

The Testing of Speaking

Yumi MIZUNO

0. Introduction

Morrow (1977:23) observes a potential imbalance between the concerns of language teaching and the resources of language testing and states 'Increasingly language teaching is becoming concerned with communication and its objectives are being reassessed in these terms; language testing, however, has still by and large failed to develop techniques for measuring effectively the use which is made of language in a communicative situation, preferring procedures which relate to the candidate's knowledge of and ability to manipulate the grammatical and phonologically systems of a language.'

In his study, Morrow considers that the following seven features are important in the use of language in a communicative situation: a) interaction-based, b) unpredictability, c) context, d) purpose, e) performance, f) authenticity, and g) behaviour-based.

The organisation of this essay is: in section one we examine Morrow's seven features of communication; in section two we take a speaking test from *Eiken ni kyu* and describe the background to the teaching/testing situation; in section three the test is analysed and evaluated according to Morrow's seven criteria; in section four an alternative test designed to meet Morrow's criteria is proposed.

1. Morrow's seven features of communication

Morrow (1982:56) mentions that language should be tested in ways and contexts which correspond to real life. We shall start our discussion by considering Morrow's seven features of communication.

a) Interaction-based

As Morrow observes, communication is a double-sided process. The form of interaction is obvious when we take short turns, where we decode the messages which are conveyed and integrate them into a response. That is, the receiver in turn becomes a producer, where not only modification of expressions and content but also a combination of receptive and productive skills are required.

Even in a seminar where a speaker sustains his/her turn, however, the language in use by the speaker is based on interaction. The speaker monitors the listeners' uptake of the message being communicated. The content of the speaker's message and the way in which it is expressed are affected by

the listeners' reaction. Thus, the speaker and the listeners bridge the information gap and accomplish the speaker's purpose (to let the listeners know what the speaker knows) and the listener's purpose (to get information from the speaker). Similarly, when listening to an announcement in an airport, we select a message which is unknown to us but necessary to accomplish our purpose (going to our destination) and re-encode it. This enables us to make our behaviour (boarding).

Thus, interaction is essential in order to accomplish the purpose of the speaker or the listener by bridging information gaps within the context.

b) Unpredictability

As Morrow states, if a listener is convinced that s/he knows what is coming, s/he does not listen to a speaker. Normally a speaker has information that a listener does not. For example, we listen to news in order to get new information. If the information is old, it does not count as communication.

However, the information is not totally unpredictable. The interaction between speaker and listener is based on a specific topic, shared knowledge between them or their schemata. Thus, in real life, a listener needs to predict, match what is actually said with what they expect to hear and interpret the overall meaning of each utterance in a short time to bridge the information gap in order to formulate an appropriate response.

Therefore in this essay, 'unpredictability' is considered as information gap to be bridged by interaction.

c) Context

As Morrow mentions, interaction is influenced by the role of the speakers, the status of speakers or the relationship between the speakers. The language between close friends is different from that between a subordinate and a boss. What is proper in one circumstance would be inappropriate in another. For example, in a shop, if the relationship between a customer and the shop assistant is very close, the customer can say 'Show me the dress' but if not, it might be inappropriate and might give an aggressive impression. Thus, it is important that language is appropriate to the sociolinguistic context in which it is uttered, such as role, status, familiarity and formality.

In addition, as Morrow observes, the mood of language users is expressed mainly by the attitudinal functions of intonation; pre-supposition is by sentence stress or grammatical/lexical devices (*John managed to solve the problem* implies that *John solved the problem.*); and previous utterances by anaphoric references. Without knowing these devices, we cannot interact with each other and the communication breaks down.

Therefore, it is important that the suitability of utterances should be tested: whether or not learners understand the appropriateness of their utterances to the situation, the mood of language users, the pre-supposition of utterances and anaphoric/cataphoric references.

d) Purpose

Morrow mentions that it is important to recognise the function of an utterance since every utterance is made for a purpose. On a logical level, a question is a means of determining what is true and what is false, however, questions can be used pragmatically for the purpose of making an offer as in *Would you like some more?*, making a suggestion as in *Why don't you come with me?* or expressing a strong feeling as in *Was it a marvellous play?* Thus, in real life, the listener would need to interpret the communicative function: whether it is an offer, a suggestion or a strong feeling.

He also mentions that we have to select a topic according to our purpose, using appropriate language to talk about the topic. For example, in order to establish our social relationship, we select phatic topics, which are neutral and non-threatening, such as the weather or family, depending on our culture. For successful interaction, the listener needs to recognise the function of utterances and the speaker needs to use appropriate language for the topic.

