

Developing an online placement test for the Japanese program at UC Berkeley: Norming test analysis

Spring 2015 BLC Fellows:
Yasuko Konno Baker
Noriko Komatsu Wallace

I. Introduction

A. Producing the online placement test

During the winter semester of 2012, we began to create an online placement test for six levels of the Japanese language program. There were two primary reasons:

- 1) an online placement test is meant to be the most objective method of measuring student language proficiency levels
- 2) there was a recommendation in the external review of our department that we replace portions of the current placement test with an online test in order to reduce instructor workload.

One step in the process of creating an accurate online placement test is to have students take “norming” tests. The norming tests are the best way to make an accurate online placement test. We administered four such norming tests, analyzing them after they were given to modify and improve test content. The last of the four was given April 28–30, 2015.

B. The six levels of the placement

The final norming test incorporated questions for all six levels of language proficiency. It had 558 questions in its database. As the table below shows, what questions assess differs according to proficiency level, with a general pattern of moving from simple, concrete items in the basic levels to more complex and more abstract items in the higher levels.

Course name	Level	aspects emphasized when designing the questions
Elementary Japanese	Level 1	things and events around the students
	Level 2	
Intermediate Japanese	Level 3	speech style, stating opinions, abstract concepts
	Level 4	
Advanced Japanese	Level 5	coherence and flow in expression, very abstract concepts, social issues, authentic materials
	Level 6	

C. The online portion of the Japanese language program placement test

This online test is one part of a three-part placement test that the Japanese language program requires of students who are joining the program from elsewhere and who do not plan to begin at first semester (Level 1) of language instruction. These three parts are:

- 1) online test
- 2) oral interview
- 3) writing sample

The questions we compose are based on our program's curriculum which uses the following textbooks:

Levels 1 and 2 — *Elementary Japanese, Volumes One and Two* (Hasegawa, et al, 2005–2006)

Levels 3 and 4 — 上級へのとびら [*Jokyu e no tobira, Tobira: Gateway to Advanced Japanese*] (Oka, Tsutsui, et al, 2009)

Levels 5 and 6 — 新 中級から上級への日本語 [*Shin chukyu kara jokyu e no Nihongo, Authentic Japanese: Progressing from Intermediate to Advanced. New edition*] (Kamata, Beuckman, et al, 2012)

However, we wish to note that we incorporate Japanese culture indirectly into the questions where possible.

The online test uses a multiple-choice format because problems can arise when we ask students to input Japanese script into their various devices. Further, we wish the test to be computer-gradable because this eases instructor workload and generates a database that can be analyzed.

II. Analysis of the Norming Test

A. What is a “norming test”?

The purpose of a placement test is to assign each student to the language course (level) best suited to his or her current language ability. Since our online placement test is in a multiple-choice format and therefore generates purely mathematical results that are not reviewed by an instructor on a question-by-question basis, it is critical that we compose multiple-choice questions that definitively test actual curriculum content at the intended target level of each question. The concern arises,

however: How do we know if the questions we have created are truly at the levels we think they are; in other words, how can we objectively conclude that they are based on the actual curriculum content that we expect students to learn at each level of the program?

Norming tests provide the data that tell us actual performance for each question in the placement test database. Through the analysis of these results, we can determine whether, in terms of its level, a question is satisfactory as is, or needs to be adjusted, or needs to be re-designated as a question at a different level, or perhaps even needs to be discarded from the database altogether for some reason. This is called “norming” the placement test questions and so the test itself is called a norming test. Norming tests are taken by students who have actually completed the target levels. Their performance will confirm whether we composed good test questions at each level. In other words, the first step in making a placement test is, of course, to compose questions for each level of the test database. However, the critical second step is to norm those questions.

But we cannot simply administer a norming test that has, for example, only Level 2 questions to students who have completed Level 2 of the curriculum. It is not possible to norm the questions with this method. Instead, the method used in a norming test to determine whether a question discriminates between levels has two important components. First, each norming test includes *two* levels of questions. Second, we must compare results across a *set* of norming tests that are taken by students completing different levels of the curriculum.

