

A Study on Oral English Examinations for Japanese EFL Learners

Koji Konishi

1. INTRODUCTION

Konishi (1995) comments that it is high time to adduce the true reasons why Communicative Language Teaching (CLT) has not rooted in the EFL classroom in Japan and why Grammar Translation Method has persisted in the Japanese EFL culture to this day.

Giving grammar-based examinations to EFL learners can be regarded as one of the principal reasons for this, because grammar-based examinations themselves have an undesirable and cramping effect on the language learning and teaching¹⁾. Therefore, to begin with, we should think over why grammar-based examinations still survive and why communicative language ones do not prevail in English Language Teaching (ELT) in Japan even today.

We surmise that grammar-based tests, which can almost count on objective marking²⁾, are easier to mark, while, communicative language tests, which usually rely on subjective marking³⁾, are rather difficult to mark. Let us consider this issue from the perspective of inter-/intra-rater reliability⁴⁾ or inter-/intra-marker reliability⁵⁾. As Bachman and Palmer (1996: 23) suggest that "a multiple-choice test of grammatical knowledge might yield very consistent or reliable scores", grammar-based tests are more reliable than communicative language tests.

Next, we need to look at this question from the practical standpoint. As far as assessment is concerned, it should be practical in terms of resources, which include human resources, material resources, and time⁶⁾. Take 'time' for example. According to Harris and McCann (1994:4), "any approach to assessment must not be too time-consuming in terms of class hours and of our own time outside the class". By contrast, as Alderson *et al.* (1995:107) have argued, subjective marking requires us to make "judgements which are more complicated than the 'right-wrong' decisions".

This means that we consume more of our time in scoring communicative language tests. On the other hand, according to Bachman and Palmer's (1996) definition of practicality, if required resources exceed the available resources, the test is not practical⁷⁾. On the basis of what we have said, we can draw a conclusion that grammar-based tests also excel communicative language tests in respect of practicality. So far, we have ascertained the cause of the long persistence of grammar-based examinations and the unwillingness to accept communicative language examinations.

Then, the next question, despite all the reasons mentioned above, concerns why we should introduce communicative language examinations in ELT. To find out this answer, we shall focus on validity⁸⁾ or construct validity⁹⁾. For example, Bachman and Palmer (1996:23) explain:

Suppose that we needed a test for placing individuals into different levels in an academic writing course. A multiple-choice test of grammatical knowledge might yield very consistent or reliable scores, but this would not be sufficient to justify using this test as a placement test for a writing course.

In other words, as Bachman and Palmer (1996:23) point out, "reliability is a necessary condition for construct validity, however, it is not a sufficient

condition for construct validity". According to their comments, grammar-based examinations have no meaning with regard to validity in CLT, because they assess only grammatical competence and do not assess EFL learners' communicative competence¹⁰⁾ as a whole.

That is to say, if we adopt CLT, it will mean cultivating and accessing EFL learners' communicative competence in the EFL classroom.

2. PURPOSE

Under circumstances in which we adopt CLT while continuing to give EFL learners grammar-based examinations, we cannot expect CLT to win the day in the EFL classroom. On the other hand, as mentioned in the Introduction, grammar-based examinations still predominate in ELT because communicative language examinations are inferior in reliability and practicality to grammar-based examinations. Therefore, in order to remedy a catch-22 situation, we should design communicative and authentic examinations that have a decided improvement on the traditional ones in reliability and practicality.

Halleck and Moder (1995) state that we must give EFL learners task-based examinations in order to design more communicative and authentic language tests. They (1995:733) also explain the tasks used for examinations as follows:

These tasks require the L2 speaker to make use of a wide variety of linguistic abilities and communicative strategies. As a result, the emphasis on authentic tasks has raised questions concerning the respective contributions of linguistic and strategic competence to test performance.

In fact, however, it is almost impossible to invent examinations which can

replicate reality fully¹¹⁾. Therefore, as Weir (1993:xi) remarks, “we must weigh up what is practical by exploring the tension between test authenticity and test ‘operationalisability’; *and also* we must determine to what extent compromises may be made in the trade-off between reliability and validity” (Italics added).

