≪ Research Note ≫

An Analysis of Effectiveness of TOEIC Practice with Computer Adaptive Testing (u-CAT) as a Web-Based CALL System

SATO, Yasuko *

Abstract

This paper investigates the effects of TOEIC practice with a Web-Based CALL System, which is called u-CAT (you, university, and ubiquitous Computer Adaptive Testing) as the way to enhance a student's motivation and to improve his/her scores. This author will define u-CAT as it is used in the Business English I to VI courses at Niigata University of International and Information Studies. This paper examines the pros and cons of u-CAT in the International Business English courses. Finally, the effectiveness of u-CAT in those courses will be discussed in this paper.

Key words：CAT, English teaching, e-learning, TOEIC TEST

Contents

1. Introduction

With the establishment of new faculty at Niigata University of International and Information Studies in 2014, new courses called International Business English I to VI have been adopted for students who chose English as one of their foreign language courses. There were four choices: Chinese, English, Korean, and Russian. In the second semester, they could start to take three required Intensive English courses called: International Business English, Media English, and Reading. In the run-up to full implementation of u-CAT (you, university, and ubiquitous Computer Adaptive Testing) in the class from 2014, the author innovated u-CAT as a Web-Based CALL system in two semesters of the International Business English I and II courses starting in the second semester of 2013.

---

* SATO, Yasuko 〔国際文化学科〕

2. What is u-CAT?

"u-CAT", which stands for you, university, and ubiquitous Computer Adaptive Testing, is an e-learning material that Asahi Press Inc. developed for the new TOEIC test with YBM SISA. COM Inc. (YBM/Si-sa) providing the TOEFL and TOEIC tests, from nonprofit ETS (Educational Testing Service) in Republic of Korea in 2007. "u-CAT" of the e-learning material for TOEIC test enabled YMB/Si-sa to receive a Korea e-Business Excellence Award by Ministry of Trade, Industry and Energy Republic of Korea. Asahi Press Inc. signed a sales contract in 2006, and systematically converted English-Korean into English-Japanese in the spring of 2007.

Originally, Linacre (1987) created Useful Computer Adaptive Testing (UCAT) with a program in Microsoft BASIC which determines examinees' competence by choosing appropriate questions to ask them from multiple choice questions about a subject. The program chooses the next question to give them based on each estimate. When the estimate is accurate enough, the test stops, and they are told how they scored. According to Mike Linacre and George Rasch (1995), the program allowed them to select and to ask customized questions to measure their ability promptly.

3. How to score u-CAT with IRT

The u-CAT program adapted Item Response Theory (IRT) known as a latent trait theory/strong true score theory. It is a paradigm for scoring of tests and questionnaires to measure abilities, attitudes, and other variables. Each question is based on the level of difficulty in a numerical way. Then the computer can give examinees a numerical estimate of their ability based on the difficulties of the questions they got right and wrong. They can directly compare this estimate with the other test takers, no matter which questions they had. (Linacre J.M., 1987) u-CAT and the TOEIC test calculate their scores altogether based on the questions from a diverse level of difficulty item pool.

4. u-CAT learning flow

u-CAT consists of seven steps, and offers eight hundred fifty questions in total for a learner. The flow of the learning begins with a diagnosis test. Assignments and trial examination are carried out three times in turn afterwards. The analysis of the learner's weakness is carried out by a diagnosis test and three trial examinations, and each result is reflected by learning contents of the homework just after that. Therefore, the most suitable questions upon individual ability are assigned by u-CAT, and the same questions to him/her are not given. The overall flow of the learning is as follows:

The System Producing Learners Effects

TRIAL EXAM

60 minutes

100 questions

Accurate assessment

Assignments

The exercises of highly-selected 150 questions to overcome their weakness

STEP 1 ～ STEP 7

STEP 1 Diagnosis test ― Judging English proficiency and Analyzing the
weakness (100 questions)

↓

STEP 2 Assignment ― Reinforcement of the weakness clarified on a diagnosis
test (50 questions × 3)

↓

STEP 3 The first trial exam ― Judging English proficiency and Analyzing the
weakness (100 questions)

↓

STEP 4 Assignment ― Reinforcement of the weakness clarified on a diagnosis
test (50 questions × 3)

↓

STEP 5 The second trial exam ― Judging English proficiency and Analyzing
the weakness (100 questions)

↓

STEP 6 Assignment ― Reinforcement of the weakness clarified on a diagnosis
test (50 questions × 3)

↓

STEP 7 The third trial exam ― The final check of study results (100 questions)

u-CAT provides a study management system to the teachers, which enables him/her to manage students ID, test/homework schedule and send class information and explanation; make an approach to their evaluations of test and assignments.

5. The pros and cons of CAT

Generally, the benefits of CAT (Computer Adaptive Testing) are more likely to be discussed, with a scaled-up version/group test simulated. Compared with those of PPT (Paper Pencil Testing) and CBT (Computer Based Testing) which all of the examinees can take the same questions in the same section, the benefits of CAT are divided broadly into the following five categories (Weiss, 1982; Green, 1983; Zickar, et al.: 1999):

1) Increased test security
2) Instantaneous score reporting
3) Variety types of item
4) Enhanced efficiency
5) Superior measurement

The first three categories 1), 2), and 3) apply to physical benefits, and are just as valid for CBT as well. Compared with those of PPT, the benefits of CBT and CAT are indicated. As one of the five benefits, category 3) is created by shifting multi-media from paper to a computer, which enables examinees to make the most of not only the characters and still images, but sound, motion pictures, and so force. The last two categories 4) & 5) include advantages of the measurement theory, which only CAT can offer them. In comparison with PPT and CBT, those provide test takers the same questions in the same section, and category 4) and 5) show the precision of the measurement without decline, even if the number of questions on CAT is reduced in a half. As for them, however, there is the aspect it is hard to be called advantages of CAT as the psychological side of examinees being taken into consideration.

