Reading-Aloud Versus Task-Based Reading Instruction

Gordon Robson, Hideko Midorikawa, Naomi Ono, Emiko Takano and Tsuneo Takanashi

Abstract

This research investigated the efficacy of two types of teaching methods in Japanese high school reading classes. An experimental class was taught using task-based materials developed from a standardized textbook with the instructor making extensive use of the target language and employing a more student-centered approach than is typical of Japanese high school classes. A pre-test was administered at the start of the semester with a post-test, regular midterm exam and English language proficiency test given at the end of the semester. The control classes were taught using a method known in Japanese as "ondoku" or reading-aloud that was combined with grammar-translation methods. The results of the post-test showed a significant difference between the groups with the experimental group outscoring all the control groups. No significant differences were found among the other tests. Finally, a questionnaire was administered to identify student reactions to the experimental class and materials. These results and educational implications are discussed.

Introduction

In Midorikawa, Ono, Robson and Takanashi (2002) we extensively discussed the necessity of acquiring automaticity in the processing of vocabulary and syntax so that reading comprehension can take place. This discussion resulted in the development of pre- and post-reading materials based on a text entitled "Heihachi’s Revolt." Midorikawa et al. (2002) concluded that the next step in our series of studies would be to use such materials in a study that had a truly experimental design to see if they would be more effective than the traditional methods of teaching reading typically used.

To determine if, as had been found through questionnaire research (Midorikawa, Ono, Robson, Takanashi and Takano, 2003, 2005), Japanese high school teachers are actually making extensive use of grammar-translation and reading-aloud, the reading classes of 38 teachers were chosen from among the 300 subjects of the questionnaires for videotaping. The analysis of the videotapes confirmed the researchers' beliefs that most Japanese high school teachers continue to use these traditional teaching methods (Midorikawa et. al., 2003, 2005).

In attempting to find the basis for the use of grammar-translation and reading-aloud methods and whether such methods have been tested in terms of their effectiveness, we came across a number of publications by Miyasato (2002, 2003, 2005, with Takatsuka, 2004 and under review a). However, before looking at whether these methods are effective or not, it
is necessary to see if they are theory-driven, what they actually entail and why they are used.

The Reading Process

A number of theories to explain the reading process have been put forward, but perhaps the most compelling is that based on cognitive psychology (Gagné, Yekovich and Yekovich, 1993). One part of this process is derived from declarative knowledge, which is composed of letters, phonemes, morphemes, words, ideas, schemas and topics. This knowledge is stored in long-term memory as propositions, images, linear orders and schemas. The second part of the reading process is derived from procedural knowledge, which is composed of production sets such as decoding, literal comprehension, inferential comprehension and comprehension monitoring. These procedural production sets access declarative knowledge interactively, which leads to successful reading comprehension. Both of these types of knowledge underlie the automated basic reading skills, conceptual understandings and cognitive and metacognitive strategies skilled readers use.

Grammar-Translation and Reading-Aloud Methods

In Horiguchi (Ed., 1991) the grammar-translation method is described as the teacher or student reading a phrase, sentence or paragraph from the textbook and then orally translating what was read aloud. Typically, the students have prepared their translations before class and then check their translations with those provided by the instructor. There are several variations to this approach such as writing the translation on the blackboard or in notebooks. These oral translation activities are followed by grammar and vocabulary explanations in Japanese. As to why grammar-translation is regularly used in Japanese classes, a JACET publication (2001) states that a teacher with an insufficient working knowledge of English and who has little time for careful preparation can easily employ such methods. There is also the widely made claim that such methods help students pass university entrance exams. The vocabulary and grammar explanations are provided in the teaching manual, along with translations of all the texts. These methods are believed to be effective for acquiring accurate knowledge about sentence structures in English and usage in terms of grammar rules. Moreover, having students read the text aloud before translating is effective for teaching the relationship between letters and speech sounds and the speech sound features of sentences (JACET, 2001). Nonetheless, no studies are provided that support these claims empirically. Moreover, given the recent trend in longer reading passages and less translation on entrance exams (Kikuchi, 2006), the above described methods are probably not much help here either.