An example might help illustrate how this works. Let’s imagine that we ask students currently completing the Level 2 curriculum to take a trial run of a placement test that has, say, mostly Level 2 questions but also some Level 3 questions. (The distribution ratio of Level 2:Level 3 questions used in our case was 4:1—too few Level 3 questions will not give us a reliable measurement but too many Level 3 questions will overly discourage the test taker during the taking of the test.) In theory, these students should be able to answer the Level 2 questions and should not be able to answer the Level 3 questions. If this is what actually occurs, we can know that the incorrectly answered Level 3 questions are *above* Level 2 and the correctly answered Level 2 questions are at Level 2 or lower. However, we know nothing more than this.

There are two things we still do not know. First, we do not know whether the *incorrectly answered* Level 3 questions are Level 3, or Level 4, or even higher. Second we do not know whether the *correctly answered* Level 2 questions are actually Level 2 or perhaps Level 1.

However, since each norming test includes two levels of questions, when the norming tests are *considered together*, we can now accurately determine a question level. (This will be called ID, “Item Discrimination” in the next section.) To illustrate how this works please consider the diagram below.

II. Analysis of the Norming Test

A. What is a “Norming Test”?

One norming test with only one level of questions		Two tests from a norming test set	
Students who have just completed the Level Two curriculum take a trial placement test for Level Two (a norming test)	...with Level Two questions	Students who have just completed the Level Two curriculum	... with Level Two and Level Three questions. Includes “Question A” (designated as a level three question)
		Students who have just completed the Level Three curriculum	... with Level Three and Level Four questions. Also includes the same “Question A”

Let’s say that “Question A” was composed by us with the intention of composing a Level 3 question and has been designated in the database as such. Let’s further posit these test outcomes from two different tests that both used the question: the question could *not* be answered by Level 2 students in one norming test, but *was* answered correctly by Level 3 students in another norming test. In this case, it is likely that the level calibration is correct. However, if the Level 3 students *also* cannot answer the question, then perhaps it should be re-designated as a Level 4 question or perhaps the question itself is not effective for some reason and needs to be reconsidered.

B. Using the data from the Norming Test

Add a New Question

This question has audio, video, image, or longer reading text ?

Title

Select from existing or enter new instructions: Frequently used instruction?

Language

Difficulty Level

Category

When creating questions, please do not add a., b., c., or 1., 2., 3., to the text of the answers: answers for each question will appear on exams in random order.

Question Type

Question Text

B-1

Figure B-1 shows the interface for creating questions using the Berkeley Online Language Testing (BOLT) application. The creators of the test need to fill in the target level, category of the content such as grammar, reading or *kanji*, instructions, question, and answer choices for a multiple-choice question. The placement test questions all contain four answer choices, which still gives us 25% accidentally correct answers, but we consider it is sufficient to evaluate the data. Instructions are given in English for all the questions, but Japanese instructions are also given for

some of the questions for which more specific instructions are required. All the answer choices and reading passages are in Japanese.

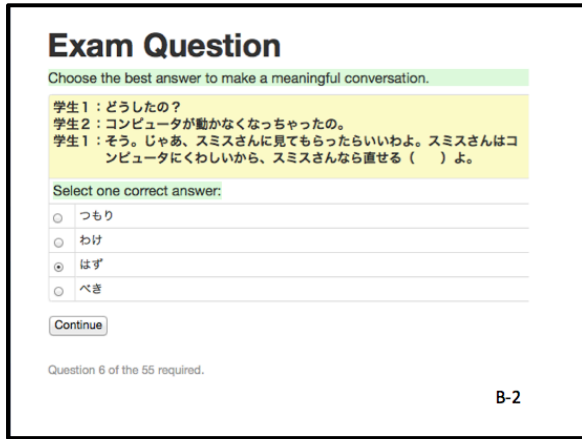


Figure B-2 shows what it looks like to the students when they take the test. They see the question part and answer choices for each question, and choose one answer by clicking the white circle. The results of the test are automatically saved in the database in the same program, and analysis reports are available in various forms.

B. Using the data from the norming test

Questions Analysis Norming Exams

4 Questions

Question ID	Prompt-Question Text	Difficulty Level	Student Level 4		Student Level 5		ID Score
			IF Score	# Responses	IF Score	# Responses	
3993	めったに忘れない 電気を消し忘れることは () ないのに、今朝起きたら電...	5	0.27	26	0.33	18	0.06
3994	悪いわけではない 私は外食がきらいな () ないが、勉強が忙しいので、い...	5	0.88	24	0.89	18	0.01

