FROM TRADITIONAL TO PERFORMANCE-BASED ASSESSMENT: STUDENT PERCEPTIONS AND IMPLICATIONS OF A NEW ACHIEVEMENT ASSESSMENT IN INTRODUCTORY JAPANESE COURSES

Namiko Uchida

Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_theses

Recommended Citation

https://docs.lib.purdue.edu/open_access_theses/1081

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
This is to certify that the thesis/dissertation prepared

By Namiko Uchida

Entitled

FROM TRADITIONAL TO PERFORMANCE-BASED ASSESSMENT: STUDENT PERCEPTIONS AND IMPLICATIONS OF A NEW ACHIEVEMENT ASSESSMENT IN INTRODUCTORY JAPANESE COURSES

For the degree of Master of Arts

Is approved by the final examining committee:

Atsushi Fukada
Chair

April Ginther
Co-chair

Mariko Moroishi Wei
Co-chair

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University’s “Policy of Integrity in Research” and the use of copyright material.

Approved by Major Professor(s): Atsushi Fukada

Approved by: Atsushi Fukada
Head of the Departmental Graduate Program 12/3/2015 Date
FROM TRADITIONAL TO PERFORMANCE-BASED ASSESSMENT: STUDENT PERCEPTIONS AND IMPLICATIONS OF A NEW ACHIEVEMENT ASSESSMENT IN INTRODUCTORY JAPANESE COURSES

A Thesis
Submitted to the Faculty
of
Purdue University
by
Namiko Uchida

In Partial Fulfillment of the Requirements for the Degree of
Master of Arts

December 2015
Purdue University
West Lafayette, Indiana
ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to the chair of my committee member and my advisor, Professor Atsushi Fukada for giving me the opportunity to write this thesis and allowing me to do research using his *Speak Everywhere* software. Without his continuous support, wisdom, patience, and guidance, this research would not have been possible. I sincerely thank him for always being there when I needed help, and for being a great advisor. I would also like to extend my appreciation to my committee members, Professor April Ginther for her knowledgeable suggestions and feedback regarding language assessment, and Mariko Moroishi Wei for her insightful feedback and constant support. I also thank the Japanese 101 and 102 instructors and fellow graduate assistants for cooperating with this research during the period of this study. Without their help, this research would not have been completed. Likewise, I thank every student who took part in this study.

Again, I would like to thank Professor Atsushi Fukada for allowing me to use screen captures from his *Speak Everywhere* software. I also would like to thank him and Ryota Kondo for allowing me to use their pictures (p.17 and p. 22) in the examples. Likewise, I would like to thank Professor Kazumi Hatasa for allowing me to use the graphic (p. 24) from his website *Friends of New Nakama*. 
I am also grateful for all the staff at the School of Languages and Cultures for supporting my growth as a researcher and as a language instructor. Further thanks go to all my friends at Purdue University for their friendship and helping me survive in the past academic years.

Lastly, I would like to thank my parents in Japan for understanding my life in the U.S.A., especially to my dad for supporting me from his sick bed. I am also deeply grateful to my husband, Masahiko, for encouraging me to pursue the Master’s Program and continued support from Georgia.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>viii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 1. INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Motivation for This Study .................................................. 1

The Need for Computer-Mediated Performance-Based Assessment .......... 3

Research Questions ....................................................................... 5

<table>
<thead>
<tr>
<th>CHAPTER 2. RESEARCH BACKGROUND AND THE PERFORMANCE-BASED TEST</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

Learners’ Needs and Goals of Foreign Language Learners .................. 7

Achievement and Proficiency Tests in a Language Course ................. 8

The Current Situation and Issues in Classroom Achievement Testing .... 9

Inconsistency with Student Needs/Goals and Classroom Focus .......... 9

Issue of Face Validity .................................................................. 9

Test Item Efficacy ....................................................................... 10

Multiple-choice Questions .......................................................... 10

Cloze Items .............................................................................. 11

Oral Test Frequency .................................................................... 12

Limitations to Conducting Oral Assessment ..................................... 13

Speaking Practice in the Classroom ............................................. 14

The Performance-Based Test (PBT) ............................................... 16

The Purpose of PBT .................................................................... 17

Administrative Procedures for the PBT ........................................... 18
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBT Tasks</td>
<td>18</td>
</tr>
<tr>
<td>Timed Dictation</td>
<td>18</td>
</tr>
<tr>
<td>Read-aloud</td>
<td>19</td>
</tr>
<tr>
<td>Monologue</td>
<td>20</td>
</tr>
<tr>
<td>Reading</td>
<td>20</td>
</tr>
<tr>
<td>Q &amp; A</td>
<td>21</td>
</tr>
<tr>
<td>Role-play</td>
<td>21</td>
</tr>
<tr>
<td>PBT Rating Criteria</td>
<td>23</td>
</tr>
<tr>
<td>Q &amp; A and Reading</td>
<td>23</td>
</tr>
<tr>
<td>Role-play</td>
<td>25</td>
</tr>
<tr>
<td>Monologue</td>
<td>25</td>
</tr>
<tr>
<td>Read-aloud</td>
<td>25</td>
</tr>
<tr>
<td>Timed Dictation</td>
<td>26</td>
</tr>
<tr>
<td>Advantages of the PBT</td>
<td>26</td>
</tr>
<tr>
<td>Chapter 2 Summary</td>
<td>29</td>
</tr>
<tr>
<td><strong>CHAPTER 3. METHODOLOGY</strong></td>
<td>30</td>
</tr>
<tr>
<td>Overview</td>
<td>30</td>
</tr>
<tr>
<td>Study 1</td>
<td>30</td>
</tr>
<tr>
<td>Participants</td>
<td>30</td>
</tr>
<tr>
<td>Procedures</td>
<td>31</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>32</td>
</tr>
<tr>
<td>Study 2</td>
<td>32</td>
</tr>
<tr>
<td>Participants</td>
<td>33</td>
</tr>
<tr>
<td>Procedures</td>
<td>33</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>33</td>
</tr>
<tr>
<td>Chapter 3 Summary</td>
<td>33</td>
</tr>
<tr>
<td><strong>CHAPTER 4. RESULTS AND DISCUSSIONS</strong></td>
<td>34</td>
</tr>
<tr>
<td>Study 1</td>
<td>34</td>
</tr>
<tr>
<td>Result of the Student Questionnaire</td>
<td>34</td>
</tr>
<tr>
<td>Research Findings</td>
<td>69</td>
</tr>
<tr>
<td>-------------------</td>
<td>----</td>
</tr>
<tr>
<td>Pedagogical Implications</td>
<td>71</td>
</tr>
<tr>
<td>Implications for Future Research</td>
<td>72</td>
</tr>
<tr>
<td>Vocabulary and Grammar Retention</td>
<td>72</td>
</tr>
<tr>
<td>Comparison to Face-to-face Interviews</td>
<td>73</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>73</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>75</td>
</tr>
</tbody>
</table>

**APPENDICES**

<table>
<thead>
<tr>
<th>Appendix A</th>
<th>PBT Rating Criteria</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B</td>
<td>Student Questionnaire</td>
<td>86</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Instructor Questionnaire</td>
<td>90</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Score Chart</td>
<td>23</td>
</tr>
<tr>
<td>Table 2. Section 1: The Skills that Students Want to Improve the Most (Study 1)</td>
<td>35</td>
</tr>
<tr>
<td>Table 3. Section 2: Students’ Perceptions toward Two Test Formats (Study 1)</td>
<td>37</td>
</tr>
<tr>
<td>Table 4. Section 3: Students’ Methods of Preparation for Exam (Study 1)</td>
<td>44</td>
</tr>
<tr>
<td>Table 5. Section 4: PBT Task Preferences (Study 1)</td>
<td>47</td>
</tr>
<tr>
<td>Table 6. Section 1: The Skills that Students Want to Improve the Most (Study 2)</td>
<td>50</td>
</tr>
<tr>
<td>Table 7. Section 2: Students’ Perceptions toward Two Test Formats (Study 2)</td>
<td>52</td>
</tr>
<tr>
<td>Table 8. Section 3: Students’ Methods of Preparation for Exam (Study 2)</td>
<td>58</td>
</tr>
<tr>
<td>Table 9. Section 4: PBT Task Preferences (Study 2)</td>
<td>61</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1. SE Exercise Screen</td>
<td>17</td>
</tr>
<tr>
<td>Figure 2. Read-Aloud Screen</td>
<td>20</td>
</tr>
<tr>
<td>Figure 3. Guided Conversation Screen-1</td>
<td>22</td>
</tr>
<tr>
<td>Figure 4. Guided Conversation Screen-2</td>
<td>22</td>
</tr>
<tr>
<td>Figure 5. An Example of Q &amp; A Picture</td>
<td>24</td>
</tr>
</tbody>
</table>
Communicative language teaching in the early 1980s led to greater emphasis on the development of communicative ability and also increased interactive practice in classroom learning. However, in many communicative language programs today, large portions of periodic assessment do not necessarily reflect this pedagogical focus; instead, the traditional pencil-and-paper format is still in common use. Traditional written tests typically emphasize grammar and written language, and multiple-choice and cloze test items are often employed. In such formats, students are not required to demonstrate the oral skills they have acquired in the classroom; therefore, traditional written tests are not appropriate indicators of the students’ ability to use the language in real-life situations. There may be one or two oral tests in a semester, but that would not be enough to outweigh the grammar/written language emphasis. Furthermore, oral assessments are not typically given with sufficient frequency to track the development of students’ oral skills because they generally require significant amount of time for administration.

To improve such situations, a computer-assisted achievement assessment called the “Performance-Based Test” (PBT) was designed at Purdue University with the focus
on promoting students’ speaking skills. In order to argue for the need for the PBT and validate it, this study compared it with the traditional written test. In this study, both tests were administered to Japanese 101 (Study 1) and 102 (Study 2) students in order for them to experience both test formats. Then, at the end of the semester, a survey was conducted to elicit students’ perceptions of the tests and how each format related to their learning goals.

Study 1 results showed that students preferred the PBT, felt motivated to prepare for it, and felt a sense of accomplishment upon completion. Students also stated that the PBT is better aligned with their learning goals and reflects their knowledge and abilities more adequately than the traditional written exam. Study 2 obtained similar results. The number of students who preferred the PBT and felt a sense of accomplishment was significant. They also answered that the PBT is better aligned with their learning goal than the traditional written exam. Based on these findings, this study suggests that the PBT is a viable alternative with a number of advantages over the traditional written exam since this method corresponds to the course objectives and students’ learning goals to improve their speaking skills.
CHAPTER 1. INTRODUCTION

Motivation for This Study

In the context of foreign language instructional programs, speaking is a vital skill that should be developed. Current practices, largely dependent on the Communicative Approach (Canale & Swain, 1980), emphasize communication in a real-world context and language learning via meaningful and communicative practice. The advent of communicative language teaching in the 1970s to early 1980s prompted language teachers to create more communicative activities (e.g. role-play, information gap, interviews, etc.) in the classrooms (Omaggio, 1986). Meanwhile, American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines (Byrnes, Child, Patrizio, Lowe, Makino, Thompson, & Walton, 1986) served as the national consensus on language proficiency goals and standards for assessing proficiency. The use of these guidelines aims to measure what a test taker can or cannot do with the language based on the criteria for the four skills. After that, oral proficiency guidelines were especially implemented in language classes and adjusted to instructors’ own practices, measurement and testing (Pino, 1989; Magnan, 1986; Fall, Adair-Hauck & Glisan, 2007). In response to these trends, the field of language assessment began a shift from the traditional pencil-and-paper format to a performance-based assessment during which learners demonstrate their acquired knowledge and skills (Norris, Brown, Hudson & Yoshioka, 1998;
McNamara, 1996; Brown, 2007). However, difficulties in the administration of such performance-based oral assessments (interviews, presentations, etc.) impeded consistent implementation because they generally require considerable time and labor (Bachman, 1990; Fukada, 2015; Omaggio, 1986; Pino, 1989).

One of the course objectives in a Japanese 101 course at a Mid-western university in the United States is to promote proficiency using the communicative approach. The stated goals of the curriculums are that, “The purpose of this course is to promote elementary level of proficiency in Japanese through a communicative approach. Students are expected to be able to understand basic spoken and written forms of modern Japanese, and to do some basic communication in Japanese upon completion of the course. In addition to being able to read/write sentences in hiragana, katakana and kanji (we are introducing 28 kanji characters), you will be able to function at a survival level in interactive situations common to daily life in Japan”. The communicative method is considered fundamental, and instructors place emphasis on speaking activities immediately after introducing new grammar; however, achievement tests, locally called “chapter tests”, occur six times throughout the term by means of a traditional written exam, whereas oral assessments occur only once or twice any given term, usually administered in the form of one-on-one oral interview. During the interview week, regular classes are canceled, and teachers allot three or four days for the interviews. The time set aside for each interview amounts to seven minutes per student. Even though the communicative approach is applied to everyday instruction, there are few integrated opportunities to assess students’ oral proficiency.
Canale and Swain (1980) stated that a communicative approach needs to be built on the learners’ communicative needs, and its goal is to facilitate the integration of grammatical competence, sociolinguistic competence, and strategic competence (i.e. communication strategies). We must take note of an inconsistency between classroom focus and actual testing. The traditional written exams neither reflect the day-to-day classroom activities, nor do they align with one of the curriculum goals, i.e. students will be able to function at a survival level in interactive situations common to daily life in Japan, because they do not provide students with the necessary speaking opportunities.

Another concern would be that when the test emphasizes grammar and vocabulary in the written mode, students would likely prepare for it by rote memorization, rather than learning to use such knowledge for communicative purposes. In addition, such traditional testing formats are unhelpful for instructors, since they cannot adequately describe how successfully students can communicate with others using correct pronunciation, grammatical structures, and fluency. With regards to this matter, a face-to-face interview is an effective method to measure students’ communication skills; however, there are serious concerns relating to the logistics and practicality of an interview as assessment.