This means that such task-based tests mentioned above should be not only suitable to ELT in terms of reliability, practicality and validity but also realisable even in the Japanese EFL culture. To meet these demands, we have tracked down a number of examinations at several local examination boards in the UK, viz. University of Cambridge Local Examinations Syndicate (UCLES) in Cambridge, University of London Examinations and Assessment Council (ULEAC) in London, and Northern Examinations and Assessment Board (NEAB) in Manchester. As a representative sample of such examinations, we shall focus on the new general certificate of education (GCE) ordinary level (“O”-level) oral English examination. The purpose of this paper is, therefore, to study this new GCE “O”-level oral English examination in detail.

3. STUDY

3.1 Problems with an oral assessment scheme

Heaton states (1990) that in testing speaking, whatever system is adopted, the marking itself is very subjective. Underhill (1987:88) echoes that “this poor reliability makes it difficult to be confident that the scores awarded in oral test are accurate and trustworthy”. Hence, first of all, we should think up oral assessment schemes which significantly lessen this subjectivity. In the words of Underhill (1987:89), “the single most effective way of getting round the central problem of lack of reliability is to use more

than one assessor"¹²⁾. In such tests, "a separate person may take on the role of 'interlocutor', eliciting language from the candidate while the examiner is freed to assess" (Alderson *et al.* 1995:115). However, in the light of the existing Japanese ELT situations, it is more likely that the examiner single-handedly gives his/her EFL learners instructions, ask questions, reacts to the candidate's contributions, and tries to assess the performance at the same time (Alderson *et al.* 1995).

In order to solve this dilemma, we need to formulate oral assessment schemes, which improve reliability even if only one assessor can be employed. In this respect, Underhill (1987:92) writes that :

Oral tests can be recorded on video or audio tape so that they can be marked later. This can be done as well as, or instead of, live marking. Taped tests are easier to mark because you can mark them when and where you want, and because you can rewind the tape and listen to what the learner says as often as you want.

He (1987:93) also lists the following three reasons why we should use marking from tapes : "(1) the first marking, where the test has been held in a language laboratory ; (2) second marking, as a check on the first marking of a live test ; and (3) re-marking tapes as an assessor-training procedure". In this connection, Brown and Yule (1983:105) suggest that "the taped performance of a student can be kept as evidence to support the teacher's judgements". While, with regard to recording, Underhill (1987) recommends using videotape recordings instead of audiotape recordings partly because we miss the visual element when judging a candidate's communicative ability from audiotape recordings. Weir (1993:41) adds that "if videos can be made of candidates performing test activities these can subsequently be used for training and standardisation of marking".

On the other hand, as Brown and Yule (1983:106) have pointed, “we are aware that maintaining a taped record and doing some assessment on the basis of this taped record will require more of the teacher’s time than has been typically devoted to spoken English assessment in the past”. Whereas, as Harris and McCann (1994:4) point out, “assessment is only one aspect of our jobs and cannot be allowed to detract from teaching or preparation time”. Taking account of the status quo in ELT in Japan, assessment mentioned above will be too high to be realised supposing we impose another big burden on the Japanese English teachers (JETs). Therefore, in order to put an oral assessment programme into practice, in the immediate future, we need to establish several examinations and assessment boards (EABs) such as NEAB, UCLES and ULEAC.

3.2 The role of examinations and assessment board

As discussed in the previous section, in testing oral communication skills, the marking itself is very subjective. Therefore, particularly in formal examinations, the principal role of EAB is to standardise and moderate examiners’ marking nationwide. The second role of EAB is to train examiners and administer examinations.

Alderson *et al.* (1995:114) comment, “arranging for either type of recordings, i. e. audio recordings or video tape recordings, to be made and edited is a complicated and time-consuming affair”. Therefore, in order for oral examinations to go smoothly in Japan, all our arguments presuppose that EABs administer the overall oral examinations in order to help relieve JETs of such troublesome affairs.