B-3

Now, we will explain what we did with the data after the norming test. Figure B-3 shows a screen shot of the Question Analysis for two grammar questions with *めったに～ない* (*mettani ~ nai*) “rarely~” and *～わけではない* (*wake dewa nai*) “it does not mean that ~”. This report shows the record for each question. Each row shows one question. The Item Facility, or IF score, is the proportion of correct answers for the question at each student level. We can see that the Item Facility score at Level 4 for this question, and the IF score for students at Level 5 for the same question. The Item Discrimination, or ID score, is the difference between IFs of students at the target level and those at one level below. The bigger the ID is, the better this question differentiates between Level 4 and Level 5 students. Ebel (1991) says, “One

of the two direct uses that can be made of indices of discrimination is in the selection of the best (that is, most highly discriminating) items for inclusion in an improved version of the test.” He suggests the indices of item discrimination as follows:

Index of Discrimination	Item Evaluation
0.40 and up	Very good items
0.30 to 0.39	Reasonably good but possibly subject to improvement
0.20 to 0.29	Marginal items, usually needing and being subject to improvement
Below 0.19	Poor items, to be rejected or improved by revision

We analyzed the IF and ID scores for each question, corrected any mistakes in the question, and made improvements in questions which had low ID scores and did not differentiate between students at the target level and those one level below. We aimed at an ID score around 0.25 for editing for improvements; however, some of the questions whose ID score were below 0.25 were kept as they are, and the decision whether they need to be edited was postponed until the next norming test results. In this regard, Brown (2011) suggests, “Of course, Ebel’s guidelines should not be used as hard and fast ‘rules,’ but rather as aids in making decisions about which items to keep and which to discard until a sufficient number of items has been found to make up whatever norm-referenced test is under development.”

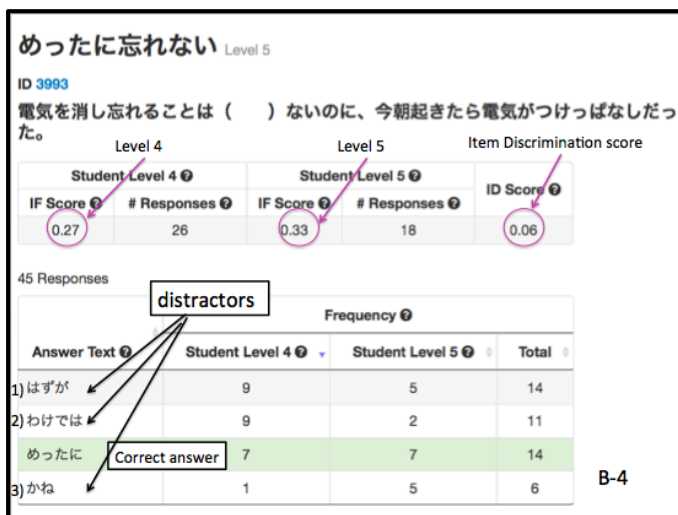


Figure B-4 shows a more detailed report for the question with *めったに～ない* (*mettani ~ nai*) “rarely~” which tells us many things about the question and so is very useful for analysis. The Item Facility score for Level 4 is 0.27 and for Level 5, 0.33, both of which are very low, which means that students at both levels did not do very well on this question. The Item Discrimination score (ID) for the same question is 0.06, which is also very low. This means that this question does not differentiate well between the students at Level 4 and those at Level 5.

This report also shows exactly how many students chose each distractor. At Level 4, nine students chose distractor #1 and another nine chose distractor #2, which is greater than the number of students who chose the correct answer. At Level 5, the correct choice was chosen most often, but two of the distractors were each chosen by five students. If we look at the totals for each answer choice, we can see that the first and second distractors are too attractive. Looking at these numbers, it is clear that this question has problems, so we need to edit it.

Figure B-5 shows the interface for editing questions in the Berkeley Online Language Testing (BOLT) application, which appears after the test has been run. It has buttons for “minor change” and “major change”.

When we found questions with problems, we could fix them in one of three ways; we had to decide whether we needed to make a “minor” change, or a “major” change, or even discard the question completely.

When we found questions which only required corrections which we believed would not greatly affect the students’ answer choices, such as a few changes in *kanji* or vocabulary, we just fixed them and marked it as a “minor change”. When we do this, the revised question appears on the next norming test, and the results from that will be added to the same record as the earlier results. In other words, after the next norming test, we will have more data for this question in addition to the data we already have. Even though the new results are attached to the same record in the database, we can still compare the new data with the accumulated data from earlier norming tests, in order to judge whether this question effectively discriminates students at one level from the other.