Regarding reading-aloud, Takanashi and Ushiro (Eds., 2000), describe these activities as chorus reading, typically after a recording of the passage has been played; group reading followed by free or buzz reading, where students individually read sections of the text together, at their own pace, aloud and repeatedly, and finally, individual students being
called on to read aloud. According to Niizato (1991), reading-aloud is the most frequently used method in Japanese English classrooms at both the junior and senior high school level. Reading-aloud is viewed as being effective for helping students connect letters with speech sounds, comprehending the text and fostering speaking abilities (Kanatani, Ed., 1995). However, reading-aloud has been criticized for preventing the reader from comprehending the text, improving reading speed or from just getting the basic idea of the text. One could argue that the same reasons for the use of grammar-translation methods can be found with the use of reading-aloud, lack of English language proficiency and lack of training. Moreover, rather than call such activities reading-aloud or “oral-reading” as Miyasato does, it would be more accurate to call this orthographic vocalization, as there seems to be little resemblance to actual reading.

Returning to Miyasato (2002, 2003, 2005, with Takatsuka, 2004 and under review a), in his literature reviews, he admits that the claims for and against reading-aloud suffer from a lack of formal theoretical and empirical support. Nonetheless, Miyasato supplies his own studies and those with Takatsuka (2004) to both provide a theory of oral reading and empirical support. For theoretical support Miyasato draws on Coltheart and Rastle's (1994) Dual Route Cascaded (DRC) model of word recognition. However, this model is concerned with how adults read and not with the knowledge required for reading, and thus has no relation to the use of reading aloud as a teaching method. Miyasato also cites Grabe's (1999) componential processing views of reading and Baddeley’s (2000) model of short term memory. However, Grabe in particular would find little use for reading aloud as a teaching method and Baddeley is not concerned with it at all.

In Miyasako (2002, 2003, 2005, with Takatsuka, 2004 and under review a), the numerous claims for the efficiency of reading-aloud are largely based on very poor research designs. Design weaknesses include lack of discussion of validity or reliability, and the use of multiple treatments with experimental groups without discussion of control group treatments. This final weakness has made any claims that reading-aloud alone resulted in better reading comprehension, higher levels of proficiency or any significant levels of learning suspect. In one study (Miyasato under review a) reading proficiency scores actual went down after treatment, and in a replication that divided the students into high and low proficiency groups, the high group’s reading scores again fell while the low group had a significant increase. Grammar and vocabulary scores also increased; however, due to the wide variety of treatments given, the lack of a control group and a non-causal based method of analysis, what led to these increases is unknown. Perhaps most surprising about Miyasato’s research is that he does not measure his subjects’ comprehension of their own high school reading materials, but concentrates exclusively on global reading proficiency. Similarly, Kanatani (1995) has attempted to support various claims for reading aloud based on equally poor research designs.
Referring to the discussions in non-Japanese sources of reading aloud, it has been noted in Griffin (1992) and Alderson (2000) that reading aloud gives teachers an opportunity to evaluate individual learner's reading skills. However, Rounds (1992) argues that reading aloud leads to a word-by-word decoding of the text thereby overloading short-term memory and resulting in a total lack of comprehension. Finally, drawing on the whole word approaches advocated by Steinberg, and fully discussed in Steinberg, Nagata and Aline (2001) concerning their work with hearing mute children, Steinberg has shown that reading does not require an ability to decode phonetically.

Task-Based Methods

As for task based teaching, this approach recognizes that students need a combination of both bottom-up and top-down reading exercises and strategy instruction (Celce-Murcia, Ed., 2001). These exercises are provided in a pre-, during- and post-reading framework, where pre-reading exercises help students access background knowledge and provide exercises for key vocabulary and syntax to improve automaticity of bottom-up processing. During reading students can look for the answers to pre-reading questions or confirm pre-reading predictions. Post-reading activities include answering comprehension questions, responding to the reading critically, or perhaps listening to a lecture on the subject (Anderson, 1999; Celce-Murcia, Ed., 2001; Day, 1993; Silberstein, 1994; Urquhart & Weir, 1998). The empirical support for task-based and strategy instruction is impressive and includes the research of such individuals as Clarke and Nation (1980), Ellis, N. (1997), Ellis, N. and Beaton (1993), Ellis, R., Tanaka and Yamazaki (1994), Foster and Skehan (1996), Joe (1995 and 1998), Knight (1994), Lawson and Hogden (1996), Paribakht and Wesche (1993 and 1997), Reed (2000) and Skehan and Foster (1997). This research typically looks at one type of task or strategy in comparison to another type to determine which is most effective. To date, as far as the researchers are aware, a study comparing traditional Japanese English language approaches with task- and strategy-based reading instruction has not been done.

