ASSESSING MODERN LANGUAGES ACHIEVEMENT: A SCOTTISH PILOT STUDY OF LATE PRIMARY AND EARLY SECONDARY PUPILS

Richard Johnstone, Jane Cavani, Lesley Low, Joanna McPake

Edited by Lesley Low

December 2000

ACKNOWLEDGEMENTS

ABBREVIATIONS/GLOSSARY

ORGANISATIONS

Scottish CILT   Scottish Centre for Information on Language Teaching and Research
SCRE   Scottish Council for Research in Education
SOEID   Scottish Office Education and Industry Department
SEED   Scottish Executive Education Department (successor to SOEID)
SCCC   Scottish Consultative Council on the Curriculum, now *Learning and Teaching Scotland*

CURRICULUM AND EXAMINATIONS

AAP   Assessment of Achievement Programme
MLPS   Modern Languages in the Primary School (usually 10-12 year olds)
5-14   Scotland's national curriculum equivalent for the primary and early secondary
                 stages
Standard Grade   Scotland's public examinations at age 16 (GCSE equivalent)
Higher Still   Scotland's programme for post-16 education

OTHER TERMS USED IN THE REPORT

P4   Primary Four (pupils aged 8/9 years)
P7   Primary Seven (pupils aged 11/12 years)
S2   Secondary 2 (pupils aged 13/14 years)
ML   Modern Language
FL   Foreign Language
FLES   Foreign Languages in the Elementary School
MLA   Meta-linguistic Awareness
NS   Native Speaker (assessor)
NSS   Non-Native Speaker (assessor)
SEN   Special Educational Needs

CONTENTS

Acknowledgements

List of abbreviations

PART ONE: THE CONTEXT

PART TWO: THE ASSESSMENTS

PART THREE: THE FINDINGS

APPENDICES

CHAPTER 1
THE SCOTTISH CONTEXT FOR MODERN LANGUAGES

PRIMARY MODERN LANGUAGES IN SCOTLAND

More than any other part of the UK, Scotland has committed itself to the introduction of a modern language at primary school. After an initial attempt to introduce French in the upper primary during the 1960s which was considered to have been less than successful (HMI, 1969), modern languages more or less disappeared from primary education from the mid-1970s to the late 1980s. The feelings of failure were reinforced by the negative evaluation of primary school French in England (Burstall et al, 1975).

However, in Scotland a more optimistic view came into being in the late 1980s, arising from a perception that the single Market (1992) was just 'around the corner' and that it would be in Scotland's interest to increase its national capability in languages.

This was to be achieved by two complementary means: an 'earlier start' which entailed beginning in Primary 6 (students aged 10) rather than in Secondary 1 (students aged 12), and a 'languages for all' extension up to and including Secondary 4 (with students aged 16). Thus, what had effectively been a 2-year experience for many students (S1-S2) became one of six years (P6-S4).

The re-introduction of languages in Scottish primary schools began with pilot projects at national and regional levels that were initiated in the late 1980s. These became known as the national and regional MLPS Pilots, with MLPS standing for Modern Languages at Primary School and 'national' referring to Scotland. The pilots were conducted from 1989 to 1995 and were independently evaluated (Low, Duffield, Brown and Johnstone, 1993; Low, Brown, Johnstone and Pirrie, 1995).

The pilots were succeeded by a generalisation phase intended to place a modern language (which could be French, German, Spanish or Italian) in all Scottish primary schools from P6 onwards. Whereas in the pilots much of the teaching had been undertaken by trained 'visiting teachers' from the neighbouring secondary school, in the generalisation phase the task fell to primary school classroom teachers themselves. In order to prepare for this new task which would be blended into their normal duties in teaching all or most of the primary curriculum to their particular classes, the teachers received in-service training amounting to 27 days distributed over four terms (approximately 1.25 years). The training was mainly geared to the development of skills in the language (in some cases the teachers were building on existing language skills and in others they were learning ab initio).

During the second half of the 1990s the generalisation phases was progressively introduced in this way to Scottish primary schools. However clear the strategy may have seemed from 'on high' in the minds of those developing it, the landing was far from soft when the plans hit the ground. The mapping exercise of provision on the ground conducted by the two National Development Officers (Tierney and De Cecco, 2000), showed how highly variable the provision was within and across schools and also in initial primary teacher education. The generalisation phase was indeed accompanied by a public and often heated debate concerning the levels of provision, training and professional development that would be needed if primary school teachers were to be able to make a real success of the job.

The debate was heightened following publication of the HMI Standards and Quality Report on Modern Languages (HMI, 1998) which for the first time commented on MLPS. Although several instances of good practice were identified, the report was widely perceived as being more negative than positive and the situation of languages was described in the Foreword to the report as 'far from reassuring'. Whereas much of the MLPS debate had centred on provision, processes and support, no evidence on the outcomes of the generalisation phase was available in terms of what students could in

fact do with the language they were learning. The then SOEID (Scottish Office Education and Industry Department) decided it would be appropriate to commission an independent study of students' attainments. This led in the first instance to the present study which takes the form of a Pilot Assessment of Achievement Programme (AAP), and subsequently to the full-scale national AAP which is due to take place in Spring 2001

AN AAP IN MODERN LANGUAGES

The purpose of the Assessment of Achievement Programme (AAP) is to provide information on the achievements of pupils across a representative sample of Scottish schools. Thus far, the curricular areas have been English, mathematics and science in a rolling programme (one year English, the next science, the next mathematics, then English again, and so on). Pupils are assessed at Primary 4 (P4), Primary 7 (P7) and Secondary 2 (S2) in order to provide cross-sectional data for the same curricular area, though a longitudinal element can be built in by tracking the same pupils from (say) P4 English to P7 English or from S2 mathematics to Standard Grade mathematics. The information allows a picture to be built up of the impact of particular within-sample factors (e.g. gender), of strengths and weaknesses in pupils' achievements within any given administration and of any national improvements (or the reverse) from one administration to the next, three years later. The information is not used for evaluation of the effectiveness of particular pupils, teachers, schools or local authorities, since it is the 'national picture' that is important.

So far as an AAP in modern languages is concerned, two key characteristics of modern languages at school in Scotland must be borne in mind:

1. Most pupils at P7 will have been learning the particular language for only a short period (mostly beginning at P6), with relatively little time devoted to it per week (detailed information on this for the pilot AAP sample of 20 primary and 20 secondary schools is given in Chapter 2). It follows that their proficiency in the language will be elementary. On the other hand, the same pupils will have had a much longer period of time for learning mathematics, science or English, and in addition these three curricular areas are much more part of our everyday culture and life in Scotland than are modern foreign languages, with the consequence that pupils may be expected to acquire relatively little 'incidental learning' of a modern language from outside their school. This contrasts starkly with primary pupils in Holland who (c.f. Blondin et al, 1998) acquire roughly half of their English from societal rather than from within-school sources.

2. Probably as a consequence of the above, pupils in the initial stages of learning their language (from P6 to S2) are heavily dependent on their teacher as the sole or the main source of foreign language input and interaction. It has been common practice among languages teachers in Scotland to devote considerable attention and energy to 'setting up' or 'scaffolding' their pupils' classroom activities, so that many of these are dependent on those that immediately precede them, with the preceding activities having introduced a great deal of prior information that is relevant to the one that is next in line. This may indeed be appropriate for purposes of learning and teaching, and indeed many observers of primary FL classes have been impressed by what pupils are able to do in the language on this carefully scaffolded, connected 'chain of activity' approach, possibly via a project on which pupils have been working for days in a variety of curricular areas. However, it poses a problem with regard to assessment, particularly when this is of the 'one-off' variety as represented by an AAP. The problem is how in a short space of time to provide a sufficient degree of 'scaffolding' that will allow pupils at an elementary level of language proficiency to 'get into' their assessment tasks while at the same time 'testing' rather than 'teaching' them, including testing their ability to summon up relevant prior information for themselves. We have to accept that AAP assessments in modern languages are unlikely to be able to cover all relevant aspects of pupils' classroom performance. What we think they do assess, however, is pupils' ability to transfer their learning from a connected classroom context to one in which there is inevitably less support. Provided that the experience is not too anxiety-inducing, this may in fact provide a better

picture of what it is that pupils have internalised to a deeper level that allows them to 'perform' when there are fewer supportive props.

For the purposes of devising assessments at the P7 and S2 stages, it is appropriate to comment on three further related problems:

1. Although national 5-14 Guidelines for Modern Languages (Scotland's equivalent of a national curriculum) had been in existence for several years, they deliberately did not include languages at the primary stages. As a consequence, there was a fair degree of uncertainty across schools as to what it was reasonable to expect pupils to achieve and how these achievements might be assessed.

2. There were uncertainties in many people's minds concerning the future of MLPS, and in particular concerns were regularly expressed concerning the provision of an appropriate number of MLPS-trained primary teachers and their further language enhancement and professional development once their MLPS training was completed. At a Scottish CILT conference in 1997, for example, when a first presentation was given of what a pilot AAP in modern languages might consist of, several reservations were expressed from the floor through questions such as 'How can this be done when there are no 5-14 Guidelines for MLPS?' 'Don't we really need to be concentrating on development rather than assessment for the next few years, in order to ensure that MLPS really beds down?' and 'What will be assessed: pupils or their teachers?'

3. There was relatively little published international research evidence available that seemed directly relevant to Scottish concerns. It is true that impressive multi-level frameworks have been developed for foreign language proficiency, e.g. in Australia and the United States, and that excellent work has been done in order to validate these frameworks and develop reliable and valid procedures for assessment tasks related to each level. None of these, however, have had MLPS learners primarily in mind. Closer to home, the Council of Europe framework has initially at least proved more relevant to the intermediate levels of language proficiency that are within the reach of Scottish pupils, e.g. Higher Still, than it has at the elementary level, where its descriptors of language performance do not appear to coincide with what Scottish MLPS pupils have been observed to do. (See Johnstone, R 2000.) [1]

An obvious conclusion from the above was that, in addition to the agreed aims of the project as set out in the next part of this chapter, an additional aim would be to work with teachers, local authority representatives and others in order to create a climate in which AAP assessments in modern languages would be viewed positively.

In this sense then, the project would have implications for teachers' professional development, and this remained an implicit aim throughout the pilot. A major argument in favour of AAP assessments would be that they would eventually provide reliable and valid information on what Scottish pupils were able to do, four and two years before they took their first national examination at Standard Grade. A key conclusion of the SOEID-funded study on factors affecting the decline in uptake of modern languages at Higher (see McPake et al 1998) centred on the importance of developing clear and shared understandings of what it was reasonable to expect pupils to achieve in the Scottish context. The researchers found that uncertainty in some cases led to false expectations and anxiety. AAP data at these two prior levels might help considerably in this respect.

---

[1] Johnstone, R 'Context-sensitive assessment of modern languages in primary (elementary) and early secondary education : Scotland and the European Experience' in Language Testing, 17, 2 pp 123-143. This is a special issue of this international research journal which is dedicated to the assessment of young learners at primary school. It indicates, however, that a substantial body of research on how to assess young second-language learners does not exist. In his paper, Johnstone (2000) outlines what the key problems are and how these have been confronted in Scotland.

AIMS OF THE AAP PILOT STUDY

The aims of the study can be summarised as:

1. to develop appropriate assessment instruments and procedures for a pilot AAP in modern languages at P7 and S2

2. to identify an appropriate sample of primary and secondary schools, drawing on the larger national AAP sample for English in 1998

3. to implement the assessments in May-June 1998

4. to evaluate the effectiveness of the instruments and procedures that had been adopted, so that these if necessary might be refined for a possible larger-scale AAP in modern languages in 2001

5. to analyse and report on the attainments of pupils at P7 and S2 in French and German, though bearing in mind that this was a pilot study only and that the sample would inevitably fall short of being nationally representative.

Two other possible 'added value' benefits were conceived:

1. to explore in a tentative way any possible links with children's attainments in English, possibly by means of assessing their metalinguistic knowledge (i.e. their implicit or explicit 'knowledge about language' including knowledge of concepts such as 'noun', 'verb', 'adjective') and their metalinguistic awareness (i.e. the extent to which they were consciously aware of possessing that knowledge); and

2. to provide initial research-based information on content coverage, assessment instruments and procedures and pupils' attainments that might be of value to the then Scottish Consultative Council on the Curriculum's (SCCC)[2] 'Review and Development Group' who were engaged in revising the national 5-14 Guidelines for Modern European Languages to include the P6 and P7 stages.

PRE-PILOT WORK (1995/97)

SOEID interest in the possibility of an AAP pilot for modern languages was first expressed in 1995 following the English survey of that year. The modern languages departments in the secondary schools that had taken part in the English survey were contacted and invited to take part in a small scale project based at Scottish CILT. The aim of the project was to consider the broad issue of assessment at the S2 stage with a view to developing assessments to inform any future survey in modern languages. Given that AAP was not known among secondary teachers of languages, the response was not high, but from a core of about 20 schools, 13 went on to take part in the project, which lasted over two years from autumn 1995 to 1997.

The modern languages departments in the 13 schools were visited during November-December 1995 by a researcher who met the staff concerned, collected S2 assessment materials and copies of assessment policy and procedures and gauged the extent to which 5-14 Guidelines informed assessment practice in modern languages in those schools. The main work of the project was achieved through a series of full day meetings which were held 2 or 3 times a year at Scottish CILT and attended by the principal or assistant principal teachers from the 13 schools. The teacher group

---

[2] SCCC has now been integrated along with the Scottish Council for Educational Technology into a new organisation entitled Learning and Teaching Scotland.

discussed assessment issues in general and looked in particular at the template devised for the English AAP assessments in order to gauge its appropriateness as a model for modern languages. This involved selecting topic areas which would form the basis for a set of assessment tasks in the four language skills with an non-assessed bridging activity to link the comprehension tasks of listening and reading with the productive tasks of speaking and writing.

Three sub-groups of teachers were formed based on geographical proximity, and each group devised a set of assessment tasks covering the 4 language skills including procedures and criteria for assessment. The assessment tasks and the bridging activity were tried out with S2 groups across the various schools by individual teachers and sometimes with the researcher present. The outcomes of these trials, particularly the working of the procedures and set tasks in practice, provided a body of valuable evidence on which the subsequent AAP pilot could draw. The teachers who had taken part in the pre-pilot work could legitimately be considered an 'expert group' and as such they were invited to attend two full day meetings at Scottish CILT to validate the final pilot S2 tests in terms of:

- suitability of the foreign language content
- appropriateness of the tasks (e.g. for different ability levels)
- feasibility of the test procedures

In addition, one of the principal teachers organised the trialling of the full range of final pilot assessments in his school across two half-day sessions. (See Chapter 4 for details of this process.)

DEVELOPMENT OF THE ASSESSMENTS

The short time scale for devising the pilot assessments (3 months, January-March 1998) and the equally tight period for validating and implementing them (April-June 1998) required the researchers to build up an extended team, many of whose members could be involved in every stage of the process. At the outset, the research team had decided on a visiting assessor approach, which meant that members of the extended team would go out to schools to conduct the assessments rather than ask teachers to conduct the tests in their own or others' schools. The extended team was made up of:

- researchers and the project director at Scottish CILT
- researchers from SCRE, both modern linguists with previous or current secondary teaching experience
- native speakers, one French and one German, who were teaching in secondary schools
- native speakers, one French and one German, who were tutors on the MLPS national training course

The involvement of native speakers in the test development and implementation process was designed to bring authenticity, both to the foreign language content of the tests and in the interaction with pupils during the speaking tests at the implementation phase. Also, in evaluating pupil performance, native speakers are widely considered to take a more generous (more realistic?) view of pupil performance than non-native speakers and this was felt to be a valuable perspective to include as part of the pilot.

The assessment writing team consisted of all of the above and the design and production of the full range of assessment tasks was achieved through a series of regular twilight meetings during the first three months of the project. The native speaker members of the team devised individual tasks in the four skill areas based on a framework which had been previously agreed by the full team.

The decision to opt for visiting assessors to each of the pilot schools imposed a time limit of a day in which to conduct the full battery of tests. There was therefore a tension between the need to

maximise the number of students taking part in the tests, especially for the speaking, and the time available. A number of options were considered for the P7 tests including assessing the students in small groups, which would have the additional advantages of permitting some interaction among the students themselves and might be less daunting for the children concerned, who would be interacting with two unknown adults. However, it was felt that there would need to be some measure of individual student performance even at the P7 stage and that groups of three or four would make this very difficult. It was decided to opt for pairs of pupils to try to meet the various needs of students and assessors.

In secondary, however, it was felt that students would be able to cope with an individual speaking test and that it was important to gain a measure of each student's performance. In order to maximise the very tight time constraints, the students would do certain tasks with the native speaker assessor (NS) and then move to the non-native speaker assessor (NNS) to do others. Only half of the S2 sample would be involved in the speaking tests, usually 12 per school.

Another choice which had some implications for the development of the tests was how they would be recorded. Video-recording of the P7 tests and the speaking tests at S2 was seriously considered but then rejected for the pilot because it might prove too obtrusive and have a detrimental effect on student performance. There was also the additional technical burden which video-recording was bound to impose on the visiting assessors and the schools themselves, and so it was decided to audio-record the tests for the purposes of post-hoc scoring and ask the assessors to do some real-time scoring of student performance during or after the tasks. Transcription was considered to be easier from audio-recorded rather than video-recorded material, although any non-verbal forms of communication between students and assessors could not be captured.

LAYOUT OF THIS REPORT

The first two chapters of the report, including this introductory chapter, set the scene and context for the teaching, learning and assessment of modern languages in the primary and early secondary sectors in Scotland. Chapter Two gives details of the school and pupil samples involved in the pilot and reports teachers' and pupils' views gathered by questionnaire on a wide range of issues pertaining to current provision and practice of ML teaching in Scotland.

The second part of the report deals with the development, trialling and evaluation of the assessments and associated procedures. It draws on feedback from the assessors, schools and pupils. Chapter 3 covers the Primary 7 assessments and Chapter 4 the Secondary 2 assessments.

The final part of the report presents the detailed findings from the pilot. Chapters 5-8 deal with the skills of listening, reading, speaking and writing respectively. The final Chapter provides a summary of the findings and a set of conclusions. These are reproduced in the form in which they were originally presented to the SOEID in April 1999.

IMPACT OF THE PILOT AAP REPORT

Although it is for others to assess the impact of our present study on general thinking and practice in relation to the assessment of modern language learners in late primary and early secondary education, we ourselves are able to point to what we consider to be three positive outcomes.

First, the original report (which was submitted to SOEID in April 1999) did serve its stated purpose in that it demonstrated to the satisfaction of the national authorities (SOEID and now SEED) that it was indeed feasible as well as desirable to assess learners at this stage of their languages development,

despite circumstances on the ground being far from ideal. This led to the commissioning of Scottish CILT to undertake the full-scale national AAP Survey due to take place in Spring 2001.

Second, the report was made available to the national working group that was developing revised Guidelines for Modern Languages at 5-14. The summary data on what students were able to do at different levels in Listening, Speaking, Reading and Writing[3] assisted the group in devising their descriptors for 5-14 levels C to E. This is in fact rather unique, since most national levels are not grounded in relevant prior research but are conceived by professionals drawing on their experience. In the case of the 5-14 Guidelines, both of these processes of research and professional experience were engaged.

Finally, the report was presented to the Ministerial Action Group on Languages[4] while it was preparing its report and recommendations concerning modern languages in primary and secondary schools. The Action Group's report, published by the Scottish Executive in December 2000, acknowledges the formative influence which the present study had on its thinking and influenced the

---

[3] These summary tables can be found at the very end of each of the Findings Chapters (Chapters 5-8).

[4] The Action Group on Languages was set up in November 1998 by the then Minister for Education in Scotland, Helen Liddell, with the remit to secure the place of modern languages in the Scottish school curriculum, following the concerns raised by HMI in their report on Standards and Quality in Modern Languages (1998).

REFERENCES IN THIS CHAPTER

Blondin, C., Candelier, M., Edelenbos, P., Johnstone, R., Kubanek-German, A., Taeschner, T., (1998). Foreign Languages in Primary and Pre-School Education : A Review of Recent Research within the European Union. London. CILT

Burstall, C., Jamieson, M., Cohen, S. and Hargreaves, M. (1974). Primary French in the balance. Slough, National Foundation for Educational Research.

H M Inspectors of Schools. (1998). Standards and Quality. Primary and Secondary Schools. 1994-98: in Modern Languages. Edinburgh: HMSO for the Scottish Office.

H M Inspectors of Schools. (1969). French in the primary school. Edinburgh, HMSO.

Johnstone, R. (2000). Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience. Language Testing, 17, 2, 124-144.

Low, L., Duffield, J., Brown, S. and Johnstone, R. (1993). Evaluating foreign languages in Scottish primary schools. Report to Scottish Office. 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. Stirling, Scottish CILT.

McPake, J., Johnstone, R., Low, L. and Lyall, L. (1999) Foreign languages in the Upper Secondary School, A study of the Causes of Decline, SCRE Research Report No 91, Edinburgh, Scottish Council for Research in Education.

Scottish Office Education Department (1993). Curriculum and Assessment in Scotland. National Guidelines, Modern European Languages 5-14. Edinburgh. HMSO.

Tierney, D. and De Cecco, J. (1999). Modern languages in the primary school. Research - Final Report. Glasgow, University of Strathclyde.

Action Group in concluding what kinds of entitlement for students, and training and development for teachers, would be needed if modern languages were to become a successful reality.

CHAPTER 2
THE SCHOOL CONTEXT

This chapter presents data on the schools and pupil samples involved in the pilot AAP, followed by contextual data gathered from the schools and the pupils themselves.

SAMPLE SCHOOLS AND PUPILS

The goal of the sampling procedure was to identify a group of pupils representative of the full range of schools and pupils across Scotland, in order to ensure realistic implementation and evaluation of the pilot assessment instruments and procedures.

THE SCHOOL SAMPLE

For logistical reasons, it was necessary to use a sub-sample of the 1998 AAP English language survey. This consisted of forty schools: ten schools for each language at each of the two stages involved. The school sample is represented in diagram form in Table 2a below:

Figure 2a: The school sample

	Primary 7	**Secondary 2**	**Total**
French	10 schools	10 schools	20 schools
German	10 schools	10 schools	20 schools
Total	20 schools	20 schools	40 schools

While the procedures for selecting the English language school sample ensured that the sample was nationally representative, the selection of the modern languages school sample was clearly constrained, since each school had to fulfil a number of conditions:

- be a participant in the 1998 English language survey;
- deliver French or German in sufficient numbers to provide a viable pupil sample;
- display a willingness to participate in the pilot project.

Consequently the pilot project school sample was selected in such a way that not all schools had an equal probability of being included. However care was taken to ensure that the sample included a range of schools in terms of location and size. Schools from 23 different local education authorities participated in the pilot project: 14 different authorities were represented for the primary sector and 16 different authorities for the secondary sector. Within the primary sector school size varied from 40 pupils to 550 pupils. Within the secondary sector it varied from 170 pupils to 1400 pupils.

THE PUPIL SAMPLE

Within each sample primary school, the intention was to select 12 pupils for involvement in the pilot. In each sample secondary school, it was intended that 24 pupils would participate. This 'ideal' sample is represented in Table 2b, below.

Figure 2b: The 'ideal' pupil sample

	Primary 7	**Secondary 2**	**Total**
French	10 schools *120 pupils*	10 schools *240 pupils*	20 schools *360 pupils*
German	10 schools *120 pupils*	10 schools *240 pupils*	20 schools *360 pupils*
Total	20 schools *240 pupils*	20 schools *480 pupils*	40 schools *720 pupils*

However, it was not always possible to assess the full number of pupils in each school, for a variety of reasons: school size, pupil absence, etc.[5] The 'actual' sample is represented in Table 2c.

Figure 2c: The 'actual' pupil sample

	Primary 7	**Secondary 2**	**Total**
French	10 schools *106 pupils*	10 schools *235 pupils*	20 schools *341 pupils*
German	10 schools *110 pupils*	10 schools *238 pupils*	20 schools *348 pupils*
Total	20 schools *216 pupils*	20 schools *473 pupils*	40 schools *689 pupils*

The intention was to pilot the modern languages assessments with pupils who had participated in the English language survey. The English language sample had been selected to be representative of pupils in all mainstream Scottish schools. The modern languages sample was therefore to be a sub-sample of this group.

---

[5] In the primary school sample for French, one of the schools could only provide three pupils, due to the small size of the P7 cohort. Another school could provide only seven pupils. The remaining eight schools provided the full 12-pupil sample. In the primary sample for German, one of the schools could provide only eight pupils, due to the small size of the P7 cohort. A further two schools could only provide nine pupils. The remaining seven schools provided the full 12-pupil sample. In the secondary school sample for French, one of the schools could only provide a sample of 19 pupils, while the remaining nine schools provided the full 24-pupil sample. In the secondary sample for German, two of the schools could provide a sample of only 23 pupils, while the remaining eight schools provided the full 24-pupil sample.

In selecting the sub-sample, it was decided to include only pupils with at least two years experience in the same foreign language: P6 and P7 for the P7 sample pupils, and S1 and S2 for the S2 sample pupils.

This criterion created some difficulties in the selection of pupils, particularly those studying German. Only some of the primary schools, and none of the secondary schools involved, delivered German to the entire cohort over the relevant two-year periods. In these schools, it was necessary to 'top up' the sample to the required number by requesting that the schools **involved select pupils of like gender** and ability from pupils not included in the English language sample. In a few cases, 'topping up' was also required in schools delivering French.

Thus approximately 87% of the primary pupils and 94% of the secondary pupils participating in the pilot AAP in French were also English language survey sample pupils.[6] The proportion of pupils participating in the pilot AAP in German who were also English language survey sample pupils was lower, most significantly in the secondary sector: 75% of the primary pupils involved and only 38% of the secondary pupils.[7]

All pupils did not, therefore, have an equal probability of being included in the sample. Nevertheless care was taken to ensure that both genders were, as far as possible, represented in equal numbers. This was not always possible normally due to the occasional need to use reserve pupils to replace absent pupils on the day of the assessments. Thus sample was slightly skewed, with more boys than girls in P7, and in S2 German. S2 French had slightly more girls. The gender balance of the sample is represented in Table 2d.

Table 2d: Gender balance of the sample

P7 FRENCH		P7 GERMAN	
M	F	M	F
54	52	59	51

S2 FRENCH		S2 GERMAN	
M	F	M	F
114	121	121	117

In addition any 'topping-up' procedures implemented aimed to take account of the need for the final sample to include the full range of ability, given that the main purpose of the pilot project was to evaluate the assessment instruments and procedures. Thus all ability levels were represented, although the P7 and S2 samples for both languages contained more pupils of middle ability, than high

---

[6] It should be noted, however, that in one of the sample primary schools for French the pupils had been learning the foreign language in P7 only, while in another two they had been learning since P5, and in one very small school since P4 (in a composite P4-P7 class). Also in the secondary sample for French two schools had adopted a system whereby the S2 cohort studied both French **and** another foreign language throughout S1 and S2, while in a further two schools a second foreign language had been studied alongside French in S2 only. Also, one secondary school participated in the project at a later stage than the others: at the start of the 1998-99 session, rather than at the end of the 1997-98 session. These factors may have implications for the pupil findings, analysed in Chapter 5 - 8.

[7] It should be noted that in three of the primary schools for German the pupils had experience of the foreign language at P7 only, and in one of those primary schools the pupils had studied both German **and** French. Also in the secondary sample for German one school had adopted a system whereby the S2 cohort studied both German **and** another foreign language throughout S1 and S2, while in a further three schools a second foreign language had been studied alongside German in S2 only. Again these factors may have implications for the pupil findings, analysed in Chapter 5 - 8.

or low, with the number of high ability pupils also outweighing the number of low ability pupils[8]. The ability level balance of the sample is represented in Table 2e.

Table 2e: Ability level balance of the sample

P7 FRENCH			P7 GERMAN		
H	**M**	**L**	**H**	**M**	**L**
41	**48**	**17**	**43**	**54**	**13**

S2 FRENCH			S2 GERMAN		
H	**M**	**L**	**H**	**M**	**L**
72	108	55	79	88	71

If we then combine gender and ability, we find that the samples for both languages at both stages contained more high ability girls and more low ability boys, although the differences were more marked at S2 than at P7 as can be seen from Table 2f.

---

[8] It should be noted that the allocation of sample pupils to ability groupings was carried out by the relevant class teacher, prior to the implementation of the pilot.

Figure 2f : Balance of the sample by gender and ability level

P7 FRENCH						P7 GERMAN					
H		M		L		H		M		L	
M	F	M	F	M	F	M	F	M	F	M	F
20	21	24	24	10	7	21	22	26	28	12	1

S2 FRENCH						S2 GERMAN					
H		M		L		H		M		L	
M	F	M	F	M	F	M	F	M	F	M	F
21	51	55	53	38	17	34	45	41	47	46	25

A total of 40 schools and 689 pupils participated in the pilot survey. See the summary in Table 2g for a simple breakdown of the sample. (The column headed E. PUPILS indicates the number and percentage of pupils who were also part of the English language survey.)

Table 2g: Summary of sample

STAGE	LANGUAGE	NO. SCHOOLS	NO. PUPILS	NO. E. PUPILS
P7	French	10	106 (54B, 52G)	92 (84%)
P7	German	10	110 (59B, 51G)	82 (75%)
S2	French	10	235 (114B, 121G)	222 (94%)
S2	German	10	238 (121B, 117G)	91 (38%)

SCHOOL CONTEXTUAL DATA

Questionnaires were circulated to all of the sample schools. The questionnaires dealt with a number of elements relevant to the pilot project: school size, languages taught, staffing, provision type, time allocation, class size, materials, skill areas assessed, topics covered, teacher views on issues relating to the situation of modern languages in Scottish schools. A 100% return rate was achieved, thus enabling the project team to compile a very detailed picture of the language learning experience of pupils in each of the sample schools. The sampling constraints outlined above militated against the inclusion of secondary schools and any of their associated primaries in this sample.

A copy of the school feedback questionnaire is provided in Appendix 1.

PRIMARY

School Size
French sample schools varied in size from 48 pupils to 420 pupils
German sample schools varied in size from 40 to 550 pupils.
The average school size was 246: 222 for French and 271 for German.

MLPS BACKGROUND INFORMATION

Languages taught in P6 & P7
French was taught in 14 of the 20 sample schools – the ten French sample schools and four of the German sample schools

German was taught in ten of the 20 sample schools – the ten German sample schools
Gaelic was taught in one of the French sample schools.
Neither Spanish nor Italian was taught in any of the schools.
No other languages were taught in any of the sample schools.

Length of time involved
The length of time that schools had been involved in the MLPS programme varied greatly.
In the **French schools** it varied from seven years to one year: One school had been involved since 1991, two since 1993, two since 1994, four since 1995 and one since 1997.
In the **German schools** it varied from nine years to one year: One school had been involved since 1989, one since 1990, one since 1992, two since 1994, one since 1995, two since 1996 and two since 1997.

Staffing
The number of primary teachers who had undergone the national MLPS training course and were still in post varied.
In the **French schools** it ranged from one teacher only in five schools, to two teachers in four schools to three teachers in one school only. In addition one trained teacher was no longer in post.
In the **German schools** it ranged from no teacher in one school, to one teacher in four schools, to two teachers in two schools, to three teachers and four teachers in one school only. In addition two trained teachers were no longer in post.
In terms of **language support** received since completion of training, six of the French schools and five of the German schools had benefited from it in various forms:
• refresher/in-service/twilight courses
• visits to the local secondary school
• visits from MLPS tutors.

Primary-Secondary liaison
Eight of the French schools and all ten German schools had established a co-operative link with their associated secondary. The link took various forms:

• cluster or individual planning meetings with associated secondary to discuss pace and content
• cluster meetings (primary only)
• visits from secondary ML staff
• agreed syllabus
• transfer of information (plans, link sheets outlining language areas covered, pupil files)
• joint projects with P7 pupils for secondary induction week
• agreed bridging lesson at the start of S1

MLPS PROVISION

MLPS class start
In a majority of the **French schools** the pupils had started their foreign language learning in P6 (6 schools), while two schools had started in P5 and a further very small school in P4 (in a composite P4-P7 class). In one school the pupils had been learning in P7 only.

Similarly in a majority of the **German schools** the pupils had been learning the foreign language since P6 (7 schools), while three schools started in P7 only. In one of the latter schools, the pupils had been learning French **and** German throughout P7.

Provision type
Most teaching in both P7 and P6 was carried out either by the class teacher or by a drop-in teacher. (A drop-in teacher is usually a primary teacher colleague from the same school who has undergone the national training program for MLPS.)

In six of the **French schools** the sample pupils had been taught throughout P7 by their own class teacher, while in the remaining four they had been taught by a drop-in teacher. In P6 they had been taught by their class teacher in only four of the schools, and by a drop-in teacher in five schools. The remaining school did not yet teach the foreign language in P6.

In two of the **German schools** the sample pupils had been taught throughout P7 by their own class teacher, while in five they had been taught by a drop-in teacher, and in one by a peripatetic secondary teacher. In the remaining two schools some of the pupils had been taught by the class teacher and some by a drop-in teacher. In P6 they had been taught by their class teacher in only two of the schools, and by a drop-in teacher in three schools. In one school some of the pupils had been taught by the class teacher and some by a drop-in teacher. In a further one school no information was provided. The remaining three schools did not yet teach the foreign language in P6.

Time allocation
The time allocated to the teaching and learning of the foreign language in P7 and P6 varied greatly: from as little as 20 minutes/one lesson per week to as much as 90 minutes/three lessons.
In **P6 French classes** the time allocation per week varied from 20 minutes or one lesson to 80 minutes or two lessons, with the average being 61 minutes.
In **P7 French classes** the time allocation per week varied from 20 minutes or one lesson to 90 minutes or two lessons, with the average being 63 minutes.
In **P6 German classes** the time allocation per week varied from 30 minutes or one lesson to 90 minutes or three lessons, with the average being 50 minutes.
In **P7 German classes** the time allocation per week varied from 30 minutes or one lesson to 90 minutes or three lessons, with the average being 56 minutes.
In one German school two lessons per week of German had been given throughout one half of the school year, followed by two lessons per week of French for the other half of the year.

TEACHING AND LEARNING

Skill areas covered
Listening and Speaking were the main language skills covered in all of the sample schools, while Reading had been covered in three French schools and two German schools, and Writing in three French schools and only one German school.

Teaching resources/materials used
The majority of sample schools used national training materials (seven French schools, eight German schools), or regional training materials (seven French schools, seven German schools) as the main source of teaching material. Published or commercial materials were used in a smaller number of schools (two French schools and five German schools). In addition one French school and two German schools used in-house materials.

TEACHER VIEWS

Relevant information influencing achievement of sample pupils
Teachers of both languages identified a number of factors that may have had a bearing on the achievement of the sample pupils in the pilot assessments.

In presenting this information for P7 and S2, we simply list the points that were made, since each has its own degree of interest. We do not attempt to convey a picture of the overall balance between positive and negative views.

A limited number were elements that teachers felt might have a **positive effect** on their pupils' achievement:
- The early start (P5) for pupils in a limited number of French schools
- The high level of qualification of staff (three language graduates) in one German school

However the majority were elements that teachers felt might have a **negative effect** on the pupils' achievement:
- Class teacher with very little background in the language
- Lack of continuity/breaks in teaching due to staff arrangements
- Problem of pupils with learning difficulties or new to the school
- Timing of pilot not good (so near the end of the school year)

Views on MLPS
Teachers were also asked to give their views on the current situation of MLPS in their school and nationally, identifying both the advantages and disadvantages involved. Several issues were raised.

Advantages:
- Pupil and teacher enthusiasm and enjoyment
- Children introduced to ML when they are more receptive
- Early start encourages fast progress - an earlier start preferred
- Encourages acquisition of knowledge about other cultures

Disadvantages:
- Uneven priority given to MLPS nationally
- Problem if no-one within school trained or willing to be trained
- Need for ML training to form part of primary initial or pre-service teacher education
- Lack of confidence on part of primary teachers - prefer drop-in secondary specialist
- Movement of trained staff problematic
- Timing and time-tabling problems with composite or multi-composite classes
- Need for in-service training
- ML teaching methodology (teacher-led) not the normal primary method
- Need for more preparation time
- Differentiation difficult
- Lack of time for liaison
- Pupil choice of language problematic at primary or secondary

SECONDARY

SCHOOL SIZE

French sample schools varied in size from 170 pupils to 1200 pupils
German sample schools varied in size from 600 to 1400 pupils.
The average school size was 850: 716 for French and 984 for German.

MODERN LANGUAGE BACKGROUND INFORMATION

MLPS experience
In sevem of the **French schools** the sample pupils came into S1 having learned French at primary. The pupils had started the language at P6 in five schools (although in one of these schools composite classes at primary meant that pupils had often started their foreign language learning at an even earlier stage), and at P7 only in two schools.

In three of the **German schools** the sample pupils came into S1 having learned a foreign language at primary: German or French in one school, German, French or no language in another school. No information was given by the third school. The pupils had started the language at P6 in one school, and at P7 or P6 in the remaining two schools.

It should be noted that one German school did not respond to the questions relating to MLPS.

Primary-Secondary liaison
A co-operative link had been established between the secondary school and its associated primary schools in five of the seven French schools receiving pupils with experience of French, and all three of the German schools receiving pupils with experience of German or French. In a further one French school and one German school the link was on the way to being established.

The link took various forms:
- Secondary/primary meetings
- Secondary visits to primaries
- Agreed common core syllabus/materials
- Agreed system of grading
- Transfer of information: full P7 report - pupil profiles, pupil work

Languages taught in S1 and S2
French was taught in each of the 20 secondary sample schools
German was taught in 14 of the sample schools –the ten German sample schools and four of the French sample schools
Italian was taught in two of the sample schools for French
Gaelic was taught in two of the sample schools – one for French and one for German
Spanish was not taught in any of the sample schools.
No other languages were taught in any of the sample schools.

MODERN LANGUAGE PROVISION

Provision type S1 and S2
In seven of the **French schools** French was the only language on offer to **S1 pupils**. In a further school, S1 pupils studied one of the two languages on offer. In the remaining two schools, S1 pupils studied more than one language.

In five of the **French schools S2 pupils** continued with French, started in S1. In a further two schools the S2 pupils continued with the two languages they had been learning since S1. In one school they continued with French and started another language. In one school they continued with French and got a taster in another language. In the final school S2 pupils either continued with French alone or started another language also.

In nine of the **German schools S1 pupils** studied one of the two languages on offer. In the remaining school S1 pupils studied more than one language.

In six of the **German schools S2 pupils** continued with German, started in S1. In a further school the S2 pupils continued with the two languages they had been learning since S1. In two schools they continued with German and got a taster in another language. In another school S2 pupils either continued with German alone or also got a taster in another language. In the final school S2 pupils continued with German and were offered an optional taster in another language.

Class organisation
In **S1 French classes** the ten schools used a system of mixed ability groupings, with one school time-tabling the classes together one period per week to allow for extraction of the top and bottom groups in order to undertake extension and reinforcement work.

In **S2 French classes** seven schools used a system of mixed ability, with one school time-tabling the classes together one period per week to allow for extraction of the top and bottom groups in order to undertake extension and reinforcement work. A further two schools used a system of setting, and one used a system of broad ability groupings.

In **S1 German classes** all ten schools used a system of mixed ability groupings.

In **S2 German classes** seven schools used a system of mixed ability groupings, two used a system of setting, and one used a system of broad ability groupings.

Time allocation
In **S1 French classes** the time allocation per week varied from 120 minutes or three lessons to 200 minutes or five lessons, with the average being 165 minutes.

In two French schools S1 pupils also received another foreign language: 110 minutes of Gaelic in one school alongside 165 minutes of French, and 60 minutes of Italian alongside 120 minutes of French in the other.

In **S2 French classes** the time allocation per week varied from 105 minutes or two lessons to 200 minutes or five lessons, with the average being 154 minutes.

In four French schools S2 pupils also received another foreign language: 120 minutes of German or Italian alongside 200 minutes of French, 105 minutes of German alongside 105 minutes of French, 165 minutes of Gaelic alongside 110 minutes of French or 60 minutes of Italian alongside 120 minutes of French.

In **S1 German classes** the time allocation per week varied from 120 minutes or two lessons to 200 minutes or five lessons, with the average being 165 minutes.
In one German school S1 pupils also received another foreign language: 159 minutes of either French or Gaelic alongside 159 minutes of German.

In **S2 German classes** the time allocation per week varied from 120 minutes or two lessons to 212 minutes or five lessons, with the average being 166 minutes.

In four German schools S2 pupils also received another foreign language: 80 minutes of French alongside 160 minutes of German, 40 minutes of French alongside 120 minutes of German, 53 minutes of French alongside 159 or 212 minutes of German, or 159 minutes of French or Gaelic alongside 159 minutes of German.

Class size
In both **S1 and S2 French classes,** the class size varied from 16 to 32, with 26 being the average.
In both **S1 and S2 German classes**, the class size varied from 24 to 33, with 27 being the average.

TEACHING AND LEARNING

Commercial teaching materials
All 20 secondary sample schools used a commercially published course as the main source for teaching materials in S1 and S2. However in both S1 and S2 classes only three of the French schools and four of the German schools were able to provide pupils with a textbook to take home.

For **French** the most common course was *Arc-en-ciel*, used in five of the schools. Other courses used were *Avantage*, *Route Nationale*, *Tricolore* and *Spirale* (two schools).
For **German** the most common course was *Zickzack*, used in five of the schools. Other courses used were *Auf Deutsch*, *Deutsch Heute* (two schools), *Einfach Toll* and *Gute Reise*.
The point reached in these courses by the end of S2 varied to a large extent, even when the same course was used.

Supplementary teaching materials
Seven of the French schools and five of the German schools regularly supplemented the commercial course used in S1 and S2. The supplementary materials used were of various types:
- In-house materials: worksheets, extra reading, grammar, writing, differentiated materials
- Materials from other commercial courses
- Puzzles, games
- Reading materials
- Audio materials
- TV programmes or videos
- ICT materials

Topics
The topics covered and timing of that coverage varied greatly, as can be seen in Table 2h below:

Table 2h: Topics covered and timing

TOPIC	FRENCH				GERMAN			
	S1	S2	S1/S2	NONE	S1	S2	S1/S2	NONE
Alphabet	9			1	10			
Classroom language	9		1		8		2	
Classroom objects	10				10			
Clothes	2	6		2	1	5		4
Colours	6	4			5	1	3	1
Daily routine		8		2	3	3	2	2
Dates	10				8	1	1	
Directions	4	4	1	1	3	4	2	1
Family	10				7		2	1
Food/drink	4	5	1		4	4	2	
Hobbies	2	2	6		5	3	1	1
House/home	4	6			3	4	1	2
Jobs/place of work	4	2	1	3		1		9
Nationality	7	1	1	1	8	1		1
Numbers	8		2		9		1	
Parts of body	2	6		2		5		5
Personal	3		4	3	4	1	3	2

language								
Pets	10				9		1	
Physical description	1	5	1	3	2	5		3
Places in town	4	5	1		1	7	2	
School subjects	7	1	1	1	8	1		1
Time	8	2			8	1	1	
Weather	3	6		1	2	4		4
Other: Personality						1		
Making plans					1	1		
Holidays						1		
Shopping/ money		1			1			
Travel/ transport		1				1		

Assessment

The skill areas of **Listening, Reading** and **Speaking** had been regularly assessed throughout S1 and S2 in all 20 of the sample schools. However **Writing** had been assessed in only seven French and seven German schools in S1, and eight French and eight German schools in S2.

Other areas were also **assessed** to a greater or lesser degree:

- Vocabulary (eight French schools, eight German schools)
- Grammar (five French schools, three German schools)
- Extended writing (one German school)

The **source of the assessment materials** varied:

- Commercial coursebook (one French schools, one German school)
- In-house (two French schools, four German schools)
- Mixture of both of the above (seven French schools, five German schools)

TEACHER VIEWS

Relevant information influencing achievement of sample pupils
Teachers of both languages identified a number of factors that may have had a bearing on the achievement of the sample pupils in the pilot assessments.