In testing, it is difficult for learners to have a real world purpose because of the fact that they are aware that they are in a test situation. However, if the task is made realistic, the purpose could be closer to being realistic.

e) Performance

In the real world, whether a person can drive a car is judged only by driving (performance) but not his/her knowledge necessary to drive a car (competence). However, emphasis has been given to competence (knowledge on language patterns) rather than to performance (use of language) in second-generation testing.

As Morrow mentions, performance is essential since what Chomsky considers as grammatically irrelevant, such as slips of the tongue, hesitation, false-starts, repetitions or errors, certainly exists in the speaker's spontaneous production of language (Morrow 1977:26).

Thus, in the tests of speaking, performance features should not be penalised.

f) Authenticity

Morrow says that this feature is related to all the features mentioned above. In the real life situation, the listener is exposed to what is unknown and unpredicted and language is not simplified to take account of the linguistic abilities of the addressee. That is, simplified syntax and vocabulary are not appropriate even if the listener's communicative ability is low. The input texts should be produced for native speakers of English for a non-pedagogical purpose. However, it is often better not to insist on pure textual authenticity in Morrow's sense but to use authentic-like spoken or written texts which lead to authentic interaction (Widdowson 1976:165, Geddes and White 1978:138). In other words, authenticity of task which derives from the purpose for interaction is important, though Morrow did not mention this in his original 1977 report. Tasks should be ones that learners recognise as being ones that real people in the real

world really do (Morrow 1982:56-57, Weir 1988:9). If the task is realistic, it can elicit the learner's response to achieve the purpose.

g) Behaviour-based

Morrow (1977:26) states, 'The success or failure of an interaction is judged by its participants on the basis of behavioural outcomes.' For example, an interaction in which we ask our friend to bring her notebook the next day will be judged a success if the person brings her notebook the next day.

As Morrow (1977: 27) admits, however, this feature is difficult to include in testing but not impossible.

2. The situation of language teaching and testing for which the test is used

The test examined in this essay is a language proficiency test called *Eiken ni kyu*, which is held twice a year all over Japan and is widely used. It is targeted for learners who have acquired fluency in oral communication based on the knowledge taught in senior high school (see Appendix 2). Although 'oral communication' is emphasised, this test is far from communicative, which will be examined in section four. The table below explains about the test briefly:

Title	Eiken ni kyu (English proficiency-second grade)
Conducted by	Nihon Eigo Kentei Kyokai (Japan English Proficiency Test Association)
Status	A qualification, which can be written on c.v.
Aims	To provide a comprehensive general English proficiency examination
Level	An intermediate examination requiring a knowledge of all language skills. Fluency in oral communication based on the knowledge taught in senior high school is essential.
Availability	Candidates must enter through a recognised centre. Examinations are held in June (and July on oral interview) and October (and November on oral interview).
Result	A pass/fail list is issued to centres, and a certificate is provided for successful candidates. Everyone who gets seventy percent of correct answers passes.

We shall now discuss the situation for which the test is used in Japan. The result of this test has been considered significant because it is a formal qualification which is accepted by Japanese society. Therefore, the majority of the university students and some of the successful senior high school students in Japan are interested in this test. However, since this test is a general proficiency test, there is no direct washback effects on teaching. The syllabus used in a language teaching situation is not directly related to the content of this test. The most powerful test, which has a poor washback effect, is a test used as an entrance examination by all the national universities and main private ones in Japan, which consists of only discrete-point items and does not assess learners' communicative skills: there is no listening test and no speaking test, which prevents senior high school language teachers from putting an emphasis on

communication and from promoting good language use in society. Thus, it is claimed that the primary aim in language teaching is to develop learners' communicative skills and 'communicative' is frequently used to describe current approaches to language teaching, but the focal point is still on the learning of the grammatical patterns of English.

In this situation, this proficiency test plays an important role in our society not only because a good result in this test has been acclaimed in our society but also because it is trying to be communicative. It is noted that there is no exaggeration to say that this is the only test that includes a test of speaking skill in Japan. Therefore if this test is revised to become a real communicative test, therefore, it could affect society and, as a result, could affect entrance examinations. This in turn provides a good washback effect on learning and teaching and a good washforward effect, so that learners acquire real language use which they are likely to encounter in real society.

3. Analysis and evaluation of the speaking test in terms of Morrow's criteria

In this section, we shall evaluate the speaking test according to Morrow's criteria. For convenience, in this essay, we examine the test held on the 17th of July, 1994 (Appendix 1).

