

## Chapter 18

# Use of Piloting Data in an In-service Course to Help Teachers Prepare for New Examinations

Együd J. Györgyi, Gál Ildikó and Philip Glover

### About the course

International research and experience, for example in Sri Lanka, Slovenia and the Baltic States (see Wall & Alderson, 1993, Wall, 1996a, Gabrscek & Bethell, 1996, Wall, 1996b, Mazuoliene, 1996, and Együd, 1998) suggest that examination reform is more likely to achieve its aims if accompanied by effective familiarisation and in-service teacher training programmes for teachers. This chapter describes an in-service teacher training course which accompanies the Hungarian English Examination Reform Project and which aims to contribute to the effectiveness of the impact of the new examinations. It also describes how data from the pilot exams were used on the course.

### Course design and development

The development of the course has taken two years so far, and has been undertaken by an in-service (INSET) team within the Exam Reform Project. Members of the team were appointed specifically to develop INSET courses, but they have worked closely with the developers of the new examinations, have attended the same training sessions as outlined in Chapter 2, and have themselves written test items for possible piloting.

There have been four phases in the development of the INSET course:

Phase 1: Training the team of course designers and preparing the actual writing process.

Phase 2: Designing and piloting the first version of the course.

Phase 3: Revising the materials on the basis of the first pilot and piloting the second version.

Phase 4: Revising the materials again and writing the latest, third, version.

### Personnel

Pilot course writers, developers and trainers were, with one exception, members of the INSET Team who had been working with the Reform Project from the start. The reasons for this were:

- Trainers knew a great deal about the progress and development of the Examination Reform.
- They knew the course very well, having designed it.
- Training on the course gave first-hand experience that helped with later revisions.
- Teachers have many doubts and fears about the new examinations. Since the trainers were practising secondary school teachers, it was more likely that the participants of the course would find the ideas about teaching for the new exams believable and practical.

The first pilot was based at a provincial grammar school in Eger and the trainers were 3 members of the INSET Team. The second pilot was organised in co-operation with a

county pedagogical institute in Debrecen, with 2 members of the INSET Team and a local university lecturer participating as trainers on the course.

Throughout the course development period and later, the Team had two British Council advisers: Prof J Charles Alderson, and Philip Glover. Both advisers were present during the delivery of the course. At various points the team also had support from other consultants, namely from the University of Manchester (Richard West, Jane Andrews) and from Lancaster University (Dianne Wall, Caroline Clapham).

### **Type of course**

The course was an intensive 60-hour in-service training course. Its mode was a combined one: it consisted of 35 contact hours (14 + 14 + 7 with each block one month apart) and 25 hours of distance study. The aim of this organisation was to allow teachers to try out various ideas discussed during course sessions in their own classrooms and to report back on the success or problems they experienced. Three blocks were essential to cover the wide variety of material involved, and to allow teachers more time to digest and react to the course content.

### **Target population**

The course was designed for secondary school teachers of English. This is a broad category, including teachers with various qualifications and backgrounds (teachers with a teacher training college degree, teachers with a university degree, former Russian teacher re-trainees, teachers at vocational secondary schools, teachers at vocational training schools, teachers at grammar schools). However, all these groups of teachers were perceived to share a common need, namely to get as much information and help in connection with the new examinations as possible.

### **Need for the course**

There was evidence before the course design procedure started of a definite need for the course among secondary school teachers. In the course of several formal and informal events (meetings, conferences, in-service courses, workshops, professional fora) teachers had voiced their desire to know more about the planned new examination system. In addition the observations and interviews carried out during the Classroom Observation Project for the Baseline Study (Fekete et al, 1999) showed there was both a great need and a desire for exam-related training. International examples (see above) had also shown that informing and winning over the teachers would help the implementation of the new system.

### **Aims of the course**

The main aims of the course were:

1. to inform teachers about the principles, content and procedures of the likely new examination system
2. to encourage teachers to have a positive view of the likely new examinations
3. to develop teachers' awareness of teaching and testing techniques
4. to study and analyse the classroom implications of the new examination and to experiment with implementing some of these implications in the classroom.
5. to support teachers in their aim of raising standards of student performance.

Course outcomes were expected to be:

- raising teachers' awareness in terms of what skills can be tested and how they can be tested, and also how students can perform in the course of these tests

- building teachers' confidence that
  - students, even weak ones, can perform the tasks
  - new exams are being developed carefully, to professional standards with procedures that will help, not hinder students and teachers, and with due attention to the needs and wishes of teachers.
- that teachers would
  - be well-informed about the new examinations
  - be supportive of the new examinations
  - use better teaching and testing techniques (positive washback)
- that students' language skills and test performance would improve.

Since this was a pilot course, it also aimed at:

1. trialling materials and procedures devised to support the successful introduction of the new examinations in Hungary
2. enabling course designers and writers to develop materials and procedures further
3. developing trainers' training skills and providing experience that can be shared with future trainers.
4. examining the responses of a sample of secondary school teachers in a real training situation and collecting views for future use.

### **Principles relating to the overall design of the course**

The general training approach of the course was based on the cyclical process of experimental learning (Kolb, 1984), recommended by Richard West and Jane Andrews (University of Manchester). The cyclical process consisted of the following stages:

Stage 1: Concrete experience (This could be what teachers do in the classroom already, or what they already know about something.)

Stage 2: Reflective observation (Reflection on experience in the light of new exam-related information.)

Stage 3: Abstract conceptualisation (Drawing conclusions from the previous stages, for example teaching implications.)

Stage 4: Active experimentation (Involving classroom work, for example putting ideas into action.)

Each cycle contained four recurring elements:

- reference to course books and other resource books
- study of the *Working Document* (Vándor, 1998)
- classroom research and observation
- teaching implications.

