

Piloting Computer-Based Proficiency Exams in First-Semester French, German and Spanish

Laura Anderson - Foreign Languages

This project was structured around the following action steps:

- development of a set of computer-based exams that would be used to assess language proficiency at the first-semester elementary level in French, German and Spanish;
- piloting of these exams in the fall of 2001 in all sections of French 1040, German 1240 and Spanish 1840;
- evaluation of the outcome of the computer-based testing through comparisons with final semester grades and oral interview scores;
- revision of the testing instruments based on the analysis of the data collected from the pilot program.

Development of the First Computer-based Exam in French:

August/September 2001

In order to develop a computer-based exam that could be used in the foreign language lab, it was first necessary to locate a commercially produced software program that would best meet the assessment criteria established by the three faculty members (Laura Anderson, Mark Evenson and Patrick Hagen) involved in this project. Assisted by Ken Wiegman, the LAE Computer Technician, we conducted research on the availability of such a program. My own research on the Internet led me to conclude that there were very few software programs commercially available that could be used for constructing proficiency exams in a foreign language.

In fact, I found only one program, marketed by Brigham Young University that was specifically designed for an interactive assessment of oral and aural language skills. The version of this program that was listed on the BYU web site in August 2001 was a relatively unsophisticated program that allowed the examiner to create oral tests using text, graphics, audio and/or video as prompts, but did not include the capability of pulling questions randomly from a test bank. When I revisited BYU's web site later in the semester, I discovered that a new "enhanced" version had become available. The enhanced version had some of the features, including the ability to create a test bank, that we had been looking for. Unfortunately, by that time I had already ordered a copy of the first version, and the project was too far underway in French to start over with the newer version of the software program.

September/October 2001

With no other viable options available, I requested that the department purchase a copy of the "Oral Testing Software" marketed by BYU. When the materials arrived at the end of September, Ken Wiegman installed the software program on one of the computers in the foreign language lab, and we later spent an hour together working through the program options. The other project participants were notified that the program was available in the lab, and copies of the user's manual were distributed to each of them.

November 2001

Since there was no test bank option available with this software program, I created two different versions of each assessment exam that I gave in French. In this way, I would be able to ensure that students who sat next to each other in the lab would not be taking the exact same test. Both versions followed the same format, however. The first exam consisted of the following sections:

- The family

Areas assessed: vocabulary identification / sentence structure / pronunciation.

Activity: Students were given a handout of a family tree prior to beginning the exam. According to the instructions that appeared on the screen, students were asked to record as many statements in French as they could about the family from the point of view of one of the individuals included on the family tree.

- Numbers 0-60

Areas assessed: vocabulary identification / pronunciation.

Activity: Students were told to read/record the appropriate French word for each number they saw on the screen (12 numbers selected at random).

- Responding to oral questions

Areas assessed: aural comprehension / ability to respond logically / vocabulary identification / sentence structure / pronunciation.

Activity: Students were asked to record a simple response in French to the questions that they heard (no visual cues provided).

- Identifying activities

Areas assessed: vocabulary identification / sentence structure / pronunciation.

Activity: Students were given a photocopy of several drawings that illustrated some of the activities that were part of their vocabulary assignments. They were instructed to record a simple statement in French that described each picture.

For purposes of test security, the two versions of this exam were recorded onto zip disks that were then labeled with the individual student's name and the version of the exam (A or B). Using the class roster, I recorded the number of the computer each student would be using next to his/her name; this would later prove to be an important step in tracking down problems with recording due to equipment failure.

Piloting the First Exam in French:

November 2001

After having one of my students in intermediate French take the exam on a trial basis and making some revisions to the format, I administered the first exam on November 6 to the 15 students in my first-semester elementary French (FREN 1040) class. Though the students had already been given an introduction to using the lab and had completed several lab assignments by this point in the semester, this was their first exposure to this particular testing program. I conducted an informal discussion in class the next day with the students concerning their experience with this testing procedure. Reactions were mixed, with approximately half the students expressing a preference for computer-based testing and half indicating a preference for a face-to-face interview with the instructor. The main reason given for both preferences was a perception of feeling more relaxed with one particular method of assessment. Objectivity and accuracy in evaluating exam results were not mentioned by any of the students as a reason for preferring one method over the other. Students also offered comments on their experience using this particular software program. Most felt that the software was fairly "user-friendly", but they all seemed to agree that some sort of practice section would have been beneficial.

Evaluating the First Exam in French:

November 2001

I developed a scoring sheet for the first exam and spent more than two hours listening to the students' recorded responses on the individual zip disks. In a face-to-face interview, it is possible to evaluate the student's ability to negotiate meaning and communicate effectively in the target language through the give and take of an actual conversation. Since it is impossible to conduct a conversation with a recording, I found that I was really limited to an evaluation of vocabulary production, pronunciation, sentence structure and appropriateness of response: in other words, I was evaluating the student's ability to identify vocabulary and respond to isolated questions rather than his/her ability to participate in a conversation, an experience similar to grading a "fill-in-the-blank" test instead of an essay exam.