3.3 The procedure for standardisation

In the interests of standardisation, each school selects its own chief examiner (CE) within a school. EAB summons CEs to attend coordinating meetings several times a year. At the coordinating meetings, each EAB shows CEs the procedure for standardisation. It is CE's job to standardise the marking scheme within his/her school. CE is also responsible for training everyone else in his/her school. EAB sends each school under its control a standard videotape, teacher's training manuals, guideline papers and so forth several months in advance of the scheduled day when an oral examination is conducted.

In each videotape, there are sample extracts from several, usually around six, assessments made by EAB. First, EAB lets CEs look at the videotapes and access these extracts according to a rating scale without looking at the keys in the training manuals (the issue of rating scales is discussed in the next section). Later CEs compare their marks with the keys and study their resultant marks (see appendix 1). Next, each CE asks his/her colleagues in his/her school to mark the same extracts according to the rating scale in a similar fashion. If there is a big gap between the marks of CE and those of his/her colleagues, they discuss their marks with each other in order to make up the gap. If there is a vast gap between the marks of a particular school and the keys — i. e. assessments made by EAB. EAB sends one of its representatives to the school and has a lively discussion on the problems so as to bridge the gap.

According to Alderson *et al.* (1995), in order to ensure reliable administration and marking, a complete day should be set aside for a standardisation meeting and in addition to this meeting, half a day should also be set aside for a separate training session, which allows the examiners to 'walk

through' the testing situation. "Although this is likely to be expensive, it is the safest way of ensuring that enough discussion will take place for all examiners to understand thoroughly the rating scale and the procedures for marking" (Alderson *et al.* 1995:111).

3.4 The need for rating scales

Upshur and Turner (1995:5) show us that "within the CLT, teachers and other professionals using tests in instructional situations need feasible rating scales to assess student use of language".

Heaton (1990:68) also argues that we are "strongly advised to use a scale for grading students' performances on speaking tests rather than a marking scheme". He (1990:69) adds that "it is much better to use a rating scale containing short descriptions of each grade in the scale". Alderson *et al.* (1995) call these short descriptions 'descriptors'. The main reason why we adopt a rating scale is that "when we grade a student's speaking ability, we simply read through the scale and choose the most appropriate description for the particular student" (Heaton 1990:69). What he means is that using a rating scale is eminently practical. Besides, Alderson *et al.* (1995:111) elaborate on this point from the reliability viewpoint as follows:

We would recommend never using a scale which contains numbers only or one where the descriptors are simply one-word statements like 'Excellent', 'Very good', etc., as these statements can be interpreted in different ways by different examiners. We would recommend scales with no more than about seven points, as it is difficult to make much finer distinctions, and we would also recommend that explicit descriptors accompany most of the points on the scale.

To borrow Weir's (1993:40) words, "in the measurement of spoken language

there is a need to establish clear criteria for assessment and to standardise examiners in their use of these criteria”.

3.5 The new GCE “O”-level oral English examination

As mentioned in the Purpose, the purpose of this paper is to study the new GCE “O”-level oral English examination (see appendix 2). Therefore, let us go on to detail this examination.

3.5.1 Candidates

This examination is designed for 16-year-old students who, although they may be using English in school, are living in a community where one or more other languages are widely used (e. g. in Singapore).

3.5.2 The components

See Table 1.

Table 1. The new GCE “O”-level oral English examination items

Part 1	Reading aloud	(2-3 minutes)
Part 2	Picture discussion : — describing, interpreting and responding to a picture given by the examiner	(3 minutes — maximum)
Part 3	Conversation : — responding to some issues raised in the passage and the picture in conversation with the examiner and perhaps developing other related issues	(4 minutes — maximum)

Source: UCLES, *The new general certificate of education ordinary level oral English examination : Teacher's training manual* (Cambridge: UCLES ; 1995), p. 3.

3.5.3 Objectives

See Table 2.