### **Structure of the course**

The two main components were:

- a) 35 hours of contact mode (spent in training sessions) in 5 days over 3 blocks, divided as follows:

Days 1-2 : 14 sessions, Days 3-4 : 14 sessions, Day 5: 7 sessions

- b) 25 hours of distance mode between Days 1-2, Days 3-4 and Day 5 (spent on assignments which required classroom investigation and application).

## Content of the course

The course content consisted of detailed study of the *Working Document* (Vándor, 1998), reference to course books and other resource books, classroom research and observation and teaching implications.

Following the content of the proposed School-leaving examination, the course consisted of 5 core modules, 4 of which were skill-based modules (Reading, Listening, Writing, Speaking), and a Use of English module. Each module repeated the four stages and the elements of the cyclical process described earlier.

Besides the 5 core modules of the course, the contact sessions included introductory, review and concluding sessions as well.

## How the pilot data were used

### a) Reading

#### *i) Discovering and evaluating tasks and text types in the Working Document*

At the time of the Eger course no pilot tasks were available, so other examples were used. By the time of the Debrecen course, we were provided not only with the piloted exam tasks but empirical results from the pilots as well, so we could make good use of these on the course.

Participants looked at three sample tasks from the pilot exams: Sample task A *Fatherhood has transformed me*, B *Advertisements* and C *Tadpoles*. (See Chapter 13 in this volume.)

In one of the sessions participants were given this task:

Task 2 – analysing the tasks

- a) Look at the examination document and identify text and task types.
- b) Decide which of the reading skills we looked at in session 1 are needed to complete the task.
- c) Give your opinion of these tasks. How difficult are they? Are they suitable for any of your students? Are they interesting for your students?

Having identified the text and task types, participants discussed the questions. When sharing their ideas with the whole group, participants found the topic of sample A challenging and interesting. They expressed the fear that the rubrics would be difficult for their students and felt the layout and example were confusing. Most participants suggested in this session or later that the level of all three tasks was Intermediate. Very few participants suggested that sample task A was Advanced.

After the discussion participants were shown the mean scores from the pilot exams.

Pilot exam data- Reading Paper		
Mean score of all students		
Sample task A	Sample task B	Sample task C
29%	80%	76%
Intended level: Basic	Basic	Basic

This made it possible for participants to compare intuitions, feelings and fears about the sample tasks with the empirical findings of the pilots. The most striking difference between the teachers' intuitions and real data was about the level of the tasks. Pilot data suggests that sample task B and C are suitable for Basic level students and sample task A for Advanced students. This showed how intuitions can be wrong when it comes to

assessing the difficulty of tests, tasks and texts. It also made the point that the level of difficulty is determined primarily by the task, not by the text.

### **ii) Teaching implications**

Having looked at examples from the pilot exams, teachers were asked to relate the new exams to the classroom and course books and work out teaching implications.

Suggestions made by participants included the following:

1. When we test reading it should be as close as possible to real life reading, using authentic materials.
2. When we teach reading it should be as close as possible to real life reading, using authentic materials.
3. You can make the task fit the student, even with a difficult text.
4. Reading skills can be developed in class...
  - a) by encouraging students to read as much as possible.
  - b) by raising students' awareness of the sub-skills.
  - c) by discussing reading strategies in class.
  - d) by encouraging good readers to share their secrets with other students.
  - e) by teaching vocabulary and grammar from a text after students have read it.
  - f) by ensuring students have a realistic reason to read.
  - g) by ensuring there is a result or outcome from the reading.
  - h) by integrating reading with other skills.

### **iii) Trying out sample tasks in the real classroom through the homework assignments**

Participants were asked to look at the results of the reading sessions in the light of their own classroom experiences through the reading assignment. In Reading session 5 (Day 3) participants were asked to discuss the experience of trying out the exam tasks with their students. The task instructions were:

Task 1

- a) In a group discuss the Reading tests you gave your students.
- b) What were your expectations before giving the students the test?
- c) How did the students perform in the test, and what did they say about it afterwards?

After carrying out the task a number of pleasant surprises were noted by most course participants. One commented 'I had more problems with the rubrics than the students'. The assignment sheets, on which teachers recorded their students' results, showed that the students performed better than their teachers had expected, confirming what had been suggested by the pilot exam results. Some students did find the layout of the tasks rather confusing and several complained about dictionaries not being allowed. One teacher summed up her experience like this, *'Most students complained about the layout, the novelty of the task. Some felt frightened by the unfamiliar vocabulary, some others enjoyed the challenge that they had to use their logic and intelligence as well.'*

### **b) Listening**

#### **i) Discovering and evaluating tasks and text types in the Working Document**

Out of the five sessions on Listening, originally one session was designed to deal with the pilot exam tasks: Listening session 3. However, as pilot tests were not available, instead, as part of their assignment, participants were expected to do either a listening or a reading task from the course books they used with their students and they had to write a self-observation report about that lesson.

By the time the Debrecen course started the pilot exam tasks were available and were copied onto cassettes for each participant, so it was possible for the teachers to trial the tasks with their own pupils as part of their assignment. This made it necessary to have one more session on pilot exam tasks where participants could discuss their findings, so in Debrecen there was an activity in Listening session 5 Day 3 devoted to this purpose.

As things keep developing, and we now have the results of the pilot exams, it seems advisable to design a further activity to see and discuss the results of the pilots with the teachers as was done in Debrecen with the Reading tasks. Here is one idea how this could be done on future courses:

### **Listening session 3**

#### **Task 3 – discussion** Level of difficulty

- i) Look at each sample task and each question. Decide on the order of difficulty of each question. Number the most difficult question 1, the next most difficult question 2 and so on. After you have finished, compare your answers with the results on page....\* Do not look before!
- ii) What does the mean score suggest about the level of the tasks?
- iii) Give some reasons for this mean score.
- iv) What conclusions do your answers suggest?