However, some questions needed larger corrections, such as replacing key *kanji* or vocabulary, or replacing a large part of the question. In those cases, which we believed *would* affect the students’ answer choices, we marked it as a “major change”. A question marked as a major change goes into the archive and will not appear in the next norming test, but the revised question will appear instead of the

archived one. However, the data from the archived question is still available within the database so that we can refer to it later.

Questions Analysis

No more new data

Question ID	Prompt-Question Text	Difficulty Level	Student Level 4		Student Level 5		ID Score
			IF Score	# Responses	IF Score	Responses	
3993	めったに忘れない 電気を消し忘れることは（ ）ないのに、今朝起きたら 電...	5	0.27	26	0.33	18	0.06
3994	嫌いなわけではない 私は外食がきらいな（ ）な いが、勉強が忙しいので、い...	5	0.88	24	0.89	18	0.01
4638	めったに忘れない 電気を消し忘れることは（ ）ないのに、今朝起きたら 電...	5					
4639	嫌いなわけではない 私は外食がきらいな（ ）な いが、勉強が忙しいので、い...	5					

Archived questions

Revised questions

New data will be added here.

B-6

In this case, the data in the archived questions is still shown in the database as shown in Figure B-6. But new data from the next administration of the norming test will not be added to them. The lower part of Figure B-6 shows the revised questions, but no data has been added to them yet. That part of the database will be filled in the next time the norming test is given.

The last choice in editing the test was to delete problematic questions. When the question ID was very low and it did not differentiate between student levels, and we also did not see a way to improve the question, we simply had to delete it.

We went through this process on each of the questions twice. As a result, we have deleted 37 questions, have made 57 major changes, and have made minor changes on about 70% of the rest of the questions.

C. Discoveries while analyzing norming tests

Analyzing norming tests was a very interesting learning process. We made discoveries we did not anticipate. We would like to share two question-design errors we noticed that we had made, and some new ways of thinking about our future curriculum that were also the result of analyzing the norming tests.

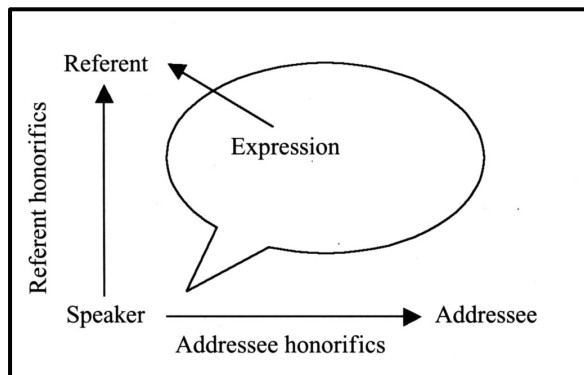
1) *We should test only fully practiced content; we should avoid testing content that has only been introduced.*

The first of these question-design errors was not keeping in mind the important distinction between content that is only introduced to students at any given level and content that is fully practiced by students at a level. Some test questions were built around content that is only introduced at a certain level but not yet fully

practiced. The poor performance on certain questions, we realized, was a result of us not making a clear distinction between content that had been only introduced and content that had been well practiced through drills and such.

As an example, we would like to share an issue we had with a sub-category called “Polite expressions for interaction”. This sub-category tests a complicated aspect of Japanese, and many students find the concepts related to it quite challenging. Therefore, we begin to introduce it at the start of Level 1. For example, on the first day of this elementary class, we teach greetings. And, during the process of teaching these basic greetings, we also teach the difference between 「おはよう」 (ohayo) which means “good morning” and 「おはようございます」 (ohayo gozaimasu) which is a polite way to say “good morning”, in other words a “polite expression for interaction”.

In this way, students learn, beginning with the early stages of studying Japanese, that “polite expressions for interaction” is a key cultural concept. At Level 1, we introduce polite expressions in ways such as the above. At Level 2, we add specific honorific expressions such as two types of honorifics. The less complicated one is “addressee honorifics” used when the speaker wishes to show respect directly to the person to whom he or she is talking. The more challenging type of honorifics is when the speaker wishes to show that he or she has respect towards someone who is referred to within something he or she is talking about. This is called “referent honorifics”. The diagram below from Hasegawa (2005) illustrates the difference:



This is an example from the test database in the sub-category, given in a Level 2 test. On the above diagram, “Smith” is the Speaker, “Yamada-san” is the Addressee, and “teacher” is the Referent:

Example 1

Choose the best answer to make a meaningful conversation.
Smith and Yamada are classmates.

