato, R., Antonek, J.L. and Tucker, G.R.** 1994: A multiple perspectives analysis of a Japanese FLES program. *Foreign Language Annals*, 27, 365–78.
- 1996: Monitoring and assessing a Japanese FLES program: ambience and achievement. *Language Learning* 46, 3, 497–528.

- Edelenbos, P. and Suhre, C.** 1996: English in Dutch primary education. In Edelenbos, P. and Johnstone, R., editors, *Researching languages at primary school: some European perspectives*. London: CILT, 47–57.
- European Commission.** 1995: *White Paper: teaching and learning: towards the learning society*. Luxembourg: Office for Official Publications of the European Communities.
- Genlot, S.** 1996: *L'enseignement des langues à l'école primaire: quels acquis pour quels effets au collège? Éléments d'évaluation: le cas de l'anglais* (Teaching languages at primary school: what gains for what effects at secondary school?). Les Notes de l'Iredu. Dijon: Institut de Recherche sur l'Économie de l'Éducation.
- Giovanazzi, A.** 1993: Sense or sensibility: the organisational imperative. In Arndsorf, D., Boyle, H., Chaix, P. and O'Neil, C., editors, *Foreign language learning in the primary school*. Triangle 11. Paris: Didier Erudition, 21–28.
- Helfrich, H.** 1995: Evaluation des Modellversuchs "Integrierte Fremdspracharbeit an Grundschulen in Rheinland-Pfalz" (Evaluation of the pilot scheme "Integrated foreign language learning in primary schools in Rhineland-Palatinate"). In *Entwicklung und Erprobung eines didaktischen Konzepts zur Fremdsprachenarbeit in der Grundschule: Abschlußbericht* (State Institute for Teachers Education and Professional Development: Development and Trialling of a didactic approach for foreign language learning in the primary school: Final Report). Speyer: Staatliches Institut für Lehrerfort- und -weiterbildung, 97–122.
- HM Inspectorate of Schools.** 1969: *French in the primary school*. Edinburgh: Her Majesty's Stationery Office.
- Johnstone, R., Low, L. and Brown, S.** 1996: Foreign language attainments in the Scottish pilot. In Edelenbos, P. and Johnstone R., editors, *Researching languages at primary school: some European perspectives*. London: CILT, 58–73.
- Kahl, P. and Knebler, U.** 1996: *Englisch in der Grundschule, und dann? Evaluation des Hamburger Schulversuchs* (English at primary school: what next? Evaluation of the Hamburg pilot scheme). Berlin: Cornelsen.
- Kubanek-German, A.** 1996: Research into primary foreign-language learning in Germany: a trend towards qualitative studies. In Edelenbos, P. and Johnstone, R., editors, *Researching languages at primary school: some European perspectives*. London: CILT, 3–15.
- Low, L., editor** 1997: *Modern languages at primary and early secondary school: support for the teacher*. Proceedings of the conference held at Scottish CILT, March 1997. University of Stirling: Scottish CILT.
- Low, L., Duffield, J., Brown, S. and Johnstone, R.** 1993: *Evaluating foreign languages in Scottish primary schools: report to Scottish Office*. University of Stirling: Scottish CILT.
- Low, L., Brown, S., Johnstone, R. and Pirrie, A.** 1995: *Foreign languages in Scottish primary schools: evaluation of the Scottish pilot projects: final report to Scottish Office*. University of Stirling: Scottish CILT.
- Luc, C.** 1992: *Approche d'une langue étrangère à l'école: perspectives de*



- l'apprentissage learning* (Teaching a foreign language at school: perspectives on L2 learning), Volume 1. Paris: INRP.
- 1996: *Première année d'initiation à une langue étrangère au cours élémentaire: constats, analyses, propositions* (First year of initiation into learning foreign language at primary school: findings, analyses, proposals). Paris: INRP.
- Ministry Group of Experts (Groupe national d'Experts)** 1996: *L'initiation à une langue au cours élémentaire, première année: rapport d'étape* (Initiation into a language at elementary school, first year: report on the first stage). Paris: Ministère de l'Éducation.
- Mitchell, R. and Martin, C.** 1997: Rote-learning, creativity and 'understanding' in classroom foreign language teaching. *Language Teaching Research* 1, 1–27.
- Nagy, C.** 1996: L'éveil au langage: contribution à une étude de l'activité métalangagière de l'enfant à l'école élémentaire (An awakening to language: contribution to a study of the metalinguistic activity of children at primary school). Unpublished PhD thesis, University of Grenoble.
- Peltzer-Karpf, A., Hasiba, U. and Zangl, R.** 1996: *Lollipop Program: Psycholinguistische Untersuchung zur vierten Klasse* (Lollipop Program: psycholinguistic investigation up to Class 4). Graz: Karl-Franzens-Universität.
- Peltzer-Karpf, A. and Zangl, R.** 1997: *Vier Jahre Vienna Bilingual Schooling: Eine Langzeitstudie*. (Four years of Vienna bilingual schooling: a longitudinal study). Vienna: Bundesministerium für Unterricht und kulturelle Angelegenheiten, Abteilung 1/1.
- Pinto, M.A.** 1993: Le développement métalinguistique chez les enfants bilingues: Problématiques théoriques et résultats de recherche (Metalinguistic development in bilingual children: theoretical issues and research findings). *Scientia Paedagogica Experimentalis* 1, 119–48.
- Shohamy, E.** 1994: The role of language tests in the construction and validation of second-language acquisition theories. In Tarone, E., Gass, S. and Cohen, A., editors, *Research methodology in second-language acquisition*. Hillsdale NJ: Lawrence Erlbaum, 133–42.
- SOEID (Scottish Office Education and Industry Department)** 1997: *How good is our school? Self-evaluation using performance indicators*. Edinburgh: Stationery Office for HMI Audit Unit.