

The 7th JFKL Conference on Japanese Language Education
in collaboration with Ambang Asuhan Jepun, Pusat Asasi Sains, Universiti Malaya
and
Urawa Teacher Training Program Reports

第7回 JFKL 日本語教育研究発表会・浦和研修報告会

PROCEEDINGS

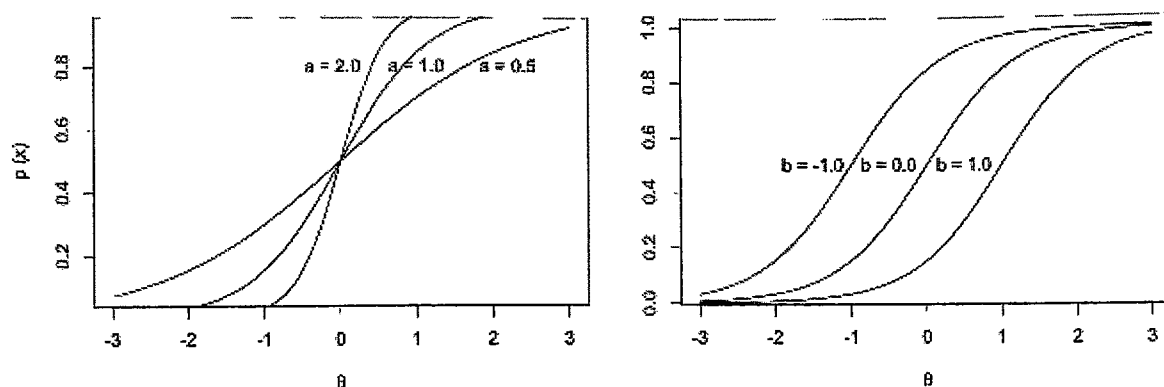
October 9-10, 2010
Ambang Asuhan Jepun, Pusat Asasi Sains, Universiti Malaya



基調講演

KEYNOTE LECTURE

Fig. 1: Item Characteristic Curve



5. A sketch of J-CAT

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 45 to 90 minutes, depending on the performance of each examinee. The CAT system selects questions items suitable to each examinee. Unnecessary items for the evaluation such as items that are too difficult or too easy for the examinee are not given.

5.1 Registration

There are two types of registration; for an individual and for a group. The individual registration is for an individual 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 2 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 3) 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 2. Opening Page of J-CAT

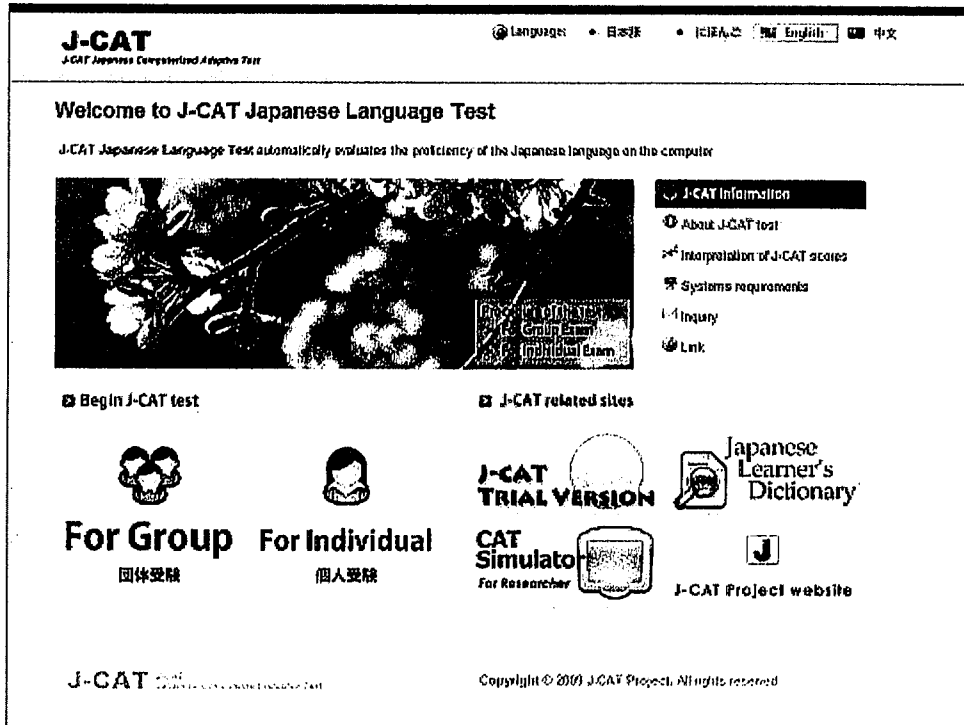


Figure 3. Registration

J-CAT
J-CAT Japanese Computerized Adaptive Test

Registration (For individual)

Registration Form
Fill in the blanks. Use alphabets and/or numbers. Do NOT use Japanese fonts.

