Assessing the Language Component of

the Manoa General Education Requirements

Report on the College of Languages, Linguistics, and Literature (LLL)
Assessment Committee's 2002-04 assessment activities

Kimi Kondo-Brown
East-Asian Languages and Literatures
University of Hawaii at Manoa
May 7, 2004

1

Defining terms

Assessment

An ongoing process aimed at evaluating and improving student learning in Hawaiian/ second languages, which involves:

- Making the expected learning outcomes of the Hawaiian/second language requirement explicit,

- Systematically collecting, analyzing, and interpreting data to determine the degree to which the actual student learning matches our expectations, and

- Using such evaluation information to improve student learning.

Tests

The use of tests is only one of the many sets of tools that can be used in assessment activities.

- We have decided to develop criterion-referenced, faculty-made achievement tests.

- The primary purpose of CRT is to measure the amount of learning that a student has accomplished on given objectives, very different from NRT designed to measure more global abilities of individual students, and interpreted with reference to all other students' abilities.

2

Purpose and significance of the present assessment project

The language requirement in the UHM general education descriptions

" . . . proficiency in Hawaiian or a second language is an integral part of the university's mission to prepare students to function effectively in a global society to preserve and promulgate Hawaiian, Asian, and Pacific language, history, culture . . . . before graduation all students must show competency at the 202 level."

What is meant by "202-level competency" in different language programs?

The present assessment project was intended to redefine future learning objectives for the core Hawaiian/second language programs and develop assessment instruments and procedures to measure the effectiveness of these programs in achieving these objectives.

3

Steps for product-oriented approaches

The present assessment project falls into the category of "product-oriented approaches" (Hammond, 1973).

1. Identifying precisely what is to be evaluated
2. Defining the descriptive variables
3. Stating objectives in behavioral terms
4. Assessing the behavior described in the objectives
5. Analyzing the results and determining the effectiveness of the program

4

An overview of five steps for the LLL assessment project

1. **Identify** (a) key concrete and measurable learning outcomes and (b) preferred assessment tools and procedures

2. **Develop** assessment instruments and procedures to assess one of the identified objectives in different languages for the 2003-04 pilot testing

3. **Implement** the instruments and procedures to measure the target objective in different languages

4. Compile and **analyze** the data from the participating language programs

5. **Present** the results of the data analyses to the faculty members and plan for future actions

5

Step 1: Identify key concrete and measurable learning outcomes and preferred assessment tools and procedures

The LLL assessment planning survey (participants and instruments)

Participants

All teachers in the Hawaii/Second language programs were invited to participate. 92 teachers from 22 language programs responded to the survey (47.9% return rate).

Instruments

The questionnaire was developed primarily based on the information obtained through a preliminary electronic survey among program coordinators. The draft was revised based on the feedback from the LLL Assessment Committee.

The final version had four sections.

- Section 1: Participant background information
- Section 2: The degree to which they agree with 42 statements as learning outcomes for their students who complete a fourth-semester course
- Section 3: Assessment instruments for measuring the identified outcomes
- Section 4: Assessment procedures (e.g., when and how often should we assess, who should be assessed, and who should do the assessment, etc.)

6

Step 1: Identify key concrete and measurable learning outcomes and preferred assessment tools and procedures

The LLL assessment planning survey (Survey results 1)

Recommended learning outcomes

1. Understand conversations about everyday experiences (e.g., school, work, interests, preferences)
2. Understand factual content of paragraph-length descriptions/narratives on familiar topics (e.g., recorded telephone instructions, announcements in public areas)
3. Perform a variety of "real-life" tasks in common social and transactional situations (e.g., shopping, making hotel reservations)
4. Sustain conversations/interviews about self, family, experiences, interests, and preferences
5. Understand fully paragraph-length texts dealing with personal and social needs such as personal letters, messages, and memos
6. Get main ideas from authentic everyday practical materials written entirely in the target language (e.g., menus, ads for products)
7. Meet practical writing needs and social demands by writing paragraph-length personal letters, messages, applications, and journals
8. Demonstrate understanding of holidays and traditions celebrated in the target culture

7

Step 1: Identify key concrete and measurable learning outcomes and preferred assessment tools and procedures

The LLL assessment planning survey (Survey results 2)

Preferred assessment tools and procedures

Preferred assessment tools

There was strong interest in developing faculty-made achievement tests embedded in final exams across various language programs. For example, more than 75% of the participants chose "faculty-made paper-and-pencil achievement test embedded in the final exam" for measuring reading skills.

Preferred assessment procedures

More than half of the participants think that assessment should be conducted every semester, at the end of second-year courses among all target students in all languages. Opinions as to who should do the assessment seemed divided.

8

Step 2: Develop assessment instruments and procedures to assess one of the identified objectives for the 2003-04 pilot testing

Steps taken and decisions made in Spring 2003

1. The results of the LLL Assessment Survey were presented to all relevant administrators in the College of LLL to discuss strategies for continuing this college-wide assessment project.
2. The LLL Assessment Committee members and administrators decided to conduct a pilot assessment of selected programs during the 2003-04 academic year.
3. The LLL Assessment Committee Chair announced the next phase of this project to the entire faculty in the involved departments and sent them the major findings of the LLL assessment survey.
4. The LLL Assessment Committee decided to focus on one of the eight strongly recommended learning outcomes: "Understand fully paragraph-length texts dealing with personal and social needs such as personal letters, messages, and memos."
5. The representatives of each language program agreed to develop a single set of multiple-choice reading comprehension test items in order to assess this learning outcome.

