

A comparative study of the effectiveness of the language laboratory in school

P.S. Green, Language Teaching Centre, University of York, England

This is a short report on a comparative study of the effectiveness of the language laboratory in school.* It was carried out between 1967 and 1972 by the Language Teaching Centre of the University of York, England, in collaboration with the German teachers of Archbishop Holgate's Grammar School, York¹.

BACKGROUND

The study began right in the middle of a decade of very rapid expansion of language laboratory facilities from the first installation of a school laboratory in Britain in 1962 to the situation in 1972 when a third of secondary schools teaching languages had laboratories.

It was a time of great enthusiasm, when there was much discussion of the merits of different equipment and much attention given to the programming of material for the lab. The effectiveness of the lab was not so often called into question for its advantages seemed self-evident. The research evidence in 1966 when we began to plan our study was scant and conflicting.

Keating (1963) had found significant differences favouring no-lab groups, whereas Lorge (1964) concluded that the more time spent in the lab the greater the gains in speech, and the more complete the equipment (audio-active-record vs. audio-active) the better the results. The massive Pennsylvania Project, which involved a comparison of three 'laboratory systems' (classroom tape-recorder, audio-active laboratory, audio-active-record laboratory), had only just begun. It was to report later (Smith 1970) that the lab had no discernible effect on achievement.

All these studies referred to an American context and none was, in our opinion, entirely satisfactory as regards the control of variables other than the language laboratory, and the testing of speech production. There was a need for a carefully controlled study of the effectiveness of the language laboratory in the context of the British school. This need was recognised by the Department of Education and Science who supported the study with a grant of £ 2,100.

*The complete report will be published in 1975 by Oliver & Boyd, Edinburgh.

¹ The team members were:

Eric Hawkins, University of York (research design); John Caley and David Lloyd, Archbishop Holgate's Grammar School, York, and Peter Green, University of York (teaching and test construction); Paul Barber, Birkbeck College, London (statistical treatment); Norman Rea, University of York (IQ and attitude tests).

AIMS

In formulating our principal aim of studying the effectiveness of the lab we were well aware of the danger of confusing method and equipment. The language laboratory does not of itself constitute a method: the taped dialogues, structure drills, etc. which form the "software" of labs may have been inspired by labs, but they can be and indeed *are* used in the classroom with, or even without, a tape-recorder. The lab is a technical teaching aid, providing above all individualisation and intensification of practice. The question was not, therefore, "Is the language laboratory effective?" but "Is the teaching in a given situation made more effective if the language laboratory is used?"

The specific situation in which we proposed to evaluate the effectiveness of the lab was the following: 100 boys in a grammar school were to learn German for five years, from 11 to 16, using modern commercial courses with accompanying pre-recorded tapes; as the school did not (then) have a lab, the tapes would be used with a classroom tape-recorder. How would the learning be affected if part of the practice with tapes took place instead in the audio-active-record lab of the nearby Language Teaching Centre? The school was prepared to keep intact for three years any groups established for comparison purposes. This specific situation might be characterised as *the use of the lab for one period a week with junior forms using commercial tapes*. It is one of a very large number of different possible uses of the lab, but it is one which we considered to be typical of British schools and therefore worthy of investigation. To what extent this situation really was typical will be considered later.

A secondary aim of the study arose from the need to establish groups of pupils of comparable language-learning potential. To do this meant testing for aptitude, since the pupils' previous experience of foreign-language learning was too diverse to serve as a basis for comparison. We decided to make use of a number of different potential predictors of success in foreign-language learning, and to compare later how well they fared when measured against actual achievement.

For the school, 100 pupils meant three groups, and this gave the possibility of another comparison in addition to the lab versus no-lab comparison. We decided, therefore, to compare also the achievement of pupils using a predominantly audio-visual course with that of pupils using a predominantly audio-lingual course (both using a language laboratory).

DESIGN

The basic strategy was to compare the mean achievement of a lab and a no-lab group over a period of three years. Such comparison could only be meaningful if two basic conditions were fulfilled:

- 1) if all other variables between the groups except that of the language laboratory were eliminated or controlled.
- 2) if the achievement compared was relevant achievement, measured by valid and reliable tests.

Three groups were established and designated L, T and N:

Group L used an audio-lingual course with a classroom tape-recorder, and spent one of its weekly language lessons in an audio-active-record laboratory. *Group T* used the same audio-lingual course, also with a classroom tape-recorder, but had no access to the lab.

Group N used an audio-visual course, with a classroom tape-recorder, and spent one of its weekly lessons in the lab.

Thus all groups used a classroom tape-recorder. It was not considered realistic to compare a group using a laboratory with one using no equipment at all. In our experience use of courses with accompanying tapes was by no means restricted to schools with labs: many schools used them relying on a classroom tape-recorder. The frequency and duration of use of the tape-recorder was not controlled in the study but left to the teacher's discretion. The lab, however, was used regularly at a fixed point in the timetable for one out of five weekly periods in the first year, and one out of four weekly periods in the second and third years. The tapes used in the laboratory were the same as those used with the classroom tape-recorder, namely the pre-recorded tapes accompanying the two commercial courses². The courses concerned were both well-known, up-to-date courses which emphasised the spoken language used in everyday situations, limited the use of English, taught grammar largely inductively, and attempted to teach, in addition to the foreign language, something about the country, its people and institutions. The principal differences between the two courses were that the A-V course used filmstrip pictures to contextualise and semanticise language, whereas the A-L course initially used English to help the understanding of new language, and that the A-V course moved rather more slowly in the initial stages than the A-L course. The tapes for both courses contained dialogues for practice,

² Nuffield: *Vorwärts*, Arnold (audio-visual); Creese & Green: *German, A Structural Approach*, Oliver & Boyd (audio-lingual)

structure drills (rather more contextualised in the A-V course), and comprehension exercises.

The following diagram shows the treatment of the three groups (differences are picked out in capitals):

Group	Course material	Language laboratory
L	audio-lingual	yes
T	audio-lingual	NO
N	AUDIO-VISUAL	yes

Cross-group comparison reveals, in the case of L vs T, any differences ascribable to the lab variable, and, in the case of L vs N, any differences ascribable to the course variable. The comparison T vs N is not meaningful because two variables are involved — lab and course.

CONTROL OF OTHER VARIABLES

If the comparison between groups on the variables of language laboratory and course material was to be valid, then all other variables had to be held constant. Other possible variables were pupil potential, teacher potential, and exposure to the foreign language.

Pupil potential

Pupils may have differing success in learning a foreign language for a number of reasons. One pupil may be more intelligent, harder-working, better motivated, have more flair, more experience of language learning, or come from a more encouraging home than another. He may respond better to one particular method of teaching or to the personality of a particular teacher better than another. In a large-scale study, such as that of Keating, these factors may perhaps average out: in a study such as ours, involving only 100 pupils, they must be taken into account. As we did not have to work with ready-made groups of pupils but were free to determine our own grouping, we attempted to assess the following four factors and then establish matched groups accordingly: intelligence, language aptitude, previous experience of foreign-language learning, and parental encouragement.

1. *Intelligence*

A test of verbal reasoning was given to all the pupils: the National Foundation for Educational Research Primary Verbal Test 3. The scores ranged from 95 to 140 with a mean of 117³.