\* On this page the table of Pilot exam data can be found.

When designing the material for the Listening sessions, special care was taken to avoid repetition of task types. As a result, the task types which had been identified when discussing course book tasks in Listening session 2 or when doing loop-input tasks based on a recorded interview, were purposefully different from the ones in the sample exam tasks. Consequently, most of the potential exam task types on the list in the Working Document were dealt with in some way during the 5 listening sessions.

The pilot exam tasks used in both courses were the following:

Sample task A *Harry and Sue* – following routes on a map/ identifying the order of events

Sample task B *A special object* – completing text

Sample task C *Rudolf Zukol* – short answers

(See Chapter 12 in this volume)

As with the Reading tasks, participants were asked to carry out and evaluate sample exam tasks. These are the activities from Listening session 3:

Task 1 Carry out the sample exam tasks

Task 2 Analyse the tasks

In your group comment on the sample exam tasks as directed. After sharing ideas in your group, a volunteer from each group will be asked to report to the others.

Group A Did you like the tasks? Which did you like most/ least? Why?

Group B Identify text/ task types in the Working Document pages 11-12.

Group C For which level do you think each task was intended?

Group D Comment on the rubrics.

These tasks enabled participants to become familiar with the pilot exam tasks. It was easier for participants to identify the text and task types of the sample tasks than it had been in the case of the Reading tasks.

Unlike with the Reading tasks, the levels of the Listening tasks were put approximately where the pilot exam results suggest. However, it should be noted that the quality and variety of Listening tasks available was much more limited than for the Reading exam.

When discussing the questions about the pilot exam tasks there was general approval of having the rubrics in English, but a few participants said they did not like the layout of the tasks, because they thought they might be confusing for the candidates.

Both in Eger and Debrecen participants were against the idea of one *single listening to a text* (sample task C). They found this idea unfair, contrary to pedagogic practice and likely to increase feelings of anxiety.

A participant in Eger remarked, *'I am absolutely against the idea that on Advanced level gist listening may be tested with one playing. I suppose the Advanced level listening bits will be more difficult than the Intermediate listening ones. Why should we make these bits even more difficult for the candidates at the exam.'*

Another similarity between the opinions of the participants on the two courses was that Sample Task A was preferred most, while Task C was the least popular among the teachers, due to its topic (politics) and the single listening opportunity.

### **ii) Teaching implications**

As had been done in the reading sessions, participants were asked to draw conclusions for teaching practice based on what they had seen of the new exams. Ideas were stimulated through the mechanism of a question box, where participants had to answer questions from imagined colleagues. Discussions were often lively, particularly concerning the pre-teaching issue. Questions and suggestions included:

1. *My students stop listening when they hear someone speaking at normal speed, or as soon as they do not understand something. What can I do?*

Suggested answer: They just need to get used to the sounds and speed through practice. The more they listen, the easier it will get. Tell them they are not expected to understand every single word. Warn them beforehand not to worry about it. Tell them to go on answering questions that they can understand. If they know they will have a chance for a second or third listening, it may help to ease the stress. Sometimes we can wind the tape back and stop at a particular word and listen to it a few times until someone catches it.

2. *How should I check listening activity answers? Is it a good idea to break up a listening text into sentences to help the students?*

Compare answers in pairs. Read out individually. Group work. Try not to listen to the text sentence by sentence or word by word, we need to persuade students that even without understanding every word/ sentence they will be able to deal with the questions. The students need to develop the skill of listening to extended speech.

3. *How can I set a listening homework?*

Cable TV, films, music, radio, cassettes, find and interview local speakers of English.

4. *Is it a good idea to use authentic texts with Basic level students?*

Yes, it is. They can listen to numbers, the time, weather forecast, simple conversations. Let the task fit the student, even if the text does not seem to.

5. *Shall I teach the new words before or after the listening activity?*

It is better to teach words after the task. In exams and real-life students do not have someone to explain the words they are going to meet. Pre-teaching all the time may prevent students from developing essential language skills, and should be used sparingly if at all.

6. *Do you think that in a lesson there should be exclusively Listening tasks?*

It depends on the listening text. We can begin with a pre-listening task, then the while-listening and finally the post-listening activity, then all skills can be integrated.

7. *Do you think it is a good idea to make the students repeat each sentence after the tape?*

It depends on the level of students, with beginners it may be a good idea occasionally, but not for most students.

8. *Do you think it is a good idea to teach listening with recorded real-life conversations?*

Yes, authentic texts seem to be more motivating and challenging for students. Those texts can be graded as well. Students can listen to utterances where accuracy does not always matter.

9. *Is the tape the only way students can improve their listening?*

Certainly not. It is essential for students to see listening as part of real communication in the classroom, for instructions, requests and so on. Students should be able to listen to the teacher, to each other, to a video/ TV/ radio/ films, to native speakers or foreigners.

10. *Course book listening tasks often have a pre-listening activity that leads into the task.*

*The students can't do this in exams, can they?*

Yes, they can. In exams they can use the clues in the title, instructions and questions to anticipate the content of the task. This is an important skill for students to develop.

### ***iii) Trying out sample tasks in the real classroom through the homework assignments***

In the listening assignments participants tried out at least one pilot task with their own students. These were handed in after Listening session 5 on day 3. In that session participants had an opportunity to share their thoughts in groups of four or five then in the whole group.

Participants were given the following task:

In your group discuss the Listening tasks that your students did.  
How well did the students do the tasks?  
How do you feel about the tasks now?  
Summarise your discussion on an OHT and report to the whole group.

