ABSTRACT
This paper discusses which type of computer-based questions could accurately predict the English reading comprehension levels of EFL (English as a Foreign Language) learners: traditional multiple-choice questions or write-in-answer questions (writing answers in the blank on the screen). Both types of questions were used in our surveys conducted on EFL learners in Japan. We conducted two types of computer-based English comprehension tests along with semi-structured interviews on how students tackled those tasks. Descriptive and qualitative analyses indicated that the write-in-answer questions had more explanatory power to identify learners’ weaknesses than multiple-choice questions did. They also showed that both the learners’ levels of vocabulary knowledge and awareness of their own strategies had some influence on their reading comprehension levels. The analysis suggested that educators need to take the time to judge individual learner’s performances in foreign language testing based not solely on the results of common computer-based multiple-choice type questions, but also on painstaking analysis of individual write-in-answers collected by computer. This could provide an opportunity to avoid the stagnancy in advancing the learners’ levels of comprehension. In accordance with our current conclusion, we will incorporate the suitable type of test in the present web-based training system, named WebIts, developed by the author’s group.

KEY WORDS
write-in-answer, multiple-choice, computer-based, explanatory power, reading comprehension, semi-structured interview

1. Introduction
Educators and researchers in EFL and ESL (English as a Second Language) have devoted enormous energy for many years to improve reading and listening skills of EFL learners (e.g. Birch, 2007[1]; Koda, 2005[2]); however, many Japanese EFL learners have been struggling to improve their English proficiency levels. This is because, as it has been long noted, Japanese learners have few opportunities to use English in actual situations. In other words, it is uncommon for them to have an immediate need to use English in daily life. It is not an overstatement to say that term examinations and entrance examinations are their only stimuli in learning English.

It is also true that not a few Japanese EFL learners have been successful in mastering English. In the course of learning English, those successful learners happened to use or intentionally sought optimum approaches and strategies in mastering English. It is said that the Japanese educational environment is rather uniform and ruled by university entrance examinations; however, the actual learning environment is non-uniformly arranged. Some students do their work in an educational environment where everything is graded for entrance examinations, such as grammar-translation methods, but others enjoy a communicative approach or an exploratory approach to learning.

Conducting the present research survey, we discovered the importance of the roles of educators and researchers as advisers, especially in the EFL setting. They should provide optimum information for individual learners.

2. Research Rationale
2.1 Theoretical Background
For our purposes, the fundamental standpoint for educators should be found in Vygotsky's theory: that the process of human concept development is not simply a stimuli-response, one-way act process, but is a complex, mediated act process which involves a dynamic process of reciprocal relations between cultural, historical, societal human acts in real environments which surround people (Vygotsky, 1987[3], 1997[4]). Accordingly, in EFL learning environments, it is indispensable to have a computer-aided learning system in order to provide even slight more opportunities to use the target language.

Moll (1998 [5] ) summarized a part of Vygotsky’s concept, the zone of proximal development, as follows: “he would offer the subjects a second set of stimuli in
order to study how the subjects accomplished the task with the aid of new auxiliary means (p.4)." It is essential to give tasks beyond, but not far beyond and not too simple for, learners’ present abilities.

In view of this, it is essential to corroborate the explanatory power of questions generated by computer. Those questions should be appropriately shaped to help Japanese EFL learners have a better chance of experiencing the dynamism of reciprocal relations through web-based training.

2.2 Purpose of the Present Research

The present study addressed the research question below in order to find the appropriate type of computer-based question to predict Japanese EFL learners’ reading comprehension level. We conducted two types of English comprehension tests: Type A test – multiple-choice questions; and Type B test – write-in-answer questions. We also distributed a questionnaire, and administered semi-structured interview surveys on these two types of comprehension tests.

Research Question:
Which type of comprehension test has more explanatory power to predict learners’ reading comprehension levels: multiple-choice questions (Type A test) or write-in-answer questions (Type B test)?

3. Method

3.1 Participants

Five male and six female university students (a total of eleven) in their second to fourth year participated in the present research survey. None were English majors. One group consisting of five (two male and three female, belonging to the department of International Studies) had English classes four times a week, and another group consisting of six (two male and four female; belonging to other departments) had English classes two to three times a week. Each participant had had few opportunities to use English outside of their classes.