2. *Language aptitude*

Two commercial test batteries were available: the Carroll-Sapon "Modern Language Aptitude Test" (MLAT) and the Pimsleur "Language Aptitude Battery" (LAB). The latter was chosen, mainly because it was the more recent. Two parts of the battery were given (tapes for the other parts did not arrive in time for the testing session). *Part 3* is a test of English vocabulary using multiple-choice synonyms. *Part 4* is a multiple-choice test of language analysis: pupils are asked to construct new sentences in an unknown language (Kabardian) on the basis of model sentences with translations. Pimsleur says that "the scores from these two parts may be combined to give a single verbal ability score". As the LAB was intended for American pupils rather older than ours, we designed three further tests to measure pupils' abilities in three areas relevant to language learning: sound discrimination, vocabulary retention and grammatical analysis. *Sound discrimination* was a multiple-choice test based on minimal pairs of Swedish words, some distinguished by vowel quality, some by tone. *Vocabulary retention* required pupils to learn some nouns and verbs in Danish and Spanish with their English equivalents, using whatever technique they wished. *Grammatical analysis* was a longish test in which forms of Swedish nouns and verbs were presented with English equivalents, and pupils were asked to produce further forms by analogy. It was not a multiple-choice test: pupils wrote words, phrases and sentences. A standardised total of all the aptitude tests was calculated for each pupil. This total was used in allocating pupils to groups.

3. *Previous experience of foreign-language learning*

The pupils were to learn German as their only modern foreign language (some would later add French, but not until the fourth year, when the controlled part of the study was completed). However, for many pupils it was not the first experience of learning a foreign language: about two-thirds of them had learned French in the primary school for periods ranging from three months to four years. Coming as they did from no less than 39 different primary schools, they could be assessed only on whether or not they had learned French and for how long, not on what kind of experience or success they had had.

3 The pupils in the study were thus of above-average intelligence. This is because the grammar school takes the top 20/25 % of the ability range.

4. Parental encouragement

In an attempt to get some information on the important but inaccessible factor of parental encouragement, a confidential letter was sent to the 39 primary-school headmasters asking them to rate on a five-point scale the support given to each pupil by his home. This proved to be a very crude test, as the categories did not discriminate finely enough at the top end of the scale: if parental encouragement does indeed influence achievement in school subjects, then children in a selective school like the grammar school should enjoy above-average parental support. In fact, the bottom two categories were hardly used at all and the scale was in effect reduced to a three-point scale.

(Retrospectively, when the pupils were already in the third year of language learning, a more ambitious attempt was made to evaluate the home background. The method on this occasion was a lengthy questionnaire similar to that used for the Plowden report on primary schools (1967). It was administered by an interviewer to each pupil's mother in the home. In the fifth year an attempt was made to add still further to our knowledge of parental encouragement. This took the form once again of a five-point rating scale, but it was applied this time by the pupils' grammar school form masters and headmaster (independently). To ensure that all the categories in the scale were used, the judges were instructed to use a normal curve distribution.)

In establishing the groups, we attempted to achieve a balanced distribution of all four factors in each of the three groups. Pupils were sorted first according to intelligence, then readjustments were made to ensure an even balance of language aptitude without upsetting the distribution of IQ, and so on. Table I shows that good matching of the factors was achieved:

Table I: Comparability of groups on matching factors

Group		L	T	N
Verbal IQ	Mean	116.7	117.6	116.8
	S.D.	9.58	10.40	8.21
Language aptitude (Standardised total)	Mean	55.4	57.6	56.3
	S.D.	9.44	7.41	8.08
Previous experience (0-4 years)	Mean	1.29	1.32	1.24
	S.D.	1.22	1.17	1.17
Parental encouragement (Grades 1-5)	Mean	1.75	1.71	1.65
	S.D.	0.95	0.80	0.91

Teacher potential

Although all three teachers who were to teach the groups were experienced in the classroom, competent in their subject and of similar outlook as regards method, nevertheless one of them might be more successful with a given group of pupils than another, not necessarily because he was the more skilful teacher but perhaps because the interaction between his particular personality and those pupils created a favourable learning atmosphere.

To neutralise the specific influence of the teacher, we had all the teachers teach all the groups for the same length of time. This was achieved by rotating the three teachers between the three groups at the end of each of the three terms in a school year. Thus, at the end of the three-year study, each teacher had taught each group for a total of one year, and in any one year, each group had encountered each of the teachers in turn. Doubts were expressed by the teachers at the beginning of the study about the unsettling effect this constant change might have on pupils. At the end of the study, all were agreed that their fears had been unfounded and no ill effects on the pupils had been observable.

Exposure to the foreign language

The total amount of time spent in the classroom only, or in the classroom and the lab together, was exactly the same for all groups, namely five lessons of 35-40 mins each per week in the first year, and four lessons per week in

the second and third years. Almost equally important, the lessons always took place at the same time of day in the three groups. The classroom was to some extent an uncontrolled variable. Groups L and T had no permanent base and had to be in different rooms on different days. Group N was not only always in the same room, but that room was in fact the "German Room", where there were posters and wall displays, books and magazines, all dealing with German or Germany. The reason for this was that Group N used the audio-visual course and the German Room had black-out and projection facilities. The comparison of Groups L and T for the language laboratory variable is not, however, affected, since neither group had access (in class time) to the German Room.

An important variable which could not be controlled in this or in any other study of fixed groups is that of the "group dynamic", the unforeseeable and indeed often unseen way in which the pupils in a group interact to produce an atmosphere which may or may not be conducive to learning a particular subject or learning with a particular teacher. How far the results of this study may have been affected by favourable or unfavourable group dynamics it is impossible to say. However, any group dynamic effect was probably mitigated by the fact that the groups described here existed only for German. For the greater part of their time in school the pupils were in forms, each of which contained members of all three German groups.

ATTAINMENT AND ATTITUDE TESTS

It has already been suggested that earlier comparative studies of the language laboratory were often unsatisfactory from the point of view of the tests with which they measured pupils' attainment. Keating, for instance, used the same tests at different levels of French, where they could hardly be expected to have the same power of discrimination. His so-called test of 'speech production' was no more than a test of the ability to produce ten sounds of French accurately. The same test was given to all four levels tested.

The tests used in our study set out to give full and accurate feedback on all aspects of pupils' attainment in German at regular intervals throughout the three years of the controlled part of the study. To achieve this, the pupils were tested at the end of each term in the skills of listening comprehension, speaking and reading/writing. In the first eight terms, these tests were internal tests of achievement, i.e. they were tests devised by the teachers on the syllabus taught. They were, as far as possible, objective tests: however, "objective" tests remain subjective as regards the sampling of the syllabus. Therefore, in the ninth and final term, external tests were used which were

not tied to any syllabus but were rather tests of general proficiency in German. These tests had the advantage of permitting comparisons between all the groups, whereas the internal tests of achievement, being syllabus-based, had to be different in content for Group N with the audio-visual course, and thus only allowed comparisons between Groups L and T.

Internal tests of achievement

1. *Listening comprehension*

These tests were multiple-choice tests in which both the stem and the choices were presented orally (pre-recorded on tape). They lasted 30-35 mins.

2. *Speaking*

The oral tests were administered to all the pupils in small groups of five or six at a time by their own teacher, who asked each individual five or six scripted questions of an autobiographical kind, on general topics, on pictures, etc. The answers to the questions were assessed by a teacher from another group and a native speaker of German, sitting in the same room. Although there was a well-defined marking scale, this was nevertheless the least objective part of our testing. Therefore, to check the reliability of our assessment procedures, all the oral tests were recorded on videotape for reassessment at the end of the study by a panel of five independent judges. The high correlation between the judges themselves, and between the mean of their assessments and our own assessments, suggests that the oral marking was in fact reliable.

3. *Reading/writing*

As the written word was not introduced until about the end of the first term, there were no tests of reading/writing before the third term. Thereafter, there was a reading/writing test lasting about the length of a lesson at the end of each term. Pupils were generally asked to supply appropriate words for blanks in continuous texts of German.

In addition to the above tests there were, in the first term only, tests of sound discrimination (English-German and German-German) and of pronunciation (sentences repeated after a model and recorded on tape).

