Features of J-CAT (Japanese Computerized Adaptive Test)

Shingo Imai
Tsukuba University

Sukero Ito
Tokyo University of Foreign Studies

Yoichi Nakamura
Seisen Jogakuin College

Kenichi Kikuchi
Toho University

Yayoi Akagi
Yamaguchi University

Hiromi Nakasono
Shimane University

Akiko Honda
Ritsumeikan Asia Pacific University

Takekatsu Hiramura
Tokyo Institute of Technology

2009 GMAC® Conference on Computerized Adaptive Testing

Abstract

J-CAT (Japanese computerized adaptive test) has been developed and used as a proficiency test of the Japanese language. It is accessible from anywhere and any place through the Internet. The test consists of four sections—vocabulary, grammar, reading, and listening. It uses sounds, pictures, and movies in order to increase the authenticity and validity of the test. The score is calculated according to item response theory, and converted to 0-100 points in each section and 0-400 points in total. J-CAT provides a score sheet in a PDF file immediately after completing the test; examinees may download it on their computer and/or print it.

Acknowledgment

Copyright © 2009 by the Authors

Citation

Imai, S., Ito, S., Nakamura, Y., Kikuchi, K., Akagi, Y., Nakasono, H., Honda, A., & Hiramura, T. (2009). Features of J-CAT (Japanese Computerized Adaptive Test). In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

Shingo Imai, University of Tsukuba, Graduate School of Humanities and Social Sciences, and International Student Center, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, JAPAN.
E-mail: imai@intersc.tsukuba.ac.jp

Features of J-CAT (Japanese Computerized Adaptive Test)

The J-CAT (Japanese Computerized Adaptive Test) has been developed to diagnose the proficiency level of Japanese as a second/foreign language. It has a two-way delivery system; by Internet and by LAN. The test consists of four sections: Listening, Vocabulary, Grammar, and Reading. Test durations vary from 40 to 90 minutes, depending on the performance of each examinee. The J-CAT system is built with two main parts—the engine and the item bank. By changing the item bank with other subjects, the system is applicable to any language test or subject, provided that the test format is multiple choice.

Most of the foreign students who enter a Japanese university take Japanese language courses. Placement tests are required for the allocation of students to their suitable level of classes. Development and operation of paper-and-pencil placement tests is a time-consuming task for language teachers every year. In practice, the same test is usually used year after year in most institutions to avoid the cumbersome procedure of making a new test, although this is not the ideal situation for test security. However, if a new test is given each year, it would not be certain if the different sets of tests are compatible with each other, unless test items are equated/linked based on item response theory (IRT). The J-CAT project was launched to provide the solution to these problems and to achieve the accountability of the test, which is increasingly demanded in the present society.