The Need for Computer-Mediated Performance-Based Assessment

The utility of language tests and assessments is considerably higher if the measures reflect course objectives and classroom instruction (Bachman, 1991; Gareis & Grant, 2015; Glaser & Silver, 1994; Jacobs & Chase, 1992). Glaser & Silver (1994) state that “first, the outcomes being tested must be recognized and accepted as important
objectives of the instructional program. Second, achievement assessment must be planned and implemented as an integral part of the curriculum and program of instruction” in order to promote learning (P.411). Furthermore, the characteristics of language tests affect students’ preparation, motivations, performances and scores. Therefore, a carefully-designed test should reflect our language priorities: the particular abilities that we want our students to improve (Hughes, 2003; Wiggins, 1989). An ideal assessment, then, features exercises to appropriately measure those abilities (e.g., speaking and communication skills). In the case of a proficiency-based communicative curriculum, an aligned practical test design and increased opportunities for oral assessments are crucial in order for instructors to observe students’ step-by-step verbal language acquisition of materials. In addition, it is important to verify that test formats are valid, reliable, and practical (Bachman & Palmar, 1996) as well as the alignment with learners’ goals. Worth noting here too, is the washback effect: that testing is likely to affect teaching and learning, including curriculum design, classroom practice, and student preparation (Alderson, 1991).

At the aforementioned mid-western university, large portions of the exams are conducted in the paper and pencil format. In order to address the issue of assessment alignment with course objectives, this study presents a new assessment called the Performance-Based Test (PBT), which prioritizes oral language development with frequent oral assessments. It is implemented via Speak Everywhere, a computer-assisted program, and it is designed as a complete replacement for the traditional written exam and oral tests. In order to establish the need for and examine the validity of the PBT,
surveys were conducted at the end of the semester to elicit students’ perceptions and opinions regarding the traditional written exams and the PBT.

**Research Questions**

The purpose of this study is to present the Performance-Based Test (PBT) as a new assessment alternative in foreign language courses. In order to examine the need for and the validity of the PBT, the following research questions will be investigated.

1. Do students of Japanese 101 and 102 at the Mid-western university prefer the style of traditional written testing or PBT testing?

2. Considering both traditional written testing and the computer-assisted PBT testing, which of the two test formats do the students think better aligns with their learning goals?

3. What are the students’ perceptions (e.g. with regard to anxiety, motivation, fairness) of the two test formats?

4. Did the PBT format change the way the students studied Japanese in the course?

5. What effect, if any, did the introduction of the PBT assessment have on student learning?

These questions will be addressed through survey studies. The surveys evaluated the test’s appeal, students’ perceptions of its value and utility, and the relation of the formats to their learning goals. The questionnaires consisted of four parts: (1) students’ reasons for taking the Japanese course; (2) a comparison between the traditional pencil-and-paper test and the performance-based assessments; (3) students’ preparation for the
exams; and (4) PBT task preferences. All research questions will be examined by a quantitative analysis of Likert-scale survey items and a qualitative analysis of students’ comments.
CHAPTER 2. RESEARCH BACKGROUND AND THE PERFORMANCE-BASED TEST

The goal of this study is to introduce the Performance-Based Test (PBT) and advocate its use as a new achievement assessment option for foreign language courses. In order to understand the value of the PBT, this study examined students’ perceptions of it as compared with the traditional written test.

This chapter first discusses language learners’ goals and expectations found in previous studies. Next, it will address various issues regarding the current status of foreign language achievement tests used in courses. Lastly, the PBT will be described in detail, including its concept, implementation, administration, and grading.

Learners’ Needs and Goals of Foreign Language Learners

The needs and goals of foreign language learners have been investigated by many researchers with the objectives of creating curricula and assessment instruments (e.g. Iwai, Kondo, Lim, Ray, Shimizu, & Brown, 1999). An analysis of students’ needs and goals helps instructors understand their motivation, effort, participation in the classroom, and attitude toward learning (Gardner, 1985; Schumann, 1997).

In language learning, communication in the target language is central. As Brown (2007) notes, “The benchmark of successful language acquisition is almost always the demonstration of ability to accomplish pragmatic goals through interactive discourse with
other speakers of the language” (p. 322). In fact, Alalou (2001) reports that over 90% of the students enrolled in French, German, and Spanish consider speaking as the most important skills in Modern Language studies. Similarly, Thomas (2010) reports that students enrolled in a Japanese language program considered verbal communicative dimension as their primary motivation to choose the course. Liu and Shibata (2008) found that integrative orientation (i.e. a desire to learn the language and socialize with native speakers) and cultural understandings were the most common reasons why students enrolled in Japanese. A study by Tse (2000) reports that limited oral communication training persists as a common failure of foreign language programs. These studies suggest that language learners typically value the development of their speaking skills over all the other skills. The current study is also designed to enhance student motivations for learning Japanese.

**Achievement and Proficiency Tests in a Language Course**

In this study, the traditional written chapter tests were replaced by the PBT. Such chapter tests, quizzes and final exams are categorized as achievement tests. An achievement test is commonly defined as a test aimed at diagnosing whether or not the examinees mastered specific linguistic features (e.g. verb forms, vocabulary items, or syntactical patterns) taught within a particular unit of study. Therefore, students are able to prepare for such tests by reviewing specified materials introduced and practiced in class. Achievement tests serve to discover students’ gaps in understanding, as well as ascertain their attainment of the learning goals of the course. In this regard, it can also be referred to as a syllabus-based test. Meanwhile, a proficiency test is used to measure a
broad range of global competencies acquired by the learner regardless of the course, curriculum, or instructor. Unlike achievement tests, it is difficult to prepare for proficiency tests in the same way as one might study for an achievement test, since the tasks or materials are not predefined in a specified range of materials (Bachman, 1990; Brown, 2007; Clark, 1979; Omaggio, 1986).

The Current Situation and Issues in Classroom Achievement Testing

What follows is a list of issues pertaining to the classroom achievement tests that are commonly used in language curricula:

**Inconsistency with student needs/goals and classroom focus.** In the Japanese course where this study was conducted, chapter tests occur six times throughout the semester, whereas oral assessments occur only once or twice. One of the concerns is that all achievement tests are completed in a pencil-and-paper format. Given that the classroom focus is to learn materials through oral practice (e.g. call-and-response and communicative activities), the format of a traditional achievement test is inconsistent with its focus because such formats do not properly measure students’ oral language development, but instead measure vocabulary and grammar knowledge that students learned. Therefore, this format is not sufficiently in line with either the classroom focus or the aforementioned student needs/goals.

**Issue of face validity.** Written test formats may be effective assessments of listening and reading comprehension, and therefore reflect the abilities that the instructors want to measure, such as learners’ control of particular grammatical structures (Hughes, 2003). However, when it comes to assessing speaking, these formats are less valid. Their
limited scope does not necessarily provide evidence that learners will be able to use the language in real-life situations. Such tests are called indirect speaking tests because they assess oral skills indirectly. Although the tests ostensibly assess oral skills, the indirect format does not provide opportunities for students to display language skills that they want to develop. Thus, indirect speaking tests are considerably lacking in terms of face validity. Test takers might even be discouraged when their speaking abilities are measured via pencil-and-paper (Clark, 1979). Given such indirect measures, it is difficult to measure such aspects of learners’ oral proficiency as fluency, accuracy, and complexity in speech.

**Test item efficacy.** Pencil-and-paper assessment often utilizes a combination of fill-in-the-blanks, multiple-choice, and true or false items. These formats are preferred because of scoring reliability and practicality. It is less time-consuming to assess and interpret achievement. However, researchers criticize these methods as artificial and superficial; above all, they rarely include oral language activities (Bachman, 1990).

**Multiple-choice questions.** Multiple-choice questions are the most prevalent selection-type test questions. These items are favored, especially in large-scale test administration due to efficiency in administration and scoring. Multiple-choice questions are effective for recall or recognition; however, the construct validity of such questions is often challenged. For example, a test taker may encounter a multiple-choice question below:

How do you greet someone you meet for the first time? Choose the appropriate answer from a) to c).

a) こんにちは konnichiwa (Hello)
b) ありがとう arigatou (Thank you)

c) はじめまして hajimemashite (Nice to meet you)

Selecting the correct option does not assure that he can pronounce the expressions correctly or use it in actual communication. Moreover, it is quite possible that students relied on guessing when answering multiple-choice items. Therefore, such question types may lead to negative washback on learning when students’ attention is paid to test taking strategies, and eventually create a gap between their knowledge and language use (Hughes, 2003). Furthermore, it is possible that multiple-choice test items might result in harmful washback on teaching—such as Lake Wobegon effect—when teaching becomes too closely tailored to the test style or the overuse of some tasks (like multiple-choice items) in order to earn higher scores on standardized tests (Norris, Brown, Hudson & Yoshioka, 1998). Scouller (1989) found a negative influence on the quality of learning that a multiple choice format might bring about. It was found that students are more likely to employ surface strategies and motives when preparing for Multiple-Choice Question (MCQ) examinations. Surface strategies refer to when learners summarize materials without understanding them well while preparing for the assessment. Surface motives indicate that learners choose topics that they think they can do well on rather than those they are really interested in when preparing for the assessment. Students also perceived that the MCQ examination assesses knowledge-based or lower level cognitive abilities.

**Cloze items.** Cloze items (i.e. fill-in-the-blank items) are also widely used in written exams. In a dialogue completion test, for example, students read a conversation between two or more speakers and complete the conversation with appropriate phrases
and sentences in the target language. Grammar usage, vocabulary, and spelling are scored. The cloze procedure might be effective in tests of grammar and vocabulary learning (Talebzadeh & Bagheri, 2012), or reading and listening if answers can be found in the given text (Hughes, 2003), but it does not address speaking traits (e.g. pronunciation). Considering the many facets of communication, such a format also does not precisely diagnose test-takers’ weaknesses or strengths in oral proficiency (Shohamy, 1982).

It follows that high scores on these written format exams are simply not reliable indications that students have adequate oral skills (e.g. fluent speech and correct pronunciation) (Brown, 2007; Qian 2009). Indirect speaking tests like the written tests are not only ineffective for assessing oral skills, but they also do not reflect current language learning pedagogies, which are more focused on interactive and communicative methods (Bachman, 1990). In addition, negative washback should be taken into account as previously discussed. Studies show that test characteristics can affect students’ experience of preparing for and taking tests (Alderson & Wall, 1993; Bachman & Palmer, 1996; Cheng, 1998; Hirai, Fujita, Ito, & O’ki, 2013; Xie & Andrew, 2012). When a test is administered in a written format, students tailor their learning accordingly, by memorizing materials rather than speaking and listening, which might also promote surface learning. Such format, therefore, may hinder the development of learners’ communication skills.

**Oral test frequency.** Fukada (2015) reports on a survey conducted with the American Association of Teachers of Japanese (AATJ) members in 2013 to which 61 institutions (28 K-12 and 33 universities) responded. This survey revealed that the
average frequency of oral assessments was 2.73 per semester, while the average frequency of the written exams was 3.75 per semester. Only 11 institutions conducted oral assessments more frequently than written exams. Fukada pointed out that although the development of students’ speaking skills is emphasized in language education, oral assessments are infrequent. Regular oral assessments are ideal for tracking the development of students’ oral skills from one textbook chapter test to the next. Frequent and formative assessments benefit both teachers and students. For teachers, formative assessment helps them to monitor students’ progress and modify instruction based on students’ learning needs. For students, the results of oral assessments provide important feedback that can bridge the gap between their current skills and desired proficiency (Black & Williams, 1998a, 1998b; Tuttle & Tuttle, 2013). However, frequent oral examinations are difficult to integrate into the regular assessment schedule due to administrative constraints.

**Limitations to Conducting Oral Assessment**

There are many tasks used in oral assessment, including one-on-one interviews, paired-interviews, role-play, and narrative tasks (Luoma, 2004). In the survey by Fukada (2015), 39 out of 61 institutions reported that they conduct one-on-one interviews as an oral assessment. Thirty-five institutions use paired or group performance. Interviews are the direct, communicative, and authentic tests that most closely reflect real-life language use (Bachman, 1990; Clark, 1979); nevertheless, oral interviews and other types of oral exams are not frequent as Fukada pointed out in his study.
One major obstacle of oral assessments lies in the debate of practicality. The oral assessments mentioned above require significant time and costs in administering, training raters, rating, and analyzing the scores (Bachman & Palmer, 1996; Clark, 1979; Kenyon & Malabonga, 2001; Norris, Brown, Hudson & Yoshioka, 1998; Pino, 1989). Moreover, since an interview is administered individually, regular classes might have to be compromised or cancelled so that the students can avoid conflicts with other classes (Fukada, 2015). Some teachers might allocate their office hours for oral assessments, but scheduling a large number of appointments may still prove arduous. Furthermore, appointment slots allocated to each student might be insufficient. Therefore, it is important to have an alternative method that overcomes these practical issues.

**Speaking Practice in the Classroom**

In a language class, a pedagogy that integrates the four skills (listening, speaking, reading, and writing) is of paramount importance. However, especially in the first two years of instruction (novice to intermediate), the curricular focus is more on oral proficiency development than writing and reading proficiency development. The day-to-day focus in the classroom also reflects this. Similarly, students’ learning goals are to develop oral communication skills as previous studies found. Given that both students’ and day to day classroom focus are oral communication development, listening and speaking skills should be a focus in the classroom.

Communicative Language Teaching is a main approach in current language curricula and many teachers are seeking in-class activities and tasks to improve their students’ communicative skills. However, classroom time is consumed by other tasks—
presenting vocabulary, grammar, culture, and projects. Furthermore, it is impossible for a single teacher to monitor all students’ oral production and offer individual feedback in the course of a single class session. (Fukada, 2015). According to Fukada and Wei (2013), speaking practice was the focus of 42% of any given class session (i.e. 25 minutes/hour). However, actual speaking time per person amounts to 10 minutes or less when they practice in pairs, small groups, or call-and-response with the instructor. Class time is inarguably limited; classroom experience alone does not provide sufficient time for a novice language learner to develop into a fluent speaker. A solution to this issue is to increase students’ speaking and listening practice time outside of class. This study suggests that if the currently utilized achievement exams were replaced by the oral assessment-based PBT, speaking practice would increase both within the classroom and outside the classroom in preparation for the PBT. This format would encourage students to practice and review the materials covered in class and lead to stronger oral skills than in the written format.

In summary, the major concerns discussed above include: (1) in current practices, the test format is not aligned well with classroom focus, (2) regular oral assessment is ideal, but various practical issues impede the implementation of such an assessment program and (3) it is also difficult to allocate time for additional speaking practice in class. To overcome these issues, a new assessment method, the PBT, was designed.
The Performance-Based Test (PBT)

Performance assessments are widely defined as direct, authentic, and simulating real-world tasks (Brown, 2000; McNamara, 1996; Norris, Brown, Hudson & Yoshioka, 1998; Wiggins, 1989). Tasks in performance assessments imitate the actual situations where candidates must demonstrate their skills that they acquired, and those skills will be developed based on the prediction that the candidates will use the skills in the real-world in the future. The performance-based testing in this study requires students to demonstrate their mastery of skills through actual oral performances, rather than measuring students’ underlying knowledge by means of pencil-and-paper tests (McNamara, 1996). The PBT aims to create oral tasks that students would require in the real world situations to a greater degree than the traditional written tests.

