Chapter 5

Sampling Schools and Administering Tasks

Együd J. Györgyi and Öveges Enikő

The December 1998 pilots of the Speaking and Writing Tasks

We had hoped to be able to pilot all the draft tasks we had produced and edited, some time in October or November 1998. However, it proved impossible for OKI Budapest to make the necessary arrangements in time, and so we had to postpone a large-scale piloting until suitable arrangements could be made. Nevertheless, we were keen to try out at least some of our tasks, and as we felt that the Writing and Speaking tasks were relatively straightforward to administer, in the case of the Writing tasks, or that we only needed a relatively small number of students, in the case of the Speaking tests, we decided to go ahead and pilot the tasks before the Christmas recess, 1998. In any case, we needed sample performances to take to Manchester in January 1999, for the second overseas training session, which was planned to deal with the development of rating scales for the productive skills.

Criteria for choosing the participating schools
We intended to video the Speaking tests, and so the availability of film crews was a key factor in deciding where to pilot the tests. In addition, we wanted to have at least some groups who were not in Budapest. Moreover, we needed in the short space of time available to identify schools that were conveniently located for those colleagues who would administer the tests. Finally, we needed where possible to have a mixture of school types, and a spread of ability levels if at all possible. Consequently, we decided on Budapest and Szeged for the Speaking tests and for the Writing tests, schools in Budapest, Szeged and Pécs. We had a mix of grammar, vocational and combined grammar/ vocational schools. We do not, however, wish to make any claims at all for the representative nature of this sample.

Contacting and preparing the schools for the pilots
The participating schools were contacted by the local members of the pilot team. These members informed the school management and the Head of English about the Examination Reform Project and about the aim of piloting the tasks. Afterwards the detailed arrangements were made with the help of the Head of English and the class teachers.

Selecting candidates for the Speaking tests
The candidates were chosen from Year 10 and Year 12 students by the teachers of the schools, on a voluntary basis. The pilot team wanted to get a sample that was reasonably representative of each particular school, so the teachers were asked to encourage and motivate weak students as well as strong ones to participate. Pairing up the students was important, since some of the pilot tests were to be conducted in paired format. Test administrators discussed this issue with the teachers in advance to achieve the most ideal

pairings in the circumstances. In most cases, students were paired with friends of roughly the same ability level.

Training test administrators, interlocutors and assessors
The term 'test administrator' and 'interlocutor' in most cases refers to the person who actually conducted the oral test, but there were a few occasions when the administrator was a different person. Usually several members of the pilot team were present in each school where the pilot tests were administered, and they acted as interlocutors, assessors and observers in turn according to a previously set timetable. In addition, one person acted as usher, in order to entertain waiting students, to reassure them and to administer the questionnaire as students emerged from their test session. The test administrators and the interlocutors received detailed instructions before the pilots, and they also had a package of test booklets, so the booklets were available while they were preparing. The instructions gave clear and detailed guidelines about the following:

- how to prepare the speaking test booklets for the pilot
- the content of each booklet
- how to arrange the timetable of the tests (paired vs. individual interviews)
- the usher's tasks
- practical issues (location and arrangement of the examination room, waiting room, labels, etc.)
- how to conduct the test from the very beginning to the very end
- how to assess the performances
- how to get students to complete the questionnaires
- how to keep a record of all that happened during the tests.

Before each test day the team also held a preparatory meeting, where they went through the whole process again and discussed possible problem points.

Technical preparations
As the tests were to be filmed, technical preparations presented a considerable problem. In each school the administrators had to find a suitable room which was quiet enough and allowed the camera crew to move around while being as unobstrusive as possible. Seating arrangements were crucially important.

The tests
When the candidate(s) entered the room, the interlocutor introduced him/herself and asked them to be seated. Then s/he checked the candidate's name against the previously prepared list, and began a general conversation. The aim of this conversation was to relax the candidate(s) as much as possible, so it took as long as was necessary to achieve this (maximum 5 minutes). The interlocutors received clear guidelines about what kind of questions to ask. The topics of the three tasks to be administered had to be avoided. When the candidate(s) seemed relaxed, they were given the first task.