The two groups had exactly the same assignments. Each was administered the survey at the same place, on the same day, and their data was analyzed in the same manner after completing all assignments.

All the participants took the computer-based tests at their own pace. They were informed of the contents and procedures of the research survey beforehand and agreed to cooperate on the research survey and signed a consent form.

3.2 Materials

We have developed computer-based English comprehension tests of two types, incorporating vocabulary taken from the JACET (Japan Association of College English Teachers) 8000 vocabulary level list into the question items. The reasons why we adopted the JACET 8000 list were: some of the usual wording commonly used by native speakers of English is not familiar to the Japanese EFL learner community, and the JACET 8000 list was made in consideration of such a community.

The steps we took in making the tests were as follows. First, we chose to use the JACET 1000 level through 3000 level vocabulary in each sentence. Second, we selected ten sentence patterns. (Both types of tests have ten parallel patterns.) Third, ten sentences for each level (JACET 1000, 2000, 3000) were generated. (Each sentence in both types of tests had a different context.) The proportion of target word class from the JACET list for each sentence should be the same as Schmitt, et al. (2001): noun 3, verb 2, adjective 1. However, in the process of modifying sentences to match the Japanese EFL learners’ actual learning environment, the verbal portion was increased.

The ten sentence patterns were as follows:

1) SVC, interrogative
2) SVO + infinitive, nominal
3) SVOO, past tense
4) SVOC, past tense
5) relative pronoun, nominative
6) relative pronoun, objective, reduced
7) passive voice
8) present participle, postmodification
9) complex sentence, reduced, past tense
10) prepositional phrase, perfect form

These two types of tests were conducted on computer, in a way which is similar to SuperLab type psychology tests. The participants could not reread the target sentence on the screen once they chose to click the button to go forward. The elapsed time for the following were recorded automatically: reading, choosing options, filling out the blanks with Japanese equivalent meaning, and rest periods between trials and answering the questionnaire. Both Type tests were prefaced with practice session pages (Apx.1).

Type A test (Apx.2) is a multiple-choice question type. It consists of the target sentence written in English and three options. (All answers are written in Japanese.) Type B test is a write-in-answer type. It consists of the target sentence written in English and a blank which should be filled in in Japanese. Both have a questionnaire (Apx.3) after every sentence task. Questions are presented in random order.

An interview on these two types of computer-based English comprehension tests was conducted when each student completed the tests individually for the purpose of finding out how the participant tackled the tests. During the interview, voices were recorded to be analyzed after
the experiment. All participants were asked to fill out the face sheet at the beginning of the present study. The sheet asked them to give their demographic data and previous English learning experience throughout their lives.

3.3 Procedure

Face Sheet:
The participants filled out the face sheet after receiving instructions on the whole procedure.

Type A test:
The participant read the sentence at his/her own pace and clicked the button, and then on the next screen he/she chose one answer out of three options. The participant was not allowed to reread the sentence, and after each sentence trial, on the next screen, the participant was asked about his/her confidence in answering (choosing from four options: enough confidence, not sure about, no confidence, and answered incorrectly). No time limitation was set, but all the participants finished the task in less than 50 minutes.

Type B test:
The participant read the English sentence and on the next screen he/she filled in the blank with a Japanese translation. The participant was not allowed to reread the sentence, and after each sentence trial, on the next screen, the participant was asked about his/her confidence in answering (choosing from four options : enough confidence, not sure about, or no confidence, and “forgot the target sentence before filling out”; they are free to fill out the blank with their additional comments ). No time limitation was set, but all the participants finished the task in less than 50 minutes.

Interview:
The participants were asked about their trials. The interview was semi-structured.

Data Analysis:
Participants were divided into five levels in three categories for the qualitative analysis: beginners (A & B), intermediate (D & I), and advanced (K). The data were analyzed descriptively and qualitatively.