The PBT is a computer-assisted test using a software system called Speak Everywhere (SE). Speak Everywhere is an online oral practice/instruction/assessment platform for foreign language teaching developed by the Center for Technology-Enhanced Language in the School of Languages and Cultures at Purdue University. It is a user-friendly, web-based server software program and makes large-scale language practice and testing much easier. The program allows language instructors to efficiently create exercises, grade, and give feedback to their students. Students’ oral practice is the focus of the program. This program will increase students’ oral practices outside the classroom when the instructor assigns Speak Everywhere exercises as homework. This online program enables teachers to monitor their students’ individual speech recordings and to rate their oral performance (Fukada, 2013).
Figure 1 shows an example of a PBT task. The left side of the screen is an instruction panel where students read instructions and a passage. Pictures and tables can also be placed in the same area. The right portion of the screen is the video display area and a control panel.

The purpose of PBT. The central purpose of the PBT is to develop students’ communication skills in Japanese so that each student can use Japanese in a real conversation. In order to attain this goal, we think that it is crucial to increase opportunities for oral assessment and speaking practice. Therefore, this study replaced
the traditional written exam with the PBT and aims to improve students’ oral performance.

**Administrative procedures for the PBT.** First, PBT tasks are made available online at the beginning of each chapter so that students are able to practice at their convenience. At this time, model audio samples are presented to show to the students what level of performance is expected. This openness is also expected to reduce test anxiety caused by not knowing what will be on the test. Second, teachers inform students that they will be graded not on whether or not they can complete the tasks, but on how well they can perform them. Third, a PBT scoring rubric is also made public so that students have a clear picture of what to expect in scoring (e.g. fluency, promptness of responses, grammatical accuracy, etc.). Lastly, on the test day, students are instructed to record their responses as many times as necessary and to choose the best recording for their submission.

**PBT tasks.** The PBT was primarily designed to strengthen oral production skills. PBT tasks are contextualized with content that students are likely to encounter in real-world situations. The PBT is composed of six sections: timed dictation, read-aloud, monologue, reading, Q & A, and role-play. These tasks were designed to promote students’ integrated performance so that students can improve four skills: speaking, listening, writing, and reading. All tasks but timed dictation are implemented on *Speak Everywhere*. The tasks were made accessible to students at the beginning of each chapter. For example, tasks for Chapter 2 are made accessible when the class starts studying Chapter 2 of the textbook.
Timed dictation. Timed dictation was devised to foster listening comprehension, and writing fluency and accuracy. The students were instructed to listen to and write down ten sentences. Each sentence consisted of new words and grammar constructions they learned in each chapter. Sentences were presented via an audio clip. The audio file for practice was available online so that students can practice both inside and outside the classroom. On the test day, students were given a chance to listen to each sentence twice. The length of the pause after a sentence was determined by considering the length of the sentence, the number of kanji characters, and the students’ writing fluency.

Read-aloud. Figure 2 shows a read-aloud screen. The objectives of the read-aloud task are to enhance students’ pronunciation (pitch accents and intonation), and oral fluency, but it has been shown to also improve listening, reading comprehension, and vocabulary retention (Kobayashi, 2006). A paragraph of text was created using the grammar and vocabulary introduced in each chapter. Model audio was also available in the display area so that students could listen on demand.
Monologue. This task provides students with a particular topic or a scenario to elicit their free-form responses. For example; “Imagine that you just met a Japanese student on campus. The student asks you what you usually do on a weekend. Describe your typical weekend using a variety of words and expressions you’ve learned in the chapter.” Students begin recording once they are ready to speak.

Reading. In this section, students were directed to read a short passage, listen to a comprehension question, and answer it orally. It was created to assess students’ integrated skills of reading, listening, and speaking.
**Q & A.** Q&A was another integrated task where subjects were directed to listen to a series of 10 questions and answer them orally (e.g. What do you often do on weekends?) one at the time. A chart or a picture accompanied some of the questions.

**Role-play.** Students participated in various role-play activities. A dialogue from each chapter of the textbook was typically used. The dialogue usually featured two characters. Students memorized the lines of each character and took turns playing the role of each character. When the dialogue was not suitable for role-play, a conversation—called a guided conversation—was created, during which students carried on a conversation with a person on the screen by following textual cues. Figure 3 presents an example screen of guided conversation. The instruction panel on the left provides the context for the conversation. The yellow box below is a textual cue. A test taker reads these instructions first, and clicks the start button to play the first video. A Japanese student in the video starts introducing himself in Japanese. When the playback is finished, the test taker is to introduce himself/herself as the cue instructs. Then, on the next screen (Figure 4), the Japanese student in the video asks “What is Purdue University like?” The test taker is to describe the university as instructed by the cue. The conversation goes on in this way until it ends. This task was designed to allow students to have a communicative interaction with an interlocutor recorded in the video.
Figure 3. Guided Conversation Screen-1, showing Ryota Kondo participant. Screen capture from *Speak Everywhere* developed by Atsushi Fukada. Used by permission.

Figure 4. Guided Conversation Screen-2, showing Ryota Kondo participant. Screen capture from *Speak Everywhere* developed by Atsushi Fukada. Used by permission.
**PBT rating criteria.** Rating criteria were developed to rate students’ oral performances. Students’ performances were evaluated on these aspects: grammatical accuracy, vocabulary usage, structure (discourse structure), promptness of speech, fluency, pronunciation, and reading accuracy (reading mistakes in read-aloud tasks). Structure, promptness of a response, fluency, pronunciation, and reading accuracy were scored on a five-point scale respectively (see Appendix A). Using Kormos and Denes (2004) and Kondo (2014) as references, criteria for the structure, fluency, pronunciation, and accuracy components were developed. Scoring of timed-dictation and criteria for the vocabulary & grammar and promptness components were developed through discussion with experienced Japanese language instructors. The vocabulary & grammar component was allotted 15 points: three mistakes would result in zero, and five points were deducted for each mistake per question (see Table 1). Since each task had its own set of components to rate, the rating procedure for each is described below.

<table>
<thead>
<tr>
<th>Table 1. Score Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary Grammar</td>
</tr>
<tr>
<td>15</td>
</tr>
</tbody>
</table>

**Q & A and reading.** In Q & A and reading tasks, vocabulary & grammar (15 pts), promptness (5 pts), fluency (5 pts), and pronunciation (5 pts) were evaluated for every question. For example, in one Q & A item, a student looks at a picture below (see Figure 5) and is asked, “Where is a cat?” The answer is, “The cat is on the bed.” If the student
makes a mistake by saying, “The cat is under the bed” with correct pronunciation, adequate fluency, and sufficient promptness, the student receives five points each in pronunciation, fluency, and promptness. However, since he used the wrong word, he would get a five-point deduction on grammar for improper use of a preposition, bringing the total score down to 25 out of 30. As this example shows, one vocabulary & grammar mistake is equivalent to the maximum score of the promptness, fluency, and pronunciation components. This deduction seems rather large; however, students had the benefit of sufficient practice time because all PBT tasks were made available before the exam. Students were also allowed to have unlimited tries before submission. Since students could prepare for their answers in advance, we assumed their likelihood of making vocabulary/grammar errors would be lower. Anticipating the criticism that students might just memorize all responses and recite them without understanding them, the items in these sections are randomized. Reading tasks were scored in the same way as Q & A tasks.

![Figure 5. An Example of Q & A Picture. The graphic from Friends of New Nakama, developed by Kazumi Hatasa. Used by permission.](Image)
**Role-play.** For this task, vocabulary & grammar, promptness, fluency, and pronunciation were evaluated. In case of guided conversation tasks, the scores were distributed as in Q & A and reading: vocabulary & grammar (15 pts), promptness (5 pts), fluency (5 pts), and pronunciation (5 pts), but an evaluator listened to student’s responses in each turn and rated all turns as a whole. In order to properly weigh this task in relation to the other tasks, the score was multiplied by eight (=240 pts). This score amounted to approximately 18 % of the entire score.

When a dialogue from a chapter of the textbook was used, students were required to memorize the lines; therefore, chances of making vocabulary/grammar errors would be low. For this reason, the vocabulary & grammar component was reduced to five points for each speaker’s turn: vocabulary & grammar (5 pts), promptness (5 pts), fluency (5 pts), and pronunciation (5 pts) for one speaker and the same distribution for another speaker, and one point was deducted per mistake on vocabulary & grammar. The score was also multiplied by eight (=240 pts).

**Monologue.** Vocabulary & grammar, structure, fluency, and pronunciation were evaluated and scored. Students were asked to deliver their speech about given topics and could obtain a maximum of 30 points for this task. An evaluator listened to a student’s entire speech and rated it as a whole. One point, instead of 5 points, was deducted per mistake on vocabulary & grammar. In order to properly weigh this task in relation to the other tasks, the score was multiplied by eight (=240 pts). This score amounted to approximately 18 % of the entire score.

**Read-aloud.** In Read-aloud tasks, students were asked to read paragraph(s), and fluency, reading accuracy, and pronunciation were evaluated. Reading accuracy measures
how faithfully a student reproduces the test orally, whereas pronunciation measures segmental accuracy (e.g. the distinction between voiced vs. voiceless obstruents especially problematic for L1 Chinese learners) and prosodic accuracy (e.g. pitch accents). Each component had a maximum of five points; therefore, the subtotal would be 15 points. An evaluator listened to a student’s entire speech and rated it as a whole. To adjust its weight with the other tasks, this score was multiplied by eight (=120 pts). This score amounted to approximately nine percent of the entire score.

**Timed dictation.** An evaluator gave one point per one letter correctly transcribed, including small letters in palatalized sounds (ゃ, ゑ, ィ), geminate consonant (っ), and long vowel (ー). For example, if a student could write 198 letters correctly out of 203 letters possible, then 198 will be the student’s score.

The students’ final scores were presented by percentage: the total number of scores was summed up and divided by the maximum possible number. Since there were multiple sections and multiple instructors, the instructors had a meeting to agree on grading-criteria prior to rating their students’ performance. Each instructor was responsible for rating their own students.

**Advantages of the PBT.** Tests can be classified according to the degree of directness with which ability is measured: indirect, semi-direct, and direct testing.

Indirect tests are used to estimate test takers’ abilities that underlie skills in development (Hughes, 2003). In terms of assessing oral proficiency, indirect testing assesses oral skills indirectly; such tests do not require test takers to produce oral language (e.g. written cloze tests). Direct testing refers to all procedures where the test taker demonstrates any related skills or performance the test developers want to evaluate.
Direct speaking tests, therefore, include face-to-face interviews where test takers produce speech during communicative interaction with one or more human interlocutors. A semi-direct speaking test, on the other hand, refers to those tests during which the test taker is required to deliver speech in response to the interlocutor by means of a recording or computer (Clark, 1979). The PBT is categorized as semi-direct.

The computer-mediated semi-direct testing like the PBT offers a number of advantages over direct test. First, oral assessment can be efficiently administered to a large group of test takers as long as appropriate equipment and facilities are available; thus, instructors can assess oral language frequently and efficiently. Second, unlike oral interviews, test reliability and fairness will be maintained because students receive the identical instructions standardized by the computer, and students’ scores and performance are not likely to be affected by the interviewers’ skills or inconsistent guidance. In addition, Shohamy (1994) reported that language samples obtained from the computer-mediated methods are more formal, cohesive, and longer than face-to-face interviews. Test takers using the semi-direct formats seemed to pay more attention to linguistic accuracy; they frequently and voluntarily used self-correction and paraphrasing (Clarks, 1979; Fukada, 2015; Hughes, 2003; McNamara, 1997; Norris, Brown, Hudson & Yoshioka, 1998; Quian, 2009; Shohamy, 1994; Shohamy & Reves, 1985; Stansfield, 1990a, 1990b; Stansfield & Kenyon, 1992).

Aside from practicality and reliability, the PBT offers advantages over indirect testing; it promotes alignment between classroom instruction, course objectives, and assessment. Also, the PBT has higher face validity than the written tests and provides more useful tasks that students can apply in the real-life situations since they allow
students to demonstrate speaking skills while listening to an interlocutor’s speech and responding to a real-time utterance. Accordingly, their preparation time for speaking will be increased. Furthermore, since the PBT utilizes the online software system, instructors can provide the PBT as a take-home exam; also, the PBT can be adapted to online courses.

Critics of semi-direct assessments like the PBT claim that the format cannot duplicate the authentic interactive nature of a face-to-face interview; these nuances include negotiation for meaning, expanding topics, and requesting clarification due to the absence of an interlocutor. Although it may appear that direct testing more closely reflects this authenticity, the interview format is also challenged by researchers who claim that conversation emerging from an interview is likely to be manipulated by the interviewer and thus does not reflect “real-world” conversations (Clark, 1979; Ginther, 2004; Hughes, 2003; Shohamy & Reves, 1985; Spolsky, 1985, Stansfield, 1990).

At the current study site, achievement tests for every unit were conducted in the form of the traditional written exams over many years. Clark (1979) suggests that semi-direct testing is suitable for measuring students’ diagnostic language achievement given the specific linguistic features that have been taught in the classroom, and is especially effective for students at the lower levels. The PBT is a viable alternative to paper and pencil chapter tests, because it measures the learners’ pronunciation, vocabulary, fluency, grammatical control, and promptness of responses. Furthermore, the PBT can include tasks that correspond to lesson objectives. Thus, this study replaced periodic written tests with the PBT in order to provide constant oral assessments.
Although there are many studies which transitioned entirely from a traditional paper-and-pencil format to a computer-based format (Choi, Kim, and Boo, 2003; Clariana & Wallace, 2002; Jones & Maycock, 2007; Noubandegani, 2012; Sawaki, 2001), few studies converted the paper-based tests to a computer-assisted format that includes the number and variety of language tasks featured in this study. Accordingly, little is known about students’ perceptions of the change from an indirect test to semi-direct performance-based formats and its influence on learning. Bachman and Palmer (1996) state that having test takers involved in the design and development of the tests and collecting information about their perceptions about the tests and test tasks (e.g. Erickson, 2010; Erickson & Gustafsson 2005) are beneficial because the more they tend to view assessment positively, the more they are likely to be motivated and perform better. Considering this potential positive washback, it is necessary to understand students’ preferences and perceptions (e.g. anxiety, sense of achievement, and motivation) of the two tests, as well as explore the learning effects of the two test formats on the students.