All three groups did the same number of tests (23) and the same types of test. Group N's tests, however, were based on different course material and were therefore different in linguistic content.

External tests of proficiency

In the ninth and final term of the controlled part of the study, all three groups took the same external test of proficiency in German, namely the Pimsleur German Proficiency Test, First Level. However, as the speaking part of the test (Part 2) is conducted in the language laboratory, and we wished to maintain our practice of having face-to-face oral tests, a separate external oral test was given. It was designed and conducted by Mr. Walter Grauberg, Director of the Language Centre at the University of Nottingham, who spent three days at the school examining each pupil individually.

1. Pimsleur German Proficiency, First Level

Three of the four parts were taken. *Part 1* is a multiple-choice test of sound discrimination followed by a multiple-choice test of listening comprehension. *Part 3* is a multiple-choice test of reading comprehension. *Part 4* is a test of writing proficiency consisting of blank-filling items, sentence transformations, and free writing in response to pictures.

2. Oral Test

The oral test had three sub-sections: a simple text was *read aloud* after a few minutes' study, as a test of the accuracy of sound-symbol association; *general questions* were asked, starting with simple biographical questions as a "warm-up", and continuing with more advanced questions requiring longer answers; pupils *played a role* with the examiner, working on cues suggested by a picture studied in advance.

Attitude tests

Both the audio-visual and the audio-lingual course aimed not only to teach the German language, but also to widen the horizons of the pupils a little by introducing them to Germany and to its people and their customs. It was of interest, therefore, to test not only attainment in the language but also attitude to it and to the people and the country. A test was given in the final term of the controlled part of the study in which pupils were asked to agree or disagree with a number of statements about German, Germany and the Germans. The statements had previously been rated on an approval-disapproval scale. (See Lovell (1969, 312) for a description of the design of the test.)

In the fourth year, i.e. after completion of the controlled part of the study, a questionnaire was given to the pupils, testing their attitude to the language laboratory itself. All the pupils completed the questionnaire since all had by then had experience of the lab.

Table II: Tests

PREDICTIVE TESTS	Intelligence	Pimsleur Language Aptitude Battery	NFER Primary Verbal Test 3
	Language aptitude	York language aptitude tests	Part 3: Vocabulary Part 4: Language analysis Sound discrimination Vocabulary retention Grammatical analysis
ATTAINMENT TESTS	Internal tests (terms 1-8)		Sound discrimination Pronunciation Listening comprehension (7 tests) Oral production (8 tests) Reading/writing (6 tests)
	External tests (term 9)	Pimsleur German Proficiency Test, First Level External examiner	Test 1: Listening comprehension Test 3: Reading comprehension Test 4: Writing proficiency Oral production (3 sub-tests)
ATTITUDE TESTS			Attitude to language, people and country Attitude to language laboratory

RESULTS

Effectiveness of the language laboratory

The 23 internal tests permit comparisons only between Groups L and T, who followed the same course. Means and standard deviations were calculated for them and compared. In all cases two-tailed tests were used, since there was no prediction of the direction of any advantage.

There seems to be little evidence of a pattern of differences in means (or variances). In 15 cases out of 23, the mean of Group L is numerically the higher; on the other hand, there are only four statistically significant differences (at the 0.05 level) and in all four cases Group T is favoured. The lack of pattern applies both in terms of time (first year, second year etc.) and in terms of skills (listening, speaking etc.). For instance, at the beginning of year two, Group L was marginally ahead in listening comprehension and reading/writing; by the end of the year, Group T was clearly ahead in these two tests; in year three, there were no statistically significant differences. Interestingly, not one of the comparisons in the oral tests produced a statistically significant result.

As a further basis for comparison, annual percentage totals were worked out for each pupil (giving equal weighting to the different tests) and group means and standard deviations calculated accordingly. As can be seen from Table III, the difference in percentage scores is less than 1% in favour of Group L in year one, about 2 1/2% in favour of Group T in year two, and about 2% in favour of Group L again in year three. To be judged significant at the 0.05 level, percentage differences of about 3.5, 6.8 and 7.5 respectively were needed (at the more stringent 0.01 level, 4.7, 7.8 and 10.5). What has been said about means applies also to differences in variability.

Table III: Annual percentage totals for Groups L and T

	Group L		Group T	
	Mean	Standard deviation	Mean	Standard deviation
Year 1	66.84	6.83	65.96	7.82
Year 2	61.25	11.24	63.74	12.66
Year 3	59.46	15.94	57.22	13.79

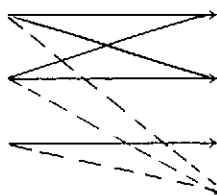
The external assessments of the final term were common to all groups and therefore allow comparisons to be made also with Group N. Like the internal oral tests, the final orals failed to yield any significant differences between the groups. The three sub-sections of the Pimsleur Proficiency Test produced one difference that was just significant at the 0.05 level: this was a significant superiority of Group N over Group T in Reading Comprehension (the mean for Group L lay between the other two and was not significantly different from either). No particular importance can be attached to this isolated and barely significant difference between two groups that differed on two variables (lab and course material). A composite standard score was computed for the three sub-tests: it failed to reveal any significant difference between the groups.

The overall picture emerging from the statistical analysis of both the internal and the external tests is one of balance between the groups, with any differences going now in one direction, now another, but remaining in any case negligible. One is forced to the conclusion that neither the language laboratory nor the course material had any markedly different effect on attainment, at any rate over a three-year period.

Reference has been made a number of times to the "controlled part of the study", that is to say the first three years from 1967-1970 when the pupils were in matched groups and variables such as the teacher could be controlled. This was the period during which the school had guaranteed to hold the matched groups intact, a guarantee which was honoured. The results presented so far refer to this period.

However, the pupils continued to learn German for at least two further years, at the end of which most of them took the public examination for the General Certificate of Education (GCE) in German. At the beginning of the fourth year some reorganisation of the *forms* took place which affected our groups since some boys from each of the groups found themselves in a slow "stream" called JM. JM's different timetable meant that they could not learn German at the same time as their colleagues, and so a fourth group for German came into being. The remainder of Groups L and T were combined and then divided into ability sets A and B. The remainder of Group N continued as an intact group. These changes are shown below in diagrammatic form:

End of year three	
Group	No. of pupils
L	34
T	36
N	33



Beginning of year four	
Group	No. of pupils
A	29 (13 L + 16 T)
B	27 (15 L + 12 T)
N	29
JM	14 (7 L + 5 T + 2 N)

The new groups continued up to the end of the fifth year with minor changes (some interchanges between A and B and boys arriving and leaving). The teachers were the same as in the first three years but they no longer changed groups. Course materials remained unchanged for pupils in A, B and N, but pupils in JM started an entirely new course as they came from both the audio-lingual and the audio-visual groups. All the groups now used the language laboratory if the teacher wished (the school had in the meantime installed its own 18-booth AAR lab).

Thus the differences in the treatment of the groups were now blurred and the results of the GCE examination in 1972 are of interest only in so far as they might show a long-term effect of the exposure or non-exposure to a language laboratory in the first three years. In fact, neither in terms of pass rate nor in terms of mean performance is there any significant difference between the original Groups L and T. (Group N took a GCE examination with a different syllabus reflecting the different course.)

Attitude tests

1. The test of attitude to the language, people and country was readministered after an interval of one week in order to estimate its reliability. The coefficient of 0.74 thus obtained suggests that the test was reliable.

Attitudes in all three groups tended towards the favourable end of the scale. Inter-group differences in attitude are not statistically significant for L vs T or L vs N, but there is a significant difference (at the 0.001 level) between Groups N and T in favour of Group N. This cannot be attributed to any one variable, but may be a result of the unique combination of variables distinguishing the treatment of the two groups — language laboratory, course material and classroom.