スミス : 山田さん、先生はもう

(Sumisu : Yamada-san, sensei wa mo)
 Smith : Yamada-san, has the teacher already gone home?

- a. お帰りにおなりになりましたか。
(okaeri ni onari ni narimashita ka?)
- b. 帰りましたか。
(kaerimashita ka?)
- c. 帰りになりましたか。
(kaeri ni narimashita ka?)
- d. お帰りになりましたか。
(okaeri ni narimashita ka?)

山田 : さあ、ちょっと分かりませんが。
 (Yamada : Sa ..., cyotto wakarimasen ga ...)
 Yamada : Hmm, I'm not sure.

Smith is asking his friend Yamada-san where the teacher is. The teacher is the Referent and Smith is showing his respect towards him or her while talking to Yamada-san. It is, in other words, a “referent honorifics” question. We do introduce these types of referent honorific expressions in the Level 2 curriculum but we only *practice* addressee honorifics. When we composed this question (and some of the others as well), we did not notice this distinction. When we analyzed the results of the norming test, we realized we had tested the students on a grammar point that we had never had them practice. Even though the textbook introduced it, students had not had the opportunity to learn it.

In the case of Level 2, we made 8 items for the sub-category “polite expressions for interactions”. These were the results of the analysis of these 8 items in the norming test:

Item Facility (IF) Rate and whether it indicates a usable question for the database	Number of items
0.7–0.8 (satisfactory)	3
0.5–0.7 (marginal)	3
0.1–0.4 (unacceptable)	2

The large number of marginal and unacceptable items indicates that the questions have not been well matched to the proper proficiency level. Clearly, it is important to stay alert to the difference between material that is introduced and material practiced sufficiently enough that we can expect the student to have learned it.

2) Achievement-oriented questions are not appropriate for placement tests.

A second error in designing questions was not keeping in mind the difference in goals between an achievement-oriented test and a placement test. Achievement-oriented questions are meant to be a reach for some students but might be successfully answered by stronger students. A placement test need not try to identify this difference between the average and stronger students inside the level. It should instead offer questions that are squarely within a certain level. What we came to understand when we analyzed the norming tests is that we must be careful at all times to compose placement-style questions and not accidentally compose achievement-style questions.

Example 2 and Example 3, below, are questions on a Level 3 test. The first is too complicated for a placement test question, although it might be good for an achievement test. The second seems to be very good for the placement test.

In Level 3, students learn a variety of speech styles and also different levels of writing style. In the chapter that teaches different speech styles, we introduce humble expressions and very polite (double polite) expressions for interaction. So at this level students begin to understand well the concept of the polite expressions for interactions. Theoretically, Level 3 students can use complex polite (double polite) expressions, but of course it is not so easy. When we teach this chapter, we do not focus on double polite usage because we do not have enough time to teach these complicated expressions. Also, we think it does not need to be a high priority for this level of students. We think understanding the concept of polite expressions for interactions is important but perfect usage is not yet necessary. Because of time limits, students cannot get enough usage drills so we try to mention the concept of these types of complicated polite expressions whenever they occur in textbooks or video clips.

Example 2 is a good example of a double polite expression that is probably too challenging for students who are taking the placement test (but might be good for an achievement test that seeks to identify the best students in a specific pool of students). The ID score is very low, 0.07, meaning that it is not matched well to the level assigned to the question. Again, if we refer back to Ebel's table, this ID score falls in the category he describes as "Poor items, to be rejected or improved by revision."

Example 2

Choose the best answer to make a meaningful conversation.

リー : 明日、小山先生は日本にお帰りになると上田先生に*
(Lee : Ashita, Koyama-sensei wa Nihon ni okaeri ni naru to Ueda-sensi ni)
(Student) Lee : I heard from Teacher Ueda that (you) Teacher Koyama will go back to Japan tomorrow ...

- a. うかがいましたが…。 (humble for "to hear")

(ukagaimashita ga...)

- b. 聞きましたが…。 (simple polite “to hear”)
(kikimashita ga...)
- c. 行きましたが…。 (pronunciation is similar to “to hear”)
(ikimashita ga...)
- d. 来ましたが…。 (pronunciation is similar to “to hear”)
(kimashita ga...)

小山 : はい、そうです。
(Koyama : Hai, so desu.)
Koyama : Yes, that’s right.

