

Chapter 9

Feedback from Observers, Teachers and Students

Öveges Enikő

Analysing the Report Sheets (observers' and teachers' comments on the tasks and the administration of the Written test)

Teachers and observers were asked to complete a sheet reporting on the test administration, their reactions to the pilot tests, and their perceptions of the students' reaction. The idea behind using this Report Sheet was twofold. Apart from wishing to get concrete data such as the number of students taking the test or the name of the class in order to confirm the data we already had on schools and planned numbers of students, we wished to gather subjective descriptions of the atmosphere of the testing situation from the observers and we hoped to involve the EFL class teachers whose role in the piloting was otherwise minimal. As the observers had all received training in testing, it seemed sensible to take their observations into account, and the teachers' evaluation of the tasks, their level, or the organisation of the test would provide us with guidance for our future work.

The Report Sheet was sent to the observers along with instructions for the administration of the exam. It requested concrete data such as the name of the observer, name of school and group, number of students taking the test, the exact time and date of the test and how many students took the test at any one time in a classroom. Despite the fact that this was not a genuine test-taking situation, we thought it would be useful for future discussions if we gathered information on what time the first and the last students submitted their papers. We expected a short paragraph on the students' questions and reactions during the test, and another one on the teachers' opinions.

Some of the data was merely used to confirm that the piloting went in accordance with our original plans and they also helped us to spot missing groups, extra groups or groups that were larger than expected.

Unfortunately, the questions asking for subjective judgements of the students' and teachers' reactions were not clear enough to give guidance on what comments were expected. Consequently, the observers interpreted them in a variety of ways, thus providing less useful information than expected.

What follows describes the students' and their teachers' remarks with respect to the booklets and the administration of the test. The comments are listed without comment, as an evaluation comes at the end of this section of the chapter. The lists progress from the positive to the more critical comments. The number of schools expressing similar views is provided in brackets.

Grammar school students' questions, reactions

Regarding rubrics

- No assessable remarks.

Regarding tasks

- ‘Tasks were good, varied and interesting. It was nice to be instructed in which lines the answer should be looked for.’ (2)
- Students seemed to be frustrated by the thought of having to do such tasks in the érettségi as well. (7)

Regarding administration

- There are too many questions in the questionnaire. Students did not seem to remember how many contact hours they had years before.
- Having to write the time they started and finished the tasks on the task sheet did not allow students to come back to and think about them again.
- Were they allowed to leave the room as soon as they had submitted their papers?
- Did the tasks have to be solved in the order they were presented in the booklets?
- They would like to get feedback. (almost all)

Grammar school teachers’ questions, reactions

Regarding rubrics

- No assessable remarks.

Regarding tasks

- ‘Our students were familiar with all the task types, and I think, the booklets were perfectly suitable for assessing language proficiency.’ (3)
- The tasks were lifelike and motivating. (2)
- There were several task types in the booklets that the students had been unfamiliar with, e.g. cloze test. (3)
- The extremely wide vocabulary required was the main reason why students were not too successful in doing the tasks. In fact, there was too much emphasis on vocabulary. (6)
- Students would need more time to read the questions in the listening tasks beforehand. (3)
- Teachers appreciated that the booklets seemed to start with easier tasks, from there going on to more complicated ones.
- Although the booklets contained tasks perfectly appropriate to assess students’ language proficiency, the present conditions of language teaching in Hungary make it an unrealistic aim to use tasks of these types and this level in a high-stakes exam. (2)
- With easier or more common tasks (shorter texts), fewer items would have been left unanswered.
- With fewer tasks, students would have been able to devote more attention to the last tasks and would have made fewer guesses. Some students found the test highly exhausting due to its level and length. (4)

Regarding administration

- The poor quality of the recordings had an impact on the students’ results. (5)
- Being allowed to use dictionaries would have been helpful.
- Feedback on their students’ work is definitely expected. (almost all)

Vocational secondary school students' questions and reactions

Regarding rubrics

- In some places students' reactions indicated that they did not understand the instructions (they seemed to encounter too many unknown words in them), they would have preferred them in Hungarian. Sometimes they did not even start doing the tasks because of the complicated way of formulating what to do. (9)

Regarding tasks

- Students said that they had found the listening tasks extremely difficult due to both their level (fast pace) and the quality of the recording (heavy background noise). Some of them seemed rather frustrated by the end. (4)
- Students rated the 'spot the error' tasks the least manageable.