2. No test-retest reliability coefficient is available for the language-laboratory questionnaire, but a test of internal consistency gave an index of 0.755 suggesting satisfactory reliability.

As with the other test, attitude in all groups tended towards the favourable end of the scale. The pupils claimed they enjoyed their lab sessions and found them useful; they thought that the lab helped their pronunciation and general fluency, but were non-committal when asked if it helped their grammatical accuracy. The two lab groups L and N did not differ significantly in their attitude towards the lab.

VALIDITY OF THE SITUATION INVESTIGATED

The results just presented refer to a controlled study of the effectiveness of the language laboratory used in a quite specific situation: with junior forms (= 11–13 year olds), for one timetabled lesson a week, using commercial pre-recorded tapes. This we considered to be a common pattern of usage in British schools. In order to assess how far our supposition was borne out by the facts, we sent out two questionnaires in January 1972. The first went to individual schools in all parts of the country, believed by us to have language laboratories. The second was sent to modern language advisers in different education authorities, who might be expected to have a good knowledge of how the language laboratories were used in the schools in their areas.

Questionnaire to schools

74 replies were received to the questionnaire sent to schools. They came from all parts of the country and represented urban and rural areas and comprehensive and selective schools. One of the schools no longer used its lab and so the summary statistics presented in the table refer to 73 schools.

Most schools used their lab at all levels in the school; only four did not use it with junior forms. Though 28 out of 68 schools did not give their junior forms one lesson a week in the lab, only five of them gave them longer. The question about tapes was badly formulated in that it asked in general whether the school used mainly commercial or home-made tapes. It should have restricted the question to junior forms. As some schools pointed out, commercial tapes are generally used with junior forms, home-made tapes with senior forms, simply because there are far fewer commercial tapes available at the advanced level. Nevertheless, a majority of schools used mainly commercial tapes.

Table IV: Summary of replies to schools questionnaire

Is the lab used with junior forms?		How much time do junior forms spend in the lab?		Does the school use commercial or home-made tapes?	
Yes	69	1 lesson/week	40	Mainly commercial	41
No	4	Less than 1 l./week	23	Mainly home-made	19
		More than 1 l./week	5	Both equally	13
Total	73	Total	68*	Total	73

* one school omitted to answer this question.

For this limited sample of schools, at any rate, our chosen pattern of lab usage seems to have been typical.

Questionnaire to language advisers

34 replies were received to the questionnaire sent to language advisers. Five replied that no schools in their area had language laboratories. Statistics are thus based on 29 replies. The advisers were simply asked to agree or disagree with the proposition that our pattern of usage (which was carefully defined) "could be regarded as a common pattern of usage in the schools with laboratories" in the area. They were then asked to give information about the total number of secondary schools teaching languages in their area, the number with labs, what types of labs there were, etc.

Table V: Summary of replies to language advisers' questionnaire

	No. of language advisers	Secondary schools represented	Secondary schools with labs
Total	29	1327	442
		%	%
Typical	22	80	75
Not typical	5	12.5	12.5

Thus 22 of the advisers representing 80 % of the total schools and 75 % of the labs agreed that ours was a typical pattern of usage. Five disagreed, one was doubtful and one did not have labs in schools with pupils in our age-range.

It is interesting that almost exactly one-third of the secondary schools teaching languages had labs. Though there are no national statistics on the number of labs in schools, our finding substantiates that of an enquiry conducted by the Association of Teachers of German on the current position of German in schools (Thompson, 1971). 39 % of the teachers replying had access to a language laboratory.

VALIDITY OF THE PREDICTORS

Predictive information was of four kinds: previous experience of learning a foreign language (expressed in time), parental encouragement (on a five-point scale), verbal IQ, and language aptitude (Pimsleur LAB and York tests). The results of all these predictors were correlated with all the internal tests (and annual totals) and all the external tests. The resultant correlation matrix is very large and not easy to summarise in the limited space available here. As an indication of the comparative validity of the various predictors the numbers of significant correlations with the annual percentage totals are presented in Table VI. As there were three annual totals for each of the three groups, there are nine correlations with each predictor.

As can be seen, fairly good prediction was achieved with two of the four measures (verbal IQ and language aptitude). The grammatical analysis sub-test was the best single predictor, not only with the annual totals but throughout the correlation matrix. Its success in the very limited scope of this study has encouraged us to initiate further trials with it in a much wider context.

CONCLUSIONS

Research into language teaching, like other educational research, faces two broad problems. The first is the difficulty of ensuring that results are reflections of the factors under study and not of something else, i.e. that any interpretation of results is valid within the specific context of the research: it has *internal validity*. The second is the difficulty of generalising from internally valid results, i.e. of venturing outside the specific context of the research to an interpretation that has *external validity*.

Table VI: Correlations between predictors and attainment (annual totals)

Predictor	No. of significant correlations (9 possible)			Size of significant correlations	
	Total	0.01 level	0.05 level	Largest	Smallest
PREVIOUS EXPERIENCE OF FL	3	2	1	0.58	0.41
PARENTAL ENCOURAGEMENT	2	0	2	0.37	0.36
VERBAL IQ	8	6	2	0.58	0.29
LANGUAGE APTITUDE (TOTAL)	8	6	2	0.68	0.33
Pimsleur LAB Pts 3 + 4	5	2	3	0.55	0.31
Sound discrimination	2	0	2	0.31	0.30
Vocabulary retention	4	2	2	0.60	0.29
Grammatical analysis	8	7	1	0.72	0.37

In this paper so far I have attempted to give an outline of how pupil, teacher and situational variables were controlled and adequate testing of attainment in German was achieved. Then, on the basis of a rigorous statistical analysis of all our data, I drew the conclusion that we held to be *internally valid*, i.e. to be a justifiable interpretation of our results in the specific context of our research.

That conclusion was that a group of pupils using a language laboratory as an aid in their learning of German showed no detectable difference over a period of three years, in either performance or attitude, from a matched group of pupils that did not use the language laboratory. To turn now to the *external validity* of that conclusion: what, if anything, does it tell us about the language laboratory outside its own context?

One restriction on the general applicability of our findings we have already considered, namely the quite specific limited way in which the lab was used. Evidence was presented, however, to show that it is a very widespread pattern of usage in British schools.

Other restrictions to be considered are the size of the York sample and the representativeness of the York pupils.

Size of the York sample

Keating's research involved some 5000 pupils in 21 school districts, the Pennsylvania Project 105 classes in 58 high schools; in the York study there was only one school, three classes and 101 pupils. However, against the smallness of the sample must be set the high degree of control which was achieved over variables other than the lab, and the fact that matched groups were kept intact for three years and the pupils studied for five years.

The representativeness of the York pupils

The pupils in the York sample were untypical of their age group as a whole in that they were all boys, and boys who had been selected for grammar school education, i.e. from the top 25 % of the ability range.

The absence of girls from the York study may be serious since all the research evidence seems to indicate that there is a difference between boys and girls in attitudes and ability when learning languages (see, for example, Burstall, 1970), and therefore their response to work in a language laboratory might be different.

The absence of less able pupils from the York study is certainly serious: French in the primary school and the introduction of comprehensive secondary schools have meant that for the first time in Britain modern

languages are being taught to children of all abilities⁴. Knowledge of the role of the language laboratory in this situation would clearly be of great value.

How valid, then, is our result for more able pupils in other schools using the lab in the way ours did? It must be remembered that by using the university laboratory we enjoyed more favourable conditions than many schools:

1. We always had enough lab positions to accommodate the whole class.
2. The lab had a split-monitoring facility, and by using the German assistant we were usually able to double the intensity of monitoring.
3. A staff of full-time technicians meant that faults could be dealt with instantly.
4. Pre-recording of the pupil tapes was often possible, allowing the pupils to work independently right from the start of the lab session.
5. The fact that the pupils knew they were the object of a research investigation and that they made weekly visits to the university may have heightened their motivation (Hawthorne effect).

