Fundamental Issues Surrounding Integrated Tests in Terms of Assessment Literacy
- The Case of Integrated Speaking Tests -

Adam Murray[1], Tomoyasu Akiyama[2],

Kahoko Matsumoto[3], Kei Miyazaki[4], Yuji Nakamura[5]

and Taiko Tsuchihira[6]

[1] [3]Tokai University, [2]Bunkyo University, [4]Keio High School, [5]Keio University, and [6] University of Tsukuba

murray@scc.u-tokai.ac.jp, akitomo@koshigaya.bunkyo.ac.jp,

mkahoko@tsc.u-tokai.ac.jp, kei@hs.keio.ac.jp, nkyj@flet.keio.ac.jp,

t-tsuchihira@nifty.com

Abstract

This paper discusses fundamental issues to consider when designing and implementing a speaking test in the context of integrated-skills language assessment, which has become increasingly necessary, as the high school course of study in Japan has been directed toward integrated teaching and learning based on the new guidelines of MEXT (the Ministry of Education, Culture, Sports, Science, and Technology). The issues and problems will be considered in light of a recent concept, assessment literacy (Inbar-Lourie, 2008; Taylor, 2009) for teacher training: namely, what teachers should know about testing and assessment to make well-informed decisions in their teaching practices in order to optimize the improvement of their students. The paper indicates the necessity of raising the awareness of pre-service and in-service teachers' assessment literacy and concludes with suggestions for further research on integrated speaking tests and further work on workshop practices.

Keywords

speaking evaluation, integrated-skills tests, and teacher training

1 Introduction

As the high school course of study in Japan has shifted toward integrated teaching and learning based on the new Ministry of Education, Culture, Sports, Science, and Technology (MEXT) guidelines, some important issues that arise are related to the difficulty of making valid, reliable, and feasible "integrated" performance tests (speaking and writing assessment). There have been various teaching methods which combine input and output in the fields of TESOL and second language acquisition, but little attention has been paid to assessing achievement in such integrated-skills courses. For instance, when a student does a presentation based on some reading or listening activities, what aspects of the presentation should be treated as the exhibition of reading or listening ability and how should his or her speaking ability be assessed? What kind of assessment tool fits the evaluation of learning outcomes of such integrated teaching? Facing this new requirement, high school teachers seem to be either at a loss or simply follow traditional ways by assessing different skills separately. Also, university teachers who should guide or train prospective high school teachers are not well equipped to address this demand. In fact, the new direction requires all English teachers to acquire more assessment literacy in addition to teaching skills for this recently emerged need.

2 Purpose of the study

The purpose of our research is to make a feasible proposition for dealing with fundamental issues teachers face when creating and implementing integrated-skills performance tests, and in this study the focus was placed on speaking tests. Compared to writing tests, the performance of which is left as an artifact, speaking evaluation has always been more difficult because of its elusive nature and the many factors involved in its administration. The issues will be examined from a number of perspectives in the following sections: (a) a literature review on the issues related to making effective speaking tests, (b) an analysis of the speaking section of TOEFL iBT® and IELTS®, focusing on their rubrics, constructs, and methods of assessment, (c) an analysis of a needs survey for in-service teachers (see Appendix A), (d) an analysis of teacher qualification examinations for English teachers in Japan, and (e) observations and participant responses collected at prospective teacher training sessions on assessment literacy conducted by the Testing SIG of the Japan Association of College English Teachers (JACET).

2.1 Literature review

Taylor (2009) states that the term *assessment literacy* refers to the knowledge essential to the assessment process. This could also include the level of skill, knowledge and understanding of assessment principles needed by various stakeholders such as government officials, policy planners, the media, and the general public.

Inbar-Lourie (2008) claims that learning and assessment are intertwined. It is important to understand the vital role of assessment in the learning cycle and also the dual roles of a teacher as an instructor and assessor of progress. She also suggests that the ability to do this successfully requires a teacher to be 'assessment-literate.' In other words, teachers should have the ability to pose and answer critical questions about the purpose of assessment, the tools being used, testing conditions, and future implications.