4. Results

Table 1 shows the results of Type A and B tests for the five participants along with their TOEIC scores. Tables 2 & 3 show the number of incorrect answers and participants’ levels of confidence in their answers for both tests.
In Table 1, “RT” means the total elapsed time for reading comprehension of thirty target sentences for Type A or Type B tests. “Ans” is the number of correct answers by each participant. “AcR” is the accuracy rate, which shows the percentage of correct answers by each: “Ans” was divided by 30 (full marks). “Eff” means the individual efficiency of the reading comprehension task for each Type test: “Ans” was divided by “RT” for reading of each Type test. The larger the number of “Eff”, the greater the accuracy and rapidity of reading comprehension of the target sentences. “Ans - Cnf” means the difference between the number of correct answers for each Type test and the number of participants’ “Confidence level 1” choices on the questionnaire asking their confidence level in their answers. The plus figure shows that participants could choose correct answers even when they were not entirely sure of their answers. The minus figure shows that they made mistakes even when they had confidence in their answers.

In Tables 2 and 3, “level” means the difficulty level of the vocabulary list for Japanese EFL learners created by JACET (as noted in Introduction). The first 2000-level vocabulary is designed to include approximately 75 percent of the English vocabulary found in English newspapers concerning basic information from daily life. The 3000-level is approximately the TOEIC score 550-level. The individual number of incorrect answers to JACET 1000 through 3000 levels questions and participants’ confidence levels for those incorrect answers are shown in Tables 2 and 3.

In Table 3, “*errs” means the detailed information about the errors for each level: “.5” shows the partial incorrectness of the answer. The task for Type B was write-in-answer type (making a Japanese translation). An insignificant mistake in translation was counted as a minus .5 score.

As for the data from the semi-structured interview, section 5 contains the detailed information necessary to discuss the research question. Scores from comprehension tests alone cannot sufficiently support the assertion that a specific type of test is suitable for inclusion in the present web-based training system. In order to decide which test is best, it is essential to understand the process by which participants determine their answers. And the only way to understand this process is by interviewing the student on the web page immediately after each question by means of the questionnaire.

In Table 3, “ID” denotes the number of incorrect answers and participants’ levels of confidence in their answers for Type B

<table>
<thead>
<tr>
<th>ID</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>I</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>level</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>errors</td>
<td>4.5</td>
<td>8.5</td>
<td>9</td>
<td>6</td>
<td>4.5</td>
<td>7</td>
</tr>
<tr>
<td>*errs</td>
<td>5*1+4</td>
<td>5*1+8</td>
<td>9</td>
<td>6.5*1+4</td>
<td>5*2+6</td>
<td>5*2+1</td>
</tr>
<tr>
<td>Cnf</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cnt</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Fgt</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*errs: details for the number of incorrect answers

Fgt: forgot the target sentence

Figure 1 Graph of the explanatory power of Type A and Type B tests

Figure 1 shows the comparison of participants’ TOEIC scores to their accuracy rates on Type A and Type B tests. In order to make an easy-to-read radar chart, the TOEIC score was divided by 5.

5. Discussion

As Table 1 shows, participants’ accuracy rates were 70 to 87 % for Type A test, with no significant differences among them. But there were significant differences for Type B test: 27 to 87 %. This seems to provide the first indication that Type B test has greater explanatory power to differentiate participants’ proficiency levels.

The figures of “Eff” – efficiency of reading comprehension task – only provide an indication of differences in participants’ reading proficiency levels. These figures in themselves have no absolute value. A large number indicates the participant’s high level of efficiency in reading target sentences. In Type B test, each participant had a larger number of “Eff” which correlated with their TOEIC test scores, but this was not the case for Type A test. This could be the second indication that Type B test has greater explanatory power to differentiate participants’ proficiency levels.

As “Ans - Cnf” in Table 1 shows, beginner-level participants were able to make a good guess in choosing the correct answer out of three options for Type A test but
not for Type B test. Type B test required them to write in their own translation, so this test style avoided correct answers by chance. This might be the third indication of the explanatory power of Type B test.

Tables 2 and 3 also show the number of incorrect answers. In Type A test, the incidence of errors did not parallel the JACET 1000, 2000, and 3000-levels. In particular participant “A” did not make any errors in the 3000-level questions. In the semi-structured interview, he said it was far easier for him to choose one answer among options than to write in translation. But in Type B test, the incidence of errors almost paralleled the JACET levels. It was difficult for participant “A” to decipher thirty questions, especially the JACET 2000 and 3000 vocabulary level sentences: 8.5 errors and 9 errors out of ten questions respectively. This might provide the fourth indication of the adequacy of Type B test to differentiate learners’ proficiency levels.

