SOOT Ratings Vary by Context & Gender

by Dr. Paul Schnorr, Assistant Professor of Sociology & Dr. Matthew Stollak, Assistant Professor of Business Administration

Editor’s Note: This is the first of two issues of Assessment News devoted to research on the Student Opinion of Teaching (SOOT) and the process by which we currently evaluate teaching. This month’s issue presents a statistical analysis of 5 years of SOOT data undertaken by Drs. Paul Schnorr and Matthew Stollak with support from the Office of Institutional Effectiveness. Next month’s issue will be devoted to an analysis of faculty and student perceptions of the purpose and usefulness of the SOOT process. Schnorr & Stollak’s statistical analysis refers to some 135 pages of tables contained in 14 appendices. The appendices are not reproduced here, but can be reviewed on the OIE web site under “Reports & Presentations.” https://www.snc.edu/oie/

Data and Methods

The data for this project was drawn from several sources. The largest section of data are the SOOT responses from students for the years 1999 through Fall 2004. Faculty were sent an email informing them about the research and asking them to contact the Office of Institutional Effectiveness if they did not want to participate. Only one faculty member chose not to participate. Therefore, the data set is a complete set of SOOT scores of virtually every faculty member at St. Norbert over five years (53,178 individual records from 3183 classes). In addition to the evaluation data, the course GPA was also added to each record. For example, if a student was in a particular class with 25 SOOTs returned, there would be 25 records for that class with the overall class GPA being the same for each of the 25 records. Therefore, any analysis that includes GPA is operationalizing this variable not with individual student GPA, but with the GPA for all students in that course. (Continued on Page 2)

SCOLA Conference Report

by Dr. Tom Conner

Since spring 2004 I have attended the yearly foreign language conference organized by SCOLA, a non-profit consortium offering retransmissions of foreign television news broadcasts 24/7 to educators and the community at large. I wanted to learn more about how to use SCOLA in our curriculum, specifically with regard to the assessment of cultural proficiency. Beginning last year, I used SCOLA news broadcasts to assess cultural proficiency in French 375 (Introduction to French Civilization). In this time I have come to realize that assessment can actually work for us, insofar as it helps teachers realize some of our course objectives: I mean, what more logical venue for assessing cultural proficiency than a course on French civilization?

SCOLA now offers more programming, including documentaries and films, and plans to make these available through a new technology called “video streaming” or “webstreaming,” making it possible to download programs from the Web. Working with the good people in Media and Computer Services, I plan to implement this exciting new technology into my curriculum, making it possible for students to log in from their residence if they chose.

I set out to achieve several different objectives at this year’s SCOLA conference. I still had many nuts and bolts type of questions about webstreaming, and finally had a chance to get some answers from SCOLA staff. Second, I was actually a featured speaker at this year’s event and participated in a series of talks and a round-table discussion on assessing cultural awareness. In connection with this roundtable I met a number of interesting people both in education and in business who helped me think about various ways in which I can improve our assessment tools at St. Norbert College. In fact, I have already been retained as an “expert” on cultural assessment and will be making a few trips this summer to other institutions to talk about cultural assessment. Lastly, I finished my article on SCOLA to be published in The NECTFL Review. (Continued on Page 6)

Assessment Conference Possibilities

Indiana University-Purdue University Indianapolis, The 2006 Assessment Institute, October 29-31, 2006.
AIRUM 2006 Annual Conference, November 2-3, 2006 at Sheraton Hotel, Bloomington, MN.