A limited number were elements that teachers felt might have a **positive effect** on their pupils' achievement:

- Pupils had just completed attainment tests covering all S1 and S2 material

However the majority were elements that teachers felt might have a **negative effect** on the pupils' achievement:

- The courses used sometimes did not cover all of the topics included in the pilot assessments
- The timing of the pilot visits occasionally clashed with the introduction of the new school time-table, resulting in problems of administration
- The timing late in the school session also resulted in a negative response from some pupils
- Some parents and pupils re-acted negatively to the pilot project, given the need for the same pupils to be tested in both English and the foreign language

Views on MLPS

Secondary teachers were asked to give their views on MLPS, identifying both the advantages and disadvantages involved. Several issues were raised.

Advantages:
- Early start
- Increased pupil confidence
- Increased pupil enthusiasm
- Increased pupil receptiveness
- One less new subject in S1
- Faster pace in S1

Disadvantages:
- Poor overall planning leading to inconsistency
- Need for prior and on-going consultation between associated primaries and secondary
- Training problems: level reached not high, limited number of languages represented
- Staffing problems in primary sector: high turnover of trained staff, resulting in uneven learning experience for some pupils
- Preference for visiting secondary model
- Adverse effect on diversification
- Adverse effect on composition of S1 classes
- Need for national syllabus
- Differing primary experiences of pupils leading to problems of continuity and progression
- More 'fun' topics covered in primary
- Decreased pupil enthusiasm
- Decreased pupil confidence

Views on ML teaching S1-S2

Teachers were also asked to give their views on the current situation of ML teaching of S1-S2 in their school and nationally, identifying both the advantages and disadvantages involved. Several issues were raised:

Advantages:
- Commitment of teachers to promote general language awareness
- Ability to build on primary experience
- Broad-banding in S1/S2
- Blocking of timetable to allow for extraction of more or less able
- Taster course in S2 to improve pupil motivation

Disadvantages:
- Uncertain future of diversification – problem of continuity from primary
- Uncertainty regarding how best to teach SEN pupils
- Need for structured syllabus, revision of 5-14 Guidelines
- Shortage of money for resources
- Large class sizes
- Mixed-ability teaching problematic – setting preferred
- Time-tabling resulting in no possibility of setting/banding
- Lack of time for primary-secondary liaison
- Lack of time per week – need to bring ML into line with other core subjects
- Need to improve overall profile of modern languages

PUPIL CONTEXTUAL DATA

Questionnaires were circulated to each of the sample pupils after each of the tests. Thus, P7 pupils completed one questionnaire only, while S2 pupils either completed two questionnaires (after the Reading/Writing Test and Listening Test), or three questionnaires if they had also been selected to sit the Speaking Test. The main purpose of these post-test surveys was to gather data regarding pupil views on the assessment procedures and the prototype assessments themselves. This very specific information will be analysed in more detail in Chapters 3 and 4 of the present report[9]. However a secondary aim was to collect data of a more general contextual nature, regarding the number and gender of participating pupils, their linguistic background and some of their basic attitudes towards foreign language learning. Copies of the S2 and P7 feedback questionnaire are provided in Appendix 2.

PRIMARY

Number of pupils
106 pupils took part in the French tests and all completed a post-test feedback sheet. The German sample was marginally bigger with 110 pupils of which 108 completed the post-test feedback sheet.

Gender of pupils
The gender balance of the two language samples was fairly even; there were 53 boys (50%) and 51 girls (48%) in the French sample (2 pupils did not respond to this item) and 55 boys (51%) and 49 girls (45%) in the German sample (6 pupils did not respond to this item).

Start of ML learning
Over half of the French P7 sample (55%) had started learning French in P6 and nearly a quarter in P7 (24%). Some had started younger with 17% indicating at P4 or P5 start and 2% at the earlier primary stages and 2% pre-primary.

Just under half of the German P7 sample (47%) had started learning German at the P6 stage and nearly a third had started in P7 (31%). The remaining 22% had begun learning German at the P4 or P5 stage.

VIEWS ON THE LEVEL OF DIFFICULTY OF THE FL
Just over half of the French P7 pupils (51%) said they found the language of average difficulty at school, and just under a third (31%) found French easy and 7% said it was very easy. For 10% of the sample, French was considered difficult.

Over half the German P7 pupils (56%) said they found German to be of average difficulty at school, a third found it easy and 6% said the language was very easy. Only 5% considered the language difficult.

language to be studied at Secondary
Of the French P7 pupil sample, 82% said they would be continuing with French when they transferred to secondary school. A further 11% said they would be taking German, 2% would take Spanish and the remainder other languages.

A larger proportion of the German P7 sample were continuing with their primary language in S1 (92%) with only 7% saying they would be learning French.

---

[9] Chapter 3 provides an account of the development, implementation and evaluation of the primary assessments while Chapter 4 provides a similar account of the S2 assessments.

Visits to the foreign country
Over half the P7 French sample (55%) had never visited France or a French-speaking country, but nearly a quarter had (23%) and a similar proportion had visited France on several occasions.
By contrast, three-quarters of the German sample (75%) had never visited Germany or a German-speaking country; 16% had made one such visit and 10% had made several visits to Germany.

SECONDARY

Number of pupils
The S2 French sample was made up of 235 pupils and the German sample of 238 pupils.

Gender of pupils
There were 114 boys (49%) and 121 girls (51%) in the French S2 sample and 120 boys (50%) and 118 girls (50%) in the S2 German sample.

Start of ML learning
There was a considerable difference in the French and German samples with regard to when they had started learning the foreign language concerned. Over two-thirds of the German sample (67%) had not begun learning German until S1, with only 19% beginning at the P6 stage, 9% at the P7 stage and 2% at the P5 stage. By contrast, only 22% of the French S2 sample had started learning that language in S1. Over a third had begun in P6 (35%) and a quarter at the P7 stage (25%) and the rest had started even earlier.

Views on the level of difficulty of the FL
Just over half of the French and German S2 cohorts found the foreign language of average difficulty at school (52% and 51% respectively). Only 4% of the French and 3% of the German samples found their foreign language very easy, although more of the German pupils found the subject easier than their French counterparts (26% compared with 19%). Nearly a quarter of the French sample (24%) found the language difficult or very difficult compared with 18% of the German sample.

Views on the level of interest of the FL
About a third of the pupils studying French and German said the language was of average interest as a subject at school (33% and 32% respectively). Slightly more of the German sample found the language interesting or very interesting than their French counterparts (36% compared with 30%). Over a third of the S2 French cohort said they found the language boring or very boring compared with 19% in the German S2 sample.

Self-evaluation of competence in the FL
The pupils taking German tended to rate their competence in the language slightly higher than those taking French. Only 3% of each cohort felt they were very good, but 32% of the French sample thought they were average compared with 40% of the German sample. Nearly a third of the French sample (32%) said they were not good or not at all good at French compared with 19% of the German sample.

Visits to the foreign country
Over a third of the S2 French sample had visited France or another French-speaking country, with 21% having visited once, 10% twice and 2% several times. Just under a quarter (24%) of the S2 German sample had visited Germany or another German-speaking country, 17% having visited once, 4% twice and 3% several times. Some 59% of the French and 70% of the German samples had never visited a country where the foreign language they were learning is spoken.

USE OF THE FL OUTSIDE OF THE CLASSROOM

Listening
Over three-quarters of the French S2 sample (76%) and nearly two-thirds of the German S2 sample (64%) said they never listened to the foreign language outside of school. Some 19% of the German sample and 13% of the French sample said they did so very occasionally, but only 3% of each sample listened to the foreign language once or twice a month or once or twice a week. A tiny minority of 1% in each sample said they listened to the foreign language nearly every day.

Speaking
The picture was somewhat more encouraging when it came to speaking the foreign language, for although 60% of the French sample and 48% of the German sample said they never spoke the foreign language outside of school, 18% and 28% respectively said that they did so very occasionally and 3% did so once or twice a month or once or twice a week. Some 4% of the French and 5% of the German sample said they spoke the foreign language almost every day.

About a fifth of each sample (19% for French and 21% for German) said they spoke the foreign language with members of their family. They also spoke with friends (8% of the sample for French and 14% for German).

Reading
Reading the foreign language was less frequent an occurrence and 70% of the French and 60% of the German samples never did so outside of school. A fifth of the German sample (20%) said they read German very occasionally compared with 14% of the French sample. Reading the foreign language once or twice a month or once or twice a week was done by only about 10% of each sample. A small minority of the German sample (2%) said that they read German nearly every day.

Writing
The incidence of writing in the foreign language was proportionally very similar in both samples to that of reading in French or German with 72% of the French and 59% of the German samples never doing so outside of school (not including homework). Writing very occasionally in the foreign language was done by 13% of the French and 18% of the German samples, with marginally more writing in German once or twice a month or week (9% and 5% respectively) than in French (6% and 2% respectively). A very small minority in both samples said they wrote in the foreign language almost every day (1% for French and 2% for German).

OVERALL

The contextual data gathered for the present study suggested strongly that a national policy for MLPS was being implemented on the ground in highly diverse ways. The relative lack of commonality across schools allied to the lack of national guidelines would undoubtedly pose problems for the researchers in the development of the pilot assessments.

Our contextual data shows that provision for modern languages at primary school varies enormously and therefore pupils are likely to have had very different experiences. Variation in performance at primary level could better be explained by the very different circumstances in which children are learning another language than by other variables such as gender or differences in ability. Analysis of AAP results by school, however, lies outside the remit of AAP in general, as these data are used to assess the national picture performance in a given subject.

CHAPTER 3
THE PRIMARY 7 ASSESSMENTS

SPECIFICATION OF THE DOMAIN

It was extremely important to begin by developing a clear idea of what could and should be assessed. In this way the validity of the assessment instruments would be strengthened. The research team was aware of the great diversity of experience at the P6/P7 stages depending on a number of factors: how long the school had been involved in MLPS; whether the teacher was the class teacher or a drop-in teacher; the amount of time allocated to the foreign language; the extent of continuity of provision; the availability of regional or local authority support materials. The research team had decided against trying to construct assessments which would test P7 pupils' knowledge of a common corpus of language, but instead construct tasks which would be sufficiently flexible for children to bring to them the foreign language they knew. It was decided to liaise with each of the 20 primary schools independently by telephone to gather information regarding specific topic areas covered by those particular schools during P6 and P7, so that assessors could arrive prepared to tailor the assessment tasks to individual contexts.

TOPICS COVERED
The topics covered and timing of that coverage varied greatly, as can be seen in Table 3a below:

Table 3a: Topics covered and timing

TOPIC	FRENCH SCHOOLS				GERMAN SCHOOLS			
	P6	P7	P6/P7	NONE	P6	P7	P6/P7	NONE
Alphabet	5	3	2		7	3		
Animals	2	4	2	2	3	5		2
Classroom language	3	2	5		7	2	1	
Classroom objects	3	1	3	3	6	3		1
Clothes		7	1	2	2	5		3
Colours	5	2	3		8	2		
Dates	4	4		2	6	4		
Family	1	6	2	1	2	7	1	
Food/drink		5		5	1	5		4
Hobbies		5	1	4	1	6		3
House/home		4		6	2	5		3
Nationality		6		4		3		7
Numbers	5	2	3		7	2	1	
Parts of body	3	4	2	1	5	4		1
Personal language	5	2	3		7	3		
Physical description		6		4	3	3		4
Places in town		1		9	1	2		7
Time		4		6	1	5		4
Weather	3	2	3	2	5	5		
Other: Greetings Easter Christmas			1		1 1	1		

TEACHING RESOURCES/MATERIALS USED

The majority of sample schools used national training materials (seven French schools, eight German schools), or regional training materials (seven French schools, seven German schools) as the main source of teaching material. Published or commercial materials were used in a smaller number of schools (two French schools and five German schools). In addition one French school and two German schools used in-house materials.

STUDY OF RELEVANT GUIDELINES/OFFICIAL DOCUMENTS

The MLPS Advice to Schools, MLPS Topic Frameworks and the 5-14 Guidelines for English and Modern Languages were also analysed. Again this contributed to the identification of a common-core of language for P6 to S2 (topic areas, grammatical structures and communicative functions) that would form the basis on which to construct the pilot assessments. In addition the analysis of the 5-14 Guidelines for Modern Languages enabled the project team to develop a clear idea of the types of task pupils could reasonably be expected to undertake at the two stages involved.

P7 TEST DEVELOPMENT

The feasibility of undertaking tests at the P7 stage was a major consideration, given that there was no substantial body of prior work on which to draw. Instead of opting for a piloting of the first draft of the tests with schools outside the sample of 20 schools, a decision was taken to review the tests and procedures after the first four implementations and then review the content and procedures as necessary.

FIRST DRAFT

A provisional framework was decided upon, which was to be based mainly on listening and speaking but with possible reading and writing elements included.

Task 1 Speaking - social chat in FL with native speaker assessor (NS)
Task 2 Listening - pupils respond to verbal instructions, as a pair and individually
Task 3 Listening - vocabulary recognition; identifying items in pictures
Task 4 Listening - understanding a simple narrative - ticking the box which matches the appropriate part of the narrative
Task 5 Listening - understanding simple dialogues and their location. Pupils tick box to indicate where conversation is taking place
Task 6 Speaking - pupils describe animals from a set of pictures (name, age, colour, etc.)
Task 7 Reading - pupils match a visual with a phrase in the FL
Task 8 Listening/Writing/Reading - NS spells out four numbers in the FL. Pupils must write out the word as dictated, then the actual number
Task 9 Writing - pupils write any five words they know in the FL
Task 10 Reading/ MLA pupils are given a card with sentences in FL. They are asked about their knowledge of the FL by the non-native speaker assessor (NNS) (adjectives, verbs, nouns, etc.)

FINDINGS FROM INITIAL TESTS

After the first four administrations of the P7 assessments (two schools for French and two for German involving the whole core team) a meeting was held to review how the assessments were performing in the light of actual experience. Students were not registering any undue anxiety about taking the tests and some very interesting data were emerging, but nonetheless some problems had been identified, namely :

- The process was taking too much time for each pair of pupils.
- Reading Task 7 (matching pictures with phrases/sentences) did not seem to be sufficiently discriminating.
- Pupils generally seemed to be producing very short 'minimal' responses.
- Listening task 4 (understanding a simple narrative, with visual support) seemed to be testing vocabulary rather than more inferential listening comprehension. If the pupils knew a particular form (e.g. *Es ist sonnig*), they could recognise it, but were not able to recognise it in another form (*Die Sonne scheint*).

Generally, the pupils seemed tied to the precise ways in which their teacher had operated and to the precise forms of language they had been taught. In many cases they also seemed to be very tied to what they had recently been learning. If a test did not seem to cover this, some of the pupils found it difficult to cope.

It was agreed after discussion that the following modifications should be made:

- All 12 pupils would be tested only if it proved possible in a particular school for the team to arrive earlier, before the morning interval. If this did not prove possible, then the number would be reduced to ten or eight as time permitted. A decision on this would be taken in advance, once the time of arrival at the school was confirmed, since it would not be appropriate for two-four pupils to spend the entire day expecting to be tested and then to find there was insufficient time after all.

- A new task would be introduced at the start, consisting of '*Can you tell us what you have been learning in the past few weeks?*' This would allow pupils to summon up those topics and aspects of language that were reasonably fresh in their minds. The NS would then try to build on what the pupils had said, encouraging them to produce words, phrases and even strings of phrases in the FL where appropriate. This activity would also allow pupils to engage in some form of social learning, i.e. what one pupil remembered might enable the other one to recall something else.

- Speaking Task 1 (social chat with assessor) would be reduced in length to the more obvious and basic personal questions, though pupils would still be given the opportunity to ask questions as well as to answer them. Additional social chat questions would only be put to those pupils who seemed to be doing well and who might be further challenged. The NS would use their judgement as to whether to ask these additional questions as part of this task or to return to them later during the test.

- The Listening Task 4 (understanding a simple narrative) would no longer be based on visuals (which encouraged pupils to focus on specific hit or miss vocabulary) but on simple narratives read by the NS. The NS would tell story X to the pair and ask Pupil A to say in his/her own words (in English) what he/she thought he/she had understood. Pupil B would then be given an opportunity to 'fill in'. The NS would then tell story Y, this time beginning with Pupil B, with Pupil A being given the opportunity to 'fill in'. The activity would be done orally. The NS would use eye-to-eye contact and body language as appropriate. This test would now be geared to gist extraction and more global comprehension of the discourse as a whole. It would also be more focused on comprehension of verbal meanings.

- The final task of Reading/Metalinguistic awareness (MLA) would now be used as a two-phase task. First, the sentences would be used for reading comprehension. Then, the MLA element would take place. The Reading Comprehension part would be done orally.

- Certain tasks would be treated as optional, to be done only if time permitted. These would be:
    Task 3 - Listening (vocabulary recognition: identifying items on pictures)
    Task 5 - Listening (understanding a simple dialogue).

- Certain tasks would no longer feature. These would be:
    Task 2 - Listening (pupils responding to verbal instructions)
    Task 7 - Reading (Matching pictures with phrases/sentences)
    Task 8 - Listening/writing/reading (spelling numbers - dictation)
    Task 9 - Writing (Writing any five words).

- Student feedback questionnaire would be simplified and a simplified scoring system for the real-time coding of student performance would be devised.

The basic outline of each test is described below. A more detailed set of outlines is to be found in Appendix 3.

FINAL P7 TEST OUTLINE

P7 Test   Tasks 1-6: Native Speaker Assessor
                      Task 7: Non-native Speaker Assessor

Task 1: Speaking	Vocabulary recall via discussion of recently studied topics
Task 2: Speaking	Spontaneous question-and-answer session with the assessor
Task 3: Speaking	Description of a colour visual
Task 4: Listening	Understanding the subject and message of a short narrative
Task 5: Listening	Understanding short dialogues
Task 6: Listening	Vocabulary recognition
Task 7: Reading, Translation, Metalinguistic Discussion	Reading aloud three short sentences, understanding them and discussing their linguistic content in English

P7 TEST PROCEDURES

CONTACT WITH SCHOOLS

As previously mentioned, schools had initially been contacted by Scottish CILT and asked to provide them with details of the topics which had been covered by their pupils in modern languages.

Further contact by telephone was also made by most of the assessors in order to confirm suitable times to visit, as well as to establish personal contact with the head teacher. Schools were also asked to choose which pupils would take part in the pilot tests (wherever possible from within the AAP sample for English), and to indicate each pupil's ability level.

EVALUATIONOF THE P7 TESTS

ROLE OF NS/NNS IN ASSESSING PUPILS

The NS carried out most of the tests. The NNS assessed Task 7 (Reading/Metalinguistic Awareness), and took part in the dialogues (Task 5).

Assessment sheets were developed in the course of the pilot phase. While the NS was carrying out the tests with the pupils, the role of the NNS was to take notes on what the pupils were saying, recording memorable phrases and jotting down observations. This was reversed for Task 7, where the NNS conducted the test and the NS took notes. At the end of the tests an overall grade was awarded for each pupil, as agreed by both NS and NNS.

The tests were evaluated by three different groups, all of which had been involved in the primary assessments and associated procedures. These were:

- The native and non-native speaker assessors
- The P7 pupils who had taken the tests
- The head teachers in the primary schools taking part

Each group was given a specific feedback questionnaire to complete. Copies of the school and pupil feedback questionnaires can be found in Appendices 1 and 2.

ASSESSOR FEEDBACK ON THE P7 TESTS

The assessors provided very detailed and rich feedback on the tests and associated procedures. In Table 3b below we report a summary of their views on the implementation and logistics of the tests. For their detailed comments on each individual test and how they felt the speaking tests and scoring systems worked, please see Appendix 4.

Table 3b: Implementation and Logistics

	COMMENTS
What worked well	*Several testers commented that the tests worked well in general terms and that the format was generally well-understood.* *Where the teacher had devised pairings, this had been successful. Pairs seemed to work better than trios.* *One tester commented: Original version of the tests did not work well. Revised version much better in terms both of timing and content.*
What did not work well	*Individual testers experienced particular problems in particular schools. One commented: None of the tasks worked well in School X!* *Another experienced difficulties with accommodation: Most things to do with pupils' performances their attitudes and the schools' co-operation worked well. However the availability of suitable accommodation for uninterrupted interviews was a problem in both schools. One had **hired** the nearby village hall (excellent room for interview). In the other we were in a tiled, tiny, medical room which made recording quality very poor and organisation of papers and grids a nightmare.* *More general comments related to the demanding and time-consuming nature of the tests, particularly if administered at the end of the school day when children are less receptive and more tired. One tester noted general failings among this group of pupils: Some personal information questions were not handled confidently if the **form** in which they were put departed from the **familiar** formulae (or even sequence). Third person questions were not familiar. Certain elements of the reading (some verbs, for example) were neither known, nor guessed without much prompting.*
Suggested improvements	*Testers suggested either more time to administer the tests, or else cutting some of the items: Testing 6 pairs of children took a whole day, no matter how early we started. And it was difficult to shorten it as I found that the 'best able pair' had to be 'stretched' in Tasks 1-2-3 and the 'less able ones' had to be given Tasks 5-6 as a way to finish on a good performance.* *One tester suggested: Could the pupil questionnaire be administered to the whole group of sample pupils at the end of the day – so saving time? Another suggested an additional test was needed: New task?? A reading task, similar to Task 4 listening.*

PRIMARY SCHOOL FEEDBACK

At the end of the pilot phase all the schools which took part were asked to fill in a questionnaire. The contextual data gathered from this survey have already been analysed in detail in Chapter 2. A very limited number of responses related specifically to the implementation of the pilot and highlighted aspects deemed negative by the schools concerned:

- The timing of the pilot, so close to the end of the school session, was criticised by a number of schools in both the primary and secondary sectors.

P7 PUPIL FEEDBACK

A questionnaire was given to each pupil at the end of each test in P7. The findings from these surveys have already been analysed in detail in Chapter 2 of the present report. A very limited number of responses related specifically to the implementation of the pilot:

ANXIETY CAUSED BY THE TEST
Pupils were asked how anxious or relaxed they were before and after the tests had taken place. The following table summarises their responses :

Table 3c: Before the test started

	very anxious	anxious	average	relaxed	very relaxed
French	13%	34%	26%	22%	5%
German	7%	45%	24%	21%	3%

Table 3d: After the test

	very anxious	anxious	average	relaxed	very relaxed
French	0%	2%	13%	56%	29%
German	2%	4%	15%	51%	29%

From Table 3d above it can be deduced that levels of anxiety were slightly higher among the German P7 pupils than among their French counterparts (52% compared with 47%), but that after the tests only a small minority remained anxious with 85% of the French sample and 80% of the German sample reporting that they felt relaxed or very relaxed.

DIFFICULTY OF THE TEST
Pupils were also asked how easy or difficult they felt the tests had been. Table 3e below summarises their responses to this question :

Table 3e: How easy was the test?

	very easy	easy	average	difficult	very difficult
French	2%	13%	79%	6%	0%
German	2%	15%	72%	11%	0%

It was encouraging to note that none of the P7 pupils felt that the test had been very difficult, although a minority in each sample had found the tests difficult. The number finding the German tests difficult was almost twice as high as that for French and may reflect the fact that a slightly larger proportion had only started learning the foreign language in P7 rather than P6 (a 7% difference).

It was interesting to note that a big majority in each sample felt that the tests had been of average difficulty, given that it was very likely the first time that they had been tested in the foreign language under formal conditions.

CHAPTER 4
THE SECONDARY 2 ASSESSMENTS

SPECIFICATION OF THE DOMAIN

From the pre-pilot work with the secondary teacher group it was clear that modern languages were being delivered at the S1/S2 stages using a variety of published materials and class organisation arrangements and differing time allocations. A common language syllabus and common point reached by the end of S2 could therefore not be assumed. However, in order to construct the listening and reading tests in particular, a common body of vocabulary, structures and language functions had to be established. This was achieved via a variety of methods outlined below.

TELEPHONE SURVEY
A total of 34 secondary schools participated in a telephone survey, the main aim of which was to gather information on the modern language provision and teaching materials used in S1 and S2 French and German classes. Thirty one local authorities and one school from the independent sector were represented.

Information was gathered on two main areas:
- Language provision - languages taught, diversification, systems of language provision, classroom organisation.
- Teaching materials – commercial courses used, point reached by the end of S2; use of and sources for supplementary material.

Although the information gathered in these two areas was far from homogenous, it did provide the project team with a clearer picture of the likely pattern of modern language learning experience of S2 pupils. A summary of the main information gathered can be found below:

LANGUAGE(S) TAUGHT	French	34 schools
	German	24 schools
	Italian	1 school
	Spanish	2 schools
	Gaelic	1 school
Language provision	No diversification (French)	7 schools
	French/German (split)	15 schools
	French/German (both)	8 schools
	French/German/Spanish (all study 2 of 3)	1 school
	French/Italian (both)	1 school
	French/Spanish (split)	1 school
	French/Gaelic (both)	1 school

Coursebook – French S1	*Arc-en-ciel 1*	9 schools
	Avantage 1	6 schools
	Encore Tricolore 1	3 schools
	Pyramide 1	1 school
	Route Nationale 1	6 schools
	Spirale 1	4 schools
	Tricolore 1A	1 school
	Tricolore – Encore Tricolore	1 school
	Departmental booklets	3 schools
Coursebook – French S2	*Arc-en-ciel 1*	3 schools
	Arc-en-ciel 1-2	5 schools
	Arc-en-ciel 2	1 school
	Avantage 1	3 schools
	Avantage 1-2	3 schools
	Encore Tricolore 1	1 school
	Encore Tricolore 1-2	1 school
	Encore Tricolore 2	1 school
	Pyramide 2	1 school
	Route Nationale 1-2	4 schools
	Route Nationale 2	2 schools
	Spirale 1-2	2 schools
	Spirale 2	3 schools
	Tricolore 1B	1 school
	Departmental booklets	3 schools
Coursebook – German S1	*Auf Deutsch 1*	4 schools
	Deutsche Heute 1	1 school
	Gute Reise 1	1 school
	Zickzack 1	3 schools
	Zickzack 1A	4 schools
	Zickzack Neu 1	4 schools
	Departmental booklets	3 schools
Coursebook – German S2	*Auf Deutsch 1*	1 school
	Auf Deutsch 1-2	3 schools
	Deutsche Heute 1	1 school
	Deutsche Heute 1 (S2 start)	1 school
	Gute Reise 1-2	1 school
	Lernpunkt Deutch 1 (S2 start)	2 schools
	Zickzack 1	3 schools
	Zickzack 1A-1B	1 school
	Zickzack 1B	3 schools
	Zickzack 1-2	1 school
	Zickzack Neu 1 (S2 start)	1 school
	Zickzack Neu 1-2	1 school
	Zickzack Neu 2	2 schools
	Departmental booklets	3 schools

Supplementary material – French	Reading	*À la carte, Autolire; Bibliobus; Carte blanche; C'est facile; Escalade; Everyday French; Lire d'avantage; Satellite; Vu et lu.*
	Audio	*Au secours; Bien entendu; Eurolab; Eurolab Junior; J'aime écouter; Steps to listening.*
	Video	*Arc-en-ciel; Avantage; Carousel; Channel Hopping; Clémentine; Global; Le Club; Ici Paris; Jeunes francophones; Quinze minutes; Quinze minutes +; Route Nationale; See you, see me, see France; Spirale; Vidéothèque.*
Supplementary material - German	Reading Audio Video	*Lesekiste; Mücke Hör zu; Super. Auf Deutsch; Global; Hallo aus Berlin; Lernexpress; Partner; Projekt Deutsch; Willkommen; Zickzack Neu.*

ANALYSIS OF MOST COMMONLY USED COURSE BOOKS

The telephone survey was then followed by the analysis of the coursebooks most commonly used in S1 and S2 French classes (*Arc-en-ciel, Avantage, Encore Tricolore, Escalier, Route Nationale, Spirale*) and German classes (*Auf Deutsch, Deutsch Heute, Zickzack, Zickzack Neu*). It was consequently possible to delineate the likely common core content for S1 and S2 French and German classes: topic areas, grammar and structures, communicative functions. These are as follows:

Topic	Grammar	Function
All coursebooks: Age	Adjectives - agreement and position	Accepting/refusing Agreeing/disagreeing
Animals/Pets	Alphabet	Asking/saying names
Birthdays Classroom language/objects	Articles - definite/indefinite Gender	Asking/saying what something is
Countries	Imperative	Choosing
Currency	Infinitives	Defining location/position of
Dates - days of the week/ months of the year	Interrogatives	person/object
Directions	Negatives	Describing
Family	Partitives	Expressing likes/dislikes
Food and drink	Plurals	Giving instructions/directions
Gifts/presents	Possessives	Greetings
Greetings/introductions	Prepositions - most common	Introductions Opinions
Leisure activities/hobbies	Pronouns - subject	Preferences
(incl. cinema, music, sport,	Pronouns - object	Questions - asking/answering
television)	Tense - present; also some	Statements
	introduction to perfect tense;	Talking about feelings
Numbers	limited conditional; and	Talking about present

Parts of the day	future using 'to go' + infinitive	activities
Self		
Shopping/prices	Verbs - regular (mainly -er	
Time	group) and some common	
Town - places/signs	irregular, some reflexives,	
Towns/cities	some modal, some	
Weather	impersonal	
Some coursebooks:	Word order	
Alphabet		
Clothes		
Colours		
Daily routine		
Festivals		
Furniture		
Home/housing/rooms in the house		
Methods of transport		
Nationalities		
Parts of the body/illness		
Restaurants		
School - subjects/timetable		
Seasons		
Travel/holidays		
Work/jobs/workplaces		

STUDY OF RELEVANT GUIDELINES/OFFICIAL DOCUMENTS

The MLPS Advice to Schools, MLPS Topic Frameworks and the 5-14 Guidelines for English and Modern Languages were also analysed. Again this contributed to the identification of a common-core of language for P6 to S2 (topic areas, grammatical structures and communicative functions) that would form the basis on which to construct the pilot assessments. In addition the analysis of the 5-14 Guidelines for Modern Languages enabled the project team to develop a clear idea of the types of task pupils could reasonably be expected to undertake at the two stages involved.

S2 TEST DEVELOPMENT

In the initial stages of the S2 test development, different formats were considered, but ultimately not taken up. These included :

- listening and speaking to be assessed together
- listening to be assessed along with reading and writing
- listening to be assessed via live input or video input
- pupils to be assessed in groups or pairs

Had listening and speaking been linked as in the primary tests, then only half of the S2 sample would have been tested in this skill whereas all would have done writing. It was felt that, given the balance of emphasis on the four language skills in early secondary, this would not have been appropriate.

Had the speaking test been conducted in the manner envisaged in primary, then this would not have allowed for any prepared speaking tasks, and all the speaking elements would have been done more or less spontaneously. It was felt that this did not reflect normal classroom activity and assessment and

might not allow students to perform as well as they might on assessment criteria such as fluency and accuracy.

The input for the listening test was a specially recorded audio-tape using a range of young men and women who were native speakers of French and German. The use of an audio-tape was agreed upon in preference to live input from the native speaker assessor in order to standardise the test across the schools. The use of video for this purpose was also considered, but time and resources did not make this a viable option. It was also felt that students at the S2 stage were used to audio-taped listening comprehension tasks and that listening comprehension via video was perhaps something which they might not be familiar with in the foreign language.

In the final test format, it was agreed that S2 students would be assessed individually in all 4 skill areas with the full sample (24 students) in each school taking a reading/writing test and a listening test, and half of that sample would do the speaking test. A non-assessed bridging activity would be undertaken as a form of preparation for the speaking task during which the native speaker and non-native speaker assessors would work with the 12 students concerned. This activity would serve the purposes of preparing the students for the tasks ahead and allow them to get to know and interact with the assessors in advance of the tests.

TRIAL

TRIAL FORMAT

Formal trialling of the procedures, format and content of prototype assessments were undertaken for the S2 tests. A member of the teacher panel from the pre-pilot stage kindly offered to pilot the tests with S2 students in his school and two half day visits were made by the French assessor team. On the first occasion the reading/writing test was conducted with 12 students and on the second, the listening test, bridging activity and speaking tests, again with 12 students.

It was important during the trial that assessors gained some experience in getting the balance between putting the students at their ease whilst still maintaining a test atmosphere. Also, the timing and management of the speaking test procedures was quite complex, with two assessors and two students operating in the same room whilst another student prepared and another completed an evaluation form. A dry run was needed in order to ascertain the feasibility of such procedures.

TRIAL OUTCOMES

The trials of the S2 tests did not lead to any significant changes to the tests themselves although in the final listening task (task 5) an additional third listening to all three sections was introduced to give students greater opportunity to cope with the demands of the task.

A simplified system for the real-time coding of student performance in the speaking test by the native speaker assessor was introduced because the demands of conducting the tests and noting student performance after each task had proved to be difficult to manage and detracted from the interaction between assessor and student.

During the trialling of the bridging activity, one higher ability student became quite distressed and said that she could not cope with the demands of the task. She was persuaded to attempt some of the tasks with the native speaker which she managed quite well, so that she did not leave the test with a sense of failure. However, this incident did alert the assessors to the anxiety which some students feel at having to speak the foreign language, particularly in an unknown test with unknown adults. The trial of the bridging activity did show that most students overcame their nerves through the informal interaction with the assessors during this half hour period. However, in order to prevent similar

scenes of distress, and given that the main aim of the pilot was to test the assessment instruments, it was agreed that teachers would be asked to select the 12 students from the 24 doing the tests in the other skill areas in their school. They were asked to ensure a balance of gender and ability levels in their choice of the sub-sample for speaking in selecting students whom they felt could cope with the demands of the situation, not simply the demands of the tasks.

FINAL S2 TEST OUTLINE

S2 Test A: Reading and Writing

Task 1: Reading	Discourse connection: matching six stimulus questions to the appropriate response
Task 2: Reading	Vocabulary recognition: multiple-choice questions
Task 3: Reading	Understanding the message of a number of short texts: open-ended questions

Task 4: Reading and Writing	Extracting specific information from a number of more extended texts drawn from an authentic source: open-ended questions and vocabulary retrieval
Task 5: Reading	Understanding the central message of an extended narrative text: (i) multiple choice (ii) open-ended questions
Task 6: Reading and Writing	Problem-solving and vocabulary recall at the single word level: comprehension and gap-filling
Task 7: Writing	(i) labelling and listing of nouns and articles with the aid of visual stimuli (ii) written response to questions using fixed visual stimuli
Task 8: Reading and Writing, Metalinguistic Knowledge	Comprehension and gap-filling: (i) copying from a list of possible responses (ii) vocabulary recall
Task 9: Writing	Open-ended writing task with the aid of visual stimuli

S2 Test B: Listening   Audio-Recorded Material, Native Speakers

Task 1	Recognition of single words via numbering of words in English
Task 2	Recognition of short phrases via numbering of visuals
Task 3	Understanding the subject and setting of short dialogues via grid-ticking exercise
Task 4	Extracting specific information from short monologues via grid-completion exercise
Task 5	Understanding the subject and message of more extended monologues via open-ended questions

S2 Test C: Speaking   Tasks 1-4: Native Speaker Assessor
                                 Tasks 5-6: Non-native Speaker Assessor

Task 1	Prepared talk on a topic
Task 2	Prepared semi-structured dialogue
Task 3	Part-prepared, part-spontaneous question-and-answer session
Task 4	Spontaneous description/narration of a composite colour visual
Task 5	Reading aloud a short text in the foreign language
Task 6	Metalinguistic discussion in English

S2 ASSESSMENT PROCEDURES

CONTACT WITH SCHOOLS

Prior to the visits the schools were usually contacted by telephone so that the Principal Teacher of Modern Languages could be made fully aware of the test procedures and arrangements. It was also essential to ascertain whether the school would be able to provide a playback machine for the listening tests or provide a room with PALE units.

With the enormous variation in schools in the timing and format of their school day it could prove to be a problem for the assessors to carry out the tests within the allotted time. With that in mind, it proved extremely useful for all parties involved if the NNS could draw up a draft timetable prior to the visit, so that the timing of the tests could be planned to fit round the school day.

Before the visits the NNS was issued with a list of the 24 students who would take part, including the 12 who would be involved in the speaking test and the list of 'reserves', were any of the original students to be absent on the day of the visit. The NNS was also given an indication of each student's ability level.

TIME IN SCHOOL

It was clear from the outset that the tests would be likely to last for a whole school day. In general the NNS arrived at the beginning of the day to set up the tests and to check the administrative arrangements with the Principal Teacher of Modern Languages. The NNS administered the Reading, Writing and Listening components of the test, with the NS arriving in time for the 'Bridging activity' prior to the speaking tests. Both NS and NNS were involved in assessing aspects of the speaking test.

PROCEDURE FOR READING/WRITING TEST

This combined Reading and Writing Test was administered first, and all 24 students were involved. Each student was allocated a number and that number written on his/her question paper. The student would retain it for the duration of the tests. Approximately 70 minutes was allocated for the whole test.

The procedures for Tasks 1-4 were explained by the NNS and a check made to ensure that all students understood. 20 minutes were allocated to this section of the test. Students were advised not to continue further with the test if they had completed these tasks, but were told simply to check over their work before they would be given instructions on Tasks 5-7. There was a break of 5 minutes between Tasks 1-4 and Tasks 5-7.

The same procedure was undertaken for Tasks 5-7. with 20-25 minutes being allocated to this section of the test. Again there was a break of 5 minutes between Tasks 5-7 and Tasks 8 and 9. This last section was allocated 20 minutes.

At the end of the test, the students were asked to fill in the 'Reading/Writing' section of their evaluation booklet. This booklet would be retained by the student and filled in after each test was completed. This would provide a record of how the student felt he/she had done in each task, and provide feedback on how fair or how easy each task had seemed to the students.

Prior to the Listening Test students were usually given a break of 10-15 minutes. If this happened to coincide with the school's morning interval, then the students were allowed to take advantage of this.

PROCEDURE FOR LISTENING TEST
This test followed the Reading/Writing Test, and again all 24 students were involved. 45 minutes were allocated to this test and its evaluation, although in reality the test tended to take around 30 minutes to complete. Students sat in the same seats as before and were allocated the same 'candidate number' as for the Reading/Writing Test.

Again the NNS went over each task in the Listening test, playing the appropriate section(s) from the tape. At the end of the test the students were again asked to fill in the appropriate section of their evaluation booklet.

At this point the students were told which of them would be taking part in the Speaking tests. The 12 who were **not** taking part were allowed to leave the room, and the 12 remaining students began the preparation phase, known as the 'Bridging activity'. For this part of the assessment both NNS and NS were involved.

BRIDGING ACTIVITY
The time allocated to this activity was 30 minutes. Of the twelve students involved, six received Speaking Booklet A and 6 Booklet B. They were also each given three 'Joker cards.'

The six tasks which they would be expected to carry out during the speaking test were explained to the students, and the opportunity given to them to prepare Tasks 1 and 2 and part of Task 3. Both NS and NNS were there to help. The Joker card system was also explained: if the students needed to use a word or phrase which they did not know or did not remember, they could ask the NS or NNS for the answer.

However, in asking for help, one of their Joker cards would have to be 'forfeited', thus leaving them with only two more opportunities for help. In reality, most students were able to find the answers for themselves, with a little prompting from the assessors, and very few students used up all three Joker cards.

Each student was then allocated a time for his/her Speaking test. It was explained to the group that when they came back to do their test they would have their Booklet returned to them and a further five minutes' preparation time, before doing Tasks 1-4 with NS (10 minutes) and Tasks 5 and 6 with NNS (five minutes).

TIMING AND ADMINISTRATION OF SPEAKING TESTS
In order to complete the tests in time, it was important to keep to a strict timetable. Each student spent a total of around 25 minutes doing this part of the assessment, including five minutes at the beginning and at the end for preparation and evaluation. The students arrived at ten-minute intervals, so this would mean that two, or even three students might be in the assessment room at any one time. With this is mind, it was important a) for the students to arrive on time, and b) for the assessors to

keep to a very strict timetable. The afternoon timetable below demonstrates how tight the schedule actually was:

ROLE OF NS/NNS IN ASSESSING SPEAKING TESTS
Both NS and NNS had their own role to play in assessing the speaking tests. They were responsible for recording each student's performance on tape, and to make sure that the timing of 'their' section of the test was strictly adhered to.

Tasks 1-4
The NS was asked to judge how well each student coped with each of the four tasks separately. He or she was then asked to give his/her overall impression of the student's pronunciation, fluency, accuracy and range of language. In both cases there were four 'categories' which the NS could use to describe the student's performance: *limited, adequate, good* and *very good*. At the bottom of the score sheet there was space for additional notes.

Tasks 5 and 6
The NNS was asked to rate the student's performance for each of the tasks on a scale of 1 to 4, with 1 being the lowest and 4 being the highest. Under each task section there was space for the NNS to take notes.

EVALUATION OF THE S2 TESTS

The tests were evaluated by three different groups, all of which had been involved in the primary assessments and associated procedures. These were:

- The native and non-native speaker assessors
- The S2 pupils who had taken the tests
- The principal teachers of modern languages in the secondary schools taking part

Each group was given a specific feedback questionnaire to complete. Copies of the school and pupil feedback questionnaires can be found in Appendices 1 and 2.

ASSESSOR FEEDBACK ON SECONDARY TESTS

The assessors provided very detailed and rich feedback on the tests and associated procedures. In Table 4a below we report a summary of their views on the implementation and logistics of the tests. For their detailed comments on each individual test and how they felt the speaking tests and scoring systems worked, please see Appendix 4.

Table 4a: Implementation and Logistics

What worked well	*Necessary to have two assessors throughout the whole evaluation process (morning and afternoon). It did make a **big** difference.* *Schedule of assessments well-planned.* *The test procedures generally worked well and the timings in terms of length of the tests were fine.* *The actual organisation of pre-Speaking Test preparation, followed by speaking tasks 1-4 then tasks 5-6, and finally evaluation went smoothly, with all pupils co-operating and being considerate.* *Task 2 worked very well although the pupils sometimes suggested a place instead of asking me **where** to meet.* *Task 3 – They seemed to like this task and often told me that the questions were easier than expected.*
What did not work well	*Planning of school visits quite vague and too tight.* *Exact role and involvement of external assessors who did not know enough of the project.* *The different timings of the tests vis-à-vis the school day. The early break did not present problems for the school or the pupils, although this meant that the listening test was taking place during school break and noise was a potential problem depending on the location of the room. The long school morning and short school afternoons meant that some pupils had to volunteer to come back early from lunch. Also there were no breaks in the speaking for the assessors, as some tests overran.* *The Bridging Activity caused alarm amongst pupils in all except one school. After a very long morning the pupils were unprepared by the organisation prior to the tests for having to stay for an additional half-hour. Some were quite hostile at the time.*
Suggested improvements	*Couple of days training **in teams** for **all** the people involved in the project.* *Definite school visit days (difficult to manage last minute changes).* *In terms of improvements/facilitating the implementation of the tests, I found it helpful if the 2 assessors were there from the start; if the assessors set out the ground rules in a no-nonsense manner; if the pupils were seated by the assessors boy/girl/boy etc (preferably low ability boy next to high ability girl).* *At this stage, I am not sure how best to amend the timings of each test to fit the prevailing school day. I don't think we can base a system around pupil volunteers.*

SECONDARY SCHOOL FEEDBACK

At the end of the pilot phase all the schools which took part were asked to fill in a questionnaire. The contextual data gathered from this survey have already been analysed in detail in Chapter 2. A very limited number of responses related specifically to the implementation of the pilot and highlighted aspects deemed negative by the schools concerned:

- The timing of the pilot, so close to the end of the school session, was criticised by a number of schools in both the primary and secondary sectors.
- A related issue raised by a limited number of secondary schools was the clash between the pilot and the introduction of the new school timetable.
- One secondary school also raised the issue of lack of parental and pupil support for the pilot, due to the need for the same pupils to be tested in both English and the foreign language.