Several of the final scores on this first exam were significantly lower than I had anticipated, based on my knowledge of the students' classroom work. Technical difficulties in recording responses via the computers in the lab were largely responsible for these low scores. Some examples of the difficulties encountered are listed below.

- The wires connecting the headphones/microphones to the computers were loose on several of the computers, resulting in a distortion of sound and background "clutter" noise that lowered, sometimes significantly, the quality of the recording.
- Even though most students did not sit immediately next to another student in the lab, I discovered when listening to the recordings that the microphones still picked up the voices of other students as they recorded their responses, as well as my voice as I circulated among the students.
- Several students commented that they were having trouble hearing the recording, even though the volume settings for both the headphones and the computer were set at their highest level. It was not possible to fully resolve this issue during testing, but all students were eventually able to hear the recording well enough to complete the exam.
- There is a short sound check built into the testing program that students are to complete before beginning the exam. Using this sound check, we discovered problems with two of the computers. Fortunately, I had a small enough group that I was able to move the two students to different computers. Unfortunately, there were two other students who reported no problems with the sound check and who turned in what they believed to be a completed exam. However, when I began the evaluation process, I quickly discovered that they had recorded absolutely nothing, resulting in a test score of 0. A malfunction in the computers they were using was later identified.

I had originally intended on averaging the scores of the two computer-based exams given during the course of the semester in order to determine the final grade for oral proficiency. However, because there were so many technical difficulties encountered during the piloting of the first exam, I informed the students that I would use only the higher of the two scores. As it turned out, the score on the second exam was the higher score for all students.

Development of the Second Exam in French:

December 2001

After spending an hour in the foreign language lab with Ken Wiegman checking on equipment, I spent another two and a half hours developing and recording two versions of the second exam and copying them to the students' zip disks. Again, the format for both versions was the same. The second exam consisted of the following sections:

- Telling time

Areas assessed: vocabulary identification / sentence structure / pronunciation

Activity: Students were instructed to read/record the times they saw on the screen, using the conventional (conversational) style of telling time.

- Dates

Areas assessed: vocabulary identification / sentence structure / pronunciation

Activity: Students were instructed to read/record the dates they saw on the screen.

- Expressions using the verb faire

Areas assessed: vocabulary / sentence structure/ pronunciation

Activity: Students were given a photocopy of several drawings that illustrated some of the activities associated with the verb faire that were part of their vocabulary assignments. They were instructed to record a simple statement in French that described each picture.

- Reaction statements

Areas assessed: aural comprehension / ability to give an appropriate response / vocabulary / sentence structure / pronunciation

Activity: Using vocabulary expressions that indicated surprise, disappointment, indifference, etc., students were asked to record a simple "reaction" in French to the statements that they heard (no visual cues provided).

- Responding to oral questions

Areas assessed: aural comprehension / ability to give an appropriate response / vocabulary / sentence structure / pronunciation

Activity: Students were asked to record a simple response in French to the questions that they heard (no visual cues provided).

Based on student comments given after the first exam, I extended the amount of time that the students had to read the instructions on the screen, listen to the oral cues and record their responses (a clock on each screen counts down the time remaining for each question segment of the exam so the students are always aware of how much time they have in which to reflect and then record their answers).

Piloting the Second Exam in French:

December 2001

The second computer-based exam was given on December 12 in lieu of the face-to-face interview that I normally administer at this point in the semester. Because the first round of testing had identified certain technical difficulties associated with using the computers in the foreign language lab for the purposes of assessment, piloting the second exam did not prove as difficult, and the resulting scores were significantly higher for several students. I had the students use a different sound check system that more clearly identified problems with recording and better enabled them to adjust the sound levels before beginning the exam. I also reminded the students that they needed to speak up when recording and watch the clock so that they would not run out of time on any given question. Because the students had already had the experience of using the program once, they all indicated that they felt more comfortable with the recording process, and there were no problems with incomplete recordings during this second round of testing.

Evaluating the Second Exam in French:

December 2001

I developed a scoring sheet for the second exam and again spent more than two hours listening to the students' recorded responses on the individual zip disks. As noted above, the scores for this exam were significantly higher for several students, largely due to the elimination of major technical difficulties with recording via the computers in the lab. The students were also more at ease with using the testing program and seemed to navigate through the exam with more confidence.

I did, however, discover a significant drawback to using zip disks to record the students' responses. Since the foreign language program has a limited number of zip disks available for classroom use, I had to spend almost an hour moving the student recordings from the disks to my computer's hard drive and then erasing the disks so that they could be reused for subsequent testing. Even if students are required in the future to supply their own zip disks, it will still be necessary to transfer the recorded material to the instructor's hard drive if he or she wishes to retain a copy of the exam.

Comparison of Computer-based Exams and Final Grades:

Below is a listing for each student enrolled in FREN 1040 during the fall semester 2001.