As a result of the above discussion, the researchers find it necessary to ask the following research questions.

1. What are the relationships among the pre- and post-tests, the mid-term test and the proficiency tests?
2. Will there be systematic differences in the English language reading abilities, mid-term test scores and English language proficiency of students using task-based, pre- and post-reading materials and those students using the present, traditional approaches to reading instruction?
3. Will there be general satisfaction with the materials among the students using the task-based pre- and post-reading materials?

The alpha level for all statistical decisions was set at .05.
METHODOLOGY

The subjects in this study were all second year Japanese senior high school female students. Their average age was 17 years old, and they were all native speakers of Japanese. These students were in six intact English reading classes with an average class size of 16. The classes met twice a week for 50 minutes. The textbook used for these classes, *Sunshine English Course I*, had been approved by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Class number two was chosen as the experimental group due to its Japanese female instructor who was somewhat less typical of many Japanese high school English teachers. First, the instructor had a master’s degree, and second, she was rather young (late twenties) and relatively inexperienced (three or four years of teaching). Unfortunately, this teacher’s master’s degree was in English and American literature rather than in Teaching English as a Foreign Language (TEFL), although she had taken some TEFL courses during her graduate training. Nonetheless, the researchers felt confident they could provide this instructor with sufficient guidance to overcome this lack of training. Moreover, due to the instructor’s age and inexperience, the researchers felt that she would still be open-minded and flexible in her teaching. All the teachers in the control classes had been videotaped prior to the start of the experiment, and the class number two teacher who taught the experimental class was video taped before and during the experiment.

Three of the lessons that were going to be used in the second half of the school year (from September to January) were selected for adaptation. These lessons were Lesson 5, “Observing the Influence of Acid Rain,” Lesson 8, “Will We Live to be 150?” and Lesson 9, “The Right Machine for the Job.” The materials developed for these lessons were very similar to those introduced in Midorikawa et al. (2002) with pre-reading vocabulary, collocations, syntax and background information exercises followed by post-reading comprehension questions, and vocabulary and grammar exercises. With Lessons 5 and 8, the students were not allowed to use their bilingual dictionaries; however, some students complained that without a Japanese translation they could not understand the vocabulary and, as a result, with Lesson 9 this policy was dropped.

A review of the pre-experiment videos showed that the control groups’ classes followed a predictable pattern of techniques. A recording of the reading text would be played and after each sentence the students would chorally read aloud. The instructor would then explain the grammar and vocabulary in Japanese. This was followed by the students reading aloud translations of the text with the instructor critiquing the translations and making comments and corrections. Then this process was repeated.

In the experimental class, the topic was introduced through visual, schema stimulating materials, and then a vocabulary in context task where target vocabulary items were matched with their meanings was worked on. The contexts were all directly from the
textbook reading text that had been broken down into three parts. Next, the students filled in blanks in the textbook text with the previously practiced vocabulary. This process was repeated for all the vocabulary in that section. Then a grammar point was introduced with a consciousness raising exercise whereby the students were given a number of example sentences and asked to underline a given grammatical feature. They were then given an explanation of that feature and asked to identify the correct or incorrect application of that rule. If the rule was applied incorrectly, the students were asked to correct the sentence. They were then given a task that required them to create their own sentences using this grammatical feature. After the grammar task the students worked in groups on two pre-reading activities, a questionnaire asking for agreement or disagreement to statements dealing with the reading such as, “Air pollution comes from burning garbage in Japan.” This task was followed by a discussion task where students worked together to answer two questions directly related to the reading. An example would be: “What are the main causes of air pollution?” The students then read the text and answered a series of true or false comprehension questions. They also did a fill in the blank comprehension exercise and a true or false task that required the students to draw inferences from the text. Finally, two additional vocabulary tasks and an additional grammar task were completed. This pattern of instruction was then repeated for the next section of the text.