The interview starts with testing pronunciation by reading a text aloud, followed by testing speaking skills by an oral interview. This test is not a communicative test for the following reasons. Firstly, this test is not a pure speaking test in that it requires listening skill and especially the skill of reading a passage (not an authentic passage) in a short time. Secondly, there is no actual interaction between a testee and an assessor. Apparently there is an interaction, but the interaction is dominated by the interviewer, which causes inequality between speakers and unnaturalness of interaction. The speakers cannot negotiate the message using turn-taking or back-channel feedback. Clarification or hesitation is penalised. Thirdly, testees must base their responses on the language used in questions asked by the examiner; no spontaneous production is required; and accuracy is more essential than fluency. The language used in no.1-no.4 is not beyond sentence level in that it is based on the interviewer's language and even the task to retell the story (no.5) would be based on the written text because testees are allowed to see the text to answer the questions. It is clear from the previous discussion that only 'linguistic competence', such as grammar or vocabulary, is assessed among the four components of communicative competence which Canale and Swain (1980) identified. 'Sociolinguistic competence', such as appropriacy of language to situation and interlocutor, 'discourse competence', such as the use of cohesion and coherence in organising a spoken scripts and 'strategic competence', such as repair strategies or communicative effectiveness, are not assessed. Fourthly, there is no information gap to be bridged in the task and testees do not have a real world purpose. Thus, this kind of oral test where testees are asked to read aloud a passage and then answer a series of questions on it would not meet the criterion of authenticity (Morrow 1982:57).

4. A test designed to meet Morrow's seven features

In this section, a revised test is proposed as a substitute for the proficiency test which we have so far examined.

4.1 Arranging the interaction

In the test, two learners speak to each other and the assessor and the interlocutor are largely outside the interaction. This reduces stress and unnaturalness of interaction. This increases reliability because the assessor can focus on assessment. However, if one learner is more advanced or more outgoing, there is a tendency for him/her to dominate the interaction. Consequently the other learner cannot display his/her skills fully. It is for this reason that two -on-two (two learners and assessors) is used in each test.

4.2 Speaking test

The test is the pre-arranged information gap to spot the difference carried out in pairs. Learner A has a picture which is related but different from B's (Appendix 2). They need to find out the similarities and differences, which becomes outcome.

Interlocutor explains each role. Candidate A is a boss and Candidate B is a subordinate. Candidate B has to put the furniture as Candidate A says. Interlocutor gives pictures. The pictures are similar but different in the position or the colour of the furniture.

Stage 1. Interlocutor asks Candidate A to describe the furniture of his/her picture to Candidate B, who has a picture which is slightly different.

Stage 2. Interlocutor asks Candidate B to tell what s/he should do (the difference between pictures).

Stage 3. Interlocutor asks Candidate B to describe his/her picture to Candidate A.

Stage 4. Interlocutor allows Candidate A to ask short questions. There is an interaction between learners. Candidate A asks Candidate B some questions.

Stage 5. Interlocutor asks Candidate A to say what Candidate B should do (the difference between the pictures).

Stage 6. Interlocutor allows learners to discuss until they reach an agreement. There is an interaction between learners.

Stage 7. Interlocutor asks for their conclusion. Candidate A speaks to an interlocutor; and Candidate B speaks to an assessor.

This task achievement (find out the differences) is one of the criteria for scoring. The interaction requires individual learner's to make transactional long turn rather than a short turn. During the interaction, they cannot show their pictures to each other, in order to make the task purposive and challenging. At stage 6, learners can ask questions to check or clarify the information. The interlocutor delegates in an individual task to each candidate and helps the learners only if the interaction breaks down, so that the interaction is

natural and equal. Even if one learner is dominated by the other at stage 6, each learner has an opportunity to show their performance at stage 7.

4.3 Rationale for the testing of speaking

a) Interaction-based

The learners are to interact with each other to bridge the gap. Candidate A gives orders and Candidate B understands the orders to find out differences.

b) Unpredictability

There is a genuine information gap to be bridged between speakers because learner A does not know what learner B knows and vice versa.

c) Context

A setting and the role of the learners are given. The setting is a company and the roles for the learners are a boss and a subordinate.

d) Purpose

The learners are given a topic and a language function. The topic is 'changing the position of furniture', the function is 'ordering where to put furniture in the room' and 'questioning the position of the furniture'.

e) Performance

The following is assessed: accuracy of grammar, vocabulary, etc.; appropriacy of language according to their roles; the ability to organise ideas and to make a coherent output; effectiveness and clarity of message are all tested. Turn-taking or negotiation skills are also measured. The learners' success in bridging the gap is taken as an outcome and an indication of their oral performance.

f) Authenticity

The task is authentic and communicative in that both learners have a purpose to transfer information across the gap between them by interaction so that they can succeed in their task; and authentic in that we need to describe objects accurately in real life.

g) Behaviour-based

The test is behaviour-based in that it is a preparation for changing the position of the furniture.