Table 2. Objectives of the new GCE "O"-level oral English examination

<p>Part 1 Reading aloud</p> <ol style="list-style-type: none"> (1) Pronunciation and articulation to read a passage with good pronunciation and clear articulation of the words (2) Rhythm and fluency to read with appropriate rhythm and stress to achieve a well-paced, fluent rendering of a passage (3) Expressiveness to read with appropriate variation of pitch and tone in order to convey the information, ideas and feelings in a passage
<p>Part 2 Picture discussion</p> <ol style="list-style-type: none"> (1) Description to describe scenes of general human interest through the use of language (2) Interpretation and explanation to interpret and explain the situations in the picture (3) Language to use a range of appropriate vocabulary and structures to complete the task (4) Coherence to develop ideas in a clear and coherent manner
<p>Part 3 Conversation</p> <ol style="list-style-type: none"> (1) Personal response to give a personal response to the theme of the picture and the passage (2) Clarity of expression to express oneself clearly and succinctly in a conversation, using appropriate vocabulary and structures (3) Engagement in conversation to discuss issues with the examiner stemming from the picture and the passage

Source: UCLES, *The new general certificate of education ordinary level oral English examination: Teacher's training manual* (Cambridge: UCLES; 1995), pp. 5-9.

3.5.4 Assessment criteria

3.5.4.1 Reading aloud

See Table 3.1.

Table 3.1 Assessment criteria for reading aloud

Band	Mark	Descriptor
1	10-12	<ul style="list-style-type: none"> • Reads the passage with near-perfect pronunciation and very clear articulation • Reads with fluency and good pace, using appropriate rhythm and stress • Varies the pitch and tone in order to convey the information, ideas and feelings in a passage
2	7-9	<ul style="list-style-type: none"> • Generally clear articulation with occasional errors in pronunciation • A mainly fluent reading with some mistakes in stress and rhythm • Some variation of pitch and tone for expressive effect, but not always appropriate to the passage
3	4-6	<ul style="list-style-type: none"> • Pronunciation of words is rather inaccurate, but some attempt to maintain clarity of articulation • Reading is hesitant with frequent mistakes in stress and rhythm • A slight attempt to achieve expressiveness through variation of pitch and tone. Uses largely inappropriate variations
4	0-3	<ul style="list-style-type: none"> • Very weak pronunciation and little clarity of articulation • Very hesitant reading, full of rhythm and stress errors • Reads in a monotone or with inappropriate tones

Source: UCLES, *The new general certificate of education ordinary level oral English examination: Teacher's training manual* (Cambridge: UCLES; 1995), p. 6.

3.5.4.2 Picture discussion

See Table 3.2.

Table 3.2 Assessment criteria for picture discussion

Band	Mark	Descriptor
1	10-12	<ul style="list-style-type: none"> • Makes detailed observations • Combines these intelligently with explanations and interpretations • Uses a wide and suitable range of vocabulary and structures • Develops the description coherently and satisfyingly on any aspect of the picture
2	7-9	<ul style="list-style-type: none"> • Makes observations on the picture • Links these with explanation and interpretation • Uses an adequate range of appropriate vocabulary and structures • Describes fairly coherently, though some ideas may be thrown in as an afterthought
3	4-6	<ul style="list-style-type: none"> • Describes picture superficially • Makes some attempt to link this with explanation and interpretation • Uses a basic, or somewhat inappropriate, vocabulary and simple structures • Describes picture with little coherence
4	0-3	<ul style="list-style-type: none"> • Description is inadequate • Interprets with little awareness of what is happening • Uses very limited, or inappropriate, vocabulary and structures are faulty • Requires constant prompting

Source: UCLES, *The new general certificate of education ordinary level oral English examination: Teacher's training manual* (Cambridge: UCLES; 1995), p. 8.

3.5.4.3 Conversation

See Table 3.3.

