

The Evaluation of JACET Test

Kaori Tsuchimochi

Introduction

The JACET Listening Comprehension Test was developed by the Japan Association of College English Teachers (JACET) as a standardized test to assess listening proficiency of Japanese university and college students. JACET offers two parallel tests called Form A and Form B. Form A is given in May, and Form B is in October. In the Kagoshima Prefectural College these two tests are given to the English major students as a proficiency test by which the teacher can assess the level of students' listening ability and check each student's progress by comparing their results with those of past tests. Apparently many students are also motivated by these tests to achieve higher scores and consequently these tests can be said to be helpful to both the teacher and the students.

On the other hand, in the wake of development of communicative language competence in the last decade, the emphasis in listening comprehension tests has moved away from linguistic knowledge towards functional global decoding in contextualized real – life situations. Accordingly, the JACET Test, developed approximately two decades ago, should be evaluated under both orthodox criteria and current criteria in light of communicative language competence.

First, I would like to evaluate the extent of the usefulness of the JACET Test for both teachers and students. Second, some of the variables which increase the reliability of the JACET Test will be examined. Finally, the level of validity of the contents of the JACET Test shall be evaluated with reference to the current emphasis in listening comprehension tests.

I. How results of the JACET Test can be utilized

Generally tests are seen as a means of reflecting teaching ; in other words, testing is always expected to have a 'washback' effect with regard to the teaching. In my teaching situation, however, the JACET Test is given twice a year to evaluate the students' listening proficiencies and their growth, and also to raise the students' awareness of their own listening ability. On the other hand, this test does not necessarily seem to bring beneficial effects. Therefore, in this section, I would like to evaluate the actual usefulness of the JACET Test.

First, positive feedback for teachers and the students will be examined. It is suggested that results of this test are opportunity to motivate the students by increasing awareness of their own listening level.

Two reasons are assumed for this effect. One is that the students can not be made explicitly aware of progress in their own listening ability by results of achievement tests given at the end of semester, since in these tests they are mostly tested on items of the textbook used in the class. The other reason is that generally Japanese students seem to be encouraged to study positively to challenge some tests, because they have become accustomed to such a way of studying up to secondary institution level. Although it may be debatable whether such instrumental use of the test is good pedagogy or not, this is the real situation of the students. As listening ability is of an implicit nature likewise reading ability, it is very difficult for students to measure it without taking any tests to show them specific results (i. e. scores) or levels.

Next I would like to show what sort of information teachers can attain from results of the JACET Test. As Davies suggested, testing can be used as ‘a method of providing information that may be used for teaching and other purposes’ (Davies, 1990, p24). The JACET Test also can provide beneficial information for language, language learning and teaching. The teacher can measure both the present level of individual student’s listening proficiency and the progress of her listening proficiency by comparison with the last or past results. In addition to the data about the students, JACET provides teachers not only with the scores of individual students but also with detailed data such as average score, standard deviation of the total candidates, and other statistics according to the grade of university and college. Second, as JACET also recommends, the teacher can make use of the data for their own research by analysis of test items and students’ results for their teaching and investigation of language learning. For example, Ushio (1982) made an insightful research paper based on analysis of the JACET Test and data about the relationship between college – students’ listening progress and factors which affect listening competence.

However, while acknowledging the beneficial information provided by the JACET Test, we also need to recognize insufficient aspects of this information. Whereas the JACET Test can provide evidence of the level of the achievement of the students, it will not provide any relevant information as to the specific areas in terms of students’ difficulties in listening comprehension. Weir argues against the use of such indirect tests are, ‘only of limited use where the interest is in what the individual student can or cannot do in terms of the various language tasks he may face in real life situations’ (Weir, 1990, p6).