There seems then no reason to suppose that other schools who use their language laboratories as we used ours are obtaining better results than we did. In other words, a substantial number of schools may well be achieving no more with a costly language laboratory than they could achieve with a simple tape-recorder used in the classroom.

To judge from the continued growth of language laboratory installations in schools, both teachers and administrators assume that they are beneficial to pupils. Our study does not justify the assumption and points to the need for more experimentation in types of installation, patterns of usage and kinds of recorded material. Such experimentation should be accompanied by controlled tests of its effectiveness.

In too many schools, in our experience, a language laboratory exists side by side with totally inadequate facilities for playing tapes and projecting visuals. Projectors and tape-recorders have to be carried from room to room (which generally means that the sound is relayed through the tape-recorder's own small speaker often pointed at the teacher rather than the pupils). Rarely are all the rooms where languages are taught capable of being blacked out.

The present concept of a language laboratory as a separate room to which classes have to be timetabled may prove to be an inefficient one. For the same cost several ordinary classrooms can be converted into language-teaching

4 Whereas in 1964 only about 25 % of secondary pupils learnt a foreign language, by 1972 the figure was about 60 % (Report to the Council of Europe Symposium, 1973, by M. V. Salter, HMI — mimeographed).

rooms, each equipped with tape-recorder and linked set of audio-active headsets, extension loudspeaker, slide/filmstrip projector, overhead projector, screen, blackout facilities, magnet board or flannelgraph, and versatile furniture. Further experimentation may show this kind of installation to be better value for money.

Bibliography

- Burstall, Clare (1970): *French in the Primary School — Attitudes and Achievement*, National Foundation for Educational Research in England and Wales.
- Carroll, John B. & Sapon, Stanley M. (1958, 1959): *Modern Language Aptitude Test Form A*, New York, The Psychological Corporation.
- Central Advisory Council for Education (England) (1967): *Children and their Primary Schools, Vol. 2*. London, HMSO.
- Keating, Raymond F. (1963): *A Study of the Effectiveness of Language Laboratories*, The Institute of Administrative Research, Teachers' College, Columbia University.
- Loge, Sarah W. (1964): "Language Laboratory Research Studies in New York City High Schools: A Discussion of the Program and the Findings", *Modern Language Journal* 48, No. 7, 409—419.
- Lovell, K. (1969): *Educational Psychology and Children*, London, University of London Press.
- Pimsleur, Paul (1964): *Language Aptitude Battery*, New York, Harcourt, Brace & World.
- Pimsleur, Paul (1967): *German Proficiency Test, First Level*, New York, Harcourt, Brace & World.
- Smith, Philip D. Jr. (1970): *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction — The Pennsylvania Foreign Language Project*, Philadelphia, Center for Curriculum Development; London, Harrap.
- Thompson, Miles C. (1971): "Report on the ATG Questionnaire on the Position of German in Schools", *Treffpunkt* 4, No. 1.

Commentary by J. Ankers, *Etudes pédagogiques de l'enseignement secondaire*, Geneva

Preface

The author of the following lines teaches English as a foreign language at secondary school level and is responsible for the training — methodology, not language — of graduate teachers of English.

His experience of language laboratory work has been primarily with 15—17 year-olds in their first two years of English, all pupils using the lab 1 period per week, in groups of not more than 15 (half the class). On IQ ratings the pupils would be in the top 30 % of their age group.

The following remarks no doubt reflect — more or less directly — the writers preoccupations and experience, which, in turn, fix the limits of his competence and reliability as a critic.

1. Expectations

Most of those interested in the language laboratory had had wind of the project long before the results of the York research were published. The hopes and expectations thus raised will, in the main, have been disappointed. Firstly, because the report is neither encouraging nor helpful to those who *have* a lab (i.e. the vast majority of those interested in the problem) and secondly because of the strictly limited scope of the investigation. Under the circumstances there will be a strong temptation either to ignore the report or, more probably, to try to apply the results and conclusions within far too wide a framework.

The present commentary will, in fact, be more concerned to stress the danger of unjustified generalization than to call into question the findings of the York team.

2. Internal validity

Mr Green states his terms of reference approximately as follows. "Is the teaching in a given situation — use of the lab for one period a week with junior forms using commercial tapes — more effective because the lab is used? "

Given this limited aim, what factors might invalidate the findings?

2.1. Limited numbers

This risk is mentioned by Mr Green, who tends to discount it in view of the precautions taken to eliminate all the conceivable variables. The fact remains that the effectiveness of the lab is judged on the performance of some 60 boys, and that this number is reduced to 30 if one considers that comparison between groups N and T is not possible because of the introduction of a second variable (audio-lingual / audio-visual). Admittedly, the groups were not arbitrarily formed.

2.2. Course material

The comparison is based on one course-book. It can be valid only insofar as that course-book is typical of good, modern, commercially available teaching material.

Moreover, if the course, whatever it purports to do, actually teaches knowledge rather than skills, the means by which the material is fed to the pupils (lab or not) is unlikely to have a significant bearing on the results of

skills tests. There must be a realistic link not only between the lexical and structural contents of the course and that of the tests but also between the capacities which the course *really* develops and the type of tests used.

2.3. Frequency of lab work

The frequency of 1 period out of 5, in the first year, may well be typical of the way in which the language lab is used in British secondary schools. It must, however, pose problems of integration and must surely be close to the limit below which the use of the language lab would cease to be a significant variable. To that extent it casts doubts on the internal validity of the research as well as on the conclusions — replacement of the lab by something else, rather than using it differently — which Mr Green draws from his findings.

2.4. Conditions of use

It would appear that the whole class was in the lab at the same time, monitored "usually" by two people.

The pupils' tapes, we are told, were "often" pre-recorded. This situation prompts the following remarks.

- Although they were at an age when they might be expected to need considerable teacher support the pupils were numerous in the lab (even 16 is a lot to monitor in 35 minutes) and were occasionally monitored by one person only.
- When there were two people, one — the German assistant — was not really in a position to make the explicit, ideally pupil-by-pupil link between classroom instruction and lab work which integration of the latter demands. Moreover, the teacher's personality and his familiar presence are very strong motivating factors at this age. When he is monitoring a small group he can be more present for each individual in the lab than in the classroom. If he is monitoring a large group he tends to fade into anonymity.
- The fact that recording onto student tapes not infrequently took place during the lab period implies one of two things. Either the amount of actual practice time in the lab was even further reduced (see 2.3. above) and the pupils performed other activities while the tapes were being copied, or they did the exercises during the recording process, so that, for a considerable proportion of the time available, the lab was operating simply as an AA installation with headphones and a central tape-recorder¹.

¹ In fact, the second solution was the one generally chosen.

The objection here is the same as that raised in connexion with the frequency of lab work. Under conditions so far from optimal can the language laboratory still be considered as a significant variable?

3. External validity

Taking all due precautions, Mr Green makes a number of tentative suggestions about the relevance of his findings to other situations. His reservations concern essentially the type of pupil observed during the York inquiry. The following paragraphs attempt a more comprehensive survey of the factors to be taken into account before any generalization can be made about the effectiveness of the language laboratory.

3.1. The pupils

The York pupils were all boys, belonging to a specific age-group and a specific ability-group. They were attending a certain type of school, learning L₂ in an L₁ environment (L₁ and L₂ being relatively similar languages) and, setting aside the possibility of the research itself creating a Hawthorne effect (Do '0' level results throw any light on this?), had probably no real motivation for learning German beyond the desire for success in a school subject.

