University Speaking Assessment as Collaborative and Self-Reflective Skill Development

James Herbach, University of Shizuoka, Japan
Kinsella Valies, University of Shizuoka, Japan

Abstract
The impetus of this research is deeply rooted within the immediate needs of students at the University of Shizuoka and their future professional goals, as suggested by student comments gathered from previous Language and Communication Research Center workshops and course evaluations. The aim of this action research is to improve confidence and communicative ability by implementing classroom-based speaking assessment tools that provide students with opportunities for ongoing speaking advancements and feedback. Research questions are: 1) To what extent can speaking assessments help students better express themselves in English? 2) To what extent can teaching towards this assessment help improve self-efficacy and confidence in students' ability? 3)To what extent can the speaking assessment tool aid in advancing students' critical thinking skills? This study took a mixed method approach where data from both student grades and survey results were collected and analyzed. Results showed that student confidence in their ability to express themselves in English increased significantly. The median grade differential between finals and midterms was approximately 6%. Future research will focus on the design and implementation of similar speaking assessment tools for first graders across all departments.

Keywords: EFL, Speaking, Assessment, University, Japan

iafor

The International Academic Forum
www.iafor.org

Introduction

The motivation for this project emerged from the recent interest in the idea of supporting the speaking skills of learners in the field of English language education as well as from the management of the Language and Communication Research Center (LCRC) at the University of Shizuoka. As Reinbold (2017) states, "in the current world, many non-native speakers are using English as a tool for communication (p.246). Reading comprehension and essay writing skills have traditionally been assessed so as to provide both students and instructors with concrete data showing students' learning progress. The objective of this project was to, first, afford students and instructors with tangible evidence of speaking development on three specific categories of skills and abilities and, second, evaluate speaking assessment tools created inside the department. Both objectives support students who in the future may be faced with authentic situations in which they may have to interact with non-Japanese speakers and express themselves effectively. The three categories which were targeted for this project included Reading a text aloud, Describing a picture and Giving an opinion.

The decision to target the three aforementioned speaking categories was based upon the large class sizes involved, up to 35 students per class, as well as the amount of time required to assess students' speaking abilities in a face-to-face environment. In order to provide results of speaking progress both a mid-term and a final speaking test were scheduled. The project initially targeted one group of second year pharmacy students which four members of the LCRC all taught during the same day and period. Focusing on this particular group of students provided instructors with the opportunity to assess each other's classes so as to remain as objective in the assessment process as possible.

Along with the rational for initiating this project, the methodology, the steps to production and the implementation of the assessments are also addressed. In addition, the results of the assessments, the limitations of the project and future implementations are further investigated. As the project is now an ongoing element in the LCRC curriculum, many additional revisions have been and continue to be made.

The research questions that guided our investigation were:
1)	To what extent can speaking assessments help students better express themselves in English?
2)	To what extent can teaching towards this assessment help improve self-efficacy and confidence in students' ability?
3)	To what extent can the speaking assessment tool aid in advancing students' critical thinking?

Methodology

The researchers probed tertiary foreign language students of English regarding their experiences taking an English Communication course and interpreted forthcoming answers making connections to the RQs that guide this study.

The paradigm at the base of this study is an interpretivist one. Additionally, this study was both classroom-based and collaborative as all the data was collected by Language and Communication Research Center's (LCRC) instructors teaching in classrooms and

working together to create and implement assessment tools and distribute the surveys to the participants.

The methodology supporting the project is both quantitative and qualitative in nature. Students' test scores were compiled to provide quantitative results of their speaking abilities, while pre- and post-assessment surveys were implemented to offer qualitative results for both the overall course satisfaction of students and the speaking assessment in particular.

Participants

Based on the similar scheduling of both day and time for English Communication classes, one group of second year pharmacy students were targeted for the project. Four native members of the LCRC all taught one class of approximately 29 to 35 students representing a total of 146 participants. All of the classes used the Lecture Ready textbooks; however, students had been divided into two groups based on their TOEIC test scores the previous semester. The participants starting scores ranged between 500-600 points. The equivalent CEFR level was between A1 and A2, in other words, they can express themselves "in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters (Council of Europe, 2001,p. 74)." As a result, two of the groups were assigned with the Lecture Ready 1 textbook while the others used Lecture Ready 2. Clear and concise instructions were drafted for all three sections of the assessment and all members of the LCRC distributed the information to their students in an attempt to ensure consistency. Practice and preparation of students towards the assessments, however, were left to the discretion of the individual department members.

