

Office of Assessment

2006 Course Evaluation Pilot

Summary Results

Prepared by Kathryn Doherty

During the Spring 2006 semester, a pilot course evaluation instrument was administered in four departments of the College of Liberal Arts: Geography, History, Psychology and Sociology/Anthropology. This pilot was in follow-up to the Fall 2004 course evaluation pilot conducted in the College of Business and Economics. Initial analyses of student response and of the instrument suggest response results consistent with anecdotal and research-based evidence on course evaluations and support both the validity and the reliability of the instrument itself.

Background

Using statistical analyses of the Fall 2004 results, faculty feedback, and Promotion and Tenure Committee input, a revised course evaluation instrument was developed for the Spring 2006 pilot. This instrument has 25 items: 13 course, instructor, and curriculum-related items; 4 overall rating items; 5 student demographic items; and 3 open-ended questions. The open-ended responses were not analyzed as part of this pilot but were returned directly to faculty. Within the four departments that participated in this pilot, there were 52 course sections, 50 faculty, and 1785 registered students identified for the pilot. The response rate was 71% or 1268 of the 1785 registered students. The following table presents this sample by department:

Department	Sections	Faculty	Registered	Responding	Rate
CLA	52	50	1785	1268	71%
Geography	7	7	248	174	70%
History	10	10	357	214	60%
Psychology	20	19	522	401	77%
SOAN	15	14	658	479	73%

Analysis

Analysis of the pilot data was conducted in two parts. The first part was an analysis of the student responses. This analysis was conducted by course, by department, and by college. Summary reports were prepared and distributed to the faculty who participated in the pilot, to their Department Chairs, and to the CLA Dean. Faculty reports contained the instructor's results, overall department results, and overall college results. Chair reports contained course results for each participating course in the department, overall department results, and overall college results. The Dean's report contained overall results for each participating department and overall college results.

The second part of the analysis was a statistical analysis of the instrument itself. This consisted of factor analysis, reliability analysis, and comparison of means. The factor analysis was conducted to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis was used to identify a small number of factors that explain most of the variance observed in a much larger number of variables. Reliability analysis is used to study the properties of the scales created through factor analysis and the items that make up these scales. Reliability analysis helps in determining whether responses across

multiple items are consistent and also provides information about the relationships between individual items in the scale. Comparison of means is used to explore patterns of relationships between variables and to look at whether different groups or categories of respondents are responding in different ways. Each of these analyses is discussed below.

Factor Analysis

Factor analysis of the 17 non-demographic variables on the pilot course evaluation instrument was conducted using varimax extraction and a rotated correlation matrix solution that produced two component factors (or cluster of individual variables). The first factor had 13 strongly correlated variables. These variables were related to pedagogy and curriculum. The second factor had 4 strongly correlated variables. These variables were related to student learning. Variables within each factor are listed in the following table. These factors are in order of the strength of their correlation.

Factor 1: Pedagogy and Curriculum		Factor 2: Student Learning
The instructor provided clear grading standards		I worked hard to meet the requirements of this course
The instructor explained important ideas clearly		Course requirements challenged me intellectually
The instructor overall		I learned a great deal in this course
The instructor’s contribution to my understanding of course concepts		Assignments asked me to integrate information from various sources.
Methods for evaluating my work were fair		
The course as a whole		
The instructor treated me with respect		
The instructor gave prompt feedback on my academic performance		
The course content overall		
The instructor was accessible outside of class for help with this course		
Tests and assignments covered the most important points of the course		
Course objectives were clearly stated on the syllabus		
The instructor displayed enthusiasm about the subject		

Correlation analysis of the two factors produced a statistically significant correlation statistic of .734, indicating a strong, positive relationship between the two factors. The strength of the correlations in the factor matrix indicate consistent grouping of related variables around each of the two factors.

Reliability Analysis and Scaling

The second analysis that was conducted on the evaluation results was a reliability analysis. Two scaled variables were created from all of the non-demographic variables using the factor analysis results as a guide. Scaling is a computation process that creates new variables based on grouping contiguous values of existing variables into a limited number of distinct categories. Correlation analysis of the two scaled variables demonstrated an even stronger correlation between the two scaled variables than between the two factors upon which the scales were based, further supporting the strength of the relationship among the clustered variables. Reliability analysis on the scales showed very high levels of reliability, using Cronbach's alpha to demonstrate internal consistency of both the scaled variables and the individual responses from which these variables had been generated.

The following table presents the alpha reliability for each scaled variable.

Scaled Variable	Alpha Reliability
Pedagogy and Curriculum	.94
Student Learning	.82
Overall Response (from Pedagogy and Curriculum)	.93
Pedagogy and Curriculum without Overall Response	.91

Alpha reliability results close to 1 indicate a strong relationship among individual variables within the scale and consistency across response on these items. Analysis of individual variables within each scale showed that in no case did the alpha reliability decrease with the removal of any variable from the scale; thus, the scaled variable was stronger by keeping each variable originally identified through factor analysis than by removing any one variable. These results support the reliability of the course evaluation instrument.

Comparison of Means

The third analysis that was conducted on the pilot course evaluation result was a comparison of mean response by demographic group. Results from these analyses support intuitive, anecdotal, and research-based analyses of expected student response, and thus support the validity of the evaluation instrument. Five demographic variables were identified as independent variables through which to compare student response as represented by the scaled variables generated using factor analysis. These independent variables include: reason for enrolling in course, current year in college, expected grade in course, attendance record for course, hours per week preparing for course. Results from each are summarized below.

In the first analysis of mean, the two primary scaled variables (Student Learning and Pedagogy/Curriculum) were compared using main reason for enrolling as the independent variable. The results (mean response) are summarized in the following table.