*This sentence structure lacks some clarity and might have affected student understanding.

We think students do not need to know these high level polite expressions at Level 3. This is not an achievement test; it is a placement test.

Next we would like to describe an item that we think *is* appropriate for the placement test. Here is the item:

Example 3

Choose the best answer to make a meaningful conversation.

学生 : 先生はもう「XXX」という映画をごらんになりましたか。
(Gakusei : Sensei wa mo “XXX” to iu eiga o goran ni narimashitaka?)
Student : Teacher, have you already seen the movie “XXX”?
先生 : はい、 ...
(Sensei : hai ...)
Teacher : Yes ...

- a. 見ました。 (simple polite sentence)
(mimashita)
- b. はいけんしました。 (humble expression)
(haikenshimashita)
- c. ごらんになりました。 (honorific expression)
(goran ni narimashita)
- d. はいけんになりました。 (incorrect humble expression)
(haiken ni narimashita)

先生 : とても面白い映画ですね。
(Sensei : (totemo omoshiroi eiga desu ne)

Teacher : It is a very interesting movie, isn't it?

This item was a Level 3 question on the norming test. The sentence spoken by the student in this example has Level 2 grammar. But the humble expression that is one of the choice options in the teacher's part is taught in Level 3. That is why we put this question in this level. However, here is the data on this item:

Across Item Discrimination Rate (ID)	Item Facility (IF) Rate		
	at Level 2	at Level 3	At Level 4
0.08	0.83	0.91	1.0

Please notice the correct answer score (IF). For Level 2 it is 0.83, and then increases for Level 3 to 0.91, and increases again at Level 4 to a perfect 1.0. This question is performing well in the norming test environment. Students are getting used to the concept of the "polite expression for interaction". It appears that they can understand this aspect of Japanese culture better over time. Besides being able to confirm the appropriateness of the question, we have been vividly reminded of the importance of continuing to repeat practicing this concept as well as mentioning it as often as possible. We should already have known this, but sometimes we just forgot and sometimes time pressure kept us from doing this. Making the placement test and analyzing the norming test data has reminded us of the need to continue offering practice opportunity to students.

But, although we might gain some satisfaction with regard to the IF score, which indicates that over the course of the curriculum we are successfully teaching a complicated facet of the Japanese language, as you can see, the ID score is very low, well below our 0.22 minimum as an acceptable score. This means that this question is too easy as a Level 3 question. So, we are thinking we will try putting this question in Level 2 for the next norming test and check the result. If Level 1 students cannot answer the item and Level 2 students can, then this item is good for Level 2.

We are checking all the items in the norming tests. It is a valuable opportunity for rethinking the purpose of the placement test. It is sometimes a challenge because we are used to making achievement tests not placement tests. And sometimes it is difficult simply to compose four choices for an item, which is now necessary because of the multiple-choice format. Still, it is a rich learning experience and has helped us remember some important basic teaching principles that we had lost focus on from being so busy simply teaching day to day.

3) Considerations for the future curriculum

In addition to analyzing the results of the norming test in order to improve the placement tests, we also realized that the results might help us to improve the Japanese curriculum in the future.

After examining the results of the norming tests, we realized that in some cases, even though a certain structure may be considered appropriate to teach at a certain level, the students do not always master it completely at that level. For example, Japanese subordinate clauses with [時_に] (*tokini*, equivalent to English clauses introduced by *when* or *if*) are introduced in the first semester in our program, but in the test, the Level 1 students did not do well on them. Even Level 2 and 3 students scored low when the question contained this structure. The IF Score for the Subcategory 'tokini' was 0.59, the lowest IF score among all the Level 1 subcategories, and among the lowest on the entire test. We conclude that it might be beneficial to introduce these structures later, or teach them repeatedly in the following levels so that the students will get more practice using them.

Another finding concerns lists of vocabulary the students are assigned to memorize. In each lesson of the textbook, there is a vocabulary list to help the students understand the reading material, and the students are required to memorize them. But when we wrote questions for the norming test, we found that we needed different vocabulary items in order to talk about the topic of each lesson in depth, but these items were not on those lists. Perhaps when the instructors create the curriculum, they could consider what kind of discussion each lesson should include, and make their own vocabulary lists for the students to use, instead of relying on the vocabulary list in the textbook. In this way, the students could learn the vocabulary most useful for discussing the topic, and they will not have to memorize the vocabulary items that are only useful for understanding the reading material.

