統合的スピーキングテストの理論的・実践的問題
Theoretical and Practical Issues of Integrated Speaking Tests

中村　優治 NAKAMURA, Yuji

● 慶應義塾大学
　Keio University

ABSTRACT

　外国語教育の分野で急速に取り上げられつつあるのが内容・言語統合型学習である。ところが統合型教育の重要性が強く叫ばれても，これと表裏一体をなす統合型評価に関しては充実した議論の展開が見られないようである。学生がテキストを読み，講義を聴き，メモを取ってスピーチに活かすことは技能統合型の活動であるが，この評価をする教師の方の受け入れ態勢がまだ整っていないため，伝統的な技能別，さらに内容と言語別々の採点がなされているのが現状である。本稿では統合型スピーキングテストの諸問題を4つの視点（TOEFL iBT，指導要領，統合型テスト理論，実践例）から議論し解決への一提案を試みようとするものである。

　The content and language integrated learning approach has been gaining popularity in second/foreign language education in many places in the world. While an integrated approach to language and content learning and instruction has been promoted, the aspect of integrated language and content assessment has not been sufficiently discussed. When a student makes a speech, based on the lecture he/she listens to, language and content are integrated. However, teachers find it difficult to assess language and content in an integrated way since they are not accustomed to doing so. Accordingly, they maintain the traditional practice of assessing language and content separately. The purpose of the paper is to discuss the fundamental issues involved when implementing a speaking test in the scope of integrated test. The problems will be raised and discussed from

four aspects: a) integrated tests in terms of TOEFL iBT; b) Plakan's idea regarding integrated assessment and the ideas of MEXT (the Course of Study); c) Case Study: JACET Workshop administered by the JACET Testing SIG. Finally, the paper will conclude with some ideas on this important testing issue.

The content and language integrated learning (CLIL) approach has been gaining popularity in second/foreign language education in many places in the world (Mohan, Leung and Slater, 2010). While an integrated approach to language and content learning and instruction has been promoted, the aspect of integrated language and content assessment has not been sufficiently discussed. When a student makes a speech based on the lecture he/she listens to, language and content are integrated. However, teachers find it difficult to assess language and content in an integrated way since they are not accustomed to doing so. Accordingly, they keep the traditional practice of assessing language and content separately (Mohan, Leung and Slater, 2010).

The purpose of the paper is to discuss the fundamental issues when implementing a speaking test in the scope of integrated test. The problems will be raised and discussed from four aspects: a) integrated tests in terms of TOEFL iBT, b) Plakan's idea regarding integrated assessment and the ideas of MEXT (the Course of Study); c) Case Study: JACET Workshop administered by the JACET Testing SIG. Finally, the paper will conclude with some ideas on this important testing issue.

1. Integrated tests in terms of iBT TOEFL

One issue in the testing of oral skills is the degree to which it is possible to isolate speech from other skills in test design. This is known as the distinction between integrated versus discrete skills testing (Hughes, 2002).

The Test of English as a Foreign Language Internet-based test (TOEFL iBT), for instance, includes speaking as a mandatory section. The TOEFL iBT Speaking test has been designed to measure candidates' ability to communicate orally in English within an academic environment (Xi, 2007).

The TOEFL iBT includes tasks that resemble those performed by students in North American colleges and universities. These 'integrated skills' tasks require test takers to use the information provided in the reading and listening passages to compose essays and/or spoken responses, in addition to an independent task (cf. Jamieson, 2005).

McGinley (2006) describes how integrative tests have developed along with changes to language teaching methods; from grammar-translation, to audio-lingual and from structural to communicative. Generally speaking, the first generation saw the use of non-authentic texts, tasks that lacked context, and subjective scoring. Second generation tests corrected the subjectivity of the earlier tests through the use of objective type questions on discrete aspects of language. The third generation tests were a reaction against previous tests and tasks, and were authentic (from real-world sources) and contextualized (simulated real tasks) (McGinley, 2006). Because of this, assessment is necessarily of integrative language use (of language and other skills). This style of assessment has been achieved in the case of productive skills such as speaking (cf. McGinley, 2006).

Types of integrated tasks require integration of two language skills, as in listening/speaking and reading/speaking tasks. Integrated tasks have been advocated for two main reasons (cf. Lewkowicz, 1997):

> 1） Test-takers are less likely to be disadvantaged due to lack of information on which to base their argument.

2）The validity can be enhanced by simulating real-life communication tasks in academic contexts.