	Student Learning	Pedagogy/Curriculum
Overall Mean	4.15	4.34
Someone recommended course or instructor	4.53	4.70
Gen Ed requirement	3.97	4.20

Looking at the mean response, where 5.00 is high (strongly agree, excellent, etc.), results using reason for enrollment as an independent variable highlight the differences in student evaluation of a course or instructor based on the reasons the student is taking the course. All results are statistically significant at $p \leq 0.000$. Based on an overall mean score of 4.15 for student learning, students who enrolled in a course because it was recommended responded more favorably on the course evaluation than students who enrolled because the course was a general education requirement. The same statistically significant results occurred for the scaled variable of pedagogy/curriculum. Overall mean was 4.34. Students enrolled because of a recommendation were statistically more likely to give a favorable evaluation than students enrolled to fulfill a general education requirement. These results should be considered when interpreting student response within individual courses.

When comparing means on the two primary scaled variables (Student Learning and Pedagogy/Curriculum) using current year in college as the independent variable, the following results (mean response) were evident:

	Student Learning	Pedagogy/Curriculum
Overall Mean	4.16	4.35
First Year Students	4.05	4.19
Seniors	4.23	4.47

As with the results for main reason for enrollment, these results are not unexpected. Looking at the Student Learning variable and using the overall mean of 4.16 as the gauge, first year student responses are less favorable than senior student responses ($p < 0.000$), 4.05 compared to 4.23. Looking at the Pedagogy/Curriculum variable with an overall mean of 4.25, first year response was again less favorable than senior response. There are many ways that these data might be interpreted (e.g., student experience, expectations, understanding, etc.) and this should be taken into consideration when analyzing course evaluation results. For instance, if first year students are statistically more likely to give lower (less favorable) responses on the evaluations, perhaps courses with high enrollments of first year students might have lower overall evaluation results. These considerations become especially important in light of the number of students enrolled in general education courses.

The third independent variable used to compare the mean response on the two scaled variables was expected grade in course. Again, results match anecdotal experience and are summarized in the following table:

	Student Learning	Pedagogy/Curriculum
Overall Mean	4.16	4.35
Grade of "A"	4.27	4.53
Grade of "D"*	3.46	3.63

*Please note that the sample size for "D" grades is small compared to the overall sample so results should be retested in future administrations of the course evaluation.

Using an overall mean of 4.16 on the Student Learning variable, students expecting a grade of A are statistically more likely ($p \leq 0.000$) to respond more favorably on student learning items than students expecting a grade of D. This also holds true for the Pedagogy/Curriculum variable. These results should be considered when looking at overall course evaluation results for courses in which students typically receive lower grades (e.g., upper level courses, courses in specific disciplines, etc.).

Using attendance record for the course as the fourth independent variable when comparing mean response on the two scaled variables, results indicated that students who attend class more often are statistically more likely ($p \leq 0.000$) to respond favorably on course evaluation items. These results are summarized below:

	Student Learning	Pedagogy/Curriculum
Overall Mean	4.15	4.34
Never missed a class	4.27	4.40
Missed 5 or more classes	3.79	4.11

Attendance results support both anecdotal evidence and evidence in the literature that students who attend more often are more likely to be engaged in a course, successful in a course, and offer positive evaluation of the course. These results should be considered in light of existing research and experience, as well as within attendance policy requirements and attendance records for individual courses.

The final independent variable used to compare mean response on the two scaled variables was preparation time for the course. The following table summarizes these results:

	Student Learning	Pedagogy/Curriculum
Overall Mean	4.15	4.34
3-4 hours preparation per week	4.34	4.43
7 or more hours	3.80	
0-1 hours preparation per week		4.12

When looking at student response on Student Learning variables, students who spend about 3-4 hours preparing for the course each week respond more favorably ($p \leq 0.000$) when evaluating a course. Students who spend 7 or more hours per week preparing for a course respond less favorably. This may suggest that students who are spending more hours than average studying for the course are possibly struggling with the course and may thus evaluate the course or the instructor less favorably. When looking at student response on Pedagogy/Curriculum, students who spend 3-4 hours a week preparing for the course are again responding more favorably. Students who spend 1 hour or less preparing for the course respond less favorably. Here, one might interpret these results as indicating students who are less engaged in a course are less likely to offer more positive evaluations of the course. Additionally, 3-4 hours of preparation time may be a time commitment that fosters engagement without producing the overload of preparation time (7 hours or more) that may be associated with less student satisfaction with the course and the instructor.

The comparison of mean response on the Spring 2006 course evaluation pilot supports anecdotal and intuitive understanding of the ways in which students respond on course evaluations. These results also support the research and the literature on course evaluations which suggest that the ways in which students respond may be impacted by student demographic factors (i.e., reason for enrolling, expected grade, etc.). Results should be considered when reviewing the evaluation results from individual courses.

Summary

Four departments, 50 faculty, 52 sections, and 1785 registered students participated in the Spring 2006 course evaluation pilot project. Analysis of the data supports the validity and reliability of the instrument and the intuitive, anecdotal, and research-based experience related to student response on course evaluations. Factor analysis produced two clusters of variables that were scaled into categorical variables with strong correlations. These variables were analyzed for reliability and differences in mean response. Reliability analysis of the scaled variables supported the reliability of the course evaluation instrument. Comparison of means points to statistically significant differences in student response based on reason for enrolling in the course, year in college, expected grade, attendance record, and hours of preparation per week. This pilot will be expanded during the Fall 2006 and Spring 2007 semester. Results will be analyzed and compared to the Spring 2006 findings.