After the third task, the interlocutor thanked the candidate(s) for participating in the test, and asked them to complete a questionnaire in Hungarian. This happened outside, in the waiting room with the assistance of the usher. Meanwhile, in the examination room preparations were being made for the next candidate(s). In addition, the assessor and the interlocutor briefly discussed the previous performance and the results were written up. These were later made available to the class teacher.

At the end of each test day the team had a meeting with the local teachers, informed them about their students' performances, and answered questions about the planned exam reform. These meetings appeared to be very motivating for the classroom teachers, and many schools applied to participate in the second round of the pilots (in April 1999) because of this positive experience.

Oral feedback (Speaking tests)

After each pilot we had the opportunity to talk with the students about their immediate impressions. They responded positively, saying that they found the tests interesting. The unusual circumstances of the oral test (being filmed), naturally, caused excitement and nervousness in many cases, but despite this they felt the advantages of this new way of testing their speaking skills. A lot of students had very positive opinions about the paired format. They also found the tasks very life-like. Generally, they found the prompts helpful, and expressed their wish to have them in their own forthcoming school-leaving examinations in English. Chapter 11 contains a summary of the written feedback from students.

Administration of the Writing tests

Selecting candidates

In many cases, there was no need to select candidates again, since the same students took the test who volunteered to participate in the Speaking pilot. In the case of those schools which had not participated in the Speaking test, intact classes were identified in Years 10 and 12.

Training test administrators and administering the test

No special arrangements were necessary, as schools are used to dealing with written tests. Administrators were given detailed instructions on seating arrangements, the distribution of booklets and questionnaires, invigilation and collection of completed Writing tasks and questionnaires. Students were seated individually, with desks a reasonable distance apart, and were told not to copy from their neighbours. All four booklets were distributed in sequence in any one class, so that neighbouring students received a different booklet to reduce cheating, and in order to ensure that booklets would be distributed equally across classes (distributing Booklet 1 to one class, Booklet 2 to another class and so on would have resulted in booklet samples biased by the ability of the particular class). This way we hoped to avoid major biases in our sampling across classes.

The actual administration of the tasks was scheduled to take a double lesson of 90 minutes but in fact everybody completed well within that time, and students were allowed to leave the room once they had completed the questionnaire, provided that they did not disturb those still writing.

Feedback from the schools

The schools responded to the pilot tests positively. The students, the teachers and the management of the participating schools felt it was a great privilege to take part in the Examination Reform in this way. They could test their own students' real performances in an exam-like situation, and at the same time they got valuable information about the currently developing new school-leaving language examination system. They greatly appreciated the fact that they could contribute by giving their opinions about the tasks. The large number of schools applying for participation in the second round of the pilots

provided convincing evidence about the importance of involving stakeholders in the process of developing the new examinations.

The April 1999 pilots of Listening, Reading and Use of English

Rationale for sampling schools

The main factors affecting arrangements for the administration of the pilot tasks were the following:

1. There would be six different booklets in a total of four versions, as the Listening booklets were to be combined with a Reading/Use of English booklet and the two Reading/ Use of English booklets that were not combined with Listening booklets would contain a larger number of tasks (for the rationale, see Chapter 4).

2. All test papers were to be taken by 250 students, one-third of whom should be Year 10 students, and the rest Year 12. There were several reasons why we decided to organize the piloting in this way.

First, all the items we had to pilot had been produced within the joint project between OKI Budapest, OKI Szeged and the British Council. It was left to the itemwriters to decide which of the three levels (Basic for the Basic exam, Intermediate or Advanced for the School-leaving exam) they intended their tasks to be at. Due to the lack of empirical data, however, they could only rely on their knowledge and experience of their students and teaching materials. Therefore, one of the main aims of the piloting was to gain information on the level of the tasks, thus providing the itemwriters with guidance for their work in the future as well as getting statistical results for the further development of the exams.

Second, we did not know much about the level of the target population of either exam, and for this reason it seemed sensible to assess their performances separately and in comparison as well. Randomization of the booklets within classes (two versions in a class) would also contribute to our getting a clearer picture of the Hungarian reality.

3. The sample of approximately 1000 students had to include schools that took part in the CITO project of 1995, as items used and calibrated then were anchor items in our tests. This would allow us to compare those schools' performances on the same items then and now.