Volume 6, Issue 2
October 9, 2006
SOOT Ratings Vary by Context & Gender

(Continued from Page 1)

To provide the greatest degree of anonymity possible, the investigators first received a data file with course numbers and names of faculty and faculty ID, but no evaluative data. From this data variables were created for the time slot of the course, the length of service at St. Norbert for each instructor, and the tenure status of faculty members. Each faculty member was then assigned a 5 digit random ID number. This data was then sent back to computer services and a new data file was created that included the new variables along with the evaluative data, but which did not include course numbers, names of faculty, or the institutional ID of faculty. Therefore, it would be very difficult, if not impossible, to identify many faculty members.1

Most analysis in this report are simple comparisons of means and correlation analyses that compare the SOOT ratings under different conditions (type of class, length of class, etc.) and for different groups or conditions (male/female, small class/large class, general education/elective/major/minor, etc.). Our analyses of the data drawn from SOOT forms does not include any discussion of statistical significance because our data set is a population rather than a sample. In other words, we have all of the SOOT data gathered between Fall 1999 and Fall 2004 and, therefore, all differences in any summary statistics are characteristic of the population with no sampling error. The substantive significance of differences in mean scores or correlation coefficients will be noted by highlighting the range of variation. For example, when the difference between male and female SOOT scores for overall quality of a course is only .03 on a 5 point scale, this does not seem to be a large difference. However, when the range of mean SOOT scores for overall quality of course between small courses and large courses taught by women is .46, this is almost half a point and appears to be something that is substantively significant. Moreover, when one considers that the data is skewed with a high concentration at the high end, a change of .46 means that one can drop over one quartile in a distribution of means for overall course or instructor rating. (See Appendix 3)

The report concentrates on two particular measures – Overall Rating of Course and Overall Rating of Instructor. The decision to concentrate on these measures reflects both analytical and practical logic. Analytically, these two measures are very highly correlated with most other items measuring perceived instructor preparation and classroom performance, with coefficients over .5. (See Appendix 1) The only items that correlate with these two measures that are below .5 are the following:

- Student Well Prepared
- Student Actively Participates
- Student Does Part to Learn
- Student Can Meet with Instructor Outside of Class
- College Resources are Sufficient

With the exception of the perceived workload and ability to meet with instructor outside of class, these items primarily measure student engagement rather than characteristics of the course or instructor. Further, it is likely that the lack of correlation of these items with Overall Scores is a result of the lack of variation on these items (See Appendix 2) with the Workload item having the smallest standard deviation (.337 on a three point scale) and Availability being generally rated high with a small standard deviation (.57 on 4 point scale). The relatively low correlation of the items measuring overall course and instructor quality with the student self-rating items may reflect that the student items may not capture characteristics of courses or instructors, but students. While the case could be made that good teaching elevates student engagement, instructors are limited in the control they have over student interest and behavior.

In practical terms, the Overall Course Rating and Overall Instructor Rating items are often the items that many instructors look to first when viewing SOOT results. These items are the most comprehensive measures on the SOOT form and by using these measures for many analyses it is possible to report findings in a more parsimonious fashion.

Findings

Overall Measures of Student Evaluation of Teaching

The data from 1999 to 2004 give an overall picture of relatively high ratings from students for instruction and courses. The table in Appendix 2 illustrates that with the exception of an item over which instructors have relatively no control (College Resources are Sufficient), the specific items that either rate instructor preparation or classroom performance are 3.3 or higher on a 4 point scale. (Those who lament grade inflation for students should perhaps take note that if assessment of teaching by students was normally distributed around an “average” of 2.5, many instructors would find themselves looking at much lower SOOT numbers).

SOOT Scores and Length of Class Period

St. Norbert class periods are either 50 minutes, 70 minutes, 2 hours or a combination of two 50 minute periods and one 2 hour session (the morning floater). The length of a class period and the SOOT scores generally do not vary much on specific evaluative items, but the variation is very consistent. The graphics (see Appendix 3) indicate that ratings of 50 minute periods and 70 minute periods are generally higher than the morning floater and the two hour periods. Interestingly, the greatest variation occurs in the overall evaluation of course and instructor. The morning floater receives the lowest overall score for both course and instructor (3.80/5 and 4.0/5 respectively) while 70 minute classes receive the highest scores for course and instructor (3.99/5 and 4.19/5 respectively). In both cases, the .19 difference is relatively small, but still enough to cause the scores to vary from above the college average to below the college average. One other interesting fact is that students tend to rate their own preparation and work in class to be highest in the 2 hour periods. (Continued on Page 3)