Recently, the feasibility of integrated-skills speaking tests in Japan has gained the attention of researchers and practitioners. In particular, the sections related to speaking tests of major studies on assessment have been reviewed with the Japanese EFL context in mind. For example, Ito, Nakamura, Kimura, Tsuchihira, Murray, Okada, and Matsumoto (2011) examined English teacher education textbooks published in Japan, particularly from a testing perspective, and identified what testing-related topics and contents are generally covered in teacher education in Japan.

In Murray, Ito, Kimura, Matsumoto, Nakamura, and Okada (2011), textbooks written in English were examined in comparison with 10 domestically published textbooks to evaluate differences in testing-related concepts that are covered. Both studies also investigated the current situation regarding teacher training on testing and assessment (including that of integrated-skills speaking tests) through questionnaires. The findings from the textbook analysis and the survey results provided information about how testing concepts (e.g. creating rubrics for integrated-skills speaking tests) should be covered in teacher-training programs and textbooks to help teachers establish a meaningful connection between teaching and evaluation in terms of assessment literacy.

In addition to these studies, Taylor (2011) and Plakans (2011) provide an invaluable theoretical perspective and insight into the various ways to validate the constructs and rubrics of integrated-skills tests.

Taylor (2011) introduces an interactionalist perspective that views the construct of speaking as the interactions among underlying cognitive ability, the context of use, and the process of scoring. She claims that, in general, at the heart of any language testing activity, we can conceive a triangular relationship between the following critical components: a) the test taker's cognitive abilities, b) the task and context, and c) the scoring process.

Taylor further maintains that these three components, which are related to cognitive validity, context validity, and scoring validity respectively, offer an important perspective on construct validity, which has both theoretical and direct practical relevance for test developers and producers. In addition, the interactions between, and especially within, these aspects of validity may eventually offer further insight into a more accurate definition of the different levels of task difficulty (Taylor, 2011).

On the other hand, Plakans (2011) describes the constructs of integrated assessment, the difference between integrated and individual skill tests, and recommendations about integrated assessment. In addition, she claims that there are benefits of implementing integrated assessment, such as task authenticity and increased motivation and inspiration for students, despite some challenges pertaining to topic effect, task effect, and rater reliability.

2.2 Analysis of existing tests

The two most widely used examples of integrated speaking tests are probably the speaking sections of IELTS® and TOEFL iBT®. However, there are big differences in the way these two tests integrate multiple skills.

The IELTS® speaking section is integrated in its administration as test-takers interact with the interviewer in two conversational parts (so listening and speaking are integrated) and one monologue part which requires test-takers to read the prompt and take notes before speaking (thus reading, writing and speaking are integrated). However, its rubric mostly measures speaking abilities. Though the strength of the IELTS®

speaking section is its use of authentic, natural conversational situations, the assessment is done on a single-skill basis.

The TOEFL iBT® is a computer-based test, and its speaking section includes both independent and integrated items. The two independent items require reading and responding to short prompts, but the four integrated items are truly integrated in nature: two items require summaries of the main points of a substantial reading and a listening passage while the other two items require the test-taker to express their opinion after listening to a passage and a discussion. Thus, the rubric for the independent items mostly assesses speaking abilities, but the one for the integrated items encompasses reading and listening abilities. The test-taker's speech should include "relevant information" from what they have read and listened to, and occasional mitigation that happens in the speech by recalling important ideas is accepted. For instance, even the highest-score (Score 4) rubric states "Pace may vary at time as speakers recall information", and "It includes appropriate details though it may have minor errors or minor omissions" (Educational Testing Service, 2012). The strength of the TOEFL iBT® is its truly-integrated assessment as well as administration, but it has met criticism because it is somewhat artificial given the fact that it is a computer-based test requiring test-takers to produce only monologues and talk to a computer. Also, integrated assessment needs a lot more training on the part of raters.

From the analysis of the speaking sections of these two exemplary tests, there are valuable lessons in the different ways of combining skills in creating integrated-skills tests and producing valid and reliable rubrics. When dealing with many students in a regular high school or university classroom, a recording device or computer may be necessary. More importantly, it is very difficult to come up with a valid, reliable rubric for any integrated-skills speaking test which fits the purposes of different learning activities.