Chapter 2 Summary

In this chapter, current issues in achievement testing in Japanese education and an overview of the PBT were discussed while reviewing previous studies. The next chapter explains the methodology of this study.
CHAPTER 3. METHODOLOGY

Overview

The purpose of this study is to present a new assessment method, the Performance-Based Test (PBT), for use in foreign language courses. In order to examine the effectiveness of the PBT, this study surveyed students’ perceptions of two test formats used in classroom achievement testing: the traditional written test and the PBT.

In order to compare the two different test formats, students experienced both the written and the PBT testing over the course of one semester. At the end of the semester, students were asked to fill out a survey. The survey was conducted twice: once to Japanese 101 students in the Fall semester of 2014 (Study 1), and once to Japanese 102 students in the Spring semester of 2015 (Study 2). The results of the survey were analyzed to understand students’ preferences between and perceptions of the two test formats. An instructor questionnaire was also created, and all instructors participated in the survey.

Study 1

Participants. The participants in Study 1 (N=75) were comprised of students enrolled in five sections of Japanese 101 (first-year first-semester course) during the Fall
2014 semester at Purdue University. The instructors of these sections also participated. Two types of questionnaires were administered: one for the students and one for the instructors. A total of 75 students out of 80 students and four instructors participated in the study. The students’ first languages were Chinese (N=44), English (N=28), and Korean (N=3). There were 48 males and 27 females whose mean age was 20.2 years old. The teachers were all native Japanese speakers.

**Procedures.** Written exams were replaced by PBT at the midterm point of the semester in order for the students to experience both types of testing.

In Japanese 101, six chapters were covered, and students took six chapter exams and one final exam. The textbook used for the course was *Nakama 1: Japanese communication, culture, context* (Hatasa, Hatasa & Makino, 2010). In order to compare the written exam and the PBT, the written exam was replaced by the PBT midway through the semester: during the first half of the semester, the students took a traditional paper and pencil test at the end of each chapter; during the second half of the semester, they completed the PBT for chapters four through six. The final exam was also the PBT. The first PBT was conducted in a computer laboratory, but in an effort to reduce any possibility of technical issues and to reduce the subjects’ anxiety, it was then changed to a take-home exam for all tasks except for timed dictation. Timed dictation took place in the classroom one day before the PBT exam. On the day of the final exam, students were invited to participate in the student survey. The purpose of the survey was announced, and the survey was distributed in hard copy. It was explained that its aim was not only to ask for their preferences or thoughts on the two styles of tests, but also to improve the overall quality of future sections of Japanese 101.
**Questionnaires.** A questionnaire was developed to elicit students’ preference of and views on the two formats. The student questionnaire (see Appendix B) was anonymous and consisted of a background survey (e.g. first language, age, gender, etc.) and four sections: (1) students’ reasons for taking the Japanese course; (2) a comparison between the traditional pencil-and-paper test and the performance-based assessments; (3) students’ preparation for the exams; and (4) PBT task preferences. The survey included Likert scale items and open-ended questions. For example, “Which style of evaluation/test did you like better?” asked students to rate the degree to which they preferred the PBT or the traditional written test. The students could choose from the following statements: liked traditional much better, liked traditional somewhat better, about the same, liked performance-based somewhat better, and liked performance-based much better. The z-test was used to determine whether two populations (e.g. students who preferred the PBT and students who preferred the traditional written exams) differ significantly on a single question. The level of significance was set to 0.05.

The instructor questionnaire (see Appendix C) similarly elicited their thoughts and commented on the PBT with more emphasis on changes in their teaching methods as well as their emphasis on teaching.

**Study 2**

**Participants.** The participants in Study 2 (N=34) were comprised of students enrolled in three sections of Japanese 102 (first-year second-semester course) during the Spring 2015 semester. The instructors of these sections also participated. As in Study 1, both a student and an instructor questionnaire were administered. A total of 34 out of 35
students and four teachers participated in the study. The students’ first languages were Chinese (N=13), English (N=20), and Korean (N=1). There were 25 males and 9 females whose mean age was 20.5 years old. The teachers were all native Japanese speakers.

Procedure. In Japanese 102, students proceeded to the next six chapters of *Nakama 1* and took six chapter exams and the final exam. During this semester and study, students took two written exams (the midterm and final exam) and five PBT exams. As in Study 1, all the PBT tasks (except for timed dictation) were made available online at the beginning of each chapter, and the grading criteria were clearly announced. Students took all the PBTs in the form of a take-home exam. All tasks were done at home except for the timed dictation part. On the day of the final exam, students were asked to fill out the student survey in class.

Questionnaires. The same questionnaires that were used in Study 1 were administered to both students and instructors in Japanese 102. One question was added asking whether or not participants had previously taken the PBT exam in Japanese 101 in order to pick out those that experienced both test formats.

Chapter 3 Summary

In this chapter, the methodology used for this present study was explained. In the next chapter, the research questions will be answered by summarizing the results of student questionnaires in both Study 1 and Study 2.
CHAPTER 4. RESULTS AND DISCUSSION

In this chapter, the data from surveys administered to the students and the instructors of Japanese 101 and 102 are analyzed and discussed. The student survey will be discussed first and then the instructor survey.

Study 1

A total 75 out of 80 students of Japanese 101 responded to the student survey (see Appendix B). In order for students to experience both the written exam and the PBT, the written exam was replaced by the PBT at the halfway point of the semester.

Result of the Student Questionnaire. Tables 2 to 5 report results from Likert scale items.

Section 1. The objective of this section is to find out students’ purpose(s) in taking the Japanese course. Table 2 summarizes Section 1 results.
Table 2. *Section1: The skills that students want to improve the most (Study 1)*

<table>
<thead>
<tr>
<th>Questions</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. What is your purpose in taking the Japanese language course?</td>
<td>To learn the language (26). To fulfill a language requirement (22). An interest in culture/subculture (17). To travel and to study abroad (12).</td>
</tr>
<tr>
<td>Q3. Among the 4 skills of Speaking, Listening, Writing, and Reading, which skill do you want to improve the most?</td>
<td>Speaking (%): 50.6 Listening (%): 33.3 Writing (%): 6.6 Reading (%): 14.6</td>
</tr>
<tr>
<td>Q4. What is/are your goal(s) for taking the Japanese course? What do you want to be able to do in Japanese by the end of the semester?</td>
<td>To have a basic conversation (31). To be able to speak fluently (20). To be able to read and write (14).</td>
</tr>
</tbody>
</table>

Note. The reason why the percentages for Q3 do not add up to 100% is that one student circled more than one choice.

Q1: “What is your purpose in taking the Japanese language course?”

Twenty-six out of 75 (35%) students articulated interest in learning the language by answering “To learn Japanese.”, while 22 students (29%) stated the necessity to fulfill a language requirement. Seventeen students (22%) showed interest in Japanese culture, including some areas such as anime, manga, and TV shows, and 12 (16%) students expressed interest in traveling or studying abroad.

Q2: “Why did you choose the Japanese language over the others?”

Similarly, most students (N=44) expressed interest in learning the language and a fascination with the Japanese culture/subculture. In other words, their motivation to select Japanese included both learning the language and a desire to participate in Japanese culture. Eight students mentioned a language similarity between their first language and the Japanese language.
Q 3: “Among the 4 skills of speaking, listening, writing, and reading, which skills do you want to improve the most?”

As Table 2 summarizes, half of the students desire to improve speaking skills (50.6%) followed by listening skills (33.3%).

Q 4: “What is/are your goal(s) for taking the Japanese language course? What do you want to be able to do in Japanese by the end of the semester?”

A large number of students (67%) mentioned communicative facets (e.g. having basic conversation and being able to speak fluently), while 14 students (19%) indicated that they want to read/write in Japanese. Based on this result, overall communication skills and oral communication fluency were dominant goals for students in Japanese language learning.

On the whole, the findings summarized above are very similar to those of Liu and Shibata (2008), which state that cultural understanding and integrative orientation (e.g. personal interests and a desire to interact with people who speak the target language) are the two foremost motivational factors in selecting a Japanese course.

Section 2. This section was designed to identify students’ perceptions toward the two test formats by comparing the traditional written exam given in the first half of the semester to the PBT exam in the second half of the semester. Responses are summarized in Table 3. Questions 1 through 6 use a five-point Likert scale along with a space for students to write rationale for their choices. The z-test was used to determine whether the two groups (e.g. students who preferred the PBT and students who preferred the traditional written exams) differ significantly. The level of significance was set to 0.05.
Table 3. Section 2: Students’ Perceptions toward Two Test Formats (Study 1)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Much more PBT (%)</th>
<th>Somewhat more PBT (%)</th>
<th>Both (%)</th>
<th>Somewhat more Paper (%)</th>
<th>Much more Paper (%)</th>
<th>N/A (%)</th>
<th>z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Which style of evaluation/test did you like better?</td>
<td>26.7</td>
<td>25.3</td>
<td>25.3</td>
<td>6.7</td>
<td>16.0</td>
<td>0.0</td>
<td>3.1255*</td>
</tr>
<tr>
<td>Q2. Which caused you more anxiety?</td>
<td>21.3</td>
<td>21.3</td>
<td>25.3</td>
<td>14.7</td>
<td>17.3</td>
<td>0.0</td>
<td>1.0773</td>
</tr>
<tr>
<td>Q3. Which gave you a greater sense of accomplishment?</td>
<td>32.0</td>
<td>21.3</td>
<td>25.3</td>
<td>8.0</td>
<td>12.0</td>
<td>1.3</td>
<td>3.6596*</td>
</tr>
<tr>
<td>Q4. Which test did you think was more challenging?</td>
<td>26.7</td>
<td>18.7</td>
<td>34.7</td>
<td>10.7</td>
<td>9.3</td>
<td>0.0</td>
<td>2.8583*</td>
</tr>
<tr>
<td>Q5. Which style of test motivated you more to learn?</td>
<td>24.0</td>
<td>26.7</td>
<td>25.3</td>
<td>9.3</td>
<td>14.7</td>
<td>0.0</td>
<td>2.8098*</td>
</tr>
<tr>
<td>Q6. Which style of test did you think was fairer?</td>
<td>9.3</td>
<td>16.0</td>
<td>37.3</td>
<td>20.0</td>
<td>17.3</td>
<td>0.0</td>
<td>1.3281</td>
</tr>
</tbody>
</table>

*p<0.05

Q1: “Which style of evaluation/test did you like better?”

This question aims to identify students’ exam preferences. In response to this question, 52% of students (26.7% much better; 25.3% somewhat better) showed their preference for the PBT, whereas 22.7% (16% much better; 6.7% somewhat better) preferred the traditional exam. Statistically, a significant difference ($z = 3.1255, p<0.05$) was found between the two groups of responses. In the comparison of students’ first language, 26 out of 44 Chinese speakers (59%) answered that they preferred the PBT either much or somewhat better, while 10 out of 28 English speakers (35.7%) answered that they preferred either much or somewhat better. Given that those Chinese students have taken computer-assisted language testing, such as TOEFL iBT or IELT, in order to
enter American universities, such education background might have influenced the preference result. They might be familiar with a computer-assisted language exam and feel comfortable of talking to a computer. Twelve students who preferred the PBT mentioned the importance of speaking skills. This result was consistent with their responses to Question 4 in Section 1 that indicated their desire to communicate in Japanese. The following comments reflect shared common opinions among the students who were in favor of the PBT.

“Speaking is more important when it comes to language” (Student #5).

“Because I can improve my speaking/listening through this” (Student #34).

“The performance-based assessments can improve to ‘say’ and ‘listen’” (Student #43).

Some students referenced the usefulness of the PBT method.

“I think SE is useful for real life and so not have that much pressure” (Student #39).

“Traditional test doesn’t test what you need in real life” (Student #41).

“It is more versatile and shows talented not just a score” (Student #54).

“I thought traditional was easier, but performance-based helps us learn more effectively” (Student #62).

Some students might have even enjoyed taking the PBT exam, as they reported that it was “interesting.”

Q 2: “Which caused you more anxiety?”

This question was designed to find out students’ test anxiety prior to and/or during the exams. The results show that 42.6% of students (21.3% much more; 21.3%
somewhat more) felt more anxious about the PBT, while 32% (17.3% much more; 14.7%
somewhat more) felt more anxious about the traditional written exam. However, no
significant difference was found between the groups ($z = 1.0773$). Comments from the
students who felt anxious about the PBT include:

“Sometimes I need to do it over and over again” (Student #30).

“I’m afraid of making simple mistake and I really can’t read or repeat fluently in
a certain time” (Student #68).

“I feel that if I don’t get it perfect I won’t do as well” (Student #12).

Since the PBT exam required students to demonstrate their optimal performance and
display their prompt responses, these grading criteria might have caused anxiety. As a
result, they felt somewhat uncertain about whether or not they could achieve a
satisfactory performance. Other comments indicated that students felt inferior because of
their imperfect pronunciation and intonation. Such language concerns might not occur in
the case of a written test. The following are examples of students’ statements regarding
their anxiety:

“It was easy for me to study written Japanese, but I get points off for my accent
on performance-based. It has been very hard for me to get pitches correct”
(Student #64).

“Have to memorize the vocabs, and pronounce it clearly” (Student #31).

On the other hand, students who felt anxious about the traditional written exams
commonly expressed that they had difficulties writing *Hiragana* and *Katakana* precisely.
Q 3: “Which gave you a greater sense of accomplishment?”

This question aims to identify students’ sense of accomplishment following the exams. As a result of this question, 53.3% of students (32.0% much greater; 21.3% somewhat greater) answered that the PBT gave them a greater sense of accomplishment than the traditional exam, while 20% felt the opposite (12.0% much greater; 8% somewhat greater). There was a significant difference between the two groups of responses ($z = 3.6596, p<0.05$). For those who wanted to improve their speaking skills (N=38), half felt accomplished when they were able to do the PBT tasks successfully. In this regard, the PBT is a more effective method for the simulation of a verbal communication.

“Performance was harder to me, so if I did it well, of course I would feel better” (Student #1).

“Though stressful, it is more realistic to using Japanese in real conversation” (Student #13).

“I get more practice and that make me feel more confident” (Student #18).

“I am able to speak Japanese. This makes me excited” (Student #26).

Q 4: “Which test did you think was more challenging?”