A pre-test based on questions from the three lessons to be taught was developed and administered to all six classes. 25 items were selected that tested for vocabulary, grammar and reading comprehension. After administration, the test results were analyzed using an ANOVA statistic to determine whether the six classes were equivalent in their knowledge of the material. With the exception of a normal distribution, all the assumptions underlying the ANOVA statistic were met. The results are shown in Tables 1 and 2.

As there was a significant difference for group with the ANOVA statistic of $F=2.99$, a

<table>
<thead>
<tr>
<th>Table 1: Pre-Test ANOVA Results $N=98$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect</strong></td>
</tr>
<tr>
<td>Class</td>
</tr>
<tr>
<td>*=p&gt;.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Post Hoc Comparisons among Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class Mean</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>*=p&gt;.05</td>
</tr>
</tbody>
</table>
Scheffé test was run and this significant difference was found between class one and class four as shown in Table 2. As can be seen from the different groups’ means, class one was rather lower in comparison to the other classes. Due to this one difference, we decided to exclude class one from the experiment.

All of the students took the regular mid-term exam. This exam was independently developed by the teachers in the high school. The test, as with all the other measures used in this study, was examined using descriptive statistics, correlations using Pearson R and for mean differences using the ANOVA statistic. If significant differences were found with the ANOVA statistic, an LSD Test or Planned Comparisons Test was used for making post hoc comparisons. At the very end of the academic year, all of the students were given an English proficiency test entitled “Eigo Nöryoku Hantei Tesuto,” Test B or “English Proficiency Test,” Version B. This test is part of the STEP test series, and it tests proficiency in vocabulary, syntax, reading comprehension and listening. It is popularly known in Japan as the Eiken, and while it does have a numerical score, students typically receive a level designation such as “passed Step Level 2.” Although we are not concerned with these levels here, the test used included levels three, pre-two and two. Finally, a post-test, which was the same as the pre-test, was given at the end of the school year to determine if there were systematic differences among the groups.

Finally, the researchers created a simple questionnaire that was administered after the completion of the lessons. The first three items on the questionnaire were open-ended questions that allowed the students to respond in anyway they wished. These questions were: 1) What is different from the previous way you were taught; 2) Compared with your previous lessons, what do you understand more; and 3) Compared with your previous lessons, what causes you more trouble? The next questions gave four possible choices from 1=It does not help at all; 2=It does not help much; 3=It helps to some extent; and 4=It helps a lot. The questions were: 4) Reading English texts without Japanese translation helps comprehension; 5) Learning vocabulary in English helps with reading in English; 6) Grammar explanations before reading helps with reading comprehension; 7) This new study method helps me with understanding grammar and vocabulary; and 8) I enjoyed this new method of studying. This last item was 1=Did not enjoy at all; 2=Did not enjoy much; 3=Enjoyed somewhat; and 4=Enjoyed very much.

RESULTS

The descriptive statistics on all measures are shown in Table 3. The different n-sizes were due to not all the students taking the tests. Those students with missing data were excluded from the experiment.

The descriptive statistics show that the experimental groups did rather well on all of the tests. Moreover, they made fairly clear gains on the post-test with the scores moving
from a mean of 9.94 to 12.87. The mid-term results were better for this group at 66.61 (compared to the entire second year at 54.21), but far too low for what was described by the high school faculty as an advanced group of learners. In terms of proficiency, this group is rather good with a mean of 388.09, and with a fair number of students (eight) scoring perfect on the EIKEN. The grammar scores were the lowest, but still in general, rather good. Vocabulary and reading scores were outstanding at 95.23 and 91.64 respectively out of 100 points. The distribution for the pre- has a high positive skew, while the sub-tests on the EIKEN are rather negatively skewed. The post-test has an almost normal distribution, while the mid-term and EIKEN are normal.

Table 4 shows the correlation matrix for all the tests given, including the sub-tests on the EIKEN. The assumptions underlying the Pearson R statistic were largely met, although, as noted above, several distributions were not normal. These distributions did not seem to have negatively affected the results.