2.3 Needs analysis for in-service teachers

Although the new MEXT directive requires teachers to integrate the teaching and assessment of multiple language skills in their classes, it is unclear how prepared classroom teachers are to make this shift from the current single-skill approach to an integrated-skills one. In order to better understand the current reality of classroom teachers, a needs analysis of in-service junior high school and high school teachers was conducted in the spring of 2012. The focus of the needs analysis was on their experiences with and plans for using integrated-skills speaking tests, along with the problems and concerns they have. The results clarified how in-service teachers tackle with making integrated-skills tests and evaluating multiple skills.

To be specific, the needs analysis study investigates the current situation of integrated teaching and the teachers' needs for integrated-skills tests, especially for evaluations on speaking. For this purpose, a questionnaire (Appendix A) was constructed which consisted of two main parts: (1) questions about "integrated teaching" in terms of MEXT's official guidelines, and (2) questions on "integrated-skills evaluations on speaking". For some questions, the respondents were asked to choose their answers from given multiple-choices, while other questions requested them to write short comments. A total of 54 teachers responded, with approximately 70% of them being employed in the public sector and the remaining 30% in the private sector. About 40% of the respondents were in their 20s, 25% in their 30s, 25% in their 40s, and about 10% were over 50 years old.

The first section of the questionnaire contains statements about integrated teaching in terms of MEXT's official guidelines. On a four-point scale, a response ranged from strongly disagree (1) to strongly agree (4). Table 1 shows the distribution of responses to each statement. To the first statement, almost all of the respondents felt that integrated teaching would improve English teaching. On the contrary, in response to the second statement, more than 70% of the respondents reported that integrated teaching had not actually been incorporated into real classroom settings. Regarding the third statement, the opinions were divided. A half of the respondents implemented integrated teaching; the other half did not. However, in response to the fourth statement, more than 80% of the respondents showed positive attitudes toward the future of integrated teaching in classes. It should be noted that the majority of such positive responses came from the teachers in their 20s and 30s. These results clearly indicate that in-service teachers place an importance on integrated teaching, but it seems hard for them to put it into practice.

Table 1: Integrated Teaching in Accordance to MEXT's Official Guidelines

Statement	Strongly agree	Agree	Disagree	Strongly disagree
1. Integrated teaching will improve English Teaching.	26	25	2	0
2. Integrated teaching is well recognized in the	2	11	31	10

field of English language teaching.				
3. At the present, I am teaching English integratively.	13	20	17	4
4. From now on, I am going to teach English integratively.	25	22	7	0

Note: n = 54.

Then, using open-ended questions, the respondents were asked to explain their responses to the four statements about integrated teaching with MEXT's official guidelines in mind. Table 2 shows the reasons why the respondents chose positive answers (agree or strongly agree) to the statement that integrated teaching will improve English teaching. The majority of the respondents realize that the complex interaction of four skills makes up language use or communicative ability. The second most common reason "It is effective to teach the interaction between input and output" may partly overlap with the first reason. The third most common reason "Integrated teaching will develop learners' ability from language knowledge to language use" all came from the respondents employed in public junior high schools. This reason coincides with one of the aims of MEXT's Course of Study.

Table 2: Reasons Why Integrated Teaching Will Improve English Teaching

Reason	Number	Percentage
The four skills are inseparable in communication.	28	57.0
It is effective to teach the interaction between input and output.	9	16.0
Integrated teaching will develop learners' ability from language knowledge to language use.	5	13.0
There is high demand for production skills.	4	8.0
Integrated teaching may lead to improvement, but not sure.	3	6.0

Note: These reasons were given by the respondents who agreed or strongly agreed with Statement 1.

The respondents were also asked to give the reasons for their response to the statement that integrated teaching is well recognized. For this statement, 70% responded negatively (disagree or strongly disagree). Their reasons are shown in Table 3. Although the most common reason was related to the infeasibility due to the influence of the Grammar Translation Method which is still predominant on entrance examinations, the other three reasons were highly related to the fact that teachers are hesitant about the recent MEXT initiatives to start integrated teaching, such as "Teachers don't know how to do integrated teaching" or "Tests and texts do not reflect integrated teaching." These responses obviously come from the lack of practical instructions to guide teachers in the new direction.

