Developing A Computer Adaptive ESP Placement Test Using Moodle

Yukie KOYAMA  
Nagoya Institute of Technology  
Japan  
koyama@nitech.ac.jp

Minoru AKIYAMA  
e-learning Service CO. LTD.  
Japan  
akiyama@e-learning-service.co.jp

Abstract: In this study, we introduce a Computer Adaptive Test (CAT) module for creating an English for Specific Purposes (ESP) placement test in a Learning Management System (LMS), Moodle. The module is based on Item Response Theory, One-Parameter Rasch Model in this case, embedded in Moodle. One of the advantages of this system is that it enables each institute to create an inhouse-CAT, which is usually considered to be difficult to realize. In addition, the results of the CAT placement test are shown to discuss the validity and efficiency of the CAT system in comparison with other pencil & paper tests.

1. Introduction

Computer-adaptive testing (CAT), in which test items are selected based on the estimated ability of the test taker, has become one of the most useful and advanced assessment methods in this decade. In CAT the first item to be given to the test taker is usually of medium difficulty. If the first item is not answered correctly, an easier item will be provided, and if it is answered correctly, a more difficult one will be given next. This item selection and the terminating condition are based on a certain computer algorithm, and Item Response Theory is applied for the preliminary item analysis to determine item difficulty of each item.

Computerized testing, in general, benefits many testing areas including individuality and time-independence (Brown, 1997). Moreover, CAT has the major advantages of eliminating test items of inappropriate difficulty and saving testing time. However, as Dunkel (1999) points out, both test developers and users need to fully understand the underlying testing model, Item Response Theory (IRT), in order to avoid misuse and to maximize the advantages of CAT.

At many institutions placement tests are given to ensure efficient language teaching in terms of proficiency level, but typically there are various constraints. For example, the testing time is limited to class time, usually 90 minutes in Japan, immediate feedback is necessary, and the test has to be operated in a consistent environment. In addition to these conditions, very few English for Specific Purposes (ESP) tests are available, especially in science and technology. The word ESP is often used in contrast with General English, and ESP means teaching or learning English aimed at a specific discipline such as English for medical science. In the last a few decades ESP has become one of the most prominent methodologies in the teaching English as a Foreign Language. Therefore, the introduction of an ESP CAT using Moodle, an Learning Management System, offers one promising solution to the placement testing problems at an institution with the test constraints mentioned above.

2. Objective: Designing a CAT module in Moodle

In this study, we introduce a CAT module for creating an ESP placement test. The module is based on Item Response Theory, One-Parameter Rasch Model in this case, and this module is embedded in Moodle. In addition, by analyzing the results of the CAT, the validity and efficacy of this test are explained in comparison with other pencil & paper tests.
3. Moodle and the CAT module

As is generally known to educators around the world, Moodle is a free open source LMS developed by Martin Dougiamas in 2002. Dougiamas developed Moodle based on the philosophy of social constructivism, as a tool for collaborative learning among teachers and learners. In 2008, there were more than 48000 Moodle sites in 203 nations, and approximately 29 million users. However, according to Moodle Statistics, (http://moodle.org/stats/) these numbers are still increasing each day, which demonstrates the fact that Moodle plays a substantial role in the e-learning world.

This study uses the Quiz function of Moodle to make a CAT module for a placement test. After conducting preliminary tests online using the Moodle quiz function, results are analyzed using the Rasch model so that unfit items can be excluded. In the Moodle quiz function, various question types are included, such as multiple-choice question, text and Cloze. Therefore, the question type used for the CAT module can be any one of these types. The Test Data Analysis Program (TDAP, Ohtomo et al. 2006) is used to analyze the test data and to store them in the item bank with item difficulty indices. Only the items with item fit indices indices between -2 and +2 are saved in the bank because items less than -2 are considered to be overfitting and those more than +2 to be misfitting.
Next, using TDAP, several items are chosen from the item bank to constitute one testlet. In terms of the function, the testlet corresponds to a Quiz in the Moodle system. When one testlet is completed, the test information function of the testlet is shown as a bell-curve graph in the TDAP screen embedded in Moodle. It has to be considered that the spread range of maximum values of the test information function in all the testlets should cover the whole range of the test takers' ability the test set aims to measure, and they should be placed at equal distance in difficulty. To make a whole test set in the CAT module developed by Akiyama (2008), a group of testlets are chosen according to the principle explained above. A simulation in the module is conducted at this point in order to find the appropriateness of the termination condition of the test in terms of standard error of the mean.

The algorithm used in this CAT module is as follows:

1) The first testlet to be given is arbitrary assuming the size of the item bank is large enough.
2) The ability of a test taker is estimated by the answering pattern and difficulty levels of the items.
3) The second testlet to be given is the one with the closest difficulty level to the estimated ability of the test taker.
4) The ability of the test taker is again estimated by a combination of the answering patterns of the first and the second testlet.