All participants commented on the poor quality of the recordings, which definitely should be improved for future courses and pilots. When discussing students' performance of Sample Task A and B and comparing that with previous pilot data, teachers agreed that recording quality caused serious problems and said they were convinced that the results would have been very different (better) with an audible tape.

It was also suggested that some of the poor student performances (poorer than the reading, for example) could be explained by the absence of listening in the current exams or the lack of practice in listening in class.

### **c) Use of English**

Use of English was delivered in 3 sessions only, both in Eger and in Debrecen, because it will only be tested in the Advanced level exam, and because course designers felt that

Use of English would be the most familiar component of the new exam since the current exam contains a grammar/ usage test.

Nevertheless, the Use of English pilot exam tasks proved to be difficult or unfamiliar to most of the participants' students, at least partly because Use of English is tested only on Advanced level and the majority of secondary school students are likely to be on either Basic or Intermediate levels.

Participants were asked to do the following tasks for familiarisation and evaluation purposes in Use of English session 2 (Day 4):

Task 1	Complete the following tasks. Identify the task and text types from the Working Document pp 20-21. Identify which language area they are intended to test. Give your opinion of the tasks. Compare your answers with a partner. Discuss your answers with whole group.
--------	---

Task 2	
A)	In a group make a list of tips for teachers and students preparing for the new exam.
B)	Report your suggestions to the rest of the group.

Participants were shown three pilot exam tasks:

- Sample task A *What on earth* - gap filling (prepositions)  
Sample task B *After flying royal visit* - gap filling (words)  
Sample task C *Underlined phrases* - spot the error (correcting)  
(See Chapter 14, this volume)

Participants found the tasks difficult, and recognised that the need to teach discourse and text-level grammar would require considerable changes to current practice, which is based on sentence-level grammar.

According to participants' reports in their assignments, their students complained about very difficult vocabulary and they also seemed to be lacking strategies for solving the tasks. This fact suggested that there must be problems with the current practice of teaching grammar in spite of the fact that most of the participants declared they liked teaching grammar when they were asked about it at the beginning of Use of English session 1.

## **d) Writing**

### ***i) Discovering and evaluating tasks and text types in the Working Document***

Pilot exam tasks as well as students' pieces of writing composed in answer to those tasks were used in the Writing sessions in both pilot courses. In Eger only one session out of six was planned about the pilot exam tasks and grading, which did not really give enough time.

Having evaluated the Eger course, course designers realised that there was a strong need for revision so that participants would have enough (or at least more) time to familiarise themselves more thoroughly with the rating scale (Írásbeli feladatok értékelési szempontrendszer – see Chapter 10) too.

As a result of this revision, two more sessions were allotted to Writing. So in Debrecen four out of eight sessions were used, fully or in part, for getting acquainted with the pilot exam tasks as well as the analytic rating scale. The extra amount of time made it possible

for the participants to try applying the rating scale when marking students' pieces of writing. Moreover, they had the chance to compare the marks they gave when doing impression marking with those they gave when using the analytic rating scales. As a result of these activities, participants could see for themselves how unreliable impression marking might be.

In Debrecen two Writing sessions were delivered on Day 2, and the pilot tasks were introduced in the second session.

The following sample tasks were used:

Sample Task A	<i>Advertisement</i>	– a letter of inquiry and request
Sample Task B	<i>Letter to a friend</i>	– personal letter
Sample Task C	<i>Letter to host family</i>	– formal letter/ descriptions/ giving info

(See Chapter 10, this volume)

First, participants used the tasks to identify text and task types in the Working Document. Then they were asked to think about how well their students would be able to do these tasks.

Thirdly a set of 10 scripts written by students in the pilot exams was handed out. Participants read them and graded them as if they were marking for the current érettségi, on a scale of 1 to 5. Participants put their grades in a table in their course booklet for later use.

On Day 3 the impression grading from Day 2 served as the basis for an intra- and inter-rater reliability task. Participants were asked to grade the same 10 scripts as they had graded on Day 2. They compared their own grades on the two different days (in between Day 2 and 3 a month had passed), then compared their grades with a colleague's.

The results were very interesting and successfully demonstrated how unreliable impression marking is. Participants found they had given significantly different grades to the same students. There was a variety of responses. Most participants were surprised, even astonished, at the differences between Day 2 and Day 3, but recognised the significance of the results, realising that impression grading is totally inadequate for fair and consistent grading. Some even refused to believe that the scripts were exactly the same, and others claimed that the results were not the same because the grading had been done in a hurry on Day 2.

Participants were less surprised to see that there were even more differences between grades given by different raters for the same script. This confirmed what had already been stated in discussions, namely that érettségi grades are arbitrary and standards and expectations can vary enormously between schools and even between different teachers in the same school.

In discussions participants felt that more than one rater is needed, preferably 3 and that there is a need for clear instructions, guidelines and training for all raters.

This led neatly into the rating scales session (Writing session 4, Day 3), in which participants used the Basic exam rating scales to grade a number of scripts in two subsequent activities.

Participants' responses to the principle and practice of the rating scale were very positive. They were pleased to find that the scales helped them reach closer agreement on the grades. Most were also pleased with the results the scales produced. They were happy to realise that the grades tend to improve when more criteria than just accuracy were taken into account.

A number of ways were suggested in which the scales could be improved, for example the wording of the criteria and the descriptors should be more precise, examples of different student performances could be provided, and rater training would also help the relative terms to be more understandable.

### **ii) Teaching implications**

Activities looking at the teaching implications of the new Writing exam centred on correction and grading strategies. Most of these activities have already been discussed above as they were related to the grading procedures of the new exams. A number of participants felt that making students aware of grading criteria would help students improve their performance.