3.2. Teacher attitudes

The fact that such material was available, and that each teacher taking part in the experiment used the lab 2 terms out of 3, presumably meant that the general teaching strategy was geared to the use of lab-type exercises. Moreover, each experimenter being aware that classroom tape-recorder and language laboratory were being compared, probably inclined, whether consciously or not, to make optimal use of the former as a lab substitute. Such an approach is not necessarily typical of the average teacher, since the lab, by its nature, tends to impose certain activities in a way that a classroom, however well equipped, does not.

3.3. 'Richness' of lab use

By 'richness' we mean the extent to which the inbuilt possibilities of the language lab are exploited.

This will depend, essentially, on two inter-related factors, materials (what types of exercises are available) and methodology (how the exercises are used). The following table will give an idea of the complexity of these two factors.

<i>Materials:</i> What sorts of exercise are available?	<i>Methodology:</i> How are the exercises used?
	Frequency and length of lab sessions
	Numbers in lab and monitoring arrangements
	Degree of pre- and post-integration ('presentation' and responsibility in organisation of lab work (programmed instruction?))
	Library system / broadcast system with pre-recording / broadcast system without pre-recording
<i>'Pronunciation' exercises</i> — discrimination — production — phonemic difficulties — phonological factors — based on L ₁ / L ₂ comparison	In early stages only Initial learning or remedial In situation With visual support Problems made explicit Comparison written code/oral code
<i>Structure drills</i> — on what sorts of difficulty (L ₁ /L ₂) — grading: exercise to exercise and within each exercise — situational . in what sense . to what extent . visual support . sound effects . course-book related . personal-experience related — 2/3/4/5 phase	Degree of pre- and post-integration ('presentation', 'exploitation') No of exercises per session and no of repetitions per exercise Extent to which pupils use 'comparative' element of AAC lab Exs used for initial training, immediate reinforcement or revision Tapescript available for student consultation
<i>Comprehension exercises</i> — frequency — length — amount of unfamiliar lexical and/or structural material they contain — speech rate — variety of voices and accents	With/without text and at what stage Possibility of repeating section by section Nature of comprehension control — question and answer (L ₁ or L ₂) — multiple-choice questionnaire — restitution
<i>Open-ended conversation exercises</i> — requiring immediate, short answers or permitting development of response — grading — subjects relating to students' own experience or to knowledge of L ₂ 'culture'	

<i>Dialogue memorisation</i>
Small-group communication
Oral composition

N.B.

- a) Each item should be followed by a question mark. If this has been omitted it is simply a question of typographical convenience.
- b) Methods and materials are to a considerable extent interdependent and the dividing line between the two is necessarily arbitrary. Moreover they are valid only as a part of a pedagogical strategy adapted to the particular situation in which they are to be used.
- c) Factors listed only in connexion with one type of exercise are often applicable to other types as well.
- d) Neither the catalogue of types of exercise nor the list of items in each category is exhaustive.

3.4. General and administrative factors

On the one hand these have a bearing on the extent to which the findings of any educational research project can be generalized; on the other hand their probable influence on — or the likelihood of their being influenced by — any *modified* situation must be assessed before recommending, on the basis of the results of such a research, a change in policy (in this case, abandoning the language laboratory in favour of something else).

3.4.1. Administrative attitudes

The degree of administrative inertia, indifference, inefficiency or resistance to any given teaching strategy is neither incidental nor accidental to the pedagogical situation but an inherent constraint upon it. Arguably it is measurable and relatively constant in relation to a limited number of internal or external pressures. It is a phenomenon of the same type as the high proportion of drop-outs in any non-compulsory instructional programme or the over-estimation of Mme Harding's students as to the number of hours they would spend using the facilities of the bibliothèque sonore². While the phenomenon is largely absent from an experimental framework as limited as

² See the paper "La bibliothèque sonore: Implications pédagogiques" in this volume.

the York research (thus limiting the *external* validity of the findings) it works as a factor of *internal* invalidation in anything as vast as the Pennsylvania project.

3.4.2. Practical constraints

The wide variations in results which can be observed, even in the same school, as between different classes and/or different subjects, suggests that the influence on performance of practical constraints can be significant. In a research project, however, the necessary elimination of the more important variables tends to neutralize factors such as the following, thus rendering the situation non-typical.

Material conditions, in particular

- class numbers (including size of lab groups where the lab is used)
- constant use of same classroom or movement from one room to another
- quantity and quality of classroom (and lab) equipment, including acoustics etc
- regularity and quality of servicing of this equipment
- external noise

Timetabling factors

- spread of modern language periods over the week
- time of day at which ML periods take place
- types of activity preceding or following ML periods
- frequency and length of lab periods

The teacher

- personality
- knowledge of L₂
- training
- experience
- attitude towards course-book, methods and stated objectives

The class

- degree of homogeneity as regards
 - intelligence
 - language aptitude
 - knowledge of L₂ (possibly of a L₃)
 - motivation
- 'group personality'
- acceptance of teacher

4. Validity of the recommendations contained in the report

Even if one admits (a) that the York research is internally valid and (b) that the results are generalizable to a greater or lesser degree, the acceptability of the recommendations Mr Green makes in the concluding paragraphs of his paper must depend on the various interactions of the factors listed under 3.3.1. and 3.3.2. above. For example, to what extent can the language laboratory compensate for inadequate classroom teaching with pupils of a given age, ability and L₂ level? Or again, doesn't the fact that the lab is *expensive* and an *entity* condition administrative attitudes and action in a way favourable to modern language teaching in general? (Generally speaking, far more importance is accorded to lab servicing than to looking after classroom equipment, and how many authorities would be prepared to split a class in two once a week for language instruction if they hadn't invested in a lab?)

Even within his own, limited framework, one wonders whether Mr Green's proposal to abandon the AAC lab in favour of an AA, recorder-plus-headphones system is not ill-advised. Surely, with classes of 30 pupils or more, imaginative use of the lab could give each student much more meaningful speaking time and a much greater degree of individual attention than he could get in the classroom. Perhaps the first step would be to try to create a material which, unlike commercial tapes — bi-valent more or less by definition — really sought to take maximum advantage of the unique, specific possibilities of the AAC lab. An important merit of the York research is that it underlines this need and that it might serve to persuade educational administrators to make the necessary timetabling and other concessions, so that such a material, once it existed, might be used in truly favourable conditions.

Intervention de R. Jeanneret et de G. Merkt, Centre de linguistique appliquée de l'Université de Neuchâtel

Une expérience sur l'efficacité du laboratoire de langues dans l'enseignement de l'allemand dans les écoles secondaires neuchâteloises

1. Le cadre historique

Le premier laboratoire de langues digne de ce nom a été installé en 1962 à l'Université de Neuchâtel. Deux ans plus tard, l'École supérieure de Commerce de Neuchâtel, préoccupée par le renouvellement de l'enseignement

du français langue 2 à ses nombreux élèves suisses alémaniques et étrangers, faisait également l'acquisition d'un laboratoire CEDAMEL de 24 cabines. Sur la base des expériences réalisées dans cette Ecole et à l'Université, et suivant les conseils de la CILA en la matière, le département de l'Instruction publique décidait, au début de 1969, de favoriser l'implantation généralisée de laboratoires de langues dans les écoles secondaires du canton.

Avant même de fixer son choix sur un type particulier de matériel, la Commission cantonale "Laboratoires de langues" se préoccupa de mettre aussi rapidement que possible à disposition des écoles des programmes permettant une fréquentation soutenue au LL. Or, les programmes alors disponibles sur le marché s'ajustaient mal au manuel officiel, en raison d'une progression différente du matériel lexical et des structures grammaticales. De plus, les séries de bandes étaient trop courtes et ne permettaient pas une fréquentation suffisante du LL.