1It may be possible to mine the data to identify some members if one were to take the time to look at the gender and length of service along with type of class taught during a semester and then comparing that with timetables from the registrar’s office. This process would require SPSS skills and access to the data. The data set, however, is only in the hands of the two researchers and will not be made publicly available to prevent this sort of laborious identification.
SOOT Ratings Vary by Context & Gender
(Continued from Page 2)

SOOT Scores and Time of Class Period
The tables and graphs in Appendix 4 indicate that the time of class sessions has a minor influence on SOOT scores, with a few notable exceptions. First, time slots “K,” “L,” and “M,” which are the evening class meeting times are generally rated higher than the day-time slots. These courses have very small numbers (less than 10% of any of the day-time slots) and so any comparisons with the other time slots should be done with caution.

The second exception that appears to be consistent is that the ratings of courses and instructors in the “C” (morning floater), “I” (MW 3-5) and “J” (T 1-3, Th 3-5) time slot are slightly lower than other time periods. The “G” (T 11-1, Th 1-3) slot is also somewhat lower in student evaluations, but not as low as the “C,” “I,” and “J” time slots.

“ratings of courses and instructors in the “C” (morning floater), “I” (MW 3-5) and “J” (T 1-3, Th 3-5) time slot are slightly lower than other time periods. The “G” (T 11-1, Th 1-3) slot is also somewhat lower in student evaluations, but not as low as the “C,” “I,” and “J” time slots.”

SOOT Scores and Type of Course
The data regarding the type of course is somewhat problematic because students self-report this information and some may have been confused regarding how to mark courses. For example, many major, minor and elective courses are cross-listed as general education courses. Therefore, unless students are instructed carefully, they may not have understood how to classify a course.

However, with no way to determine if there is a systematic misclassification of courses, we simply report the results with a cautionary statement that the data probably includes some error that may affect the findings. Students’ evaluation of different types of courses – electives, general studies courses, require major courses and required minor courses – show clear patterns. The range of variation on different evaluative items varies, but the pattern is very robust – electives are rated higher than major or minor courses which, in turn, are evaluated more positively than general education courses (see tables and graphs in Appendix 5). Overall Rating of Course shows .25 points of variation between electives and general education courses with mean scores of 4.09/5 and 3.85/5 respectively. This .24 difference is greater than the variation for instructors, which is 4.26 for electives and 4.08 for instructors (a .18 difference). This may indicate that while student’s satisfaction with required general education courses reflects an aversion to courses that may not match their interest, they still respect good teaching. We note below, however, that the relationship between SOOT scores and type of course tends to vary more for women than for men.

“electives are rated higher than major or minor courses which, in turn, are evaluated more positively than general education courses”

SOOT Scores and Class Size
The relationship between SOOT scores and class size shows that, generally speaking, larger courses tend to receive lower evaluation scores than do small courses. While the data for course size is not perfect because we use the number of SOOTs returned as proxy for class size, it was our available measure. By looking at the SOOT summaries from F2002 until S2005, the years that both researchers administered SOOTs, the average college response rate was between 83% and 91%. Therefore, the data probably understates class size, and classes that are near cut points (10,20,30) may be misclassified. This problem is probably most acute with classes of 31+ students being classified as classes of 21-30 because the range of classes in this category are probably between 31 and 35 rather than having a full 10 point range as in the other categories. The relatively small differences between the categories of 21-30 and 31+ may reflect the fact that many 31+ classes are influencing the 21-30 category. Or, alternatively, the lack of variation between these categories may indicate that after a class gets over a certain size, the influence of the class size doesn’t change much as the class grows larger. Regardless, we caution the reader to recognize that there may be some noise in the data.

“generally speaking, larger courses tend to receive lower evaluation scores than do small courses.”

