Faculty Assessment of the Effects of a Freshman Chemistry Course

Susan B. Millar, Steve Kosciuk, Debra Penberthy, and John C. Wright
Learning through Evaluation, Adaptation and Dissemination (LEAD) Center
University of Wisconsin-Madison

Abstract

A unique assessment of the effects of a freshman chemistry course was conducted at the University of Wisconsin-Madison during spring, 1995. Twenty-five faculty from eleven departments outside the Chemistry Department participated as assessors, interviewing some 200 students from an experimental and a comparison version of the course. Results show notable differences. Moreover, the experience of participating in the assessment has generated significant interest in new approaches to teaching among the twenty-five faculty participants.

Background: Development of Structured Active Learning Strategies

Chemistry 110 is the second course in a two-semester sequence designed for "fast-track" first-semester students, most of whom are science and engineering majors. One of the sections of Chemistry 110 has been taught for many years by John Wright, the Chemistry Department faculty member who initiated the Chemistry 110 assessment project reported here. After devoting serious effort to teaching the course for some 20 years, Wright realized that, while most students acquired enough command of the material to perform well on exams, they rarely made the connections between lecture material and lab applications. Drawing on input from Chem 110 students and a senior colleague, he decided in 1992 to try to create a learning environment that would allow students to understand the connections among concepts and between theory and applications. His overarching strategy for achieving this goal was to give the students more responsibility for their learning while at the same time providing the necessary tools and support for solving more difficult research problems. Each year since 1992, he introduced and used informal classroom assessment techniques until he had a set of "structured active learning" (SAL) strategies that he believed achieve his newly articulated goals.[1] These SAL strategies include:

- an absolute grading scale that replaced the "curve";
- student lab groups that completed three open-ended laboratory projects (during the first 6 weeks of the semester, students completed standard lab experiments individually);
- student lab groups that read and analyzed research papers;
- interactive techniques in the lecture, including think-pair-share, collaborative problem solving, concept tests, and list generation exercises;

- differing types of assessment measures including in-class individual exams and difficult cooperative take-home exams as well as oral exams with lab groups;
- spreadsheet programs 'for homework and laboratory problems;
- a student board of directors that advised on all aspects of the course.

Observing the effects of the revised course on students, Wright was very impressed. As he put it,

> I was blown away by the phenomenal creativity and competence some of these students brought to course projects. They had a deep command of the material, and could apply it in different contexts in very creative ways. In short, student performance was at a level far above what I' d seen before.

The Need for Assessment

Upon describing his students' learning outcomes to his faculty colleagues, Wright found them generally supportive but not interested in trying the SAL approach themselves. He realized that many of his colleagues were unconvinced that teaching methods had significant effects on student learning, as they believe that student effort and aptitude are the primary determinants of learning outcomes. To convince others to create similar learning environments in their own classes, he therefore would need to identify the new course environment as a primary determinant of student learning. Upon being asked what would convince them, several of his colleagues advised that they would require a comparative study that assessed learning outcomes for matched groups of Chem 110 students, one of which was taught in the traditional way by a top teacher in the Department, and one of which was enrolled in Wright's section. In particular, they advised that the best form of assessment would be to have faculty outside the Chemistry Department conduct oral examinations with of all the Chem 110 students, evaluating their ability to perform critical thinking and problem solving. Thus, the idea for the "faculty assessment" study was born.

The Chem 110 Faculty Assessment Design

With this idea in mind, Wright sought resources to help him conduct and finance this study. Fortunately, he was able to tap into grants funded by national and campus agencies (see Acknowledgements). For help in conducting the study, he turned to a third-party evaluation research group, the UW-Madison Learning through Evaluation, Adaptation, and Dissemination (LEAD) Center. LEAD and Wright designed a two-pronged study, with the faculty assessor study evaluating student learning outcomes, and a second study evaluating student learning processes. LEAD took full responsibility for the design and conduct of the study of student learning processes, which relied on qualitative research methods and is not described here. LEAD also refined and implemented the faculty assessment study. Important findings of the qualitative learning process study include that Wright and the comparison professor each implemented their chosen teaching methods at a high level of performance, and that student experiences of the learning processes in the two different sections were strongly consistent with the findings of the faculty assessor study.