This process continues until the standard error of the mean becomes less than the pre-set value.

Thus, the process of the CAT begins with the choosing the first testlet from the item bank. Therefore, the quantity, the quality and the structure of the item bank has a critical influence on the CAT process. The criteria of the item choice in this study have already been explained in relation to the use of TDAP above, but the following points should be noted in the item choice: the efficiency, the appropriateness and the precision. Considering these points, the test becomes an optimal one that can measure the test taker’s ability with the smallest number of appropriate items. (Downing, S.M. & Haladyna, 2006) Furthermore, in order to construct a well balanced CAT, this has to be done on the basis of the test specification, which clearly specifies the content, the form of the items, and in case of the conventional type of the test, the number of the items. With this well-written test specification, the validity and the reliability can be realized in a test.

4. Method

After an item bank of appropriate questions was constructed, a total of 57 testlets were made. The first testlet that students take contains 15 items. The remaining 56 testlets are composed of 3 items each. This CAT is to be used as a placement test for English classes at an engineering university. Since the items cover the following three areas, vocabulary, reading and grammar, each testlet is designed to include these three different types of questions. The CAT was taken by 112 students in July, 2009. The difficulty level of the first testlet was set for medium, that is the 0 of the testlet is 0, because the total number of testlets is 57 and not so large. The termination condition of the test is set at 0.40 in terms of standard error of the mean.

5. Results

First, as shown in fig. 5, the total number of items the students answered in the CAT range from approximately 15 to 42 except one student who completed large number of items. The average number of items needed to complete the test was 22.7 and much fewer than the regular final examinations given at the institute which are composed of 100 items to be answered within 90 minutes. Furthermore when compared with TOEIC (Test Of English for International Communication) which has been used as a placement test, the difference is even more unequivocal because TOEIC has 200 items to answer and the time duration is two hours.

Secondly, the time the students spent in the CAT was also much shorter than in the case of the other two examinations as shown in fig. 6. Except one person, all the students finished answering in less than 50 minutes; more than 90 % of the students finished in less than 2000 seconds (33 minutes).
Finally, in order to examine the validity of the CAT, the correlation coefficients between the CAT and the other two tests, an in-house final examination of the institute and TOEIC, have to be calculated. The correlation coefficient between the CAT and the final examination is 0.65, and that with TOEIC is 0.56. Since both of these coefficients are relatively high, the CAT can be concluded as an appropriate alternative for the other two tests if it were used as a placement test in this institute.
The first testlet and the result page are shown in fig. 9 and 10 respectively. The summary of the CAT in comparison with the final exam and TOEIC is given in the following table.

![ComputerAdaptive Test](image)

Summary Table

<table>
<thead>
<tr>
<th></th>
<th>items</th>
<th>Time (min.)</th>
<th>correlation with CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOEIC</td>
<td>200</td>
<td>120</td>
<td>0.56</td>
</tr>
<tr>
<td>Final Exam</td>
<td>100</td>
<td>90</td>
<td>0.65</td>
</tr>
<tr>
<td>CAT (ave.)</td>
<td>22.7</td>
<td>18.8</td>
<td></td>
</tr>
</tbody>
</table>

6. Conclusion

Moodle is already in use in many institutions in Japan, including high schools, universities and other institutes of higher education. In addition, the need for the efficient, cheap and reliable placement test is high because of the diverse English ability of the students. Some students have lived in English speaking countries for a long time and have very high English ability, and others are still in the lower intermediate levels even when entering university. As this study shows, the CAT made in the module for Moodle can be used by teachers at any institute in order to make a relatively valid in-house test. The CAT module has the potential to reduce the work load related to the placement test, and yet to provide accurate measurements of test takers.
However, some limitations of the CAT used in this experiment should also be reported. Since the number of items appropriate for the CAT is not large enough, two potential problems are observed. One is the termination condition of the test. Though the standard error of the mean is used in the experiment, some students had to needlessly answer many questions. Therefore, some other conditions such as the increasing rate of information quantity should be considered. Another problem is that the difficulty level of the items mismatches the level of some students. Because the items are relatively easy for some students, these students finished the CAT by answering all the items in the first testlet correctly to meet the terminating condition. In addition, the number of the items in the first testlet should be decreased so that these students can challenge more difficult testlets even when they have answered all the items in the first testlet correctly, which could lead to more accurate measurement of their ability. Therefore, both the quantity and quality of items are the critical conditions to construct a reliable CAT.

Because Moodle is provided freely as Open Source software under General Public License (GPL), anyone can obtain the data from existing Moodle sites that have CAT modules. One can also adjust these modules in the appropriate manner for their own purposes. Therefore, it is expected that the development of a CAT system in Moodle will stimulate CAT research, resulting in a better CAT system for varied applications.

References