The highest independent correlation was between the EIKEN and the post-test at .61 or 37 percent covariance. The other high correlations were between sub-sections of the EIKEN and the overall test scores, which was to be expected. There was also a strong correlation between the reading sub-test of the EIKEN and the post-test at .50. Interestingly, the vocabulary scores did not correlate with the post-test very highly at only .39, which is
Table 5: Post-Test ANOVA Results $N=82$

<table>
<thead>
<tr>
<th>Effect</th>
<th>Degrees of Freedom</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>4</td>
<td>4.08*</td>
</tr>
</tbody>
</table>

* = $p > .05$

Table 6: Post Hoc Comparisons among Groups

<table>
<thead>
<tr>
<th>Group Means</th>
<th>2</th>
<th>16.13</th>
<th>3</th>
<th>12.69</th>
<th>4</th>
<th>12.95</th>
<th>5</th>
<th>11.18</th>
<th>6</th>
<th>11.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.01*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.02*</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>.01*</td>
<td></td>
<td>.27</td>
<td></td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.01*</td>
<td></td>
<td>.39</td>
<td></td>
<td>.29</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = $p > .05$

Table 7: Mid-Term ANOVA Results for Experimental Groups $N=82$

<table>
<thead>
<tr>
<th>Effect</th>
<th>Degrees of Freedom</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>4</td>
<td>.32</td>
</tr>
</tbody>
</table>

* = $p > .05$

surprising given the number of vocabulary items on the test. The mid-term also correlated rather highly with both the EIKEN at .57 and the pre- and post-tests at .52. The mid-term had strong correlations with the EIKEN vocabulary subtest at .50, while the post-test had fairly strong correlations with the reading and grammar subtests at .50 for both. We can see that all of these tests are tapping into some kind of general English language proficiency.

Table 5 shows the post-test ANOVA results to determine if there were systematic differences among the five classes. All the underlying assumptions for use of the ANOVA statistics were met, including a normal distribution.

There was a significant difference for group with the ANOVA statistic and the difference was found between class two and all the other classes using the LSD Test for post hoc comparisons. Thus, class two did significantly better on the post-test with a mean of 16.13 than did the other four classes in the experiment (see Table 6).

Table 7 gives the ANOVA results on the mid-term test for the five classes. All the assumptions for use of the ANOVA statistic were met.

There was no significant difference for group with the ANOVA statistic. An examination of each groups’ means shows that group 2 had a mean of 65.35, group 3, 67.44, group 4, 66.76, group 5, 65.94 and group 6, 65.88.

Table 8 shows the ANOVA results for the EIKEN test for the five classes. Again, all of the assumptions underlying use of the ANOVA statistic were met.

There was no significant difference for group with the ANOVA statistic for the EIKEN.
Table 8: EIKEN ANOVA Results $N=79$

<table>
<thead>
<tr>
<th>Effect</th>
<th>Degrees of Freedom</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>4</td>
<td>1.58</td>
</tr>
</tbody>
</table>

* = p > .05

Table 9: Questionnaire Responses $N=17$

<table>
<thead>
<tr>
<th>Rank</th>
<th>Rank</th>
<th>Rank</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Items</td>
<td>$N$</td>
<td>$N$</td>
<td>$N$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>6.5*</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4.5*</td>
<td>10.5*</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>5</td>
<td>9.5*</td>
</tr>
<tr>
<td>8**</td>
<td>2</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

* Responses scored .5 indicate a student placing their response between the two ranks.
** For Item 8, one student did not respond.

The means for each class on this test were 389.82 for class two, 379.0 for class three, 406.69 for class four, 389.75 for class five and 370.82 for class six. Moreover, there were no significant differences among the sub-tests on the EIKEN.