Regarding administration

- They would like to get feedback on their achievement. (most of them)
- Filling in the questionnaire was hard because of inadequate instructions.
- In some schools, no student questions or reactions were reported. (7)

Vocational school teachers' questions and reactions

Regarding rubrics

- The role of the examples was highly acknowledged and appreciated.
- The instructions should be in Hungarian for vocational school students.

Regarding tasks

- Both the vocabulary and the grammar content of the tasks seem to be well above or irrelevant (e.g. the topic of the Zukol Listening text is not appropriate for this age group) to the language proficiency of even the best students in these schools. The same applies to some of the strategies necessary to do the tasks, for example the 'spot the error' ones. Similar comments appeared on most of the Report Sheets. (10)
- However, there was one teacher who thought that the Reading and Use of English tasks were built on a wide range of vocabulary, and were easy to solve.
- There should be more, but shorter tasks (including fewer items) and the options should always be provided.
- Teachers felt that it would be advisable to let students listen to the texts twice even at the Advanced level (2). They felt that at the very least we should have experimented to see to what extent students' results would have been better after two listenings.

Regarding administration

- There were no instructions regarding seating arrangements or classroom size.
- Teachers lacked guidance on what to do when a student would like to change an answer (e.g. are they allowed to cross it out?).
- The quality of the recordings was annoyingly low (even in special language labs), thus making the tasks more difficult than they would have been otherwise. Such comments were common in the Report Sheets.
- One school expressed misgivings about asking Year 10 and Year 12 students to do the same tests.
- Detailed feedback on the students' results is expected.

Implications for the future

According to some teachers, the fact that the tasks were skills-based and integrative and therefore required the students to apply their lexical and grammatical knowledge seemed to contradict the present teaching practice in many Hungarian secondary schools. Others appreciated the partly innovative nature of the test, thus showing the obvious variety of different ways of teaching English in Hungary as evidenced by the reactions to the task types or the vocabulary required in the tasks.

Some reactions show that the aims of this piloting were not made clear enough. Several observers and teachers said that the booklets should include tasks of the same level of difficulty or that the tasks were too difficult, forgetting that our intention was to assess their difficulty level and that, therefore, the tasks had to be taken by students of different levels and backgrounds.

The Report Sheets drew our attention to some concrete errors such as misleading or inadequate instructions (e.g. in Task Two, Booklet 4) or a missing example in Task Four, Booklet 4.

More consideration has to be given to how the quality of the recordings can be significantly improved in the future.

Some of the criticisms were due to the fact that this was an experimental pilot situation (e.g. the long questionnaire or the starting/finishing time) and such problems will obviously not be present in the actual exam.

Students asked questions we definitely had not expected, for example whether they could give their short answers to the Listening task in Hungarian. Questions like this will only be eliminated when familiarity with this exam increases, and the exam is contrasted with the present State Foreign Language Examination where doing some listening tasks in Hungarian is part of the requirements.

Teacher training and familiarisation with the new examination is clearly essential in order to ensure that the communicative approach we wish to see reflected in the new exam is widespread and common in Hungarian language teaching practice as well.

Students' questionnaires

In addition to a questionnaire targeted at students' biodata and language learning experience, we included a further brief questionnaire at the end of the test administration that gathered students' opinions of and familiarity with the tasks they had taken. The questionnaire consisted of two main parts (see the translation below). The first focused on how difficult students thought the tasks were, and the second one was a simple Yes/No question asking whether they had encountered the given task types before.

Table 9.1: Translation of the original chart.

The table below is intended to gather information about your opinion of the particular tasks. Please note how difficult you found the tasks, and in the last column, whether you have done tasks of similar types.

	very easy	easy	manageable	difficult	very difficult		Have you ever done similar tasks? Yes/No
Part 1						Part 1	
Part ...						Part ...	