II. What elements are included in the JACET Test to increase reliability

In order to develop better tests, it is crucial to keep the proper degree of tension between ‘reliability’ and ‘validity’. In this section, ‘reliability’ of the JACET Test will be examined by checking some variables that can increase reliability of listening tests. Hughes (1989) suggests that there are two origins of unreliability, that is, features of the test itself and the way it is scored. Therefore, we need to consider

how appropriately these factors are dealt with in order to keep the test consistent.

1. Some features that can help JACET Test to be reliable

In this section, I will take some specific features and evaluate whether they operate to make the test reliable. These features are also important in providing the students with consistent conditions in which they can work properly as suggested by the tasks. As stated earlier in this paper, the JACET Test is given through test – tape ; therefore, instructions, samples, and test tasks are all given through listening.

(1) Clarity of task

The test task of the JACET Test is explicit and easy to understand, because all questions are answered by marking the answer sheet. This test is organized in three parts, and each task is as follows:

Part I : true – false test with 20 questions;

Part II : same – different test with 20 questions;

Part III : multiple – choice test with 10 questions;

Students are familiar with these types of tests, since most of the students have experience of such multiple – choice type tests from their preparation for entrance examinations. For this reason, the JACET Test can be said to be a candidate – friendly type test.

(2) Clarity of instructions to testee

Since the students taking the listening test tend to be more nervous and have to work under mental pressure because of the transient nature of listening process (Yoshida, 1993), complex instructions will add more load on their listening process. Moreover, as Heaton (1990) suggested, rubric should not in itself become a test of listening comprehension. In order to avoid these problems, in monolingual situations giving instructions in the students' first language is suggested (Heaton, 1990 ; Weir, 1993). In the JACET Test instructions of tasks are given clearly in the students' first language (i. e. Japanese) at the beginning of each part.

(3) Provisions of sample items

Provision of a clear sample as an example of the task is recommended and this preparation will prevent the candidates' scores being unfavorably affected by the test in which they are about to participate (Heaton, 1990 ; Weir, 1993). In the JACET Test two examples are presented in each part with a clear instruction about the task and the reason why that answer is appropriate. Such examples will make the students familiar with the tape – recorded voice and the fairly rapid speed of utterances which possibly

influence students' performances.

(4) Large numbers of test items

It has been argued that the addition of further items will make a test more reliable and this has been demonstrated empirically (Hughes, 1989). Based on these grounds he suggested the likely greater reliability of multiple – choice tests:

A further considerable advantage is that, since in order to respond the candidate only has to make a mark on the paper, it is possible to include more items than would otherwise be possible in a given period of time.

(ibid. p6, pp.59 – 60)

In the JACET Test fifty test items are included within the test timeframe of approximately twenty minutes. Although it is difficult to judge whether the above number of test items of the JACET Test is enough or not, we may conclude that a fair amount of test items are included in this test according to Hughes.

2. How scoring of the JACET Test is reliable

The JACET Test is a multiple – choice type test including a 'true – false test' and a 'same – different test' whose greatest advantage is that scoring can be perfectly reliable because it cannot be affected by the scorer's judgement (e. g., Hughes, 1989 ; Weir, 1990).

Here I will examine how multiple – choice type scoring of the JACET Test is likely to be reliable. The JACET Test adopts a unique scoring system to increase reliability as follows:

Scoring of Part I and Part II

$$(R - W) \times 2 = \text{scores}$$

Scoring of Part III

$$(R - W / 2) \times 2 \times 2 = (2R - w) \times 2 = \text{scores}$$

Total scores = Part I scores + Part II scores + Part III scores

R : total number of a right answer W : total number of a wrong answer

It is often argued that the scores gained in multiple – choice tests may be suspect because the testee might have guessed answers. In order to reduce the possibility of such unreliability, the JACET Test adopts the above scoring system. It has been demonstrated statistically by the use of 'Kuder – Richardson formulas

20 or 21' that the 'coefficient of reliability' by the scoring system shown above was higher (0.864, 0.699) than that (0.983, 0.963) of the normal scoring system (Yoshida, 1993). The instruction of the test – tape also directs the students not to mark if they are not sure of their answers. Thus, the scoring system adopted by the JACET Test can reduce the unreliability caused by testees' guessing to some degree. Therefore, we can say that the JACET Test has high reliability for scoring.