Finally, regarding the results of the questionnaire, there was only one negative comment for the first item (differences between the present class’ learning style and previous class’ learning style), which was that the teacher spoke too quickly. The other comments showed that the students were rather pleased with the new methodology and materials. Their comments included statements such as, “I can understand English without translating;” “I get less tired of class;” “I can learn more vocabulary;” “I have to think more by myself;” “We use more English and less Japanese;” “The class puts more emphasis on vocabulary than grammar;” “I practiced a lot of exercises with the printed materials.” On the second item regarding increases in understanding, there were two negative comments and a large number of positive comments. One student claimed to have no increased understanding and another student claimed not to be able to understand anything at all. The positive comments included “I can learn the vocabulary more easily;” “I can understand grammar more easily;” “I can comprehend the text more easily;” “I can understand the English text without translating;” “The exercises and grammar explanations are easier to understand;” “Repeating the exercises helps me understand better;” “It is getting easier to understand the vocabulary and sentences.” As for item three regarding what caused problems for the students, the responses were very revealing. One student complained that the pace of the lessons was too fast and another that it was too slow. Some of the students struggled to understand the grammar and others to understand the vocabulary. Several students did not like all the extra work and one student had no problems at all. Table 9 shows the results for items four through eight.
These responses show the students had mixed negative and positive feelings about the materials and methodology in general. Regarding item 4, reading English texts without Japanese translation helps comprehension; the students were evenly split with 8 negative and 9 positive responses. A similar situation of 8 negatives and 9 positives was found with item 5, learning vocabulary in English helps with reading in English. With item 6, grammar explanations before reading helps with reading comprehension; the students were largely positive with 4.5 negative responses and 12.5 positive responses. This was also true with item 7 with 6 negatives and 11 positives; this new study method helps me with understanding grammar and vocabulary. Unfortunately, the students did not seem to be enjoying the new study method as shown by the responses to item 8 with 10 negatives and 6 positives.

DISCUSSION

As was noted in the methodology section, there was one significant difference in scores on the pre-test between classes one and four. Because of this, class one was excluded from the study. Moreover, a comparison of class one's mean pre-test score of 7.69 and the mean pre-test scores for the six classes in the experiment, which was 9.94 as shown in Table 3, lends support to this decision. It is clear that this class had a significantly lower level of ability. As these level designations have been based on achievement rather than proficiency, this is perhaps not surprising; however, it calls into question the wisdom of using achievement scores for leveling.

The first research question, "What are the relationships among the pre- and post-tests, the mid-term test and the proficiency tests," was answered in Table 4. The highest correlation was between the post-test and the mid-term at .61. The correlation between the reading subtest of the EIKEN and the post-test at .50 shows that the post-test was a relatively good measure of the students' overall reading proficiency. The post-test correlation with the overall EIKEN score also showed that general English language proficiency was an important factor in these scores. The rather low correlation between the post-test and the vocabulary scores at only .39 was surprising given the number of vocabulary items on the test. This may have been due to the differences in the way vocabulary was tested on the two examinations. Grammar and reading proficiency were much more important for these students' achievement. The mid-term also correlated rather highly with both the EIKEN at .57 and the pre- and post-tests at .52. As noted above, all of these tests seem to be tapping into some measure of general English language proficiency. Interestingly, the mid-term seemed to have been a better measure of vocabulary at .50 than the pre-post-test in relation to the EIKEN.

The answer to the second research question as to whether the experimental class using the task-based materials and learner-centered teaching methodology was better is found in Tables 5, 6, 7 and 9. The ANOVA statistic found a significant difference in the groups, and the post-hoc comparisons found that class two was significantly better on the post-test than
the other four classes. This should not be surprising as there was a close fit between the contents of the test and the teaching and practicing styles. Although there were some students who had trouble with the new style materials as shown by the questionnaire responses, the majority successfully passed the post-test. There were several exceptions to this, but, in comparison to the results seen on the regular mid-term exam, these scores are far more positive. Thus, we can see that when there is a clear relationship between the materials being studied, taught and tested, students will be successful. The purpose of schooling is to provide learners with opportunities to be successful. Learners who consistently fail will soon view themselves as failures and will give up. Once a learner has given up, it becomes very difficult to change that learner’s behavior. Moreover, it is important for student motivation and self-esteem to be given opportunities to succeed. Although some students in the experimental class struggled with the material, when it came to taking the post-test, most of these students did rather well. The most important thing in curriculum development and testing is to ensure that there is an extremely close fit between the materials, the teaching methodology, and the tests. Students who are taught and then tested in a manner that does not match are doomed to failure, and repeated failure leads to a loss of motivation and self-esteem (see Gagné, Yekovich and Yekovich, 1993, pages 425-448 for a full discussion on motivation and failure).