Some concerns can be raised about speaking assessments composed of integrated tasks. A primary concern is the general applicability of a task and its dependency across test sections. These concerns are also related to the role of input stimuli in eliciting examinees' spoken responses. A claim can also be made that each of these different speaking task types are somewhat distinct in construct, therefore separate scores should be reported for each of these distinctions when two types of integrated tasks are dissimilar in input stimuli characteristics (a reading passage and auditory text), as test takers might use different cognitive skills and processes in responding to these different types of speaking tasks (Lee, 2006).

A similar argument can be made about the rating process for examinees' speech samples. When rating examinee responses from integrated tasks, raters must also attend to content accuracy to make sure that the examinees have adequately understood what has been presented in the text or lecture (Lee, 2006).

2. Plakans' idea regarding integrated assessment and the ideas of MEXT (the Course of Study)

Plakans (2011) discusses integrated assessment in the following way:

First, she defines integrated assessment as the use of tests that combine two or more skills such as reading/writing or reading/listening/speaking. Usually the receptive sections lead to a final (writing or speaking) performance section. There are various relationships between the receptive and the productive sections: the performance could be a summary of or an opinion on the topic, or it could simply involve responding to a related topic, where as the receptive sections are used for inspiration.

Integrated tasks require examinees to integrate multiple language skills in a substantial way to complete whatever speaking task is at hand, e.g. to understand academic texts and lectures and create spoken responses that demonstrate understanding of those texts and lectures. "Integrated skills" tasks thus require test takers to use the information provided in the reading and listening passages in essay and/or spoken responses.

Second, Plakans (2011) discusses the difference between integrated and individual skill testing. She claims that integrated tests are more complex, containing tasks requiring "multiple steps or a longer process." Independent speaking tasks are tasks based on a stand-alone prompt or visual, while integrated tasks involve both listening and speaking, or both reading and speaking. While the integrated tasks provide the information about which examinees will speak, the independent tasks usually require examinees to rely on their personal experience or general knowledge to complete the task.

Integrated tasks are more complicated to construct, requiring that test-makers first develop an input task and then ensure that that the performance task is appropriate for the input. Plakans also states that the scoring may be different; it could rely at least in part on an assessment of whether the test-taker has used the input material appropriately.

Third, Plakans (2011) introduces several methods of integrated assessment. She recommends the use of multiple measures, i.e. pair integrated tasks with independent tasks, or utilizing more than one integrated task. She also emphasizes the importance of pilot testing because both test takers and raters are unfamiliar with this test format. It is important to look at how it is handled.

Fourth, Plakans (2011) describes the challenges of integrated assessment. She indicates that the main issues are: 1) the construct of speaking ability, 2) the tasks (test tasks and response tasks), and 3) the scoring and raters. It is important that the skill an

integrated task measures be considered. It is crucial that, as the test format is unfamiliar, instructions be very clear and succinct. When rating examinee responses from integrated tasks, raters must attend to content accuracy in order to ensure that the examinees have adequately understood what is presented in the test or lecture. Additionally, raters must consider the source material when scoring.

Finally, Plakans (2011) stresses the benefits of integrated assessment. Plakans claims that, in spite of the complexity and difficulty in its nature, integrated assessment has the following benefits: 1) Integrated assessment is less reliant on the 'four skills' model of language and appreciates that language is not a unitary construct; language skills interact with one another. 2) Integrated assessment has more authenticity: these tasks are similar to those encountered in real-world language use. 3) Performance tasks in integrated assessment are easier with an input task as inspiration, and having read or heard something on the topic gives test takers more confidence in their performance. 4) Students receive positive wash-back effects from these test tasks. 5) The purpose of integrated assessment is aliged with the objectives of the Course of Study for Japanese high school students (cf. MEXT). The gist of this objective is as follows:

- Understanding information, ideas, etc., and grasping the outline and main points by listening to introductions to specified topics, dialogues, etc.
- Understanding information, ideas, etc., and grasping the outline and main points by reading explanations, stories, etc. Reading passages aloud so that the meaning of the content is expressed.
- Discussing and exchanging opinions on information, ideas, etc., based on what one has heard, read, learned and experienced.

3. Case Study: JACET workshop that was administered by the JACET Testing SIG

A general workshop of testing is as follows: In the first stage, a lecture was delivered on testing theory and relevant testing terminology such as validity, reliability, practicality, and test specifications. In the second stage, teaching, a lesson plan was made that deals with one of the four language skills. In the third stage, test making, the participants gained invaluable hands-on experience constructing tests covering the lesson content of the previous stage. At this point, the strengths and weaknesses of various test types and item types were discussed. Additionally, topics such as rating scales, rating criteria, and grading rubrics were covered. The fourth stage, test data analysis, was also hands-on in approach. Here, statistical concepts, such as means, standard deviations, histograms, standard scores and correlations, were examined in a meaningful context. Furthermore, more practical aspects of item analysis, such as item difficulty and item discrimination, were dealt with. After completing the training, the pre-service and in-service teachers would then apply these concepts and use the associated techniques in their classes.