S2 PUPIL FEEDBACK
A questionnaire was given to each pupil at the end of each test in S2. The findings from these surveys have already been analysed in some detail in Chapter 2 of the present report. A very limited number of responses related specifically to the implementation of the pilot:

Anxiety caused by tests
It was important to try to establish what impact if any, excessive anxiety or nerves might have had on pupil performance during the tests and so pupils were asked to give a 'before and after' snapshot of how they felt about each test.

Reading/Writing
Pupils were asked how anxious or relaxed they were before and after the reading/writing tests had taken place. Table 4b below summarises their responses:

Table 4b: Before the test started

	very anxious	anxious	average	relaxed	very relaxed
French	3%	20%	26%	40%	11%
German	4%	16%	26%	35%	18%

Table 4c: After the test

	very anxious	anxious	average	relaxed	very relaxed
French	3%	9%	24%	43%	21%
German	3%	10%	29%	37%	20%

From Table 4c above it can be surmised that anxiety levels among both samples decreased after the reading/writing test although 12% of the French and 13% of the German samples remained anxious or very anxious. It was encouraging to note that over half of each sample felt relaxed or very relaxed at the start of the tests with a further quarter in a neutral state of mind. The visiting assessors had gone to considerable lengths to stress the nature of the tests and the fact that individual measures of attainment would neither be recorded nor published.

Listening
Table 4d below summarises how students were feeling both before and after the listening test :

Table 4d: Before the test started

	very anxious	anxious	average	relaxed	very relaxed
French	2%	13%	34%	34%	16%
German	3%	11%	26%	37%	21%

Table 4e: After the test

	very anxious	anxious	average	relaxed	very relaxed
French	2%	7%	24%	45%	22%
German	2%	7%	27%	38%	24%

From Table 4e above, it can be seen that anxiety levels across both samples decreased once the listening test was over, and although 9% of each sample remained anxious about the test, this proportion was somewhat less than after the reading/writing test.

Speaking
Only 50% of those taking the S2 tests in reading/writing and listening then went on to take the speaking tests. Tables 4f and 4g below summarise how anxious students were about before and after the speaking tests:

Table 4f: Before the test started

	very anxious	anxious	average	relaxed	very relaxed
French	18%	45%	27%	8%	3%
German	11%	38%	28%	12%	7%

Table 4g: After the test

	very anxious	anxious	average	relaxed	very relaxed
French	2%	8%	29%	47%	13%
German	3%	8%	30%	38%	18%

It is perhaps not surprising that a far higher proportion of S2 pupils felt anxious or very anxious before the speaking test (63% for French and 49% for German compared with 15% and 14% for the listening and the reading/writing tests). However, by the end of the speaking tests 60% of the French and 56% of the German samples felt relaxed or very relaxed (broadly comparable with the proportions for reading/writing). Of course, this could be due to the sheer relief of knowing that the tests were over, but it may also be down to the very careful preparation of the speaking tasks by the two visiting assessors, native speaker and non-native speaker, during the bridging activity, and also the interaction with a real native speaker of the language concerned.

DIFFICULTY OF TEST

READING/WRITING
Pupils were asked to give an overall judgement of the difficulty of each of the tests and these results are summarised in tables 4h - 4j below.

Table 4h: How easy was the reading/writing test?

	very easy	easy	average	difficult	very difficult
French	1%	11%	60%	23%	5%
German	2%	8%	61%	22%	6%

Attitudes to the reading/writing tests were broadly similar across the two language samples.

Listening

Table 4i: How easy was the test?

	very easy	easy	average	difficult	very difficult
French	3%	17%	42%	28%	9%
German	3%	25%	45%	21%	4%

The German sample found the listening test marginally easier than the French sample (28% saying it was easy or very easy compared with 20%). Over a third of the French sample (37%) found the listening test difficult compared with only a quarter of the German sample 25%).

Speaking

Table 4j: How easy was the test?

	very easy	easy	average	difficult	very difficult
French	0%	15%	42%	38%	4%
German	6%	27%	45%	15%	2%

Although roughly the same proportion of pupils in the two samples found the speaking test of average difficulty, there was a noticeable difference between the French and German samples about how easy the test was. A third of the German sample (33%) found the speaking test easy or very easy compared with only 15% of the French sample. At the other end of the scale 42% of the French sample found the speaking difficult or very difficult compared with only 17% of the German sample.

FAIRNESS OF TESTS

It was assumed that S2 pupils would be used to assessments and tests as part of their foreign language learning at that stage and so it seemed appropriate to ask them how fair they felt each test had been. Their views are expressed in tables 4k - 4m below.

Reading/Writing

Table 4k: How fair was the reading/writing test?

	very fair	fair	average	unfair	very unfair
French	9%	43%	40%	4%	3%
German	5%	42%	42%	8%	3%

A slightly higher proportion of German pupils found the tests unfair or very unfair (11% compared with 7% for French) and this may reflect the fact that more pupils in the German sample commented in the open-ended views on each task, that they did not know the specific vocabulary required. However, around 50% of each sample thought that the test was fair or very fair and around 40% of each sample said the tests were of 'average fairness'.

Listening

Table 4l: How fair was the listening test?

	very fair	fair	average	unfair	very unfair
French	7%	38%	42%	10%	3%
German	5%	37%	39%	13%	4%

Listening was regarded by both samples as the least fair of the tests with 13% of the French and 17% of the German samples considering the test unfair or very unfair. Again, the German sample felt slightly more strongly on this issue. (See below for comments on the individual tasks.)

Speaking

Table 4m: How fair was the speaking test?

	very fair	fair	average	unfair	very unfair
French	14%	50%	33%	3%	1%
German	26%	44%	25%	1%	1%

Both samples agreed that the speaking test was the fairest test with 70% of the German sample and 64% of the French sample indicating that the test was either very fair or fair. Only a small minority (4% French and 2% German) thought that this test was in any way unfair. This was an interesting finding, given the level of anxiety which pupils in both samples expressed before the speaking test.

TASK COMMENTS

Pupils were given an opportunity to make a comment on each of the individual tasks making up the three tests of reading/writing, listening and speaking. Most commented on the degree of difficulty of each task, although a minority expressed other views relevant to an evaluation of the tests. The comments were coded and then processed.

Reading/Writing
The first four tasks of the reading test were graded in difficulty starting with the easiest task first (recognition of short sentence/phrase). Only 2% of the French and 1% of the German sample said they found this first task in any way difficult. Similarly, only 6% of the French and 2% of the German samples had any difficulty with Task 2 which involved reading short paragraphs.

However, the more extended written passages in Tasks 3 and 4 did cause difficulty for 45% of the French and 21% of the German samples. Task 3 involved reading short extracts of authentic material taken from the Internet and answering questions in English or identifying certain words in the texts. This was considered likely to be a less familiar task and type of reading material. Task 4 was a longer passage in the form of a letter with multiple choice an open-ended questions in English and although challenging in content was considered to be a more typical format for S2 reading and assessment material.

It was interesting to note that a similar proportion of pupils had difficulties with these tasks. Within each sample, 3% said they did not know the vocabulary for Task 3, and 4% of the French and 5% of the German samples said the same for Task 4. Task 3 was enjoyed by 5% of the French sample and 2% of the German sample whilst Task 4 was enjoyed by 2% and 3% respectively.

Task 5 was a combined reading and writing task which involved writing in school subjects in the foreign language into a timetable. Over a quarter of the French sample (27%) and a third of the German sample (33%) said they had some difficulty with this task, although 35% of the French S2 pupils said it was easy. Only 8% of the German sample held this latter view. In their comments 7% of the German sample said they did not know the necessary vocabulary, compared with only 1% of the French cohort.

Task 6 involved adding to a list of classroom objects in the foreign language, using an appropriate article. Twice as many French students (38%) as German students (19%) said this task was easy although a similar number in each sample expressed some difficulties with the task (24% and 20% respectively). Of the German sample, 9% said they did not know the vocabulary of classroom objects (compared with only 1% of the French sample). This may be a feature of the fact that far fewer of the

German S2 sample had started their foreign language learning in primary (30% compared with 77% of the French sample).

Task 7 involved writing three complete sentences or phrases to express the date, time and weather using the foreign language. Nearly a third of the French sample (32%) reported that this was easy or quite easy compared with only 12% of the German sample. However, 28% of the French sample and 20% of the German sample said that they found this task difficult. Again, 4% of the German cohort claimed not to know the relevant vocabulary.

Task 8 required the pupils to fill in gaps in a foreign language text. The German pupils found this task less demanding than their French counterparts, 23% said it was easy compared with 18% of the French sample, whereas 44% of the French sample reported some difficulty with the task compared with only 14% of the German S2 cohort.

Task 9 involved guided/open-ended writing and whereas 55% of the S2 French sample said they found this task difficult in varying degrees, only 26% of their German counterparts expressed the same view, although roughly the same proportion found the task easy (13% for French and 15% for German). Yet 4% of the German cohort said they did not know the necessary vocabulary for the task compared with only 1% of the French cohort.

Listening
The five listening tasks were graded in difficulty beginning with the easiest task, Task 1 which required pupils to recognise single words only to do with items of clothing. However, 5% of the French sample and 24% of the German sample indicated that they did not know the required vocabulary. That said, only 11% of the French and 13% of the German samples said they had any difficulty with the task. A small proportion of each sample (3% French and 4% German) said they found the tape too fast.

Task 2 involved recognising short phrases to do with free time and leisure activities and over half of the French sample (54%) and a third of the German sample (33%) found it easy. Although 2% of the German cohort maintained that they did not know the vocabulary, 3% said they enjoyed this task as did 5% of the French sample.

Task 3 which required pupils to identify the places where dialogues were taking place was found to be easy by 51% of the French sample and around a third of the German sample (32%). Only 17% of the French cohort reported having difficulty with this task compared with 9% of the German sample. A small proportion of the German sample felt that the tape was hard to make out.

Task 4 involved identifying items of food and drink and likes and dislikes associated with them. This task was considered easy by 17% of the French and 10% of the German samples, but 45% and 25% of the respective samples found it difficult. One possible explanation came from 14% of the French and 17% of the German samples who said they found the tape either too fast or too difficult to make out.

Task 5 involved listening to an extended narrative in three sections and more than twice the number of French S2 pupils found this difficult compared with their German counterparts (60% and 28% respectively). A similar proportion in both samples (15% and 16%) considered the tape too fast or too difficult to make out and 3% of the German sample said they did not know the necessary vocabulary.

Speaking
The speaking tasks consisted of some prepared and some spontaneous tasks. Task 1 was a prepared narrative and 43% of the French sample commented that this was easy, as did 25% of the German sample. By contrast 26% and 15% respectively in the two samples found the task difficult to some degree. A small proportion of the German sample (3%) said they did not know the necessary vocabulary. This task was enjoyed by 5% of the French and 6% of the German S2 pupils.

Task 2 involved a prepared role-play and was found easy by 50% of the French sample and 29% of their German counterparts. A smaller proportion in each sample found the task hard (15% and 7% respectively). This task was enjoyed by 10% of the French and 8% of the German samples.

Task 3 consisted of a question and answer session between the pupil and the native speaker assessor and 25% of the French and 18% of the German samples found this task easy. Three times as many French pupils reported difficulty with this task compared with the German sample (39% and 13% respectively), although once again 3% of the German cohort claimed not to know the necessary vocabulary. This task was enjoyed by 10% of the French and 7% of the German samples.

Task 4 required pupils to answer questions or make comments on a composite visual. This task was found to be the most difficult with 70% of the French sample reporting problems with the task. By contrast, only 34% of the German sample had such difficulties, although 11% again stated that they did not have the required vocabulary for the task. This task was considered easy by 10% of the French and 8% of the German samples and 2% and 3% of each sample respectively said they enjoyed the task.

Tasks 5 and 6 were carried out with the non-native speaker assessors. Task 5 involved pupils in reading aloud a short passage after a brief time to read it through. This task was found easy by 43% of the French and 31% of the German samples, whilst 21% and 10% respectively found it difficult. This task was enjoyed by 5% of the French sample and 3% of their German counterparts.

Task 6 involved the pupil in a discussion about metalinguistic awareness based on the passage they had read aloud. Roughly equal numbers of students found this task either easy (21% French/19% German) or difficult (22%/17% respectively) and 12% of the French cohort said they enjoyed this task, as did 5% of the German sample.

CHAPTER 5
FINDINGS ON PUPIL ACHIEVEMENT: LISTENING

This chapter describes the findings on pupil performance in the pilot assessments carried out in P7 and S2 in the receptive skill of Listening. Achievement at P7 will be examined first, followed by achievement at S2. The data on which the section draws derive from Tasks 4-6 of the P7 test and from Test B, Tasks 1-5 of the S2 Tests (see Appendix 5 for details).

A variety of different listening skills was assessed through a range of varied tasks and these can be divided into different types, each of which fall within the boundaries of the Listening for Information strand outlined in the 5-14 Guidelines for Modern Languages (1993):

- Vocabulary recognition
- Discourse recognition
- Information extraction/problem solving (S2 only)
- Central message extraction/recall

Within each of these sub-strands, a description is given of the scores achieved in each related task, followed by an analysis of full competence as compared to non-response and levels of partial competence. In each case, achievement in French is studied alongside achievement in German. This is possible since the majority of tasks from each test are identical in content and format. In addition, the scores of the S2 pupils are analysed by gender and level of ability.

At the end of this chapter the range of achievement in the listening is summarised in tabular form.

PRIMARY 7

All input was 'live', provided principally by the native-speaker assessor (NS), with limited assistance from the non-native speaker assessor (NNS). Each input was repeated twice.

Comprehension in the form of the 5-14 strand Listening to Establish Relationships with Others is included in the communicative Speaking tasks also forming part of the P7 Test, and will be analysed in Chapter 7.

VOCABULARY RECOGNITION

One task included an element of vocabulary recognition:
- Task 6, which involved the identification of five items from a possible five. A variety of domains were offered: clothes, food and drink, parts of the body, pets, weather. The audio stimulus material comprised of a series of five short sentences, while the written stimulus material comprised of a grid with headings in the form of a series of visuals.

Figure 5a: Breakdown of results for Task 6 - percentage of pupils scoring 0-5:
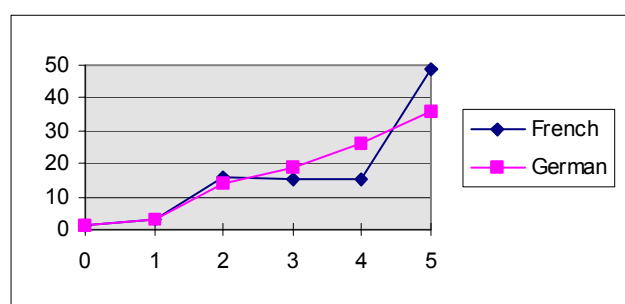
Full competence/Non-response

92 pupils of French sat this task. Each of the pupils involved attempted all of the questions. Of them, almost half scored full marks (48.9%), with just under two thirds scoring 80% or more. Only one pupil scored zero. A smaller number of pupils of German sat this task (only 72), with each pupil again attempting all of the questions. Of them, well over a third scored full marks (36.1%), with again just under two thirds scoring 80% or over. Similarly only one pupil scored zero.

Analysis of the cross-tabulations between achievement and the domain of language assessed would seem to suggest that for French pupils, the weather and clothes were the more difficult domains and food and drink and pets the more easy, while for German pupils the weather and food and drink were the more difficult domains and clothes, parts of the body and pets the more easy. Recent familiarity would seem to lead to a higher level of achievement, with the exception of the domain of the weather, which many pupils found difficult. This may be explained by the more complex nature of weather expressions, more demanding than the single word recognition required by the other domains.

DISCOURSE RECOGNITION

One task included an element of discourse recognition:
- Task 5, which involved the understanding and identification of the location of three dialogues from a possible six. The domains covered were places in town and simple transactional language. The audio stimulus material comprised of a series of three short dialogues, while the written stimulus material comprised of a grid with headings in English.

Figure 5b: Breakdown of results for Task 5 - percentage of pupils scoring 0-3:
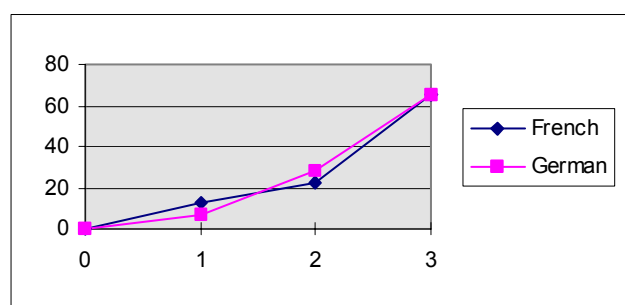
Full competence/Non-response

96 pupils of French sat this task. Each of the pupils involved attempted all of the questions. Over two thirds of pupils scored full marks (64.6%), with no pupil scoring zero. A similar number of pupils of German sat this task (94). Again each of the pupils involved attempted all of the questions. Again well over two thirds scored full marks (64.9%), with no pupil scoring zero. Slightly more pupils of German scored 2/3 than did pupils of French.

CENTRAL MESSAGE EXTRACTION/RECALL

One task included an element of central message extraction and recall of a longer input:

- Task 4, which involved the extraction and recalling of the central message of a short story. The audio stimulus material comprised of a short narrative.

For this task a second listening to a sample of performances and subsequent discussion among the native and non-native speakers who had conducted the tests, led to the establishment of three levels for this task: 1 to 3 (with 3 as the highest mark). A description of the levels is given below:

Level 1 – weakest students
- comprehend/retain very little: only understand individual items of vocabulary
- use random guessing to supply answers
- use generalisations like 'someone is doing something'
- need lots of support/prompting to answer questions on the story

Level 2 – average students
- need some prompting to get at exact meanings
- need quite slow pace, marked intonation, repetition of parts of story
- initially pick up on English vocabulary, or very familiar, basic information (name, age, etc.)
- pick up bits of the story but do not really understand the sequence of events fully
- understand/remember less information

Level 3 – excellent students
- require little prompting
- understand the sequence of the story for the most part
- display accuracy in terms of main elements of the narrative: place, people, time, action and some details: descriptions
- guess astutely
- problems experienced appear to be more of memory than of comprehension

OVERALL P7 LISTENING SCORES

In terms of overall achievement in listening at P7, it would appear that the levels reached in both French and German were of a similar high level. Achievement in the vocabulary recognition task in French was higher than that in German - a fact perhaps explained by the fact that a greater number of German pupils had been learning the foreign language in P7 only. It should also be noted that at P7 there was total participation on the part of the sample pupils: all of the questions were attempted by each of the pupils involved. It was therefore neither possible nor necessary to analyse levels of non-response.

SECONDARY 2

A variety of different listening skills was assessed via the varied tasks set as part of S2 Test B. These can be divided into different types, each of which fall within the boundaries of the Listening for Information strand outlined in the 5-14 Guidelines for Modern Languages (1993):
- Vocabulary Recognition
- Discourse Recognition
- Information Extraction/Problem-solving
- Central Message Extraction/Recall

All Listening Test material was professionally audio-recorded by a team of native-speakers. Each input was repeated twice with appropriate pauses between each reading to allow time for question completion. Task 5 was the exception to this, since here each extract was heard three times.

Comprehension in the form of the 5-14 strand Listening to Establish Relationships with Others is included in the communicative Speaking tasks forming part of S2 Test C, and will be analysed in Chapter 7.

VOCABULARY RECOGNITION

Two different tasks included an element of vocabulary recognition:
- Task 1, which involved the identification and numbering of six items from a possible ten drawn from the domains of clothing and adjectives/descriptions. The audio stimulus material comprised of a series of six short sentences, while the written stimulus material comprised of ten boxes containing single words in English, each accompanied by a blank numbering box.
- Task 2, which involved the identification and numbering of six items from a possible ten items drawn from the domains of leisure activities and opinions. The audio stimulus material comprised of a series of six very short monologues, while the written stimulus material comprised of ten boxes containing visuals, each accompanied by a blank numbering box.

Task 1
French
Breakdown of results for Task 1 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Dress	65.1	22.4	
2 Hat	45.3	39.2	Scarf (9.9%)
3 Skirt	51.7	32.8	
4 Shirt	30.6	24.6	Pyjamas (11.6%) Trousers (17.7%)
5 Jacket	32.8	44.4	Shirt (31.5%) Pyjamas (12.1%)
6 Trousers	56.5	14.2	Socks (9.1%)

German
Breakdown of results for Task 1 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Dress	9.2	51.5	Shirt (13.8%)
2 Hat	36	44.8	Coat (11.3%)
3 Skirt	37.7	38.1	Dress (17.2%)
4 Shirt	24.7	46	Hat (10%) Trousers (10.5%)
5 Coat	27.6	36	Dress (11.7%)
6 Trousers	41.4	32.6	None

Figure 5c: Breakdown of numbers of correct responses for Task 1 - percentage of students scoring 0-6:
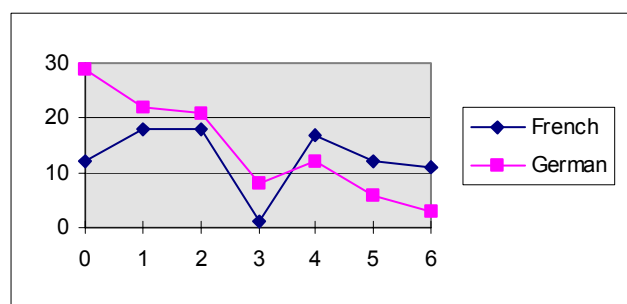
Full competence/Non-response
11.2% fully correct for French and only 1.3% fully correct for German.

In French, the largest percentage of correct scores was 1/6 or 2/6 – 18.1%, with the results quite evenly spread over each of the possible scores. In German the largest percentage of correct scores was 0/6 (29.3%), 1/6 (21.8%) or 2/6 (20.6%). Here, therefore the results were quite obviously clustered towards the lower end of the scale. On average, just under half of the French students (47%) successfully identified each item, as compared with only just over a quarter (29.4%) of the German students.

In both languages quite a high proportion of the students opted to leave the question blank: for each item in the task, on average 29.6% of French students and 41.5% of German students opted to make no response.

Partial competence
Some distractors (Item 4, shirt instead of Item 5, jacket in the French Test) may indicate that a certain number of students experienced some difficulty with the format of the task: the need to keep pace with the numbered items. However the very wide range of distractors would also seem to suggest that a relatively high proportion of students resorted to guessing at the answer.

It can be surmised, therefore, that students may not have been as familiar with this domain of language as had been expected. Indeed study of the school contextual data reveals that in six of the sample schools the students had not studied items of clothing in either S1 or S2, while in a further three schools, the topic had not been studied since S1. The fact that four of the former group of schools were German sample schools, may explain the poorer performance in German than in French for this particular task.

Task 2
French
Breakdown of results for Task 2 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Reading	87.1	7.8	
2 Cards	81	13.8	
3 Cycling	58.2	35.3	Dancing (12.1%) Painting (9.1%)
4 Television	97	1.3	
5 Walk	48.3	39.2	Dancing (16.4%) Painting (10.3%)
6 Swimming	91.8	6.5	

German
Breakdown of results for Task 2 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Reading	91.2	5.4	
2 Cards	81.6	15.1	
3 Walk	86.2	11.3	
4 Television	44.4	49.8	Swim (20.9%) Paint (9.2%)
5 Dancing	81.2	10.5	
6 Cycling	84.1	7.1	

Figure 5d: Breakdown of numbers of correct responses for Task 2 - percentage of students scoring 0-6:
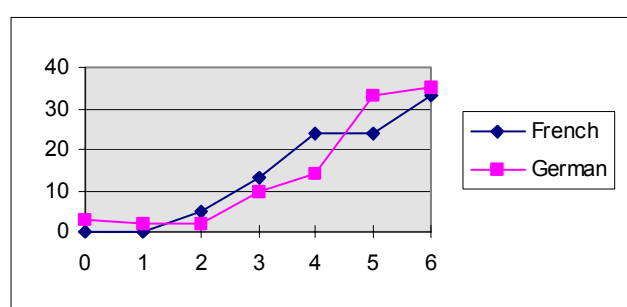
Full competence/Non-response
Task 2: 33.2% fully correct for French and 35.1% fully correct for German.

The largest percentage of correct answer was 6/6 (33.2% for French and 35.3% for German), 5/6 (23.7% for French and 32.8% for German) and 4/6 (24.1% for French and 14.3% for German). It would seem that comprehension was easier when words resembling English cognates were included in the audio input (*magazines, cartes, télévision, Comics, Karten, Disko*) or when the word was very familiar (*natation, piscine, Rad*). This reflects the findings of the APU report of 1986[10]. It would appear that this task was completed successfully by a high percentage of students of both languages: on average just over three quarters of students successfully identified each item.

Consequently a smaller proportion left the questions blank than had been the case for Task 1: for each item in Task 2, on average only 17.3% of French students and 16.5% of German students opted to make no response.

Partial competence
In some cases the nature of the distractors seems to suggest that the students were able to pick out verbal clues relating to types of activity: walking and cycling confused with other physical activities such as dancing. However in other cases it is difficult to interpret the significance of the distractors. It may thus be surmised that once again a certain proportion of students resorted to guessing at the answer.

---

[10] *Foreign Language Performance in Schools: Report on 1984 survey of French, German and Spanish* by Peter Dickson, Christopher Boyce, Barbara Lee, Matthew Portal and Malcolm Smith:
It will come as no surprise that in cases where the key French words have English cognates and sound like those cognates, pupils did well […] This also applied in German and Spanish. (p.34)

It can therefore be concluded that Task 2 was more successfully completed than Task 1. Possible reasons for this are better familiarity with the domain of language being tested since hobbies/leisure activities had been studied by all but one of the sample schools at some point during S1 and/or S2. Also, ease of comprehension of a slightly longer input may explain the difference in levels of achievement between the two tasks. Again this would mirror the findings of the APU report of 1986.[11]

DISCOURSE RECOGNITION

One task included an element of discourse recognition:
- Task 3, which involved the understanding and identification of the location of four dialogues from a possible seven. The domains covered were places in town and simple transactional language. The audio stimulus material comprised of a series of four short dialogues, while the written stimulus material comprised of a grid with headings in English.

French
Breakdown of results for Task 3 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Café	83.2	9.5	Shop (6.5%)
2 Zoo	89.7	9.1	
3 School	90.1	7.3	
4 Railway station	69	27.6	Hotel (6.5%) Shop (15.9%)

German
Breakdown of results for Task 3 - per item to be identified:

Item	% Correct	% Blank	Significant distractors
1 Café	95.8	1.7	
2 Zoo	63.6	34.3	Shop (27.2%)
3 School	87.9	9.2	
4 Railway station	82.8	14.2	Shop (8.8%)

Figure 5e: Breakdown of numbers of correct responses for Task 3 percentage of students scoring 0-4:
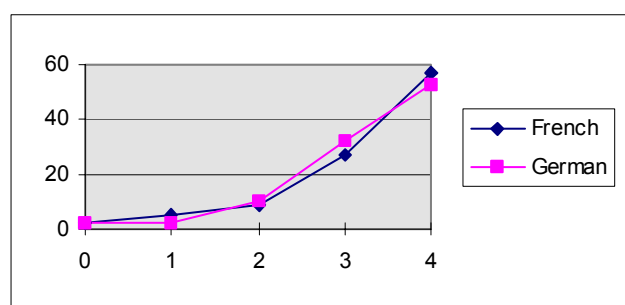
Full competence/Non-response
The largest percentage of correct answers in both languages was 4/4 (57.3% for French and 53.1% for German) and 3/4 (26.7% for French and 32.2% for German). Again it would seem that levels of comprehension were higher when words resembling English cognates served as contextual clues in

---

[11] *Ibid*:
In some cases such items (i.e. longer items) may be easier than those in category 1 (shorter items) – the longer text provides the run-in and the contextual clues absent in category 1, and so may give pupils a better opportunity to complete the item successfully. (p.33)

the audio input (*Orangina, éléphants, girafes, Pizza, Cola,*), or when the discourse was very familiar (classroom language). It would therefore appear that this task was again completed successfully by a high percentage of students of both languages: on average well over three quarters of all students successfully identified each item.

This interpretation is further supported by the fact that a smaller number of students were inclined to make no response than in either Task 1 or Task 2: on average only 13.4% of French students and 14.8% of German students left each item blank.

Partial competence
The nature of the distractors seems to show that the students were aware of the difference between transactional language, classroom language and general conversation. Generally locations involving a financial transaction of some kind were confused with other locations of the same type: café and railway station confused with shop and hotel. The railway station dialogue posed most problems to the French students: this type of role-play may be a topic area not covered by the end of S2. The confusion by more than a quarter of the German students between zoo and shop, while apparently more puzzling, may quite easily be explained by the reference in the German zoo dialogue to *Brot und Kekse* – items for sale in a shop.

INFORMATION EXTRACTION/PROBLEM-SOLVING

One task included an element of information extraction and problem-solving:
- Task 4, which involved the identification and allocation to either a 'Likes' or 'Dislikes' column of a number of items of food and drink. The domains covered were clearly food and drink and expressing likes and dislikes. The stimulus audio material comprised of four short monologues by different speakers. The stimulus written material comprised of a blank grid.

French
Breakdown of results for Task 4 – per item to be identified:

| Item	% Correct	% No mention/ no response	Partial competence
1 Pizza	92.2	6.5	
2 Coke	78	20.7	
3 Tea	24.6	74.1	
4 Meat	17.2	82.3	
5 Steak	31.9	28.9	Beef only, correct position (37.1%)
6 Chips	75	23.3	
7 Peas	20.3	64.2	
8 Orange juice	23.3	37.1	Orange juice, wrong position (15.9%) Orange only, correct position (12.9%) Orange only, wrong position (9.9%)
9 Fish	55.6	38.4	
10 Green beans	3.9	92.7	
11 Chicken	43.1	50	
12 Dessert	13.8	84.9	
13 Chocolate cake	60.3	3.9	Chocolate only, correct position (17.7%)
14 Tomato salad	31.5	14.2	Salad only, correct position (17.2%)

			Tomato/salad separate (25.4%)
15 Pâté	57.3	41.4	
16 Potato soup	19.8	18.1	Soup only, correct position (37.1%) Soup only, wrong position (12.9%)
17 Fruit	1.7	97.8	
18 Bananas	63.4	27.6	
19 Apples	49.1	42.7	

German

Breakdown of results for Task 4 – per item to be identified:

Item	% Correct	% No mention/ no response	Partial competence
1 Sausage sandwich	8	55	Sausage/cold meat only, correct position (14.7%) Bread/roll only, correct position (13.4%)
2 Apples	79	14.7	
3 Bananas	83.2	11.3	
4 Cheese	29	48.3	Wrong position (18.9%)
5 Oranges	13.9	69.7	Apple something, correct position (10.9%)
6 Hamburger	89.9	8.4	
7 Chips	84.5	13	
8 Mayonnaise	43.7	54.6	
9 Cold milk	14.3	23.1	Milk only, correct position (45.4%) Item correct, wrong position (8%)
10 Ketchup	63.9	19.3	Item correct, wrong position (16.4%)
11 Tomatoes	61.3	33.6	
12 Vegetables	28.6	63.9	Vegetarians, correct position (5.9%)
13 Spinach	7.1	91.6	
14 Cabbage	2.5	95.4	
15 Meat	55.5	37.8	
16 Chocolate	73.5	17.2	
17 Orange Juice	72.3	17.2	
18 Potatoes	58	38.7	
19 Rice	22.3	76.5	
20 Coke and lemonade	71	9.2	1 item, correct position (8.4%)
21 Tea and coffee	50	22.7	Items correct, wrong position (18.5%); 1 item, correct position (5.9%)
22 Hot drinks	0.8	99.2	

Figure 5f: Breakdown of numbers of correct responses for Task 4 - percentage of students scoring 0-19 for French or 0-22 for German:
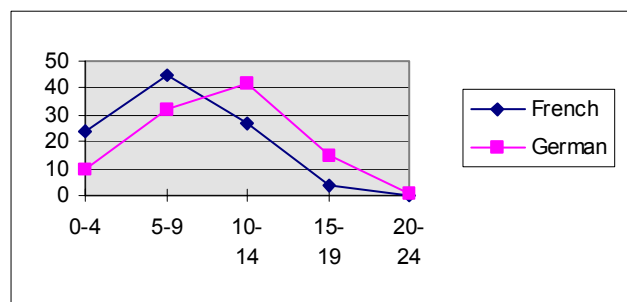
Full competence/Non-response
No students in the French sample managed to achieve the top score of 19 items correctly identified and positioned on the grid. The top score was 17 items, achieved by only 0.9% of the sample. Similarly no students of German achieved the top score of 22 items, although 0.4% did manage to score 21, 20 and 19. The score achieved by the highest proportion of students in the French sample was nine items correct (12.1% of the sample), while in the German sample it was 11 items (11.3%). This equates to approximately half of the number of possible items in each case. In fact on average each item was successfully identified by 40.1% of French students, as compared to 46% of German students.

Only a very small number of students failed to score: 1.7% of students for both French and German. The method used for analysing results for this particular task (no mention of an item and non-response calculated together) does make it difficult to calculate levels of non-response. However, since the number of students scoring zero is extremely low (1.7%) for both languages, it can be concluded that an even lower number of students failed to attempt this task. This may be directly related to the fact that the domain being tested (food and drink) had been studied by six of French and six German sample schools in S2 and four of each in S1. It was thus a familiar topic, recently studied by the majority of sample students.

Partial competence
Frequently students were able accurately to position the item on the grid, thus showing a fair understanding of the functional language used to express likes and dislikes. The most common error was placing the correct item in the wrong column of the grid (on average, per item, 3.6% of French students and 4.7% of German students committed this error). Thus it can be concluded that the students displayed a greater competence in identifying each item of vocabulary than the functional language attached to it. Nevertheless a much greater number in every case was able accurately to identify **and** position each item (on average, per item, 40.1% of French students and 46% of German students did so).

Certain item types were more difficult for the students to identify: compound items (*jus d'orange, salade de tomates, soupe de pommes de terre, Wurstbrot, eiskalte Milch*) and generic groups (*les desserts, les fruits, Gemüse, warme Getränke*). While some students successfully identified the items, others could not or could identify only one part of a compound item. Other items appear simply to have been less familiar to the students (*haricots verts, Spinat, Kohl*).

In addition error analysis revealed interesting levels of partial competence. Students, in their desire to note as many items of food and drink as possible sometimes confused target language words with other similar-sounding target language words: *fruits* was noted as chips – *frites,* and *tous les jours* as yoghurt – *yaourts*. Students also confused target language words with similar-sounding, although not equivalent, English words: *parce que* understood as pasta, *la même chose* as lemon/lemonade, *gar nicht* as garnish.

CENTRAL MESSAGE EXTRACTION/RECALL

One task included an element of central message extraction and recall of a longer input:

- Task 5, which involved recalling and extracting the global and central message from a number of longer inputs. The domains covered varied greatly. The stimulus audio material comprised of four more extended monologues from a single speaker. The stimulus written material comprised of a series of open-ended questions in English.

French

Breakdown of results for Task 5 – per item to be identified:

Item	% Correct	% No mention/ no response	Partial competence
1 Annie	75.4	0.4	Anne (14.7%)
2 14	87.5	6.5	
3 South of France	18.1	9.9	France only (57.3%)
4 Only child	52.2	6.5	
5 2 dogs	64.2	8.2	Correct animal, wrong number (18.1%)
6 1 horse	55.2	23.2	Wrong animal, correct number (18.5%)
7 Jean	19.6	47.8	
8 Police inspector	93.1	6.9	
9 Big	31.5	67.7	
10 Nurse	8.6	56.9	
11 Ice-skating	31.5	66.4	
12 Listening to music	64.7	6	Music only (29.3%)
13 Playing guitar	67.2	10.8	Guitar only (22%)
14 Shopping	90.9	8.6	
15 On Saturday	6.9	93.1	
16 With mother	37.1	62.9	
17 Boring/not like	50.9	23.7	
18 Art	48.3	50.9	
19 P.E.	42.7	13.8	Physics (43.1%)
20 Monday	65.5	23.3	In French (10.8%)
21 Saturday	59.9	30.2	In French (9.9%)
22 8.15	21.6	43.1	8.00 (22%)
23 Tuesday	49.6	42.2	In French (8.2%)
24 Friday	48.7	40.9	In French (10.3%)
25 3.40	10.3	53.4	4.00 (10.3%) 4.20 (5.6%)
26 Living room	52.2	40.1	In French (7.8%)
27 Dining room	28	69	
28 2 bedrooms	42.7	36.7	Correct item, wrong number (11.6%)
29 Parents' study	6	66.8	Parents only (12.9%) Study only (12.1%)
30 Big garden	55.6	32.7	Garden only (10.8%)
31 In front of house	1.3	96.1	
32 Hot	33.2	66.8	
33 Sunny	56.9	42.3	
34 Every day	5.6	90.1	
35 Write soon	7.8	75	
36 180	16.4	37.1	

German

Breakdown of results for Task 5 – per item to be identified:

Item	% Correct	% No mention/ no response	Partial competence
1 Thomas	97.1	0	
2 13	96.2	0.4	
Item	**% Correct**	**% No mention/ no response**	**Partial competence**
3 South of Germany	66	3.8	Germany only (20.2%)
4 Only child	72.3	2.1	Sisters, no 'no' (20.6%)
5 2 dogs	83.6	8.8	
6 1 budgie	64.3	30.7	
7 Karl	51.3	48.7	
8 Bus driver	25.6	74.3	
9 37	79.8	20.2	
10 Teacher	58.8	25.6	
11 Cycling	79.4	20.6	
12 Skiing	32.4	67.6	
13 Listening to music	60.5	10.5	Music only (28.6%)
14 Collecting stamps	47.9	49.5	
15 Boring	47.5	52.1	
16 History	53.8	45.8	
17 Geography	47.9	50.8	
18 Wednesday	76.5	20.6	
19 Thursday	72.7	23.9	
20 Lesson 1	46.6	52.5	
21 Tuesday	67.6	29.8	
22 Friday	81.1	16	
23 Lesson 3	54.2	45	
24 Living room	55	43.7	
25 Dining room	26.9	72.7	
26 2 bedrooms	54.6	34	Correct item, wrong number (9.7%)
27 Guest room	73.1	26.4	
28 Big garden	52.9	34.4	Garden only (12.2%)
29 In front of house	0.4	99.5	
30 Lots of snow	13.9	61.8	Snow only (24.4%)
31 In winter	47.1	53	
32 Write soon	30.7	50.8	
33 57	71.8	10.1	

Figure 5g: Breakdown of numbers of correct responses for Task 5 - percentage of students scoring 0-36 for French or 0-33 for German:
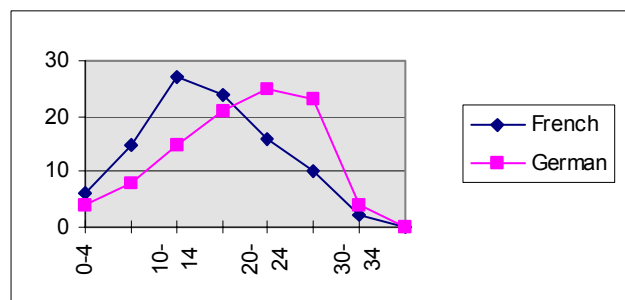
Full competence/Non-response

No students in the French sample managed to achieve the top score of 36 items correctly identified. The top score was 33 items, achieved by only 0.9% of students. Similarly for German no student achieved the top score of 33 items correctly identified. The top score was 31 items, achieved by only 1.7% of students. The score achieved by the highest proportion of students in the French sample was 14 (5.6% of the sample), while in the German sample it was 21 or 23 (6.3% of the sample). This equates to under half of the number of possible items for the French task, but well over it for the German task. In fact on average 41.9% of French students, as compared to 57.2% of German students, successfully identified each item.

No student failed to score. As for Task 4 the method used for analysing results for this particular task (no mention of an item and non-response calculated together) does make it difficult to calculate levels of non-response for individual items. However, since the number of students failing to score was zero for both languages, it is clear that all students attempted the task. This may be directly related to the fact that the many of the domains being tested (personal language, family, pets, hobbies, school, house, weather, numbers) were very familiar and had been studied by the sample schools in both S1 and S2. Indeed the items with the highest scores tended to be very familiar items (spelled out proper nouns, simple numbers, pets) or words with similar-sounding English cognates (*musique, guitare, shopping*).

It should be noted that certain items with particularly low scores (In French Task 5: item 3 - South of France, item 25 - 3.40pm, item 29 - parents' study, item 31 - in front of the house and item 34 - every day. In German Task 5: item 30 - lots of snow, item 29 - in front of the house) were actually either difficult compound items, or value-added items that it would be unreasonable to expect the majority of students to identify given the wording of the questions. It is therefore interesting that a small number was nevertheless able to do so in each case. Also the order in which some of the information required was given did not always correspond to the order of the questions set. However a relatively high percentage of students was still able to identify the correct response: on average over 40% of both French and German students did so for these particular items.

Partial competence

It would seem that many students were able to identify certain linguistic types of item easily: in particular nouns, less so verbs and prepositions. Thus with compound items such as south of France, listening to music, playing guitar some students were able to identify the noun only. Times also revealed levels of partial competence in that a certain percentage of students could successfully identify the hour, but not the minutes. Also longer numbers, while quite successfully identified in German, with over 70% of students identifying both 37 and 57, were less successfully completed in French: only 16.4% of students successfully identified the number 180. However some of the numbers suggested did reveal that the students had been able to identify certain elements of the number: 524, 104, 420, 5420.

BREAKDOWN OF OVERALL LISTENING SCORES

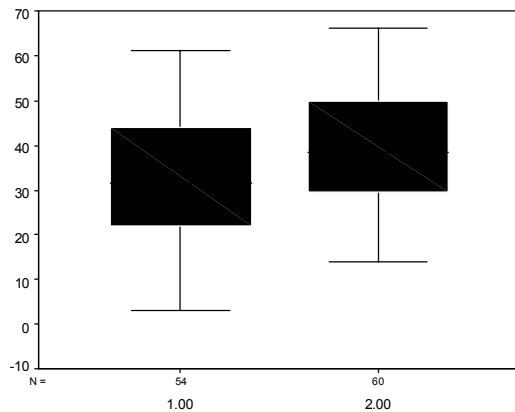**Figure 5h: Tasks 1–5 - percentage of students scoring 0-71:**
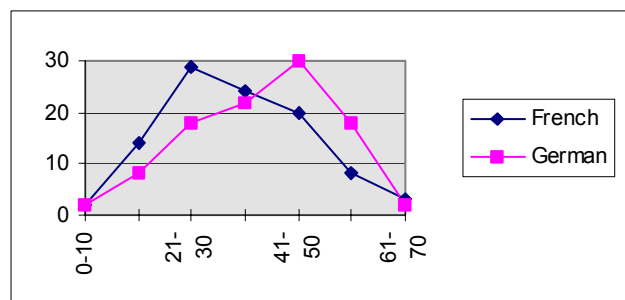
In terms of overall achievement in listening at S2, it would appear that S2 students were able to complete more cognitively complex tasks and deal with more extended and linguistically difficult audio-recorded input than pupils at P7. However, while all P7 sample pupils opted to participate fully in the pilot assessments, by the end of S2 a certain proportion of the students had decided that non-response was a better option than risk-taking or guesswork.

If we compare achievement in both languages we find that students of German achieved slightly better results than students of French. Although the points at either end of the scale were very similar for each language with regard to average score and percentage of students achieving that score, a higher percentage of German students scored a higher top score. In particular German students achieved higher scores in the tasks involving problem-solving/information extraction and central message extraction. In other areas achievement in each language was of a very similar level: in the skills of vocabulary and discourse recognition. The sole exception to this was the first vocabulary recognition exercise on the domain of clothes – a higher level of achievement by students of French may be explained by their better familiarity with this particular domain.

GENDER DIFFERENCES

In terms of the levels of achievement of each gender, analysis would seem to suggest that in both the French and German listening assessments girls outperformed boys in Tasks 1, 2, 4, and 5, while the achievement of each gender was virtually identical in Task 3 only (simple discourse recognition). Thus, overall, girls achieved at a slightly higher level than boys. This is outlined in the boxplots below. However it is not clear at this stage whether or not these differences are an artefact of the sample since, as was outlined in Chapter 2 of the present report, the samples were slightly skewed in terms of gender.