There may have been some concern that the students in class two would be disadvantaged on the mid-term due to the different materials and teaching methods used. However, this group did just as well (or poorly) as the other groups. These scores were extremely low and probably do not reflect the students’ true abilities in terms of the ways in which these materials were used. We have already seen that the mean for the entire second year was 52.21, which means that well over half of all these students failed the test. This test clearly contained too many items, and the items themselves were very difficult. The contrast between this test and the pre- and post-tests shows these tests to be far more reasonable and less demanding. Nonetheless, the mid-term test items did reflect the textbook rather well. However, the textbook used for this class was very difficult with a large number of new vocabulary items introduced each lesson as well as a good deal of new syntax. Interestingly, the Flesh-Kincaid reading level for the Lesson 5 text was 6.8 or almost a seventh grade reading level, while that of Lesson 8 was 9.0 or two grade levels higher. With Lesson 9, the grade level drops back to down to 6.4. Even so, these are very challenging reading levels with wide ranges of difficulty from lesson to lesson. In addition, if learners are being taught through such methodologies as grammar explanations in Japanese, choral reading aloud and translation, then these methodologies will not prepare students to pass a test such as this one. We should expect that those few students who did extremely well on this test did so for reasons that had little to do with the classroom teaching, and, in fact, those students who did so well on the mid-term and EIKEN were all students who had returned from stays in English
speaking countries and are near-native in their English language proficiency.

Another interesting aspect of the mid-term results was that the distribution of scores was almost normal, which should be entirely unexpected in an achievement test. This is another sign of concern. The mid-term test should be an achievement test. As such, we would expect a strong negative skew to the distribution where most of the students had done very well. However, this test is acting like a proficiency measure where the distribution is rather normal. In fact, the distribution of scores on the mid-term is more normal than that of the EIKEN, which actually is a proficiency measure. Both these and the previous results indicate very serious problems with this mid-term exam.

In regards to the EIKEN and its sub-tests, here too no significant differences were found among the five classes. It would be rather miraculous if a single semester where new materials and teaching methodologies were used would result in significant proficiency gains. Language proficiency is cumulative and requires a great deal of time for there to be an effect that can be based on classroom learning. Seen in a different light, it would seem that the task-based materials and the learner-centered teaching methodology did not disadvantage these students either. Moreover, this test seems to have been a rather good measure of these students' English language proficiency. The distribution was a little bit negatively skewed, which means that the students did rather well on it, but it was not skewed enough to indicate that the test was too easy. As opposed to the use of course grades for placement purposes, as was done with these classes, this test would be much better for that purpose, and it could also be used for program evaluation purposes.

The final research question asked about student satisfaction. Here the results were a bit confused. When the students responded to the open-ended items, they were generally positive. However, with the Likert type items, there was a bit more ambiguity. In terms of not using bilingual dictionaries to help with understanding and vocabulary learning, the students were evenly split. However, with grammar and the use of the new study method, the students were very positive. This contrasted with the large number who reported that they did not like this new method. Most students agreed that they could understand the texts without using Japanese, that vocabulary learning helped, as did learning the grammar, and that the new study method was beneficial. However, when asked about enjoyment, the students were again split.

Several explanations for these results seem plausible. First, the students were unused to doing exercises and having to participate. Until the experiment began, students sat passively in class listening to their teacher explain things and occasionally doing some choral work, reading aloud and translating aloud. They never had to do any actual exercises. With the new materials, they suddenly had to do some real language work by themselves. They seem to have been overwhelmed by the amount of exercises they needed to do, and although they did see that doing these exercises was beneficial, they did not enjoy them. Regarding the use
of bilingual dictionaries, the researchers had initially believed that such dictionaries made it difficult for the students to process the texts in English and to comprehend them. However, recent research has shown that bilingual dictionaries are popular, can be a useful tool for reading comprehension and can enhance vocabulary learning (Reed, 2000). This resulted in the change of policy with Lesson 9, and the reaction was very favorable.

CONCLUSION

Although the results discussed above would appear to be fairly straightforward, there are a number of issues lurking behind the various tables. As was discussed in the Methodology section, the instructor chosen for the experimental class was young, untrained and relatively inexperienced. This resulted in a number of difficulties in using the tasks created for the experimental class and the student’s reaction to them. Most of this information was gathered using qualitative techniques and in the interactions between the researchers and the instructor. As a result, this information is necessarily anecdotal. However, it is important to discuss some of these issues despite the fact that they are not reflected in the quantitative data.