Steps to Production

Timeline

The impetus for the project was initiated by Dr. Atsushi Fujimori and was originally presented to the LCRC members approximately two weeks before the beginning of the semester. The target group was second-year pharmacy students who met once a week during a 15-week semester. Planning meetings began immediately in order to create a clear goal and an outline for the project's implementation. The three target areas of the assessment, reading a text aloud, describing a picture and giving an opinion were finalized. This was in part due to the fortuitous occurrence of professor Valies having previously used these categories for implementation in some of her classes. The categories have been inspired by their use in the TOEIC test, however they have been modified and revised so as to fit the semester schedule and the level of the students at the University of Shizuoka.

In addition, a rubric was drafted to provide students with specific details for what they would be required to achieve in each of the five scoring levels for all three sections of the assessment. Students would then be given a result out of a possible 15 total points. The rubric was also translated into Japanese. Copies of both the English and Japanese versions were then printed double-sided and distributed to students at the beginning of

the semester. In addition, the testing instructions were handed out at this time to all participants.

One result of the compacted timeline was that once the semester started, all of the materials for both the mid-term and final assessments had to be created in between preparing for and teaching classes. This led to a redefinition of the criteria for materials creation after the mid-term and before the implementation of the final assessment. Subsequently, some of the project limitations appeared at this time, and although addressed, will be illustrated in more detail later.

Step 1

Materials for the three target categories were created and evaluated by the four participating instructors. Each member collected six examples of texts, pictures and opinion questions. Materials were discussed during meetings and a selection was made as to which items would be used for midterms and finals on the one hand, and which would be used for in-class practice on the other. Material selection guidelines required instructors to match materials to themes found in the class textbook they used. As the speaking assessments took the shape of individual student-instructor interviews, previous students might have the opportunity to pass on information to the following test-takers. To mitigate this circumstance, six assessment packets were made for each assessment moment for the instructors to cycle through during the test period.

For the final speaking test instructors proctored each other's groups in a bid for impartiality. This added both a level of anxiety and motivation for students as they could not count on their usual teacher's familiarity with their mannerisms and word-use to make themselves understood. As a result, participants had to really attempt to express themselves clearly and succinctly.

Assessment Examples

The speaking assessments were initially designed to be implemented in a face to face environment. According to Qian, (2009), "such a test is believed to offer a much better test validity because, in a properly designed oral test in the direct testing mode, every effort is supposed to be made for the tasks to be as authentic as possible (p.115). As a result, all of the materials were printed on paper and laminated so that students could be presented with hard copies during their individually scheduled testing sessions. In addition, each instructor was provided with six different versions of each of the three parts of the assessments. In an attempt to prevent students from sharing details of the assessment materials with their classmates, a different version of the assessment materials was used after approximately five students were tested.

The printed and laminated assessment materials all included the same instructions in both English and Japanese to ensure that all students could clearly understand which skills and abilities they were being assessed on. The following are examples of part I and III of the assessments.

Part I (Read a Text Aloud)　文章の音読

In this part of the test, you will read the text aloud. You will have 45 seconds to prepare. Then you will have 45 seconds to read the text aloud.

このテストでは、下の文章を声に出して読んでもらいます。まず 45 秒間で黙読し準備をしてください。その後、45 秒間で声に出して読み上げてください。

Mass media refers to the distribution of impersonal information to a wide audience, such as what happens via television, newspapers, radio, and the Internet. With the average person spending over four hours a day in front of the TV, and children averaging even more screen time, media greatly influences social norms. People learn about objects of the material culture, like new technology and transportation options, as well as the nonmaterial culture. For example, what is true, what is important, and what is expected.

Part III (Express an Opinion)　考えを述べる

In this part of the test, you will give your opinion about a specific topic read by the instructor. Be sure to say as much as you can in the time allowed. You will have 15 seconds to prepare. Then you will have 60 seconds to speak.