Table 3: Reasons Why Integrated Teaching Is Not Well Recognized

Reason	Number	Percentage
Dominance of the Grammar Translation Method for entrance exams.	10	28.0
Teachers do not know how to do integrated teaching.	10	28.0
Tests and textbooks do not reflect integrated teaching.	6	17.0
Courses are still divided by skills.	6	16.0
Other	4	11.0

Note: These reasons were given by the respondents who disagreed or strongly disagreed with Statement 2.

Those who responded positively (agree or strongly agree) to the statement that they are teaching integratively or intend to teach integratively in the future were then asked to select which skills they usually combine or will combine. Figure 1 shows the number of responses for each combination of skills. The top four combinations preferred were related to progressing from a receptive skill to a productive skill.

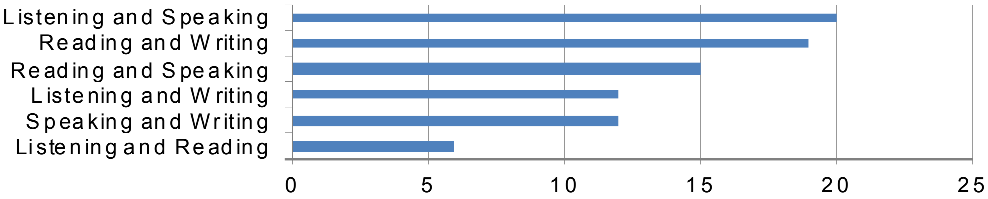

Figure 1: Combinations of skills

In a follow-up open-ended question, many respondents expressed the reasons for the different combinations of skills preferred. Table 4 shows that for the four most widely used combinations, the importance of the connecting input and output was the mostly cited reason. A high level of practicality in the classrooms was the second most common reason, but this was heavily outweighed by the importance of the input-output process.

Table 4: Reasons for the Selected Skills to be Integrated

Combination	Reasons	Percentage
Listening and Speaking	Importance of input-output connection	67.0
	High level of practicality	27.0
	Easier to motivate students	6.0
Reading and Writing	Importance of input-output connection	66.0
	Leads to effective acquisition of syntax and vocabulary	17.0
	High level of practicality	17.0
Reading and Speaking	Importance of input-output connection	73.0
	Close connection between reading aloud and speaking	18.0
	Effective for oral instruction	9.0
Listening and Writing	Importance of input-output connection	60.0
	High level of practicality	23.0
	Importance of dictation	7.0
	Other	10.0

Based on the results from this section of the questionnaire, it can be said that the in-service teachers understand why integrated teaching can be effective. Most of them think that the four skills cannot be separated and interaction between input and output is an integral part of communicative activities. However, in reality it seems difficult for them to start incorporating integrated teaching without clear guidelines and examples that are based on the theory of testing and assessment. The teachers are seeking model assessment tools which fit the evaluation of learning outcomes of such integrated teaching.

In the second section of the questionnaire, the focus was to have the respondents evaluate integrated-skills speaking tests in terms of real school environments. They were asked to answer which skill(s) should be combined with speaking in their specific teaching contexts. As Figure 2 shows, the majority of the respondents thought that listening and reading could be combined with speaking, 37% and 35% respectively. Writing followed with 15% and no response with 13%. This result may seem obvious, considering that most of the respondents are aware of the importance of the input-output process and the nature and relationship of these skills.

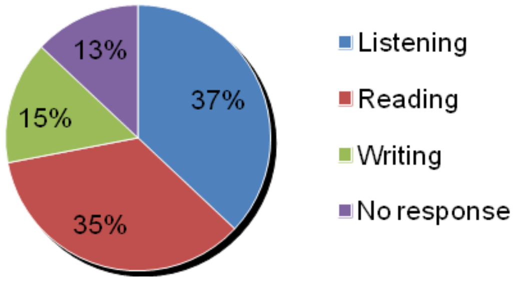

Figure 2: Skills to be combined with speaking

Table 5 presents an overview of the respondents' responses to the open-ended question "What kind of tests and evaluation should be conducted for integrated-skills speaking teaching?" Although there were some variations among the responses, about 80% of the respondents claimed that Question and Answer, fill-in-the-blanks, and summary type tests should be conducted based on the contents of reading and/or listening materials.