C'est ainsi que le Centre de linguistique appliquée, à la demande du département de l'Instruction publique, prit la direction de groupes de travail chargés de mettre sur pied un programme étroitement intégré au manuel officiel (*Wir sprechen Deutsch*, Lausanne, Payot, 1960). Tout en respectant scrupuleusement la progression lexicale et grammaticale du manuel de base, le programme était conçu de telle façon que chaque classe puisse travailler une heure par semaine au LL. En 1972, un ensemble de quelque 120 bandes d'exercices était terminé. Il s'adresse aux élèves des sections classique et scientifique des trois dernières années du cycle d'études secondaires inférieures.

Par ailleurs, en 1971, le département de l'Instruction publique a constitué une Commission de l'informatique, chargée d'étudier l'introduction de cette discipline dans l'enseignement secondaire. Un des groupes de travail dépendant de cette Commission fut chargé d'étudier la correction automatique de tests par ordinateur.

Il paraissait dès lors intéressant d'une part d'expérimenter ce système de correction automatique pour une langue vivante, d'autre part de chercher à mesurer l'efficacité du travail au LL. En effet, en 1972, au moment où fut prise la décision de tenter cette expérience, plusieurs classes du canton n'avaient pas encore la possibilité de suivre des cours au LL, tout en utilisant la méthode WSD. Cette tentative nous a paru d'autant plus justifiée que certaines critiques commençaient à s'élever, portant en particulier sur les points suivants:

— l'obligation de fréquenter le LL à heure fixe pour y étudier un programme arrêté une fois pour toutes, imposait aux classes un rythme incompatible avec l'idée que certains se faisaient de la liberté d'organisation du travail;

- la motivation des élèves (en particulier chez les sujets les plus et les moins doués) ne se maintenait pas toujours à un niveau très élevé, et retombait vite après une période d'enthousiasme initial;
- l'amélioration des aptitudes linguistiques des élèves ne correspondait pas en tous points aux espoirs qu'on avait placés dans le LL.

Par ailleurs, sur le plan scientifique, certains spécialistes en linguistique appliquée opéraient un revirement spectaculaire dans leur prise de position à l'égard du LL. De plus en plus le scepticisme était de mise, entretenu par des rapports d'expériences désenchantés et des remises en question des fondements théoriques de la pédagogie du LL.

C'est dans ce contexte général que s'insère l'expérimentation qui a été tentée au cours de l'année scolaire 1972/73.

II. Description de la batterie d'épreuves

Le programme d'allemand complémentaire au manuel WSD pour le travail en LL était, comme nous venons de le mentionner, essentiellement axé sur la fixation des structures grammaticales. En conséquence il semblait tout indiqué de contrôler l'influence du LL sur le degré d'acquisition de ces structures. Cela nous eût amenés à construire une batterie d'épreuves mesurant le degré de correction et la rapidité de réaction à des énoncés oraux. Si ces deux paramètres pouvaient faire l'objet d'une mesure objective, la correction présentait en revanche des difficultés considérables, puisqu'il aurait fallu procéder à l'enregistrement des réponses puis envisager un dépouillement nécessairement très long de ces documents. Nous cherchions une batterie de tests qui fussent à la fois objectifs et économiques; de plus, nous désirions autant que possible confier la correction à un ordinateur.

C'est la raison pour laquelle nous avons opté finalement pour une batterie d'épreuves axées sur la compréhension auditive. Il nous semblait intéressant d'étudier l'effet du laboratoire sur une aptitude linguistique fondamentale, qui, bien qu'elle ne constituât pas l'objectif premier de notre programme, se trouvait cependant impliquée par le caractère essentiellement oral du travail en LL. En tous les cas, si des progrès étaient constatés dans le domaine de la compréhension auditive, cela pouvait être mis au crédit du LL, et du même coup cela pouvait constituer un démenti aux reproches selon lesquels le travail au LL est purement formel, mécanique et dépourvu de référence à une situation de communication réelle.

La batterie de tests se composait de quatre épreuves.

1. D'abord une série de 20 items portant sur la discrimination auditive. Les élèves étaient appelés à reconnaître dans des séries de 3 mots ou groupes

de mots l'élément différent des deux autres. Par exemple: alt – halt – alt, ou bien: er sieht *wenige* Leute – er sieht *weniger* Leute – er sieht *weniger* Leute.

2. La deuxième épreuve de 16 items consistait à associer un énoncé à l'image (à choisir dans une série de 4 dessins) qui en représente l'illustration exacte. Ainsi, pour la phrase: Dort kommt unser Nachbar aus dem Gasthaus, nous proposons les 4 dessins suivants:



3. La troisième épreuve de 16 items consistait à trouver parmi quatre réponses écrites à choix, celle qui correspondait à une question présentée oralement. Ainsi, pour la question "Wohin gehen Sie am Morgen?" nous proposons les 4 réponses suivantes: "In die Schule – draussen – in der Schule – bei meiner Mutter".

Il est bien évident que dans cette épreuve intervient également à côté de la compréhension auditive la maîtrise de certaines connaissances grammaticales. Le choix entre "in die Schule" et "in der Schule" implique avant tout une compétence sur le plan grammatical.

4. La dernière épreuve consistait en l'écoute de deux anecdotes, à la suite desquelles les élèves étaient appelés à juger des énoncés s'y rapportant selon trois critères: juste — faux — "on ne peut pas le savoir", ce dernier critère dans le cas où la phrase exprimait plus que le texte de l'anecdote sans toutefois le contredire. Cette épreuve de compréhension globale comportait 24 items.

La batterie ainsi mise sur pied proposait quatre approches différentes de l'aptitude à la compréhension auditive. Cette démarche nous paraissait légitime dans la mesure où la compréhension auditive est une aptitude complexe dans laquelle interviennent entre autres certains des facteurs qui se manifestaient de façon marginale dans notre batterie: par exemple les connaissances grammaticales dans la 3ème épreuve ou la mémoire dans l'épreuve No 4. Nous étions conscients également que le degré de difficulté des quatre épreuves n'était pas absolument comparable: la discrimination auditive allait entraîner un taux de réussite bien plus élevé que l'épreuve de compréhension globale. Nous tenions cependant à la conserver dans notre batterie, même si elle n'était pas très sélective, car elle pouvait nous permettre de détecter d'éventuelles défaillances graves de l'audition chez certains élèves.

Les consignes de passation étaient minutieusement définies. Les énoncés avaient été enregistrés sur bande magnétique et ils étaient diffusés au moyen d'un magnétophone. Le temps de passation était donc rigoureusement identique pour toutes les classes. Nous avons en outre pris la précaution de confier l'enregistrement à des présentateurs "neutres", que les élèves n'avaient eu l'occasion d'entendre à nulle autre occasion.

Il est superflu de décrire dans le détail les consignes de passation. Elles ne présentaient pas de difficultés sensibles. En revanche, la procédure de notation des réponses pouvait poser quelques problèmes aux élèves. En effet, nous leur demandions d'abord de cocher une grille spécialement préparée pour chaque épreuve, puis de transcrire les réponses sur une feuille individuelle de réponses conçue en vue de la correction par ordinateur. Cela risquait de multiplier le nombre d'erreurs. Il ne semble toutefois pas que des erreurs de transcription se soient produites lors de la première passation. En revanche, par la suite, on en déplore quelques-unes, ce qui se traduit par une ouverture de l'écart-type (on passe de la valeur 6,62 à 10,61 entre la première et la troisième passation). Ce phénomène n'est pas étranger aux conditions de passation du troisième volet de l'expérience, qui, intervenant à la fin de l'année scolaire, se ressentait de la lassitude et de l'indifférence que l'on constate habituellement à cette période de l'année scolaire.