This question aims to identify the challenges students faced while preparing for and taking the exams. The results show that 45.4% of students (26.7% much more; somewhat 18.7%) felt that the PBT exam was challenging, while 20% (9.3% much more; somewhat 10.7%) felt that the traditional exam was more challenging. A significant difference was observed between the two groups of responses ($z = 2.8583, p<0.05$). The
results might stem from the mindset that “Speaking is hard.” Even though students were allowed to retry tasks on the PBT, they felt that it was difficult to be a successful speaker. Comments from students who felt difficulty in speaking include:

“Speaking is harder than writing” (Student #35).

“I have a strong accents and I find pitches confusing” (Student #69).

“I have to speak fluently while I’m not fluent at all in Japanese yet” (Student #11).

Furthermore, the PBT required students to read, to listen, and to promptly respond to what they heard and read. These tasks may have contributed to their belief that the PBT exam was challenging. The following are comments from students who thought the PBT was challenging.

“You had to hear, understand, and respond to questions instead of just circling a correct translation.” (Student #6)

“The SE had a more ‘think on your feet’ method to answering.” (Student #46)

“It is more challenging because you should make a good response in such a short time.” (Student #57)

Q 5: “Which style of test motivated you more to learn?”

In response to this question, 50.7% of students (24% much more; 26.7% somewhat more) agreed that the PBT motivated them to learn more than the written exam, while 24% felt the opposite. The difference between the two groups of responses is statistically significant ($z = 2.8098, p<0.05$). In the previous question, approximately half of the students expressed that the PBT exam was more challenging; nevertheless, the number of the students who felt the PBT was more motivating exceeded the number of
the students who felt the traditional written exam was more motivating. This result may be related to a sense of accomplishment after facing the challenge of the exam. Speaking practice was more challenging for students, yet it motivated them to practice more. Thirteen students (34%) explained how speaking practice was important and useful to them.

“Performance based motivated me to practice speaking as well as real-time recording” (Student #4).

“Speaking part is more practical and useful” (Student #68).

Q 6: “Which style of test did you think was fairer?”

This question aims to determine students’ perceptions of the fairness of the two exam types. On Question 6, 37.3% of students (17.3% much fairer; 20% somewhat fair) answered that the traditional written exam was fairer, while 25.3% (9.3% much fairer; 16% somewhat) believed that the PBT exam was fairer. However, there was no significant difference between the groups of responses. The reasons why students thought that the traditional exam was fairer seemed to be related to the test conditions. Due to technical issues in the laboratory, take-home exams were administered instead of taking the PBT at the lab. This change might have reflected unfair test conditions of take-home exams as there was no instructor supervision.

“Everyone got the same test at the same time in the same atmosphere. The other one [PBT] was taken at different times and we could use multiple unfair methods to get good grades” (Student #14).

“Traditional test ask people were all in the situation” (Student #33).
Some students commented that whether or not students had previous learning experience would affect their scores.

“Students who advanced learned JPNS would act much better in the performance-based ones” (Students #30).

“Because someone already know Japanese and people will interrupt you when you having SE test” (Student #73).

Others regarded the written exam’s more straightforward, precise answers as more preferable. For example, Student #11, who felt the written exam was fairer, commented, “It is easy to distinguish between right and wrong answers very clearly, while performance does not have as clear of a standard.” On the other hand, students who felt the PBT was fairer commented, “it [PBT] wasn’t dependent on only one thing” (Student #53) “because it is a diversification test” (Student #55). Student #50 wrote a comment that nicely sums up these opinions: “It does not only measure one’s knowledge. It seems the PBT requires a variety of skills.”

Section 3. This section was designed to investigate how students prepared for the two types of tests. The results are presented in Table 4.
**Table 4. Section 3: Students’ Methods of Preparation for Exam (Study 1)**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practicing writing (15).</td>
</tr>
<tr>
<td></td>
<td>Memorizing vocabulary and grammar (32).</td>
</tr>
<tr>
<td></td>
<td>Practicing on SE (25).</td>
</tr>
<tr>
<td>Q3. Did the performance-based method change the way you studied Japanese in the course?</td>
<td>Yes (%) 66.7</td>
</tr>
<tr>
<td></td>
<td>No (%) 33.3</td>
</tr>
<tr>
<td>Q4. Which method of preparation better aligned with your language learning goals?</td>
<td>Much more PBT (%) 29.3</td>
</tr>
<tr>
<td></td>
<td>Somewhat more PBT (%) 28.0</td>
</tr>
<tr>
<td></td>
<td>Both (%) 18.7</td>
</tr>
<tr>
<td></td>
<td>Somewhat more Paper (%) 12.0</td>
</tr>
<tr>
<td></td>
<td>Much more Paper (%) 12.0</td>
</tr>
<tr>
<td></td>
<td>N/A (%) 0.0</td>
</tr>
<tr>
<td></td>
<td>z-score 3.4449*</td>
</tr>
<tr>
<td>Q5. Which testing method better reflected what you know/can do in Japanese?</td>
<td>28.0</td>
</tr>
<tr>
<td></td>
<td>25.3</td>
</tr>
<tr>
<td></td>
<td>22.7</td>
</tr>
<tr>
<td></td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>z-score 3.2544*</td>
</tr>
</tbody>
</table>

* *p<0.05

**Q 1: “How did you prepare for the traditional pen-paper test?”**

Since the test is in a written format, 33% students indicated that they prepared for the exam by reading and writing. Furthermore, 42.6% of students described their preparation with words like “memorize” or “remember.”

**Q 2: “How did you prepare for the performance-based exam?”**

The PBT required memorization to some extent, but 48% of students, in contrast, used the word “practice” to describe their preparation for the PBT exam. The following are some of their comments:
“I reacted the dialogue over and over again until I could say it without messing up” (Student #11).

“Practice SE again and again until I think I am alright” (Student #39).

“I speak to the mirror and listen to the audio” (Student #58).

Some students practiced through interacting with others.

“Practice speaking to my roommate” (Student #5).

“Talk with other people, do the practice test” (Student #22).

“Ask friends to correct my speaking” (Student #68).

Students seemed to practice speaking more often outside the classroom than they had for the written exams. This transformation corresponded to the purpose of the PBT exam, i.e. to encourage oral practice.

Q 3: “Did the performance-based method change the way you studied Japanese in the course?”

The results show that 66.7% of students (50 out of 75 students) responded “Yes” (see Table 4). Twelve students stated that they spent more time speaking, and 13 students said they focused on speaking. Three students mentioned that they paid attention to pronunciation.

Q 4: “Which method of preparation better aligned with your language learning goals that you put down in Section 1-1?”

In response to this question, 57.3% (29.3% much more; 28% somewhat better) responded that the way they prepared for the PBT was more closely aligned with their learning goals, while 24% (12 % much more; 12% somewhat better) felt the way they prepared for the written exam was closely aligned with their learning goals. The result
showed a significant difference between the two groups of responses \((z = 3.4449, p<0.05)\). Given that the PBT encouraged students to practice speaking as parts of their studies; such preparation were more aligned with students' primary learning goals. Thirty-three out of 43 students (who reported that the PBT was more aligned) seemed to appreciate the opportunities to demonstrate speaking and listening skills on the PBT exam.

“Fluency is the ultimate goal” (Student #4).

“Because I’m more confident to speak in Japanese” (Student #26).

“I want to be able to speak and listen in order to communicate” (Student #32).

“Speaking helped me out greatly in understanding” (Student #47).

This result was consistent with their desire to improve speaking and listening skills inquired in Question 3 in section 1.

**Q 5: “Which testing method better reflected what you know/can do in Japanese?”**

Similarly, 53.3% (28% much better; 25.3% somewhat better) felt that the PBT exam reflected what they know/can do in Japanese, while 22.6% (8% much more; 14.7% somewhat better) felt the written exam reflected what they know/can do. The result also showed a significant difference between the two groups of responses \((z = 3.2544, p<0.05)\). The reasons they cited for this question are also similar to the previous question. Students pointed out that the PBT exam can test multiple skills:

“I could learn almost 4 kinds of skills, listening, reading, speaking, and writing” (Student #51).
“It helps practice different kinds of skills. Therefore, it’s better for the learning progress” (Student #50).

Others touched upon the authenticity of testing:

“It’s more close to real life” (Student #6).

“Speaking a language forces one to think in more applied ways so it makes you learn how to respond naturally in ‘real-life’ situations” (Student #59).

“Language shouldn’t be just writing like in way in tradition” (Student #54).

Section 4. This section concerns students’ PBT task preferences. Results are shown in Table 5. This section has two subsections: Dictation and Speaking tasks.

Table 5. Section 4: PBT Task Preferences (Study 1)

<table>
<thead>
<tr>
<th>Questions I</th>
<th>Yes (%)</th>
<th>No (%)</th>
<th>N/A (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did you practice for dictation in advance?</td>
<td>90.7</td>
<td>9.3</td>
<td>0</td>
</tr>
<tr>
<td>2. Did you like practicing for dictating?</td>
<td>68.0</td>
<td>30.7</td>
<td>1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question II</th>
<th>Monologue (N)</th>
<th>Read aloud (N)</th>
<th>Q &amp; A Reading (N)</th>
<th>Role-play (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Which activities did you like? You can choose as many as you want.</td>
<td>36</td>
<td>45</td>
<td>34</td>
<td>27</td>
</tr>
</tbody>
</table>

Dictation.

Q1: Did you practice for dictation in advance?

In response to this question, 90.6% of students said “Yes”. Dictation was the only task that took place in the classroom. Unlike the other tasks, students were not
allowed to try as many times as they wanted. Therefore, it is assumed that many students practiced for dictation before the test day.

**Q2: Did you like practicing for dictation?**

In response to this question, 68% of students answered that they liked practicing for dictation. The most common reason was that it helped to improve their listening skills.

*Speaking tasks.*

**Q1. Which activities did you like? You can choose as many as you want.**

Many students preferred the read-aloud task to others. This task is simple, but expected to improve vocabulary retention, fluency, pronunciation, and reading comprehension. Actually, Student # 55 commented, “It makes me remember lots of words and makes me practice my speaking.” Student # 11 also wrote, “I liked being able to read a paragraph in Japanese and understand what it said.”

Major findings from Study 1 are as follows. First, Section 1 indicated that students selected the Japanese course to learn the language and culture, and they expected to be able to use Japanese for communicative purposes or interactive discourse. Second, based on the data from Section 2, students perceived that the PBT was challenging, but motivated them to study. Overall, students who conveyed positive perceptions of the PBT outnumbered those who did not approve of the PBT. The results of Section 3 show that students allocated time for speaking, which implies that there was an increase in students’ speaking practice outside the classroom. Furthermore, students felt their preparation methods were more closely related to their learning goals than those for the traditional
written exam. Lastly, the results of Section 4 indicate that most students practiced for
timed-dictation in advance and favored the read-aloud task.

**The Instructor Questionnaire.** The instructor questionnaire aims to identify
instructors’ views on applying the PBT to their classroom instruction. Four instructors
responded to the survey (see Appendix C).

In terms of the implementation of the PBT, all four instructors expressed positive
attitudes. Instructor 1 stated that oral practice became more meaningful, especially tasks
requiring students to speak with their own words in Japanese. Such tasks encouraged
student creativity in their responses. Instructor 2 stated that the PBT was useful for
students to remember target phrases for their own use because there was not sufficient
time for oral practice in the classroom. Instructors 3 and 4 observed that students paid
more attention to their pronunciation than before. As for the changes in classroom
instruction, three instructors stated that they tried emphasizing pronunciation and
intonation, which tends to be less of a focus for a written exam. While three instructors
admitted that the PBT had advantages with improving communicative skills, two
instructors expressed concerns about deemphasizing writing skills. Instructor 2 was
doubtful about the students’ ability to demonstrate their oral skills in a natural
conversation without rehearsing.

Based on the adjustment of their teaching methods, the PBT obviously
influenced what skills the teachers focused on in the classroom, especially in relation to
language production and delivery. The PBT preparation also appeared to be in line with
learning through the Communicative Approach, which served as a course objective.
Study 2

A total of 34 out of 35 students of Japanese 102 participated in the student survey. In order to compare the two exams, students experienced two written exams (the midterm and final exam) and five PBT exams. Since the results are generally similar to those of Study 1, we will focus on the items that are different from Study 1.

Results of the Student Questionnaire. Tables 6 to 9 report results from Likert scale items.

Section 1. This section aims to identify students’ purpose(s) in taking the Japanese course (see Table 6).

Table 6. Section 1: The Skills that Students Want to Improve the Most (Study 2)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. What is your purpose in taking the Japanese language course?</td>
<td>To learn the language (10).</td>
</tr>
<tr>
<td></td>
<td>To fulfill a language requirement (8).</td>
</tr>
<tr>
<td></td>
<td>An interest in culture/subculture (8).</td>
</tr>
<tr>
<td>Q2. Why did you choose the Japanese language over the others?</td>
<td>An interest in the Japanese language (8).</td>
</tr>
<tr>
<td></td>
<td>An interest in culture/subculture (19).</td>
</tr>
<tr>
<td>Q3. Among the 4 skills of Speaking, Listening, Writing, and Reading, which skill do you want to improve the most?</td>
<td>Speaking (%) 52.9</td>
</tr>
<tr>
<td>Q4. What is/are your goal(s) for taking the Japanese course?</td>
<td>To have a basic conversation (15).</td>
</tr>
<tr>
<td>What do you want to be able to do in Japanese by the end of the semester?</td>
<td>To be able to speak fluently (10).</td>
</tr>
<tr>
<td></td>
<td>To gain a basic understanding (9).</td>
</tr>
</tbody>
</table>

Note. The reason why the percentages for Q3 do not add up to 100% is that one student circled more than one choices.

Q 1: “What is your purpose in taking the Japanese language course?”

Students’ purpose(s) were very similar to Study 1. Students were interested in learning Japanese language for the following reasons: being able to speak, having
conversation, and studying in Japan. Others mentioned their language requirement and an interest in culture.

**Q 2: “Why did you choose the Japanese language over the others?”**

In response to this question, 55.8% of the students stated that they were interested in Japanese culture and subculture (e.g. anime, music). It is highly likely that the wide variety of cultural entertainment is one of the primary motivations to choose Japanese (Liu & Shibata, 2008).

**Q 4: “What is/are your goal(s) for taking the Japanese language course? What do you want to be able to do in Japanese by the end of the semester?”**

The most frequent responses were the desire to have a conversation and to speak fluently. The results, again, were similar to Study 1. Being able to speak and have an interaction in Japanese appear to be the main goals for students.