Major problems with the CAT which an examiner and examinees often face will be classified into the next four categories (Kimura, T. 2013):

1) Implementation of a prior test; needing to take too much time and effort on the management of the analysis of an item pool
2) Computer illiteracy and feelings of unfamiliarity with answer methods used
3) Feelings of discriminatory treatment; feeling that not every examinee receives exactly the same questions
4) Psychological influence, discouragement, lowering the ratio of correct answers for enforced items

Category 1) indicates a problem wherein a prior test has already prepared and conducted. In order to raise the effectiveness and measurement accuracy of the CAT, the number of prepared questions should be greater than that of a non-adaptive test. Also, there may be a problem wherein some questions show more frequently when examinees take the CAT in a series. Therefore, new and diverse questions should be added to the item pool with appropriate time given so as not to require too much time. Category 2) points out a serious issue regarding computer literacy, which arises from low computing skills. In addition, unlike PPT and CAT does not allow examinees to go back to previous questions because of a rule forcing them to answer the test questions in a specific order. According to the result of a survey conducted in an International Business English I class in January of 2013, more than half of first-year students felt uneasy about using a computer or unconfident in their computing skills and answering methods. Those issues cannot be ignored since they affect the test results, even though they may occur in only a few cases and vary depending on the extent to which students are uneasy using a computer. Also, due to CAT's answer process, it is difficult for them to evade CAT's algorithm. Therefore, the author expects that giving well-informed instruction to test takers in advance and giving them a chance to take the CAT frequently may be helpful in easing their fear of using or taking a test on a computer. Category 3) indicates that as well as the instruction listed for category 2), it is crucial for an examiner to provide well-informed instruction about questions and marking methods. These

problems can be solved by helping examinees understand CAT well. The last category, 4), shows that taking the test can provide motivation for learning and self-efficacy. From this, the author found that fear and other expected problems can be solved or managed by an examiner.

6. Conclusion

According to the results of one-year use of u-CAT by fifty six freshmen majoring in English from September 2013 through July 2014, 88.8% of students improved the final score on the fourth TOEIC test compared to the first one or second attempt. Approximately 85% thought that exercises were very effective in improving their TOEIC score. Regarding the last question about whether they were willing to use this program in the next semester, 70 % gave an affirmative reply. However, educators who will use or are using u-CAT will need to give thoughtful consideration to the four major problems pinpointed in the section 5, and correct those negative aspects by judging the situation in the class. Also, considering the fact that detailed findings have not yet been obtained in this research, extensive future work should be done on the analysis of effectiveness of TOEIC practice with u-CAT, while implementing the web-based CALL system.

References

Calcote, A. S., et al. (2012). *E-learning heiyou: Shin TOEIC test.* (Multi-strategy Learning for the TOEIC Test, 44-79). Tokyo: Asahi Press.

Cho, I. (2007). *Kakushindo tesuto-how to koumoku hannou riron.* (Measurement Testing and IRT Theory). Tokyo: Univ. of Tokyo Press.

Green, B. F., Jr. (1983). The promise of tailored test. In: Wainer, H. and Messick, S. (Eds.), *Principles of Modern Psychological Measurement* (pp. 69-80). Hillsdale, NJ: Laurence Erlbaum Associates.

Kimura, T. & Ohnishi, A. (2011). Moodle UCAT beta version: a computer-adaptive test module based on Rasch model, 22-25. JALT CALL 2011.

Kimura, T. (2013). Senzai-ranku riron wo mochiita konpyuta tekiou-gata tesuto no tame no arugorizumu no teian to jitsusou. (Proposition and Implementation of Algorithm for Computer Adaptive Test Based on Latent Rank Theory, 10-34). The dissertation of PhD in Human Sciences of Waseda University, Tokyo.

Koyama, Y & Akiyama, M. (2009). Developing a Computer Adaptive ESP Placement Test Using Moodle. eLEARN 2009, 940-945.

Linacre, J. M. (1987). UCAT: a BASIC Computer-Adaptive Testing Program. MESA Psychometric Laboratory. (ERIC ED 280 895).

Linacre, J. M. (1995). Computer-Adaptive Testing, CAT: A Bayesian approach. *Rasch Measurement Transactions*, 9:1, 412.

Shiba, S. (1991). *Koumoku hannou riron – kiso to ouyou.* (IRT Theory: Basic and Application). Tokyo: Univ. of Tokyo Press.

Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing.

Applied Psychological Measurement, 6(4), 473-492.

Zickar, M. J., et al. (1999). The development of a computerized adaptive selection system for computer programmers in a financial services company. In F. Drasgow and J. B. Olsen (Eds.), *Innovations in Computerized Assessment* (pp. 7-33). Mahwah, NJ: Erlbaum.

Questionnaires for u-CAT users

Q1 Do you think this tool is useful? Yes/No

Q2 For those who answered YES to Q1, write any idea.

_____

_____

_____

Q3 For those who answered NO to Q1, write any idea.

_____

_____

_____

Q4 Do you think the exercises are effective to improve your TOEIC score?

_____

_____

_____

Q5 Were there any changes to the score on your fourth test comparing the first/second one?

_____

_____

_____

Q6 Are you willing to use this program in the next semester again? Write any idea.

_____

_____

_____