III. How the JACET Test is valid as a listening comprehension test

1. Validity as a discrete – skill test

The JACET Test is a discrete – skill test which requires testee (i. e. students) to use just listening skill to answer questions. We can say that such a discrete – skill test will be able to make listening tests valid for the following reason. If the test requires the student to use other language skills besides listening, it is very difficult to judge whether or not the causes of the incorrect answers are due to the student's insufficient listening ability. In short, other language skill can interfere with the assessment of listening comprehension in integrative tests. The JACET Test which is a discrete – skill and multiple – choice test can avoid this particular difficulty in assessment.

2. Content and construct validity of the JACET Test

In this section, the content and construct validity of the JACET Test shall be evaluated. Accordance with the current popularity of communicative language teaching and learning, English classes given in the language laboratory have also gradually changed toward the development of communicative competence. Not every student in another university taking the JACET Test is necessarily taught in the class where they can develop listening skill available in real life situation. On the other hand, we can easily assume that the JACET Test, completed more than fifteen years ago when the psychometric – structuralist and statistically driven approach were the concerns of testing, will not satisfactorily meet the criteria of content and construct validity in the communicative language teaching and learning which is at the psycholinguistic – sociolinguistic stage. Therefore, we need to know the basic elements which should be introduced into the current standardized listening test. Weir stated the change of the focus of listening tests briefly:

The emphasis has shifted to contextualised tests of listening comprehension, incorporating all the additional redundant features that facilitate comprehension. The concern now is with testing the communication of meaning as against structural understanding.

(Weir, 1993, p99)

Rost also includes the importance of context in listening tests as criteria for assessment validity:

- (1) Items should be in contexts similar to those experienced in the target settings ; language used should reflect natural conditions of use.
- (2) The test – taker should have the opportunity to utilize contextual knowledge in making sense of the items and response tasks.

(Rost, 1990, p185)

Thus, I would like to evaluate content and construct validity of the JACET Test principally in light of ‘contextualization’.

First of all, we will examine validity of the type of response required in the JACET Test. It is apparent that the JACET Test consisting of multiple – choice type tests, including a true – false test and a same – different test, will require only a single correct response. Weir (1990) shows considerable doubt about the validity of multiple – choice items as measures of language ability, since selecting only one correct answer is an unreal task in real – life situations.

Next, we shall evaluate individual parts of the JACET Test by giving particular consideration to the importance of contextualization of test items discussed above. Firstly, let us evaluate the validity of Part I. Part I requires the testee to choose ‘true’ or ‘false’ after hearing two utterances:¹

1. It’s just nine o’clock, but my watch says nine to nine.

So it’s nine minutes fast.

2. Suppose you are flying west.

You can see the north side view through the left – hand – side window.

The two statements in each item give the testee contextual clue to judge them. Moreover, the testee may need to use referential knowledge or his own non – linguistic knowledge to judge the items. Emery admits the validity of this type of true – false tests:

In everyday conversation, one use of the listening skill is to determine the correctness or otherwise of other speakers’ statements. An adult native speaker of English will be able to say without hesitation that a statement like ‘The sun rises in the west’ is false, while ‘A square has more sides than a triangle’ is true. Therefore the degree to which a foreign learner approximates to the native speaker’s skill in evaluating such statements might enable us to assess his listening ability.

¹Examples presented in this paper are partly modified statements in the real questions used in the JACET Test.

(Emery, 1980, p97)

Secondly, in Part II, the testee needs to judge if the meaning of the statements is the same or different after hearing a set of two utterances:

(a) He continued working until six.

(b) He worked even after six.

(a) If she comes before I get back, ask her to wait.