In brief, the Chem 110 faculty assessment study was designed to determine whether faculty assessors could perceive differences between the competence of students in Wright's "structured active learning" (SAL) section and the comparison section. The design required faculty outside of the Chemistry Department to conduct oral exams with groups of students drawn from both sections. The faculty were not aware of the section in which the interviewed students were enrolled. Moreover, the great majority of the participating faculty were unaware of the methods used by either lecturers. All students enrolled in the course were assessed (103 from the SAL section and 77 from the comparison section).

Students divided by octile based on past performance

It was determined that the students in the two Chem 110 sections had similar statistical patterns in Chem 109 performance. This made it possible to minimize the differences attributable to Chem 109 performance that any one faculty assessor would see. This was accomplished by combining the students from both Chem 110 sections into one list, sorting them by Chem 109 performance, and dividing the resulting list into octiles. Thus each octile contained approximately 24 students who had performed at essentially the same level in 109. Each octile was divided, in turn, into three approximately equal groups, each of which had proportionate representation from each Chem 110 section. Each of these groups of approximately 8 students were then assigned to a faculty assessor. During the last two weeks of the semester, the students in each faculty assessor group were interviewed individually by the faculty member assigned to their group.

Faculty drawn from diverse engineering, science and math disciplines

All the faculty assessors were invited to participate in the study through a letter signed by both Chern-110 lecturers. The assessors were drawn, with the exception of four mathematicians, from disciplines that rely on chemistry, e.g. biochemistry, chemical engineering, pharmacy, soil science, and geology. In other words, most of the faculty assessors were "clients" of Chem 110, in that they taught upper level courses requiring the knowledge of analytical chemistry and solution equilibria that Chem 110 students are expected to know. Upon being interviewed afterward about their assessment work, most indicated that they participated in the study because they had a vested interest in students' learning experience in Chem 110. The four mathematicians were asked to participate in order to provide an additional interdisciplinary perspective.

Syllabi and textbook materials provided to the faculty

Enclosed with a second letter sent by the two Chem 110 lecturers to each faculty assessor was a brief syllabus for each section of Chem 110, plus copies of the table of contents of the course textbook and selected chapters on acid/base, complexation, and oxidation/reduction. These were intended to provide the participating faculty members information about the material covered in Chem 110.

Individual faculty developed their own criteria and exam

Each faculty member was provided a packet of materials including information about how to proceed with their oral exams and document their assessments. The instructions requested each to develop his/her own criteria for "assessing the competence" of the chemistry students. To both guide and document faculty efforts to design their exams, each assessor completed an "Oral Exam Preparation Exercise," which asked the faculty to "List and describe factors you will look for in your dialogue with each student to assess his/her competence." Prior to their interviews some of the assessors attended one of three orientation sessions held by a science faculty member from a department outside the Chemistry Department who did not act as a faculty assessor, but who was familiar with the experiment. These sessions served to familiarize the assessors with the established exam procedure and provided a forum in which the assessors could brainstorm and discuss their ideas for the exams.

Implementation of the Faculty Assessment

The LEAD Center scheduled the time for each oral exam and informed each assessor and student of the exam time and place (the faculty assessor's office). Faculty were instructed not to seek to learn from students which section they were in, and students were instructed not to disclose this information either. Interviews generally lasted thirty minutes with some as short as twenty minutes and some as long as forty-five minutes. The faculty ranked the students in their group both relatively and on their own absolute scale. Information about the students' and the faculty's perceptions of the oral exams was gathered through several different methods.