Using the number of SOOTs returned and correlating with Overall rating of Course and Overall rating of Instructor we find a negative relationship with coefficients of -.094 and -.085 respectively. While these are not strong correlations, in light of the relatively high scores for SOOTs overall and skew of the data, they indicate that class size does have some relationship with scores.

To explore this relationship more fully, the number of returned SOOTs in each class was collapsed into a four category variable with values of 1to 10 students, 11 to 20 students, 21 to 30 students and over 30 students. Despite the fact that courses with only one SOOT returned may not be typical courses (many may be Independent Study), these are included in the category of 1 to 10 students because only 115 of the 5326 courses in this category are courses of size one (about 2%) and, therefore, should not skew the results significantly. Mean scores were then generated for Overall Course Rating and Overall Instructor Rating. (See Appendix 6) The mean score for both course and instructor decline as class size increases until classes get over 30 and then the scores show a slight increase over the 21 to 30 category. The difference between the highest Overall Rating of Course is for classes of size 1 to 10 with a mean score of 4.19 while the lowest score is for classes with 21 to 30 students with a mean score of 3.86 (a .33 difference). The range of difference in mean scores for Overall Rating of Instructor is similar with a .31 range between the high of 4.36 for classes with 1 to 10 students while mean evaluation score for instructors teaching classes of size 21 to 30 is 4.05.

To explore whether these differences simply reflect different ability of instructors teaching smaller courses vs. those teaching larger courses, we constructed an analysis of the Overall Course Rating (Continued on Page 4)
SOOT Ratings Vary by Context & Gender
(Continued from Page 3)

and Overall Instructor Rating for individual instructors teaching classes of different sizes. Again, the classes were grouped into sizes of 1-10, 11-20, 21-30 and 31+ and means were calculated from all the SOOTs returned in classes of each size for each instructor. For example, if an instructor had taught 300 students in classes in size 21-30, these three hundred SOOT responses were averaged to create a single score for that instructor’s Overall Course Rating and Overall Instructor Rating in class size 21-30. Similarly, if the same instructor had taught 45 students in classes in size of 1-10, these 45 SOOT responses were averaged to compute the means for class size 1-10. This allows direct comparison of Overall Course Rating and Overall Instructor Rating for individual instructors teaching different size classes. The differences in means between different scores from different size classes was then computed for each instructor and these differences are the basis for an analysis that compares scores of different class size while controlling for instructor.

The findings reported in Appendix 7 and 8 suggest that for the majority of instructors, teaching larger classes result in lower Overall Course Rating and Overall Instructor Rating. In terms of Overall Course Rating, the mean difference in score between classes of 1-10 and 11-20 is .13, the difference between classes of size 11-20 and 21-30 is .11 and the difference between classes of size 21-30 and 31+ is .01. Most Notable is the difference between classes of size 1-10 and 31+, which is .36. In other words, the same instructor teaching very small classes will, on average, receive scores for Course Overall that are .36 higher than when she/he teaches very large classes. The numbers are similar for Overall Instructor Rating, with difference in mean scores between teaching classes of size 1-10 and 11-20 being .10, the difference between classes of size 11-20 and 21-30 being .13 and the difference between classes of size 21-30 and 31+ being .01. Again, the same instructor teaching small classes, on average, earns SOOT scores for Overall Rating of Instructor that are .33 higher than when the same instructor teaches classes of 31+ students.

A First Look at the Graduates of 2006

Since 2000, the College has administered the College Student Survey (CSS) to our graduating seniors. Last spring all graduating seniors were invited to complete the survey. We do not know for certain if the 229 who did so are representative of the entire class of graduates. However, it is possible to view their survey responses simply as the statements of a substantial number of seniors, keeping this perspective in mind while reviewing the findings below.