Table 5: Preferred Forms of Testing for Integrated-skills Speaking Instruction

Combination	Forms of testing and evaluation	Percentage
Listening and Speaking	Question and Answer, Fill-in-the-blanks, and/or summary type tests based on what students have listened to	75.0
	Pair work speaking test	10.0
	Interview	10.0
	Dictation and Recitation	5.0
Reading and Speaking	Question and Answer, Fill-in-the-blanks, and/or summary type tests based on what students have read	82.0
	Opinion statement	12.0
	Recitation	6.0

Regarding the evaluation points for integrated-skills speaking tests, Table 6 shows that the majority of the responses can be grouped into the following six categories: Pronunciation / Intonation (24%), Listening / Reading comprehension (19%), Accuracy (including sentence structure and grammar) (18%), Vocabulary / Word choice (13%), Logic / Coherence in interaction (13%), and Positive attitude toward communication (9%).

Table 6: Evaluation Points for Integrated-skills Speaking Tests

Category	Percentage
Pronunciation / Intonation	24.0
Listening / Reading comprehension	19.0
Accuracy	18.0
Vocabulary / Word choice	13.0
Logic / Coherence in interaction	13.0
Positive attitude toward communication	9.0
Other	2.0
No response	2.0

In addition, a small number of the respondents (2%) left this section blank. Another 2% indicated that integrated-skills evaluation can only be applied to a limited number of students with comments such as "Evaluating recitation is important, but it is relatively useful for only basic level students," and "Only advanced level students can be properly tested on their opinion statements."

It is a little surprising that Pronunciation / Intonation was higher than content-based evaluation points in the integrated-skills tests. However, almost all the categorized points overlap with those that appear in the rubrics of TOEFL iBT® and IELTS® integrated-skills speaking tests. That is, the respondents are in some way conscious of the appropriate evaluation categories for integrated-skills speaking tests. "Positive attitude for communication” is not included in the rubrics of TOEFL® and IELTS®, but this might come from the Course of Study as stipulated by MEXT, which encourages teachers to evaluate students’ attitudes as one of the evaluation categories.

Table 7 summarizes the major responses to an open-ended question about problems or concerns with integrated-skills evaluations in real school settings. As had been predicted, the responses to this question fell into more than one category, as they were written from different viewpoints. The responses can be divided into the following three categories: those from the teacher’s viewpoint, from the students’ viewpoint, and from the administration’s viewpoint.

Table 7: Problems or Concerns about Integrated-skills Evaluations

Perspective	Problems or concerns	Percentage
From teacher’s Viewpoint	Heavy work load on teachers (because they do not know much about evaluation models such as criteria or rating scales)	28.0
	Inter-rater reliability: teachers wonder if they can establish and share common evaluation standards	23.0
	Teachers do not know how to evaluate multiple skills in one test.	17.0
From students’ Viewpoint	Feedback should be given to each skill rather than integrated-skills.	14.0
	Evaluation methods should be notified in advance.	6.0
	Influence of prompt and topic of the tests	3.0
From administration’s viewpoint	Environmental or instrumental difficulties with direct speaking test	5.0
	Much time and cost needed for rater training	3.0
	Other	1.0

From the teacher’s viewpoint, “Setting standards or rating scales puts a heavy burden on teachers” was the biggest concern with 28%. The next biggest concern was “It is difficult to take inter-rater reliability into consideration” with 23%. The final concern with 17% was “It is difficult for a teacher to measure multiple skills in one test”.

From the viewpoint of the students, the strongest opinion was “Feedback should be given to each skill rather than integrated-skills” with 14%. There were also opinions concerning test context validity such as “Evaluation method should be announced to students in advance” (6%) and “Topic and prompt have influences on students’ evaluations” (3%).

A small number of responses were from the viewpoint of administration. Examples of responses relate to “environmental problems in direct integrative speaking test”, “difficulty of developing can-do statements for evaluation”, and “time and cost associated with rater training”.