Students' familiarity with the different task types

Table 9.2: Listening tasks

Task type	Percentage of students who have done a similar task
B1/1. multiple choice	54.1
B1/2. multiple matching	34.4
B1/3. TRUE/FALSE	52.9
B2/1. multiple choice	58.0
B2/2. Short answer with cue	45.7
B2/3. Short answer	44.2

Approximately half the sample population reported that they had met the task types involved. The two most widely known types proved to be as expected, i.e. multiple choice and True/False, possibly because both of them are part of the written State Foreign Language Exam and findings of earlier research have shown that listening comprehension in the classroom is still mainly focused on preparing for that exam.

However, the fact that more creative task types such as multiple matching or giving short answers in the target language are only slightly less familiar may suggest that more realistic and communicative tasks are also being taught in schools. Including a wider range of tasks like these in future pilotings will allow us to explore further the reported familiarity of task types.

Table 9.3: Reading tasks

Task type	Percentage of students who have done a similar task
B1,B2,B3,B4/P1, multiple choice	61.2, 68.0, 72.6, 67.1
P2, ordering	50.4
P3, ordering	41.1
B2/P2, ordering	33.6
P3, ordering	60.5
B3/P2, multiple matching	52.9
P3, reading for specific information	24.0
P4,multiple matching	28.2
P5, ordering	57.1
B4/P2, ordering	44.4
P3, multiple matching	35.9
P4, multiple choice	72.6
P5, multiple matching	35.5

Table 9.4: Use of English tasks

Task type	Percentage of students who have done a similar task
B1, B4/P4, spot the error	36.8, 36.0
B1/P5, B2/P5, B3/P7, B4/P7, sequencing	59.3, 62.1, 81.5, 76.9
B1/P6, spot the error	27.9
B2/P6, gap-filling	64.4
B3/P8, gap-filling	67.6
B4/P8, spot the error	26.9
P9, multiple choice	79.9
P10, transformation	63.2

The most widely known tasks in Reading and Use of English appear to be those that are relevant to exam preparation (both for external exams such as the State Foreign Language Exam or international language exams, or the current school-leaving exam with its non-communicative, unrealistic tasks):

- multiple choice (e.g. Part 1/Reading/'Practice makes perfect' is familiar to at least 65% of respondents, B2/Part 9/UE/mc – 79.9%),
- sequencing/ordering (for Reading, it is familiar to between 41% and 60.5%, for Use of English it seems that students are more familiar with the task type where no clue is given – B3/Part 7/UE/Compaq: 71.5%)
- gap-filling (B2/part 6/UE/What on Earth...?: 64.4%)
- transformation (B4/Part 10/UE: 63.2%).

Familiarity with multiple-matching tasks for assessing reading varies from 28.2% to 52.9% (matching statements with advertisement with five distractors at one extreme and matching missing bits with advertisements but with one distractor at the other extreme). This task type is mainly used in recently published, more communicative textbooks, and so the fact that at least one-third of the students reported to be familiar with it is noteworthy.

The least familiar task was the spot the error tasks (26.9 – 36.8%). This was the task type that most students and teachers also claimed to be new in the feedback sheets. Despite appearing in several modern coursebook series used in Hungary (e.g. Headway), it appears to have remained either unknown or unpopular with Hungarian teachers of EFL.

Students' opinion of the difficulty level of the individual tasks

Students were asked to assess on a five-point scale (from very easy to very difficult, see Table 9.1 above) how difficult they found the particular tasks.

Listening tasks

Multiple choice. Approximately half the students found such tasks manageable, i.e. of mid-level difficulty. This result could have been influenced by the topic of the text (it was built on cultural facts that some students might have encountered before), and the fact that the answers had to be given (by circling the best option) during the pauses in the text.

True/False and Short answer. Both tasks were based on a text that was probably not appropriate to this age group and whose recording was of relatively poor quality. These two factors might be the reason why students reported these tasks to be either very difficult or difficult (short answer: 95%, T/F: 70%).