Nous avons prévu d'échelonner l'expérience sur une année scolaire complète (l'année 1972/73) et de procéder à 3 passations des épreuves, d'abord en septembre 1972, c'est-à-dire dans les premiers jours de la rentrée

scolaire, puis en février/mars 1973, et enfin en juin/juillet 1973, peu avant la fin de l'année scolaire.

Précisons encore que lors de la première passation la distinction entre classes avec et sans LL n'était pas pertinente, puisque aucun des groupes n'avait eu l'occasion de suivre un enseignement au LL au cours de la scolarité précédente. Ainsi la première passation n'avait pratiquement pas d'autre fonction que de déterminer le niveau de performances respectif des groupes soumis à l'expérimentation.

Pour les 3 passations prévues, nous avons élaboré deux versions différentes de la batterie d'épreuves décrite ci-dessus. Nous avons fait intervenir la deuxième version pour les passations de février et de juin 1973. Elle était semblable quant au type d'épreuves, aux consignes et au nombre d'items à la première batterie; elle différait en ce sens que le matériel lexical et grammatical était adapté au niveau atteint par les élèves en février.

III. Résultats statistiques

1. Méthode et objet

Deux groupes ont été considérés: classique C et scientifique S. Au sein de chacun de ces groupes, on a isolé deux sous-groupes:

classique sans laboratoire:	CSL
classique avec laboratoire:	CAL
scientifique sans laboratoire:	SSL
scientifique avec laboratoire:	SAL

Nous avons tenu compte également des dates de passation des épreuves:

- 1: septembre 1972
- 2: février 1973
- 3: juin 1973.

De ce fait, nous avons affaire à 12 catégories, soient: CSL 1, CSL 2, CSL 3, CAL 1, . . . etc.

Trois sortes de comparaison ont été envisagées:

1.1. A l'intérieur d'un même groupe (C ou S), significativité des différences entre les moyennes des résultats des classes avec laboratoire et sans laboratoire.

1.2. A l'intérieur d'un même groupe (C ou S) et d'un même sous-groupe (CSL; CAL; SSL; SAL), significativité entre les moyennes des résultats aux différentes épreuves (1, 2 et 3).

1.3. Comparaison des résultats obtenus par les sections C et S dans les classes avec laboratoire et sans laboratoire, aux différentes épreuves (CSL 1, SSL 1 – CAL 1, SAL 1 – ... etc.).

2. Résultats

2.1. Résultats bruts

Classes	Effectif	Moyenne	Ecart-type
CSL 1	101	51,40	7,6
CAL 1	145	52,17	6,62
SSL 1	125	48,86	7,28
SAL 1	155	47,69	7,46
CSL 2	102	55,69	7,70
CAL 2	170	58,32	7,70
SSL 2	119	53,98	8,94
SAL 2	202	55,05	8,94
CSL 3	108	59,54	7,72
CAL 3	161	60,16	8,66
SSL 3	123	54,13	9,51
SAL 3	173	56,19	10,61

2.2. Significativité des différences

2.2.1: Significativité des différences entre les moyennes des résultats des classes avec et sans laboratoire, à l'intérieur d'un même groupe (C ou S).

CSL 1	0,87 NS	CSL 2	2,72 TS	CSL 3	0,59 NS
CAL 1		CAL 2		CAL 3	
SSL 1	1,31 NS	SSL 2	1,05 NS	SSL 3	1,71 NS
SAL 1		SAL 2		SAL 3	

Commentaire:

NS : non significatif $t < 1,98$
 S : significatif $t \geq 1,98$ sécurité 95 %
 TS : très significatif $t \geq 2,6$ sécurité 99 %

Lors de la première passation, il n'y a pas de différence significative entre les classes utilisant le laboratoire et les autres; il s'agit d'un phénomène absolument normal, puisque le début de l'année scolaire est fixé fin août et

que les classes avec labo n'ont utilisé cet auxiliaire qu'à deux ou trois reprises seulement. On notera par contre que la moyenne de la section scientifique est un peu plus faible que celle de la section littéraire après une année d'enseignement de l'allemand. Au sein de la section scientifique, au départ, le groupe avec labo est un peu plus faible que l'autre.

Au cours de l'expérience, la moyenne de tous les sous-groupes augmente régulièrement, mais plus entre la première passation et la deuxième qu'entre la deuxième et la troisième. Par ailleurs, l'écart-type augmente considérablement d'une passation à l'autre (surtout lors de la troisième passation). On peut penser que les consignes d'application ont été moins bien observées en juin qu'en février. De plus, la période de juin, qui précède les vacances, s'est révélée peu favorable: élèves mal motivés, fatigués par les travaux de fin d'année scolaire, etc. Il convient donc de ne pas attribuer une trop grande signification à cette troisième épreuve qui, rappelons-le, était une reprise exacte de la deuxième.

On notera la différence très significative des résultats entre classes avec labo et sans labo dans la section classique, 2ème passation. De plus, dans la section scientifique, les classes avec labo obtiennent de meilleurs résultats que les classes sans labo, malgré leur handicap de départ. Cette tendance se confirme à la troisième passation (C et S).

2.2.2. Significativité entre les moyennes des résultats aux différentes épreuves (1, 2 et 3) à l'intérieur d'un même groupe (C ou S) et d'un même sous-groupe (CSL; CAL; SSL; SAL)

CSL 1	CSL 2
4,12 TS	3,60 TS
CSL 2	CSL 3

CAL 1	CAL 2
7,51 TS	2,04 S
CAL 2	CAL 3

SSL 1	SSL 2
4,89 TS	0,13 NS
SSL 2	SSL 3

SAL 1	SAL 2
8,43 TS	1,14 NS
SAL 2	SAL 3

Commentaire:

L'accroissement des moyennes est très significatif entre les passations 1 et 2, et nettement plus considérable dans les classes avec labo que dans les classes sans labo.

Entre 2 et 3, cette différence est significative dans la section classique, mais pas dans la section scientifique, quoique la tendance soit encore en faveur du LL (0,13 contre 1,14).

Comme 2 et 3 sont semblables, on mesure probablement plus la rétention de certaines connaissances qu'un véritable apprentissage, ce qui était le cas entre 1 et 2. On pourrait en conclure que le laboratoire favorise dans tous les cas l'apprentissage, surtout dans ses débuts; il est finalement regrettable que la troisième batterie de tests n'ait pas porté sur la matière étudiée entre février et juin.

2.2.3. Comparaison des résultats obtenus par les sections C et S dans les classes avec laboratoire et sans laboratoire, aux différentes épreuves (CSL 1, SSL 1 – CAL 1, SAL 1...)

CSL 1	CAL 1	CSL 2	CAL 2
2,63 TS	5,47 TS	1,50 NS	3,80 TS
SSL 1	SAL 1	SSL 2	SAL 2

CSL 3	CAL 3
4,68 TS	3,72 TS
SSL 3	SAL 3

Commentaire:

Cette comparaison entre section classique et scientifique avec ou sans labo est très significative, en faveur des élèves de la section classique, sauf lors de la deuxième passation pour les classes sans labo (CSL 2 et SSL 2).

On pourrait en conclure que le travail au laboratoire est mieux adapté aux élèves de la section classique qu'à ceux de la section scientifique, ou que le test était mieux conçu pour des classiques (mieux motivés pour l'apprentissage des langues) que pour les scientifiques. D'autre part, l'accroissement de l'écart-type, lié à une augmentation des moyennes, pourrait signifier que le laboratoire favorise les meilleurs élèves.

IV. Conclusion

L'expérience relatée ci-dessus nous semble prouver de façon irréfutable l'efficacité du laboratoire de langues dans le domaine de la compréhension auditive.

NB: La correction automatique des tests a été réalisée au Centre de Recherches psycho-pédagogiques du Cycle d'orientation de Genève, et l'analyse des résultats par le service de statistique scolaire du département de l'Instruction publique du canton de Neuchâtel.