<table>
<thead>
<tr>
<th>SNC</th>
<th>Cath. 4-yr Coll. Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raise a family</td>
<td>81.2%</td>
</tr>
<tr>
<td>Be very well-off financially</td>
<td>42.9%</td>
</tr>
<tr>
<td>Influence political structure</td>
<td>18.8%</td>
</tr>
<tr>
<td>Help others in difficulty</td>
<td>70.4%</td>
</tr>
<tr>
<td>Promote racial understanding</td>
<td>40.3%</td>
</tr>
<tr>
<td>Become community leader</td>
<td>38.6%</td>
</tr>
<tr>
<td>Become authority in my field</td>
<td>59.7%</td>
</tr>
</tbody>
</table>

Opinions (“Somewhat/Strongly Agree”)

<table>
<thead>
<tr>
<th>SNC</th>
<th>Cath. 4-yr Coll. Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase Military Spending</td>
<td>24.3%</td>
</tr>
<tr>
<td>Abolish Aff. Act. in admissions</td>
<td>46.0%</td>
</tr>
<tr>
<td>Legalize abortion</td>
<td>52.3%</td>
</tr>
<tr>
<td>Legalize marijuana</td>
<td>44.4%</td>
</tr>
</tbody>
</table>
In summary, as class size increases, the general trend is that the Overall Course Rating and Overall Instructor Rating declines. Further, when controlling for individual instructors, this general trend holds for most instructors, and the difference in ratings for some individual instructors teaching small courses vs. large courses is fairly dramatic. There are a few notable exceptions to this general rule, with the most common set of exceptions being people who receive very high evaluations regardless of class size. A less common type of exception are those who receive very poor evaluations regardless of class size.

Gender and Student Opinion of Teaching

At first glance, the data seem to suggest that there are only slight differences in student evaluation of male and female instructors. The table in Appendix 9 illustrates that the differences in mean scores between males and females is never more than .06 on either the 4 or 5 point scales and usually is .04 or less. However, when controlling for type of class and class size we see evidence that suggest that the relationship between SOOT scores and these variables are more pronounced for women than for men.

Gender and Type of Class

The pattern of students’ evaluating electives most highly, followed by minor courses, major courses and then general education courses is consistent for both men and women, however the range of these different evaluations appears to be greater for female faculty then for male faculty (See Appendix 10). This disparity is very apparent in the overall evaluation of courses and instructors. While the difference in overall course ratings between the highest rated electives and general education courses is .20 for men (4.1 vs. 3.9), the difference for women is .34 (4.07 vs. 3.73). And while the difference in mean scores for the electives is not large, the average score for women teaching general studies courses is .17 less than the male average. These differences also appear in the overall rating of instructors with the difference between electives and general education courses being .13 (4.27 vs. 4.14) for men while the difference for women is over twice as large, being .30 (4.23 vs. 3.93).

Gender and Class Size

The relationship between SOOT scores and class size also becomes more complex when controlled for gender. As indicated above, as class size increases the SOOT scores tend to decrease. While this trend appears for both men and women, the differences are much greater for women than for men. (See Appendix 11) In terms of Overall Rating of Course, the difference between the highest rated small courses (1-10 students) and the largest (30+) is .15 for men (4.18 vs. 4.03) while the difference for women is .48 (4.22 vs. 3.74). The range of scores for the Overall Rating of Instructor item is similar with the difference between small courses and large courses being .09 (4.34 vs. 4.23) for men while the difference for women is over 5 times as great, being .46 (4.38 vs. 3.92). The greater range reflects women’s slightly higher scores for both course and instructor in smaller classes (20 or less), but lower scores in larger classes, particularly classes with more than 30 students.

Grades and Student Opinion of Teaching

The relationship between SOOT scores and grades was investigated by examining student evaluations in reference to the overall class GPA for each of the classes in the data set. While the ideal data set would provide the grade of each student along with their individual evaluation of an instructor, this was not possible. Therefore, we used the class GPA with the assumption that the higher overall grades reflect more high individual grades. Overall, the data suggest that there is a slight positive relationship between higher grades and higher SOOT scores. The table in Appendix 12 illustrates that the correlation between class GPA and Overall Rating of Course and Overall Rating of Instructor are .19 and .16 respectively. In other words, as course grades increase the SOOT scores tend to be slightly higher. These correlation coefficients are not large, but they do suggest that students evaluate courses more highly when the lass GPA is higher.