### **iii) Trying out sample tasks in the real classroom through the homework assignments**

It is important to note that only after familiarising them with and discussing the use of rating scales in the Writing sessions on Day 3 were participants asked to carry out their assignment in distance mode by Day 5.

The assignment was the following:

- a. Grade your students' scripts according to the rating scales in your materials package, Day 3 page 6. Write the grades in the table below.
- b. Write a brief description of your thoughts and feelings on how you used the scales with these scripts. Write about 100 words.
- c. Discuss the task and the grading with your students and write down their views below.

Judging from the style and the length of their reports in their assignments, it is obvious that participants were really interested in trying out the pilot exam tasks as well as the rating scales with their own students.

All liked the idea of using the rating scales for a number of different reasons. Here are some opinions:

*'The scales proved to be great help when grading the compositions. I felt very comfortable to have them near at hand all the time.'*

*'The scale measures writing more objectively because it is based on testing a lot of skills. It is not so grammar-centred.'*

Some were so impressed by the scale that they decided to use it in their own grading practice.

Some confessed that they had had doubts about the usefulness of the scales, some had been afraid that it would be more time-consuming than their existing impression grading routine.

A teacher admitted, *'I was a bit scared at the beginning as I had never seen such a scale before the course, but after reading it thoroughly, I found it very interesting.'*

Many appeared to have realised the drawbacks of their own teaching practice.

One of the participants wrote, *'I liked using the rating scales. I think that there are a number of advantages of analytic scoring. First I have to consider aspects which I might otherwise have ignored. Secondly, the fact that I have to give a number of scores will tend to make the scoring more reliable. Finally I can judge each of the aspects independently of the others.'*

As for the pilot exam tasks, both teachers and students liked the task types. Some thought they were too easy for an exam and some students would have liked to use dictionaries. However, not only the teachers but the students as well complained about the limit of 100 words, considering it too short for a letter. A teacher remarked, *'I found that a short piece of writing is barely enough to judge the richness of vocabulary or their use of English.'*

## **e) Speaking exam**

### ***i) Discovering and evaluating tasks and text types in the Working Document***

One of the most important aims of the course was familiarising the participants with the possible future task types and text types of the Speaking Paper. Another important aim was to test teachers' and students' responses and opinions concerning the piloted tasks.

Compared with the current érettségi, the most radical change in the planned exam is the fact that candidates might be exposed to a paired examination. Another radical change is the way of evaluating the candidates' performance: the use of external as well as internal examiners (interlocutor and assessor). Participants were shown examples of these completely new elements in the exam.

Three tasks were chosen from the piloted ones, and the video recordings of the pilots were used during the Speaking sessions of the course:

*Sample Task A: Michael's Restaurant* required the candidate(s) to take part in a role-play, with either the interlocutor or another candidate. The prompt used in this case was an advertisement.

*Sample Task B: Taking out a penfriend from Britain* was a paired problem-solving task including negotiation / discussion.

*Sample Task C: The International Student Centre* was another paired problem-solving task in which the candidates had to bridge an information gap. The visual prompts in this case were maps. (See Chapter 11, this volume.)

The sample exam tasks were used for different purposes at various stages during the sessions.

### ***a) Identifying the text type, the task type and the purpose of the task with the help of the Working Document, introducing the idea of interlocutor – external assessor:***

#### ***Sample Exam Task A: Michael's Restaurant***

This task was used in the course as Task 1 in Session 24, Speaking session 3, Day 4. Using the Working Document participants became familiar with the possible future exam task and text types. They found the task life-like, suitable for an examination. The role of the interlocutor was familiar to them, since the current érettségi offers the same type of role to the teacher. However, they found it interesting that an external assessor was assessing the student's performance. Advantages (objectivity, less burden on teacher) and disadvantages (stress for students) were listed and discussed.

### ***b) Finding examples of the different types of speaking, introducing the idea of paired examination:***

#### ***Sample Exam Task B: Taking out a penfriend from Britain***

This task was used in the course as Task 2 in Session 24, Speaking session 3. After watching the paired sample exam task, participants identified types of speaking used by the two candidates in the sample. The task showed a good example of paired interaction. Participants were very satisfied with the candidates' performances, and discussed the advantages and disadvantages of pairing up students. They agreed that well-prepared pairing can help students to perform better.

### ***c) Assessing the students' oral performances by impression grading and by using an analytic rating scale:***

#### ***Sample Exam Task C: The International Student Centre***

This sample task was used in the course in Speaking sessions 4 and 5, Day 4 to provide evidence for the shortcoming of the current 1-5 scale grading system, and to introduce the idea of using rating scales.

Participants discussed the criteria according to which they grade their students currently. Several different things were mentioned here. One participant commented that students do not listen to feedback and they are only interested in grades. This suggests that comments should be linked to grades and used to help students improve their performance.

This discussion led into grading the sample performances using the current 1-5 impression scale. Great differences were apparent in the grades given for the same performances. Several participants suggested one performance should receive a 4 in the érettségi, but only a 3 in class. Another commented that teachers can raise a student's overall English grade by giving higher marks in the oral compared with their real ability. Also, participants seemed to be giving lower grades than deserved based not on the language produced, but on the uncertainties shown and expressed when doing the task.

Participants gave a highly positive response to the use of speaking rating scales, just as they did with the writing ones. Considering that they were using them for the first time their performance was very gratifying. They reported a high level of agreement, commenting that *'it was a great help, although it seemed difficult at first because of so many criteria.'* The issue was raised whether students should use the scales in class. There was support for this idea, as students need to know what is important and how they are graded. Most participants applied the scales and this idea in particular in their speaking assignments.

### ***ii) Teaching implications***

As with the writing sessions, teaching implications were closely bound up with the grading procedures for the new exams.