III. Operational Problems

In addition to the findings in the results of the norming test, we encountered some operational problems during the testing.

In the first two norming tests, we made taking the test voluntary, and gave an extra 1% to each student who took it in addition to the 100% for the course grade. However, some students did not take it seriously and just tried to finish all the questions without caring about their answers, and some good students did not even bother to take it, since they already knew that they were going to receive an "A" for the course. In order to avoid this happening again, for the third and fourth norming tests, we made the norming test mandatory and gave 2% out of 100% of the final grade if they completed all the questions and scored 60% or better. As a result, the participation was much better, and the results looked more reasonable.

We also ran into another operational problem. The students were able to access the BLC norming test website at any time and from anywhere during the three-day test period. As a result, some students copied the questions and even brought them to the instructors' office hours and asked them to explain them, even before the test period was over. It had never occurred to us that the current students would make copies of the questions. Even though each student receives a different set of

questions from a big database, if they share questions, the results of the norming test will not be reliable. Our solution is that we made a handout of procedures and had the students sign it, promising not to copy and share the test items. This is not a perfect solution, but we hope it will at least make it less likely that students will copy or give copies to their friends. We were very upset about this, but it may not be a serious problem, because when it is in actual use as placement test, the students who are taking the placement test are unlikely share the questions with their friends or look them up in the dictionary, since they usually do not want to do better than their actual level and thus be placed in more difficult level. Whenever any test is given online without any supervision, it is inevitable this kind of operational problem will occur; we should not give up but just try our best to avoid it.

IV. Conclusion

We learned a lot from giving these tests and analyzing the results. We believe that using the norming data to improve the questions on the placement test will significantly improve its accuracy. In addition, we also realized some things we had not been aware of, and ways in which we might be able to improve our teaching.

After editing these tests one more time during the summer, we will use them for placement testing of incoming students in August this year. Our placement test does not have listening, speaking or writing parts because we strongly believe that it is necessary for us to talk to the students individually to find out their background and how much they can communicate orally and in writing in Japanese, and to judge which course is most appropriate for them. Therefore, we will use this placement test together with a writing sample and a short oral interview.

This is not the end of the development of the placement test. No placement test can be used forever. Every time we change the textbook, curriculum, or lesson plans, we will need to give a norming test and examine the results, then edit the questions. This will ensure that the questions are always appropriate to place the incoming students in our program

We are the first language program to complete this process. If you are interested in our work and wish to make a placement test of your own using this program, we will be happy to offer suggestions. We also want to make it clear that the template created by the BLC which was used for this placement test can be used for any language, and not only for placement tests, but also for formative tests and achievement tests.

Acknowledgements

We would like to thank the BLC for giving us the BLC fellowship twice, and thus giving us a chance to complete an online placement test. Thanks to Professor Kern and Professor Kaiser for their useful advice and guidance. Thanks to the

programmers, Ms. Tolpygo and Ms. Suzuki, for their hard work in making the program available for us. Thanks to all the BLC fellowship participants for their patience in listening to our talks and giving us useful suggestions. We also thank Professor Hasegawa for her support, and all the Japanese language instructors for their cooperation and help in carrying out this project.

References

- Brown, H. Douglas and Priyanvada Abeywickrama. (2010). *Language Assessment: Principles and Classroom Practices*. NY: Pearson/Longman.
- Brown, James Dean. (2011). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment. New Edition*. NY: McGraw-Hill.
- Ebel, Robert L. and David A Frisbie. (1991). *Essentials of Educational Measurement*. Fifth Edition. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Hasegawa, Yoko et al. (2005-2006). *Elementary Japanese Volume One and Two*. Rutland, VT: Tuttle Publishing.
- Kamata, Osamu, Fumiko Beuckman et al. (2012). 新 中級から上級への日本語 (*Shin chukyu kara jokyū e no Nihongo, Authentic Japanese: Progressing from Intermediate to Advanced. New Edition*). Tokyo: The Japan Times.
- Miura, Akira and Naomi Hanaoka McGloin. (2008). *An Integrated Approach to Intermediate Japanese (Revised Edition)*. Tokyo: The Japan Times.
- Oka, Mayumi, Michio Tsutsui and et al. (2009). 上級へのとびら (*Jokyū e no tobira, Tobira: Gateway to Advanced Japanese*). Tokyo: Kuroshio Publisher.