In this paper, we will discuss some features of the J-CAT, focused on the viewpoint of test administrators, including teachers. The features discussed include registration method, item bank management, algorithm management, and information of test results. We will illustrate how this system registers examinees and authenticates them. We will also discuss how we manage the item bank, such as uploading items, setting IRT parameters, and setting answering time limits for each item. The testing algorithm, such as the starting rule and stopping rule, is manageable according to the administrator's requirements. The J-CAT provides useful information to analyze the result of a test. We highlight some features of data available for an administrator and how an administrator might utilize the result information. We also show a sample score sheet which is automatically produced at the end of a test. The examinee can either print the report or save it in PDF format on his/her computer.
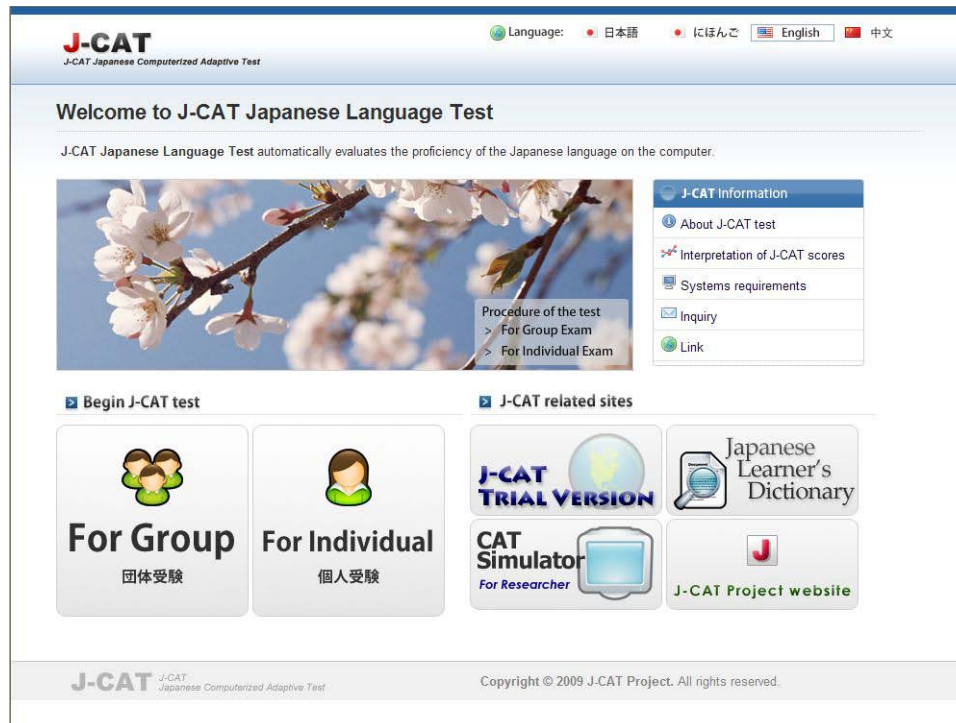
Registration

There are two types of registrations—for an individual and for a group. Individual registration is for an examinee who wants to check his or her own proficiency level of Japanese. Group registration is for institutions that want use the J-CAT as a placement test or a proficiency test and administer the test in a computer-equipped room on a specific date. Figure 1 shows the opening page of the J-CAT system.

An individual examinee can access the J-CAT URL from anywhere in the world at any time. Before registration, the system and Web connection of the client's computer (i.e., the examinee) is checked to see if they satisfy the requirements. The J-CAT requires Internet Explorer 6.0+ or Mozilla Firefox 1.5+. The monitor must be 1024 x 768 pixels of resolution or higher. If the resolution is poor, items will not fit on the screen. Adobe Flash Player must be installed. The computer must be able to run Java scripts. As the J-CAT has a listening section, it requires a decent sound system for clear sounds. All of these requirements are checked automatically before registration, and warnings are displayed if the examinee's computer does not meet the

requirements. After the checking, the volume setting is prompted with a sample sound. A registration window (Figure 2) asks the registrant to fill out their personal information: name, email address, mother tongue, organization/institution, and if he or she has previously taken other tests of the Japanese language and, if yes, their results. These data are useful for an administrator to examine the background of examinees. Optionally, an examinee can fill out his/her birth date, which will appear on the score report at the end of the test. This allows the examinee to keep the test score for proof of their level of proficiency.

Figure 1. Opening Page of J-CAT

Figure 2. Registration

The registration information is stored in the database and the request for approval is automatically sent to the administrator by email. The administrator can approve the request by clicking the URL automatically created by the server, which appears in the email. If there is any doubt of misconduct in the request, the administrator can either reject the request or send a warning message to the registrant's email address. Once the request is approved by the administrator, a password is created and sent to the registrant's email address automatically by the system. Then the registrant may login and start the test by using the password and their ID, which is the registrant's email address; the system recognizes and manages each examinee by his/her email address. This system prevents the same examinee from accessing the test contents repeatedly for inappropriate purposes. Disguised email addresses can also be detected by the administrator since the IP address appears on the request email sent to the administrator.