In summary, the results of the survey suggest that many junior high and high school teachers think that integrated teaching is an effective method for improving English classes, and as to integrated-skills speaking tests and evaluations, they give favorable comments on input-output interactions. On the other hand, they are facing practical problems and concerns as a result of complex school contexts. However, the most significant finding is that while in-service teachers are anxious about implementing a new integrated-skills teaching method, they are generally positive about it, seeking helpful practical models which include criteria and rating scales for testing and evaluating speaking as integrated with another skill. Therefore, it is essential to design and offer proper, practical assessment models for integrated-skills speaking tests so that teachers can smoothly and confidently put integrated teaching into practice in their classrooms.

2.4 Investigation of the English Teacher’s Employment Examinations (E-TEEs)

The English Teacher’s Employment Examinations (E-TEEs) are compulsory certification examinations for prospective English teachers in Japan. E-TEEs are developed by local educational boards and administered by 47 prefectures and 17 cities in July every year (Ministry of Education, 2011).

To better understand what knowledge newly certified teachers must possess, a thorough analysis of E-TEEs that were used in recent years is essential. To be specific, the focus of the inquiry was on question types and what knowledge is actually tested on assessment-related questions. A detailed analysis of these questions may have implications for aspects that are lacking in prospective English teachers in terms of teacher assessment literacy as explained previously.

Two separate analyses were conducted on all the past questions which were published in annual workbooks covering E-TEEs of the 47 prefectures and 17 cities published by Kyodo Shuppan in 2006 and 2010. Although both of these analyses were done in a similar manner, the approach used in each analysis was slightly different. For the analysis of the 2006 workbooks, every question was examined. However, in the case of the 2010 analysis, the focus was on larger sections (Daimon) rather than on individual questions. One of the main reasons for this was that the 2006 analysis was too time-consuming and the approach for the 2010 analysis was more time-effective.

For the 2006 analysis, all questions (2,096) were categorized into three types: 1) proficiency-based questions (PBQ), 2) teaching-based questions (TBQ) and 3) other (applied linguistics related ones). These three categories were defined by the researchers after examining each of the 2,096 test items. PBQs refer to items assessing general English ability, including reading, listening, writing, vocabulary, and grammar. These questions were similar to the items in TOEIC® and STEP EIKEN. TBQs include questions about the Course of Study published by the Ministry of Education and dealt with terminology frequently used in teaching methods and in making teaching plans. Also included are items related to applied linguistics such as second language acquisition, learning theories, and motivation theories. Testing-related questions, the main focus of the analyses, were also included in this category.

The majority of the questions on the E-TEEs, 90% or 1,886 of the test items, were PBQs. TBQs (9%) and other question types (1%) accounted for the remaining test items. Of the 1,886 PBQs, 27% of them were categorized as “integrative reading” items (a mixture of various types of questions using reading passages), followed by listening (22%), reading (18%), vocabulary and idiomatic phrases (14%), and grammar (8%). However, it is surprising that only 4 out of 2,096 questions (approximately 0.2%) were testing-related questions. These four questions were used in the examinations of three prefectures. This suggests that assessment literacy for English teachers is devalued.

The 2010 analysis shows similar results to those of 2006, although as previously stated, the method of counting the number of questions was different. Approximately 80% (442 sections) were categorized as PBQ, followed by 18% (97 sections) for TBQ. However, the percentage of assessment-related sections was only 0.7 % (4 sections), which means that assessment literacy for English teachers continues to be underappreciated.

Despite a close relationship between teaching and assessment, the results from the two analyses imply that assessment literacy is not reflected in the expected requirements of “English teachers” in Japan. In conclusion, various types of E-TEEs are developed by the local boards of education in cities and prefectures, showing that each E-TEE may take different approaches to assessing prospective English teachers’ abilities. It is important to note that PBQs far outnumbered TBQs, which seems to indicate that “English ability” is far more valued than the knowledge of teaching English and assessment-related concepts.