Table 3.3 Assessment criteria for conversation

Band	Mark	Descriptor
1	13-16	<ul style="list-style-type: none"> • Gives and develops intelligent personal responses to the theme • Expresses and develops ideas clearly, succinctly and naturally using appropriate vocabulary and structures • Shares ideas and opinions with the examiner • Introduces new ideas or initiates discussion of relevant issues
2	9-12	<ul style="list-style-type: none"> • Offers some personal responses to the theme, with some elaboration • Is generally clear and coherent, using largely appropriate vocabulary and structures. • Responds well to examiner's prompts, but shows less initiative
3	5-8	<ul style="list-style-type: none"> • Gives simple personal responses with little development • Makes disjointed comments which may be unclear, but with some attempts to use appropriate vocabulary and structures • Depends heavily on the examiner for encouragement and prompting
4	0-4	<ul style="list-style-type: none"> • Offers hardly any personal response or development • Offers ideas in disconnected single sentences, phrases or even single words • Finds it difficult to maintain any sustained interaction, even with repeated prompting

Source: UCLES, *The new general certificate of education ordinary level oral English examination: Teacher's training manual* (Cambridge: UCLES; 1995), p. 10.

4. CONCLUSION

Harris and McCann (1994) claim that if tests contain authentic, real-life examples of the type of tasks which our learners will need to perform in the future, they can have a positive influence on teaching and learning, while, if tests contain artificial tasks not linked to real future needs, they can have a negative influence on teaching and learning. They (1994) label such a influence of tests on teaching and learning as the 'washback effect', which is called the 'backwash effect' by Prodromou (1995).

Then, the question is, if we look at grammar-based examinations from a backwash viewpoint, whether they have a negative influence on teaching and learning. Prodromou (1995:19) states that "by not encouraging learners to learn from their mistakes and work out the rules of the language for themselves, conventional testing prevents the full development of the cognitive aspects of learning, thereby contradicting what we have come to consider as good teaching practice". What he means is that grammar-based examinations have a deep-seated tendency to focus on accuracy, with the result that they affect teaching and learning in a negative manner. On the other hand, when we "have tests which adopt techniques more in line with communicative teaching, it is possible for testing procedures to have a positive effect on classroom practice" (Prodromou 1995:15). Needless to say, if grammar-based examinations are given in the right proportions, they have a valuable contribution to make in assessing learners' achievement (Prodromou 1995). However, we cannot help admitting that resorting constantly to grammar-based examinations in ELT in Japan discourages the Japanese EFL learners from taking an active part in communicative activities. In CLT, as Weir (1993:31) points out, "we are no longer

interested in testing whether candidates merely know how to assemble sentences in the abstract”.

Now that we have got to know that grammar-based examinations are quite inadequate to measure a candidate's communicative competence as a whole, we should design examinations which can access his/her general language ability. In this respect, Bachman and Palmer (1996:75-76) point out that “language ability has traditionally been considered by language teachers to consist of four skills; listening, reading, speaking, and writing”. However, they insist that we should not “consider language skills to be part of language ability, but to be the contextualised realisation of the ability to use language in the performance of specific language use tasks”. Weir (1993:31) develops this point as follows:

To test whether learners can speak, it is necessary to get them to take part in direct spoken language activities. we want candidates to perform relevant language tasks and adapt their speech to the circumstances, making decisions under time pressure, implementing them fluently, and making any necessary adjustments as unexpected problems arise.

In this paper, we have studied the new GCE “O”-level oral English examination designed by UCLES because we are sure that it meets our demands mentioned above. Moreover, this examination has constructional similarity to the one conducted by the Society for Testing English Proficiency, Inc. (STEP), which prevails throughout Japan, although this GCE “O”-level oral English examination excels remarkably in assessing strategic competence¹³⁾, which “enables an individual to make the most effective use of available abilities in carrying out a given task” (Bachman 1990:106). Hence, when we use this examination as a formal

oral examination for Japanese EFL learners, we can say that it is more realisable in practice than any other oral English examination in the past from the viewpoint of validity, reliability, and practicality.