Information from the faculty

Faculty "Survey Sheets"

The faculty completed a "Survey Sheet" for each student. They used these sheets to assess each student individually prior to determining the relative order in which they should be ranked. The sheet contained four Likert scale statements/questions and an open-ended question, as follows:

1. During the oral exarn/interview this students appeared at ease.
2. This student is well-prepared for introductory courses in science majors.
3. I am confident that this student's performance on this oral exam reflects his/her true competence.
4. Taking into account all the criteria formulated in my oral exam preparation exercise, this student demonstrated overall competence.

The open-ended question asked the assessor to describe the factors that reflected on the student's overall competence. The assessors also indicated the order in which students were interviewed.

Faculty "Summary Questionnaire"

After interviewing all of their students, each faculty member completed the Summary Questionnaire, which requested them to rank the students in two ways and asked open-ended questions about the interview process and outcomes.

Absolute and relative rankings. The faculty assessors assigned their students a relative ranking from 1 to 8, number 1 being the student whom they considered the most competent in their group. They also indicated the spread in their competency ratings on an absolute scale to show how much difference they perceived among students.

Open-ended questions about the oral exam process. One open-ended question asked faculty if, as they proceeded through the interviews, they had modified their understanding of "competence" as defined in their Preparation Exercise, and if so, how. A second requested them to comment on the entire process, including the evaluation sheets and the rankings. Other questions asked them to state any prior knowledge they had of the teaching methods utilized by either Chem 110 lecturer and to describe their own teaching "methods.

Open-ended interviews with the faculty about the oral exams

After the faculty had completed interviews with all students, LEAD Center staff interviewed each of the faculty assessors. The interviews focused primarily on the faculty's criteria for assessing the students and the process that they used in the oral exams. Among the many findings from these interviews is that the experience of serving as an assessor for this study left many of the faculty participants with a keen interest in the teaching methods used by John Wright and the outcomes of the study. Based on remarks made by many of the assessors, both LEAD Center researchers and members of the chemistry education reform group associated with the NSF-funded New Traditions Chemistry Consortium believe that an important side-effect of the assessment study is that it provided 25 science, engineering and math faculty at UW-Madison with a meaningful introduction to this experiment in improving student learning.

Information from students: Student Survey

Directly after the oral exam, the faculty member handed each student a survey from the packet provided by the LEAD Center, and asked the student to complete, seal it in an envelope, and return it. The survey, which included Likert scale, multiple choice and open-ended questions, focused on the students' experiences in both the oral exams and the course overall.

Student experiences during the oral exam process

Questions about the oral exam process were designed to elicit information about the student's comfort level, self-assessment of the quality of their performance during the oral exam, and perceptions of the criteria the faculty intended to employ to assess the students. Self-assessment questions included the four Likert scale statements/questions below:

During this exam I demonstrated what I learned in the course.
In the interview I demonstrated the ability to relate my knowledge to new contexts.
In the interview I demonstrated that I am knowledgeable of chemistry.
In the interview I was fluent in responding to the examiner's questions.

The students also were asked to give themselves a numerical grade on a 100 point scale based on their performance during the exam.

Students then described the criteria that they believed the professor who interviewed them used to evaluate the students' performance on the exam.

Student experiences in the course overall

One of the Likert scale questions about the course overall asked students how well prepared they felt for other introductory science courses. Another asked students to compare what they had learned in

Chem 110 with other courses and to indicate whether or not they would "rate Chem 110 at the top" of their learning experiences.

'—Students "indicated the number of hours per week they spent on Chem 110 "doing out-of-class work" and the percentage of those hours that they spent working with other students. Another question asked students if they intentionally sought to enroll in either Chem 110 section, and if so, why. Finally, students were asked an open-ended question: "Please comment on your experiences in Chemistry 110."