9

Step 2: Develop assessment instruments and procedures to assess one of the identified objectives for the 2003-04 pilot testing

Common characteristics of the test across various language programs

Test format: one page test with a text of a few short paragraphs that the students can finish in a short time (i.e., 10 minutes or so).

Prompt: A personal letter where the sender describes (a) what he or she did during the thanksgiving and (b) his or her plans for the Christmas holidays.

Questions: Multiple-choice test with several questions (four options for each question).

Administration: The test will be given to 202 and/or 201 students (depending on the courses offered in each language program) at the end of the semester.

Use of the test scores: Test scores will not be included in the student's final grade, but their performance may be taken into consideration as extra credit.

After the draft test developed by each language program was created, each program presented the test at a meeting of all representatives in order to ensure that the items were as similar as possible across all language programs

10

Step 3: Implement the instruments and procedures to measure the target objective in different languages

- It turned out that five language programs participated in the 2003-04 pilot assessment project.

- In the case of the Japanese section, the Japanese section head sent a memo to all 201 and 202 instructors to seek volunteers to let their students participate in the pilot assessment. It turned out that 37.3% of the target students participated in the pilot testing.

- The test was given in class during the final week of instruction (the Japanese section). The tests were distributed to teachers in an envelope before the scheduled administration.

- With each envelope, several instructions for the participating instructors were given to maximize the validity of the obtained data such as opening the envelope on the day you administer the test, allowing exactly 10 minutes to complete the test, encourage your students to do their best, etc.

11

Step 4: Compile and analyze the data

Each language program was asked to send me the tabulated EXCEL data as an email attachment. In this way, data from 521 students in five language programs were received.

Example test item analysis to examine the effectiveness of the test

In this language group, there were six items and each item had four choices. The first four items were ones that their 202 students were expected to be able to answer. The last two items included some third-level materials for experimental purposes. Therefore, the cut point of this test was set at <u>four</u> points.

The table below shows the results of one type of test item analysis using the "*B*-index" that indicates the degree to which a group of students who were at the mastery level (i.e., those who scored four or above) outperformed the ones who were not at that level on each item.

Table 1. Test item analysis 2 (the *B*-index)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
IFpass	0.96	0.89	0.92	0.81	0.72	0.63
IFfail	0.49	0.58	0.27	0.27	0.26	0.16
B-index	0.47	0.31	0.65	0.54	0.46	0.47

12

6

The K-R21 reliability index (i.e., a conservative estimate of the *phi* dependability index for criterion-referenced test) for the present reading test was .576, which is reasonably high for such a short instrument, but in order to improve the reliability of the test, adding more items is recommended for the next round of piloting.

For example, the Spearman-Brown Prophecy formula allows us to calculate what the reliability would be if six similar items were added for a total of 12 items, which, in this case, turned out to be .73.

13

Example performance analysis to estimate the degree of learning that occurred

The mean and standard deviation for the 201 group were 3.1 and 1.6, respectively, and the mean and standard deviation for the 202 group were 4.2 and 1.5, respectively.   A *t*-test performed on these two sets of scores suggested that the mean for the 202 group was significantly higher than that for the 201 group (p < .000). In other words, more learning seemed to be demonstrated for the 202 students (who had [almost] finished their language requirements) than for the 201 students (who had not yet finished).

Figure 1. Comparisons of test score distributions for the 201 and 202 groups

14

7
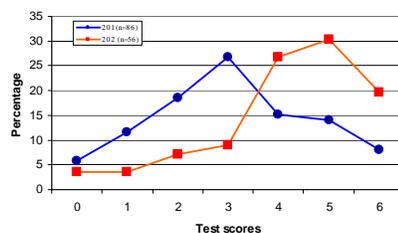
Step 5: Plan for future actions

For this semester (Spring, 2004), we have decided to continue focusing on the same objective and obtain the data using the revised tests, i.e., tests improved based on the previous year's test item analysis (e.g., adding a test item for a better reliability, revised distractors that do not attract any respondents, etc.).

In addition, in order to avoid a practice effect, in the case of Japanese, the revised version (Form B) was developed by making minor changes to the original one (Form A) (e.g., changing certain lexical items such as names, places, and actions, and changing the order of options).

15

What factors have contributed to the successful initiation and maintenance of the project to date, and how will we go beyond the present stage?

1. This project was initially funded by the 2002-03 UH Assessment Fund. It would help if the university would keep providing this kind of money for future assessment activities. I hope the university or college will invest more in existing assessment activities like this one.
2. The initiation and continuation of the present project has been possible because our deans and chairs have been so willing to get involved in the project. Without it, it would be difficult to keeping teachers involved.
3. The heads of the participating language programs were both supportive of this project and helpful. I hope the heads of language programs will continue to understand the project and be supportive of the activities involved.
4. An acceptable number of teachers actually volunteered to participate in the first round of pilot testing. Without their willingness to cooperate and participate in the testing, no assessment activity would be possible.
5. We need to discuss how best to incorporate a college-wide assessment activity like this into the existing second language curricula. It seems that testing time is a critical factor in teachers' deciding whether or not they will participate.
6. Someone must compile and analyze the obtained data and report the results. If this project is meant to go beyond the current pilot stage and expand to an acceptable level as a sound assessment activity (longer tests measuring more outcomes with more participants in more languages), we need a plan for how best to manage and analyze the *large* amount of data.

16