このテストは、講師が読んだある話題について意見を述べてもらうものです。時間内にできるだけ多くのことを言うようにしてください。スピーチの前に 15 秒間で準備をしてください。その後、60 秒間でスピーチをしてください。

1.　　Do you think that the government should add a 10% tax on all fast-food restaurants to pay for national health costs?

Grading system

All students in the four separate second year pharmacy classes were evaluated using the same grading system. Both the mid-term and the final speaking tests were assigned a value of 20% of the final grade, composing a total value of 40% of the class grades for students' speaking abilities. Additional assessment included comprehension quizzes at 20%, weekly writing journals at 20% as well as out of class assignments and participation which were both valued at 10% respectively. The following table (1) illustrates the breakdown of grades for the class.

Assignments	% of Final Grade
Mid-term Speaking Test	**20%**
Comprehension Quiz	20%
Out of Class Assignment	10%
Final Exam Speaking Test	**20%**
Weekly Writing Journal	20%
Participation	10%
Total	**100%**

Table 1: Course grade breakdown

Rubric

Each section of the test required the students to attain a certain number of speaking goals to be awarded maximum points. This was done supplying the instructors with a bilingual English-Japanese grading rubric. It was modified for implementation and has been under continuous review resulting in several rewrites since its first implementation. Most adjustments were made to facilitate fairness and give students clearer insight into what aspects of their speaking they can work on to improve proficiency. It is a 15-point rubric (see appendix A), distributed to students as a double-sided handout. Each category is rated up to five points. Students were expected to range between 3-4 with 3 being the average. Students were assessed by their peers during in class practice as well as their instructors using the same rubric.

Step 2

In-class instructions were focused on the upcoming mid-term and final assessments and required students to produce specific types of speech acts. Information was given both in oral and written format. Class handouts described the test format providing some instructions for the test day.

Both the mid-term and the final speaking tests were proctored using the instructors' offices to afford students with privacy during the assessment process. All students were given a testing schedule which divided the entire assessment slate into 30-minute blocks. The result was to limit the number of students waiting to be assessed outside of instructors' offices to four at any one time. In addition, each instructor examined a colleague's class which were comprised of a different level of students in an attempt to retain objectivity in the assessment process.

Both the midterm and the final were scheduled on the same days and the same class periods for all four groups. The assessment schedule required the use of two class

periods as well as three lunch periods in order to accommodate the larger class sizes. Some additional time was also deemed necessary in the case of emergency absences. The result was that four classes out of a 15-class semester were needed for the implementation of both the mid-term and final speaking tests.

Clear and concise testing instructions were drafted and distributed to all students across the four second year pharmacy classes. In addition, the assessment schedules along with a detailed description of the testing environment were handed out to each student about 2 weeks in advance of the assessment dates.

Implementation

Instructors worked autonomously in class to practice and prepare students for the speaking assessments. However, all four instructors used the textbooks to target the three specific parts of the text to support students' knowledge and abilities and to maximize their individual results on the assessments. In order to implement the assessments, instructors used voice recorders and timers to collect data for the project.

Through the use of group work and task-based activities targeting the three categories of the speaking assessments, students were able to receive feedback from both their instructors as well as their classmates as to their knowledge and abilities while practicing in class. In addition, subsequent to the mid-term speaking test students were given their scores on a printed copy of the test rubric. The combination of in-class feedback and the mid-term speaking assessment feedback afforded students the opportunity to target areas in need of improvement for the final speaking test.

Instructors met post mid-term assessment to discuss the implementation of the assessment as well as the results for their respective classes. One outcome of these discussions was that all instructors agreed to clarify and modify the rubric in order to provide students with a more detailed range for their performances on the test. It was also agreed that the statements for the giving your opinion section of the assessment be trimmed down to a one sentence explanation. Revisions were then undertaken to modify and revise materials for the final speaking test. The use of the number of class periods necessary for the implementation of the speaking assessments and the ensuing loss of class instruction time was also a concern for all instructors.