### ***iii) Trying out sample tasks in the real classroom through the homework assignments (between Days 4 and 5)***

The participants' classroom-based speaking assignment in the second distance module had a crucial significance in three ways:

- a) applying the information about the new examination and the different new ideas
- b) provided by the face-to-face speaking module in the teachers' everyday classroom reality
- c) getting further data about how these piloted tasks work and how the teachers and students react to them
- d) introducing and promoting the idea of peer observation.

We received 20 speaking assignments, and on the basis of the submitted Speaking Task Report Forms, the classroom observation reports and the participants' oral feedback in the concluding speaking sessions the following conclusions can be drawn:

#### **Sample Task A: Michael's Restaurant**

2 participants chose this task. Their reason for choosing this task was:

*'I could adjust it to the exam topic 'Eating out', and I found the level of difficulty appropriate for an intermediate group.'*

*'We had been dealing with this topic for two weeks.'*

The level of the classes the task was administered to was Intermediate (Year 12) in both cases.

The main aims of the lessons the task was incorporated into were:

- vocabulary revision on the topic *'Meals, eating out'*
- preparing student for the oral exam
- developing speaking skills
- practising vocabulary.

The purpose of using this Speaking task in the lesson was:

- providing a real-life situation,
- helping students to get used to these kinds of situation.

The teachers used role play cards, vocabulary lists and handouts in the lessons.

In both cases there was an observer present in the lesson, who also acted as an external assessor.

The instructions were given in English and the students were able to follow them.

The teachers used the following formats:

1 student + 1 interlocutor + 1 assessor

2 students + 1 interlocutor + 1 assessor

One teacher tried both formats, and found that the students preferred the paired one.

The teachers anticipated the following problems:

*'I expected them to miss some aspects of the advertisement, and some of them really did so.'*

*'I thought the students would be less co-operative and active because of the visitor.'*

The latter did not cause serious problems in the lesson.

The pairing of the students proved to be successful. One explanation for this was:

*'They had been working in the same pairing for a long time.'*

Teachers think this task can be used for Intermediate or higher levels.

According to the teachers' reports, the students used the following types of speaking most frequently:

1. asking for information
2. asking for clarification
3. expressing opinion.

Teachers and observers reported the following difficulties:

a) in one class the students had difficulty with understanding the expression *'dining experience'* and the teacher had to paraphrase it

b) in one class the interlocutor spoke too much.

The rating scale was used in both cases and the students were also involved in the assessing process:

a) the teacher made the students familiar with the criteria

b) volunteers did the task in front of the class, and the other students participated in the evaluation of their performances through a whole-group discussion afterwards.

The students were willing to act out the task and most of them really enjoyed it.

The teacher's opinion about the task was the following:

*'A really creative task but it can only be carried out with good students.'*

*'I liked it, I always appreciate my students' enthusiasm.'*

### **Sample Task B: Taking out a penfriend from Britain**

This task proved to be the most popular among the participants of the course. 12 people chose it. Reasons for choosing this task were:

*'I found it the most interesting.'*

*'It is lifelike, relevant to Ss' level, Ss are familiar with the task type.'*

*'It encourages discussion, problem-solving.'*

*'I supposed these roles would be familiar to my students.'*

*'Because of the topic.'*

*'It was a good way of revising the studied material.'*

*'My students like discussion very much.'*

*'It fits in the material of the textbook.'*

*'The topic itself is well manageable at a pre-intermediate level, and it fits well into the curriculum.'*

*'It seemed to be perfect for revising suggestions.'*

The task was administered to different levels and different age-groups:

- a) elementary (Year 9)
- b) pre-intermediate and intermediate (Year 10)
- c) intermediate (Year 11)
- d) upper-intermediate (Year 12)
- e) intermediate (Year 12).

The main aims of the lessons the task was incorporated into were:

- a) practising speaking
- b) using functions for making plans to organise free-time activities
- c) trying out a new oral exam task type
- d) revising the communicative functions studied before
- e) eliciting vocabulary of free-time activities.

The purpose of using this Speaking task in this lesson was developing speaking skills, namely developing arguing, negotiating, persuading sub-skills, practising making suggestions.

The following materials were used by the teachers:

- a) textbooks
- b) pictures
- c) cassettes
- d) local 'What's On'.

In most cases there was an observer present in the lesson, who also acted as an external assessor.

The students were given instructions in English. The teachers found it natural, necessary and useful:

*'At their level they are supposed to understand instructions in L2. They are accustomed to being given them in English.'*

*'In the 12<sup>th</sup> year students can understand instructions and explanations in English.'*

*'I always give instructions in English.'*

*'My students were elementary ones. I wanted to see whether they could manage.'*

The format varied as follows:

2 students + 1 interlocutor + 1 assessor

Reason: *'More students could have the possibility to talk. No influence on the part of the teacher.'*

*'To avoid subjectivity as far as possible, to simulate real exam situation.'*

*'My students are used to acting out role plays.'*

2 students + 1 interlocutor

Reason: *'Two students can practise at the same time.'*

2 students + the teacher, who acted as interlocutor and assessor

Reason: *'It was the only suitable format according to the type of the task.'*

2 students + 2 assessors (the teacher and the observer)

Reason: *'I did not want to take part in the conversation at all, I wanted to try out the new rating scale. During the role play the rest of the students also marked the performances according to the scale.'*

Teachers anticipated the following problems:

*'My students would not be creative enough.'*

*'Lack of tolerance. Being too dominant / too shy.'*

*'Not a wide scale of appropriate and stylish expressions at this level'*

*'The role players will be too embarrassed to perform well.'*

*'My students won't be very talkative, they won't use a wide range of structures and vocabulary.'*

*'Reluctance to speak.'*

*'Shortage of time, difficulties in understanding the instructions.'*

*'Too much use of mother tongue.'* (elementary students)

The pairing of the students was successful according to the reports, since the students could co-operate quite well. They attributed it to the following facts:

- a) The students were on the same level.
- b) They could choose their partner.