Student	First Computer-based Exam	Second Computer-based Exam	Final Grade
1.	80%	84%	B
2.	33%	0 (student stopped attending class)	F
3.	0 (nothing recorded)	49%	F

4.	71%	74%	B
5.	83%	84%	B
6.	57%	78%	C
7.	69%	89%	B
8.	67%	95%	A
9.	0 (nothing recorded)	90%	A
10.	86%	88%	A
11.	71%	83%	B
12.	57%	80%	C
13.	71%	88%	B
14.	35%	57%	D

As previously noted, only the score of the second exam was used in the calculation of the final grade for each student. For most students, the score on the second exam roughly correlates to their final grade for the course. It should be noted, however, that the computer-based exams were not the only means by which proficiency in the target language was evaluated. Quizzes, lab assignments and chapter exams were also used to assess oral and/or aural communication skills. Moreover, 10% of the student's final grade was based on participation in class activities in the target language.

Ten students from FREN 1040 also completed the next course in the sequence (FREN 1140) during the 2002 spring semester. There were no computer-based exams given in FREN 1140, but all students did meet with me individually at the end of the semester for a face-to-face conversation (10% of their final grade). Students were told in advance that the interview would be conducted exclusively in French, that they would be graded on the conversation as a whole (rather than on a percentage of errors made) and that I would be using the following list of criteria in determining their grade:

- appropriate use of vocabulary,
- correct grammar and syntax,
- accurate pronunciation,
- "flow" of speech,
- originality, and
- ability to understand and correctly respond to questions.

The following shows a comparison of the second computer-based exam score from fall 2001 and the oral interview letter grade from spring 2002 for the ten students who took both FREN 1040 and FREN 1140:

Student	Second Exam (computer-based)	Oral Interview
4.	74%	A-
5.	84%	B
6.	78%	B-

7.	89%	B-
8.	95%	A
9.	90%	A
10.	88%	B-
12.	80%	C-
13.	88%	B-
14.	57%	D-

It is not possible to draw any concrete conclusions from such a comparison since the exams were given at the end of two different semesters in a sequenced set of courses. However, it is interesting to note that the two scores for many of the students fall roughly within the same range of grades. Since these were both small classes, I had the opportunity to become very familiar with the personalities and learning styles of the students I was testing. I was not surprised, therefore, that some discrepancies in scoring did occur.

Student D, for example, did not appear to be very comfortable with recording answers on the computer (very soft-spoken and hesitant), but was able to participate well in a conversation where non-verbal communication skills were important. Student J, however, noted in a class discussion that she felt very nervous during an interview situation and much preferred recording answers via computer; her score for the oral interview was indeed lower (though still in the same letter range) than her score on the second computer-based exam.

Revision of the Piloted Computer-Based Exams in French:

Summer/Fall 2002

As noted earlier in this report, Brigham Young University now offers an "enhanced" version of the oral testing software that I used for this project. Having reviewed the information available on-line and having spoken with the language lab director at BYU about this enhanced version, I believe it would be in the best interests of our on-going assessment project to purchase new software for use this fall, even though it will make revising the computer-based exams that I have already created difficult if not impossible; revising the exams may very well mean creating completely new ones. While I was satisfied with the basic format of the exams I gave, I would like to add more test items to each section. Hopefully, the enhanced software program will make this a less cumbersome process.

Closing Comments:

In conducting this pilot program in French, I discovered that the process of creating, administering and evaluating one computer-based exam in one class of 15 students took a total of approximately six hours to complete; hardly a "time-saving" substitute for the one and a half or two hours it typically takes to conduct oral interviews with a class that size.

The students also encountered a wide-range of technical difficulties when recording exam responses through the computers, problems that simply do not occur when conducting a face-to-face interview in one's office. Any breakdown in communication during an interview is usually resolved by asking the student to repeat what he/she has said, to clarify his/her point, to speak more loudly, etc. It was frustrating to me not to have the same flexibility when evaluating the recorded responses of the computer-based exams: when the recording was unclear, the student lost points.

Based on comments from my students and on my own experiences thus far with computer-based assessment, I would conclude that such testing has the advantage of making some students feel less anxious about the concept of assessing oral and aural skills in the target language than a face-to-face interview does.

Moreover, while it is impossible to construct a computer-based exam that in any way truly assesses the student's ability to participate in a conversation in the target language, it is possible to use computer-based testing to evaluate certain aspects of language production, such as pronunciation and intonation, as well as aural

comprehension and, in a limited way, knowledge of vocabulary, grammar and syntax. Since I believe that the development of oral communication skills in French is an important objective at all levels of language instruction, I am not prepared at this point to eliminate oral interviews from my curriculum. I do not feel that computer-based testing is an appropriate means of assessing a student's ability to communicate effectively in a casual conversation in French. Even participation in an on-line discussion or conversation is not the equivalent of face-to-face communication.

However, since some students do find computer-based testing to be a "less stressful" experience and since it can be used to assess certain skills, I believe it is worth the time and effort to continue working with this type of assessment tool. I therefore intend on using computer-based testing in conjunction with individual oral interviews this fall in FREN 1040. Eventually, these methods of testing will likely be used in some form at all levels of instruction to assess oral language proficiency in French.

Assessment Activity Fund grant report 2002