2.5 Participant responses to our workshops

In order to have a more comprehensive perspective of the impact of integrated-skills teaching and assessment, the opinions and concerns of future English teachers must also be taken into account. This perspective is useful because it shows the extent to which teacher training programs are evolving and it shows the direction in which teacher training programs should evolve in response to MEXT’s new guidelines and how prepared the next generation of teachers are in facing such demands.

On an annual basis, the Japan Association of College English Teachers’ Testing Special Interest Group (JACET Testing SIG) has been holding one-day training sessions on assessment literacy for prospective teachers. For the 2012 workshop, the topic of the training sessions was on integrated-skills assessment, with a focus on speaking tests. Each of the training sessions consisted of five parts:

a) Lecture (General testing concepts)
b) Teaching (Lesson plan)
c) Test Making (Hands-on experience)
d) Test Data Analysis (Hands-on experience)
e) Teaching (Application in the classroom)

These five parts were chosen based on the premise that classroom teaching has significant interactions with assessment (Figure 3). In other words, formative assessment, which is carried out during classroom teaching, is analyzed and discussed so that the results can be used to provide the opportunities to review and improve classroom practices.

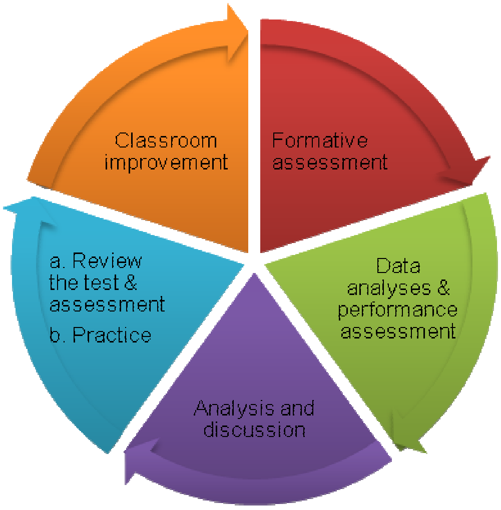

Figure 3: Relationship between classroom teaching and assessment

Every year, upon completion of the one-day workshops, the participants completed anonymous questionnaires. The questionnaires were distributed at the end of the sessions to the participants, and most participants spent about 30 minutes completing them. The questions start with impressions of each section of the workshop including difficulties, time allocation and the appropriateness of the materials used, and end with asking about the importance of this kind of workshops. Generally, the responses of the participants coincided with the opinions of the in-service teachers explained in the previous section and the general observations made by the workshop facilitators. The participants felt the need to learn more about test making and analysis methods since formative assessment is a part of everyday practice. The hands-on practice provided in the workshop in those areas was most favorably received. They even hoped to spend more time on making tests and to have more advice from the facilitators during the process. Also, they expressed a strong interest in learning about performance assessment, the possibilities of which were not explored much in various teacher-training courses they had previously taken. For the data analysis session, some participants wanted to see and experience the real processes of statistical analyses using computers.

3 Conclusion

The present study aimed to make a proposition for dealing with fundamental issues teachers face when integrated speaking tests are created and implemented. The needs analysis of in-service teachers on integrated-skills teaching and testing, the analysis of the E-TEEs questions, and the feedback from the annual workshop participants revealed that while assessment literacy seems to be undervalued in the E-TEEs, there is a need among in-service teachers, at least those who participated in our study, for assessment models, examples, and tools that can be applied to their classroom teaching. This suggests that it is indispensable for them to acquire higher assessment literacy when making integrated-skills tests. In our annual workshops, critiquing the existing test items and hands-on experience of test making and analysis has proven very effective in raising the participants' assessment literacy. Thus by attending training sessions in assessment for pre-service and in-service teachers, they can gain more knowledge and experience, especially on how to combine different constructs and how to validate tests statistically and qualitatively. However, as the type of integrated speaking test item depends on the purpose and nature of assessment, for example, placement or diagnostic feedback, different models should be established and demonstrated in teacher training. We also have to consider "context and cognitive validity" (Field, 2011; Galaczi & ffrench, 2011) as well as "scoring, consequential and criterion-related validity" (Hawkey, 2011; Khalifa & Salamoura, 2011; Taylor & Galaczi, 2011), prioritizing the more important ones based on the purpose and aim of assessment.