The additional information collected by scrutinizing the causes for participants’ errors for Type B test and through the interview might also support Type B’s adequacy. The beginner-level participants “A” and “B” made similar errors caused by their lack of vocabulary for each level. Especially in the JACET 2000- and 3000-level questions, and even in some of the 1000-level questions, they experienced confusion deciphering sentences, accelerated by their lack of sentence structures. The intermediate level participants “D” and “I” had fewer difficulties deciphering questions in levels 1000 and 2000, caused by their lack of vocabulary, but they experienced more difficulty grasping the meaning of sentences in the 3000-level, caused by both their lack of vocabulary and their weak knowledge of sentence structures.

However, the advanced participant “K”’s errors were of a bit different kind. For two of the errors in levels 1000 and 2000, she created a second sentence for each after taking a look at the wording in the target sentences. Four of the errors in level 3000 were partial misunderstandings: the target words had more than one meaning, prompting her to make predictable mistakes. It is not possible to detect these detailed aspects of errors in Type A results because of the test type – multiple-choice questions. These facts also support the idea that Type B test could better estimate the learners’ proficiency in reading. Figure 1 clearly indicates the high probability of Type B results.

Based on these findings, Type B questions should be incorporated into our present web-based training system in order to better the present learners’ learning environment.

6. Conclusion & Limitation

As stated above, in answer to the research question, Type B – write-in-answer – computer-based test has more explanatory power to predict learners’ actual reading comprehension levels. At the present time, many Japanese EFL learners have difficulties improving their English proficiency levels even though they have had much English education, usually for six years to ten years. What is needed is an individual support system using computer-based tests along with appropriate advice by specialists in the field. In order to realize this kind of environment, not the usual computer-based multiple-choice question style, but a write-in-answer style test, is preferable.

It is not feasible to realize this kind of educational sea change in Japan today. However, an individual advisory system could be reasonably provided using computer-based write-in-answer type tests followed by a semi-structured interview asking how students tackled those tasks through web-based training system.

As for a detailed analysis of the results of the semi-structured interview and the concrete process of incorporating these tasks, another paper will discuss these in order to provide more productive advice for Japanese EFL learners.

Appendices

Appendix 1
Reduced screens of TYPE A practice question

<table>
<thead>
<tr>
<th>Aタイプ練習（1）</th>
</tr>
</thead>
<tbody>
<tr>
<td>I want to eat an apple pie.</td>
</tr>
<tr>
<td>&lt;進む&gt;</td>
</tr>
</tbody>
</table>

英文に合う正しい和文を選択し、ボタンをクリックする。クリックするとすぐに次の画面に移る。

Choose the proper answer. After clicking the button, you will move to the next page.

私はアップルパイが食べたい。
私はアップルパイを食べた。
私はアップルパイを食べに行った。

the first page preceded by explanatory pages

Appendix 2
Type A questions – JACET 1000 level
1) Is it an important plan?
2) I forgot to go to the bank.
3) He showed me some pictures.
4) I found her latest novel interesting.
5) There is a national park which protects nature.
6) There were books my husband ordered.
7) Books and CDs are sold at that store.
8) It is natural for boys to respect fathers having a good influence on them.
9) I didn’t think the test was easy.
10) I searched the house for the missing key.

the second page (English translation added)
Appendix 3
Questionnaire screen after the Type B questions (English translation added, reduced screen)

休憩画面
(take a break)

直前の問題について、以下の当てはまる項目を選んでください。複数選択も可能です。
Choose the one that best fits your feeling:

● 自信がある Confident
● あいまい Not sure
● 理解できなかった Can’t understand
● 入力中に英文を忘れForgot the English sentence
● キー操作に手間取Input difficulty
● その他（other）

次の提示文に進む準備が整ったら、＜進む＞ボタンをクリックする。If you’re ready for the next question, press the <button>:

＜進む＞
N question items to go

Acknowledgements

This study was partially supported by the Japan Society for the Promotion of Science, Grants-in Aid for Scientific Research (C) (No. 23501188), given to the author.

Special gratitude is due to the eleven participants who made this study meaningful. I am grateful to Mr. Ochi, who created the application software for this study. I am also grateful to Ms. Nishida, Ms. Tsurii, and Mr. Hawthorne for their painstaking help and comments.

References