In the case of group registration, one common password is provided to the group to skip the approval processes for each examinee. Examinees are requested to fill out parts of the personal information. Email addresses are not requested since the password is already provided and not sent by email. Each examinee in the group must provide a non-identical name since the examinee's ID is distinguishable solely by name in group registration.

Item Bank Management

Uploading Items

Items can be uploaded either individually or as a set. The system supports plain text, picture, sound, and movie formats. Some items of the J-CAT require "ruby" or annotation placed above characters to show pronunciation. There is a ruby tag in HTML; however, the maintaining of correct alignments and adequate readability in HTML format is not easily achieved if the resolution of monitors varies from one examinee to another. The J-CAT avoids this problem by converting text items into GIF format. The use of a picture format also prevents the problem of inconsistency of fonts used/installed in client terminals.

Parameter Setting

There are three slots to fill in with item parameters for each item. The J-CAT supports one-parameter, two-parameter, and three-parameter IRT models used for ability estimation. When all three parameters for items are provided, i.e., all three slots are filled, ability estimation is done based on the three-parameter model. If the difficulty parameters and discrimination parameters are provided while 0.0 is given to the guessing parameters of all items, estimation is done based on the two-parameter model. The one-parameter model is applied, if the difficulty parameters are provided for each item, while the discrimination parameters are set as 1.0 and the guessing parameters are set as 0.0 for all items.

The administrator can manage the characteristics of the item bank either by distributing the items with a wide range of difficulty levels to make the target band of examinees wide or by concentrating on items with similar difficulty parameters if the test focuses on a particular level of ability, or a cut score.

Time Limit Setting

To avoid unnecessary extension of the test duration caused by examinees who linger over particular items for an excessively long time, the administrator can set a time limit for answering the questions. The time limit can be set differently for each item. For instance, the administrator can allot 30 seconds for a short item and five minutes for a long item.

Copy Protection

Using Adobe Flash and a PHP program with an HTML header in the system prevents copying, downloading, and caching at the user terminal. The image of this system is shown in Figure 3. The outputter.php, which is called by Flash files, opens files in the Data folder and returns data to the Flash files. The HTTP header in the outputter.php prevents the data from being kept in the cache.

Figure 3. Copy Protection
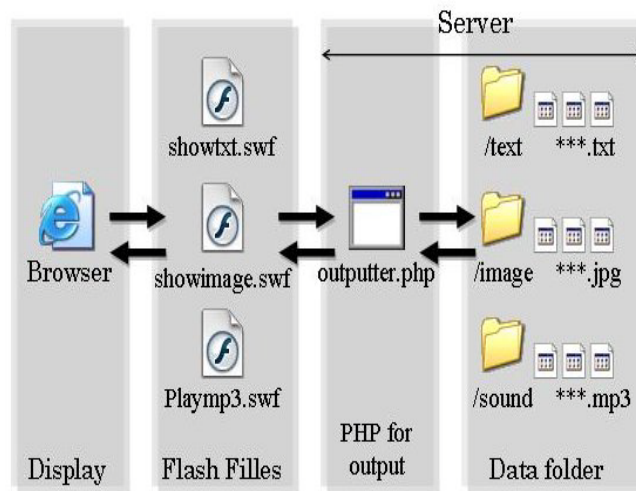
Algorithm Management

Starting Rule

The starting rule can be set by the administrator. It is possible to start from the item of middle difficulty, or from a test with a few items with a range of difficulties and tentatively estimate the examinee's ability in order to decide the starting level. The latter method can lower the exposure rate of the items of middle difficulty and also reduce the test length by adjusting the tentative, but suitable level of the examinee.