The applied workshop on Integrated assessment, focusing on speaking ability, proceeded as follows (using junior high school textbook):

Procedures:
a. Sample of an integrated skill lesson
   A sample of an integrated skill lesson was conducted by a university teacher.
   Textbook: New Horizon 3 (pp.48-49)
   Students: Workshop participants—prospective English teachers, teacher trainees, and university students majoring in English Teaching Methodology
   Topic: Asking for directions and giving directions

b. Making an integrated test of speaking ability

Students design a test based on the lesson they took.

Items to be taken into account when making and editing a test were explained first as follows:

a. What is the purpose of the test?

b. What skill is to be measured?

c. What kind of integrated tasks can be constructed?

d. Who is the rater?

e. What are the evaluation items and what is the rating criteria?

f. How long is the testing time?

g. How is the test administered?

c. Testing (examining) an integrated test

Students (workshop participants) examine their own tests one by one through the group discussion with the assistance of the lecturer.

Items to be taken into account when evaluating the tests are as follows:

a. Is the test measuring what it was intended to measure?

b. Are the rubrics clear and succinct?

c. Are the evaluation items appropriate?

d. Is the test practical?

4. Conclusion

In addition to these three different aspects of integrated assessment, Taylor's (2011) suggestions also deserve attention. She claims that the test taker's cognitive abilities, the tasks and content of the test, and the scoring process of the test are three triangularly related components critical to any language-testing activity. Focusing on these three components offers a theoretically sound and directly practical concept of construct validity in terms of testing.

In conclusion, when developing a new test, just as when developing integrated assessment instruments,

the following ideas should be paid special attention: (cf. Council of Europe, 2001; O'Sullivan and Nakatsuhara, 2011):

a. What is to be assessed (construct and content)

b. The employment of a range of tasks to provide the test takers with opportunities to perform their best (test method)

c. Indication of how the performance is scored (rating criteria and raters)

d. Explanation of how the test result is interpreted (score interpretation)

e. Separate scores must be reported for each task, not a single combined score (test result reporting)

f. The development of locally sensitive instruments (practicality)

References

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

JACET Testing SIG (2011,September). 2011 JACET Testing SIG Summer Workshop program. The Japan Association of College English Teachers (JACET) Testing SIG.

Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, *25*, 228–242.

Lee, Y-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, *23*(2), 131–166.

Lewkowicz, J. (2000). Authenticity in language testing: some outstanding questions. *Language Testing*, *17*(1), 43–64.

Hughes, R. (2002). *Teaching and Researching Speaking*. London: Pearson Education.

McGinley, K. (2006) The 'Test of Interactive English'—from conception to implementation. *ELT Journal*, *60*(4), 374–381.

(MEXT) The Ministry of Education, Culture, Sports, Science & Technology. (2012, February 29). Retrieved February 29, 2012 from http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/eiyaku/1298353.htm

Mohan, B., Leung, C. and Slaiter, T. (2010). Assessing Language and Content: A Functional Perspective.

In Paran, A. and Sercu, L. (ed.) *Testing the Untestable in Language Education* (chapter 11). Toronto, Canada.

Murray, A., Ito, Y., Kimura, K., Matsumoto, K., Nakamura, Y., & Okada, A. (2011). *Current trends in language testing education in Japan*. In A. Stewart (Ed.), JALT2010 Conference Proceedings. Tokyo: JALT.

O'Sullivan, B and Nakatsuhara, F. (2011). Quantifying Conversational Styles in Group Oral Test Discourse. In O'Sullivan, B. (ed). *Language Testing: Theories and Practices* (chapter 8). New York: Palgrave Macmillan.

O'Sullivan, B. (ed). (2011). *Language Testing: Theories and Practices*. Palgrave Advances.

Plakans, L. (2011, September 5). Integrated Assessment. Retrieved September 5, 2011, from http://languagetesting.info/video/main.html

Taylor, L. (2011). Introduction. In L. Taylor (ed.), Examining Speaking: Research and practice in assessing second language speaking. Cambridge, UK: Cambridge University Press.

TOEFL (2012). TOEFL iBT Tips. Retrieved March 9, 2012 from the TOEFL Web site http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf

Xi, X. (2007). Evaluating analytic scoring for the TOEFL Academic Speaking Test (TAST) for operational use. *Language Testing*, *24*(2), 251–286.