

Assessment that Teaches: An Experiment in Just Evaluation

James D. Wilkins
Lee University
Cleveland, Tennessee

Reflective pedagogical practice is something I have always championed and striven to do. After reading several thought-provoking works on issues and goals of pedagogical practice, I began to reflect deeply and critically on the topic of traditional assessment and grading, and their role in the development of L2 proficiency. A student of mine in an intermediate-level course served to heighten my interest in this topic. She seemed confidently engaged in the learning of French, though when it came to assessment, she succeeded minimally, receiving average and frequently below-average grades. Why was this student, who I *knew* was developing functional proficiency in French, based on class performance, being defined by letter grades as a C or C- student? What did that mean? What did those letters say about what she was able to do or not do in French? More troubling, how could I make the case for a constructivist pedagogy whose end was world justice if the grading system used to indicate functional proficiency was non-informing, misinforming, or worse... unjust? Having taught for 30 years, and being tenured, I decided it was time to risk an experiment.

I began in the fall semester by educating my students on the topic of knowing *about* a language vs. functional proficiency. This involved giving examples of familiar traditional assessments, then asking students to identify the goals of such testing and what they reveal about a person's ability to function in the target language. Together, we were amazed at the conclusions, but excited about transforming our thinking about language learning and how it is assessed. Following several more discussions, including proposing a pilot for the following semester, all my students whom I would have in the sequel course, were willing to participate. The pilot

involved rethinking the nature of material covered, and the way that learning that material would be assessed. Rather than traditional granting of points or letter grades, task-based assignments and assessments would be evaluated by clear, attainable goals established for each. I was inviting my students, and forcing myself, to look at learning a world language as a goal-driven enterprise, rather than one defined by an ambiguous collection of points.

At the beginning of the pilot semester, the first three days of class were spent doing two things: establishing goals in speaking and writing, and developing an evaluation policy for the course with appropriate assessment rubrics. A good deal of time during these first days was spent helping students understand formative vs. summative assessment and language-learning tasks. This level of involvement alone was a major educational benefit to the students, as it enabled us together to examine more closely second-language acquisition.

Assessment during the semester was two-pronged: by individual assignments and by holistic assessment at the end of the semester. During the semester, each summative assignment or test was given a goal-based rationale, i.e. "the purpose of this assignment/task is to..." Then, completed tasks were evaluated by rubrics created to inform students to what degree they had met the goal(s) of each assignment. For example, an oral task's assessment goals might be 1) ability to speak in complete sentences, 2) ability to create personal meaning with the language, and 3) ability to give a simple, coherent description. The student was evaluated as 1) having exceeded the goal(s), 2) having fully met the goal(s), 3) having partially met the goal(s) and 4) having not met the goal(s). Rather than maintaining a grade sheet with columns of points and grades, copies of all evaluation rubrics were placed in a portfolio for each student. Students thus had an ongoing record of assignment with goal-based comments built into the rubrics, which were generally much more substantively productive than a simple "bien, pas mal" or "excellent."

The final grade for the course was based on a final summative evaluation that took two days to administer, consisting of a series of speaking and writing tasks. The speaking tasks were designed to be done in pairs or small groups (i.e. to assess in a manner that reflected classroom language practice), others involved face-to-face exchanges with the instructor. The writing tasks were a series of easy to more difficult tasks, with things like writing a post card to a friend, or writing a past narrative. The goal of the

tasks and order of presentation were designed with the goal of helping determine a pattern of linguistic strengths and weaknesses, all the while maintaining focus on content over form. Evaluations were holistic. Rather than tallying the number of mistakes, an assessment of what the students were able *to do* functionally French was given priority. Based on whether or not students had met the established proficiency goals for the course, they were given a letter grade equivalent to represent their achievements on their academic transcript (equivalents that had been established early in the semester). For example, a final designation of *intermediate-mid* in speaking was awarded an A, a B in writing.

Results were mixed. Some who had done poorly on more traditionally-designed assessments, for example, did better in meeting or exceeding clearly-defined goals. Others, who were accustomed to receiving high grades on traditional evaluations, were surprised to find they had only minimally met the functional proficiency goals. In student course evaluations, most students highlighted the lessened pressure of point-evaluated exams, citing the more positive and helpful feedback of rubrics. Some students complained of “not knowing their grade” during the semester. I anticipated this, as the traditional grading system is so engrained in grades 1-12, and without it, some students would sense a great deal of frustration, as their evaluations were not being articulated by letters or points as they were in their other courses. In semesters since the pilot, however, there have been a number of student comments that I should return to the “no-grade” format of course evaluation, due to its lessening of the “affective filter,” especially for those students with high test anxiety. Final grades were better on the whole at the end of the semester. There were no disputed grades. I attribute this to the fact that students had shifted focus toward what they could and could not do in French, thus having a more realistic picture of their own abilities throughout the semester.

Would I do this again? I would, with more careful preparation of assessment rubrics, and detailed planning of assessments. More attention would be given to reading and listening assessment, to better equip students to “hear” Others. Teaching and testing in this manner did free me from the constraints of a textbook-prescribed syllabus, and seemed to motivate students to take greater advantage of in-class practice and conversations and contact with native speakers in the community.

Finally, teaching at a university where the integration of faith and learning is a foundational enterprise, this pilot forced both my students and

me to see this restructuring as an exercise in justice. *Just* in the sense that expectations were clear from the start and accepted by everyone. *Just* in the sense that final grades, if not perfectly, at least more closely represented what a student was actually able *to do* with the language. It has also fostered reflection on how Kingdom principles affect those governing good teaching and effective learning. Not everyone who does seat time, giving discrete professions of knowledge, will truly acquire the ability to communicate with Others, nor be prepared to receive what Others have to say. They know about language, but they do not develop the functional ability necessary for interaction with people, to communicate with them nor to receive communication from them. As in the Kingdom, the blessings belong to those who *function* as Christ's disciples, not to those who are only cognizant of His teachings (Matthew 25:34-46). These principles and the values they imply played a major role in our decisions during this course pilot. It made for a more just, less-anxious learning environment, a learning experience that was goal- rather than prize-driven, and students who have developed new strategies for learning and integrating those strategies into lives of action.