Figure 5i: French Listening Total Gender Difference:

Figure 5j: German Listening Total Gender Difference:


1: Male   2: Female
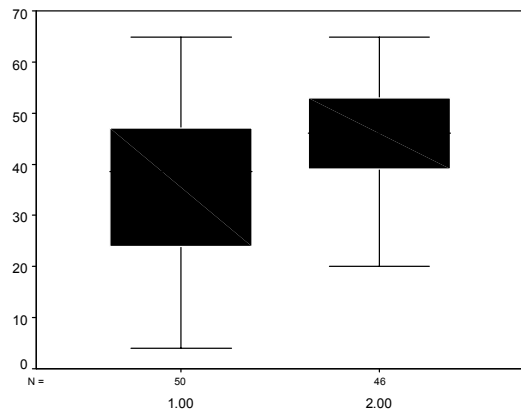
The French listening scores show that the extent of the cluster range for girls and boys is the same (20 points) with boys (scoring between 23 and 43 points) 7 points lower than girls (scoring between 30 and 50 points). The German scores show a wider range for boys than for girls, with girls scoring higher up the scale: between 23 and 46 points for boys, between 40 and 53 points for girls.


ABILITY DIFFERENCES

Ability differences in listening were analysed in a similar way. For both French and German, the expected 'staircase' distribution is found, but the results show that the cluster range within the top ability group is narrower for German than for French, while they are wider within the middle and bottom ability groups. In addition, there is a degree of overlap from group to group in French and German, in every case, which is more marked for German than for French. The figures below show these patterns. At the moment it is not clear what the significance of these differences between French and German may be, although the overall results appear to confirm teachers' own assessment of student ability.

Figure 5k: French Listening Total Ability Difference:
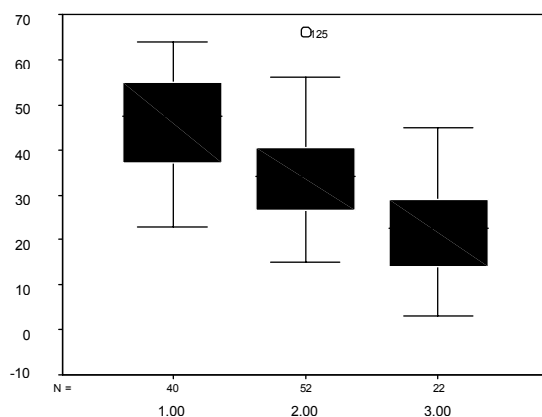
Figure 5l: German Listening Total Ability Difference:

1: High   2: Middle   3: Low

71
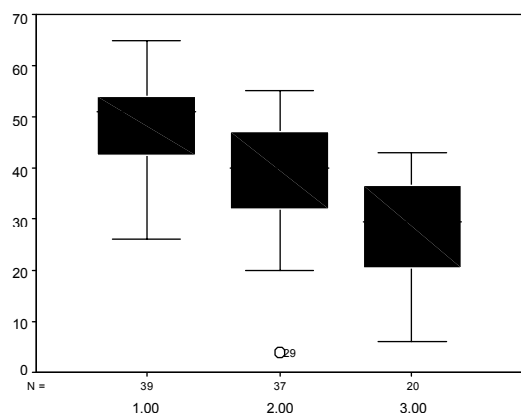
For the listening task scores in French, the highest ability group scores clustered between 37 and 55 points, while the other two groups clustered within narrower ranges: the middle ability group scored between 27 and 40 points, and the bottom group between 14 and 29 points. There is therefore a 3-point overlap between the middle and top groups, and a slightly narrower 2-point overlap between bottom and middle.

The listening task scores for German show a similar pattern. However, while the range within the middle and bottom groups is similar to those for French although higher up the scale, the range within the top ability group is narrower: only 11 points as compared to 18 points for the French top group. Also there is a slightly wider 4-point overlap in each case. The top group scores clustered between 43 and 54 points, the middle between 32 and 47 points, and the bottom between 20 and 36 points.

RANGE OF AURAL SKILLS - P7 AND S2

LISTENING SKILL	Excellent students can …	Average students can …	The weakest students can …
VOCABULARY RECOGNITION	**in P7** … successfully identify items of vocabulary from a given domain when no distractors are present - often after a single reading. **additionally in S2** … deal with more cognitively difficult tasks: successfully number a list of items containing distractors.	**in P7** … successfully identify most items of vocabulary from a given domain when no distractors are present - usually after 2 readings. **additionally in S2** … successfully number most elements from a list of items containing distractors and make reasonable guesses at the others.	**in P7** … successfully identify some items of vocabulary from a given domain when no distractors are present - after 2 readings. **additionally in S2** … successfully number some elements from a list containing distractors, and make usually random guesses at the others.
DISCOURSE RECOGNITION	**in P7/S2** … successfully identify the location of a number of dialogues – often after a single reading.	**in P7/S2** … successfully identify the location of most dialogues after 2 readings.	**in P7/S2** … successfully identify the location of some dialogues after 2 readings and make reasonable guesses at the others.
INFORMATION EXTRACTION/ Problem-solving	**in S2** … successfully identify and position items on a grid, including composite, generic and less familiar items; … sometimes confuse items with phonetically similar items from the target language- e.g. *fruits* confused with *frites*	**in S2** …successfully identify and position some items on a grid; … sometimes successfully identify items, but position them incorrectly; … sometimes confuse items with similar-sounding items in English- e.g. *parce que* confused with *pasta*	**in S2** **… successfully identify a limited number of items;** **… sometimes position them correctly on a grid, in particular items with phonetically similar English cognates-** e.g. *pizza, coca, bananas, Apfel, Hamburger, Schokolade*
CENTRAL MESSAGE EXTRACTION/ Recall	**in P7** … require little prompting; … understand the sequence of the story for the most part; … display accuracy in terms of main elements of the narrative: place, people, time, action and some details and descriptions; … guess astutely; … experience more problems of memory than of comprehension. **additionally in S2** … understand the majority of elements from more extended monologues – virtual global comprehension; … understand elements not strictly required by the questions, elements given in an order different to that of the questions set and more complex elements.	**in P7** … need some prompting to get at exact meanings; … need quite slow pace, marked intonation, repetition of parts of story; … initially pick up on English vocabulary, or very familiar, basic information; … pick up bits of the story but do not really understand the sequence of events fully; … understand/remember less information. **additionally in S2** … understand some elements from more extended monologues, without going beyond the information required by the questions; … achieve partial understanding of more complex elements.	**in P7** … comprehend/retain very little: only understand individual items of vocabulary; … use random guessing to supply answers; … use generalisations like '*someone is doing something*'; … need lots of support/prompting to answer questions on the story. **additionally in S2** … understand a limited number of elements, particularly the very familiar or less complex: simple numbers, words with similar-sounding English cognates.

CHAPTER 6
FINDINGS ON PUPIL ACHIEVEMENT: READING

This chapter describes the findings on pupil performance in the pilot assessments carried out in S2 in the receptive skill of Reading. Reading and Writing will be analysed at S2 only, since neither of these skill areas was extensively assessed at P7. The data on which the chapter draws derive from Test A, Tasks 1-5 and parts of Tasks 6-8 of the S2 Tests (see Appendix 5 for details).

A variety of different reading skills was assessed via the varied tasks set and these can be divided into different types, each of which fall within the boundaries of the Reading for Information strand outlined in the 5-14 Guidelines for Modern Languages:

- Vocabulary Identification
- Discourse Connection
- Central Message Extraction

The 5-14 strand Pronunciation and the Written Word was included in Task 5 of S2 Test C. However it will not be possible to analyse this aspect of reading ability within the present report. The 5-14 strand Reading for Enjoyment was not included in the pilot assessments, since this strand may be best assessed within the classroom through extended contact with the class teacher.

Within each of the three strands covered in the tests, a description is given of the scores achieved in each related task, followed by an analysis of full competence as compared to non-response and levels of partial competence. In each case, achievement in French is studied alongside achievement in German. This is possible since the majority of tasks from each test are identical in content and format. In addition, the scores of the S2 pupils are analysed by gender and level of ability.

At the end of this chapter the range of achievement in the reading is summarised in tabular form.

SECONDARY 2

VOCABULARY IDENTIFICATION

Three different tasks included an element of vocabulary identification:
- Task 2, which involved the identification of four items, each from a possible four - drawn from the domains of pets and weather/seasons. The written stimulus material comprised of four visuals each accompanied by four phrases from which the correct response should be selected;
- Task 4, part of which involved the identification and copying of items of vocabulary (two single words and a phrase for French, and two single words for German) from a continuous text;
- Task 8 (Part 1), which involved text identification and completion – six blank spaces to be filled using an option list containing ten items of vocabulary.

Clearly for the two latter tasks an element of the writing skill of copying is also involved. This particular aspect will be analysed in Chapter 8. For the purposes of this section the element of reading comprehension only will be examined.

Task 2
French
Breakdown of results for Task 2 - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 Dog	90.6	0	
2 Rabbit	92.8	0	
3 Snow	74	0.4	
4 Sunny	86.8	0	Cold in summer (7.2%)

German
Breakdown of results for Task 2 - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 Dog	99.2	0	
2 Rabbit	87.4	1.3	
3 Snow	87.8	0	
4 Sunny	73.1	0.4	Stormy (20.6%)

Breakdown of numbers of correct responses for Task 2 - percentage of students scoring 0-4:
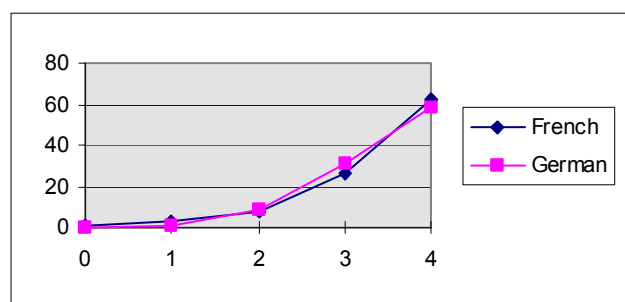
Full competence/Non-response
The largest percentage of correct answer in both languages was 4/4 (61.7% for French and 58.8% for German) and 3/4 (26% for French and 31.1% for German). This task was clearly completed successfully by a high percentage of students, since on average 86% of French students and 86.9% of German students successfully identified each item. It would therefore seem that the domains of language being assessed were very familiar and that the multiple-choice format was highly accessible.

This interpretation is further supported by the fact that a very small number of students was inclined to make no response: in most cases all students attempted each item, with only a negligible number failing to do so in one case for French and 2 cases for German.

Partial competence
There were very few distractors of any significance. The nature of the distractor in French would seem to imply that students were aware of the season mentioned (*été*) and allowed this knowledge to determine their response.

Task 4
French
Breakdown of results for Tasks 4 - per item to be identified and copied:

Word/phrase	% Correct identification/ copying	% Correct identification/ incorrect copying	% Partial identification/ correct copying	% Partial identification/ incorrect copying	% No response
Allez les voir	15.7	0.9	21.3	13.2	6.8
pauvre	47.7	1.3	0.9	0	8.9
naufrage	65.1	0.9	1.3	0	8.1

German
Breakdown of results for Task 4 - per item to be identified and copied:

Word/phrase	% Correct identification/ copying	% Correct identification/ incorrect copying	% Partial identification/ correct copying	% Partial identification/ incorrect copying	% No response
traurig	15.5	0.8	2.5	0.8	8.4
tragisch	26.1	1.3	4.6	0	18.1

Full competence/Non-response
In French the highest level of correct identification was for single words as compared to a phrase (*Allez les voir*). In German the level of correct identification of single words was lower. This was due to the relatively high number of students who confused the two quite similar words required (*tragisch* and *traurig*) – on average a quarter of German students did so.

The confusion outlined above also resulted in a higher average level of non-response for German than for French in this particular aspect of Task 4.

Partial competence
Levels of partial competence were discernible in a very small number of cases where students either identified only part of the phrase required, or included the word/phrase in a longer phrase than was required. This occurred more frequently in the French task, which involved the identification of a phrase: 34.5% of French students partially identified the correct phrase irrespective of their ability to copy it accurately.

Task 8
French
Breakdown of results for Tasks 8 (Part 1) - per item to be identified and copied:

Word/phrase	% Correct identification/ correct copying	% Correct identification/ incorrect copying	% Correct identification	% No response
m'appelle	86.8	5.5	92.3	1.3
ai	69.4	0.4	69.8	2.6
il	57.4	0	57.4	6
blonds	74	0.8	74.8	4.7
petite	48.1	6.8	54.9	3.8
est	29.8	0	29.8	3.8

German

Breakdown of results for Task 8 (Part 1) - per item to be identified and copied:

Word/phrase	% Correct identification/ correct copying	% Correct identification/ incorrect copying	% Correct identification	% No response
heiße	83.2	2.8	86	1.3
habe	89.9	0.8	90.7	0.4
er	79.4	1.2	80.6	3.8
blonde	87.8	1.2	89	2.9
kleine	40.8	1.2	42	3.4
ist	66	0.4	66.4	5

Breakdown of numbers of correct responses for Task 8 (Part 1) - percentage of students scoring 0-6:
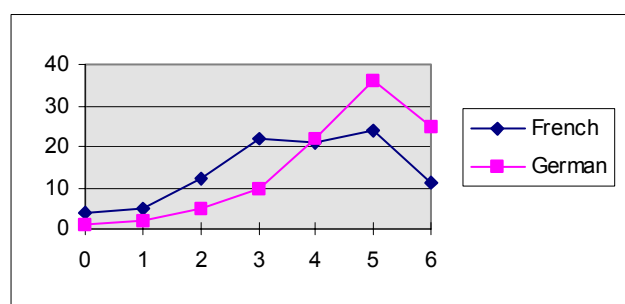
Full competence/Non-response

11% fully correct for French and 25% fully correct for German – if we consider both the skill of identification and the skill of copying together. In French, the largest percentage of correct scores was 3/6 (22%), 4/6 (21%) or 5/6 (24%). In German the largest percentage of correct scores was 4/6 (22%), 5/6 (36%) or 6/6 (25%). Here, therefore the results were quite obviously clustered towards the upper end of the scale.

If we consider the skill of vocabulary identification alone, irrespective of correct copying, on average almost two thirds of French students (63.2%) successfully identified each item, as compared with just over three quarters (75.8%) of German students.

In addition in both languages a very low proportion of the students opted to leave the question blank: for each item in the task on average 3.7% of French students and 2.8% of German students opted to make no response.

Partial competence

Certain words posed more of a problem than others, with a below average number of students successfully identifying them: *est*, *petite* and *il* for French; *kleine* and *ist* for German. Some explanation is possible for these low scores. For example in the case of the correct answer *est* in the French task, successfully identified and copied by only 29.8% of students, a large number of students (27.2%) selected the word *a*, thus correctly recognising that a verb was required, but selecting a word which they possibly believed to be phonetically correct. Similarly in the case of the correct response *kleine* in the German task, successfully identified and copied by only 45.4% of students, a large number of students (40.8%) selected the word *groß*, thus correctly recognising that an adjective was required, but failing to recognise the need for the adjective to agree with the noun.

DISCOURSE CONNECTION

One task included an element of discourse connection:

- Task 1, which involved the understanding and connecting of six stimulus questions to the correct response from a possible 12. The domains covered were personal language, likes and dislikes. The stimulus material comprised of two groups of three numbered stimulus questions in the target language, each attached to a group of six possible lettered responses in the target language.

Task 1
French
Breakdown of results for Task 1 - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 Name	89.4	0.4	
2 Age	86.4	0	*Elle a 13 ans* (12.8%)
3 Home	94.5	1.3	
4 Like sport	79.1	0.4	*Non, elles détestent le sport* (11.1%) *C'est le foot* (7.7%)
5 Favourite sport	54.9	0.4	*Il préfère la natation* (39.6%)
6 Favourite food	35.3	6.4	*Je voudrais une pizza* (31.1%) *Il préfère la natation* (12.3%)

German
Breakdown of results for Task 1 - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 Name	89.9	0	*Du heißt Peter* (8.4%)
2 Age	85.7	0.4	*Wir sind 12 Jahre alt* (11.8%)
3 Home	51.3	0.4	*Er wohnt in Berlin* (46.2%)
4 Favourite food	23.1	3.8	*Ich mag The Verve lieber* (41.2%) *Sie macht Musik* (15.5%)
5 Like football?	53.8	0.4	*Er hat einen Fußball* (45.4%)
6 Like Oasis?	47.9	1.7	*Ja sehr gern* (16.8%) *Sie macht Musik* (32.8%)

Breakdown of numbers of correct responses for Task 1 - percentage of students scoring 0-6:
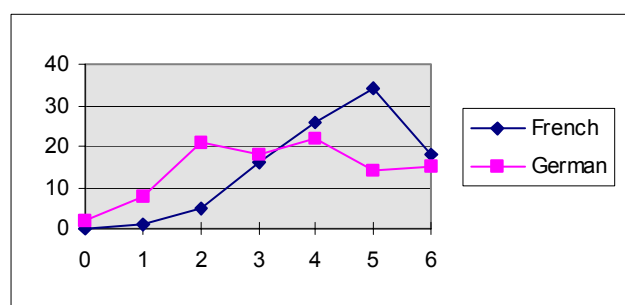
Full competence/Non-response

18% fully correct for French and 15% fully correct for German.

In French, the largest percentage of correct scores was at the upper end of the scale: 4/6 (26%), 5/6(34%) or 6/6 (18%). On average each question was connected to the correct response by 73.3% of French students. In German the largest percentage of correct scores was 2/6 (21%), 3/6 (18%) or 4/6 (22%). Here, therefore the results were quite obviously clustered lower down in the middle of the scale. On average each question was connected to the correct response by only 58.6% of German students.

In addition in both languages an extremely low proportion of the students opted to leave the question blank: for each item in the task on average 1.5% of French students and 1.1% of German students opted to make no response. Therefore there was virtual total participation in this task by all sample students. This may be due to the fact that it was the first task in the first test of the day. Also the task type was relatively simple: requiring students appropriately to letter the correct responses.

PARTIAL COMPETENCE

Certain questions and responses posed more of a problem than others, with a below average number of students successfully connecting them: *Quel est ton sport préféré?* and *Qu'est-ce que tu aimes manger?* for French; *Wo wohnst du?*, *Was ißt du am liebsten?*, *Spielst du gern Fußball?* and *Wie findest du 'Oasis'?* for German. Analysis of the distractors selected shows that in the majority of cases students selected responses with the correct general content, but failed to take note of certain grammatical features: the subject of the verb; the exact nature of the verb (e.g. confusing *vouloir* with *préférer*); the specific interrogative being used.

CENTRAL MESSAGE EXTRACTION

Six different tasks included an element of central message extraction:

- Task 3, which involved the understanding of three short texts (on average two-three sentences in length), and responding to open-ended questions in English. The domains covered were places in town and prepositions;
- Task 4, which involved the understanding of three slightly longer authentic texts drawn from the Internet (on average four-sevem sentences in length and containing some unfamiliar vocabulary), and responding to open-ended questions in English. The domains covered were cinema, likes and dislikes;
- Task 5 (Parts 1 and 2), which involved the understanding of an extended text in the form of a letter (each part four paragraphs in length and containing both unfamiliar vocabulary and grammatical structures), and responding to a series of multiple-choice questions in English (Part 1), and to a series of open-ended questions in English (Part 2). The domain covered was personal language;
- Task 6, which involved the understanding of a school morning timetable in the target language in order to complete a number of gaps. The domains covered were school subjects and the school day;
- Task 7 (Part 2), which involved understanding and responding in written form to three prompt questions in the target language using fixed visual stimuli. The domains covered were date, time, weather, interrogatives;
- Task 8 (Part 2), which involved text identification and completion – six blank spaces to be filled from memory, based on comprehension of a stimulus text. The domain covered was personal language.

For Tasks 6, 7 (Part 2) and 8 (Part 2) an element of writing skill is also involved: writing from memory at the single word or phrase/sentence level. This particular aspect will be analysed in Chapter 8, while the skill of comprehension alone will be examined here.

Task 3
French
Breakdown of results for Task 3 - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Centre	69.8	14.5	6.8	
2 Of town	44.3	18.7	6.8	Of village (22.6%)
3 Next to	28.9	63.4	6.8	
4 Market square	12.8	48.9	31.5	Market only (17.9%) In French (11.9%)
5 Is not	24.3	37.4	31.5	-ve, wrong verb (6.4%)
6 On photo	27.2	17	31.5	Photo only, not 'on' (23.4%)
7 But	7.2	60	31.5	
8 Is nearby	5.5	58.3	3.4	
9 Chemist	74	19.6	9.4	
10 Behind	23.8	65.1	9.4	
11 Chemist	68.5	21.3	9.8	
12 In	25.1	64.3	9.8	

| Item	% Correct	% No mention	% No response	Partial competence
13 Park	44.3	44.7	27.7	
14 Is not	0	72.3	27.7	
15 Far	0	71.9	27.7	
16 Go past	42.6	27.7	27.7	
17 Library	36.2	30.2	27.7	In French (6%)
18 Take second street	32.3	25.1	27.7	Take only, not 'second' (10.2%)
19 On left	44.3	22.6	27.7	
20 On right	16.6	40.4	22.6	
21 Facing	66	18.3	10.2	

German
Breakdown of results for Task 3 - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Centre	58.8	39.5	1.7	
2 Of town	78.6	10.1	1.7	*Stadtmitte* 1 word (9.2%)
3 Next to	82.8	13.9	2.1	
4 Market square	55	14.3	2.1	Market only (25.2%)
5 Is not	58.8	23.9	16.8	
6 On photo	58	22.3	16.8	
7 But	27.3	55.9	16.8	
8 Is nearby	29	41.6	16.8	Near something else (8.4%)
9 Chemist	63	15.5	6.3	In German

				(8.4%)
10 Behind	49.2	24.8	4.2	Other preposition (21.4%)
11 Chemist	58	26.9	4.2	In German (8.4%)
12 In	64.3	12.2	4.6	Other preposition (18.5%)
13 Park	84.9	10.1	4.6	
14 Is not	0.8	76.1	23.1	
15 Far	0.4	76.1	23.1	
16 Go past	17.6	47.5	23.1	Go to/from (7.6%)
17 Library	21.8	48.7	23.1	
18 Take second street	28.6	22.7	23.1	No verb (14.3%) 2 not second (8%)
19 On left	50.4	23.1	23.1	
20 On right	28.2	29.4	15.5	On right of sports centre (23.9%)
21 Facing	18.1	49.2	16	Other preposition (16.4%)

Breakdown of numbers of correct responses for Task 3 - percentage of students scoring 0-21:
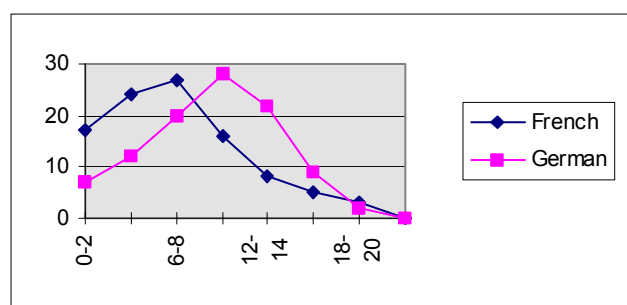
Full competence/Non-response

No students in the French sample managed to achieve the top score of 21 items correctly identified. The top score was 19 items, achieved by only 0.9% of the sample. Similarly no students of German achieved the top score, with only 0.4% managing to score 19. The score achieved by the highest proportion of students in the French sample was 6 items correct (11.1% of the sample), while in the German sample it was a significantly higher 10 items (10.1%). The latter equates to approximately half of the number of possible items. On average each item was successfully identified by only 33% of French students, as compared to 44.5% of German students.

In addition, in both languages - although more so for French than for German - a higher proportion of the students opted to leave the question blank than had been the case for the first 2 tasks: for each item in the task on average 19.8% of French students and 12.8% of German students opted to make no response. Therefore there was a higher level of non-engagement with the task than had been visible thus far, and a higher level of non-response in the French task than was visible in any of the reading tasks. This may be due to lack of recent practice in the domains of language involved: places in town and directions. Indeed analysis of the school contextual data reveals that four French schools had studied neither places in town nor directions since S1, with one school not having studied directions at all. Similarly one German school had not studied directions at all. However fewer German schools had not studied places in town or directions since S1: three schools for directions and only one for

places in town. Thus we also find a possible explanation for the slightly higher level of achievement in this task by students of German.

Partial competence

It would seem that many students were able to identify certain linguistic types of item more easily than others: in particular nouns, less so verbs and prepositions, with only a small number identifying negatives and conjunctions. Interestingly the students of German were more likely successfully to identify prepositions than were the students of French. The sole exception was *en face de*, identified as the similar-sounding English preposition 'facing' by 66% of students, while the German *gegenüber* was identified by only 18.1% of students. Some students of both languages also left certain items - place names in particular - in the target language, thus demonstrating an ability to locate the correct response without necessarily fully comprehending its meaning.

Task 4
French
Breakdown of results for Task 4 - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Special effects	90.2	3.4	2.1	
2 Extra-terrestrials	31.1	64.3	2.1	
3 Romantic	79.1	16.2	3.8	
4 Full of emotion	50.6	45.1	3.8	
Item	**% Correct**	**% No mention**	**% No response**	**Partial competence**
5 *Allez les voir*	15.7	42.1	6.8	Phrase in longer phrase, correct spelling (21.3%) Phrase in longer phrase, wrong spelling (13.2%)
6 Jack and Rose	51.5	8.1	4.3	1 correct character, no adjectives (10.6%)
7 Good actors	32.8	29.4	10.2	Actors only (26.4%)
8 Lots of	22.6	67.2	10.2	
9 Action	54.9	34.9	10.2	
10 *pauvre*	47.7	40.9	8.9	
11 *naufrage*	65.1	24.7	8.1	

German
Breakdown of results for Task 4 - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Martin	95.8	4.2	0	
2 Kathy	73.5	26.1	0.4	
3 Isabell	60.1	38.2	1.7	
4 Interesting	46.6	38.2	14.7	
5 Good love story	3.4	41.2	14.7	Action only (27.7%) History only (9.7%)
6 Good acting	73.5	5	4.2	She/it unrealistic

				(11.8%)
7 Good-looking	29	2.9	6.3	Nice/cute (19.3%) Good/very good (42%)
8 Preferred in R & J	28.2	6.3	5	He was in R&J (28.2%) She preferred R&J (26.1%)
9 *traurig*	15.5	40.3	8.4	*tragisch* (31.5%)
10 *tragisch*	26.1	30.6	18.1	*traurig* (19.3%)

Breakdown of numbers of correct responses for Task 4 - percentage of students scoring 0-11 for French or 0-10 for German:
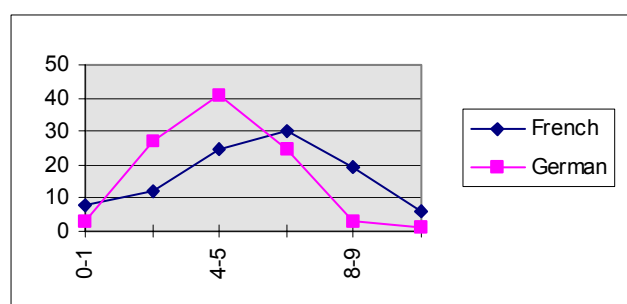
FULL COMPETENCE/NON-RESPONSE

A small number of students in the French sample managed to achieve the top score of 11 items correctly identified: only 1.3% of students. A very slightly smaller proportion of students of German achieved the top score of ten items correctly identified: only 0.4%. The score achieved by the highest proportion of students in the French sample was seven items correct (15.3% of the sample), while in the German sample it was a lower five items (21%). The latter equates to half of the number of possible items, while the French figure is higher. On average each item was successfully identified by 49.2% of French students, as compared to 45.2% of German students. In addition certain students were also able to identify value-added components not required by the question set: the adjectives used to describe items within each text.

In both languages - although slightly more so for German than for French – a quite small proportion of the students opted to leave the question blank: for each item in the task on average 6.4% of French students and 7.4% of German students. This is interesting in the light of the fact that these texts, the French text in particular, were authentic texts written by native speaker adolescents, which contained a certain amount of unfamiliar language.

Partial competence

Again students of both languages displayed a certain level of partial competence, discernible in their ability to identify at least one component of certain compound items: 'actors' only, not 'good actors' in the French text; 'preferred *Romeo and Juliet*' rather than 'preferred Leonardo Di Caprio **in** *Romeo and Juliet*' in the German text.

Task 5 (Part 1)
French
Breakdown of results for Task 5 (Part 1) - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 London	95.3	0.4	
2 End of June	94.9	0	
3 Seaside	81.3	0.4	Mountains (12.3%)
4 Sunbathe	72.3	3.4	Walk (17.4%)
5 Long walk	73.6	4.7	Run (11.9%)
6 Very hot	90.2	1.7	
7 Five	80	0.4	Twelve (13.2%)
8 Stay at beach	81.7	1.3	Eat ice-cream (12.3%)

German
Breakdown of results for Task 5 (Part 1) - per item to be identified:

Item	% Correct	% No response	Significant distractors
1 North of Scotland	98.7	0.4	
2 August	96.2	1.3	
3 Lake	64.3	2.5	Seaside (27.3%)
Item	**% Correct**	**% No response**	**Significant distractors**
4 Sunbathe	76.9	5	Walk (14.7%)
5 Climb	56.7	12.2	Short run (23.1%)
6 Very hot	89.1	5.5	
7 Five	84.9	3.8	
8 Stay at lake	71.8	6.7	Eat ice-cream (13%)

Breakdown of numbers of correct responses for Task 5 (Part 1) - percentage of students scoring 0-8:
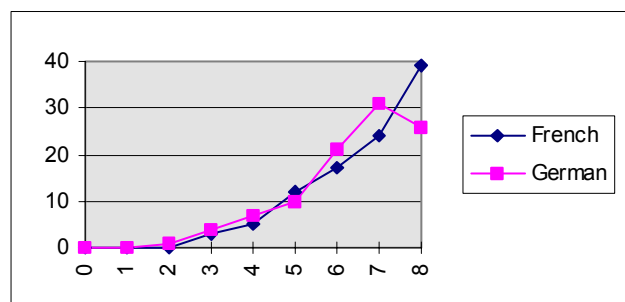
FULL COMPETENCE/NON-RESPONSE
39% FULLY CORRECT FOR FRENCH AND 26% FULLY CORRECT FOR GERMAN.
In French, the largest percentage of correct scores was at the upper end of the scale: 6/8 (17%), 7/8 (24%) or 8/8 (39%). On average each question was correctly answered by 83.7% of French students. In German the largest percentage of correct scores was also 6/8 (21%), 7/8 (31%) or 8/8 (26%), but with more students scoring 6 or 7 than the full score of 8. Here, therefore the results were clustered slightly lower down the scale. On average each question was correctly answered by 79.8% of German students.

In addition in both languages an extremely low proportion of the students opted to leave the question blank despite the length and complex nature of the text: for each question in the task on average 1.5% of French students and 1.1% of German students opted to make no response. Therefore there was virtual total participation in this task by all sample students. This may be due to the fact that the task

type was relatively simple: multiple-choice, thus providing some respite after more difficult open-ended questions.

Partial competence
In this particular task, the nature of certain distractors would appear to show some levels of partial competence. In particular several distractors demonstrate an understanding of elements actually contained within the texts, but an inability accurately to locate the appropriate response, for example 'walk' instead of 'sunbathe' for Question 4 in both the French and German texts. Other distractors would seem to indicate that guessing is taking place, but on the basis of some understanding of the gist of the text, for example 'eat ice-cream' instead of 'stay at the beach/lake' for Question 8.

Task 5 (Part 2)
French
Breakdown of results for Task 5 (Part 2) - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Cathy	94	1.7	1.7	
2 Red hair	42.6	17.9	29.4	
3 Glasses	28.9	29.4	29.4	Sunglasses (11.5)
4 Waited	0	88.5	11.5	
5 Looked for her	37.9	46	11.5	
Item	% Correct	% No mention	% No response	Partial competence
6 Shouted her name	65.5	11.9	11.5	Shouted only (11.1%)
7 Returned	5.1	82.6	11.5	
8 To the beach	13.6	74.9	11.5	
9 At top speed	0	88.5	11.5	
10 At the beach	24.3	44.7	26.8	
11 In her swimming costume	1.7	77.9	19.1	
12 Sitting	2.6	77.9	19.1	
13 In the sun	23	57.4	19.1	
14 Eating	46.8	23	19.1	Getting/buying (10.6%)
15 Vanilla ice-cream	21.7	23	19.1	Ice-cream only (35.7%)
16 Red	26.8	53.6	18.3	
17 Angry	24.7	56.6	18.3	
18 In the water	58.3	8.5	24.7	Jetty/other water source (6.8%)

German
Breakdown of results for Task 5 (Part 2) - per item to be identified:

Item	% Correct	% No mention	% No response	Partial competence
1 Nina	93.7	2.1	2.5	
2 Red hair	63.4	8	25.2	
3 Glasses	37.4	35.7	25.2	
4 Waited	1.7	79.4	18.5	
5 Teacher looked for her	0.8	43.3	18.5	Searched only (37.4%)
6 Shouted her name	52.9	9.7	18.5	Shouted only

				(18.9%)
7 Down	0	81.5	18.5	
8 To the lake	0	81.1	18.9	
9 At top speed	0	81.5	18.5	
10 At the lake	11.3	38.2	33.6	Reference to café (7.6%) At/in water (6.3%)
11 In her swimming costume	2.9	68.9	23.9	
12 Sitting	9.2	66.8	23.9	
13 In the sun	39.9	36.1	23.9	
14 Eating	52.9	23.5	23.5	
15 Potato salad	37	27.7	23.5	Salad only (10.5%)
16 Very red	0.4	71.8	27.7	
17 Angry	27.3	42.4	27.7	
18 In the water	75.6	3.4	13.9	Other water source (5.9%)

Breakdown of numbers of correct responses for Task 5 (Part 2) - percentage of students scoring 0-18:
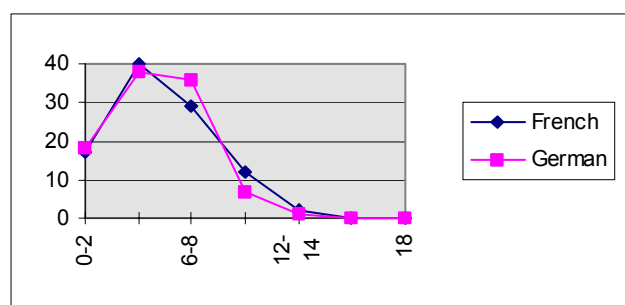
Full competence/Non-response

No students in the French sample managed to achieve the top score of 18 items correctly identified. The top score was 13 items, achieved by only 0.4% of the sample. Similarly no students of German achieved the top score, with only 0.8% managing to score 12. The single score achieved by the highest proportion of students in the French sample was three items correct (16.2% of the sample), with the largest percentage of students (40%) scoring 3-5. In the German sample the score achieved by the highest proportion of students was a significantly higher 6 items (13.4%), but with the largest percentage of students (38.2%) still scoring 3-5 – less than a third of the number of possible items. On average each item was successfully identified by only 28.8% of French students, and a very similar 28.1% of German students. However this is actually rather encouraging given the length and complexity (both linguistic and structural) of the texts involved.

In addition, in both languages - although more so for German than for French - a high proportion of the students opted to leave the question blank: for each item in the task on average 17.4% of French students and 21.4% of German students opted to make no response. Therefore for the German students there was a higher level of non-engagement with the task than had been visible in any other reading task. In the case of the French students however the level of non-response, although high, was lower than it had been for Task 3.

Partial competence

Again several levels of partial competence were discernible. Certain pupils managed to identify part of a number of compound items: 'shouted' only, instead of 'shouted her name', or 'ice-cream/salad' only, instead of 'vanilla ice-cream/potato salad'. Other items showed that the general setting of the

narrative had an influence over certain responses: 'sunglasses' instead of 'glasses' given the beach setting of the French text, 'café' instead of 'lakeside' given the reference to eating potato salad in the German text. Finally some responses demonstrated a certain confusion of target language words with similar-looking English words, but English words that matched the general setting of the narrative: *l'avons jetée à l'eau*, being understood as 'threw her off the jetty'.

Task 6
Correct understanding of a school timetable

French
Breakdown of results for Task 6 - per item to be positioned:

Item	% Appropriate, recognisable word/ correct position	% Appropriate, recognisable word/ incorrect position	% No mention/ no response
1 French	93.2	1.6	3.8
2 History	94.5	1.2	3.4
3 Computing	85.6	1.6	10.2
4 English	84.5	3.7	8.5
Item	**% Appropriate, recognisable word/ correct position**	**% Appropriate, recognisable word/ incorrect position**	**% No mention/ no response**
5 P.E.	59.7	32.1	33.6
6 Biology	93	1.2	4.7

German
Breakdown of results for Task 6 - per item to be positioned:

Item	% Appropriate recognisable word/ correct position	% Appropriate recognisable word/ incorrect position	% No mention/ no response
1 German	90.3	0.8	8.4
2 History	66.8	0.4	22.7
3 Computing	75.7	0.8	19.3
4 English	83	0	10.5
5 P.E.	68.8	16.2	29.4
6 Biology	84.3	1.2	12.2

Full competence/Non-response
If we consider both the skill of comprehension and the skill of writing from memory at the single word level together, there was 0% fully correct for French and only 3% fully correct for German. In French, the largest percentage of correct scores was 0/6 (40%), 1/6 (23%) or 2/6 (18%). In German the largest percentage of correct scores was 0/6 (29%), 1/6 (17%), 2/6 (16%) or 3/6 (20%). Here, therefore the results were clustered slightly higher up the scale.

If we consider the skill of comprehension/central message extraction alone, irrespective of correct writing from memory, on average well over three quarters of French students (85.1%), and a slightly lower proportion of German students (78.2%), correctly positioned an appropriate and recognisable, if not perfectly spelled, school subject, in such a way that demonstrated comprehension of the partially completed school timetable.

The difference in achievement between the two languages may not, in this case, be explained by the difference in the proportion of the students opting to leave the question blank or making no mention of the item, since fewer students of French failed to respond: for each item in the task on average 10.7% of French students opted to make no response, compared to 17.1% of German students.

Partial competence

Certain items posed more of a problem than others, with a number of students failing to locate them correctly: the word for PE in both languages, and the word for History in German. Some explanation is possible for these low scores. For example in the case of PE it is a compound word, which is therefore more difficult to recall, so leading to higher levels of non-response for these items. In addition the positioning of the item on the timetable was complicated by the fact that there was no Wednesday on the French timetable, leading to confusion between the days of the week. Also it was a morning timetable only leading to confusion between the breaks for interval and lunch. In the case of the German word for History it is a word that does not resemble its English counterpart, thus resulting in a lower level of accuracy.

Task 7 (Part 2)
Understanding of prompt questions

French
Breakdown of results for Task 7 - per prompt question to be understood:

Question	% Response showing understanding of prompt question	% Full of errors/wrong/English	% No response
Date	62.1	29.8	8.1
Time	43.4	35.7	20.9
Weather	52.3	23.4	22.1

German
Breakdown of results for Task 7 - per prompt question to be understood:

Question	% Response showing understanding of prompt question	% Full of errors/wrong/English	% No response
Date	68	24.4	7.6
Time	55.5	33.6	10.9
Weather	54.2	39.5	10.5

Full competence/Non-response

If we consider both the skill of comprehension and the skill of writing from memory at the single word level together, there was 1% fully correct for French and 4% fully correct for German. In French, the largest percentage of correct scores was 0/3 (58%) or 1/3 (31%). In German the largest percentage of correct scores was also 0/3 (48%) or 1/3 (32%). However 16% of students managed to score 2/3. Here, therefore the results were clustered slightly higher up the scale.

If we consider the skill of comprehension/central message extraction alone, irrespective of correct writing from memory, on average well over half of French students (52.6%), and a slightly higher proportion of German students (59.2%), produced a written response to each question which demonstrated comprehension of the prompt questions.

The difference in levels of achievement in this task between students of French and German may perhaps be explained by the difference in levels of non-response. On average more students of French opted to leave each question blank (an average of 17%) than did students of German (an average of only 9.7%).

Partial competence
Different levels of partial competence were discernible. However these relate more to the logistics of the test - a failure to follow the instructions, leading to a response not based on the visual stimulus - or to the skill of writing from memory - analysed in Chapter 8.

Task 8 (Part 2)
Understanding of stimulus text

French
Breakdown of results for Task 8 (Part 2) - per item showing understanding of stimulus text:

Item	% Correct spelling/ understanding of stimulus text	% Incorrect spelling/ understanding of stimulus text	% Understanding of stimulus text	% No response
1 *mon*	46.8	17.8	64.6	11.5
2 *est*	26.8	0.4	27.2	17
3 *travaille*/verb	13.3	8.8	22.1	41.7
4 adjective (f.)	8.5	36.5	45	43.4
5 *un*	51.5	15.7	67.2	14.5
Item	% Correct spelling/ understanding of stimulus text	% Incorrect spelling/ understanding of stimulus text	% Understanding of stimulus text	% No response
6 animal (m.)	64.6	9.2	73.8	14.9
7 *as*/verb	23	30.4	53.4	26.8
8 *sont*	1.3	36.2	37.5	33.2

German
Breakdown of results for Task 8 (Part 2) - per item showing understanding of stimulus text:

Item	% Correct word/ understanding of stimulus text	% Incorrect word/ understanding of stimulus text	% Understanding of stimulus text	% No response
1 *mein*	34	28.1	62.1	14.3
2 *ist*	45.8	2.1	47.9	24.4
3 *arbeitet*/verb	30.2	17.2	47.4	40.8
4 adjective (n.)	2.1	68.8	70.9	16.4
5 *ein*/number	57.2	29.2	86.4	10.9
6 animal (f.)	40.3	43.2	83.5	13.9
7 *hast*	60.5	8.3	68.8	21.8
8 *sind*	10.5	65	75.5	17.6

Full competence/Non-response
If we consider both the skill of comprehension and the skill of writing from memory at the single word level together, there was 0% fully correct for French or German. In French, the largest percentage of correct scores was 2/8 (24%) or 3/8 (22%). In German the largest percentage of correct scores was 3/8 (20%) or 4/8 (20%). Here, therefore the results were clustered slightly higher up the scale.

If we consider the skill of comprehension/central message extraction alone, irrespective of correct writing from memory, on average almost half of French students (48.9%), and over two-thirds of

German students (67.8%), demonstrated understanding of the stimulus text by completing each blank space with an appropriate and recognisable, if not perfectly spelled, word.

The difference in overall achievement between the two languages in this part of the task may be explained by the difference in the proportion of the students opting to leave the question blank or making no mention of the item: for each item in the task on average 25.4% of French students, as compared to 20% of German students, opted to make no response.

Partial competence
As for Part 1 of this task, analysed earlier in the present chapter, certain words posed more of a problem than others, with a below average number of students successfully identifying them: the verbs *sont* and *travaille* and the missing adjective for French. Some explanation is possible for these low scores. For example in the case of the correct answer *sont* in the French task, successfully identified and copied by only 1.3% of students, a large number of students (20.4%) selected the words *est* or *c'est*, thus correctly recognising that the verb *être* was required, but selecting the wrong form of the verb. This phenomenon was even more noticeable in the German task where 62.5% chose the singular form of *sein* and wrote *ist*, not realising that the plural form of the verb was required.