NOTES

1. If you test mainly grammar, your students will assume that this is the most important thing to learn and may make less effort during other more communicative activities (Harris and McCann 1994:27). Prodromou (1995:14) also points out how the examinations have quite an effect on the language learning in our classroom: "Many teachers, trapped in an examination preparation cycle, feel that communicative and humanistic methodologies are luxuries they cannot afford. When the market calls on teachers and institutions to produce quantifiable results, it usually means good *examination* results. Sound teaching practices are often sacrificed in an anxious attempt to 'cover' the examination syllabus, and to keep ahead of the competition."
2. Alderson *et al.* (1995:106) explain that "objective marking is used for multiple-choice, true/false, error-recognition, and other item types where the candidate is required to produce a response which can be marked as either 'correct' or 'incorrect'".
3. Subjective marking is usually used for marking tests of writing or speaking. Examiners are required to make judgements which are more complicated than the 'right-wrong' decisions.
(Alderson *et al.* 1995:107)
4. An examiner is judged to have *intra-rater reliability* if he or she gives the same set of scripts or oral performances the same marks on two different occasions. ...
Inter-rater reliability refers to the degree of similarity between different examiners.
(Alderson *et al.* 1995:129)
5. One marker will give an answer paper exactly the same score as another, assuming they can both count properly. The consistency between markers, or *inter-marker reliability*, is very high. Similarly, if one marker marks the same answer paper on two occasions, two or three weeks apart, she will give it the same score both times. In this case, the *intra-marker reliability* is very high.
(Underhill 1987:88)

6.

(1)	Human resources (e.g. test writers, scorers or raters, test administrators, and clerical support)
(2)	Material resources <i>Space</i> (e.g. rooms for test development and test administration) <i>Equipment</i> (e.g. typewriters, word processors, tape and video recorders, computers) <i>Materials</i> (e.g. paper, pictures, library resources)
(3)	Time <i>Development time</i> (time from the beginning of the test development process to the reporting of scores from the first operational administration) <i>Time for specific tasks</i> (e.g. designing, writing, administering, scoring, analysing)

Table 4. Types of resources (Bachman and Palmer 1996:37).

7.

Practicality = Available resources / Required resources
 If practicality ≥ 1 , the test development and use is practical.
 If practicality < 1 , the test development and use is not practical.

Figure 1. Practicality (Bachman and Palmer 1996:36).

8. It is also very important to be clear about what we want to assess and to ensure that we are assessing that and not something else, that our assessment has *validity*.
(Harris and McCann 1994:4)
9. The term *construct validity* is used to refer to the extent to which we can interpret a given test score as an indicator of the ability(ies), or construct(s), we want to measure.
(Bachman and Palmer 1996:21)
10. Canale and Swain (1980) describe communicative competence as four main components; grammatical, discourse, sociolinguistic, and strategic competence.
11. Weir (1993:xi) suggests that "it will not be possible to replicate reality fully because of the strong pulls of the other general principles of reliability and practicality".

12. The hiccoughs or inconsistencies in the judgement of single individual can be ironed out by combining her judgements with those of another person; the idea is that each person's hiccoughs are different so they tend to cancel each other out.
(Underhill 1987:89)
13. This examination "emphasises the sharing of personal opinions and reactions rather than allowing students to rely on ready-made attitudes", which pushes the students "to respond to points expressed by examiners quickly". This means that the students tend to exploit communication strategies to make up for a lack of knowledge of grammar or vocabulary.
(UCLES 1995a:31)

ACKNOWLEDGEMENTS

I wish to acknowledge gratitude to UCLES for their kind permission to reproduce samples of tests. I would also like to thank Dr. David Davies at ULEAC for giving me a personal interview and permission to record the interview so that I can write this paper later on.

Support for this paper was provided by a 1996 institute grant from Matsuyama University Research Centre.

REFERENCES

- Alderson, J. C., Clapham, C. and Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Brown, G. and Yule, G. (1983). Assessing spoken language.
In G. Brown and G. Yule: *Teaching spoken language*.
Cambridge: Cambridge University Press: 103-149.
- Canale, M. and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1/1: 1-47.
- Davies, D. (1996). Personal Interview. University of Cambridge Local Examinations Syndicate, London, 29 July 1996.