**Section 2.** This section aims to identify students’ perceptions toward the two test formats: traditional written exams (which included midterm and final exam) versus the PBT. Survey participants’ responses are summarized in Table 7.
Table 7. Section 2: Students’ Perceptions toward Two Test Formats (Study 2)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Much more PBT (%)</th>
<th>Somewhat more PBT (%)</th>
<th>Both (%)</th>
<th>Somewhat more Paper (%)</th>
<th>Much more Paper (%)</th>
<th>N/A (%)</th>
<th>z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Which style of evaluation/test did you like better?</td>
<td>38.2</td>
<td>32.4</td>
<td>5.9</td>
<td>20.6</td>
<td>2.9</td>
<td>0.0</td>
<td>3.2344*</td>
</tr>
<tr>
<td>Q2. Which caused you more anxiety?</td>
<td>11.8</td>
<td>8.8</td>
<td>26.5</td>
<td>29.4</td>
<td>23.5</td>
<td>0.0</td>
<td>2.3756*</td>
</tr>
<tr>
<td>Q3. Which gave you a greater sense of accomplishment?</td>
<td>29.4</td>
<td>20.6</td>
<td>32.4</td>
<td>11.8</td>
<td>5.9</td>
<td>0.0</td>
<td>2.4948*</td>
</tr>
<tr>
<td>Q4. Which test did you think was more challenging?</td>
<td>14.7</td>
<td>11.8</td>
<td>26.5</td>
<td>38.2</td>
<td>8.8</td>
<td>0.0</td>
<td>1.4422</td>
</tr>
<tr>
<td>Q5. Which style of test motivated you more to learn?</td>
<td>35.3</td>
<td>11.8</td>
<td>23.5</td>
<td>8.8</td>
<td>20.6</td>
<td>0.0</td>
<td>1.2014</td>
</tr>
<tr>
<td>Q6. Which style of test did you think was fairer?</td>
<td>5.9</td>
<td>20.6</td>
<td>41.2</td>
<td>14.7</td>
<td>17.6</td>
<td>0.0</td>
<td>1.5828</td>
</tr>
</tbody>
</table>

*p<0.05

Q1: “Which style of evaluation/test did you like better?”

In response to this question, 70.6% of students (38.2% much better; 32.4% somewhat better) indicated that they prefer the PBT exam, as compared to 23.8% (2.9% much better; 20.6% somewhat better) who prefer the traditional exam. The z-test showed a significant difference between the two groups of responses (z = 3.2344, p<0.05).

Looking at students’ first language backgrounds, we found higher percentages from both Chinese and English speakers were obtained than in Study 1: 10 out of 13 Chinese speakers (76.9%) answered that they preferred the PBT either much or somewhat better, while 17 out of 20 English speakers (85%) answered that they preferred either much or somewhat better. The common reasons for these preferences were that they could...
improve speaking skills. This result is reasonable because 52.9% of students claimed that they were most interested in improving their speaking skills, as discussed in Question 3 in Section 1. However, a few students reported that it was easier to prepare for the PBT than the written exam.

“Not so much need to be memorized.” (Student #23)

“Easier to prepare the exam, also the skill is more useful.” (Student #30)

“Because it’s easier and more effective (can practice speaking).” (Student #34)

A possible implication of these comments is that they might feel that the grading criteria for written exams are more particular than the grading criteria for the PBT, which has a more analytical scale.

Q 2: “Which caused you more anxiety?”

This question found that 52.9% of students (23.5% much more; 29.4% somewhat more) felt more anxious about the traditional written exam, while 20.6% of them (11.8% much more; 8.8% somewhat more) felt more anxious about the PBT. Statically, there was a significant difference between the two groups of responses (z = 2.3756, p<0.05). Students expressed their reasons by indicating that the written exams required a lot of memorization, and the questions were hard. In the written format, students are expected to write answers with the correct spelling and legible penmanship. For example, in the written test, if a diacritic marker on a letter is missing, points were deducted. These grading requirements might cause them anxiety. Another possible reason is that some of the tests covered a wide range of materials than in Study 1; in Study 2, students took written exams for the midterm and final exams. Those are the possible
reasons why students reported more anxiety for the written exam. The following are some free-form responses to this question.

“The writing test is really hard” (Student #30).

“It needs me to memory lots of words” (Student #33).

“The traditional exams have shorter time periods and are more harshly graded” (Student #15).

“We have to know exactly how to write and maybe the atmosphere of traditional test” (Student #32).

Q 3: “Which gave you a greater sense of accomplishment?”

Fifty percent of students (29.4% much greater; 20.6% somewhat greater) responded that PBT gave them a greater sense of accomplishment, while 17.7% of them referenced the written exams. The result of a z-test shows a significant difference between the two groups of responses. Students frequently mentioned the opportunity to demonstrate speaking skills and to improve their speaking ability.

“Because I was able to demonstrate my speaking” (Student #14).

“I could see my communication skills improve” (Student #2).

Since students were allowed to retry tasks until they were satisfied with their responses, some felt that their task processing speed became faster.

“Being able to answer all SE questions in a limited time” (Student #18).

“I feel like I got way better way faster” (Student #12).

The students who noticed improvements in their speaking and processing skills felt a sense of achievement. Providing a sense of accomplishment is one of the important factors that motivates learners to study and positively influences their learning.
Q 4: “Which test did you think was more challenging?”

Forty-seven percent of students (8.8% much more; 38.2% somewhat more) responded that the written exam was challenging, while 26.5% of them (14.7% much more; 11.8% somewhat more) marked that the PBT exam was difficult. However, a significant difference was not observed between the two groups of responses. The comments that the students gave were similar to Question 2 in Section 2. Students indicated that they felt overwhelmed by the written exam’s requirement to memorize vocabulary and grammar, as well as to write precisely. Two factors possibly contributed to these responses: first, students believed that they are scored harshly on the written exam. Second, students assumed that they are expected to demonstrate comprehensive knowledge on multiple chapters. On the other hand, two students who felt the PBT was challenging said that it required prompt responses. Five students mentioned that they are weak speakers and listeners.

Q 5: “Which style of test motivated you to learn more?”

In response to this question, 47.1% of students (35.3% much more; 11.8% somewhat more) felt that the PBT motivated them to learn, while 29.4% of them (20.6% much more; 8.8% somewhat more) felt the written exam was more motivating. However, there was no significant difference between the two groups of responses. The most frequent comments expressed that students can improve their speaking and listening skills by practicing for the PBT. A few students stated that these skills could improve learning retention.

“I was able to retain more through experience/practice rather than just read/written knowledge” (Student #2).
“I practice speaking and learning vocab much faster” (Student #12).

“I can more focus on learning skill rather than just have study for exam”

(Student #30).

Although learning strategies vary by individuals, oral production inevitably promotes foreign language conversation skills.

Q 6: “Which style of test did you think was fairer?”

The highest percentage of students (41.2%) responded that both tests were fair, with the second largest percentage (32.3%) feeling the written exam is fairer. Only 26.5% felt the PBT was fairer. There was no significant difference between the last two groups of responses. Students’ perceptions of fairness varied. The frequent comments of those students who indicated that both tests were fair revealed that their understanding of fairness depended on how much effort one put forward. For example, Student # 19 wrote, “I feel like both depend on how much effort is put towards learning the materials.”

Students who felt the written exam was fairer criticized the PBT conditions and grading criteria because it allowed students to have “infinite retries” (Student #5). Additionally, Student # 31 commented, “PBT is somewhat too free for tests.” This means that students were allowed to choose time and location for the PBT since it was a take-home exam. Some students pointed out some unexpected technical issues have occurred. Others preferred the restrictive grading criteria rather than allowing for “more than one answer on the PBT” (Student #8). However, those who favored the PBT formats referenced its conditions as flexible and positive. One student commented that the PBT was fair because it allowed for “more time to relax and think” (Student #15); another noted that its grading criteria are not based on “right or wrong” (Student #2) answers. This perception
is correct because the PBT does not contain any true or false questions or multiple choice items. Some tasks, as in the case of “Monologue” ask students to record their opinions or personal experiences. One noted that the PBT is better because it is “closer to what we did in class” (Student #14); another mentioned that writing a correct answer does not mean students “know how to say it” (Student #34). Since many students have experienced standardized exams in their lives, they expect their test performance to be assessed using theses more traditional measures. However, some students seem to realize that those standardized exams do not contribute to their speaking abilities. Lastly, it should be noted that more than 40% of students expressed that both tests were equally fair.

Section 3. This section investigates the difference between students’ preparation for both tests (see Table 8).
**Table 8. Section 3: Students’ Method of Preparation for Exam (Study 2)**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. How did you prepare for the traditional pen-paper test?</td>
<td>Reviewing materials (15)</td>
</tr>
<tr>
<td></td>
<td>Memorizing vocabulary and grammar (12).</td>
</tr>
<tr>
<td></td>
<td>Read the book (10)</td>
</tr>
<tr>
<td></td>
<td>Reviewing grammar (8).</td>
</tr>
<tr>
<td>Q3. Did the performance-based method change the way you studied Japanese in the course?</td>
<td>Yes (%) 73.5</td>
</tr>
<tr>
<td></td>
<td>No (%) 26.5</td>
</tr>
<tr>
<td>Q4. Which method of preparation better aligned with your language learning goals?</td>
<td>Much more PBT (%) 41.2</td>
</tr>
<tr>
<td></td>
<td>Somewhat more PBT (%) 23.5</td>
</tr>
<tr>
<td></td>
<td>Both (%) 11.8</td>
</tr>
<tr>
<td></td>
<td>Somewhat more Paper (%) 8.8</td>
</tr>
<tr>
<td></td>
<td>Much more Paper (%) 11.8</td>
</tr>
<tr>
<td></td>
<td>N/A (%) 2.9</td>
</tr>
<tr>
<td></td>
<td>z-score 3.1849*</td>
</tr>
<tr>
<td>Q5. Which testing method better reflected what you know/can do in Japanese?</td>
<td>Much more Paper (%) 26.5</td>
</tr>
<tr>
<td></td>
<td>Somewhat more Paper (%) 17.6</td>
</tr>
<tr>
<td></td>
<td>Both (%) 35.3</td>
</tr>
<tr>
<td></td>
<td>Somewhat more Paper (%) 8.8</td>
</tr>
<tr>
<td></td>
<td>Much more Paper (%) 11.8</td>
</tr>
<tr>
<td></td>
<td>N/A (%) 0.0</td>
</tr>
<tr>
<td></td>
<td>z-score 1.7836</td>
</tr>
</tbody>
</table>

*p<0.05

**Q 1:** “How did you prepare for the traditional pen-paper test?” Their preparation for the written exam commonly included a review of materials, reading the textbook and memorizing vocabulary and grammar, which is very similar to Study 1.

**Q 2:** “How did you prepare for the performance-based exam?”

Some described that their preparation for the PBT exam included practicing speaking, but not as often as in Study 1. What was observed was that many more students used the word “grammar” than in Study 1. This may be because Japanese 102 introduces more grammar rules than Japanese 101. These results suggest that these students practiced speaking while paying attention to grammar rules, an ideal habit to get into.
Q 3: “Did the performance-based method change the way you studied Japanese in the course?”

The result shows that 73.5% of students (25 out of 34 students) responded “Yes”, and 13 students commonly stated that they practiced speaking more; however, not everyone showed a positive attitude to this question. Two students gave negative comments:

“I didn’t study” (Student #16).

“SE meant I didn’t have to study” (Student #17).

These students preferred the written exam and claimed that this knowledge-based exam made them feel a sense of accomplishment.

Q 4: “Which method of preparation better aligned with your language learning goals that you put down in Section 1-1?”

The result shows that 64.4% of students (41.2% much more; 23.5% somewhat better) indicated that PBT aligned with their language learning goals while 20.6% (11.8% much more; 8.8% somewhat more) cited the written exam (see Table 8). This is significant because the PBT provides more opportunities to improve speaking and listening ($z = 3.1849, p<0.05$). The following are pertinent comments from students:

“All-around, it allowed me more growth. Speaking and listening required more effort” (Student #1).

“I got a lot of speaking and listening practice out of class” (Student #12).

“The PBT examine my fluency and accuracy” (Student #30).

This result shows that students seek to demonstrate communication skills on the exam.
Q 5: “Which testing method better reflected what you know/can do in Japanese?”

As shown in Table 8, 44.1% of students (26.5% much better; 17.6% somewhat better) perceived that the PBT exam better reflected their overall Japanese skills, while 20.6% of them (11.8% much better; 8.8% somewhat better) perceived that the traditional written exam was more representative. There was no significant difference between the two groups of responses. Some pointed out that PBT is a more valid way to test language than the written format.

“Allowed for more natural learning.” (Student #11)

“I don’t think you can really test language with traditional method. (Student #14)

“Speaking is always the best evidence than can reflect your result.” (Student #33)

The PBT was designed to promote oral proficiency. Students’ comments show that this is a practical assessment methods to promote oral language development.

Section 4. This section concerns students’ preference of the PBT tasks. The result is shown in Table 9. This section has two subsections: Dictation and Speaking tasks.
Table 9. Section 4: PBT Task Preferences (Study 2)

<table>
<thead>
<tr>
<th>Questions I</th>
<th>Yes (%)</th>
<th>No (%)</th>
<th>N/A (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did you practice for dictation in advance?</td>
<td>88.2</td>
<td>11.8</td>
<td>0</td>
</tr>
<tr>
<td>2. Did you like practicing for dictating?</td>
<td>50.0</td>
<td>47.1</td>
<td>3.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question II</th>
<th>Monologue (N)</th>
<th>Read aloud (N)</th>
<th>Q &amp; A Reading (N)</th>
<th>Role-play (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Which activities did you like? You can choose as many as you want.</td>
<td>21</td>
<td>21</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

Dictation.

Q1: Did you practice for dictation in advance?

In response to this question, 88.2% of students reported “Yes”.

Q2: Did you like practicing for dictation?

In response to this question, 50% students answered that they liked practicing for dictation. Students who liked practicing for dictation commented that dictation helped to improve their listening and writing skills. On the other hand, the most frequent comment from students who did not like practicing dictation was that it was hard for them to memorize all sentences.

Speaking tasks.

Q1. Which activities did you like? You can choose as many as you want.