Similarly in the case of the correct response of a French adjective in the feminine form, successfully identified and copied by only 8.5% of the students, a large number of students (34.1%) selected the words *grand* or *petit*, thus correctly recognising that an adjective was required, but failing to recognise the need for the adjective to agree with the noun. Again, this feature was more marked in the German sample where no student was able to produce a correct neuter ending on an appropriate adjective, although 55.4% selected the words *gross* or *klein* in some form. The 2.1% of fully correct responses on this item in the German test were given not for an adjective at all but for the first part of a compound noun with *Haus*, such as *Einfamilien/Reihen/Doppel*.
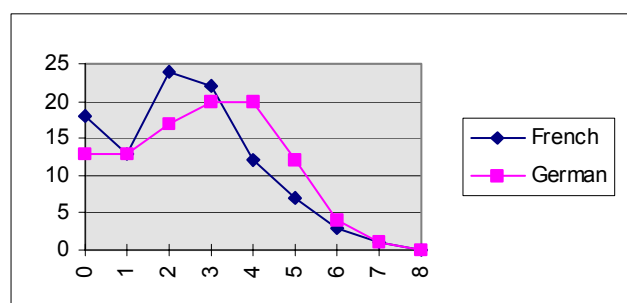
Task 8 (Part 2)
French
Breakdown of results from Task 8 (Part 2) - per item to be recalled and written:

Word	% Appropriate word/correct spelling	% Inappropriate word/correct spelling	% Correct recall any word/phrase	% No response
mon	46.8	33.5	80.3	11.5
est	26.8	42.7	69.5	17
travaille/verb	13.3	33.4	46.7	41.7
adjective (f.)	8.5	40.3	48.8	43.4
un	51.5	28.3	79.8	14.5
animal (m.)	64.6	8.2	72.8	14.9
as/verb	23	37.2	60.2	26.8
sont	1.3	48.7	50	33.2

German
Breakdown of results from Task 8 (Part 2) - per item to be recalled and written:

Word	% Appropriate word/correct spelling	% Inappropriate word/correct spelling	% Correct recall any word/phrase	% No response
mein	34	32.3	66.3	14.3
ist	45.8	24.3	70.1	24.4
arbeitet/verb	30.2	14	44.2	40.8
adjective (n.)	2.1	8.6	10.7	16.4
ein/number	57.2	2.8	60	10.9
animal (f.)	40.3	1.6	41.9	13.9

hast/verb	60.5	11.9	72.4	21.8
sind	10.5	4.5	15	17.6

Breakdown of numbers of correct responses for Task 8 (Part 2) - percentage of students scoring 0-8:

Full competence

Only one student of French and no students of German scored the top score of 8 in this task. In fact 17.9% of French students and a slightly lower number of German students (13.4%) did not score at all. However, on average, an appropriate word with the correct spelling was accurately recalled by 29.3% of French students and 35% of German students.

The exact figure for French varied from as much as 64.6% for the accurate recall of a masculine animal, to as little as 1.3% for the difficult third person plural form of the irregular verb *être*, or 8.5% for an adjective in the feminine form. For German it varied from 60.5% for the accurate recall of the second person singular of *haben* to as little as 2.1% for the first part of a compound noun with *Haus*, such as *Einfamilien/Reihen/ Doppel* which a number of ingenious students used in place of an adjective requiring agreement with the neuter noun (no student was able to produce such a form).

These more difficult items recording lower scores were also characterised by a higher level of non-response. On average just over a quarter of French students and just over a fifth of German students left each item blank.

Partial competence

In some cases students were able to recall an appropriate word, but with the wrong spelling. The system used to code the data, rendered difficult the production of exact figures for the number of students who did so, but spellings which accurately reflected the correct phonetic value of an appropriate word were relatively frequent in French (*ai* instead of *est*), as were spellings which reflected a mispronunciation of an appropriate word (*a* instead of *est*). For German, the main problem was getting the correct ending on possessive adjectives such as *mein* and the number *ein*, which also appeared with a range of different endings (*meine/eine, meinen/einen, meinem/einem)*. Such errors would seem to indicate interference from other case endings.

Students were also able accurately to recall an inappropriate word: on average, per item, 34% of French students and 12.5% of German students did so. In the case of French this was a higher figure than those who accurately recalled an appropriate word.

OVERALL READING SCORES

Tasks 1-5 - percentage of students scoring 0-68 for French and 0-67 for German:
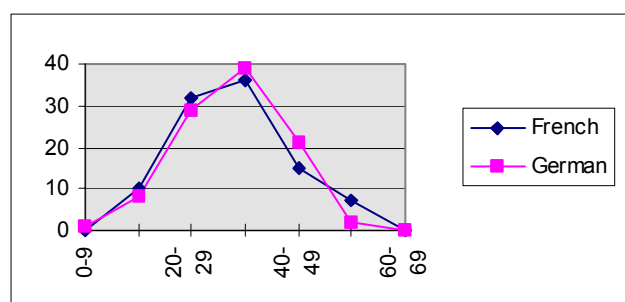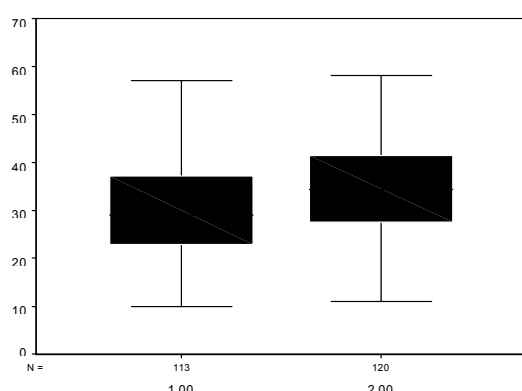

In terms of overall achievement in reading at S2, it would appear that, although results varied from task to task, achievement in French and German was of a very similar level. French achievement was slightly higher in Task 1 (discourse connection) and Task 4 (authentic, Internet-based task), while German achievement was higher in Task 3 (central message extraction centred on the domains of places in town and prepositions). A similar level was achieved in Task 2 (vocabulary identification) and Task 5 (central message extraction from an extended text, including unfamiliar vocabulary and structures).

GENDER DIFFERENCES

In terms of the levels of achievement of each gender, analysis would seem to suggest that in both the French and German reading assessments girls outperformed boys in Tasks 1, 3, 4, and 5. The achievement of each gender in each language was virtually identical in Task 2 (vocabulary identification), while boys in the German sample achieved a slightly higher level than the girls in Task 5A (central message extraction). Both of these tasks are multiple-choice exercises. Thus, overall, girls achieved slightly more than boys. This is outlined in the boxplots below. However it is not clear at this stage whether or not these differences are an artefact of the sample since, as was outlined in Chapter 2 of the present report, the samples were slightly skewed in terms of gender.

French Reading Total Gender Difference: *German Reading Total Gender Difference:*
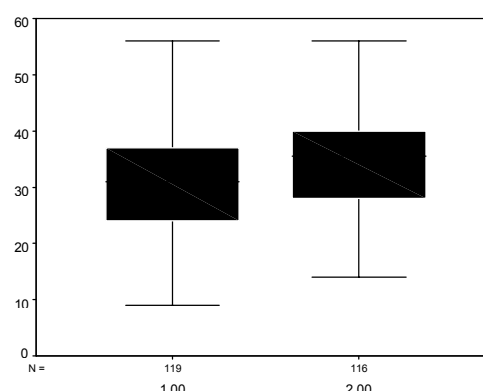

1: Male   2: Female

The French reading scores show that the extent of the cluster range for girls and boys is the same (14 points) with boys (scoring between 23 and 37 points) 5 points lower than girls (scoring between 28 and 42 points). The German scores show a slightly wider range for boys than for girls (14 points as compared with 12 points), with girls still scoring higher up the scale: between 23 and 37 points for boys, between 28 and 40 points for girls.

ABILITY DIFFERENCES

Ability differences in reading were analysed in a similar way. For both French and German, the expected 'staircase' distribution is found, but the results show that the cluster range within each ability group is narrower for German than for French. In addition, there is a degree of overlap from group to group in German, in every case, while this is less marked for French. The figures below show these patterns. At the moment it is not clear what the significance of these differences between French and German may be, although the overall results appear to confirm teachers' own assessment of student ability.

French Reading Total Ability Difference:
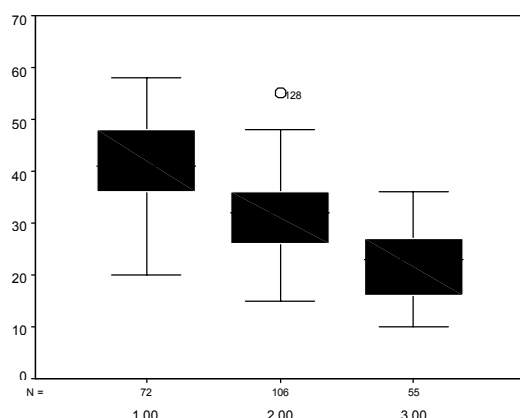
German Reading Total Ability Difference:
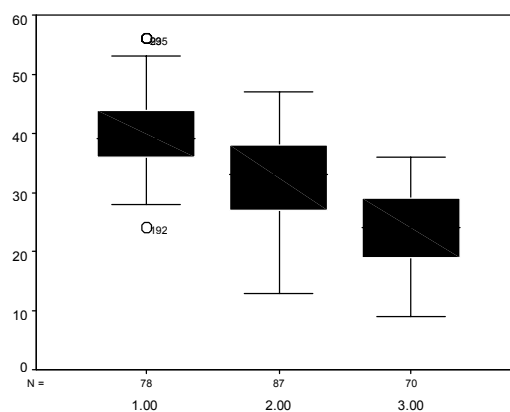
1: High   2: Middle   3: Low

For the reading task scores in French, the highest ability group scores clustered between 36 and 48 points, while the other two groups clustered within narrow ranges: the middle ability group scored between 26 and 36 points, and the bottom group between 16 and 27. There is therefore no overlap between the top and middle groups, but a 1-point overlap between middle and bottom.

The reading task scores for German show a similar pattern. However there is a narrower range, both overall and within the top and bottom ability groups, a slightly wider range within the middle ability group, and a 2 or 3-point overlap in each case. The top group scores clustered between 36 and 43 points, the middle between 26 and 38 points, and the bottom between 19 and 29 points.

RANGE OF READING SKILLS - S2 ONLY

READING SKILL	Excellent students can …	Average students can …	The weakest students can …
VOCABULARY IDENTIFICATION	… identify most single items of vocabulary/phrases from a list of distractors or a continuous text.	… identify some single items of vocabulary from a list of distractors or continuous text.	… identify a limited number of single items of vocabulary from a list of distractors and make guesses at the others.
DISCOURSE CONNECTION	… connect 2 sections of discourse exchanges involving very familiar and less familiar language; … show understanding of the subject and structures of the exchanges	… connect 2 parts of the majority of discourse exchanges involving very familiar language and some involving less familiar language; … make reasonable guesses at the others in such a way as to show understanding of the content if not the exact grammatical structures of the exchanges	… connect 2 parts of some sections of discourse involving very familiar exchanges; … make some guesses at the others, demonstrating some understanding of the general gist of the exchanges.
CENTRAL MESSAGE EXTRACTION	… understand the central meaning of texts of varying length, sometimes including unfamiliar vocabulary and structures; … identify sometimes complex/compound value-added items not required by the questions set; … deal effectively with varying formats: multiple-choice and open-ended questions in English.	… understand most of the central message of texts of varying length, but generally remaining within the limits of the questions set; … deal with varying formats: multiple-choice questions and open-ended questions in English.	… understand some very simple and familiar elements of texts of varying lengths, in particular via a multiple-choice format

CHAPTER 7
FINDINGS ON PUPIL ACHIEVEMENT : SPEAKING

This chapter describes the findings on pupil performance in the pilot assessments carried out in P7 and S2 in the skill of Speaking. The data on which the chapter draws derive from Tasks 1-4 of the P7 Test and from Test C, Tasks 1-4 of the S2 Tests (see Appendix 5 for details).

The data are reported in two ways. The first is a report of the scores which native speakers allocated to the students in S2 at the time the test took place. P7 students were also scored at the time of the tests. However, because of changes to the P7 speaking tests (see Chapter 3), these scores are not comparable across the whole cohort and have not been used in analysis. The S2 scores are analysed by task and by the discrete linguistic skills of :

- Pronunciation
- Fluency
- Accuracy
- Range

In each case, achievement in French is studied alongside achievement in German. This is possible since the majority of tasks from each test are identical in content and format. The data have also been analysed by gender and ability level.

The second is a consideration of what the tests tell us about the range of performance in P7 and S2: by the best students, average students and less able students. Analysis is based on a second listening to the tapes and discussion among the native and non-native speakers who conducted the tests.

At the end of this chapter, points made about the range of performance are summarised in tabular form covering the following aspects:

- Pronunciation and intonation
VOCABULARY
- Grammar
- Understanding
- Communicative skills
- Discourse skills
- Recall

SPEAKING TEST SCORES

S2 TASK SCORES

For S2 students, two types of judgement were made. Firstly, each of the four speaking tasks was scored on a scale of 1 to 4 (with 4 as the highest mark). Secondly, students were judged on their performance across the four tasks in relation to four discrete linguistic skills: pronunciation, fluency, accuracy and range.

Table 7a shows students' tasks in French and German on speaking tasks 1 - 4. Three French students present at the start of the day and selected for the speaking tests did not appear at the time they were due to be tested, for various reasons. Thus percentages for the French group come to just under 100%.

Table 7a: S2 Task Scores

Scores	French				German			
	T1 %	T2 %	T3 %	T4 %	T1 %	T2 %	T3 %	T4 %
1	47	22	18	49	12	13	14	19
2	24	35	43	25	33	35	24	41
3	14	27	21	16	41	42	45	31
4	13	13	15	7	14	10	17	9

Figure 7b: S2 Task 1

Task 1
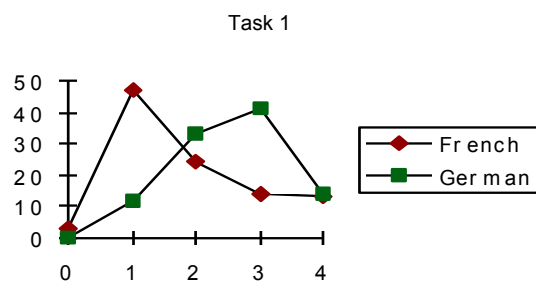
Figure 7c: S2 Task 2

Task 2
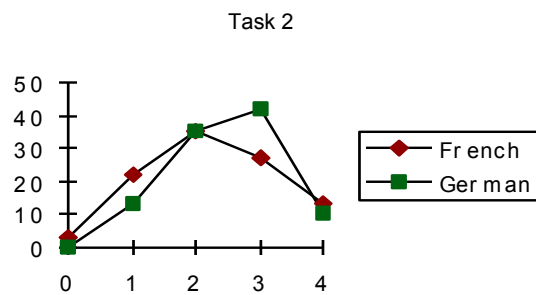
Figure 7d: S2 Task 3

Task 3
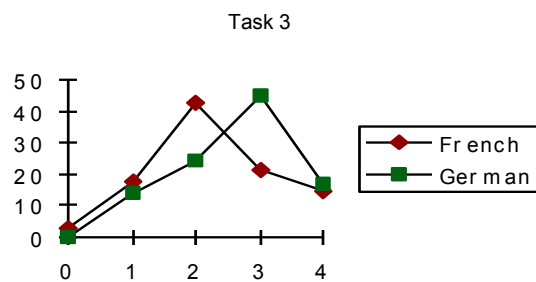
Figure 7e: S2 Task 4

Task 4
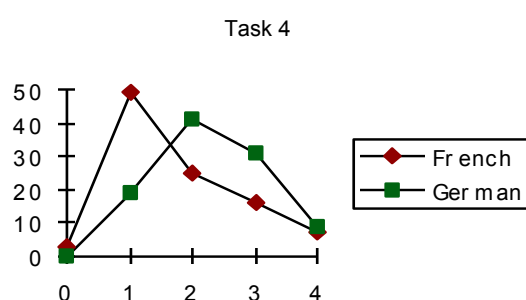
Two main issues emerge from a consideration of the task scores:

1. Is it easier to speak German than French?
The German scores suggest either that students taking German were better at speaking than those taking French, or that the German testers were more lenient in their judgements. In Tasks 1-3, more than half of the German student group scored 3 or 4, while in French, more than half scored 1 or 2.

2. Which is the most difficult task?
Task 4 (description of a composite colour visual) seems to have been more difficult than the other three tasks, for both French and German students. Almost half of the French group scored 1 on this task, while the proportion of German students scoring 3 or 4 dropped from over 50% to 40% on this task.

S2 SKILL SCORES

Table 2 shows S2 students' scores in relation to the four discrete language skills: pronunciation, fluency, accuracy and range. the highest score for each skill was 4 and the lowest was 1.

Table 7f: S2 Discrete linguistic skills

Scores	French				German			
	P* %	F %	A %	R *	P %	F %	A %	R %
1	28	41	53	48	5	17	14	15
2	35	28	25	25	29	27	40	31
3	26	19	9	15	54	43	40	41
4	9	9	10	9	12	12	5	13

*P= Pronunciation   F = Fluency   A = Accuracy   R = Range

Figure 7g: S2 Pronunciation

Pronunciation
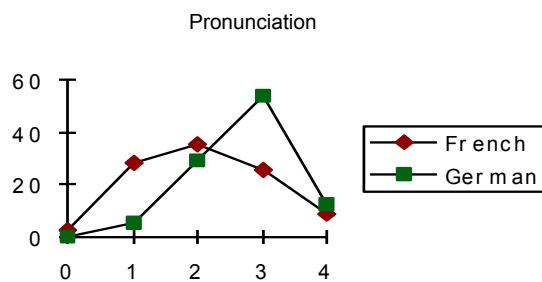
Figure 7h: S2 Fluency

Fluency

Figure 7i: S2 Accuracy
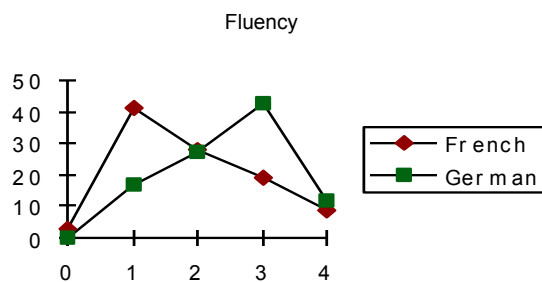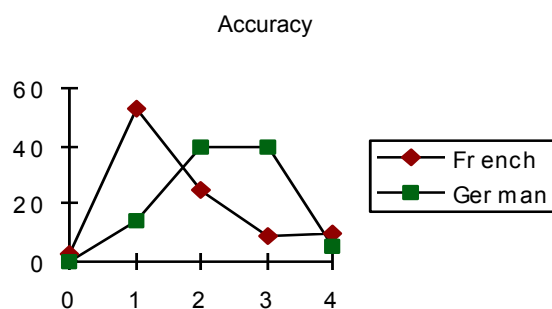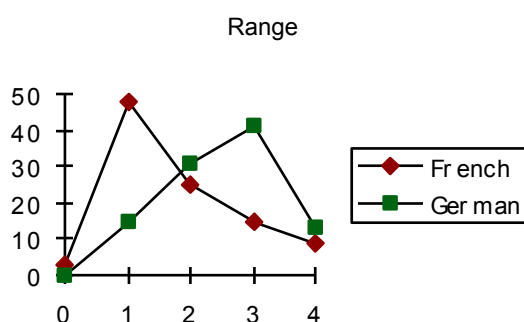
Accuracy

Figure 7j: S2 Range

Range

Again, these scores indicate that German oral skills are more advanced than French, with over half of the German student group scoring 3 or 4 on three of the four skills (they are weakest on accuracy, where the proportion scoring 3 or 4 drops to 45%). In contrast, over 60% of the French student group scored 1 or 2 on every count. Accuracy was also the area where French students were weakest: over three quarters (78%) scored 1 or 2 on this point, although it is true that slightly more French than German students achieved the highest score for accuracy (10% rather than 5%).

GENDER DIFFERENCES

Gender differences were analysed for the total 'task' scores and the total 'skill' scores. (The totals were thus out of a maximum of 16 points for each set of scores.) Figures 7k, 7l, 7m and 7n below show that girls did slightly better than boys in both French and German. However, it is not clear at this stage whether or not these differences are an artefact of the sample since, as was outlined in Chapter 2 of the present report, the samples were slightly skewed in terms of gender.

Figure 7k: Gender differences on total 'task' scores (French)

Figure 7l: Gender differences on total 'skill' scores (French)
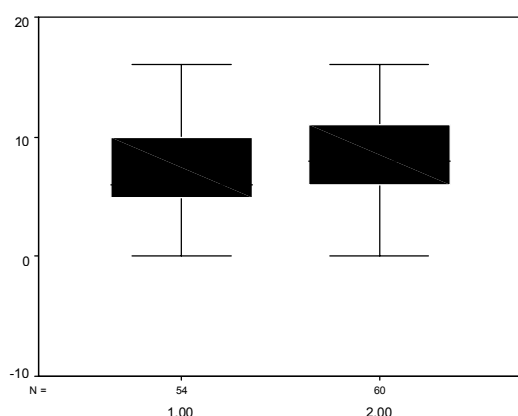
1: Male   2: Female

The French 'task' scores show that the extent of the cluster range for boys and girls is the same, with boys (scoring between 5 and 10 points) 1 point lower than girls (scoring between 6 and 11 points). The French 'skill' scores show that boys have a slightly narrower cluster range than girls, (between 4 and 8 points for boys, compared with 5 to 11 points for girls).
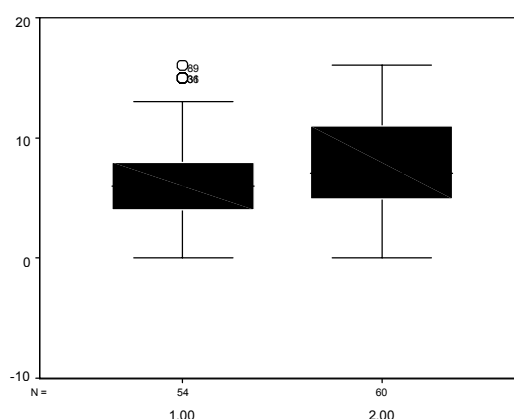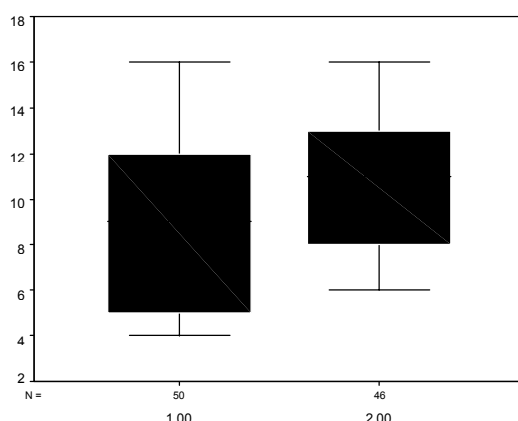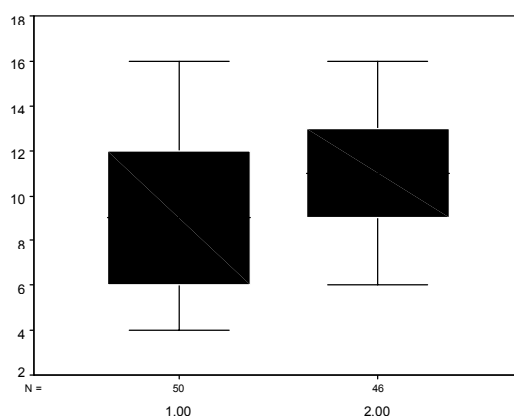
Figure 7m: Gender differences on total 'task' scores (German)

Figure 7n: Gender differences on total 'skill' scores (German)


1: Male   2: Female

The German 'task' scores show a slightly wider cluster range than with French, although gender patterns are similar: boys scored between 5 and 12 points, girls between 8 and 13. German 'skill' scores show a wider range for boys than for girls: between 6 and 12 points for boys, 9 and 13 for girls.

ABILITY DIFFERENCES

Ability differences on task and skill scores were analysed in a similar way. For both French and German, the expected 'staircase' distribution is found, but the results show that the cluster range within each ability group is wider for German than for French. In addition, there is a degree of overlap from group to group in German, in every case, while this is less marked for French. Figures 7o, 7p, 7q and 7r show these patterns. Any significance in these differences between French and German are not clear, although the overall results appear to confirm teachers' own assessment of student ability.

Figure 7o:Ability differences on total 'task' scores (French)
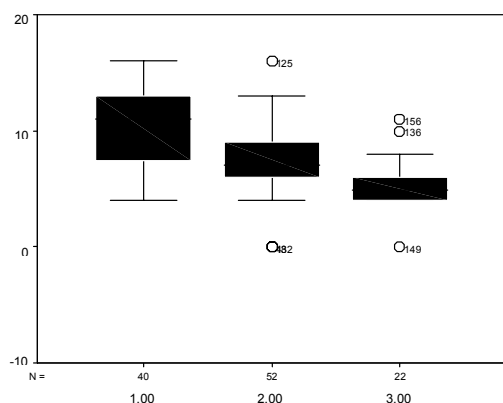
Figure 7p: Ability differences on total 'skill' scores (French)

1: High   2: Middle   3: Low
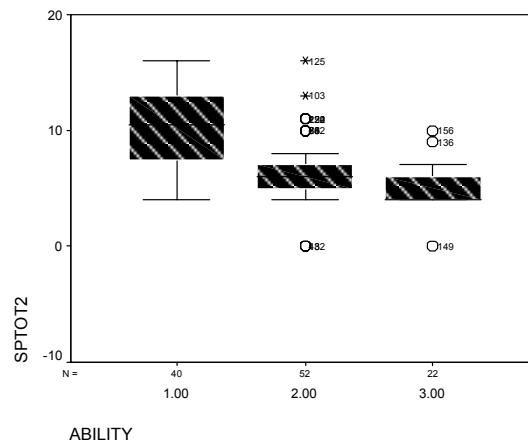
For the 'task' scores in French, the highest ability group scores clustered between 7 and 11 points, while the other two groups clustered within relatively narrow ranges: the middle ability group scored between 6 and 8 points, and the bottom group between 4 and 6. There is therefore no overlap between the bottom and middle groups, but a 1-point overlap between middle and top.

For the 'skill' scores in French, the highest ability group scores clustered between 8 and 11 points, the middle between 6 and 8 and the bottom between 5 and 7, showing a 1-point overlap between middle and bottom groups.

Figure 7q: Ability differences on total 'task' scores (German)
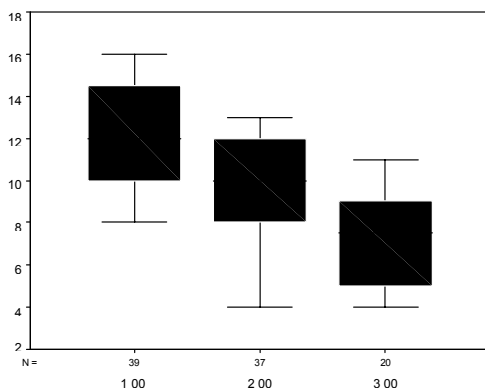
Figure 7r: Ability differences on total 'skill' scores (German)

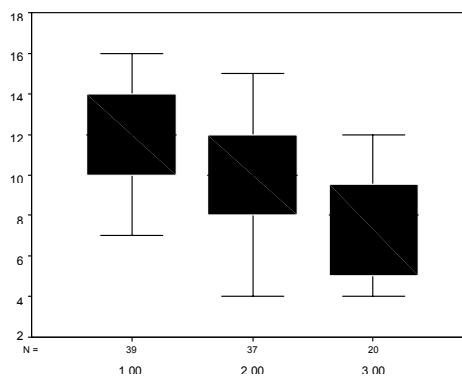1: High   2: Middle   3: Low

The 'task' and 'skill' scores for German show an identical pattern. There is a similar range within each ability group, and a 1 or 2-point overlap in each case. The top group scores clustered between 10 and 14 points, the middle between 8 and 12 points, and the bottom between 5 and 9 points, on both tasks and skills.

RANGE OF PERFORMANCE IN P7 AND S2

THE BEST STUDENTS

At the outset of the study, it was hypothesised that there might not be much difference between the oral skills of students in P7 and S2, in part because some S2 students would not have taken a language in primary school, and would therefore have spent about the same amount of time on language learning as their primary counterparts in the study, and in part because of evidence from other studies of a degree of repetition in what has been covered at primary school and what is covered in the early years of the secondary school.

However, analysis of S2 students' oral skills shows that the best students in this group perform at a significantly higher level than the best at primary school. For example, while the best P7 students showed an awareness of the distinctive features of French pronunciation and intonation and made good attempts at replicating these, some of the best S2 students had acquired convincing French accents and intonation patterns. Similarly, while both groups showed good grammatical control of what they were saying, the best students in S2 demonstrated a wider range of structures (for example reflexive verbs and phrasal verbs) and this enabled them to voice more complex accounts or ideas.

Where the best S2 students differed most markedly from the best of the P7 students was in what we have termed 'communicative' and 'discourse skills'. We look at each of these skills in turn below.

COMMUNICATIVE SKILLS
'Communicative skills' refer to aspects of communication which are not specific to the language studied - for example, confidence, interest, enthusiasm, and, in this context, a willingness to 'suspend disbelief' in the sense that the students behaved as though they were genuinely taking part in conversations, rather than being tested. Students with these qualities (which may or may not have been developed by their modern languages teachers) performed much better on all aspects of the oral tests than those who were diffident or shy. It is not possible to say whether their greater confidence enables them to grasp the technicalities of communication in another language more easily than students who are less confident, or whether the fact that they have grasped the technicalities generates a feeling of confidence.

The best students in P7 demonstrated a high level of confidence and enthusiasm for communication, and on occasion, they attempted to go beyond what they had learned to convey their own experience or ideas.

The best S2 students, perhaps because of their greater overall command of the language, demonstrated two divergent approaches to situations in which they had the opportunity to speak at some length. Some students ('defensive players') made use of structures with which they were familiar to 'frame' their speech. So, for example, the task which asked students to explain what they did throughout the day or in their free time over a week, used time phrases they had learnt to provide a clear structure into which they were able to insert a variety of statements. Such students rarely made errors and operated always within the bounds of what they knew they could do. However, their high level of competence meant that they could confidently speak at length.

Other students interpreted the demands of the task differently. Although as competent linguistically as the 'defensive players', their desire to communicate and to express their own experiences, thoughts

and ideas, meant that frequently went well beyond the 'safety' area of what they knew they knew. These students ('risk-takers') attempted levels of linguistic complexity which they had not necessarily covered in their school work. Sometimes they were successful, sometimes less so (although what they said was always comprehensible).

Both strategies are valid approaches to language learning. The 'defensive players' may be more likely to score highly in examinations, particularly those where accuracy is highly regarded; but the 'risk-takers' are perhaps more likely in the long run to achieve high levels of competence in the language because their real desire to communicate will encourage them to learn how to do this. 'Risk-takers' may turn out to be the students most likely to want to study languages to a high level and to 'become linguists' - in the sense of using languages in their adult life, for work, travel or cultural reasons. Extract 1, below, is the transcript of a 'risk-taker' response to a question about daily activities. Despite reliance on the infinitive, this response is more detailed and more engaging than the list of activities which 'defensive players' tend to produce, following the structure: *À 9 heures, je … À 10 heures, je …*, etc. The extract also shows developing awareness of the use of the past tense, even although this is not always correctly applied.

Extract 1

Je me suis levée à huit heures. Je prendre une jupe, une pull et un T-shirt. Je mange le petit déjeuner: un croissant et une crêpe. Je bu un, du chocolat chaud. Je faire une promenade chez mon ami(e) et on prendre le voiture en école. À l'école j'ai espagnol, français, anglais, maths, les sciences physiques et chimie et histoire. Mon sujet préféré est espagnol parce que j'aime le professeur et la classe est petite avec huit personnes. Je déteste les maths parce que il est très dur, et physique aussi. Quand je rentre je prendre des pantalon(s) et un T-shirt. Je fais mes devoirs et jouer avec mes soeurs. Aujourd'hui je suis allée à Guides après mon dîner. Mais en autres jours je jouer mon violon. Aussi je jouer avec mon ordinateur.

DISCOURSE SKILLS

'Discourse skills' refer to linguistic aspects of conversation which go beyond individual words or phrases. These skills require specific knowledge of how the language in question operates: for example, in order to be able to ask questions, to link sentences, to be polite, students need not only to understand the need for these elements within discourse, but also to know the appropriate structures or forms for these element in the language they are learning.

In P7, the best students voluntarily asked questions as well as answered them. They often sought to go beyond the simplest answer, providing explanations for a 'yes' or 'no' answer, or qualifying their response in some way. They also often aimed to answer using a full sentence rather than minimal responses.

The best S2 students demonstrated a wide range of discourse skills, and it seems likely that it is the development of this type of skill which distinguishes students who will do well in modern languages from the others. As well as the skills which the best P7 students had acquired, the best S2 students were able - among other things - to link sentences in varied but appropriate ways, make phatic utterances in a convincing manner, and to vary the structures used so that the tone of the conversation maintained a level of interest.

AVERAGE STUDENTS

There is also evidence to show that students defined as 'average' in S2 have a wider repertoire of skills than those in P7. For example, both groups can pronounce words clearly (though few make determined attempts to replicate the appropriate accent or intonation) and recall the main words and phrases they need for the tasks in hand. However, the vocabulary range of average students in S2 goes beyond that of the P7 students, whose descriptive vocabulary, for example, is limited mainly to

colours. In S2, average students could produce short basic phrases in order to describe a scene, such as *il fait du soleil; c'est le 10 juin; elle est rousse,* etc.

The grammatical abilities of average S2 students had also advanced. Average P7 students showed that they were aware of the importance of gender and of using the correct verb forms, but often did not know what the appropriate forms were. (For example, French students used an all-encompassing *luh* as a definite article, and often omitted verbs other than those in the first person, presumably because they were aware that the form should change but did not know the correct form.) In contrast, average S2 students distinguished between *le* and *la* and had some degree of accuracy in attribution. They could use the first and third persons of main verbs, usually accurately, although they had difficulty with more complex constructions, for example, in phrasal verb constructions which required this, following the main verb with an infinitive (e.g. *j'aime écoute la musique.*).

The confidence of the average S2 students also appeared to be somewhat higher than among the average P7 students. For example, some average S2 students attempted to make the conversation more personal by using phrases to express their own opinions and ideas - e.g. *c'est cool; c'est super, j'adore ...* - and thus going a little way beyond what was demanded by the task.

In terms of 'discourse skills', both the P7 and the S2 students were able to use 'coping strategies' in the language they were studying when they encountered difficulties: for example, one P7 student used *Comment dit [-on] ...* when she realised that she did not know the key word she needed in French; while average S2 students were able to ask the tester to repeat what s/he had said by using phrases such as *Pardon.*

LESS ABLE STUDENTS

The least able students in S2 appeared to be less able than those in a similar position in P7. It is important to bear in mind, however, as we noted earlier, that some of the S2 students we tested had not studied another language (or had studied a different language) at primary school, and therefore had little more experience of studying the language on which they were tested than their P7 counterparts.

All the students who fell into this category ('the weakest students') - whether in P7 or S2 - suffered from a very marked lack of confidence in their ability to communicate. It is not possible to determine from the tests whether students' confidence in communicating in another language is an artefact of their competence, or whether their confidence relates to other psychological factors. However, there were some students, at least, in the pilot study, who were very badly affected by 'nerves', to the extent that they were unable to produce words such as *oui* or *non*, and indeed could barely answer questions in English. Because of the low level of production, these students can only be classified as being among the 'weakest' although it may be that they are capable of much more than they demonstrated. It could be the case that these students suffered particularly from the absence of the kind of setting up and scaffolding which their teachers provide during class speaking activities.

The inability to respond because of 'nerves' was more common among S2 than P7 students, suggesting - as other studies have shown - that self-consciousness becomes more of an inhibiting factor as students progress into adolescence. One or two of the least able S2 students were unable to say very much at all. Even the weakest students in P7 were able to understand the basic questions which were put to them, and were able to produce some key words in response, usually pronounced in a way which would make them comprehensible to native speakers of the language. Among the S2 students who were able to answer the questions put to them, it was clear that some used strategies - such as hooking their comprehension of what was said to words which were the same or similar to their English counterparts - as a way of working out what was being said. In some cases, there was also evidence of a developing ability to conduct a conversation: for example, 1 student was able to

establish a sense of dialogue by judicious use of *... et toi?* to statements he made about himself, in order to draw the tester into conversation.

The tables overleaf indicates the range of performance in relation to:
- pronunciation and intonation
- vocabulary
- grammar
- understanding
- communicative skills
- discourse skills
- recall

for students in P7 and S2.

RANGE OF ORAL SKILLS - P7 AND S2

Oral skill	Excellent students can …	Average students can …	The weakest students can …
Pronunciation/ intonation	**in P7** … make a good attempt at pronouncing in a French or German way (especially key features such as, in French, 'r', nasal vowels); … in some cases, produce native-like intonation patterns. **additionally, in S2** … in some cases, produce spoken French/German close to native speaker standard in terms of pronunciation and intonation; … [in French] liaise word endings where appropriate.	**in P7/S2** … pronounce words clearly but not always consistently.	**in P7/S2** … pronounce most words sufficiently clearly for a native speaker to understand.

Vocabulary	in P7	in P7	in P7
	… recall a range of topics they have covered and many items of vocabulary within each, including appropriate questions and phrases- e.g. *j'aime la natation je bois le café j'ai mal à la tête j'ai un frère qui s'appelle ...*; … demonstrate the ability to draw on a wide range of vocabulary, phrases, etc. in response to general questions posed by assessors; … use different question words- e.g. *welche Farbe ist dein Hemd?[sic] quelle est ton adresse? qu'est-ce que tu aimes manger?*; … produce unusual adjectives- e.g. *déprimant*; ... use numbers and time phrases with confidence- e.g. *il est dix heures et demie.*	… recall some of the topics they have covered and key items of vocabulary in each, principally nouns- e.g. *la tête, les cheveux*; … use some set phrases- e.g. *Dans ma famille il y a 5 personnes*; … use numbers and colours appropriately.	… recall one or two words and phrases relating to some of the topics they have covered- e.g. *Ma famille il y a 5 personnes*; … demonstrate a basic knowledge of numbers.
	additionally, in S2 … demonstrate a very wide range of vocabulary and ability to link elements into continous conversational flow- e.g. *Ensuite je me lave ... je me maquille en arrivant au collège*; … use adverbs; … describe scenes or activities in some detail- e.g. *Le petit garçon achète une glace Il a un petit chien brun.*	**additionally, in S2** … demonstrate knowledge of key words and phrases needed for particular tasks; … describe scenes or activities- e.g. *elle écoute de la musique il fait du soleil*; … use more sophisticated set phrases than in P7- e.g. *comment ça s'écrit?*	**additionally in S2** … recall some appropriate vocabulary to describe a scene or activity.

Oral skill	Excellent students can …	Average students can …	The weakest students can …
Grammar	**in P7** … usually use correct gender, and recall nouns with appropriate article attached; … understand and use 1st and 3rd person verbs with confidence; … attempt to use different adjectival and verb endings- e.g. *sie traegt ein rotes Pullover[sic]*; … [in German] invert word order in certain contexts- e.g. *heute trage ich eine bunte Hose.* **additionally, in S2** … demonstrate excellent understanding of how gender operates, and a high level of accuracy in attributing gender; … use phrasal verbs- e.g. *J'aime manger le chocolat*; … use reflexive verbs; … use negative forms- e.g. *sie arbeitet nicht* *je n'ai pas d'argent*; … agree verb and adjective endings- e.g. *Ich spiele gern mit meinem Computer* *mes matières préférées sont …* *au centre sportif*; … use past tense- e.g. *Je suis allée en France avec l'école.*	**in P7** … show awareness of need to use articles with nouns, although article often indeterminate (e.g. in French *luh)* or incorrect; … demonstrate a degree of grammatical control; … use the first person with confidence, and show awareness of need to make changes for other persons. **additionally, in S2** … use main verbs for 1st and 3rd person in many cases; … use negative forms- e.g. *je n'ai pas de sœurs.*	
Understanding	**in P7** … understand almost anything said to them by tester, in the context of the exercise. **additionally, in S2** … demonstrate very high level of understanding of 'off the cuff' remarks made by testers; … work out meaning of difficult or unfamiliar questions- e.g. *Was liest du gern?* *Wo treffen wir uns?*	**in P7/S2** … understand most of what is said to them by testers, in the context of the exercise; … understand well when assessors slow down or prompt.	**in P7** … understand basic questions, particularly when these relate to person context (e.g. own pets, etc.). **additionally, in S2** … use English words *(hockey, football)* or words very similar to English words *(musique, cinéma)* commonly used in French as a way of understanding texts.

Oral skill	Excellent students can …	Average students can …	The weakest students can …
Communicative skills (this refers primarily to skills not specific to the language in question - qualities such as confidence, enthusiasm, pleasure in communicating, interest in others)	**in P7** … respond quickly and confidently; … enjoy demonstrating skills and attempt to extend what they know; … show empathy with speaker- e.g. by repeating *ja* while assessor speaks; … provide support for partner when partner in difficulty; … use humour in the foreign language- e.g. when speaking of kilts, *keine Unterhose!*	**in P7** … respond to set questions quickly and with some confidence; … respond to less predictable questions/ situations with some prompting.	**in P7** … respond tentatively to basic questions, often with considerable encouragement/ prompting; … demonstrate some ability to ask questions.
	additionally, in S2 … show a willingness to communicate complex events or ideas, taking risks with complicated linguistic structures in order to express themselves; … go beyond the demands of the task; … respond to 'off the cuff' comments (not necessarily task related) made by testers; … show a wider range of empathy strategies than in P7- e.g. *Oh ja!* *Ich auch* *Moi aussi j'aime bien ça*.	**additionally, in S2** … express own opinions and attempt to personalise conversation (particularly in dialogue task 3)- e.g. *il est pénible*; … express basic ideas clearly, although with some grammatical errors- e.g. *J'aime écoute la musique* *Le samedi aller au cinéma* [instead of *je vais*]; … show a willingness to communicate in spite of linguistic problems this can present.	**additionally, in S2** … in some cases, demonstrate simple strategies to draw the interlocutor into the discussion- e.g. *j'aime football - et toi?*

Discourse skills (these are skills which transcend the use of individual words or phrases - e.g. the ability to ask questions as well as answer them; recognising the limits of own linguistic capabilities and asking for help in the language, etc.)	in P7	in P7	in P7/S2
	… show that they are equally confident asking or answering questions; … deal with unexpected questions; … frequently reply using full sentences; … in some cases, use polite *(Sie/vous* form) with assesor- e.g. *Wie heissen Sie?*; … seek to provide extended answers to questions ostensibly requiring only yes/no answers; … link sentences with *und/et* (and use this as a way of gaining thinking time); … sometimes take the initiative in dialogue; …exhibit the ability to self-correct, on occasion.	… answer questions with a degree of confidence, and produce some basic questions, using set phrases on occasion- e.g. *Comment tu t'appelles?* *Quel âge as-tu?* *Où habites-tu?*; … sometimes vary sentence structure; … make use of certain 'coping' strategies- e.g. *Comment dit [-on] ...?* *Pardon?* (with French accent)	… use some set phrases to communicate, albeit with errors in some cases- e.g. *J'habite à ...* *je nager* *J'aime cinéma* *je préfère restaurant*

Oral skill	Excellent students can …	Average students can …	The weakest students can …
Discourse skills cont. (these are skills which transcend the use of individual words or phrases - e.g. the ability to ask questions as well as answer them; recognising the limits of own linguistic capabilities and asking for help in the language, etc.)	**additionally, in S2** … link sentences appropriately; … ask follow up questions of testers; … answer questions using a different structure to that used in the question- e.g. *Qu'est-ce que tu détestes manger?* *Je n'aime pas les pizzas*; … make use of structuring devices as a means of support- e.g. *A 8 heures, je me lève* *A 9 heures, je vais au lycée* *Lundi je vais à la classe de danse* *Mardi je reste à la maison*; … use appropriate polite phrases- e.g. *je voudrais*; … recognise when conversation has taken an unexpected turn, and adapt response to suit; … self-correct (particularly noun gender, when thinking takes a different turn mid-sentence and a new noun is substituted for that which was originally intended); … express own opinions, using appropriate phrases- e.g. *ich finde ihn gut* *Mathe und Hauswirtschaft gefallen mir nicht.*	**additionally, in S2** … in some cases make interjections- e.g. *C'est cool. C'est super*; … produce full sentences on occasion; … sometimes improvise questions- e.g. *tu vas avec moi?*	
Recall (this relates to pupil recounting of a story narrated by the tester)	**in P7 only** … recall the beginning and end of the story accurately after one reading; … recall the middle and certain details of a story after the second reading; … make intelligent guesses about unknown vocabulary.	**in P7 only** … recall basic elements of the story; … work out the meaning of elements of the story which involved English vocabulary (i.e. English words commonly used in French); … establish details with some repetition and prompting; … show improved grasp of text when sentences read one by one.	**in P7 only** … recall meaning of individual vocabulary items; … work out some elements of text with considerable support/ prompting.