Assessment Findings

With no knowledge of students' Chem 110 section, or even that they were evaluating students from the same octile based on Chem 109 performance, the faculty assessors evaluated the competence of the students in their groups. The quantitative responses provided on the Summary Questionnaires and Faculty Surveys were then analyzed using a variety of statistical methods, with the assumption that if significant differences were found, it was reasonable to attribute them to differences in the learning environments in the two sections.

All statistical tests showed that the faculty assessors ranked the SAL students as more competent. For example, in response to the question, "Taking into account the criteria formulated in my preparation exercise, this student demonstrated overall competence," on a scale of 6, the mean response for SAL students was 4.79, while the mean response for comparison students was 4.17. These mean responses are extremely unlikely to have occurred by chance (p value .0013). Likewise, where a relative rank of 1 means "most competent," the mean ranks assigned to SAL and comparison students, respectively, were 3.68 and 4.80. These outcomes are extremely unlikely to occur by chance (Signed Test p value .0227; Wilcoxon Matched-Pair Signed Rank p value .0066). With respect to the measure of absolute rank, where 1 is highest, the mean ranks assigned to SAL and comparison students respectively were 1.77 and 2.22 (Mann-Whitney p value .0002). By the same token, faculty responses to the question of whether students appeared at ease during the oral exam showed no significant difference between SAL and comparison students, indicating that student nervousness was not a factor in the differential performance of students in the two groups.

An additional analysis was performed to determine if the 67 percent of the SAL students and the 20 percent of the comparison students who intentionally registered in their respective lecture sections performed differently than those who had no preference. This analysis showed that the SAL students who did *not* intentionally choose the SAL section performed somewhat better in faculty assessor ranks than those who self-selected into this section. Taken together, these findings lead to the conclusion that the SAL teaching methods, not student aptitude or attitude, had significant effects on student learning, learning outcomes. These findings are supported and extended by the qualitative study of student learning processes. Presentation of these latter findings goes beyond the scope of this paper.

Findings from the student survey responses were consistent with these faculty assessor responses and support the same conclusion. For example, while there was no significant difference between SAL and comparison students in student responses to the questions about how at ease they felt during the oral exams, there were significant differences in SAL and comparison student responses to questions about whether they demonstrated the ability to relate their knowledge in new contexts and whether they felt well prepared for other science courses.

Acknowledgements

The Chem 110 evaluation research was funded by the ARPA TRP "Diversity and Cultural Change: Manufacturing Engineering Education for the Future" grant, EEC8721545 (administered by the NSF), the New Traditions: Revitalizing the Curriculum" grant from the NSF's Course and Curriculum Development: Systemic Changes in the Chemistry Curriculum program (Grant No. Due-9455928), and a Hilldale Foundation grant from the University of Wisconsin-Madison Chancellor to the LEAD Center.

References

1. For a thorough description of the course design and the underlying rationale, see "Authentic Learning Environment in Analytical Chemistry using Cooperative Methods and Open-Ended Laboratories in Large Lecture Courses," Journal of Chemical Education, John C. Wright (accepted for publication).

Biographical Information

SUSAN B. MILLAR is Director of the Learning through Evaluation, Adaptation and Dissemination (LEAD) Center at the University of Wisconsin-Madison. The LEAD Center primarily conducts evaluation research in support of UW-Madison faculty efforts to improve undergraduate and graduate student learning. Dr. Millar received her Ph.D. in anthropology from Cornell University.

STEVE KOSCIUK is a Researcher at the LEAD Center, University of Wisconsin-Madison. He is a mathematician (Ph.D. University of Wisconsin-Madison) with professional experience in statistical analysis and teaching.

DEBRA PENBERTHY (B. S.) is a Researcher at the LEAD Center, University of Wisconsin-Madison.

JOHN C. WRIGHT is the Evan Helfaer Professor of Chemistry at the University of Wisconsin-Madison. An analytical chemist with a research interest in spectroscopy, he received his Ph.D. in Physics from Johns Hopkins University, and has received numerous teaching and research awards.