Many students chose the monologue and read-aloud tasks. As in Study 1, the read-aloud task was a favorite among students. This task is simple, yet it seems to empower many students. Student #13 stated, “I can practice my reading skills, comprehension and fluency all at once.” Additionally, Student #7 commented, “[it is]
kind of fun to read it fast.” Monologue tasks require students to create their own answers when prompted with a given topic or a situation. Student #6 who chose both the monologue and the read-aloud stated, “They were completely up to me to make up or it was simple to figure out what I was supposed to do or say”. Student #9 wrote, “Monologue I can make up what I want.” This task helped students organize their answers using grammar that they learned in each chapter. On the semi-direct testing, the tasks that require test-takers to describe and/or report on things, as elicited in the monologue task, are likely to be criticized because of its unauthentic nature (Clark, 1979). Anticipating this criticism, this task was designed with situations that naturally require monologic productions (e.g. when introducing themselves, describing their vacations, leaving messages on a voice mail system). Considering these situations, students #18 chose the monologue tasks because he could “get to practice real life scenarios”.

Overall, based on these results, students’ perceptions towards the PBT were relatively positive. However, it should be noted that there were a small number of supporters of the written exam, and they pointed out some drawbacks of PBT. This will be discussed in a later section.

The Instructor Questionnaire. Four instructors responded to the survey. Instructor 1 stated that PBT is an effective assessment method to improve students’ oral proficiency, and students participated in the classroom activities more actively. Instructor 2 stated that she could monitor student’s individual pronunciation through SE. Instructor 3 noticed that students were more comfortable speaking, and their in-class participation became active. As for negative comments, three of them mentioned that students did not
seem to prepare for the PBT tasks beforehand even if the tasks were made accessible online. Instructor 4 heard from a student that studying is not necessary to achieve a high grade on the PBT. Instructor 1 claimed that some supplemental materials are needed to promote writing skills. Instructor 2 expressed uncertainty about vocabulary retention because she was unable to measure the amount of vocabulary that students had acquired and retained by the PBT format. Two instructors suggested combining the written exam and the PBT exam. When asked about the changes in their teaching methods, three of them gave similar statements that they used class time to brainstorm with students the range of possible, appropriate answers to PBT questions.

Discussion

The following research questions will be answered by summarizing student questionnaires in both Study 1 and Study 2.

Research Question 1: Do students prefer the style of traditional written testing or PBT testing? The students were asked, “Which style of evaluation/test did you like better?” (Question 1 in Section 2). In response to this question, 52% (Study 1) and 70.6% (Study 2) of students reported their preference for the PBT exam with either much better or somewhat better, while 22.7% (Study 1) and 23.8% (Study 2) prefer the traditional written exam (see Table 3 and 7). Statistically speaking, a significant difference was observed in both Study 1 and Study 2. Students’ comments, that speaking skills are the most important and the PBT provides more opportunities for oral production, support this conclusion. From these qualitative and quantitative data, it is confirmed that the PBT format was favored over the traditional written exam.
Research Question 2: Considering both traditional written testing and computer-assisted PBT testing, which of the two test formats do the students think better aligns with their learning goals? Questions 3 and 4 in Section 1 concerned the skills that students most wanted to improve and their goals for taking the Japanese course, respectively. On Question 3, 50.6% (Study 1) and 52.9% (Study 2) of students valued “speaking” the most (see Table 2 and 6). As for learning goals, the majority of students indicated a desire to be able to carry on conversations and become fluent speakers. These findings are in line with those of Alalou (2001) and Thomas (2010), who found speaking skills and communication were considered to be primary goals by students in learning languages.

In Question 1 and 2 in Section 3, students were asked to write about their preparation for the written exam and for the PBT exam. Students’ preparation for the exams were quite different. For the written exams, students were most likely to adopt memorization strategies by reading the textbook and reviewing vocabulary and grammar constructions, whereas they were most likely to spend time practicing speaking for the PBT. As shown by these responses, 57.3% (Study 1) and 67.7% (Study 2) of the students indicated that their preparation for the PBT aligned either much better or somewhat better with their language learning goals than did their preparation for the written exams. The two comparisons were both statistically significant (see Tables 4 and 8). From this, we can conclude that the students perceived the PBT to be better aligned with their learning goals.

Research Question 3: What are the students’ perceptions of the two test formats? In Study 1, most students selected the PBT as the test that gave them greater
anxiety (Q2), accomplishment (Q3), challenge (Q4), motivation (Q5) and reflection (Q8), as shown in Tables 3 and 4. These results can be attributed to the orientation of the PBT tasks and its grading criteria. In contrast to the written exams, the PBT has integrated tasks in which students demonstrate reading, listening, and speaking skills. Furthermore, in order to meet the PBT criteria, students must respond promptly with fluent speech, grammatical accuracy, and correct pronunciation. These features might have contributed to their thinking that the PBT was challenging. However, since most students’ learning goals were to improve speaking skills and hold a basic conversation, the PBT served as an opportunity to practice speaking. Even though they felt challenged, the students were rewarded with a sense of accomplishment upon completing the exam.

In Study 2, perceptions of the PBT were more positive than those of the written exams (see Table 7 and 8). Similar to Study 1, most students selected the Japanese course for improving communication. In this regard, the PBT is seen as a more effective method to meet their communicative goals compared to the traditional written exams. In addition, they felt a sense of accomplishment when they could demonstrate their communicative skills. Based on the results of the surveys, the PBT increased students’ confidence in speaking.

Research Question 4: Did the PBT format change the way they studied Japanese in the course? Students were asked, “Did the performance-based method change the way you studied Japanese in the course?” In response to this question, 66.7% (Study 1) and 73.5% (Study 2) of the students answered “Yes” (See Table 4 and 8). In Study 1, 12 out of 50 students who answered “Yes” stated that they spent more time speaking, and 13 students commonly stated that they “focused on speaking”. In Study 2,
13 out of 25 students who answered “Yes” stated that they practiced speaking more.

These comments suggest that students were willing to allocate more time to practice speaking outside the classroom in preparation for the PBT. In the classroom, as some instructors observed, students became more active participants, and more attention was given to their prosodic accuracy. Since the PBT emphasizes oral proficiency, students’ speaking practice became more frequent; it follows, then, that this test encouraged additional language practice. Therefore, it is concluded that the PBT brought about beneficial changes to students’ attitudes towards learning.

**Research Question 5: What effect, if any, did the PBT assessment have on student learning?** This study found several positive attitudes toward the PBT. First, most students welcomed the PBT because it aligned with their learning goals better than written testing. Since the PBT was designed to develop students’ oral skills, it encouraged students to allot more time to speaking practice. Furthermore, the PBT raised both students’ and instructors’ phonetic awareness and encouraged students to speak. This led to students’ active participation in class. Considering these findings, the PBT had desirable effects on student learning by providing opportunities to practice speaking. In summary, the PBT had a positive effect, especially on students who valued oral communication.

**Concerns.** Although the PBT had a positive effect on most students, minor concerns were raised. In this section, these concerns are discussed.

*“PBT is easy.”* This type of comment was seen in Japanese 102 students’ responses. Actually, a few students wrote that they did not prepare for the PBT because studying was unnecessary in order to get a good score. Student #14 wrote, “Felt it was
much, much easier to pass PB tests.” Students #17 commented that the PBT “encouraged laziness.” Possible reasons for these remarks would be: (1) instructors might have been too lenient in their grading, (2) such students were already accomplished speakers, (3) they thought it was easier than the memorization of vocabulary, grammar, and spelling required by the written exams, or (4) they did not practice speaking as a part of their studying for the PBT. Looking at these students’ exam results may clarify these assumptions. If the first reason is true; however, it is important for instructors not to be too lenient in grading. Students need to be informed that unnatural speech will result in a deduction of points. In addition, the current PBT rating criteria may have to be improved based on audio samples collected from the past student responses to provide instructors with more detailed guidelines for assessing students' performances. Along with the new criteria, instructors can carefully choose audio samples and share the recordings in the classroom. Such recordings might represent a model of “A-level performance” or “B-level performance;” this would provide students a clearer understanding of their instructors’ expectations and evaluation. Also, instructors provide each student with explicit feedback suggesting areas of improvement both before and after the PBT.

**Limited feedback.** On the PBT, students’ performances were recorded on a score sheet. However, some students implied that receiving a mere number does not help them recognize their mistakes and areas of improvement as mentioned earlier. Student #14 stated, “Traditional showed me what I know and can improve on, performance only confused me and gave me no feedback on what I could do better.” Student #24 wrote, “[On the written exams] I can see where exactly I got things right or wrong.” Therefore,
comments, advice, and praise from teachers are necessary to enhance students’ learning and self-assessment (Black & William, 1998a, 1998b; Tuttle & Tuttle, 2013).

**Unfairness.** Students who preferred the traditional written exams perceived that the PBT was unfair because it allowed them to have infinite retries. However, this is one of the purposes of PBT, meaning that instructors will measure how well students perform rather than whether they could simply do the tasks or not. Therefore, students need to be reminded of these grading criteria clearly in class, especially for those who did not prepare for the PBT.

Technical issues were another factor that contributed to the perceived unfairness of the PBT. Student #6 commented that the written exams were fairer because he doesn’t “have to worry about everything submitting [correctly] or timed tests or any digital malfunctions.” If Speak Everywhere fails to receive students’ recordings, they have to do the same task once again. This problem was alleviated in Study 2; however, tackling such technical issues is crucial in the administration of computer-assisted assessment.

**Summary of Chapter 4**

This chapter presented research findings and discussed the possible reasons for the obtained results. In the final chapter, pedagogical implications and directions for future studies are provided.
CHAPTER 5. CONCLUSION

Introduction

The goal of this study is to recommend the Performance-Based Test (PBT) as a new assessment option in foreign language courses. In order to examine the value of the PBT, this study conducted a survey to investigate students’ perceptions of the following two test formats: a traditional written test and the computer-assisted PBT, implemented via the online platform *Speak Everywhere*. The survey examined students’ test format preferences, their perceptions of a formats’ value and the alignment of the formants with their learning goals. This final chapter will summarize the research findings, discuss the pedagogical implications of the study, and suggest directions for future studies.

Research Findings

In Study 1, 75 out of 80 students and four instructors of Japanese 101 participated in the survey in this Fall 2014 survey. Section 1 results from the student questionnaire revealed that students selected the Japanese course to primarily learn the language, culture, and improve their speaking skills for communicative purposes. In Section 2, a comparative analysis of the PBT and the traditional written exam, significant differences were observed regarding students’ test format preferences, sense of accomplishment, challenge, and motivation. The data revealed that 52% of the students favored the PBT
format because they recognized the importance of speaking skills, whereas 22.7% favored the written exams. Students perceived that the PBT was challenging, and felt motivated to study for the exam. In Section 3, statistical analysis showed a significant difference in the degree to which each test format is aligned to students learning goals, as stated in their related reflections on acquired knowledge and ability. Students indicated that their preparation methods for the PBT were more closely related to their learning goals than those for the written exams. Furthermore, the administration of the PBT increased opportunities for speaking practice and encouraged students to value precise pronunciation and fluent speech. As for the PBT tasks, Section 4 showed that most students practiced for timed dictation in advance and favored the read-aloud task. On the whole, Study 1 results showed that the implementation of the PBT positively benefitted most students.

With respect to the results from the instructor questionnaire, all four instructors wrote positive comments regarding the implementation of the PBT. Each shared that the PBT is a useful method for students to improve their communication skills. Three instructors stated that they emphasized pronunciation and intonation. A concern was expressed about making light of an emphasis on writing skills, given that the administration of the PBT changed instructors’ focus to oral fluency.

Regarding the student questionnaire in Study 2, 34 out of 35 students and four instructors of Japanese 102 completed the Spring 2015 survey. Overall, very similar results to Study 1 were obtained. Section 1 indicated that students sought the development of speaking skills; being able to speak and have a conversation in Japanese were their primary goals. In Section 2, their positive perceptions toward the PBT were
significant compared to the traditional written exams. Seventy percent of students preferred the PBT format, and students who stated that the PBT motivated them to learn and helped them accomplish their goals outnumbered those who supported the written exam. In Section 3, most students felt that preparation required by the PBT were aligned with their learning goals.

The instructor questionnaire in Study 2 revealed various opinions about the PBT. Two instructors observed that students became more active participants. One stated that the PBT is an effective assessment method to improve students’ oral proficiency; another stated that she could monitor student’s individual pronunciation through Speak Everywhere. However, one expressed a concern about decreased opportunities for writing, and another questioned vocabulary retention. Three reported that they spent class time discussing their expectations for the range of possible student responses allowed for by the PBT.

Based on the results of the student questionnaire in Studies 1 and 2, we can say that students expressed a generally favorable attitude toward the PBT. Moreover, this study suggests that the PBT is more effective than the traditional written exam given students’ speaking and listening needs and goals.

**Pedagogical Implications**

When the primary curricular goal is to promote language learners’ proficiency through the communicative method, both the day-to-day classroom instruction and assessment method should reflect this priority. Furthermore, frequent oral assessments are ideal because it helps gauge whether or not students can verbally utilize the new
vocabulary and grammar introduced in each chapter. Semi-direct methods like the PBT make it possible to provide efficient and frequent administration instead of face-to-face interviews. These methods provide a more precise indication of students’ oral production than the written exams. However, there were some concerns raised by the students in their comments, as mentioned in Chapter 4 (e.g. “PBT is easy”). It is, therefore, necessary for instructors to periodically review their expectations for PBT performance with in-class audio samples and rubric scores.

Finally, it needs to be emphasized that the curriculum goals, instruction, assessment and students’ needs, should be integrated components of pedagogy. Teaching, assessment, and learning should not be discrete. Ideally, this integration brings about a positive pedagogical outcome, in particular, a positive washback effect. When it comes to assessment administration, the following characteristics should be considered: construct relevance, face validity, reliability, and feasibility. Both qualitative and quantitative data from this study will hopefully be an essential resource for future test development and Japanese language instruction at Purdue University.

Implications for Future Research

For future studies, the following topics can be investigated:

**Vocabulary and grammar retention.** The present study focused on students’ perceptions of the two test formats. Future studies should investigate students’ vocabulary and grammar retention by comparing those who took the PBT to those who only took the written exams. Since the PBT targets oral proficiency, students practiced speaking more frequently. On the PBT, all questions used in the practice mode were
identical to those in the test mode. Therefore, students might largely rely on memorization. A resulting concern is whether or not students acquired vocabulary and grammar items introduced in the chapter. In fact, one instructor expressed uncertainty about the PBT format and vocabulary retention. Future studies should address these concerns with an investigation of students’ oral proficiency progress and a study of how the PBT promotes students’ vocabulary and grammar retention via oral production and practice. Such investigations would also validate that the PBT is an effective test to be utilized as an achievement test.