CHAPTER 8
FINDINGS ON PUPIL ACHIEVEMENT: WRITING

This chapter describes the findings on pupil performance in the productive skill of Writing. Reading and Writing have been analysed at S2 only, since neither of these skill areas was extensively assessed at P7. The data on which the section draws derive from Test A, Tasks 6-9 and part of Task 4 (part) of the S2 Tests (see Appendix 5 for details).

A variety of different writing skills was assessed via the different tasks set as part of S2 Test A - Reading and Writing. These can be divided into different types, each of which reflects several of the strands outlined in the 5-14 Guidelines for Modern Languages:

- Copying
- Writing from memory
- Continuous writing

Within each of the strands, a description is given of the scores achieved in each related task. In the case of 'writing from memory' this is followed by an analysis of full competence as compared to non-response and levels of partial competence. For the continuous writing task analysis is by the following criteria descriptors :

- Volume
- Task coverage
- Linguistic range
- Accuracy

Scores based on the criteria descriptors are then analysed.

In each case, achievement in French is studied alongside achievement in German. This is possible since the majority of tasks from each test are identical in content and format. A breakdown of the overall total scores in writing are then presented by gender and level of ability.

At the end of the chapter points made about the range of achievement in writing are summarised in tabular form.

S2 WRITING

At the outset it was hypothesised that performance in writing may be quite limited since, generally, only a restricted amount of S1 and S2 teaching time has been dedicated to the learning of writing skills. Indeed analysis of the school contextual data revealed that in six of the sample schools (three French, three German) writing skills had not been routinely assessed in S1. Of those schools, four (two French, two German) still had not assessed writing by the end of S2.

COPYING

Two different tasks included an element of copying:

- Task 4, which involved the identification and copying of items of vocabulary (two single words and a phrase for French, and two single words for German) from a continuous text;
- Task 8 (Part 1), which involved text completion – six blank spaces to be filled using an option list containing ten items of vocabulary.

Both of these tasks obviously included an element of comprehension work. This aspect has already been analysed in the Chapter 6 outlining student achievement in the reading skill of vocabulary identification.

Task 4
French
Breakdown of results for Task 4 - per item to be identified and copied:

Word/phrase	% Correct identification/ copying	% Partial identification/ correct copying	% Incorrect identification/ correct copying	% Correct copying any word phrase	% No response
Allez les voir	15.7	21.3	25.5	62.5	6.8
pauvre	47.7	0.9	34.9	83.5	8.9
naufrage	65.1	1.3	16.6	83	8.1

German
Breakdown of results for Task 4 - per item to be identified and copied:

Word/phrase	% Correct identification/ copying	% Partial identification/ correct copying	% Incorrect identification/ correct copying	% Correct copying any word phrase	% No response
traurig	15.5	2.5	66.8	84.8	8.4
tragisch	26.1	4.6	44.9	75.6	18.1

Task 8 (Part 1)
French
Breakdown of results for Task 8 (Part 1) - per item to be identified and copied:

Word/phrase	% Correct identification/ copying	% Partial identification/ correct copying	% Incorrect identification/ correct copying	% Correct copying any word phrase	% No response
m'appelle	86.8	N/A	5.1	91.9	1.3
ai	69.4	N/A	24.5	93.9	2.6
il	57.4	N/A	36.3	93.7	6
blonds	74	N/A	13.6	87.6	4.7
petite	48.1	N/A	40.4	88.5	3.8
est	29.8	N/A	63.3	93.1	3.8

German

Breakdown of results for Task 8 (Part 1) - per item to be identified and copied:

| Word/phrase	% Correct identification/ copying	% Partial identification/ correct copying	% Incorrect identification/ correct copying	% Correct copying any word phrase	% No response
heiße	83.2	N/A	11.3	94.5	1.3
habe	89.9	N/A	7.5	97.4	0.4
er	79.4	N/A	13	92.4	3.8
blonde	87.8	N/A	7.1	94.9	2.9
kleine	40.8	N/A	48	88.8	3.4
ist	66	N/A	26.4	92.4	5

Figure 8a: Breakdown of numbers of correct responses for Task 8 (Part 1) - percentage of students scoring 0-6:
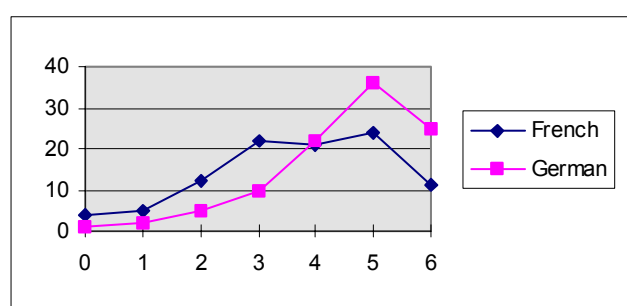
It is clear that, while a varying number of students was able correctly to identify **and** copy items of vocabulary (from 15.7% for the phrase *Allez les voir* in French and 15.5% for the word *traurig* in German, to 86.8% for the very familiar *m'appelle* in French and 89.9% for the equally familiar *habe* in German), the majority of students (well over 80 or 90% in almost all cases) was able successfully to copy items (whether correctly identified or not) from a written text.

However students would seem to experience more difficulty in copying phrases (Task 4 French table) than in copying single items of vocabulary. Students would also appear to be slightly more successful at copying single words from a list of single words (Task 8 (Part 1) tables) than from a continuous text (Task 4 tables). It is interesting to note that the great majority of students (over 90% in every case for French, and every case but one for German) were inclined to respond to this type of question, rather than to leave it blank.

If we compare overall performance in the skill of copying in French with that in German we find that on the one hand, performance in French was slightly higher where the words had to be identified and copied from a continuous text in Task 4 (an average of 42.8% of students correctly identified and copied each French item, as compared to an average of 20.8% for each German item). As explained in Chapter 6, outlining achievement in the reading skill of vocabulary identification, this may be due to the fact that the question format in the German test allowed for some confusion between the 2 quite similar words required (*traurig* and *tragisch*). On average, just under a quarter of the German sample students were caught out by this potential pitfall. On the other hand, in the task requiring the identification and copying of words from a list of options (Task 8 Part 1), performance in German was slightly higher than that in French (an average of 74.5% correctly identified and copied each German item, as compared to an average of 60.9% for each French item).

WRITING FROM MEMORY

Four different tasks included an element of writing words or single sentences/phrases from memory:
- Task 6 - single words, grid-completion of school timetable
- Task 7 (Part 1) - words with articles, listing of classroom objects
- Task 7 (Part 2) - sentences/phrases, responding to classroom questions regarding date, time, weather.
- Task 8 (Part 2) - single words, gap-filling, choosing an appropriate word to make sense of an incomplete text.

Here it is important to interpret the findings alongside the contextual information regarding topics covered by the sample students, since a relatively poor performance in a particular domain may be explained by ignorance of that topic, or lack of recent practice in it.

WORD LEVEL

The aim of Task 6, Task 7 (Part 1) and Task 8 (Part 2) was to assess students' ability to recall and accurately to reproduce in written form familiar words (nouns, articles, verbs, adjectives), prompted by clues in English, by visual stimuli or by incomplete sentences in the target language. Tasks 6 and 8 (Part 2) also include an element of comprehension work – already analysed Chapter 6 outlining student achievement in the reading skill of central message extraction.

Tasks 6 and 8 (Part 2)

The analysis of results for Tasks 6 and 8 (Part 2) is carried out first, since the language to be recalled is at the single word level only. Levels of partial competence are also discernible, since this task revealed students' ability to recall the appropriate word without the correct spelling, or to recall an inappropriate word with the correct spelling.

Task 6
French
Breakdown of results from Task 6 - per item to be recalled and written:

Word	% Correct word/ correct spelling	% Correct word/ incorrect spelling	% Incorrect word/ correct spelling	% Correct recall any word/ phrase	% Response in English	% No response
Français	20	49.9	0	20	18.7	3.8
Histoire	24.3	45.2	2.5	26.8	19.1	3.4
Informatique	5.5	8	5.8	11.3	42.1	10.2
Anglais	31.5	33.1	0.4	31.9	14	8.5
Éducation physique	8.2	17	1.7	9.9	33.2	33.6
Biologie	34.5	23.3	2.5	37	27.7	4.7

German

Breakdown of results from Task 6 - per item to be recalled and written:

Word	% Correct word/ correct spelling	% Correct word/ incorrect spelling	% Incorrect word/ correct spelling	% Correct recall any word/ phrase	% Response in English	% No response
Deutsch	50	32.4	0.4	50.4	6.7	8.4
Geschichte	11.3	19.6	0.4	11.7	28.2	22.7
Informatik	8.8	2.9	0	8.8	43.6	19.3
Englisch	52.1	7.8	0.4	52.5	22.7	10.5
Sport	32	2.8	0	32	32.4	29.4
Biologie	36.1	24.7	0	36.1	23.9	12.2

Figure 8b: Breakdown of numbers of correct responses for Task 6 - percentage of students scoring 0-6:
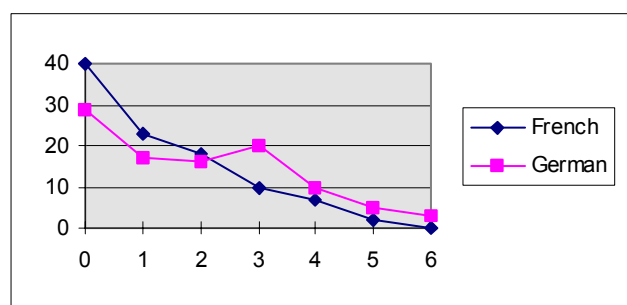
Full competence/Non-response

No student of French achieved the top score in this task, while only a very small number of students of German did so (2.9%). In fact 40.4% of French students and a slightly lower number of German students (28.6%) did not score at all. However the above tables do demonstrate that the recall of each word for a school subject was successfully completed by, on average, a fifth of the French students, and almost a third of the German students.

The exact figure for French varied from as much as just over a third for words closely resembling their English equivalents (*Biologie*) or very familiar words (*Anglais*), to 5.5% for words which were quite far removed from their English equivalent (*Informatique*), or compound words (*Éducation physique*). In the cases where lower scores were recorded, there was a more marked tendency either to give an answer in English or to make no response.

In the case of German the exact figure varied from as much as a half or just over a half for words closely resembling their English equivalents (*Englisch*) or very familiar words (*Deutsch*), to around 10% for words which were quite far removed from their English equivalent (*Geschichte*, *Informatik*). Again in the cases where a lower score was recorded there was a more marked tendency either to give an answer in English or to make no response. There was also a tendency for the word to be given in English either when the German word closely resembled the English word (*Englisch*, *Biologie*) or when students felt unable to translate the compound English word (*Physical Education*). This latter case also resulted in the highest level of non-response: 29.4%.

Partial competence

Students of French and German were often able to recall the word and its phonetic value, but experienced more difficulty in remembering the exact spelling. For example, almost a quarter of the French students omitted the accent in *Français*, while a tenth wrote *Anglais* either with the article '*l*' or beginning with the letter '*e*'. Over 10% of the German students omitted the '*s*' in *Deutsch*, while

just over 5% omitted the second 'h' in *Geschichte*. In a small number of cases the capital letter was omitted. In every case for both languages a very large variety of spellings was produced.

A much smaller number of French students (no more than 5%, and on average closer to 2%) was able accurately to recall other French words which were not exactly appropriate to the question. Very few German students (on average less than 0.5%) did so.

While, as outlined in Chapter 6, an average of 85.1% of French students and 78.2% of German students could correctly position an appropriate recognisable word, many fewer could recall that word accurately: over 80% of French students and 60% of German students scored no more than 2 out of 6, with under 20% for French and 40% for German scoring 3 or above. These relatively low results may be explained by lack of recent practice in the domain of language being assessed. Indeed, analysis of the school contextual data shows that only two of the French sample schools and one of the German sample schools had covered this domain in S2, while seven French schools and eight German schools had not studied it since S1, and one school for each language had not studied it at all.

Task 8 (Part 2)
French
Breakdown of results from Task 8 (Part 2) - per item to be recalled and written:

Word	% Appropriate word/correct spelling	% Inappropriate word/correct spelling	% Correct recall any word/phrase	% No response
mon	46.8	33.5	80.3	11.5
est	26.8	42.7	69.5	17
travaille/verb	13.3	33.4	46.7	41.7
adjective (f.)	8.5	40.3	48.8	43.4
un	51.5	28.3	79.8	14.5
animal (m.)	64.6	8.2	72.8	14.9
as/verb	23	37.2	60.2	26.8
sont	1.3	48.7	50	33.2

German
Breakdown of results from Task 8 (Part 2) - per item to be recalled and written:

Word	% Appropriate word/correct spelling	% Inappropriate word/correct spelling	% Correct recall any word/phrase	% No response
mein	34	32.3	66.3	14.3
ist	45.8	24.3	70.1	24.4
arbeitet/verb	30.2	14	44.2	40.8
adjective (n.)	2.1	8.6	10.7	16.4
ein/number	57.2	2.8	60	10.9
animal (f.)	40.3	1.6	41.9	13.9
hast/verb	60.5	11.9	72.4	21.8
sind	10.5	4.5	15	17.6

Figure 8c: Breakdown of numbers of correct responses for Task 8 (Part 2) - percentage of students scoring 0-8:

Full competence

Only one student of French and no students of German scored the top score of 8 in this task. In fact 17.9% of French students and a slightly lower number of German students (13.4%) did not score at all. However, on average, an appropriate word with the correct spelling was accurately recalled by 29.3% of French students and 35% of German students.

The exact figure for French varied from as much as 64.6% for the accurate recall of a masculine animal, to as little as 1.3% for the difficult third person plural form of the irregular verb *être*, or 8.5% for an adjective in the feminine form. For German it varied from 60.5% for the accurate recall of the second person singular of *haben* to as little as 2.1% for the first part of a compound noun with *Haus*, such as *Einfamilien/Reihen/ Doppel* which a number of ingenious students used in place of an adjective requiring agreement with the neuter noun (no student was able to produce such a form).

These more difficult items recording lower scores were also characterised by a higher level of non-response. On average just over a quarter of French students and just over a fifth of German students left each item blank.

Partial competence

In some cases students were able to recall an appropriate word, but with the wrong spelling. The system used to code the data, rendered difficult the production of exact figures for the number of students who did so, but spellings which accurately reflected the correct phonetic value of an appropriate word were relatively frequent in French (*ai* instead of *est*), as were spellings which reflected a mispronunciation of an appropriate word (*a* instead of *est*). For German, the main problem was getting the correct ending on possessive adjectives such as *mein* and the number *ein*, which also appeared with a range of different endings (*meine/eine, meinen/einen, meinem/einem)*. Such errors would seem to indicate interference from other case endings.

Students were also able accurately to recall an inappropriate word: on average, per item, 34% of French students and 12.5% of German students did so. In the case of French this was a higher figure than those who accurately recalled an appropriate word.

Task 7 (Part 1)

We will now analyse the results from Task 7 (Part 1) which assesses students' ability to recall items of vocabulary along with the appropriate article. Again levels of partial competence were discernible since students may accurately recall an appropriate word but be unable to recall the correct article. They may also recall an appropriate word, but be unable to recall the exact spelling.
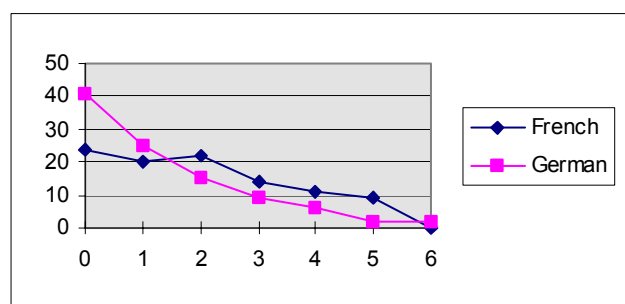
French
Breakdown of results from Task 7 (Part 1):

NO. OF ITEMS	% Fully correct	% Items added	% French items	% Correct spelling	% Correct article
0	23.4	4.3	7.2	17	11.9
1	20.4	3.8	7.7	12.8	14.9
2	21.7	6	7.7	15.3	15.7
3	14	5.1	5.5	13.6	22.1
4	11.1	10.2	14.9	19.6	20
5	8.5	7.7	17.9	18.3	10.2
6	0	61.7	38.7	3.4	5.1
7	0	1.3	0.4	0	0

German
Breakdown of results from Task 7 (Part 1):

NO. OF ITEMS	% Fully correct	% Items added	% German items	% Correct spelling	% Correct article
0	40.8	10.9	14.3	28.6	27.3
1	24.8	3.8	12.2	23.9	20.2
2	14.7	10.1	13	12.6	20.2
3	9.2	13.9	18.1	15.1	14.7
4	6.7	14.7	11.3	8.4	11.3
5	1.7	11.3	11.8	8	2.5
6	1.7	32.8	17.2	2.9	3.4
7	0	1.7	1.7	0.4	0.4
9	0	0.4	0	0	0
10	0	0.4	0.4	0	0

Figure 8d: Breakdown of results from Task 7 (Part 1) - percentage of students scoring 0-6:

Full competence/Non-response
In terms of full competence for French, no-one achieved the top score of 6 items listed with the correct spelling and correct article. Just below 10% scored 5/6, while just under a quarter scored zero. For German only a very small number achieved the top score of 6 items listed with the correct spelling and correct article (1.7%), with the same small number scoring 5/6, while well over a third scored zero. Again this may be due to lack of recent practice in the domain of language being tested, since analysis of the school contextual data reveals that classroom objects had not been studied in any of the sample schools since S1. Only 4.3% of students did not attempt the task at all in the case of French, compared to just over a tenth in the case of German (10.9%).

Partial competence

However levels of partial competence were again discernible, this time in four main areas:

- Participation in task: the majority of students were willing to attempt the task. Although no more than two thirds of French students and only a third of German students added the full six items as requested, only a small proportion failed to respond. In fact around 80% of French students and 60% of German students added four items or above, a small percentage of students actually exceeding the required number.
- Ability to add recognisable French or German words: of the French items added the majority was recognisably French, while of the German items many were recognisably German. In fact almost three quarters of students added four or more items that were recognisably French. Slightly fewer German students did so: under half added four or more recognisably German items.
- Ability to spell those words correctly irrespective of the attachment of the correct article. Of those recognisable words added, the number that was recalled accurately varied enormously, with French students outperforming German students overall. In the French sample the range of accurately recalled words was spread quite evenly between 1-5 correctly spelled items. 17% of students failed to score, while a much lower number (3.4%) managed correctly to spell all six items. In the German sample, while a similar low number (2.9%) managed correctly to spell six items, a much larger number failed to spell any words correctly (28.6%). Also the range clustered towards the lower end of the scale, with over half of the remaining students only able accurately to recall one, two or three words.
- Ability to attach the appropriate article, irrespective of the perfect spelling of the word. Again this varied to a great extent, with French students achieving at a slightly higher level than German students. In the French sample 11.9% of students failed to attach a correct article to the words added to the list. However the largest proportion of students (42.1%) recalled the correct article for three or four items. In the German sample more than double the number of students failed to recall any correct articles (27.3%), while the largest proportion of students (40.4%) recalled the correct article for one or two items only.

SENTENCE/PHRASE LEVEL

The aim of Task 7 (Part 2) was to assess students' ability to recall, in written form, familiar sentences or phrases in order accurately to convey a specified message based on visual stimuli. The task did not specify that full sentences should be used. Levels of partial competence were also discernible, since this task also revealed students' ability either to convey the specified message with some errors, or accurately to convey an incorrect message.

French

Breakdown of results in Task 7 (Part 2) - per question asked:

Question	% Correct sentence/ message	% Correct phrase/ message	% Correct phrase or sentence/ incorrect message	% Incorrect phrase or sentence/ correct message	% No response
Date	3	24.2	1.3	33.6	8.1
Time	1.7	2.1	0.9	38.7	20.9
Weather	23	N/A	5.5	26	22.1

German

Breakdown of results in Task 7 (Part 2) - per question asked:

Question	% Correct sentence/ message	% Correct phrase/ message	% Correct phrase or sentence/ incorrect message	% Incorrect phrase or sentence/ correct message	% No response
Date	2.1	22.7	2.9	40.3	7.6
Time	16.8	19.4	8.8	10.5	10.9
Weather	15.6	N/A	4.6	34	10.5

Figure 8e: Breakdown of numbers of correct responses - percentage of students scoring 0-3:
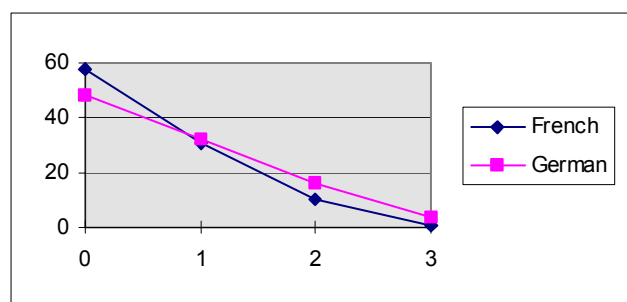
Full competence

It is clear that in terms of full competence some students of each language were able to write full sentences from memory in order accurately to convey a specified message - on average 9.2% of French students and 11.5% of German students could write a complete sentence in response to each prompt question - while a larger number could retrieve only phrases in order to do so.

Sentences:
- *C'est le 15 juin; Il est dix heures vingt; il pleut; il fait mauvais.*
- *Es ist der 15 Juni; Es ist Montag den 15 Juni; Es ist Dienstag den fünfzehnten Juni; Es ist 20 nach 10; Es ist zwanzig nach zehn; Es regnet; Es ist schlechtes Wetter.*

Phrases:
- *15 juin; le 15 juin; quinze juin; le quinze juin; dix heures vingt.*
- *15 Juni; den 15 Juni; Fünfzehn Juni; den fünfzehnten Juni; Montag den fünfzehnten Juni; 20 nach 10; zwanzig nach zehn.*

Partial competence

In terms of partial competence, a smaller number of students could retrieve phrases or sentences which, while accurate, conveyed a message not specified by the task.
- *Cinq juin; le seize juin; le quinze juillet; dix heures; Il fait beau; Il fait froid; Il fait chaud; Il fait gris; Il neige en hiver.*
- *15 Juli; Mittwoch den 3 Juni; Dienstag den 8 Juni; zwanzig vor zehn; es ist zehn Uhr; zehn minuten vor vier; Gut!!; Es ist kalt; Das Wetter ist bewölkt.*

In addition another, generally larger group of students of each language could convey the correct message, but with some errors.
- *Le quize juin; c'est la quanze juin; Dix heure vingt; dix heures et vingt; il fait pleut; il pluet.*
- *15 dem Juni; Funfzehn Juni; Es it fünfsehnten Juni; es ist zehn uhr zwanzig; Es ist zwanzig nach zehn; zehn ohr zwanzig; Es ist regnet; Das ist regnet; Das wetter ist regnet; Das Wetter ist Regen und Kalt.*

If we compare overall achievement in the skill of recalling familiar sentences in French with that in German, it can be seen that achievement in German was higher than that in French, with over double the number of students scoring full marks, and just under double getting 2 of the questions correct. Also, while just under half of the students of German scored zero, well over half of the students of French failed to score. Moreover, more students of French were inclined to leave their response blank rather than guess (an average of 17% as compared to an average of 9.6% for German).

However, in neither language were the results particularly positive: on average less than a quarter of students of French, and only a quarter of students of German were able to recall a sentence or phrase to communicate a specified message regarding, date, time or weather. This may be due to the fact that the domains of language involved were not actually as recently familiar to the majority of students as had been supposed. Indeed analysis of the school contextual data reveals information regarding the coverage of these domains:

- Date - studied in S1 in all of the sample schools for French, and in eight of the ten sample schools for German. In only two of the sample schools for German had this topic been covered in S2.
- Time - covered in S1 in 17 of the sample schools (nine French, eight German), with only three schools studying it in S2 (one French, two German).
- Weather - ten of the schools actually having covered this domain in S2 (six French, four German). However, while five sample schools had covered this topic in S1 (three French, two German), five had not covered it at all in either S1 or S2 (one French, four German). This perhaps explains the slightly lower results for German in this particular domain.

Therefore the less than encouraging performance may be explained by the fact that, in the majority of sample schools, date and time had not been overtly studied since S1, while the weather, although studied by half of the schools in S2 had been studied in S1 by a quarter of the schools and not at all by a further quarter.

CONTINUOUS WRITING

This writing skill was assessed in Test A, Task 9, which involved the writing of a short paragraph describing or narrating normal weekend activities, without the aid of a dictionary and under strict time constraints (approximately 15 minutes were allocated to the task). Students were judged on their performance in relation to four different criteria: volume, task coverage, range and accuracy, each scored on a scale 1 to 3 (with 3 as the highest mark).

CRITERIA DESCRIPTORS

The criteria definitions were as follows:

VOLUME:
This should include only those words that are recognisably, although not necessarily correct, French, and disregard English words, i.e. the volume of French words attempted.

0	Blank	Pupil writes nothing at all, or a couple of words only. To be scored 0 in each of the other criteria.
E	English	Pupil writes essay totally in English. To be scored 0 in each of the other criteria.
F/G	English and FL	Pupil writes essay in a mixture of English and the FL. To be scored 0 in each of the other criteria.
1	Low	Pupil writes 1-5 lines.
2	Medium	Pupil writes 6-10 lines.
3	High	Pupil writes 11 lines and above.

TASK COVERAGE
Account should be taken of the attempt made to incorporate three elements of the task set into the work: message format, reference to the time slot (weekend) and mention of several activities, possibly based on the picture cues.

1	Low	Covers very few or no aspects of the task set. Little or no attempt is made to tailor the material to the task set, e.g. covers one element only.
2	Medium	Covers some aspects of the task set, but may disregard others. Inconsistent in its attempt to tailor the material to the task set. May cover two elements of the task.
3	High	Covers several different aspects of the task set. An attempt is made to tailor the material written to the task set. Covers all three elements of the task as outlined above

LINGUISTIC RANGE
Again this should include only those words that are recognisably, although not necessarily correct, French, and disregard English words, i.e. the range of French words attempted.

1	Low	Demonstrates a repetitive use of language, in a limited number of linguistic categories.
2	Medium	Demonstrates some variety of linguistic categories: perhaps three or four examples in two categories, e.g. verbs and nouns. May also contain some repetition.
3	High	Demonstrates a varied use of language, in several different linguistic categories - three or more categories. Also demonstrates some degree of connectivity and sequencing.

ACCURACY:
This should include only those words that are recognisably French and disregard English words. Account should be taken only of the language attempted, and disregard the volume or range of that language.

1	Low	Demonstrates very little or no control of the grammatical features of the language being attempted, to such an extent that comprehension is impeded.
2	Medium	Demonstrates some, albeit inconsistent, control of the language being attempted. May contain some major errors, but communicates the desired message.
3	High	Demonstrates generally high, although not necessarily perfect, control of the language being attempted. Message clearly conveyed.

SKILL SCORES

An outline of performance in each of the criteria for each language is given below, followed by a commentary on performance in each criterion, and some exemplars of high performances (French only).

Scores	FRENCH				German			
	V* %	TC %	R %	A %	V %	TC %	R %	A %
0	19.6	19.6	19.6	19.6	16.8	16.8	16.8	16.8
1	46	37	42.6	34.9	35.3	29.8	9.7	6.7
2	26	37.9	28.9	39.1	23.9	33.2	44.5	52.9
3	8.5	5.5	8.9	6.4	23.9	20.2	29	23.5

*V = Volume (incl E, F/G)	TC = Task Coverage	R = RANGE	A = ACCURACY

VOLUME/NON-SCORERS

Figure 8f: Breakdown of volume scores - percentage of students scoring 0-3:
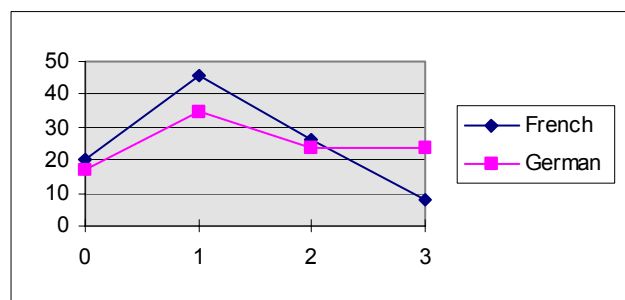
Students were encouraged to write as much as possible, with no specific guidance given as to the length of essay required. However, there were three reasons for scoring zero:
- Not attempting the task. This occurred in approximately 8% of cases in French and 4% of cases in German.
- Writing the essay entirely in English. This occurred in only 1% of cases for each language, sometimes with the pupil making the written comment, *I can't do this in French/German.*

- Writing a sufficient amount of the essay in English/another foreign language mixed with the target language to impede overall comprehension. Again this occurred in a similar number of cases for each language, although this time for a very slightly higher number of German cases (12%) than French (11%).
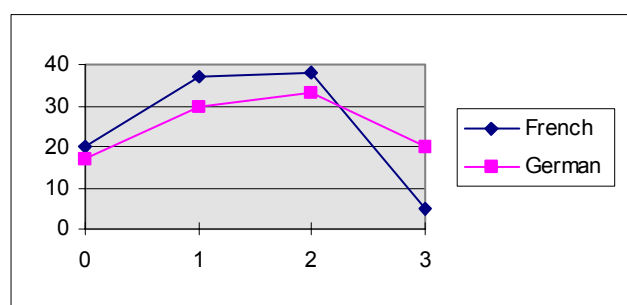
French
Just over a quarter of the students were able to write at least six lines of text, with just under 10% of all students writing more than 11 lines of text. However just under half of the students could write no more than five lines of text. The total number of students who sat the writing task in French was 235. Of those who sat the task 46 were 'non-scorers' – 19.6% of the total number.

German
Approximately half of the students were able to write at least six lines of text, with half of that number (23.9% of the total number) writing more than 11 lines. In German the total number sitting the writing task was 238. Of those who sat the task 40 were 'non-scorers' – 16.8% of the total number.

TASK COVERAGE

Figure 8g: Breakdown of task coverage scores - percentage of students scoring 0-3:

The task was as follows:
You want to describe what you do at the weekend in a message to a class in a French/German school. Write as many sentences as you can.
The completed task would thus comprise of 3 components:
- *Use of a message format.*
- *Inclusion of references to the weekend.*
- *Mention of weekend activities.*

French
Just under half of the students (43.4%) managed to cover 2 or more aspects of the task set. However, only a small minority of the overall number (5.5%) covered all three aspects of the task. In the main students failed to notice that the text was to be written as a message, while some either failed to include references to the weekend in their essay, or alluded to other days of the week.

German
Over half of the students (53.4%) managed to cover at least two aspects of the task set, with a fifth of the overall number (20.2%) covering all three aspects of the task. Again either failure to use a message format and/or to allude to the weekend were the main factors which prevented students from being awarded the top score in this category.

RANGE

Figure 8h: Breakdown of range scores - percentage of students scoring 0-3:
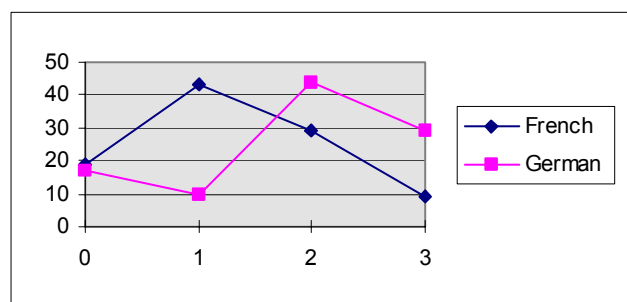
French
Over a third of the students (37.8%) were able to demonstrate at least some level of variety in their use of the target language, using a range of words and phrases which they had encountered in class (Levels 2 and 3). However of these, just 8.9% of the total number were able to demonstrate use of a wide range of target language, including an element of sequencing via the use of prepositions, connectors and time expressions (Level 3). Well over a third of students (42.6%) could only demonstrate use of a limited range of linguistic categories in the target language (Level 1).

German
Almost three quarters of students (73.5%) were able to demonstrate at least some variety in their use of the target language (Levels 2 or 3). Of these, almost a third (29%) of the overall number of students were able to link their sentences with simple conjunctions or more complex structures (Level 3). Indeed only 9.7% used a minimum of linguistic categories (Level 1). Many of the texts included some time expressions and prepositions, thus adding to the variety of language being used, although the language itself was not always accurate.

EXAMPLES

Level 1 students were able to produce fewer than four examples in, on average, only two linguistic categories. Their work was thus characterised by a great deal of repetition, and was sometimes devoid of verbs:
Vendredi: école, copains - repeated for different weekdays;
Bonjure picen, le basket, le center commercial, babyfoot, le music;
… et au café et coupins et au la lit.

Level 2 students could usually produce at least four examples from two or three linguistic categories, normally including verbs. However events tended to be listed, with no real sense of sequencing present – beyond listing of days of the week:
J'aime le chocolate et j'adore le disco;
Je visite ma grand-mere. Je travaille dans un office (bureau). Je regarde la téle. Je écoute la musique;
Samedi – je joue au foot et je joue au ball basket.

Level 3 students could produce a variety of examples in more than three linguistic categories. In addition their work tended to include some notion of sequencing via the use of connectors and more complicated time expressions and some expression of feelings and opinions:
Ensuite je téléphone mes copines et nous allons au cinéma;
Le weekend,quand il fait beau, je joue au tennis;
L'apres-midi je fait mes devoirs. Bof!

ACCURACY

Figure 8i: Breakdown of accuracy scores - percentage of students scoring 0-3:
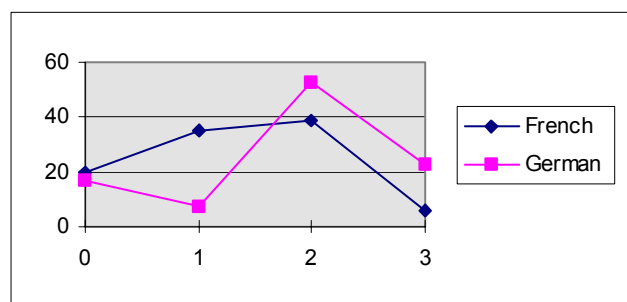
French

In almost half of the scripts (45.5%), students were able to convey the desired message to a greater or lesser extent (Levels 2 or 3). However the majority of that number demonstrated inconsistent grammatical control still containing major errors (39.1% of the overall number – Level 2), while only 6.4% of the students demonstrated high, although not perfect, control of the grammatical structures of the language (Level 3). In fact in over a third of cases (34.9%) the grammatical control of the language attempted was so limited as to impede comprehension (Level 1).

German

In a significantly higher proportion - over three quarters - of the scripts (76.4%), students were able to convey the desired message using words and phrases which they knew reasonably well (Levels 2 or 3). Given that this task was completed under time constraints without the use of a dictionary, it is not surprising that most students were careful to write what they knew, rather than use unknown language which would lead to mistakes being made. As a result, many of the scripts, whether short or long, were very accurate, with almost a quarter (23.5%) displaying a consistently high control of the grammatical structures of the target language (Level 3). Only 6.7% of students demonstrated minimal control of the language attempted (Level 1).

EXAMPLES

Level 1 students demonstrated very limited control of the grammatical features of the language being attempted to such an extent that comprehension was impeded:
Je me lave en ville á des onze heur. Il cenéma + le nation;
Mon weekend j'ai go un natation un cinema un Paisley;
Je jous un animals le jarden;
Fer du weekend alle alla piceine a au Ju a ou rugby.

Level 2 students demonstrated some, albeit inconsistent control of the language being attempted: verb endings, gender, spelling, accents, capital letters, word order. While sometimes containing major errors, their work nevertheless communicated the desired message:
J'ai va a la thêatre
J'adore cinema [...] J'adore un centre-ville;
Moi et mon pere regarde la tele.
Je manger la petit dejuner.

Level 3 students demonstrated generally high, although not necessarily perfect control of the language being attempted. They tended to make fewer errors in verb forms. Their message was clearly conveyed.
Le dimanche matin je vais à la piscine avec mes amis;
Samedi matin je me lève vers dix heures;
Mes copines rentrent avec moi.

GENERAL COMMENTS ON CONTINUOUS WRITING

Figure 8j: Breakdown of overall Task 9 scores - percentage of students scoring 0-12 (scores of 1, 2 and 3 were not possible):
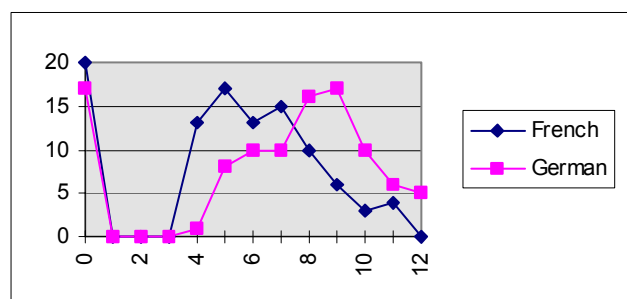
In general grades for French were lower than for German, with fewer students gaining levels 2 or 3 in each of the 4 criteria, in particular for Range and Accuracy. Also, while no student of French gained a top score of 12 over the 4 criteria, 11 students of German achieved this (almost 5%). The latter figure is impressive, particularly since some of the students had not been assessed in writing before – this was the case for two of the French sample schools and two of the German sample schools.

Indeed the range and level of German which was produced in this task was very encouraging. Most of the students, whether they wrote for two pages or for less than five lines, were able to produce a text in the target language which expressed, to a reasonable degree, what they wanted to express. Some of the texts demonstrated a high level of linguistic range and/or accuracy. The main errors appear to be those involving word order, but it was encouraging to note that many of the students were attempting to add time expressions or conjunctions to their texts in order to produce an essay which was interesting and lively for the reader.

Clearly in French the picture was slightly less positive, since fewer students of French could produce more than five lines of text, and the range and accuracy were of a lower level than that achieved by the German students. However we must remember that this non-guided writing task goes far beyond Level E of the 5-14 strand for Continuous Writing which states that pupils should 'Write a few simple sentences with support, guidance and reference materials if required, using the correct written form with increasing consistency'.

Nevertheless very good performances were still to be found in both languages.
These can be found in Appendix 6.

OVERALL WRITING SCORES

Figure 8k: Tasks 6-9 – percentage of students scoring 0-41:
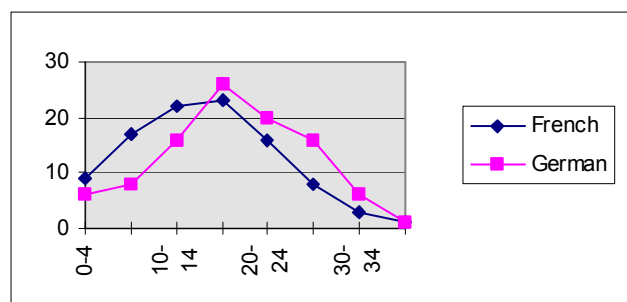
In terms of overall performance in writing at S2, it would appear that students of German achieved slightly better results than students of French. Although the points at either end of the scale were very similar for each language with regard to average score and percentage of students achieving that score, a higher percentage of German students scored a higher top score. In particular German students achieved higher scores in the tasks involving continuous writing. In other areas achievement in each language was of a very similar level: in the skills of copying and writing words, phrases and sentences from memory.

GENDER DIFFERENCES

In terms of the levels of performance of each gender, analysis would seem to suggest that in both the French and German writing assessments girls outperformed boys in Tasks 8 and 9. The level of achievement of each gender in the French sample was virtually identical in Tasks 6 and 7 (Part 2), with girls doing better in Task 7 (Part 1). In the German sample performance was virtually identical in Task 7 (Part 1) only, with girls doing better in Tasks 6 and 7 (Part 2). All of the latter tasks involved writing from memory at the word/phrase/sentence level. Thus, overall, girls performed slightly better than boys. This is outlined in the boxplots below. However it is not clear at this stage whether or not theses differences are an artefact of the sample since, as was outlined in Chapter 2, the samples were slightly skewed in terms of gender.
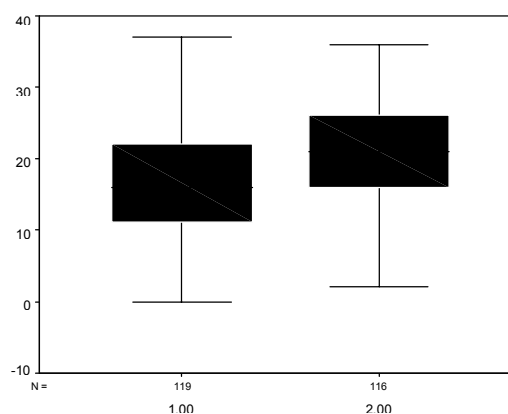
Figure 8l: French Writing Total Gender Difference (out of 41)
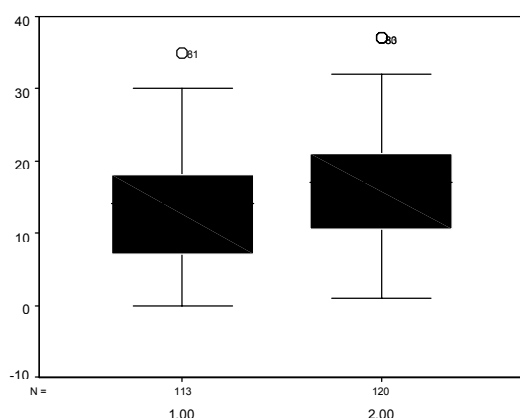
Figure 8m: German Writing Total Gender Difference (out of 41):

1: Male   2: Female

130

The French writing scores show that the extent of the cluster range for girls and boys is identical (10 points) with boys (scoring between 7 and 17 points) 3 points lower than girls (scoring between 10 and 20 points). The German scores are higher overall and show a slightly wider range for boys than for girls (11 points as compared to 9 points), with girls still scoring higher up the scale: between 11 and 22 points for boys, between 15 and 24 points for girls.

ABILITY DIFFERENCES

Ability differences in the writing tasks were analysed in a similar way. For both French and German, the expected 'staircase' distribution is found, with the cluster range within each ability group being similar for French and German. In addition, there is a degree of overlap from group to group in both languages, in every case. The figures below show these patterns. It is not clear what the significance of these differences between French and German may be, although the overall results appear to confirm teachers' own assessment of student ability.

Figure 8n: French Writing Total Ability Difference:
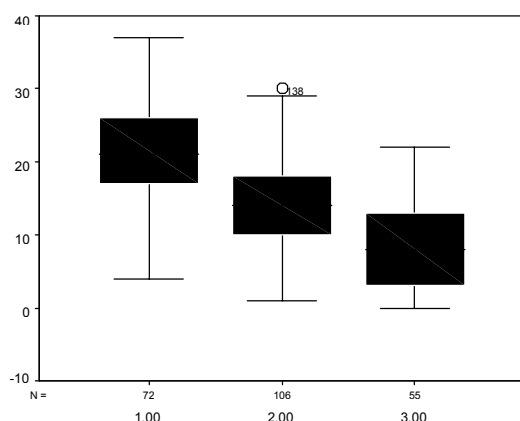
Figure 8o: German Writing Total Ability Difference:
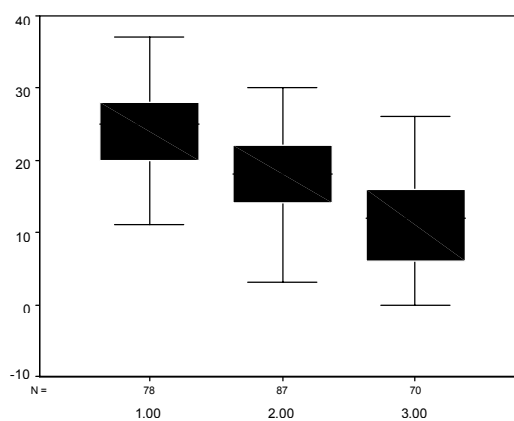
1: High   2: Middle   3: Low

For the writing task scores in French, the highest ability group scores clustered between 17 and 26 points, while the other two groups clustered within relatively narrow ranges: the middle ability group scored between 10 and 18 points, and the bottom group between 3 and 13. There is therefore a 3-point overlap between the bottom and middle groups, and a 1-point overlap between middle and top.