The pairings were decided on the spot in most cases, and were not pre-determined. One teacher suggested that *'A boy and a girl can be an ideal pairing.'*

According to the reports, this task can be used for all levels from elementary to advanced.

The following types of speaking were used by the students while doing the task:

- a) asking questions
- b) suggesting
- c) expressing opinion
- d) asking for information
- e) arguing
- f) agreeing, disagreeing
- g) summarising
- h) non-verbal reactions
- i) paraphrasing
- j) negotiating
- k) taking turns.

The students understood the prompt and used it as a guideline.

The following difficulties were reported while doing the task :

- a) The English instructions caused difficulty in some cases (elementary students).
- b) In some cases the students did not have enough vocabulary.
- c) In some cases the students first seemed a bit shy and embarrassed because of the observer.
- d) Sometimes the students used Hungarian.
- e) In some cases the students finished the conversation very quickly and not enough language could be elicited from them.

In most cases the external assessor and the teacher gave the students very similar grades. One teacher reported:

*'We (the teacher who acted as interlocutor and the observer who acted as assessor) graded the students using the rating scale, and we gave them the same grades.'*

Some people mentioned how difficult it was to use the rating scale for the first time in their lives:

*'I did not want to take part in the conversation at all, I wanted to try out the new rating scale. During the role play the rest of the students also marked the performances according to the scale. It was not so easy to use the scale, I had to concentrate on too many things (the students, the scale, the task itself) at the same time.'*

Students generally enjoyed the task, it was quite easy for most of them. They thought it was relevant to everyday life.

*'They said it was life-like and easy to do.'*

The teachers' opinion about the task:

*'I found it creative.'*

*'It was interesting for the students. It was a good task for their level as well.'*

*'I liked it, because I know my students like such problem-solving tasks.'*

*'I found it a bit artificial. I think one of the candidates should be the penfriend.'*

Some teachers made comments in connection with the task:

*'The layout of the instructions could be more spacious.'*

*'Different lead-in can be used with the same task depending on how I can embed it into the material of the textbook.'*

*'Perhaps students might listen to a short dialogue about the same topic before doing the task.'*

*'The instructions should contain some hints about the fact that the longer they go arguing the better.'*

### **Sample Task C: International Student Centre**

6 participants chose this task. Their reason for choosing this task were:

*'The students are familiar with this topic.'*

*'I had extra material in connection with it.'*

*'It was the closest to my students' level.'*

*'We are at the topic.'*

*'The students are familiar with this task type.'*

*'It's a life-like, everyday situation, still not typical and over-used in this form. It seemed to be quite interesting.'*

The level of the classes the task was administered to:

- a) intermediate (Year 12)
- b) intermediate (Year 10)
- c) pre-intermediate (Year 9, Year 11).

Main aims of the lesson the task was incorporated into:

- d) developing speaking skills
- e) teaching / revising of asking for and giving directions
- f) teaching / revising of telephoning
- g) revising vocabulary.

Purpose of using this Speaking task in this lesson:

- a) practising this topic
- b) doing free practice after the controlled and semi-controlled phase
- c) practising interactional and transactional skills.

The teachers used maps, cards, cassettes and handouts in their lessons.

Most observers in the lessons acted as external assessors as well.

Just as in the case of the other two tasks, the instructions were given in English.

*'It helps them to do the task.'*

*'Ss are used to getting instructions in English, and I did not anticipate any problems.'*  
*'I wanted to simulate a real exam situation.'*

The format of administration varied in the following ways:

2 students + 1 interlocutor + 1 assessor

*'It was the most suitable for the task.'*

*'It seemed to be more objective.'*

*'I thought it would be more exam-like in this way.'*

2 students + 1 interlocutor

*'I could concentrate only on the students' performance.'*

The following problems were anticipated by teachers:

*'vocabulary problems'*

*'I thought that Ss would be unmotivated and carry out the task as fast as possible just to get it over with.'*

*'They would have problems finding the destination because of the bad quality of the maps.'*

*'One of the students would speak much less.'*

*'The maps were not carefully designed.'*

*'Finding the ISC.'*

The pairing of the students proved to be successful with this task, too:

*'It was successful. They could do the dialogue.'*

*'It was successful. They were at the same level, they helped each other.'*

*'Yes, they both tried to speak as much as possible.'*

*'Both of them were brave enough to take risks.'*

*'The pairs had already got used to each other, so they understood each other well and could communicate well.'*

Some teachers thought the task was good for intermediate level only, but others were convinced that it could be used at any level, even elementary.

The students used the following types of speaking during the dialogues:

- a) asking for clarification
- b) taking turns
- c) negotiating
- d) summarizing
- e) asking for information
- f) taking initiative
- g) waiting time
- h) paraphrasing for clarification.

Unlike the other two tasks, in this case the prompts did not work well:

*'One of the maps was very difficult to use.'*

*'I had to show them the place of departure.'*

*'The maps and the points of orientation were not clear enough.'*

This experience clearly shows how important the quality of the prompt is.

As a consequence, the following difficulties arose:

*'They did not understand the maps first.'*

*'The map was not visible enough.'*

Other problems were caused by the students' limited vocabulary:

*'Their vocabulary was limited.'*

*'At times they had problems with paraphrasing things.'*

The problem of eliciting enough language arose when assessing the students' performance:

*'It was impossible to assess one of the students since he had said only two sentences during the entire conversation. ('Can you tell me the way...? ... Thank you.')* Can / Should I underevaluate him because of this non-communicativeness?