Stopping Rule

There are two stopping triggers: (1) the posterior standard deviation, which functions as estimation error, and (2) the maximum number of administered items. The test will stop whenever the examinee reaches either of these. For instance, if the estimation error is set as .3 and the maximum number is set as 20, whichever comes first works as the stopping trigger for the examinee. Either of the settings can be virtually nullified by setting it with an extreme value. For example, if the estimation error is set to 0.0, it is always the case that the test is terminated when the number of items reaches the maximum number. If the maximum number is set extremely highly, such as 200, it is almost the case that estimation error solely functions as a stopping rule. However, the latter type of setting is not recommended in practice unless thorough simulations are conducted beforehand to make sure that no unusual cases occur in the operation of the test. Although the estimation error usually reduces in reverse proportion to the number of items applied, it is rare but possible that an examinee might answer items in an unexpectedly bizarre manner and hardly clear the boundary of estimation error.

Item Selection

The J-CAT system can use maximum information for item selection; however, it does not in order to control item exposure. The present setting of the J-CAT selects the item based on simply the minimal difference between the estimated θ and the difficulty parameter. The simulation by using real data of the J-CAT indicated that item selection based on maximum information causes

a serious exposure problem because items with high value of the discrimination parameter are over-exposed, while items with low discriminations are rarely or never used. Item selection based on the difference between estimated θ and the difficulty parameter extends the test length, but does so only by a few items without deteriorating the exposure rate. We considered that the failure of exposure control is more harmful for the operation than a slightly longer test.

θ Estimation

Bayesian expected a posteriori (EAP) is used to estimate θ. Maximum likelihood estimation (MLE) is not used because it cannot estimate θ for all correct or incorrect responses. Bayesian estimation does not run into this problem. Therefore, from an operational point of view, it is safer to use Bayesian EAP rather than MLE. A problem of Bayesian EAP is that its estimates tend to be skewed or pulled toward the mean of the θ prior distribution. In the operation of the test, however, this tendency is considered negligible because it is reduced as the number of administered items increases.

Results Information

The score report is provided at the end of the test in PDF format. The examinee can print it and/or save it on his/her computer. The score sheet (Figure 4) provides the scores of each section and total score, with a brief explanation of how to interpret the scores. The administrator can see the results of all examinees on the administrator's page. The administrator can also download the results and information, such as profiles of examinees, in CSV format which can be used for various analyses. For instance, determining the cut score is possible by sorting the examinees by scores. Correlation analyses between profile information and test scores are also possible. Since the scores are (theoretically) invariant in IRT, it is possible to perform time-series analyses of proficiency improvements.

Figure 4. Example of Score Report

J-CAT score	Proficiency Level	JLPT 日本語能力試験
-100	Basic 初級	Level 4
100-150	Pre-Intermediate 中級前半	Level 3
150-200	Intermediate 中級	
200-250	Intermediate-High 中級後半	Level 2
250-300	Pre-Advanced 上級前半	Level 1
300-350	Advanced 上級	
350-	Near Native 日本語母語相当	

- "Novice" learners can use fixed expressions such as greetings.
- "Basic" learners can exchange basic ideas.
- "Intermediate" learners can manage daily communication.
- "Advanced" learners can manage academic and professional communication.

J-CAT Project.

Conclusions

The J-CAT is delivered both via Internet and Intranet. The Intranet version is basically a clone of the Internet version. The Internet version is required in order to make users from any places in the world accessible to the test and provide them with immediate feedback. We came a long way to reach the present system and items of the J-CAT, and there is no doubt that the development will continue on both system and contents. Authenticity of items is one of the main interests of future development. Innovative items such as using movies are also under development. Our experience so far clearly indicates that we are on the right track and the future of the J-CAT is promising.

References

Hambleton, R. K. & Swaminathan, H. (1985) *Item response theory: Principles and applications*. Kluwer-Nijhoff.

Imai, S. (2008). *Development of Japanese computerized adaptive test*. The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (pp. 821-824). Shimonoseki, Japan.

Wainer, H. (2000) *Computerized adaptive testing: A primer* (Second edition). Lawrence Erlbaum Associates.

Shiba, S. (1991) *Komoku Oto Riron: Kiso to Oyo.* Tokyo University Publishing Association.