References:

Canale, M. and Swain, M. (1980) 'Theoretical bases of communicative approaches to second language teaching', *Applied Linguistics* 1/1.

- Carroll, B. (1982) 'Language testing - is there another way?' in Heaton, J. B. (ed. 1982) *Language Testing*, MEP/Macmillan.
- Carroll, B and Hall, P. (1985) *Make your own language tests*, Pergamon, Prentice Hall.
- Carroll, B. and West, R. (1989) *ESU Framework - performance Scales for English Language Examinations*, Longman.
- Geddes, M. and White, R. (1978) 'The use of semi-scripted simulated authentic speech and listening comprehension', *Audio-Visual Language Journal* 16/3.
- Harrison, A. (1983a) *A Language Testing Handbook*, Macmillan.
- Harrison, A. (1983b) 'Communicative testing: Jam tomorrow?' in Hughes and Porter (eds. 1983), *Current Developments in Language Testing*, Academic Press.
- Heaton, J.B. (1982) *Language Testing*, MEP/Macmillan.
- Heaton, J. B. (1988) *Writing English Language Tests, 2nd ed.*, Longman.
- Hughes, A. (1989) *Testing for Language Teachers*, CUP.
- Lado, R (1961) *Language Testing*, Longman.
- McEldowney, P. L. (1982) 'A place for visuals in language testing', in Heaton (ed. 1982) *Language Testing*, MEP/Macmillan.
- Morrow, K. (1977) *Techniques of Evaluation for a Notional syllabus*, Royal Society of Arts.
- Morrow, K. (1982) 'Testing spoken language', in Heaton, J. B. (ed. 1982) *Language Testing*, MEP/Macmillan.
- Morrow, K. (1983) 'The RSA Examinations in the communicative use of EFL', in Jordan, R. (Ed. 1983), *Case studies in ELT*, Collins.
- Nation, P. (1989) 'Speaking activities: five features', *English Language Teaching Journal* 43/1
- Rea, P. (1982) 'An alternative approach to testing grammatical competence', in Heaton (ed. 1982) *Language Testing*, MEP/Macmillan.
- Underhill, N. (1982) 'The great reliability/validity trade-off', in Heaton (ed. 1982) *Language Testing*, MEP/Macmillan.
- Weir, C (1988) *Communicative Language Testing*, University of Exeter.
- Widdowson, H. (1976) 'The authenticity of language data' in Widdowson, H. (1979) *Explorations in Applied Linguistics*, Oxford University Press.
- Wingard, P. (1981) 'Assessment' in Abbott, G. and Wingard, P (ed. 1981), *The Teaching of English as an International Language*, Collins.

Test:

Eiken ni kyu (1994), Nihon Eigo Kentei Kyokai

Appendix 1

Card A

The Japanese people are internationally famous for their artistic skill with flowers. Flower arrangement, or *ikebana* as it is called, has its roots in the 6th century. Flower arrangement started in China with Buddhist monks, who used flowers in their temples. When Buddhism was introduced to Japan in the 6th century, Japanese monks gave special thought to arranging temple flowers artistically. As time went on, however, flower arrangement became the pastime of many people of the upper class such as nobles, feudal lords, and samurai warriors. In the 15th century flower arrangement developed into an aristocratic art form. Today it has become a popular pastime enjoyed by many people throughout the world.

Questions

1. What are the Japanese people internationally famous for?
2. Where did flower arrangement start originally?
3. What did Japanese monks give special thought to in the 6th century?
4. When did flower arrangement develop into an aristocratic art form?
5. Please give me an outline of this story.

Card B

The new inventions and machines of the 18th century gave people new opportunities. However, these inventions could not become fully useful until men had succeeded in finding sources of power from nature to replace their own muscles. Windmills as a source of power had been well known for centuries, but they were too unreliable to work the new machinery. Water power proved to be more reliable. But even water power had its minuses. Factories had to be built along rivers; if the rivers froze over in winter or became dry in summer, manufacturing operations had to be stopped. It was not until the steam engine came into use that man had a really reliable source of power for his machinery.

Questions

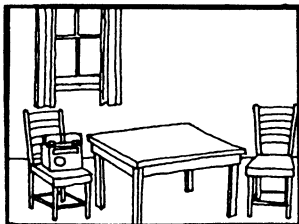
1. What gave people new opportunities in the 18th century?
2. What was wrong with windmills as a source of power?
3. When did factories run by water power have to stop?
4. What gave man a truly reliable source of power for his machinery?
5. Please give me an outline of this story/

Appendix 2

[Input]

Visual stimuli in the form of two different pictures per candidate with verbal rubrics.

Candidate A



Candidate B