4. Although the time needed to complete a task was estimated by itemwriters, it was impossible to be certain how many minutes a student would need to finish a test paper. In fact, learning more about this was another important aim of our piloting, which is why we asked the students to take notes of exactly what time they started and completed each task. Nevertheless, we estimated that 135 minutes would be ample to allow students to complete all the tasks.

5. It was important that project members administered the tests as observers, for several reasons. This was the easiest way to ensure that the test material remained confidential (the testing materials along with the instructions for the administration were posted to them a few days before the trialling period). Their presence in the classrooms together with the teachers was regarded as an appropriate means of ensuring that students took the test as seriously as possible because the observers were able to let them know the rationale, and explain the importance of the piloting. Moreover, we also believed that project members' subjective observation of and remarks on the students' and their teachers' reactions to both the tasks and the administration would give us invaluable

insight into the atmosphere and conduct of the piloting, as well as provide feedback to project members themselves on how well the tasks had suited these students.

The observers' tasks were to contact the teachers in the schools they would administer the test in and agree on the exact date of the trialling. After the administration, they had to fill in the report sheets enclosed (see Chapter 9) and send the packages back to OKI Budapest.

Sampling schools

Although our piloting was not planned to be carried out on a representative sample, we tried to keep it as close as possible to the final candidature both in terms of abilities and backgrounds. Similarly to the actual division in Hungary, the sample consisted of three main groups: grammar, vocational and combined grammar and vocational school students, each being approximately one-third of the sample population. There were low-level groups who had been studying English for 2 years, intermediate groups who had been studying for over 2 years and more advanced groups, studying for over 4 years, but we must always keep in mind that the categorisation of the classes was mainly based on the teachers' subjective judgement and whatever classification and terminology the schools usually used for a given number of hours per week and the course books used in class.

As we had decided to involve as many schools that had participated in the CITO project as possible, the first step was to compare the list of these to that of potential observers in terms of their geographical location. Contrary to expectations, this practical consideration did not limit the sampling, since the CITO project used a representative sample covering cities, towns and villages all over the country, and the same applied to our observers: they came from all parts of Hungary. After careful consideration, 28 of the original 38 CITO schools were contacted. As the majority (22 schools) were grammar schools, the rest of the schools invited consisted mainly of vocational and 'combined' schools. They were selected on the basis of two criteria: 1. their location: we chose ones near the CITO schools for practical reasons; 2. their orientation: we aimed to include a wide range of vocations, varying from future nursery school teachers to technical or commercial experts. Altogether 75 schools were contacted, out of which 39 replied positively. In the end, 26 were selected as they could provide us with one Year 10, and two Year 12 groups with the necessary number of 'Basic', 'Intermediate' and 'Advanced' students.

Each school received a letter from OKI with the confirmation of their application for the pilot. After this the observers visited the schools and informed the management, the Head of English and the teachers involved about the exact procedure. Finally, the time of the pilot was agreed.

The observers and invigilators received detailed guidelines before the tests. These guidelines included the following:
- information about the booklets
- instructions for preparation of the booklets for the test
- seating arrangements
- necessary equipment (cassette players, suitable writing equipment, scrap paper,etc.)
- number of invigilators in each room
- timing, registering time
- distributing and collecting the booklets, handling of spare copies and cassettes
- giving instructions to the students
- student questionnaires

- how to administer questionnaires.

1029 test papers were sent out to the observers. According to plans, the following number of students were to complete the six different test papers:

Booklets	List 1 + R/UE 1	List 1 + R/UE 2	List 2 + R/UE 1	List 2 + R/UE 2	R/UE 3	R/UE 4
No. of students	129	130	128	128	257	257

20 observers administered the tests in 1-3 schools each, i.e. 3-9 groups of students per observer. In the end, just under 1000 students took the test in 9 grammar schools, 10 vocational and 7 'combined' schools. The students differed in the number of years they had been studying English (2-14), the level of proficiency their teachers ascribed to them (from beginners to upper-advanced), the number of contact hours per week (from 2 to 8 and over), the size of the English class (from 8 to 26) and the course books used (Hotline, Grapevine, Access to English, New Cambridge English Course, Headway, Streetwise, etc.). See Chapter 16 for a discussion of test results in relation to these background variables.

The number of schools which volunteered to participate in our piloting shows that both school managements and English teachers are looking forward to the introduction of the new exam and are willing to take every opportunity to learn more about plans for the exam, and to prepare their students better for it.