Limitations

This study focused on the creation and implementation of speaking assessment materials to help students improve their speaking skills, increase their confidence and help them to think more critically. As for implementation of the assessments the team of instructors came across differences in grading between instructors. It was concluded that norming was required. Future norming endeavors would take the shape of having a secondary instructor evaluate group recordings of speaking tests.

Inconsistencies in text materials and instructions were found; instructions and materials printed on opposite sides, including Japanese instructions. These elements were corrected for the final assessment. Consistency in the presentation of materials also

became a concern. This concern was addressed by standardizing text-font, text-size, text-spacing, single-side / double-sided printing, paper-type.

The Describe a Picture-section also required standardization and guidelines were added for selecting photos. Important aspects to look for in an image were thought to be the amount of details in the background, ability to infer the environment, the minimum number of people in the photo, the amount of details regarding the people's appearance and actions.

One of the project limitations discovered post mid-term assessment was the quality of photos used for the describing a picture category of the test. After closer inspection and discussion, it was agreed that some of the photo examples did not provide students with enough detail to be able to describe logically and accurately. The concern was that students might unfairly receive a lower grade on this criterion of the assessment if asked to describe an inferior photo. One example is as follows;

Figure 1: Poor example   Figure 2: Good example

Students were asked to provide details of the photos in a logical order, ensuring that anyone who could not see the photo would be able to clearly understand the images through the words and phrases spoken by the students. Students were moreover instructed to provide details of each photo using three steps which included: describe the location/setting/place, describe the people/physical dimensions/clothing and describe their actions using verbs+ing.

Additionally, the test section related to critical thinking, Giving your Opinion, turned out to give students the most trouble during the midterm. Many were unable to answer the question because they either did not understand it and/or did not have enough thinking time. Instructors believed that this problem was caused by the length of the Opinion question. The questions were simplified so that students had enough time to read and understand the question and then consider their answers.

An example phrase from the midterm:

4. It is important to be mindful of gender-biased marketing laws, which are rightly getting stricter worldwide. As a result, do you think that all public services, such as train cars should be provided in a gender-neutral manner. Give reasons for your opinion.

A renegotiated phrase from the final:

> 2. Do you think that the health benefits of organic foods are worth the extra cost?

The *Do you think…?*-format proved easier to understand and yielded better results. Students seemed less stressed and fewer students ended up with a could-not-answer score of 1 on the rubric. In addition, instructors changed the timing parameters from 15 seconds thinking time to 30 seconds thinking time.

Conclusion

This project saw many beneficial additions to classroom work and adjustments to the assessment materials. Students who were part of this development reacted positively to the speaking test.

Results: Mid-term vs. Final

After the data from both the mid-term and final speaking assessments were collected and calculated it was clear that there was an improvement on students' total scores post final test. Nevertheless, some anomalies also became apparent. For the lower level group, the scores for the reading a text aloud category dropped slightly between the mid-term and the final tests. One factor resulting in this drop may have been the instructors increased familiarity with the assessment and the rubric culminating in a stricter assessment of each student. Another factor may have been that students were more focused on the describing a picture and giving an opinion sections of the test as they required more critical thinking processes.
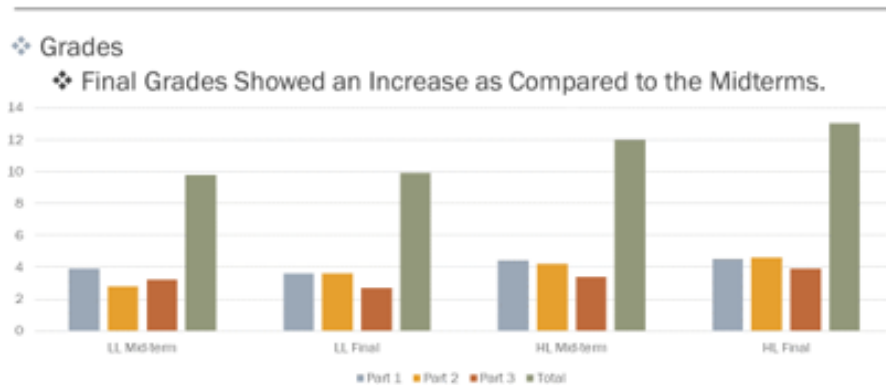

Figure 3: Grades Midterm vs. Final assessment

Results: Pre- and Post-course surveys

The findings show an increase in perceived value of speaking skills, and perceived opportunities to speak in class as well as confidence. The perceived value of speaking

skills increased by 7%. Student perception of the opportunities to speak English in-class went up by 10.2%. In addition, the most valued assessment section throughout the year was Giving your Opinion. One could conclude that students appreciate the chance to do targeted speaking practice and share their opinions and ideas (in English).