The lack of training and experience resulted in a rather negative reaction to the use of tasks and a continued desire for extensive discussions of vocabulary and syntax in Japanese on the part of the instructor. Such an approach would have resulted in little actual student practice and learning. The insistence of the researchers that the tasks and appropriate teaching methodologies should be followed resulted in a change in behavior, but not in attitude. Moreover, the students, as noted above in the questionnaire discussion, were not actually used to doing any language work in class. Their participation had been largely passive. The students’ reactions are reflected in the questionnaire results where they grudgingly admitted that the materials were effective, but they did not like the fact that they had to do them and that they had to do so many.

The conclusions we can draw from the observations of and interactions with the instructor and students in the experimental class are that both teachers and learners in high schools need appropriate training. The lack of professional language teacher training in communicative language teaching and task-based instruction made it difficult to convince the instructor to use the materials and follow the methodologies that trained language teachers use all the time. Moreover, the teacher lacked appropriate professional experiences as her colleagues all used the same methods, and these methods most likely reflected the ways in which these teachers had been taught in the past. Finally, the teacher lacked a sufficient level of English language proficiency to use the target language in a fluent and effective way. The similar lack of appropriate English language training among the learners was clearly reflected in the confused attitudes expressed in the questionnaire. Until learners are taught in appropriate and effective ways, they will continue to prefer sitting passively in

— 26 —
their classrooms, picking up what English they can incidentally and being very unsuccessful. Finally, the lack of a professionally prepared textbook reinforced many of the poor teaching and learning behaviors exhibited in the experimental class. The text lacked any effective learning materials particularly in light of the large amount of new vocabulary and syntax that was introduced in each lesson. Moreover, the teacher's manual simply provided translations, and vocabulary and grammar notes. There is no real attempt by the textbook developers to provide the teacher with adequate language learning tools. That this should be so is especially surprising given the large number of commercially available reading textbooks that can be purchased in Japan that do provide teachers with adequate materials. The result of this situation is that many learners do not have an opportunity to experience real language learning until they go on to a university.

As for the test results, the experimental group made the expected significant gains on the post-test, which showed that when there is a close fit between materials and testing, the students will be successful. The success of these students in comparison to the failure of so many on the high school's regular mid-term exam clearly shows that there were serious problems with the program of instruction and testing. Unfortunately, while the mid-term exam results showed that the students doing the task-based lessons were not disadvantaged, the use of tasks did not give them an advantage either. Given the extraordinary difficulty of the mid-term and the fact that the students had only been studying with the new materials and methods for four months, a level of improvement above the other classes is perhaps too much to expect. Language proficiency is incremental with time being one of the most important factors along with exposure and practice. Due to the limited amounts of all three of these, it is not surprising that the students learning through task-based materials did not do significantly better than the other students on both the mid-term and the proficiency test.

The implications of this study seem fairly clear. English language teachers need to be better trained, to have more appropriate teaching experiences, to know why they should use certain methods and to not just rely on the way they have been taught, and to have a level of English language proficiency that allows them to teach through the target language. Moreover, the extensive use of grammar-translation and reading-aloud in Japanese English classes is placing students at a disadvantage to those both inside and outside of Japan who are exposed to the more effective methods of task- and strategy-based instruction. When teachers have achieved the above goals, then they will be able to take the poorly constructed materials they must use and adapt them for real language teaching. Until this happens, the current situation will continue and Japanese students will not enjoy the successes that are coming very rapidly to other parts of Asia.

(This research was funded by a grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology.)
REFERENCES


language processing. Tokyo: Reading Instruction Research Group.
Miyasato, N. (2003). What relationship does the metacognition of reading-aloud have with the ability to read aloud and the English language ability of Japanese senior high school students? Language Education and Technology, 40, 121–140.
Miyasato, N. (under review a). What effect does reading instruction focused on reading-aloud have on the reading comprehension of Japanese senior high school students of English? Unpublished manuscript.
Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. Language Teaching Research, 1, 185–211.