The writing task scores for German show a similar pattern. There is a similar range within each ability group. However the clusters appear higher up the scale with a 2-point overlap in each case. The top group scores clustered between 20 and 28 points, the middle between 14 and 22 points, and the bottom between 6 and 16 points.

RANGE OF WRITING SKILLS – S2 ONLY

WRITING SKILL	Excellent students can …	Average students can …	The weakest students can …
Copying	… copy the majority of words correctly with perfect spelling, including use of accents.	… copy a large number words correctly with perfect spelling, and others with some minor spelling errors.	… copy a few words correctly, and others with both minor and major spelling errors.
Writing from memory – word level	… in the majority of cases, accurately recall appropriate single words of different types - nouns (with articles), verbs, adjectives, pronouns - in a number of different contexts.	… accurately recall some appropriate single words - usually nouns (sometimes with article), and adjectives; … recall the correct phonetic value of other appropriate single words; … accurately recall some inappropriate single words of the correct grammatical type.	… accurately recall a limited number of appropriate single words - usually nouns (without articles); … make random guesses at the others.
Writing from memory – sentence level	… recall full sentences/phrases in order accurately to convey a specified message.	… recall sentences/phrases with some errors in order to convey a specified message; … recall sentences/phrases conveying a message not specified by the task.	… recall some key words related to the message to be conveyed.
Continuous writing - Volume	… write at least 11 complete lines of text	… write 6-10 complete lines of text	… write up to 5 complete lines of text, sometimes combining the target language with words in English or another foreign language.
Continuous writing – Task Coverage	… cover all 3 elements of the task as specified in the instructions	… cover 2 elements of the task as specified in the instructions	… cover 1 element of the task as specified in the instructions
Continuous writing – Linguistic Range	… demonstrate a varied use of the target language in 3 or more different linguistic categories; … demonstrates some interconnection and sequencing by use of time phrases, prepositions, conjunctions- e.g. *Le weekend, quand il fait beau, je joue au temnnis.*	… demonstrate some variety in the use of the target language, perhaps using 3 or 4 examples in 2 linguistic categories such as verbs and nouns- e.g. *J'aime le chocolate et j'adore le disco.*	… demonstrate use of the target language in a minimum of linguistic categories: 4 examples or less within 1 linguistic category. e.g. *Bonjure picen, le basket, le center commercial, babyfoot, le music.*
Continuous writing - Accuracy	… demonstrate generally high, although not necessarily perfect, control of the language being	… demonstrate some, albeit inconsistent, control of the language being attempted; … produce a text which	… demonstrate only a minimum of control of the grammatical features being attempted-

	attempted (correct verb forms, genders); ... produce a text which clearly conveys the desired message. e.g. *Le dimanche matin je vais à la piscine avec mes amis*.	may contain major errors, but still conveys the desired message. e.g. *Je manger la petit dejuner*.	e.g. *Fer du weekend alle alla piceine a au Ju a ou rugby*.

CHAPTER 9
SUMMARY, CONCLUSIONS AND IMPLICATIONS OF THE STUDY

SUMMARY

AIM OF THE PILOT PHASE

1. The main aims were:

   - to develop, implement and evaluate assessment instruments and procedures for an AAP pilot in modern languages within the larger national AAP sample for English (1998); and
   - to develop an initial picture of pupils' achievements in French and German P7 and S2.

SAMPLE

2. Forty schools were involved, all of them on a voluntary basis, reflecting 23 different local authorities: ten schools with S2 French, ten S2 German, ten P7 French and ten P7 German. Since this was a sample within a sample (namely the 1998 full AAP sample for English), and also since German was much less frequently offered than French, it cannot be claimed that the pupils in the 40 schools constitute a nationally representative sample of pupils learning modern languages. In both P7 and S2 there was a roughly equal distribution of gender. Pupils were rated by their teachers as being of 'high', 'middle' or 'low' attainment for languages - at P7 the 'high' and 'middle' attainers considerably outnumbered the 'low' group. This also applied at S2 but to a much lesser extent. In all four samples the number of high ability girls was greater than the equivalent samples for boys, especially at S2 French, and in all four samples the number of low ability boys was greater than the equivalent samples for girls.

GATHERING CONTEXTUAL DATA

3. A large amount of contextual data was collected in order to help the team plan for the assessments. This drew attention to considerable variation across the primary schools in:

   - number of years involved in MLPS,
   - number of MLPS-trained teachers per school,
   - degree of post-training language support for teachers,
   - amount of time allocated per week to MLPS,
   - overall amount of time for MLPS at primary school over P6 and P7,
   - the range of MLPS topics covered.

Across the secondary schools there was considerable variation in:

   - extent to which they were in a position to build on MLPS,
   - extent to which languages other than French were available,
   - amount of time allocated to learning a modern language in S1-S2,
   - the teaching materials in use and the particular order in which topics were covered,
   - extent to which writing was assessed, and
   - whether S2 classes had experienced MLPS (33% of the S2 German classes as against 78% of the S2 French classes in the sample).

Pupils' perceptions of learning a modern language were also collected:

- at P7 pupils were generally enthusiastic and very few considered learning a modern language to be 'difficult' but by S2 perception of difficulty had increased (French 24% and German 18%) and boredom was setting in (French 33% and German 19%).

SPECIFYING THE COMMON DOMAINS OF LANGUAGE ACROSS SCHOOLS

4. A range of steps was taken to identify common domains of language that could be assessed across schools, including detailed analysis of national guidelines and course material plus detailed discussion with primary and secondary school teachers. This helped to validate the assessment instruments before implementation.

5. From the domains identified as above, tests were constructed for French and for German that were closely parallel to each other, so as to allow for comparison between the two languages, though it should be emphasised that the above-mentioned variation across schools allied to the nature of the sample meant that great caution would have to be applied in interpreting any possible language-specific differences in pupils' attainments.

THE 'VISITING ASSESSOR' APPROACH

6. A 'visiting assessor' approach was favoured, consisting of two members of the expanded research team (a non-native speaker and a native speaker in each case) visiting each school in order to administer all aspects of the assessments. This was considered preferable to asking teachers themselves to administer the assessments. (The P7 teachers might have felt that they themselves were being tested, since their language training had amounted to only 27 days and in some cases they may have felt insecure in their command of the language.) In adopting the visiting assessor approach it was assumed that:

- fewer demands would be made on schools (bearing in mind that they had already participated that year in AAP English),

- pupils would be given a chance to show how they could use the language with unknown persons including a native speaker, and

- it allowed for a greater degree of standardisation of procedure across schools, allowing the instruments to be more consistently evaluated.

7. The visiting assessors received training in advance of the administration, were consulted during the period of administration in order to identify any problems that may have arisen and were consulted again afterwards. This helped standardise procedures and contributed substantially to the reliability of the instruments.

THE P7 ASSESSMENT TASKS

8. The P7 assessments consisted of one overall 'test' made up of nine tasks that mainly involved aspects of Listening and Speaking but also drew to a lesser extent on Reading, Writing and Metalinguistic Awareness. Pupils came in twos, in order to minimise anxiety. It was decided to run the assessments in their initial format in the first four primary schools, then to review them. As a result, some modifications were made for the remaining 16 primary schools.

The final form of the single primary school assessment consisted of:

- Task 1 - vocabulary recall via discussion of recently studied topics.
- Task 2 - spontaneous question-answer.
- Task 3 - description of a colour visual.
- Task 4 - understanding the subject and message of a short narrative read aloud by the native speaker.
- Task 5 - understanding short dialogues.
- Task 6 - vocabulary recognition.
- Task 7 - reading aloud three short sentences, understanding them and discussing their linguistic content in English.

THE S2 ASSESSMENT TASKS

9. The S2 assessments were more extensive and detailed, and it was possible to 'trial' them in advance of the pilot administration. The pilot version consisted of:

- Test A (Reading and Writing)   nine tasks
- Test B (Listening)   five tasks
- Test C (Speaking)   six tasks

10. Each task covered a different aspect of the particular skill-area. The Reading-Writing test, for example, included some simple vocabulary-recognition activity but also some more demanding tasks such as reading a short authentic passage from the Internet, understanding a detailed narrative passage containing several words and phrases that were likely to be unknown, doing a detailed gap-filling activity and an open-ended writing task with visual stimuli. The Speaking test, administered individually rather than in pairs, included a talk on a topic (not prepared one day in advance but with a little preparation time built into the day itself), question-answer, prepared role-play, spontaneous description/narration of a coloured visual scene, reading aloud in the foreign language and a metalinguistic element consisting of talking in English about the linguistic properties of a text, e.g. nouns, verbs, agreement. The listening test was based on audio-recorded material, but listening was of course also built into the Speaking test where pupils interacted with both native-speaker and non-native speaker assessors for different purposes.

PUPILS' ATTAINMENTS ON THE ASSESSMENT TASKS

11. **Listening comprehension**: This was assessed at both P7 and S2.

- At P7 there was generally a high level of performance in both French and German, and we infer from this that on another occasion at least one more demanding task could be included.

- The S2 tasks contained a much wider spread of difficulty. On the simpler tasks (e.g. vocabulary recognition), lower-achieving pupils were able to cope, especially when the material had recently been covered and when they could draw on cognates (foreign words that sound like English). There was a much wider spread of attainment on tasks that were more cognitively complex, that required longer-term recall and that dealt with more extended and linguistically difficult input than at P7. On these, however, there was sufficient evidence to suggest progression from P7 to S2 among 'average' and 'high-achieving' pupils.

12. However, while all P7 sample pupils participated fully in the P7 assessments, by the end of S2 some pupils had decided that non-response was a better option than risk-taking or guesswork.

13. **Reading Comprehension:** This was assessed at S2. Although the results varied from task to task, achievement in French and German was of a very similar level. On the simpler tasks (e.g. identifying vocabulary and recognising different types of discourse) attainments were generally high. The more demanding tasks (e.g. Internet and detailed narrative) yielded a much wider range of attainments. On these, lower-attaining pupils were able to comprehend words and phrases with which they were very familiar and to make some intelligent guesses. 'Average' pupils were able to understand the general gist of the more demanding passages but drew meaning more from vocabulary than from grammatical structure. The highest-achieving pupils not only comprehended the familiar but were able to make excellent guesses at the unfamiliar and at times to make use of their grammatical knowledge in order to attain more precise meaning.

14. **Speaking**: This was assessed at both P7 and S2:

   - The best pupils at P7 showed confidence and enthusiasm for speaking and were prepared to 'take risks' by going beyond what they had learnt, even if this entailed making a mistake. The best pupils in S2, on the other hand, displayed two divergent strategies: some were 'risk-takers' as at P7 but others 'played safe' in order to operate accurately and safely within the bounds of what they knew to be correct - both strategies seemed valid and probably reflected different cognitive styles.

   - In the S2 test, performance in German tended to be higher than for French on all four criteria of pronunciation, fluency, accuracy and range of expression.

   - There was clear evidence that the attainments of the best and the average pupils were higher at S2 than at P7. Compared with their P7 counterparts, the best pupils at S2 had better pronunciation and intonation, showed a wider range of structure and were more able to link phrases by using connectors. On a similar comparison, 'average' pupils showed a wider range of vocabulary, made more grammatical distinctions (e.g. gender, persons), were more likely to add personal expressions and showed greater confidence.

   - A similar comparison of the P7-S2 attainments of lower-achieving pupils on the other hand showed a different profile, where progression from P7 to S2 was often not evident. Although both groups were able to recall at least some words or phrases, the P7 pupils showed better understanding of the questions that were put to them and were more able to produce key words in response. Both groups, however, showed a marked lack of confidence and some (especially at S2) were badly affected by nerves. This suggests that further work must be done in developing assessment instruments and procedures that will better encourage lower-achieving pupils to speak. It should be borne in mind that the S2 pupils either had three or four years of recall to deal with (P6/7-S2) or had made a later start in learning a foreign language (S1 as compared with P6/7).

15. **Writing**: Performance in S2 writing was generally high on tasks requiring pupils to comprehend and copy-write words or short phrases. On this, the weakest pupils were able to achieve some success though quite often with spelling errors. On word-phrase tasks involving recall (e.g. based on a visual stimulus) performance was much more variable. The continuous writing task (a paragraph, with no dictionary support, within a time-limit of fifteen minutes) produced a wide spread of attainment.

   - Overall, scores for German were somewhat higher than for French on all four criteria ('volume of language produced', 'task coverage', 'range of expression' and 'accuracy') of the more demanding continuous writing task.

   - The best pupils were able accurately to copy and recall single words and phrases and to produce continuous text (with little preparation time allowed) covering all aspects of the

specified task, demonstrating varied use of language, some interconnection and sequencing and with generally high though not perfect control of language forms (e.g. verbs, genders).

SAMPLING CONSTRAINTS

16. In considering the above levels of attainment, the following must be borne clearly in mind:

- This was the first systematic attempt at national assessment in modern languages at P7 and S2, and so there were no precedents on which researchers, teachers or pupils could build, particularly as the national 5-14 Guidelines for Modern European Languages did not cover MLPS.

- Although girls' attainments tended to be somewhat higher than those of boys, and German attainments to be somewhat higher than those in French, constraints of the sample within which the research team had to operate meant that no firm conclusions could be made about comparative attainments in respect of gender (girls v boys) or language (French v German).

17. A marked feature of the performance of most pupils was what the research team termed 'partial competence', which showed that many pupils had partly but not fully internalised many of the linguistic features they were learning, whether these applied to vocabulary, morphology, syntax or meaning.

EVALUATING THE INSTRUMENTS AND PROCEDURES

18. **Time taken**: Originally it had been hoped that the P7 tests would take 0.5 days per school and the S2 tests 1.0 days. This had to be revised in advance of the administration, to allow 1.0 days for all schools: primary and secondary. All tests were administered accordingly.

19. **Match to pupils**: The pupils' teachers had categorised those pupils taking the assessments into 'high-', middle-' and 'lower-attaining'. Generally, pupils' attainments in the assessments corresponded to those teacher-estimations. Within each test, the lower levels of task generally allowed all pupils show evidence of learning, and as such we consider that they generally 'worked'. The more demanding tasks also 'worked' in that they generated a much wider spread of attainment but gave the highest-attaining pupils an opportunity to give evidence of truly outstanding levels of performance, indicating they were much less than two years away from good performance at Standard Grade.

20. **Assessor comments.** Feedback was obtained at various points from the team of visiting assessors:

- At P7, most assessors felt the listening parts of test had generally worked well and that the format was well-enough understood by pupils. In some cases problems were identified as unsuitable accommodation or doing the test too late in the day for particular pupils who were tired. Most assessors felt the speaking parts of the tests did what was required. More work needed to be done on how to rate pupils' performance in 'real time' as the interaction proceeded.

- At S2, the procedures for the Reading-Writing test had generally worked well, though there were some problems in harmonising the timing of the test to the particular school timetable for the day. The metalinguistic awareness part had worked well with 'high achieving' pupils

but less well with the others. The speaking test was generally considered to be valid for S2 pupils and the scoring schemes were felt to be generally manageable, though improvements could be made to facilitate easier note-taking.

21. **Pupils' Comments.** Feedback was also obtained from pupils through a questionnaire which they completed once their assessments were over.

- At P7: *Anxiety*. Before the tests, levels of anxiety were slightly higher for German than for French but after taking the tests only a small minority remained anxious. *Difficulty*. None of the P7 pupils found the tests to be 'very difficult', though more found the German test to be 'difficult' as compared with the French one. A large majority in each language found the tests to be 'average'.

- At S2: *Anxiety*. In the Reading-Writing test anxiety levels decreased after the event, though some (11% French and 10% German) remained anxious afterwards. Anxiety levels decreased after the Listening test, somewhat more so than after the Reading-Writing one. For the Speaking test, anxiety levels were high beforehand (63% French and 40% German), but after the tests they had gone down considerably, though approximately 10% for each language remained 'anxious' or 'very anxious'. *Difficulty of the tests*: Most found the Reading-Writing test to be of 'average' difficulty. It was more or less the same for the Listening test, though the 'average' score lost a little in both directions towards 'easy' and also towards 'difficult'. Most found the speaking test to be of 'average' difficulty. However, 33% of German pupils found it to be 'easy' or 'very easy' whereas only 15% of French pupils found it 'easy' and none found it 'very easy'. 42% of the French pupils found it to be 'difficult' or 'very difficult' as compared with 17% of German pupils. *Fairness of the tests*: Most found the Reading-Writing test and the Listening test to be 'fair' or 'average'. The Speaking test gained the highest score on 'fairness' for both languages. This was interesting, given that anxiety levels had been quite high before taking this test.

CONCLUSIONS

PARTICIPATION OF SCHOOLS

1. The research project was successful in obtaining the willing and voluntary participation of its target number of forty schools. Indeed, had time and resources allowed, it would have been possible to add to this number. This says something positive about the interest and the professionalism of the school staff involved, both headteachers who gave their consent in principle (having already experienced AAP English that same term) and classroom teachers at P7 and S2 who were teaching their particular modern language.

2. Given the many sensitivities arising from the problematic state of modern languages in P6-S2 of Scottish schools during Summer term 1998, as briefly indicated in Chapter 1, it is highly satisfactory that in a relatively short space of time (since the Pilot phase effectively did not begin until January 1998) such a full participation rate was achieved and the tests successfully administered. It seems reasonable to conclude that the profession has moved quite a long way from its initial uncertainties on this matter and that a climate is being established in which national AAP-type assessments in modern languages at P7 and S2 may be perceived as potentially advantageous.

PARTICIPATION OF PUPILS

3.   It was also heartening that the pupils were generally not intimidated by their participation in the assessments. In this connection it is worth bearing in mind that in some respects these tests were more demanding than their Standard Grade counterparts, in that (for Speaking) pupils had much less preparation time than at Standard Grade and they also had to contend with two unknown adults, one of whom was a native speaker of the foreign language and who was instructed to talk in a natural flow.

4.   In the great majority of cases, the pupils were not 'fased' by this challenge but in fact rose to it. It says much for them, but also for the skills of the assessors and the carefully planned format of the assessments, that anxiety levels were generally much lower after the tests than beforehand.

5.   Of some concern, nonetheless, must be the lower-achieving pupils at both P7 and S2, particularly at S2, who did find their speaking assessment stressful. This suggests that ways must be found of either boosting their self-confidence or of administering the tests for them in somewhat different ways, or possibly both.

SAMPLING

6.   The pilot experience points to a problem with sampling. Given that the modern languages sample was embedded within a larger, nationally representative sample for English, and given also that German was offered in schools much less frequently than was French, it proved very difficult to achieve samples that matched each other in key respects so as to allow for comparison of pupils' performance across these two languages.

7.   Had the sample for modern languages not been constrained by having to be embedded within the national sample for English, then there might possibly have been better matching of the French and German samples. On the other hand, there was considerable merit in being within the same sampling frame as English, since this permitted the possibility of comparing pupils' performance in English and their particular foreign language, though the incentive for doing this was to some extent diminished when it was decided that AAP English would not test spoken language.

INSTRUMENTS AND PROCEDURES

8.   Given that this was a first attempt at devising and administering national assessments in modern languages at a level below Standard Grade, the evidence on the suitability of the instruments and procedures, based on feedback from assessors and from pupils, as well as on performance data from pupils, indicates clearly that generally the tests fared well. This does not mean that they were without faults or difficulties, and these are charted in detail in Chapters 3 and 4 and in Appendix 4.

9.   The pattern of results emerging from the quantitative analysis of the S2 test data tends towards a normal distribution or bell-shaped curve in each of the four skills and for both languages, which would seem to indicate that the tests have performed broadly in line with expectation.

10. It is essential to bear in mind that these were tests of individual pupils, administered out of class. As such they cannot replicate tasks that are embedded within a complex network of whole-class activity, extending backwards over several days, with the teacher playing an orchestrating role. Classroom activity of this sort, with a good teacher, can 'scaffold' impressive levels of pupil performance, but it does not systematically assess what individual pupils can do. In our case, on the other hand, pupils were confronted not by normal classroom activity but by individual assessments out of class that allowed relatively little time for advance preparation, yet these proved

'valid' in that they did give pupils a range of opportunity to show what they could do and in the process were generally judged by the pupils not to be unfair or anxiety-inducing.

11. This strongly confirmed the value of 'validating' the assessments beforehand by a variety of complementary means such as an extensive telephone survey of teachers, scrutiny of course material and of national policy documents, review of the S2 assessments by a panel of experienced teachers and discussion with the actual class-teachers in the primary schools that eventually were identified for participation.

VISITING ASSESSORS

12. The AAP pilot in modern languages made one deliberate departure from the precedent adopted for English, mathematics and science, in that a team of visiting assessors was employed to administer all elements of the assessments, rather than relying on class-teachers in the participating schools. The evidence suggests that in the circumstances of the time this was an appropriate thing to do, particularly since a key aim was to pilot the assessment instruments. Having the team of visiting assessors meant that key procedures could be planned, standardised and refined.

13. Although it became clear that a greater amount of advance training would have been beneficial, much was learnt form the experience, and the evidence suggests that positive benefits arose from having a visiting native speaker playing a key role in each Speaking assessment at both P7 and S2. There was something 'authentic' and 'natural' about the interactions to which most pupils responded positively. Although we are in no doubt that non-native speaking teachers could administer tests of this sort with a high level of professional competence, we believe there is a good case for maintaining the 'native speaker as visiting assessor' role in the future.

PUPILS' ACHIEVEMENTS

14. These have been briefly summarised in the present chapter and summarised in greater detail in the tables of 'Range Statements' for the chapters on Listening , Speaking, Reading and Writing and which show what high-, middle- and lower-achieving pupils were able to do. We emphasise that 'able to do' relates only to those areas assessed and we make no claims concerning what pupils are able to achieve in the different, and more supportive, context of everyday classroom life. Nonetheless, the assessments do cover a wide range of activity: not only do they feature activities that are high in face validity (i.e. teachers agreed in advance that they were plausible) and in content validity (i.e. they cover what has been taught), but they also feature activities that economically assess underlying knowledge, e.g. the gap-filling task in the Reading-Writing test, or that provide initial pointers towards metalinguistic knowledge (children's implicit or explicit knowledge about language) and metalinguistic awareness (their explicit awareness of possessing that knowledge).

15. The pupil sample necessarily had to be cross-sectional and could not be longitudinal, but nonetheless there were tentative pointers towards progression in their foreign language development from P7 to S2, as documented in the range statements for those skill areas (particularly Listening and Speaking) that featured in the assessments at both P7 and S2 and which show that S2 pupils were able to operate with more complex language than were pupils at P7. Although progression of this sort is to be expected, it is worth recalling that there was very much less evidence of P7 to S2 progression in the final report of the independent evaluation[12] of the national MLPS pilots, and so something may have been gained in recent years.

---

[12] Low, L, Brown, S., Johnstone, R. and Pirrie, A. (1995). *Evaluating foreign languages at primary school: Final Report.* University of Stirling: Scottish CILT.

16. A major feature of pupil performance was the large amount of what we term 'partial competence' that was displayed. This affected all levels of achievement, including the highest achievers. A common meaning of 'partial competence' nowadays, as expressed for example in Council of Europe publications, has to do with giving priority to the development of certain skill areas (e.g. listening comprehension) rather than attempting to develop a more rounded competence covering all four skill areas (listening, speaking, reading and writing) to the same level. By 'partial competence', however, we mean something different, namely partial internalisation of what has been taught, whether in one skill area or in all. We believe it to be a distinctive feature of our present report that it has documented pupils' 'partial competence' of this sort in such considerable detail. We have done so because in language performance if something is not entirely correct, it does not mean that it is necessarily wrong. Our 'partial competence' data suggest strongly that many pupils have learnt a lot, for which credit is deserved, but that over time some fine-tuning of their language systems (not only their grammatical morphology and syntax but also their spelling and their semantics) would be beneficial.

IMPLICATIONS FOR THE FUTURE

Although the sample of 20 French and 20 German classes covering P7 and S2 was too small to be nationally representative, the pilot experience has yielded a rich bank of achievement, contextual and attitudinal data.

1. **Implications for teaching.** The data from the 1998 modern languages pilot suggest that:

- Primary teachers have enabled many pupils to build up high levels of confidence and enthusiasm. Performance in the language tended to vary considerably across schools, but the performance of the best pupils was highly encouraging in their capacity to cope with native-speaker talk and in the fluency, range and flexibility of what they were able to say.

- Secondary teachers have enabled many high-achieving and average-achieving pupils at S2, whether or not they had MLPS in their secondary school language, to progress in the language well beyond the levels attained by most pupils at P7.

- At both P7 and S2, it would be desirable to explore further ways of enabling lower-achieving pupils to have a successful language-learning experience, both in regard to their performance in the language and their self-confidence as language-users.

- The high incidence of 'partial competence' may be viewed in two ways. First, it undoubtedly does provide evidence of widespread and significant learning by most pupils at both P7 and S2. The notion of 'interlanguage' (or 'interim' language) is well-attested in second language acquisition research. It confirms that language development (whether one is referring to first or second or foreign language) proceeds in an natural way through a series of 'successive approximations' before attaining more mature and 'correct' forms.[13] Second, however, there is a clear case for finding ways of helping pupils at all levels develop more accurate control: not only in relation to morphology and syntax but also in spelling, vocabulary and meaning,

---

[13] Peltzer-Karpf, A. and Zangl, R. (1998). *Vier Jahre Vienna Bilingual Schooling: Eine Langzeitstudie.* Vienna: Bundesministerium für Unterricht und kulturelle Angelegenheiten. Abteilung 1/1. This major study, based on bilingual education (a much more intensive and powerful form than MLPS or S1-S2 in Scotland) showed that all pupils after an initial silent period produced global phrases that were generally grammatically accurate but that after a while, as pupils attempted more creative language-use, their grammatical accuracy broke down and they went through a phase of 'Systemturbulenz' before (after four years or so) their grammatical system sorted itself out. In this sense, then, 'partial competence' as identified in our study is a natural phenomenon.

and also to develop techniques for recall and re-use of previously learnt material. This becomes increasingly important as S2 becomes viewed as the fourth rather than the second year of pupils' foreign-language learning experience.

- The 'Internet' task that featured in our 1998 Reading-Writing test, though containing difficult 'authentic' language, was perceived by pupils as the most popular task within that test. Both in classroom teaching and in a future AAP, there is a strong case for including more 'Internet', 'e-mail' activity and possibly also a video-element for listening comprehension, to allow pupils the opportunity to engage with foreign-language pupils of their own age.

2. **Implications for Language Awareness**. The 1998 modern languages pilot broke new ground by introducing a small and experimental element of metalinguistic awareness.

- We judge that it was useful to do so and that it proved appropriate to achieve this by talking with pupils about what they had said and read rather than by asking them to undertake a special written assessment.

- For a future AAP, we consider that MLA (metalingusitic awareness) should figure more prominently than in our pilot study, ideally based on collaboration with AAP English. A shared metalinguistic component would reflect current and emerging Guidelines for 5-14 in English and Modern European Languages, would respond to current national priorities for literacy based on knowledge of language, and would provide a measure of MLA achievement that could be correlated with performance in the various components of the modern languages and the English assessments.

3. **Other implications for a future AAP.** We believe that the 1998 pilot confirms the desirability and the feasibility of implementing a larger-scale AAP for modern languages in 2001. For this, we offer the following considerations:

- The format of the pilot tests was designed to ensure validity and reliability and proved broadly appropriate. Some further refinement is needed, particularly in inserting a more demanding listening component at P7 to ensure that the top of the range is fully tested and in finding a less stressful way of assessing lower-achieving pupils in speaking.

- Some further review of content and tasks is likely to be appropriate, once the new 5-14 guidelines have been produced, in order to ensure that the AAP fully reflects the curricular experiences that pupils are intended to have.

- The logistics of running the tests in the way we did were demanding. As our arrangements differed markedly from those of the AAP assessments in other areas (English, mathematics and science), a review of these arrangements is likely to be required, if a larger sample is indicated. There are strong arguments for retaining the native speaker as visiting assessor for the speaking part of the assessments, and costing a larger survey should take this into account.

- There were particular problems associated with the sampling procedures used for the pilot. These had no negative effects on our piloting the instruments and procedures (the first and main aim of the pilot study) but they necessitate considerable caution in interpreting any findings on pupils' attainments (the second aim of the pilot) in relation to gender, school, language or teachers' initial estimates of pupils' ability. The difficulty lay in embedding the study within the larger national sample for AAP English. As a result, the modern languages pilot, though successfully achieving its target of 40 participating schools, ended up with a skewed sample. The principle of linking the two samples needs to be reviewed before a larger AAP for modern languages is conducted in 2001. There are strong reasons for linking modern languages and English within one overall sample, but it might prove easier to select

the modern languages schools first and then to build a possibly larger English sample round this.

P7 QUESTIONNAIRE

The purpose of this questionnaire is to gather contextual data that will help us understand the sample pupils' level of achievement in the AAP pilot assessments.
Some of the requested information is of a general nature, but most relates more specifically to those particular pupils who participated in the pilot project.
All information provided will be treated in the strictest confidence.
We would be extremely grateful if you could return the form by <u>Friday 18th September</u> in the pre-paid envelope provided.

SECTION A: THE SCHOOL

NAME OF SCHOOL _____

1. How many pupils (approximately) attend your school? _____

SECTION B: ORGANISATION OF LANGUAGE TEACHING

2. Which languages are currently taught in your school?

   Please tick the appropriate box(es)

   - ❏ French
   - ❏ German
   - ❏ Spanish
   - ❏ Italian
   - ❏ Gaelic
   - ❏ Other(s)   **Please state** _____

3. How long has the school been involved in MLPS?

   Please state year when first introduced _____

   Please state number of years of involvement _____

4. How many members of staff have received national training and when was it completed?
   Please indicate below how many teachers have completed the national training programme, including any who may subsequently have moved out of post.

Trained staff	Date training completed	Currently in or out of post (Please tick as appropriate)	
Teacher 1 ❏		❏ IN	❏ OUT
Teacher 2 ❏		❏ IN	❏ OUT
Teacher 3 ❏		❏ IN	❏ OUT
Teacher 4 ❏		❏ IN	❏ OUT

5. Has any subsequent language support for the teachers trained been received since the completion of training?

    ❐   Yes
    ❐   No

If Yes, please indicate in the space below

6. Has a co-operative link been established with the associated secondary school, e.g. transfer of pupil information, agreed syllabus, etc.?

    ❐   Yes
    ❐   No

If Yes, please give details in the space below

SECTION C: AAP PILOT SAMPLE PUPILS

7. For those pupils who participated in the AAP pilot, at which point did their foreign language learning begin?

At the end of:		**In the course of**:	
❐ P7		❐ P7	
❐ P6		❐ P6	
❐ P5		❐ P5	❐
P4	❐	P4	❐
Other	❐	Other	
Please state _____		**Please state** _____	

8.   How was the foreign language teaching of the sample pupils organised during the time   they were involved in the MLPS programme?

Please tick the statement below which most closely fits the sample pupils' experience in each year group, or briefly describe the situation in the space provided

In P7:
❑   Pupils were taught by their class teacher
❑   Pupils were taught by a drop-in teacher
❑   Other system   **Please describe in the space overleaf**

```
┌──────────────────────────────────────────────────────────┐
│                                                            │
│                                                            │
│                                                            │
│                                                            │
│                                                            │
└──────────────────────────────────────────────────────────┘
```

In P6 (if applicable):
❑   Pupils were taught by their class teacher
❑   Pupils were taught by a drop-in teacher
❑   Other system   **Please describe in the space below**

```
┌──────────────────────────────────────────────────────────┐
│                                                            │
│                                                            │
│                                                            │
│                                                            │
│                                                            │
└──────────────────────────────────────────────────────────┘
```

In earlier classes (if applicable):
Please state which classes   _____

❑   Pupils were taught by their class teacher
❑   Pupils were taught by a drop-in teacher
❑   Other system   **Please describe in the space below**

```
┌──────────────────────────────────────────────────────────┐
│                                                            │
│                                                            │
│                                                            │
│                                                            │
│                                                            │
└──────────────────────────────────────────────────────────┘
```

9.   How much time did the sample pupils spend learning a foreign language, over a week   at each of the following stages?
Please write the total number of minutes and lessons per week in the space below

In P7:
_____ minutes per week for foreign language 1

_____ lessons/episodes per week for foreign language 1

_____ minutes per week for foreign language 2 (if studied)

_____ lessons/episodes per week for foreign language 2 (if studied)

In P6 (if applicable):

_____ minutes per week for foreign language 1

_____ lessons/episodes per week for foreign language 1

_____ minutes per week for foreign language 2 (if studied)

_____ lessons/episodes per week for foreign language 2 (if studied)

In earlier classes (if applicable):
Please state which classes _____

_____ minutes per week for foreign language 1

_____ lessons/episodes per week for foreign language 1

_____ minutes per week for foreign language 2 (if studied)

_____ lessons/episodes per week for foreign language 2 (if studied)

10. Which were the main language skills covered by the sample pupils?
    Please tick the box(es) below

    ❐ Listening
    ❐ Speaking
    ❐ Reading
    ❐ Writing

11. What were the main sources for the teaching material used with the sample pupils?
    Please tick as appropriate

    ❐ National training materials
    ❐ Regional training materials
    ❐ Published/commercial materials
    ❐ Other **Please give details in the space below**

12.   Which of the following topics would have been covered by the sample pupils at the time of
sitting the AAP pilot assessments, and at which stage would they have been covered: P7, P6 or earlier?
Please tick the appropriate column(s) for each topic where applicable

TOPIC	P7	P6	Earlier (Please specify)
Alphabet			
Animals/pets			
Classroom language/ instructions			
Classroom objects			
Clothes			
Colours			
Dates, birthdays			
Family			
TOPIC	P7	P6	Earlier (Please specify)
Food and drink			
Hobbies and freetime: sports, leisure, etc.			
House and home			
Nationalities			
Numbers			
Parts of the body			
Personal language: name, age, home town			
Physical descriptions			
Places in town			
Telling the time			
Weather			
Other (please give details)			

13.   Is there any other relevant information that you feel might have a bearing on the performance of the sample pupils in the AAP pilot assessments?
Please comment in the space below

14.   What are your views about the current situation of MLPS in your school and nationally?

What advantages (if any) and what problems (if any) has it brought?
Please comment in the space below.

S2 QUESTIONNAIRE

The purpose of this questionnaire is to gather contextual data that will help us understand the sample pupils' level of achievement in the AAP pilot assessments.
Some of the requested information is of a general nature, but most relates more specifically to those particular pupils who participated in the pilot project.
All information provided will be treated in the strictest confidence.
We would be extremely grateful if you could return the form by <u>Friday 18th September</u> in the pre-paid envelope provided.

SECTION A: THE SCHOOL

NAME OF SCHOOL _____

1.   How many pupils (approximately) attend your school?   _____

SECTION B: ORGANISATION OF LANGUAGE TEACHING

2.   Which languages are currently taught in S1 and S2 of your school?
                                        Please tick the appropriate box(es)

☐   French
☐   German
☐   Spanish
☐   Italian
☐   Gaelic
☐   Other(s)   **Please state**   _____

S1 PROVISION

3.   For those pupils who participated in the AAP pilot, how was their language teaching organised during S1?

Please tick the statement below which most closely fits your school's situation, or briefly describe the situation in the space provided

☐   All pupils in S1 studied the same language (e.g. everyone took French)
☐   Pupils studied one of the two languages available
☐   Pupils took more than one language in the course of S1
☐   Other system   **Please describe in the space below**

4.   How were foreign language classes organised for the sample pupils during S1?

☐   All foreign language teaching was done in mixed ability classes
☐   Setting arrangements were put into place at some point during S1
☐   Other system   **Please describe in the space below**

```


```

5.   During S1 how much time did the sample pupils spend learning a foreign language, over a week?
Please write the total number of minutes and lessons per week in the space below

_____ minutes per week for foreign language 1

_____ lessons per week for foreign language 1

_____ minutes per week for foreign language 2 (if studied)

_____ lessons per week for foreign language 2 (if studied)

6.   During S1 what was the average size of the foreign language class attended by the sample pupils?
Please write the average number of pupils in an S1 foreign language class in the space below

_____ pupils

S2 PROVISION

7.   For those pupils who participated in the AAP pilot, how was their language teaching organised during S2?

Please tick the statement below which most closely fits your school's situation, or briefly describe the situation in the space provided

☐   Pupils continued with the language they studied in S1
☐   Pupils continued with the language they studied in S1 and start a second language
☐   Pupils continued with the language they studied in S1 and also had 'tasters' in other languages offered in the school
☐   Pupils continued with the two languages they studied in S1
☐   Pupils chose one of the languages to which they were introduced in S1
☐   Other system   **Please describe in the space overleaf**

```


```

```
┌─────────────────────────────────────────────────────────┐
│                                                           │
│                                                           │
│                                                           │
└─────────────────────────────────────────────────────────┘
```

8. How were foreign language classes organised for the sample pupils during S2?

      ☐   All foreign language teaching was done in mixed ability classes
      ☐   Pupils were allocated to broad ability groupings
      ☐   A setting system was in place
      ☐   Other system   **Please describe in the space below**

```
┌─────────────────────────────────────────────────────────┐
│                                                           │
│                                                           │
│                                                           │
│                                                           │
│                                                           │
│                                                           │
└─────────────────────────────────────────────────────────┘
```

9. During S2 how much time did the sample pupils spend learning a foreign language, over a week?
    Please write the total number of minutes and lessons per week in the space below

      _____ minutes per week for foreign language 1

      _____ lessons per week for foreign language 1

      _____ minutes per week for foreign language 2 (if studied)

      _____ lessons per week for foreign language 2 (if studied)

10. During S2 what was the average size of the foreign language class attended by the sample pupils?
    Please write the average number of pupils in an S2 foreign language class in the space below

      _____ pupils

MODERN LANGUAGES IN THE PRIMARY SCHOOL

11. Did the sample pupils come into S1 having begun foreign language learning in primary?

      ☐   Yes   **Please go to Question 12**
      ☐   No   **Please go to Question 15**

12.   Which language(s) did the sample pupils study in Primary?
        Please state in the space overleaf

```
+---------------------------------------------------------------+
|                                                               |
|                                                               |
|                                                               |
|                                                               |
|                                                               |
+---------------------------------------------------------------+
```

13.   When did the sample pupils begin their foreign language learning?

        ❒   P7
        ❒   P6
        ❒   Other   **Please state** _____

14.   Has a co-operative link been established with the primary cluster, e.g. transfer of pupil information, agreed syllabus, etc.?

        ❒   Yes
        ❒   No

        If Yes, please give details in the space below

```
+---------------------------------------------------------------+
|                                                               |
|                                                               |
|                                                               |
|                                                               |
|                                                               |
|                                                               |
|                                                               |
+---------------------------------------------------------------+
```

SECTION C: MATERIALS

15.   Did you use a commercially published course, e.g. *Avantage*, *Zick Zack*, as the main source of teaching materials with the sample pupils in S1 and S2?

        ❒   Yes   **Please go to Question 16**
        ❒   No   **Please go to Question 18**

16.   Which coursebook was used with the sample pupils in the language in which they
were   assessed for the AAP pilot (either French or German)?

        In S1: _____

        In S2: _____

17.   Which point had been reached in the above S2 coursebook by the time of sitting the AAP pilot assessments in May/June of S2?
Please comment in the space below

```


```

18.   If you do not use a particular commercial course in S1 and S2, please explain in the space below what kind of materials you use.

```


```

19.   Do you, on a regular basis, have to supplement the commercial courses you use throughout S1 and S2?

    ❐   Yes   **Please go to Question 20**
    ❐   No   **Please go to Question 21**

20.   If you supplement the commercial courses used in S1 and S2, please indicate briefly how you do this in the space below.

```


```

21.   Did the sample pupils have their own textbook, or other source of reference for the foreign language, which they could take home?

In S1:
❐   Yes
❐   No

In S2:
❐   Yes
❐   No

22. Which of the following topics would have been covered by the sample pupils and at which   stage: S1 and/or S2?

Please tick the appropriate column(s) below for each topic, where applicable. The list continues overleaf.

TOPIC	S1	S2
Alphabet		
Classroom language		
Classroom objects		
Clothes		
Colours		
Daily routine		
Dates		
Directions		
Family		
Food and drink		
TOPIC	S1	S2
Hobbies and freetime		
House and home		
Jobs/places of work		
Nationalities		
Numbers		
Parts of the body		
Personal language		
Pets		
Physical descriptions		
Places in town		
School subjects		
Time		
Weather		
Other (please give details)		

23. In which of the following skill areas had the sample pupils previously been routinely assessed as part of their language learning programme?

Please tick the appropriate box(es)

In S1:
❐   Listening
❐   Speaking
❐   Reading
❐   Writing

In S2:
❐   Listening
❐   Speaking
❐   Reading
❐   Writing

24. What was the source of the assessment material, e.g. commercial coursebook, in-house materials, etc.?
Please give details in the space below

```

```

25. Did you test any other areas with the sample pupils in S1 and/or S2, e.g. vocabulary, grammar etc.?
Please describe in the space below

```

```

26. Is there any other relevant information that you feel might have a bearing on the performance of the sample pupils in the AAP pilot assessments?
Please comment in the space below

```

```

27. What are your views about MLPS?
What do you feel to be its good points (if any)? What problems does it cause (if any)?
Please comment in the space below

```

```

28. What are your views about the current situation of ML teaching of S1-S2 in your school and nationally?
What do you feel to be its good points (if any)? What problems does it cause (if any)?
Please comment in the space below

P7 QUESTIONNAIRE (FRENCH): 1998 (German version was also administered)

Many thanks for taking this test. We would be very grateful if you would take a few moments in order to give us your private views about it.

What is your school?	What is your number?	Are you male or female?
		M ☐ F ☐

Your overall impressions of the test
In each case please tick the box that corresponds most closely to your view.

1.   **How easy was the test?**

Very easy	Easy	Average	Difficult	Very difficult
☐	☐	☐	☐	☐

2.   **How anxious or relaxed were you just before the start?**

Very anxious	Anxious	Average	Relaxed	Very relaxed
☐	☐	☐	☐	☐

3.   **How anxious or relaxed were you at the end of the test?**

Very anxious	Anxious	Average	Relaxed	Very relaxed
☐	☐	☐	☐	☐

4.   **How easy or difficult do you find French at school?**

Very easy	Easy	Average	Difficult	Very difficult
☐	☐	☐	☐	☐

5.   **At what age did you begin learning French?**

Before primary	Primary 1, 2 or 3	Primary 4 or 5	Primary 6	Primary 7
☐	☐	☐	☐	☐

6.   **Do you know which language(s), in addition to English, you will be learning in first year at secondary school?**

French	German	Italian	Spanish	Gaelic
☐	☐	☐	☐	☐

If you will be learning another language, please enter it in this box. _____

7. **Which languages do you know? Make a list, including English.**

A = your strongest language.   B = your second strongest etc.

A. | C. |
B. | D. |

8. **How often have you been to France or to some other country where French is the main language?**

Never   Once   Twice   3-5 times   More than 5

9. **Do you ever use French here in Scotland outside school?**

If you do, please indicate briefly in the space below what you do (Listening? Speaking? Reading? Writing?) who you use it with and how often.

Many thanks for your co-operation. This will be very helpful.

S2 QUESTIONNAIRE - TEST A: READING/WRITING

National AAP Pilot (French): 1998 (German version was also administered)
S2 Reading and Writing Test: Pupil Feedback Sheet

Thank you for taking this test. We would be very grateful if you would take a few moments in order to give us your views about it. The information you give will be totally confidential.

What is the name of your school

What is the number of your test sheet? ☐ (It is in the top right corner of the outside page)

Are you male or female? ☐ Male ☐ Female

Your overall impressions of the test
In each case please tick the box that corresponds most closely to your view.