Insights into student confidence in-class were gleaned from the questions: "How would you rate your English-speaking ability?" and "How difficult is it to express yourself in English in class?" There was a 13.9% increase in student confidence in their ability to speak English. There was a total increase of 4.2% in student confidence in the ability to express themselves. In comparing the pre- and post-course responses one can observe that students who thought it *very difficult* decreased by 0.4%, *difficult* decreased by 4.2 %, *Somewhat difficult* increased by 2.9%. One could conclude that many just had not had a chance to practice targeted speaking in-class before and that they would assume that it was difficult. Having tried it, some found it was not as daunting, while others found their skills to be less than hoped for at this stage (See figures 4 and 5).

How difficult is it to express yourself in English in class?授業において英語で自分を表現するのは難しいですか。
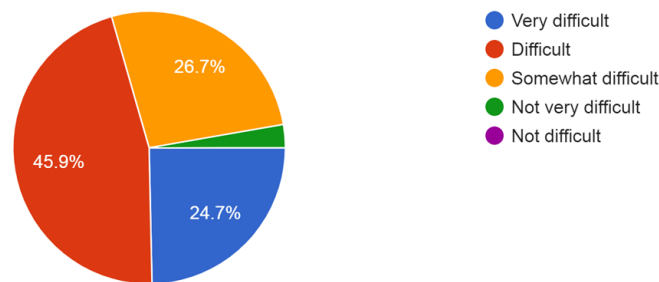146 responses

Figure 4: Pre-course survey- Difficulty expressing oneself in English.

How difficult is it to express yourself in English in class?授業において英語で自分を表現するのは難しいですか。
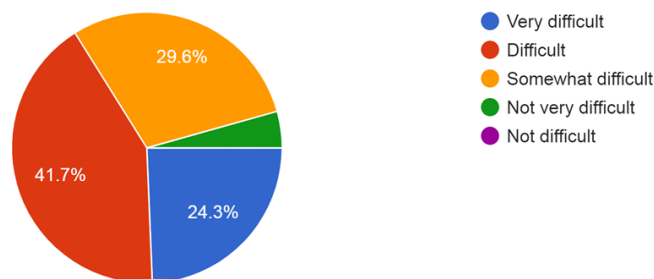115 responses

Figure 5: Post-course survey- Difficulty expressing oneself in English.

Discussion

Through the initiative of the instructors and the support of the LCRC the desire to implement a speaking assessment and ensure the equality of testing across all aspects of language learning has been successfully administered. This classroom-based project in its quantitative approach delivered positive outcomes through the acquisition of student scores on both a mid-term and a final speaking test. In addition, student survey responses were compiled which in a qualitative manner provided conclusive results in reference to student course satisfaction and an improved confidence in their speaking abilities.

Subsequent to the collection of all data, the calculation of student scores and the evaluation of student surveys the results of the project both answered and supported the original research questions. As to the question, to what extent can speaking assessments help students better express themselves in English, the data showed that students' total scores for the assessments improved from the mid-term test to the final exam by a value of 13% on average. In answer to the question, to what extend can teaching towards the assessment help students to improve their self-efficacy and confidence in their speaking abilities, the survey results detailed a 13.9% increase in their confidence levels. In addition, addressing the question of, to what extent can the speaking assessment tool aid in advancing students critical thinking skills, responses indicated that the most valued aspect of the assessment was the, Giving Your Opinion section of the test.