- Halleck, G. B. and Moder, C. L. (1995). Testing language and teaching skills of international teaching assistant: The limits of compensatory strategies. *TESOL Quarterly* 29/4: 733-758.
- Harris, M. and McCann, P. (1994). *Assessment*. Oxford: Heinemann.
- Heaton, J. B. (1990). Testing speaking skills. In J. B. Heaton: *Classroom testing*. London: Longman: 59-78.
- Konishi, K. (1995). Teaching Compensatory strategies in the Japanese EFL classroom: A core vocabulary and its pre-/post-modification — With special reference to a strategic syllabus —. *Studies in Language and Literature* 16/2: 81-112.
- Prodromou, L. (1995). The backwash effect: from testing to teaching. *ELT Journal* 49/1: 13-25.
- Thornbury, S. (1996). Teachers research teacher talk. *ELT Journal* 50/4: 279-289.
- Underhill, N. (1987). Marking system. In N. Underhill: *Testing spoken language*. Cambridge: Cambridge University Press: 88-103.
- University of Cambridge Local Examinations Syndicate. (1995a). *The new general certificate of education ordinary level oral English examination: Teacher's training manual*. Cambridge: University of Cambridge Local Examinations Syndicate.
- University of Cambridge Local Examinations Syndicate. (1995b). *General certificate of education ordinary level English language: Paper 3 oral English*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Upshur, J. A. and Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal* 49/1: 3-12.
- Weir, C. J. (1993). *Understanding and developing language tests*. Hemel Hempstead: Prentice Hall.

Appendix 1: Extracts from the sample keys attached to English teachers' training video for the new GCE "O"-level oral English examination

(A) Girl 4 (in black sweatshirt with white pattern)

(a) Reading aloud

Fluent, not many stumbles. Still a little fast. Body language issue?

Band 2/Mark 7 (see the section 3.5.4)

(b) Picture discussion

Begins well but Examiner seems rather to close her down. He asks her about relationship when she's already mentioned that it is the teacher and she is led into repeating about the boy being attentive.

Band 2/Mark 7

(c) Conversation

Doesn't respond too well to the Examiner's prompts. Passes up the opportunities.

Band 3/Mark 8

(B) Boy 5 (in white shirt)

(a) Reading aloud

Get worse. Tries to vary voice and offsets the slight jerkiness.

Band 2/Mark 7

(b) Picture discussion

Very slow and hesitant. Monosyllabic replies. Misinterpretation of figure in background. No response to question about background.

Band 4/Mark 3

(c) Conversation

A lot of closed questions from the Examiner. This candidate can read quite well but doesn't operate at all well in free talk. Long opening question brings only a one-word reply.

Band 4/Mark 3

(C) Girl 6 (in black shirt)

(a) Reading aloud

Began well but punctuation in the wrong place. Pace OK. Fairly fluent. Very few errors. Stumble at end.

Band 2/Mark 8

(b) Picture discussion

Begins with a confident overview of the picture including

background. Examiner starts to come in but she goes ahead. Fairly wide range of vocabulary and structures.

Band 1/Mark 11

(c) Conversation

Again a confident in speaking. Is willing to expand on closed questions. Pleasant attitude. Is confident enough to make a joke with Examiner (like eating, not cooking).

Band 1/Mark 15

(D) **Boy 7** (in beige patterned shirt and waistcoat)

(a) Reading aloud

Swallows word endings. Jerky.

Band 3/Mark 4

(b) Picture discussion

Fairly confident, not very fluent however. Quite a lot to say initially.

Band 3/Mark 6

(c) Conversation

Responds quite well. Expands slightly on closed questions but stops short of full expansion e.g. comment on University. Has some ideas e.g. cholesterol. The last question throws him slightly. Examiner could have been more helpful here.

Band 2/Mark 9

(E) **Girl 8** (in tartan jacket)

(a) Reading aloud

Very fluent, near perfect pronunciation, good pace.

Band 1 /Mark 12

(b) Picture discussion

Initially good description. Some good ideas about the background and situation. Good vocabulary e.g. "preoccupied", "far fetched".