Hopefully, examiner training will address this problem.

Students' opinion about the task:

*'It was challenging for them, it was something new, so they were interested.'*

*'They liked it except the quality of the maps.'*

*'They enjoyed it.'*

Teachers' opinion about the task:

*'The preparation for the task took a lot of time.'*

*'I would either supplement the written instructions, or give them extra oral instructions:*

*'Try to be as communicative as possible. Try to get **involved** in the conversation.'* etc. *If we regard this task a life-like situation we should accept that the lost person speaks very little I think.'*

*'The task itself was OK, but the layout was a serious obstacle in the way of successful student performance.'*

*'This task type is well-known and practised by the students.'*

*'A meaningful task, quite agreeable for the exam.'*

## **Outcomes of this INSET course**

The course aimed to raise awareness of task and text types, issues of language level, and the value of careful reflection based on real experiences and concrete data. In addition to the evidence given above from the various sessions and assignments, below is a selection of comments from the course feedback that show how participants appreciated the way the course dealt with these issues.

*'It is getting clearer what the requirements for the new érettségi are, and I think we all will be richer in knowledge and experience. Thank you!'*

*'I became more familiar with the new exams.'*

*'The sessions give me a lot of information about the future érettségi exams.'*

*'So far I have given little thought to the new érettségi. I have now become more conscious of what is going to happen in the future.'*

*'The course has given me a lot of new ideas. I have learnt a lot more about the new érettségi.'*

*'It was very useful. I got to know a lot of very important facts about the plans of the new érettségi.'*

*'So far the new érettségi has been a dark horse for me, but now I can see more clearly about the changes planned.'*

*'I got a lot of information about the new exam much before its introduction, so I think when the time comes I will be prepared for it.'*

*'Now I know quite a lot about the new exam. The materials were very interesting, especially the sample tasks and the rating scales.'*

*'Before the course I did not know anything about the new exam. I have learnt a lot about my new role, the new requirements, and also about what I should be careful about.'*

*'I have received very useful information on the new coming érettségi, it also made me reconsider the task-types that I use in the classroom.'*

*'I found the course very useful, some of my colleagues have already asked me several questions about the new érettségi, and I could answer them!'*

*'With the help of these discussions I can have a better insight into my own work at school.'*

The course aimed to build confidence in teachers that their students would be able to cope with the new exams, that the exams would involve carefully-produced tasks and employ fair, objective and transparent grading procedures.

*'This exam will have higher value, more prestige and appreciation than the former one.'*

*'When I first read the documents I found it difficult to imagine the new examination system, but now I am sure that it will work well.'*

*'Now we have some ideas about the requirements of the new exams. I do hope that the teachers' comments will be used while preparations for the new exams will be made.'*

*'It is really good to see a lot of things in practice, since after reading the Working Document I was really frightened. Now I am sure I will enjoy trying the tasks out with my students.'*

*'I think the scales will be very useful and will help us to be more objective.'*

*'The evaluation sheets (=rating scales) will be useful for marking both written and oral performances.'*

*'The grading system will, I think, help us to give more objective marks.'*

*'We got a lot of help how to mark/grade writing and speaking.'*

*'The sessions were enjoyable and useful, especially the rating scales.'*

The course also aimed to show teachers that teaching strategies would need to change, but that teachers' existing knowledge of good teaching practice, and many of the course books currently in use, would enable teachers to help students prepare effectively for the new exams.

Participants' feedback comments include:

*'Day 3 made me realize that I am a bit too strict or 'old-fashioned' about grammar mistakes when correcting writing exercises.'*

*'I learned that I should pay more attention to teaching writing (although correcting is the most time-consuming part of it).'*

*'I liked the video, because we could see students from other schools. I think these were very lively and real presentations.'*

*'I found it interesting to see the examination task types in connection with speaking.'*

*'I liked the video recordings. It was interesting to see this new system working.'*

*'Thank you for making me less worried about pairing the students.'*

*'My views about language teaching have been confirmed, i.e. language has to be taught in its complexity.'*

Finally, the participants expressed a high degree of approval of the approach and organisation of the course, and the trainers:

*'It was a pleasure to work with the trainers, they were very friendly, professional, business-like- very much aware of the situation and state of the old érettségi system in Hungary.'*

*'To get through a reform needs interaction and sharing which was provided.'*

*'It was good for me to see (and enjoy) how well-prepared the trainers were.'*

*'I heard about each paper in detail, trainers were very helpful and worked very hard, each participant contributed.'*

*'I liked the atmosphere, the variety of tasks, the training skills of the trainers.'*

*'I liked group work and not just lectures, discussing the up-to-date/ current problems and changing ideas with colleagues.'*

*'The course is well-organised.'*

*'The course has been very useful.'*

*'The sessions were very useful, well-organized.'*

*'Everything we had to do on these two days was very interesting and useful.'*  
*'The course is astonishingly well-prepared and built up. I have attended other courses before, and have never ever experienced such organizational skills!'*  
*'I am happy that my opinions are considered.'*  
*'The sessions were enjoyable and useful, especially the rating scales.'*  
*'Both days were well-organized.'*

**Conclusion**

The pilot courses seem to have achieved their aims, and data from the pilot examinations played a crucial role in their success. In addition, we believe that the feedback on pilot exam tasks and rating scales from course participants have proved very useful for item writers and test developers. The courses demonstrated how the chances of successful implementation of the examination reform can be greatly enhanced through in-service training.

Feedback and material produced in assignments and sessions suggest that the courses successfully informed teachers about the new exams, elicited teachers' support, raised awareness and built confidence. The overall strategy of the course, then, including the use of classroom teachers for design, development, and delivery, seems to have been successful.