Multiple matching. More than half the population found this task either difficult or very difficult (61%). This seems to correlate with the low familiarity of this task.

Short answer with cues. The majority of the students said this was either difficult or very difficult (115 + 109). If we take into account that half of them were familiar with this task type, the reported difficulty could have been caused by either the text (topic, vocabulary) or the quality of the recording.

Reading tasks

Multiple choice.

Anchor task: 'Practice makes perfect'. Highly similar results can be seen for each booklet: 50%, 54%, 55% and 55% said that the task was manageable. (It was also a very familiar task type.)

Booklet 4, Task 4: Letter to the Editor. A shorter and easier text: 80% found it either easy or manageable.

Ordering tasks. Four out of the six ordering tasks were reportedly either easy, manageable or difficult, i.e. fell into the mid-values of the scale (e.g. B1/P2: 2/91, 3/83). The only one claimed to be difficult and very difficult (4/117, 5/77) was Part 5 in Booklet 3 where the task was to organize chunks of texts. The reason for the difficulty might be the length and complexity (especially the vocabulary) of the text.

Multiple matching tasks. The reported difficulty level of these tasks matches their intended difficulty. The two tasks in Booklet 3 (Tasks 2 and 4) both seemed to be easier (3/94, 3/103) than the ones in Booklet 4 (Task 3: 4/101, 5/54, Task 5: 4/107, 5/69).

Use of English tasks

Multiple choice. Similar results were found here to the reading tasks: 50% of students thought it was a manageable task. As this is a well-known task type, the difficulty was probably caused by the content of the test.

Sequencing. Students had to order sentences/events of a story. In the booklets where the first item was given as an example the majority of students chose 'very easy' or 'easy' (1: 86, 1, 2: 82 + 86 students), while most students chose 'easy' and 'manageable' where there was no clue (2, 3: 74 + 77, 2, 3: 61 + 67).

Gap-filling. Both gap-filling tasks were claimed to be either difficult or very difficult. The task that students found easier was the one where they had limited choice, i.e. they were instructed to use only prepositions (59% said this was 'difficult'). In the other task they were allowed to use words from any part of speech, which appears to have made it more difficult for them (71% said it was very difficult).

Transformation. Despite the fact that this widely used sentence-based task tested structures commonly practised in classrooms, more students found it difficult than manageable (3: 68, 4: 80, 5: 67).

Spot the error. All three tasks proved to be difficult (Move over Webster – 4: 114, 141 students or B4/Task 8 – 4: 92, 5: 108). They were designed to be Advanced (text type, length, structures used, etc.), but they were also unknown as task types to a large number of students.

Summary of familiarity and difficulty of tasks

The following table summarises the data obtained from this questionnaire.

Table 9.5: Summary of comments on difficulty

Task type Listening	Difficult/ very difficult %	Familiarity %
Booklet 1 task 1	39	54
Booklet 1 task 2	62	34
Booklet 1 task 3	71	53
Booklet 2 task 1	33	58
Booklet 2 task 2	83	46
Booklet 2 task 3	95	44

Task type Reading	Difficult/ very difficult %	Familiarity %
Anchor task, all booklets	24	67
Booklet 1 task 2	15	50
Booklet 1 task 3	40	41
Booklet 2 task 2	14	34
Booklet 2 task 3	20	61
Booklet 3 task 2	19	53
Booklet 3 task 3	12	24
Booklet 3 task 4	29	28
Booklet 3 task 5	82	57
Booklet 4 task 2	53	44
Booklet 4 task 3	66	36
Booklet 4 task 4	16	73
Booklet 4 task 5	75	36
Task type Use of English	Difficult/ very difficult %	Familiarity %
Anchor task 1, all booklets	61	37
Anchor task 2 all booklets	18	70
Booklet 1 task 6	71	28
Booklet 2 task 6	81	64
Booklet 3 task 8	97	68
Booklet 4 task 8	86	27
Booklet 4 task 9	37	80
Booklet 4 task 10	63	63

Interestingly there is no correlation between perceived familiarity and perceived difficulty ($r = -.14$, $p = .487$).