References

Field, J. (2011). Cognitive validity, in L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing 30, Cambridge: Cambridge University Press, 65-111.

Galaczi, E., & ffrench, A. (2011). Context validity, in L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing 30, Cambridge: Cambridge University Press, 112-170.

Hawkey, R. (2011). Consequential validity, in L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing 30, Cambridge: Cambridge University Press, 234-258.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3)*,* 385-402.

Ito, Y., Nakamura, Y., Kimura, K., Tsuchihira, T., Murray, A., Okada, A., & Matsumoto, K. (2011). An analysis of English teacher education textbooks published in Japan from a testing perspective. *JACET-Kanto Journal, 7,* 27-33.

Khalifa, H., & Salamoura, A. (2011). Criterion-related validity, in L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing 30, Cambridge: Cambridge University Press, 259-292.

Murray, A., Ito, Y., Kimura, K., Matsumoto, K., Nakamura, Y., & Okada, A. (2011). Current trends in language teaching education in Japan, in A. Steward (Ed.), *JALT2010 Conference Proceedings*, 151-162.

Educational Testing Service (2012). Official guide to the TOEFL test with CD-ROM (Official Guide to the TOEFL iBT®). Educational Testing Service. McGraw-Hill.

Plakans, L. (n.d.). *Integrated Assessment*. Retrieved from http://languagetesting.info/video/main.html

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29,* 21-36.

Taylor, L., & Galaczi, E. (2011). Scoring validity, in L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing 30, Cambridge: Cambridge University Press, 171-233.

The Ministry of Education, Culture, Sports, Science & Technology in Japan (2011). *Heisei 22 nendo kouritsu gakkou kyouin saiyoushikenno jisshiyoukounitsuite* [Survey of 2010 public teacher employment examinations]. Retrieved from http://www.mext.go.jp/a_menu/shotou/senkou/1300242.htm

Appendices

Appendix A. Questionnaire on Integrative Speaking Tests (translated version)

The JACET Testing SIG has been studying evaluations and exams in English language teaching. Now, “teaching the four skills integratively” is an important keyword in MEXT’s official guidelines for school teaching. However, how to teach integratively and which skills to be combined are ongoing topics for discussion, and so are integrative evaluations. Therefore, we decided to do this questionnaire to gather opinions from many teachers in order to make evidence-based suggestions for “integrative evaluation”. Below is the questionnaire, and we are very grateful if you can kindly spare your time to complete it. Your privacy will be strictly protected, and the results will be presented at several conferences. We very much appreciate it if you could complete this by July 31st.

On “integrative teaching” in MEXT’s official guidelines for school teaching
(We define “integrative teaching” as teaching that combines more than one skill and enhances communication ability.)

Q1. “Integrative teaching” will improve English teaching.
strongly agree(4) agree(3) disagree(2) strongly disagree(1)

Q2. “Integrative teaching” is well recognized in the field of English language teaching.
strongly agree(4) agree(3) disagree(2) strongly disagree(1)

Q3. At the present, I am teaching English integratively.
strongly agree(4) agree(3) disagree(2) strongly disagree(1)

Q4. From now, I am going to teach English integratively.
strongly agree(4) agree(3) disagree(2) strongly disagree(1)

Please write the reasons for your answer to Q1.

Please write the reasons for your answer to Q2.

If you Agree or Strongly Agree in either Q3 or Q4, which skills do you or are you going to integrate in your teaching?
□ listening and writing
□ listening and reading
□ listening and speaking

□ speaking and reading
□ speaking and writing
□ reading and writing
□ others

Reason(s)

Integrative evaluations on speaking
We have been studying integrative evaluation where speaking skill is combined with other skills. Please write your ideas freely.

With which skills?　(　　　　　　) & speaking

What kind of test do you or will you use for this purpose?　Please describe it.

Please share with us your opinions on “integrative teaching” and “integrative evaluation”.

Personal information
School Location (Prefecture Only) ____________________
School Type
□ public
□ private

School Level
□ junior high school
□ high school
□ integrated junior and senior high school

Age
□ 20s
□ 30s
□ 40s
□ 50s
□ over 60