Band 1/Mark 11

(c) Conversation

Doesn't expand on the initial question but then begins to expand. Some repetition but generally confident. Is able to respond before Examiner has finished speaking. Impression of conversation rather than question and answer.

Band 1/Mark 15

Source: Unpublished documents presented by UCLES, July 1996.

Reprinted with permission of UCLES.

Appendix 2: A copy of UCLES GCE "O"-level oral English language examination

(A) Examiner's Copy

This question paper consists of 2 printed pages. No additional materials are required.

[On page 1]

Instructions to examiners

The oral test will consist of the following parts:

Part 1 — Reading aloud

Candidates will be required to read aloud the passage printed on page 2 of their paper.

Part 2 — Picture discussion

Candidates will be asked to discuss the picture printed on page 3 of their paper. You should use the prompts overleaf to initiate discussion.

Part 3 — Conversation

You will engage in a conversation with the candidate according to the prompts and guidance overleaf.

[On page 2]

Part 2 — Picture discussion

Examiner's prompts:

Examiners should initiate the discussion through two main prompts — (1) and (2) below — but may feel free to use the additional prompts to encourage the candidate to respond.

- (1) Describe, in as much detail as you can, what you see in this picture.

Additional prompts:

- Describe one or two of the children in the picture individually.
- Describe the place where this picture was taken.

- (2) Who do you think the children in this picture are?

Additional prompts:

- Why do you think they look so happy?
- What do you think will happen in the arena?

Notes: What is required here are reasonable suggestions. There are no right or wrong answers.

Part 3 — Conversation

Examiner's prompts:

After using the main prompts — (1) and (2) below — examiners should use the additional prompts at their discretion, but should also pursue any points which the candidate may mention. Examiners may refer

back to the passage and/or picture for additional prompts.

(1) Do you agree that winning in sport is all-important?

Additional prompts:

— Do you think that parents should make children sacrifice other things to become sporting champions?

— How easy do you find it to balance school and sporting commitments?

(2) What sports do you most enjoy taking part in or watching?

Additional prompts:

— Tell me something about where you play them and your team.

— What was your most satisfying sporting moment?

Source: UCLES, 'General certificate of education ordinary level English language: Paper 3 oral English' (Cambridge: UCLES; 1995).

Reprinted with permission of UCLES.

(B) Candidate's Copy

This question paper consists of 3 printed pages and 1 blank page. No additional materials are required.

[On page 1]

Instructions to candidates

The oral test you are about to take will consist of the following parts:

Part 1 — Reading aloud

You will be required to read aloud the passage printed on page 2.

Part 2 — Picture discussion

You will be asked to discuss the picture printed on page 3.

Part 3 — Conversation

You will engage in a conversation with the Examiners on a topic given by them. You may now study the passage and the picture. You have 10 minutes to do so.

[On page 2]

In this passage, the writer talks about the problem of parents being too ambitious for their children.

All parents hope that their children will grow up to be happy, healthy and successful adults. To some extent, parents can share the successes of their children and, through them, achieve some of the

dreams and ambitions they themselves had when they were young.

The father who nearly made it as a footballer will watch anxiously for signs of similar or greater talent in his son, giving encouragement and support as each step is taken along the road to success. You only have to see parents at a junior football match to witness the almost fanatical support they give their sons.

In moderation, all this pride and ambition are good, constructive elements of a happy home, but taken to extremes they can be very dangerous. Football, for small boys, should be a happy, relaxed way of learning skills and getting rid of surplus energy. Team games also teach other valuable lessons of cooperation, sportsmanship and self-control.

However, if the child feels that he has to be a sporting success to keep his parents' love; if he is afraid that a mistake on the games field will lead to a scolding at home, all the pleasure is gone, and what should be a time of enjoyment becomes a source of pain.

If a boy or girl has outstanding sporting talent, gentle encouragement is all that is needed to develop it.

[On page 3]



Source: UCLES, 'General certificate of education ordinary level English language: Paper 3 oral English' (Cambridge: UCLES; 1995).
Reprinted with permission of UCLES.