Future Implementations

This study is part of an ongoing speaking project in development within the LCRC curriculum. Future changes to second year materials include the removal of all proper nouns and slang in reading texts and opinion questions. With the next collegiate year in mind, there is a plan to expand the project to include the creation of speaking assessment tools for all first-year students across all departments. This time the assessment tools would be produced so that they could be implemented by any instructor using any textbook. As a result of the addition of first years to the project, the difficulty level of the second years materials would have to be adjusted to reflect student level progression from the previous year. Criteria in the rubric may require further adjustment as well to reflect additional speaking goals.

References

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press. Retrieved from
<http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf>

Reihbold, Lorraine (2017). English Pronunciation: Intelligibility, Comprehensibility, and Accentedness. *Faculty of Education, Hakuoh University,* Retrieved from the Internet, August 6, 2019.

Qian, David, D. (2009). Comparing Direct and Semi-Direct Modes for Speaking Assessment: Affective Effects on Test Takers. *Language Assessment Quarterly*, Retrieved August 6, 2019, from, https://doi.org/10.1080/15434300902800059

Pixabay (2019). *Couple* [Photo]. pixabay.com. Retrieved September 24, 2019 from, https://pixabay.com/

Stockvault (2019). *Granville Island* [Photo]. stockvault.net. Retrieved September 24, 2019 from, https://www.stockvault.net/

Stockvault (2019). *Cafe* [Photo]. stockvault.net. Retrieved September 24, 2019 from, https://www.stockvault.net/

Appendix A

	1	2	3	4	5
Reading Aloud	Speaks in isolated words. Consistent mispronunciation makes reading incomprehensible.	Speaks in broken phrases. Katakana English impedes understanding.	Katakana English doesn't impede understanding. Mispronunciations are frequent. Student is mostly monotone.	Speaks at sentence-level. Maintains rhythm. Some hesitation between sentences.	Clear speech, rhythm is optimal. There is little if any hesitation.
Describe a picture	Describes unimportant details w/o logical order. 7+grammar /vocab. mistakes.	Describes 1-2 aspects in logical order. 5-6 grammar/ vocab.mistakes.	Describes 3 aspects in logical order. 3-4 grammar/vocab. mistakes.	Describes 4 aspects in logical order incl. adj./v./prep. Mentions locations, persons.	Describes 5+ aspects in logical order incl. adj./v./prep. Mentions locations, persons.
Give your opinion	Unable to give an opinion/opinion is irrelevant to the question/answer is incoherent.	Gives relevant, coherent opinion. Uses text vocabulary.	Gives opinion, 1 reason, no examples. Uses text vocabulary.	Gives opinion, 2 reasons or 1 example. Uses text vocabulary.	Gives opinion, 3 reasons, 1-2 examples. Uses text vocabulary.
TOTAL					/15

Appendix B

	1	2	3	4	5
Reading Aloud	・単語で話す ・音読は発音の誤りが多く理解できない	・誤ったフレーズで話す ・カタカナ英語 のため理解しにくい	・カタカナ英語だが理解はできる ・時々発音の誤りがある ・音読が主に単調である	・文レベルで話すことができリズムも維持できる。 ・文章間で少しためらうことがある	・明確なスピーチ、リズミカルに話す ・ためらいはほとんどない
Describe a picture	・論理的な順序で説明できない ・文法/語彙の誤り7以上	・論理的な順序で特徴を1〜2点程度説明できる ・文法/語彙の誤り5〜6程度	・論理的な順序で特徴を3点程度説明できる ・文法/語彙の誤り3〜4程度	・形容詞・動詞・前置詞を使い、場所や人物について論理的な順序で4点程度説明できる	・形容詞・動詞・前置詞を使い、場所や人物について論理的な順序で5点以上説明できる
Give your opinion	・自分の意見を伝えることが出来ない ・質問に対する意見が不適切、答えがちぐはぐである	・適切な意見を分かりやすく伝える ・テキストで学んだ語彙を使うことが出来る	・自分の意見を理由付けで伝えることが出来る ・テキストで学んだ語彙を使うことが出来る	・自分の意見を理由や例を挙げて伝えることが出来る。 ・テキストで学んだ語彙を使うことが出来る	・自分の意見を複数の理由や例を挙げて伝えることが出来る。・テキストで学んだ語彙を使うことが出来る
TOTAL					/15