1.   How easy was the test?
Very easy   Easy   Average   Difficult   Very difficult
☐   ☐   ☐   ☐   ☐

2.   How fair was it?
Very fair   Fair   Average   Unfair   Very unfair
☐   ☐   ☐   ☐   ☐

3.   How anxious or relaxed were you just before the start?
Very anxious   Anxious   Average   Relaxed   Very relaxed
☐   ☐   ☐   ☐   ☐

4.   How anxious or relaxed were you at the end of the test?
Very anxious   Anxious   Average   Relaxed   Very relaxed
☐   ☐   ☐   ☐   ☐

5.   If you have any comments on particular Tasks within the test, please tell us what you think in the spaces that are provided.

Task 1	
Task 2	
Task 3	
Task 4	
Task 5	
Task 6	
Task 7	
Task 8	

Task 9	

6. How easy or difficult do you find French as a subject at school?

Very easy	Easy	Average	Difficult	Very difficult
☐	☐	☐	☐	☐

7. How interesting or uninteresting do you find French as a subject at school?

Very interesting	Interesting	Average	Boring	Very boring
☐	☐	☐	☐	☐

8. How good do you think you are at French?

Very good	Good	Average	Not good	Not at all good
☐	☐	☐	☐	☐

9. At what age did you begin learning French?

Before primary school ☐	Primary 6 ☐
Primary 1 or 2 ☐	Primary 7 ☐
Primary 3 or 4 ☐	Secondary 1 ☐
Primary 5 ☐	Secondary 2 ☐

10. Which languages do you know? Make a list, including English.
A = your strongest language. B = your second-strongest language etc.

A.	C.
B.	D.

11. How often have you been to France or some other country where French is spoken?

Never	Once	Twice	A few times	Often
☐	☐	☐	☐	☐

12. How often do you use French in Scotland outside your school, apart from homework?

	Never	Very occasionally	Once or twice a month	Once or twice a week	Just about every day
I listen to it	☐	☐	☐	☐	☐
I speak it	☐	☐	☐	☐	☐
I read it	☐	☐	☐	☐	☐
I write it	☐	☐	☐	☐	☐

13. If in 12. Above you indicated that you speak French, please write down in the space below the person or persons you speak it with

Many thanks for your co-operation. This will be very helpful.

S2 QUESTIONNAIRE - TEST B: LISTENING

National AAP Pilot (French): 1998 (German version was also administered)
Listening Test: Pupil Feedback Sheet

Thank you for taking this test. We would be very grateful if you would take a few moments in order to give us your views about it. The information you give will be totally confidential.

What is the name of your school

What is the number of your test sheet?
(It is in the top right corner of the outside page)

Are you male or female?   Male   Female
(Please tick the appropriate box)

Your overall impressions of the test
In each case please tick the box that corresponds most closely to your view.

1.   How easy was the test?
Very easy   Easy   Average   Difficult   Very difficult

2.   How fair was it?
Very fair   Fair   Average   Unfair   Very unfair

3.   How anxious or relaxed were you just before the start?
Very anxious   Anxious   Average   Relaxed   Very relaxed

4.   How anxious or relaxed were you at the end of the test?
Very anxious   Anxious   Average   Relaxed   Very relaxed

5.   If you have any comments on particular Tasks within the test, please tell us what you think in the spaces that are provided.

Task 1

Task 2

Task 3

Task 4

Task 5

Many thanks for your co-operation. This will be very helpful.

S2 QUESTIONNAIRE - TEST C: SPEAKING

National AAP Pilot (French): 1998 (German version was also administered)
Speaking Test: Pupil Feedback Sheet

Thank you for taking this test. We would be very grateful if you would take a few moments in order to give us your views about it. The information you give will be totally confidential.

What is the name of your school

What is the number of your test sheet?
(It is in the top right corner of the outside page)

Are you male or female? ☐ Male ☐ Female
(Please tick the appropriate box)

Your overall impressions of the test
In each case please tick the box that corresponds most closely to your view.

1.   How easy was the test?

Very easy	Easy	Average	Difficult	Very difficult
☐	☐	☐	☐	☐

2.   How fair was it?

Very fair	Fair	Average	Unfair	Very unfair
☐	☐	☐	☐	☐

3.   How anxious or relaxed were you just before the start?

Very anxious	Anxious	Average	Relaxed	Very relaxed
☐	☐	☐	☐	☐

4.   How anxious or relaxed were you at the end of the test?

Very anxious	Anxious	Average	Relaxed	Very relaxed
☐	☐	☐	☐	☐

5.   If you have any comments on particular Tasks within the test, please tell us what you think in the spaces that are provided.

Task 1	
Task 2	
Task 3	
Task 4	
Task 5	
Task 6	

Many thanks for your co-operation. This will be very helpful.

APPENDIX 3
ASSESSOR FEEDBACK QUESTIONNAIRE

NAME ………………………………………………………

TESTS ADMINISTERED *Please tick all those which you administered*

P7 French	☐	S2 French	☐
P7 German	☐	S2 German	☐

NB *If you administered more than one set of tests, please complete separate forms for each.*

SCHOOLS VISITED *Please list the schools you visited here*

……………………………………… ………………………………………

……………………………………… ………………………………………

……………………………………… ………………………………………

……………………………………… ………………………………………

A: THE TESTS

1. What worked well?

```



```

2. What did not work well?

3. What suggestions do you have for improving the tests?

B: JUDGING ORAL PERFORMANCE

4. Do the speaking tests test what we want to test?

5. If not, what changes need to be made?

6. How manageable are the scoring schemes for the speaking tests?

7. Do changes need to be made to these scoring schemes?

C: THE LOGISTICS

8. Was the time adequate for the tests? *Please tick the appropriate box*

about right ☐   not enough ☐   too much ☐

If you ticked 'not enough', please say whether there are particular items which take longer than expected

9. Are there any features of the organisation of the testing within schools which need attention?

10. Is any additional information required before visiting schools?

D: OTHER COMMENTS
If you have any other comments on these or other aspects of the tests, please add them here or on a separate sheet, if necessary

APPENDIX 4: ASSESSOR FEEDBACK

P7 INDIVIDUAL TASKS

TASK 1	COMMENTS
What worked well	*Worked well when pupils were asked to ask questions to native speaker.* *Worked especially well: children came up with a lot of language as it was recent in their mind. Confidence building!* *The pupils settled down quickly with the opening task asking what they had done recently. It worked well not having the most able pupils first as this allowed the NS time to assess what was meant in that particular context by having covered certain topics [and allowed them to push the able pupils further when their turn came].* *A good way to start in a relaxed way. Most of the children enjoyed asking questions after having answered what was well-rehearsed in class.* *Asking pupils to talk about what they had studied in French most recently helped them remember a number of words and phrases, though English was also used a lot.* *Known, familiar language was spoken well, with enthusiasm. I was impressed by pupils' willingness to participate and to speak.*
What did not work well	NONE
Suggested improvements	*A review of how pupils may be encouraged to give examples of language encountered*

TASK 2	COMMENTS
What worked well	*Worked well in terms of basic personal language - although a change in the order of the questions could throw some pupils. It was often not possible to go beyond simple basic language.*
What did not work well	NONE
Suggested improvements	*Add questions re weather, date, birthdays.* *Children could ask questions of native speaker.* *A review of how questions may be couched/varied/enlarged and the role played by exemplification.*

TASK 3	COMMENTS
What worked well	*Produced a lot of language although very seldom in the 3rd person*
What did not work well	*Animals and people were not always sufficient in terms of topics.*
Suggested improvements	*More options (e.g. classroom objects, parts of the body)* *More alternative drawings* *Could there be another way to encourage use of the 3rd person singular or are the majority of P7 pupils unable to use it?*

TASK 4	COMMENTS
What worked well	*All children could give good feedback whatever their ability. I also managed to make them give feedback in French: 'Yes, you are right! He has 3 brothers! How did he say that in French again?' It was an excellent task!* *Listening to stories in French and reporting back in English seemed to work well, although it often appeared to test the pupils' memory as much as their comprehension, if not more so.* *Listening done well **with prompting**.* *On the whole this was done well by nearly everyone. Most of the points were eventually offered by pupils, often after memory jogging and repetition of some phrases.*
What did not work well	*Stories were too long and content-laden to be remembered after just one hearing. In a class context they would expect to hear these **at least** twice. In fact, with a bit of prompting they managed about 70% or more of the key points, which I found impressive.*
Suggested improvements	*Marking sheets listing the key points so that the assessor can tick points as they are made. This would give us a quantifiable score instantly. Revising the tests would also help. Second reading? Shorten the stories by half and offer more. Add a reading task, or about the same length as the original stories, with oral resumés offered by pupils after reading a story on card.*

TASK 5	COMMENTS
What worked well	*Almost entirely correctly done, sections 1 and 2 especially. A few pupils missed the third one. Often a key word got it immediately. I think this section should be expanded in future.*
What did not work well	*There were not enough domains - if school had not covered 'food' or 'animals' the children could only guess. The little dialogue should have covered more vocabulary and perhaps different location (e.g. pet shop...).* *Dialogue tended to be understood by almost every pupil. We should perhaps change the format in order to make it more discriminating.*
Suggested improvements	*Okay, but I would extend it to 5 or 6 items.*

TASK 6	COMMENTS
What worked well	*We had to use 'food' almost always. We found they had done so little and couldn't tackle clothing or weather. We did animals for Task 3, so we were forced to use 'food'. Unfortunately, they didn't know that very well either and there was quite a lot of sheer guessing evident.*
What did not work well	*Caused something of a problem in terms of the vocabulary domain - in some schools pupils were quite ignorant of each of the alternatives. Also our selection of items did not always coincide with what they had been taught. It was also important to go over what the drawings represented in English.* *On the whole our test should show up the limitations of their knowledge in the different topic areas. This is not a fault of the test.* *I still feel that the grid-listening tasks do not work particularly well. They are either too easy or pupils just don't know the vocabulary, and there is nothing to help them guess. it is also difficult to stop them from copying from each other!*
Suggested improvements	*Clearer visuals for certain clothing and food items*

TASK 7	COMMENTS
What worked well	*Time consuming but enlightening!* *Worked well particularly the metalinguistic discussion. This was quite discriminating.* *I broke this down into two parts giving each pupil major responsibility for one set and the chance to assist – if requested – with the other set. The first part involved meanings of the sentences. Very few pupils could tell me about kauft or nimmt and some of the nouns were also not known: Rock, Kaminchen, Bluse. In a few cases klein and schwarz also not known. Once we had established the meanings of the sentences by a bit of give and take, we looked at the following areas: nouns, adjectives, verbs and articles, gender indicators. Most of this went well, although there were varying degrees of confidence v confusion:* • *Best known were nouns, although capital letter rule was not well known.* • *Verbs were usually known (function – **not** meaning) and identified. Kauft and nimmt okay. Lots of trouble getting hat to be a verb. It's not really a good example of their definition – a **doing** word.* • *Adjectives were often offered as adverbs, but on the whole it was okay eventually.* • *Endings meant nothing to anyone. Because of the nature of the sentences it was difficult to bring out the article/gender issue from the examples. We really needed clear eines/eine/ein or (even better) der/die/das to discuss this issue.* • *The subject/object issue was not available at this stage.*
What did not work well	*In terms of understanding, this was dependent on what pupils had covered.*
Suggested improvements	*There could perhaps be a wider variety of sentences to choose from for Task 7.* *If the use of articles to highlight gender is an important issue, the sentences need changed to provide a clear example of this. Names – lots were not sure if Hansi was a boy or a girl.*

P7 SPEAKING TESTS

Do they test what we want to test?	Most testers felt that they did. One commented: Difficult to say as the results were so patchy! But I would say yes on the whole. The fact that there were 3 different speaking activities allowed me to make the children produce as much language as they could. Task 1 was particularly good in this respect. The tasks set offer different approaches to test the command of the language e.g.: in the social chat, problems occurred with the 3rd person: 'tu as une soeur, comment s'appelle-t-elle?' Answer 'je m'appelle' or simply 'Anna'. Task no. 3 offers the possibility to use the 3rd person and test the children on that. I am not convinced that the tests succeeded **sufficiently** in eliciting the language that pupils did often seem to have. They may, for example, have more **social chat** language than the specific Task 2 questions managed to unlock. Perhaps it is not yet clear that we **know** what we **want** to test. Some pupils used very elaborate questions fast and fluently - but this was formulaic. I think so. It was an amazing experience to find out just how much the children could say, how keen they were to say it, and how relaxed they were with all the tasks. **I do** think they should be given the sentences to read – what is so magic about seeing words and reading them in S1 and not in P7? Their responses to the MLA discussion showed confidence. They can talk about language in this way. Not upset by the task as S2 pupils were. Tasks 1, 2 and 3 did not allow us to ascertain how much vocabulary could be recalled by individual pupils, their grasp of basic personal lang. (comprehension and production) and their ability (or not) to use the third person. However I did sometimes feel that we were not 'getting at' the full extent of their knowledge. I particularly liked Tasks 1 + 2, especially the way the 2 matched and followed on naturally. We invariably found that they could do far less than the official list said they could do. Have to avoid a dominant voice in each pair, but pupils tended to support each other, rather than compete. The questions were offered alternatively to pupils to avoid pupil B always copying the answers of the first pupil. We were generally able to cover the main 'social areas' – name, family, where they live, age, birthday etc and a few pupils took the opportunity to expand on their answers. Many of the pupils struggled to ask us questions, claiming they had not done it that way round.
What changes need to be made?	Probably 4 pairs of pupils per school is enough to provide the information we need. A review of how pupils may be encouraged to give examples of language encountered (Task 1). A review of how Task 2 questions may be couched/varied/enlarged and the role played by exemplification. A review of the role of 3rd person language and its place in current FL use in Primary. Could there be another way of assessing their ability to produce questions? Also their ability to respond to instructions since these were not implemented in a systematic fashion. We tended to use mostly the animal sheet for task 3. Schaf is not known – most only know about four **pets**. Hund, Katze, Hase (not Kaninchen) and (Gold) fisch. They managed name and age, but very few offered any more even after prompting. Size, colour etc ... Pronouns or der/die/das mostly not used or wrong. Change title to Can you **tell us about** any of these animals. Change Schaf to Igel or Schildkröte or Pferd. The idea of the test, however, is good and it could work well.

How manageable are the scoring schemes?	*The scoring scheme was, I thought, a bit slack – we started off with an 'overall rating for pupils on tasks 1-6'. This was far too general and did not take into account the variations in children's performances. The second one 'note sheet' involved an awful lot of writing for the N.N.S. - because the time allocated was so short, we did not have time to properly go over all these notes to agree on an overall judgement on the children's performances. I am slightly worried that the only feedback we will have (together with the tapes, of course!) to analyse the pupils performance will be a long list of note taking, written under a lot of pressure and in a minimum of time.* *The revised schedules for note-taking were okay, but I tended to use them to note down what individual pupils said at a given time (to assist the listening to the tapes). I didn't do a separate commentary – this may be a personal thing – the role of the NNS needs to be clarified here – as an interlocutor/participant. I feel the NNS needs to keep eye-contact with pupils - not be scribbling the whole time.* *I found note-taking for 2 pupils very difficult. I began by noting their utterances but with the speed some of the children were speaking and asking questions this became impossible. I can see that without making a note of who says what, following the tape and knowing when Pupil A and Pupil B are speaking especially when same sex, might be impossible.* *The initial scoring schemes for P7 were quite difficult to manage – particularly if 3 pupils were assessed at one time.* *The note-taking system worked quite well – although perhaps it would have been better to list points for each task e.g. Task 1 – type of words/phrases mentioned: nouns, verbs, adjectives, set phrases etc; range of words/phrases, etc.* *More training on the note-taking system (NNS tasks 1-6 NS task 7) would have been desirable, had time permitted.* *Two people commented that they did not use the scoring schemes.* *The original scoring scheme had too many different aspects in one category e.g. 'takes initiative and is able to use language ...'. Children sometimes took the initiative but were not able to use appropriate language.* *I have tried to link up as much of a transcript as possible of the talking tasks. Together with the cassette it should give a clear picture. There was considerable variety in the quality of the responses from one candidate to another, but all the children were very enthusiastic and friendly and seemed to be enjoying the experience. There was great willingness to experiment and invent language, but at times it degenerated into more English than German. It might be useful to draw up a scoring sheet with criteria etc as an objective reference point for the NNS to assess on the spot.*

| Do changes need to be made | Perhaps an overall rating sheet to tick with co-assessor to accompany the note taking sheets - (a different one for each task). Perhaps, half a page at the end of 'note taking' page, but for each activity. (I don't know if that is very clear! Sorry! I have enclosed an example). The scoring scheme should have been better explained at the training day! - I only discovered the first 'xxx' ('overall rating') when I arrived at my first school!
Role of NNS needs clarifying + guidelines about what should be noted – these will presumably derive from the analysis.
I am not sure how the note-taking could be made easier. I feel I missed quite a lot by looking at the paper and not always at the children – then I would see a gesture or a flash of understanding and find I had left an utterance half-written! It was absolutely essential for me to have Pupil A sitting on the left, facing my sheet for Pupil A, and Pupil B on the right or I wrote in the wrong column. No doubt practise would make things easier. I was not aware of any 'scoring scheme'. Have I missed something?
It is clearly now necessary to establish criteria and rating scales and apply them to the cassette recordings as far as possible.
The second version is more a note - taking sheet than a scoring sheet. During the visits I developed my own sheet.
Yes – clarification of what should be recorded.
I don't know if I will be involved in assessing the cassette recordings against such a written definition of a scoring scheme, but it would not be too difficult to carry out such a task. If I were scoring on a 1-5 basis (with 1 being 'high') I think most of the pupils in my three schools would have been in the range 2-4. Most coped with tasks 1 + 2 reasonably well and achieved basic communication with task 3. |

S2 INDIVIDUAL TESTS

TEST A: Reading/Writing	COMMENTS
What worked well	*All pupils appeared to understand the rubrics for the Reading and Writing test, and settled to the tasks quite readily.* *Reading and writing tests well perceived by the pupils.* *The reading tasks seemed to work well, as long as pupils queries regarding what they had to do were responded to – quietly to each individual pupil. Giving an answer to the whole class tended to lead to disruption.* *Reading and listening tests were appropriate and worked well. Pupils had no problems to follow instructions on the task sheets.*
What did not work well	*Giving the pupils breaks between batches of tasks in the R/W test caused many pupils to become restless. However it did give the assessor an opportunity to explain the tasks verbally – a real necessity in most cases. The breaks also gave some pupils an opportunity to write more/take the time they needed without feeling rushed. I'm not sure which method is best.* *Task 6 had to be explained very carefully.* *Task 7: the instructions for the 3 questions were not really clear enough - some pupils gave a phrase or a sentence. Nos. in words or digits.* *Some pupils found the change from answering the reading tasks in English to writing in French/German difficult possibly because they have not been asked to write in French/German. Those who could write French/German had no problem with understanding the task.*
Suggested improvements	*Shorten the test, by eliminating the breaks, when pupils become bored.* *I would reduce the number of short breaks in the R/W test as they can detract from pupil concentration.*

TEST B: Listening	COMMENTS
What worked well	*The listening tasks were on the whole managed by the pupils.* *The listening test tended to go quite well since the fact that the tape was not stopped between items made pupils concentrate harder.* *Reading and listening tests were appropriate and worked well. Pupils had no problems to follow instructions on the task sheets.*
What did not work well	*Listening tests too long and sometimes difficult to implement (due to discipline mainly).* *The numbering tasks (1 & 2) in the L test caused some pupils problems. Also pupils complained about the speed in some cases (Tasks 4 & 5).* *Task 1 (clothes) was very fast, difficult, partly because they hadn't done clothes, but also format of boxes quite difficult. In task with leisure activities the German gives the help of mentioning ein hund. No mention of chien in French!* *Topics were unknown to some classes (Task 1 - listening).* *The first task of the L Test is based around clothes and pupils in one school had not covered this. It can lead to a bad start to the L. Task 2 was understood better.*
Suggested improvements	*In the vocab. identification tasks (such as Task 1 clothes) do **not** have the item to be identified as the first word spoken. The pupils concentrated on the last word of the very short sentence and missed the key word.* *Review length and level of difficulty of listening tests.*

TEST C: Speaking	COMMENTS
What worked well	*The joker system seemed to work. Pupils really worked to get the answer themselves (however a clearer notion of how much help should be given would have been helpful).* *Pupil behaviour was excellent during the speaking test – individual rather than group assessment.* *The speaking tasks varied enormously from school to school. With some pupils the tasks with both the NS and NNS worked well, with others just some of the tasks.* *In the speaking tests the assessor had to show some flexibility, since some topics were unknown to pupils (e.g. Task 3, No. 9, 10).*
What did not work well	*I found the procedures for the joker cards during the bridging activity difficult to adhere to e.g. noting down the words given.* *Task 1: Very few managed to talk for the whole minute. They did not always realise that 1, 2 & 3 were suggestions to help them to compose a whole talk and tended to answer the questions.* *Task 4 was a good challenge when they knew the vocabulary but it was impossible to use future and past tenses or full sentences.* *Task 9: quite a few pupils struggled and expected help from the assessor. The bridging activity was rushed.*
Suggested improvements	*Give advance notice of the Bridging Activity to schools, so that pupils would be less hostile.* *Make clear to assessors how help from the token card is best recorded on the speaking booklet. This token seems a good idea but the less co-operative pupils during the Bridging Activity were not taking full advantage of the system.* *Have a separate sheet for noting the vocab. pupils ask for during the test rather than writing on cards. I would keep the 'joker card' idea, however, as it helped to relax the pupils.* *Task 1: Instead of writing suggestive questions, we could maybe give them an example of a 1-minute talk - text in English with the specific instruction that they are not to translate.* *Task 4: Nothing wrong with the test – Maybe a bit more training should be required from the teachers on how to comment pictures using sentences, actual facts and imagination.*

METALINGUISTIC DISCUSSION TASK	COMMENTS
What worked well	NONE
What did not work well	*The metalinguistic discussion with the NNS only worked well with those pupils considered 'High level'. Otherwise it failed and the pupils went away feeling bad at the end of the test.*
Suggested improvements	NONE

S2 SPEAKING TESTS

Do they test what we want to test?	*Overall, yes they do. They allowed me to test the pupils on various aspects of the language, but at the same time, did not force me into struggling with low ability pupils to get some 'sophisticated' dialogues from them.*
	I can really only judge tasks 1-4 from the bridging activity, however I feel that tasks 1, 3 & 4 were quite discriminating, while task 2 did not really provide much opportunity for pupils to show us what they could do, since the type of transactional language required was beyond the reach of almost all pupils I saw. This may in itself be a finding, but I feel that the amount of assessor support required for this task undermines the analysis of pupil performance. Tasks 5 & 6 did give a very interesting picture of pupils' awareness of the sound/symbol relationship and their awareness of aspects of language. I also started asking overt questions after task 5 about French/German pronunciation. This proved very interesting. Task 6 was quite discriminating. It was possible to have a quite in-depth grammatical discussion with some pupils while keeping it very basic with others. I also found it interesting to ask pupils to point out the pitfalls of the lang. or aspects that it was important to know in order to excel.
	*I am not sure what the purpose of the MLA task 6 served although I found it interesting to compare pupil performances in P7 French and S2 French. Pupils in French **had** probably not been trained to read aloud. Two schools had not learned the alphabet, so had problems with spelling the name. Is this telling us what the pupils are taught or not taught? Some P7 French pupils could spell in French! I think we should be finding out whether children can read aloud and know something about sound system, both at S2 and P7.*
	Tasks 2 and 3 worked well. For task 4 a different visual could be chosen - with more known activities for example. Some of the topics were unknown to most of the pupils (e.g. weather, describing people).
	Broadly, yes, but I think only the analysis of the assessment data will help us answer that question. I was happy with the mix of prepared and 'spontaneous' activities. In the prepared dialogue, I felt we were testing a particular structure in one of the items ('where shall we meet'), which the pupils didn't know and couldn't get round even with help during the bridging activity.
	In general yes. I just want to make a few small points from my observations:
	*In **Task 1** the children dealing with Topic 2 did less well than those with Topic 1.*
	*In **Task 2** the majority of children had to be given some help how to deal with the dialogue.*
	*In **Task 3** only a tiny minority could refer to anything outside their very personal sphere e.g. to Father, Mother, house, garden, work etc.*
	***Joker Card** was of little use as most of them didn't quite understand how it worked.*

What changes need to be made?	*Most of the S2 pupils found my time of discussion in MLA task 6 threw them off course (my fault), but I got the impression that they understand what a verb is when they are doing a familiar activity and this was all unfamiliar territory. Apart from a few exceptional children, this task made them hesitant and nervous and they made wild guesses. Do we need this task?*
	The tests could allow a little bit more freedom. Especially for task 3, I sometimes felt restricted with some of the high achievers. Task 2 (role-play) is quite artificial! I would suggest the removal of task 2 and its replacement with a spontaneous functional task or simply its removal.
	In terms of task 6 I found that the open-ended question 'what kind of word is ...' for nouns/verbs etc did not work well. It was better to ask them to say what it was first and then get them to point one out.
	Since it was an assessment I felt it better not to correct pupils overtly, although it

was possible to get them to think twice about certain answers, by giving them an example of a noun etc and asking them to say what kind of word it was.

For the prepared dialogue, change the question 'where shall we meet/ask when to meet'. In the task involving the visual, I wonder whether pupils should be allowed to describe in their own words first, before the Q + A begins by the NS - given if it is just single items of vocab., simple phrases. Then the NS could follow up with some questions. (I only observed this activity from a distance, however!)

Most of the S2 pupils found my time of discussion in M LA task 6 threw them off course (my fault), but I got the impression that they understand what a verb is when they are doing a familiar activity and this was all unfamiliar territory. Apart from a few exceptional children, this task made them hesitant and nervous and they made wild guesses. Do we need this task?

Most of the S2 pupils found my time of discussion in MLA task 6 threw them off course (my fault), but I got the impression that they understand what a verb is when they are doing a familiar activity and this was all unfamiliar territory. Apart from a few exceptional children, this task made them hesitant and nervous and they made wild guesses. Do we need this task?

The tests could allow a little bit more freedom. Especially for task 3, I sometimes felt restricted with some of the high achievers. Task 2 (role-play) is quite artificial! I would suggest the removal of task 2 and its replacement with a spontaneous functional task or simply its removal.

In terms of task 6 I found that the open-ended question 'what kind of word is ...' for nouns/verbs etc did not work well. It was better to ask them to say what it was first and then get them to point one out.

Since it was an assessment I felt it better not to correct pupils overtly, although it was possible to get them to think twice about certain answers, by giving them an example of a noun etc and asking them to say what kind of word it was.

For the prepared dialogue, change the question 'where shall we meet/ask when to meet'. In the task involving the visual, I wonder whether pupils should be allowed to describe in their own words first, before the Q + A begins by the NS - given if it is just single items of vocab., simple phrases. Then the NS could follow up with some questions. (I only observed this activity from a distance, however!)

How manageable are the scoring schemes?	*Very manageable however, it was sometimes difficult to make the difference between their language performance and the completion of the task, e.g.: what about if they fulfilled all the requirements for the task but with poor grammar and/or pronunciation.* *I administered Tasks 5 and 6 and used personal judgement rather than specific criteria for allocating a Grade 1-4. I presume my scoring will be similar to other assessors?? It was manageable.* *Manageable! A lot easier to complete while giving the actual tests, compared to the initial scoring scheme.* *The non-native speaker score sheet was quite straightforward.* *A general scope could be given and notes taken – although the latter task was quite tricky while actually implementing the tests, and tended to be jotted down quickly after the pupil had finished – thus limiting the time available for making comments. Alternatively I made the pupils aware that I was writing down lots of positive aspects of what they were saying, thus enabling me to take notes during the test, but making the test itself take a little longer.* *Scoring schemes could be shortened for each section in order to make them more manageable.* *I can only speak for the last 2 tasks - reading aloud and metalinguistic discussion. I found the rating scales for the reading aloud plus the ML discussion difficult without criteria descriptions and tended to ring 2 in some cases. The notes I made were really about what would be audible on the tape in any case.* *They are all right but it would be good if there was a short interval between each child to reflect on the scores given before moving on.*
Do changes need to be made?	*No. But it might be useful to be given the opportunity to review each pupil's results afterwards, i.e. with the hope checking what we thought at the time.* *Perhaps a more detailed score sheet should be used to facilitate the note-taking process e.g. for the reading aloud different categories:* *Spelling of the word ✓ or x* *Number ✓ or x* *Silent letters ✓ or x* *x or ✓ words mispronounced/ pronounced well etc.* *For task 6 again different categories to complete:* *Noun Pupil's comments on what it is, example given* *Plural ✓ or x (depending on whether pupils have any knowledge of them etc.)* *Gender ✓ or x* *Article ✓ or x* *Verb Example _____* *Endings ✓ or x* *Tense ✓ or x* *Infinitive ✓ or x* *Adjective Example _____* *Agreement ✓ or x* *General comments: _____* *Criteria descriptors will be needed for the reading aloud/ML discussion tasks.* *Not really, except for some time between children.*

APPENDIX 5
TEST OUTLINES

P7 Test -   **Tasks1-6, Native Speaker Assessor**
Task 7, Non-native Speaker Assessor

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 1: Speaking - Discussion of topics studied recently in the modern languages class	Vocabulary recall exercise.		Pupil-driven.	Pupil-driven.	Pronunciation and intonation. Speaking to establish relationships with others.
Task 2: Speaking - Question and answer session with the native speaker assessor	Social chat: answering and asking questions in the target language.		Personal language: name, age, home, family, nationality, date, weather, likes, dislikes.	Communication strategies, expressing likes, dislikes, interrogatives.	Pronunciation and intonation. Listening to establish relationships with others. Speaking to establish relationships with others.
Task 3: Speaking - Description of a colour visual	Description of a colour visual, comprehension of prompt questions.	*Prompt colour visual: a selection of different animals or people of varying appearance with names and ages given.*	People, pets, name, age, clothes, size, colours, physical descriptions.	Describing, use of third person, use of adjectives.	Pronunciation and intonation. Speaking on a topic. Listening to establish relationships with others.
Task 4: Listening – Understanding the subject and message of a short narrative	Understanding a simple narrative. Recalling and summarising the narrative in English.	'Live' input: a short text (from a choice of 4 texts) read aloud twice by the native speaker assessor to each pupil in turn.	Personal language.	First and third person verbal forms.	Listening for information.

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 5: Listening – Understanding short dialogues	Recognition of setting based on an understanding of 4 short dialogues between 2 speakers via selection of the 4 correct settings from a possible 6.	'Live' input: 4 short dialogues between the 2 assessors, each repeated twice.	Places in town: café, zoo, school, railway station. Numbers, prices. Transactional language. Classroom language.	Transactional language: likes, dislikes, asking questions, imperatives.	Listening for information. Classroom language.
Task 6: Listening - Vocabulary recognition exercise	Vocabulary recognition via the identification of 5 visuals mentioned in 5 short sentences.	'Live' input: 5 short sentences on a single theme each repeated twice by the native speaker assessor.	A choice of 5 themes: clothes, food and drink, parts of the body, pets, weather.	Third person verbal forms.	Listening for information.
Task 7: Reading, translation and metalinguistic discussion	Reading aloud. Understanding of sentences read. Discussion in English about language structures contained in the sentences.	3 short sentences (from a choice of 6) drawn from material used in Task 6.	Clothes, food and drink, pets.	Recognition and understanding of certain grammatical features: nouns, articles, gender, plural, verbs, adjectives - agreement, position.	Pronunciation and intonation. Pronunciation and the written word. Reading for information. Knowledge about language.

S2 Test A: Reading and Writing

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 1: Reading – Discourse connection	Reading 6 stimulus questions and matching them to an appropriate response from a list of 12.	Classroom phrases of a sort that are normally written on the blackboard, the classroom wall or in worksheets.	Personal language: name age, residence, food, leisure activities.	Opinions, discourse, subject pronouns. Verbal forms - present tense.	Reading for information.
Task 2: Reading – Vocabulary recognition	Selecting one of 4 possible phrases to match a given visual. (4 visuals). Multiple-choice exercise.	As above.	Pet animals, weather, seasons.	Descriptions, adjectives, possessives.	Reading for information.
Task 3: Reading – Understanding the message of a number of short texts.	Reading 3 short passages, each of 2-3 sentences, and finding specific information by answering open-ended questions in English.	Short, descriptive sentences of a sort that are written on the back of photo or postcard.	Places in town, directions	Prepositions, verbs indicating location, imperatives.	Reading for information.
Task 4: Reading and Writing – Extracting specific information from a number of more extended texts drawn from an authentic source	Reading 3 paragraphs of approximately 5 sentences each, including some unknown words. Answering specific questions in English. Vocabulary retrieval - identifying and copying French words from the text.	Short, authentic paragraphs from the Internet, written by French-speaking youngsters. The texts are part-description, part-opinion and address an Internet teenager readership.	Cinema, films, film-stars.	Describing people and emotions, adjectives, opinions, some vocabulary inferable from English cognates.	Reading for information. Copying.

185

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 6: Reading and Writing – Problem-solving and vocabulary recall at the single word level	Reading and understanding 6 prompt sentences in order to write specified subjects - single elements of vocabulary without articles - into the correct slot on a partially completed timetable.	Partially completed morning timetable.	School subjects, days of the week, parts of the school day.	Problem-solving, spelling.	Reading for information. Writing from memory.
Task 7: Reading and Writing – Labelling/listing, writing responses to familiar questions.	Vocabulary recall: list 6 nouns and articles representing classroom objects, using visual stimuli if desired. Writing responses to stimulus questions of a general classroom nature.	Visual stimuli depicting classroom objects, bordering a partially completed list. Stimulus questions regarding date, time and weather with fixed visual stimuli.	Classroom objects, classroom language, date, time, weather.	Spelling, gender, articles, sentence structure.	Reading for information. Writing from memory.

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 8: Reading and Writing – Text completion	Completion of 14 gaps in 12 sentences. For the first 6, 10 possible words are given from which to select. The words must then be copied correctly. For the second 8 no support is given. This assesses the pupils' knowledge about language, if not awareness of that knowledge.	Set of incomplete sentences, in the form of a letter, conveying personal information about self and family, of a sort that can be sent to a pen-friend or e-mail correspondent.	Personal language.	Knowledge of vocabulary. Spelling. Knowledge, if not awareness, of certain grammatical features: articles, gender, plural, pronouns, possessives, adjectives, verbal forms (present tense).	Copying. Writing from memory.
Task 9: Writing – Open-ended writing	Write 6-10 sentences, based on visuals and own ideas, describing weekend activities.	6 visual stimuli depicting the notion of time, weather and different activities – to be used if desired.	Weekend activities.	Ability to cover the task set, and to produce a certain volume of language, which demonstrates both linguistic range, and control of the grammatical system.	Continuous writing.

S2 Test B: Listening - Audio-recorded material, Native Speakers

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 1: Listening - recognition of single words	Vocabulary recognition via the identification of 6 items of clothing - mentioned in 6 very short sentences - from a possible 10 written items in English.	Short single sentences from 6 speakers. Repeated twice.	Clothes.	Descriptions, adjectives.	Listening for information.
Task 2: Listening - recognition of short phrases	Recognition of short phrases from slightly more extended input via the identification of 6 leisure activities - mentioned in 6 sets of a few sentences - from a possible 10 visuals.	2-3 short sentences from 6 speakers. Repeated twice.	Leisure activities.	Likes, time expressions, adjectives.	Listening for information.
Task 3: Listening - understanding the subject and setting of short dialogues	Recognition of setting based on an understanding of 4 short dialogues between 2 speakers, via a selection of the 4 correct settings from a possible 7.	4 short dialogues, each between 2 speakers. Repeated twice.	Places in town: café, zoo, school, railway station. Numbers, prices. Transactional language. Classroom language.	Transactional language: asking questions, buying items, likes, dislikes, imperatives.	Listening for information. Classroom language.

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 5: Listening – understanding the subject and message of more extended monologues	Answering specific questions in English about personal information based on 4 more extended extracts from a single speaker.	An extended monologue from 1 speaker, split into 4 separate extracts. Repeated 3 times.	Personal language: name, alphabet, age, numbers, home, family, pets, jobs, hobbies, school subjects, days of the week, time, rooms in the house, home, weather, seasons.	First and third person verbal forms in present tense, descriptions, likes, dislikes.	Listening for information.

S2 Test C: Speaking - Tasks 1-4, Native Speaker Assessor
Tasks 5-6, Non-native Speaker Assessor

Task	Task-type	Text-type	Theme/Content	Function/Grammar	5-14 Strand
Task 1: Prepared talk on a topic	Short prepared talk on a given topic.	Prompt sheet making suggestions in English as to the possible content of the talk.	Hobbies and freetime/daily routine and school.	Ability to cover the task set, and to produce a certain volume of structured input, which demonstrates both linguistic range, and control of the grammatical system.	Pronunciation and intonation. Speaking on a topic. Asking for support.
Task 2: Prepared semi-structured dialogue	Short semi-structured dialogue with a given aim: to organise an outing.	Prompt sheet in English outlining the general content of the dialogue, and providing visual stimuli to facilitate preparation.	Arranging an outing.	Transactional language: greetings, places, leisure activities, time, interrogatives.	Pronunciation and intonation. Speaking to establish relationships with others. Speaking on a topic. Asking for support.
Task 3: Part spontaneous, part prepared question-and-answer session	Social chat: answering and asking questions in the target language.	Prepared list of questions to ask the assessor.	Personal language: name, age, birthday, home, family, pets, leisure activities, school, food and drink, likes, dislikes.	Communication strategies, expressing likes, dislikes, interrogatives.	Pronunciation and intonation. Listening to establish relationships with others. Speaking to establish relationships with others.
Task 4: Spontaneous description/ narration of composite colour visual	Description/narration of a colour visual after a very short preparation time. Comprehension of prompt questions.	Composite colour visual of a beach scene.	Seaside vocabulary, leisure activities, weather, time, clothes, colours, physical appearance.	Describing, adjectives, use of third person verbal forms, recognition and/or use of different tenses.	Pronunciation and intonation. Speaking on a topic. Listening to establish relationships with others.
Task	**Task-type**	**Text-type**	**Theme/Content**	**Function/Grammar**	**5-14 Strand**
Task 6: Metalinguistic discussion	Discussion in English about language structures contained in the paragraphs and	As above.	Personal language, metalanguage.	Recognition and understanding of features of pronunciation and	Knowledge about language.

foreign language learning in general.			certain grammatical structures: noun, plural, gender, article, verb, endings, infinitive, tense, adjective - agreement and position.

CONTINUOUS WRITING EXEMPLARS OF GOOD PERFORMANCE

FRENCH

SCHOOL 4, PUPIL 1 (Male)

Le weekend, quand il fait beau, je joue au tennis. Je joue à les courts à X avec des autres garçons. Aussi, je nage dans notre piscine et promaine ma chienne.

Quand il fait mauvais, je fait mes devoirs, régarde la télé où écouter la radio. Le soir, je régarde la télé après le diné, ou fait des jeus avec ma famille. Aussi, je visite mes copains le samedi et vendredi soir. Aussi, je vais au cinéma avec mes copains et copines, ou nage à la piscine a Y.

J'aime régarder les films, au télé ou au cinéma. Mes films favorits sont: 'Full Metal Jacket', 'Independence Day' et 'Schindler's List'.

J'ai neuf cochons d'inde, et je donner au manger le matin (mon frère donner au manger le soir) et je trove des verts pour ils.

J'ai travaille dans le jardin, par example donner au manger au ois, canards et poules.

Volume
148 words in length
Correct words 119
Different errors: 20, mainly verbal - often phonetically correct

Range and Grammatical control
Different correct nouns 23 (of 28)
Different correct verbs 11 (of 15)
Different correct adjectives 3 (of 4)
Different correct pronouns/articles 11 (of 11)
Different correct prepositions/connectors 9 (of 9)
Different correct time expressions 7 (of 7)

SCHOOL 4, PUPIL 5 (Female)

Le vendredi soir, je reconte avec mes amis et nous allons au cinéma. Le samedi matin à dix heures et demi je jouer au tennis avec mon petit frère. Après ça nous avon déjeuner avec nôtre parents et nôtre soeur.

L'après-midi je fait mes devoirs. Bof!

Le soir je regarde le télé. Je regarde 'Have I got news for you?' et 'Invasion Earth'. Le dimanche matin je vais à la piscine avec mes amis. Le soir j'ai diner avec toute ma famille.

À huit heures je regarde 'Ballykissangel' au télé et à dix heures je me couche.

Volume
90 words in length (excluding English words)
Correct words: 80
Different errors: 10 (mainly verbs)

Range and Grammatical control
Different correct nouns: 19 (of 19)
Different correct verbs: 4 (of 9)
Different correct adjectives: 2 (of 2)
Different correct pronouns/articles: 7 (of 7)
Different correct prepositions/connectors: 6 (of 6)
Different correct time expressions: 7 (of 8)

GERMAN

SCHOOL 15, PUPIL 1 (Male)

Hallo Klasse! Ich heiße A! Am Wochenende spiele ich gern Tennis um der W. Tennis Centre. Es ist Wonderbar! Meinem Trainer heißt S. R. und ist sehr kommisch! Ich habe viele Freunde um meinem Tennis Class.

Ich spiele Fußball am Wochenende. Es ist nicht so gut as Tennis aber es ist gut.

Ich höre gern Musik! Es ist fantastic. Ich gern Pop Musik, Rave Music, Tanzemusic und Soul Musik.

Am Wochenende spiel gehe ich in der Stadt einkaufen! Es ist sehr gut! Ich gehe mit meinem Freunde, C. Ihr gehe gern einkaufen.

Manchmal spiele ich mit dem Computer. Es ist eine Pause für meine Hausaufgaben. Hasaufgaben ist langweilig! Das ist nicht sehr gut!

Gehe ich gern mit meinem Hund spazieren. Das ist langweilig! Aber R., der Hund, spielt gern.

Am Wochenende treffe ich mit meinem Freunde. Ihr gehe ins Kino oder einkaufen in der Stadt.

Ich lese gern! I lese Bucher gern! Es ist fantastik! Ich gern R. L. S. und C. P. Sie ist super! Danke, Klasse! 'Bye Tschs Auf Wedersehn!

Volume

154 words in length (excluding English words)

Correct words

Different errors:

Range and Grammatical control

Different correct nouns: (of)

Different correct verbs: (of)

Different correct adjectives: (of)

Different correct pronouns/articles: (of)

Different correct prepositions/connectors: (of)

Different correct time expressions: (of)

SCHOOL 19, PUPIL 21 (Female)

Hallo! Aus dem Wochenende Ich mache meine hausaufgaben. Normaleweise, das ist Mathe, Chemie oder English und dann, Ich sehe fern. Ich sehe am liebsten 'FREINDS' Es ist sehr schön. Am nechstes Tag gehe ich schwimmen um halb zehn.

Dann ich gehe zur cafe mit meinem fruende, J. Ich gehe einkaufen. Dann, gehe ich ins kino oder in der Jugendzentrum. Manchmal, gehe in die disko. Am mechtes Tag, spiele ich Fußball mit meinen Freundin und Basketball oder Straße hockey. Meine leiblingssport ist Straße hockey. Dann, wir spielt mit meinen computer. Ich habe ein 'Nintendo 64'. Es ist schön! Wir spielt 'Mario' oder 'Yosi's Story' am liebsten. Im Sommer wir gehen mit dem hund spatzieren Im Winter, gehe ich mit dem hund spatzieren. Dann, gehe ich zur meinen scloftzimmer und höre ich Musik. Ich höre gern Oasis, Blur und The Verve. Am liebsten höre ich Aqua, Catatonia oder Savage Garden. Dann, gehe ich zur Bett.

Volume

138 words in length (excluding English words)

Correct words

Different errors:

Range and Grammatical control

Different correct nouns: (of)

Different correct verbs: (of)

Different correct adjectives: (of)

Different correct pronouns/articles: (of)

Different correct prepositions/connectors: (of)

Different correct time expressions: (of)