Name:
(E.g. Ichiro Suzuki)

Email address:
(E.g. example@exam.com)

Mother tongue:
(E.g. Chinese, Chugokugo)

Affiliation:
(E.g. Asahi University, Asahi Daigaku, SEI Co. Ltd., Kobun Kagaku SEI)

Department, Section:
(E.g. Economics, Kasei, Sales Division, Eigyoebu)

Date of birth (Optional): Year: Month: Date:

Have you passed JLPT? Yes No

(ec 3)

(ec 80)

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 - the registrant's email address. The system then recognizes and manages each examinee by this ID. 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.

5.2. 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.

5.3. 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 will work 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.

5.4. Item Selection

The J-CAT system can use maximum information for item selection, yet 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 discrimination 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 slight prolonging of the test.

5.5 .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. However, In the operation of the test, this tendency is considered negligible because it is reduced as the number of administered items increases.

5.6. Result 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

成績証 Score report J-CAT® Japanese Language Test
<http://www.j-cat.org/>

あなたの J-CAT スコア は以下の通りです
Your J-CAT scores are as follows.

名前 / name		聴解 / Listening	16
生年月日 / birth date		語彙 / Vocabulary	34
所属 / affiliation	Yamaguchi University	文法 / Grammar	46
登録日 / register date	2008/04/08	読解 / Reading	20
受験日 / exam date	2008/04/08	合計 / Total	116

(watermark: 70b9f15cae0732bf8acc8c9daf12b04)
J-CAT Project.

Interpretation of J-CAT score

J-CAT score	Proficiency Level	JLPI 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	Level 2
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	Level 1
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.

6. Conclusions

We consider this test as a “service.” Our test development project incorporates an innovative idea of service science. J-CAT has been adopted in more than 30 universities, colleges, and institutions in Japan and other countries as a placement test or an achievement test. Many individuals have taken the test for self-evaluation. We receive feedback form all over the world, which we appreciate as valuable information to improve the quality of our service.

This year, we have launched a new project. This is the development of an automated speaking evaluation system by using speech recognition technologies. This adds the evaluation of production ability to J-CAT which evaluates only receptive proficiency at present. We hope J-CAT will continue to provide the high quality service for those who want to evaluate one’s own or one’s students’ proficiency of Japanese language without cost, time, and space limitations.

Notes

This paper is based on and have adopted some parts of Imai, S., et al. (2009) and Hiramura, T. et al. (to appear). J-CAT can be accessed at <http://www.j-cat.org>.

This project has been supported by Yamaguchi University Foundation The Grant-in-Aid for Scientific Research (A) (18202012) MEXT (The Ministry of Education, Culture, Sports, Science and Technology).

References

- Baker, F. and Kim, S. (2004) *Item Response Theory*. Basel, NY: Markel Dekker Inc.
- de Ayala, R.J. (2008) *The Theory and Practice of Item Response Theory*. New York, NY: Guilford Pr.
- 今井新悟・伊東祐郎・中村洋一・菊地賢一・赤木彌生・中園博美・本田明子 (2010) 『J-CAT Japanese Computerized Adaptive Test－日本語能力をコンピュータで測る－』 山口大学留学生センター
- 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*. pp.1-8.
<http://www.psych.umn.edu/psylabs/CATCentral/>
- Imai, S. (2008) ‘Development of Japanese-Computerized adaptive test.’ *The 23rd International Technical Conference on Circuits/Systems, Computers and Communications*. Shimonoseki, Japan. pp.821-824
- Hiramura, T., Imai, S., Kikuchi, K., and Mayekawa, S. (to appear) ‘Technical Problems of Internet-Based Japanese Language Test and Their Solutions.’ *E-LEARN 2010, Association for the Advancement of Computing in Education*, Florida.