**Comparison to face-to-face interviews.** Although the PBT was designed to develop students’ oral skills, one of the drawbacks of the semi-direct methods is the absence of an interlocutor in conversation. Therefore, future studies should investigate the application of students’ communication abilities acquired through the PBT method in real communication settings, such as face-to-face interviews. Such investigations would validate that the PBT practice is useful in real-time communication. These interviews should measure students’ proficiency and more holistic communication skills in real time (e.g. requesting for repetition of a statement or phrase). Moreover, this study would gauge if students are able to apply their grammar knowledge acquired from the PBT-based curriculum in real communication. It would also be interesting to compare these students’ performance to those students who took only the written exams.

**Concluding Remarks**

This study found that students positively perceived the implementation of the PBT and acknowledged the PBT method as more aligned with their learning goals than
the traditional written exam. The PBT was also aligned with course objectives. As research suggests, a semi-direct testing using a CALL system such as *Speak Everywhere* can provide efficient, feasible and frequent assessment while maintaining formality and uniformity in students’ responses. This method is an effective alternative to the written exam as an ongoing assessment, since it incorporates reading, writing, listening, and speaking in Japanese. The findings of this study suggest that the PBT is an effective and practical oral assessment option. Japanese instructors at Purdue University continue to use the PBT, and its implementation has expanded to online and advanced courses. It is, therefore, my hope that Japanese instructors continue to use the PBT in the Japanese program.
LIST OF REFERENCES


Clark, J. L., & Li, Y. C. (1986). Development, Validation, and Dissemination of a Proficiency-Based Test of Speaking Ability in Chinese and an Associated Assessment Model for Other Less Commonly Taught Languages. 1-34.


Kondo, B.K., (2012). Nihongo kyoshi notameno hyouka nyumon [Introduction to assessment for Japanese language teachers], Kuroshio Publisher


Boston, Mass.: Keinle & Heinle.


Appendix A  PBT Rating Criteria

PBT  評価基準
PBT Rating Criteria

1) Promptness of a response (5pts)

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Japanese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1秒以内に躊躇することなくすっと言える。</td>
<td>1秒以内に躊躇することなくすっと言える。</td>
<td>The student can start speaking within 1 second without hesitation.</td>
</tr>
<tr>
<td>4</td>
<td>2秒をおいてから言う。</td>
<td>2秒をおいてから言う。</td>
<td>The student takes 2 seconds to speak.</td>
</tr>
<tr>
<td>3</td>
<td>3秒をおいてから言う。</td>
<td>3秒をおいてから言う。</td>
<td>The student takes 3 seconds to speak.</td>
</tr>
<tr>
<td>2</td>
<td>4秒間をおいてから言う。</td>
<td>4秒間をおいてから言う。</td>
<td>The student takes 4 seconds to speak.</td>
</tr>
<tr>
<td>1</td>
<td>5秒以上かかる。</td>
<td>5秒以上かかる。</td>
<td>The student takes more than 5 seconds to speak.</td>
</tr>
</tbody>
</table>

2) Fluency

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Japanese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>問題なくスラスラいえる。聞く側の負担が全くない。</td>
<td>問題なくスラスラいえる。聞く側の負担が全くない。</td>
<td>The student demonstrates excellent, smooth, and continuous speech without pauses. Raters have no difficulty understanding.</td>
</tr>
<tr>
<td>4</td>
<td>1〜2回つっかえるが、聞く側の負担がほとんどない。</td>
<td>1〜2回つっかえるが、聞く側の負担がほとんどない。</td>
<td>The speaker uses a few attempts to speak correctly, but shows smooth speech. Raters require very little effort to understand.</td>
</tr>
<tr>
<td>3</td>
<td>全体的には理解できるが、速さが悪く、聞き手に少し負担を感じる。</td>
<td>全体的には理解できるが、速さが悪く、聞き手に少し負担を感じる。</td>
<td>Overall, the student can show comprehensible speech, but at an inconsistent speed. Raters require some effort to understand.</td>
</tr>
<tr>
<td>2</td>
<td>話している内容はわかるものの、聞き手に負担がかかる。練習不足と感じ る。</td>
<td>話している内容はわかるものの、聞き手に負担がかかる。練習不足と感じ る。</td>
<td>The student demonstrates comprehensible speech, but raters require much effort to understand. The student needs more practice.</td>
</tr>
<tr>
<td>1</td>
<td>単語ごとにつっかえたり、間が長かったり、非常に聞き辛く、聞き側にか なり負担がかかる。全く練習してないと感じ る。</td>
<td>単語ごとにつっかえたり、間が長かったり、非常に聞き辛く、聞き側にか なり負担がかかる。全く練習してないと感じ る。</td>
<td>The student show disjointed speech and uses pauses between every word. Raters require significant effort to understand. The student obviously did not practice at all.</td>
</tr>
</tbody>
</table>
3) Accuracy for read-aloud

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>間違いない。 The student shows no reading mistakes.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>読み方の間違いが全体で数箇所ある。 The student shows a few reading mistakes.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>全体の半分くらい読み間違いている The student shows a moderate number of reading mistakes.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ほぼ読み方が間違いている。 The student shows many reading mistakes.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>すべて間違いっている。内容がわからない。 The student shows totally inadequate reading ability. It is completely incomprehensible.</td>
<td></td>
</tr>
</tbody>
</table>

4) Pronunciation / Intonation

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>アクセントは多少あるが、十分理解できる。 It is completely comprehensible though the student has some accent.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>濁音、促音が正しく発音されていないなど直したい箇所が少しある。 The student needs some corrections for a voiced sound, a double consonant, etc.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>濁音、促音が正しく発音されていないなど直したい箇所が目出つ。 The student needs significant corrections for a voiced sound, a double consonant, etc.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>発音がわるすぎて、単語が一つか二つぐらいしか分からない。 The student’s pronunciation is not sufficient; raters understand very few words.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>何を言っているかわからない。 It is completely incomprehensible.</td>
<td></td>
</tr>
</tbody>
</table>

5) Discourse structure in the Monologue task

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>全体の構成が大変良く、文の流れがよい。新しい語彙、表現をふんだんに使用しており、内容の具体性が十分ある。否定文を使うなど、文型にもバリエティが見られる。 Well-organized with good flow; sufficient use of new vocabulary/expressions; fully substantive; wide range of sentence patterns</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>全体の構成が良く、文の流れもまあまあ良い。ほぼ新しい語彙や文型を使っているし、否定文を使うなど、文型にもバリエティが見られる。具体的な内容がだいたい述べられている。</td>
<td></td>
</tr>
</tbody>
</table>

Well-organized with fair flow; adequate use of new vocabulary/expressions; substantive; adequate range of sentence patterns

3 構成や内容の具体性にやや劣るがあるが、新しい単語や文型を使っている。
Loosely organized; inadequate development of topic; occasional use of new vocabulary/expressions

2 内容の具体性が乏しく、限られた単語や文型しか使用できない。新しい文型や語彙の使用が少ない。
Poorly organized; limited range of sentence patterns; limited use of vocabulary/expressions;

1 構成がとくにない。新しい単語を使われていても全部同じ文型で単語を変えて繰り返すだけなど、文型が単調で、内容に全く工夫がない。
No organization; monotonic sentence pattern; no creativity and development
Appendix B  Student Questionnaire

JPNS 101 Questionnaire

Your native language: ______________________
Age: _______  Sex: □ Male/□ Female
What year are you in? Please circle.
Freshman  Sophomore  Junior  Senior  Graduate
What is your major? __________________________

Section 1: Your purpose in taking the JPNS course.

1. What is your purpose in taking the Japanese language course?
2. Why did you choose the Japanese language over others?
3. Among the 4 skills of speaking, listening, writing, and reading, which skill do you want to improve the most?
   a) speaking
   b) listening
   c) writing
   d) reading
4. What is/are your goal(s) for taking the Japanese language course?
   What do you want to be able to do in Japanese by the end of the semester?

Section 2: Comparing the traditional pencil-and-paper test given in the first half of the semester and the performance-based assessments in the second half of the semester, please response to the following questions.

1. Which style of evaluation/test did you like better?
   □ liked traditional much better
   □ liked traditional somewhat better
   □ about the same
   □ liked performance-based somewhat better
   □ liked performance-based much better

Reason(s) for your choice __________________________
2. Which caused you more anxiety?
   □ traditional caused me much more anxiety
   □ traditional caused me somewhat more anxiety
   □ about the same
   □ performance-based caused me somewhat more anxiety
   □ performance-based caused me much more anxiety

   Reason(s) for your choice

3. Which gave you a greater sense of accomplishment?
   □ traditional gave me a much greater sense of accomplishment
   □ traditional gave me a somewhat greater sense of accomplishment
   □ about the same
   □ performance-based gave me a somewhat greater sense of accomplishment
   □ performance-based gave me a much greater sense of accomplishment

   Reason(s) for your choice

4. Which test did you think was more challenging?
   □ traditional was much more challenging
   □ traditional was somewhat challenging
   □ about the same
   □ performance-based was somewhat challenging
   □ performance-based much more challenging

   Reason(s) for your choice

5. Which style of test motivated you more to learn?
   □ traditional motivated me much more
   □ traditional motivated me somewhat more
   □ about the same
   □ performance-based motivated me somewhat more
   □ performance-based motivated much more

   Reason(s) for your choice
6. Which style of test did you think was fairer?

- □ traditional was much fairer
- □ traditional was somewhat fairer
- □ about the same
- □ performance-based was somewhat fairer
- □ performance-based much fairer

Reason(s) for your choice

Section 3: Your method of preparation for exam.

1. How did you prepare for the traditional pen-paper test? Please explain in detail


3. Did the performance-based method change the way you studied Japanese in the course?
   - □ No
   - □ Yes – If yes, please explain in detail how the performance-based method change the way you studied.

4. Which method of preparation better aligned with your language learning goals that you put down in section 1-1?

   - □ traditional aligned much better
   - □ traditional aligned somewhat better
   - □ about the same
   - □ performance-based aligned somewhat better
   - □ performance-based aligned much better

   Reason(s) for your choice

5. Which testing method better reflected what you know/can do in Japanese?

   - □ traditional reflected it much better
   - □ traditional reflected it somewhat better
   - □ about the same
Section 4: About the content of the performance-based exam

I. Dictation part
   1. Did you practice for dictation in advance?
      □ Yes
      □ No

   2. Did you like practicing for dictation?
      □ Yes
      □ No

Reason(s) for your choice

II. Speaking part
   1. Which tasks did you like? You can choose as many as you want.
      □ Monologue (ex, talking about your hometown, your room, and your typical weekend)
      □ Reading aloud
      □ Reading paragraph, listening to questions, and answer by speaking
      □ Role-play

Reason(s) for your choice

If you have any other comments regarding testing, feel free to write here:
Appendix C  Instructor Questionnaire

Performance-based Test 協力者評価アンケート
Survey of opinion on the Performance-Based Test for instructors

JPNS  101

この度は Performance-based Test（以下 PBT）実施にご協力いただき、誠にありがとうございます。恐れ入りますが、以下のアンケートへのご協力をお願い致します。

Thank you for trying the Performance-Based Test (PBT). We would appreciate you giving us your opinion of the PBT through the following survey.

1. PBT に関してどう思われますか。
   Overall, how would you rate the PBT?
   (☐) 大変よかった
       Excellent
   (☐) よかった
       Good
   (☐) あまりよくなかった
       Not very good
   (☐) よくなかった
       Poor

   理由をお書きください。
   What made you choose that answer?

2. PBT の内容をどう思いますか。
   How satisfactory were the contents?
   (☐) 大変満足
       Extremely satisfactory
   (☐) 満足
       Satisfactory
   (☐) あまりあま満足
       Somewhat satisfactory
   (☐) もの足りなかった
       Not satisfactory

   理由をお書きください。
   What made you choose that answer?

3. PBT を実践してみて、従来の評価法では測れていなかったところがあったと気づいたことがありますか。

   What made you choose that answer?
While trying the PBT, did you notice anything you could evaluate using the PBT that you couldn’t evaluate with the traditional pencil-and-paper test?

☐ 気づいたことがある
   Yes
☐ とくに何も気づかなかった
   No
☐ わからない
   Not sure

気づいたことがあるとお答えになった方は具体的に記してください。
If you answered “Yes”, please explain.

4. PBT で測られていないと思われることがありますか。
   Is there anything that you couldn’t evaluate with the PBT?

☐ 思ったことがある
   Yes
☐ とくに何も思わなかった
   No
☐ わからない
   Not sure

思ったことがあるとお答えになった方は具体的に記してください。
If you answered “Yes”, please explain.

5. PBT では試験問題（タスク）をあらかじめ学生に公開しましたが、その点についてどう考えますか。具体的に記してください。
   What do you think of the fact that students will know the “Tasks” ahead of time that they will have to perform on the PBT? Please explain your reasoning explicitly.

6. 総合的に従来の筆記試験と PBT を比較して、どう思いますか。
   Overall, did you prefer the PBT or the traditional pencil-and-paper tests?

☐ 筆記試験の方がいい
   Prefer the traditional pencil-and-paper tests
☐ PBT の方がいい
   Prefer the PBT
☐ どちらとも言えない
   Neither

理由をお書きください
   What made you choose that answer?
7. 筆記試験からPBTへの移行するにあたり、学生の学習態度に何か変化が見られましたか。

By switching from the pencil-and-paper test to the PBT, did you notice any difference in the way your students prepared before the test?

(□) 大変変化が見られた
Large difference
(□) 少し変化が見られた
Little difference
(□) 何も変わらなかった
No difference
(□) わからない
Not sure

もし、少しでも変化が見られたとお感じになった方は具体的にお書きください。
If you saw any difference in your students’ preparation, please explain.

8. PBTでは筆記テストにはないスピーキングを組み入れましたが、PBTに移行することであなたの教え方に何か影響がありましたか。もし、意識的に何か気をつけたことがあれば、具体的にお書きください。

The PBT has a speaking comportment not present in the pencil-and-paper test. Did the switch from a pencil-and-paper test to the PBT have an effect on your teaching? Please explain explicitly.

9. このPBTは、どうすればもっとよくなると思いますか。もし、アドバイスがあれば、お願い致します。

In your opinion, how can we improve the PBT? Please give us your input.