(b) Ask her to wait until I get back.

Two statements in a set are presented in isolation without any context, and this seems to make it difficult for the testee to both get each meaning and compare them within the limited pause (approximate 7 seconds) given to answer. Chastain also criticizes the test items given without context:

Listening comprehension test items often deviate from the real – language context by presenting only one linguistic clue to the meaning being tested. Thus, in a sense, learners are presented with comprehension tasks that are more difficult than those met in real – language situations.

(Chastain, 1979, p83)

Thus, we can say that test items of Part II have low validity.

Finally, Part III asks the testee to listen to a short passage and then choose the answer to the question given at the beginning of each passage:

Why didn't Lucy meet her friend at the airport?

Lucy went to the airport to meet her friend from New York. His plane arrived safely but he wasn't on it. Then she took out his letter from her bag. The letter said that he was arriving at 5 o'clock on the 10 th. "How silly of me !" she said to herself. "Today is the 9 th."

A. Because she didn't know the time of his arrival.

B. Because she didn't wait for her at the right place.

C. Because she went to the airport on the wrong day.

Compared with other parts, part III seems to provide more context clues in the test text. Part III includes

ten test items similar to the above example ; however, all passages consist of only descriptive statements by one person and no interpersonal utterances. Therefore, the testee will not be given the opportunity to infer the meaning of real – life utterances relating to social or situational contexts she will meet in the real world. For this reason, part III is not likely to have high validity, by judging from the viewpoint of measuring students' communicative competence.

As we have observed in this section, the JACET Test is somewhat lacking in testing students' listening ability with regard to real – life situations. Thus, content validity of this test will be lower under the criteria of current English language teaching and learning.

Concluding Remarks

In this paper, the JACET Test was evaluated against three main criteria; usefulness, reliability, and validity. Although the JACET Test consisting of a multiple – choice format is highly reliable and can provide teachers beneficial information, some test items need to be reviewed to enhance the test's validity by considering the current emphasis on the development of communicative competence. JACET seems to try to motivate students by making them take the JACET Test which is very difficult to attain high scores.

However, the JACET Test should be improved to approximate authentic conditions of communication, if JACET wants to measure students' listening ability in real – life situations.

References

- Chastain, K. D. (1979) 'Testing Listening Comprehension Tests', *TESOL Quarterly*, vol. 13, no. 1, pp.81 – 88.
- Davies, A. (1978) 'Language testing:Survey article part II', *Language Teaching and Linguistics Abstract*, vol. 11, no. 4, pp.215 – 231.
- Davies, A. (1990) *Principles of Language Testing*, Oxford, Basil Blackwell.
- Emery, P. G. (1980) 'Evaluating Spoken English : A New Approach to the Testing of Listening Comprehension', *ELT Journal*, vol. 34, no. 2, pp.96 – 98.
- Heaton, J. B. (1990) *Classroom Testing* (2 nd ed), Harlow, Longman.
- Hughes, A. (1989) *Testing for Language Teachers*, Cambridge, Cambridge University Press.
- Morrow, K. (1979) 'Communicative language testing: revolution or evaluation?', in C. J. Brumfit and K. Johnson (eds.), *The Communicative Approach to Language Teaching*, Oxford, Oxford University

Press, pp.143 – 157.

Rost, M. (1990) *Listening in Language Learning*, London, Longman.

Ushio, K. (1982) 'An analysis of listening comprehension', *Jinbun*, vol. 6, Kagoshima, Kagoshima Prefectural College, pp.41 – 49.

Weir, C. J. (1990) *Communicative Language Testing*, London, Prentice Hall.

Weir, C. J. (1993) *Understanding and Developing Language Tests*, Hemel Hempstead, Prentice Hall.

Yoshida, K. (ed.) (1984) *Eigo No Listening* [Listening in English], Tokyo, Taishukan.

(1999 年 5 月 10 日 受理)