Length of Service and SOOT Scores

The relationship between SOOT scores and years of service was analyzed by creating a variable that collapsed years of service of instructors into 5 year categories. We also created a category for less than one year because most first year faculty struggle as they move into a new teaching environment and adjunct faculty who only teach one year often experience difficulty because they have little time to adjust their courses. The table in Appendix 14 indicates that the overall scores for course and instructor are lowest for new instructors with mean scores of 3.79 for course overall and 3.94 for instructor overall. The highest rated instructors, however, are in the next category of instructors with 1 to 5 years of service with mean scores of 4.1 for course overall and 4.3 for instructor overall. After 5 years there is a general pattern of slow, slight decline until instructors are in the category of 31 to 35 years, when the trend reverses itself for that cohort while those with more than 35 years of service decline again, but are still rated more highly than those with 26 to 30 years of service.

Major Findings

• Overall Ratings of Instruction at SNC is generally high.
• Two hour class sessions tend to receive lower ratings than other class lengths, with 70 minute classes receiving the highest ratings. (Continued on Page 6)
SOOT Ratings Vary by Context & Gender
(Continued from Page 4)

• Late afternoon classes that meet for two hours and the morning floater receive lower ratings than morning and early afternoon classes.

• General Education courses tend to receive lower ratings than other types of courses. Elective courses receive the highest ratings.

• Smaller classes receive higher ratings than larger classes. This was controlled for individual instructors teaching classes of different sizes and the trend was true for approximately two-thirds of instructors teaching classes of different sizes.

• Overall differences between male and female instructors are very small.

• When specifying for type of class and size of class, the relationships between these variables and student ratings are greater for women than for men.

• There is a small correlation between higher overall class GPA and higher student evaluations.

• The correlation between overall class GPA and SOOT scores is slightly greater for women than for men.

• Instructors with between 1 and 5 years of service at St. Norbert receive the highest SOOT ratings and the trend after 5 years generally is a slow, slight decline in SOOT ratings.

The major findings above obviously include some interactions that are difficult to disentangle. For example, the type of course and size of course are related as general education and required major courses are often very large while electives tend to be smaller. Similarly, the time of day that a class meets is related to how long the class meets. This initial report is meant to simply describe relationships between class and instructor characteristics and ratings, but these interactions suggest that one use caution when interpreting SOOT ratings.

*****************************************
SCOLA Conference Report (Continued from Page 1)

Thus, my continued relationship with SCOLA has resulted in many exciting learning opportunities for me and my students. SCOLA would like to make SNC a pilot institution, which means that we would be the first to be offered new programming and/or technologies and that our students would have the possibility of participating in a program to be retransmitted worldwide on SCOLA networking. I plan to implement web-streaming next year and plan to continue strengthening my assessment activities based on what I have learned from attending the SCOLA conference.

Assessment Resources
Office of Institutional Effectiveness
Main Hall, Room 219
(Phone: 403-3855)
FAX: 403-4096
Web site: www.snc.edu/oie/

Everything You Wanted to Know about the OIE Post Grant

• The OIE will continue to exist, but with a smaller budget.

• Academic, Student Life, and Mission & Heritage programs will need to continue to assess student learning to satisfy accreditation requirements.

• Assessment of General Education courses, critical thinking and writing will continue.

• OIE will continue to support assessment of student learning, but primary responsibility for assessment will reside with each program.

• OIE will have a budget for assessment mini-grants, but the number and size of the grants will be smaller.

• OIE will have a budget for conference attendance.

• OIE will continue publishing Assessment News.

• OIE will continue to build the evidence template for the next self-study and Higher Learning Commission site visit in 2011.

• OIE will continue to carry out its institutional research functions, e.g. SNC At-a-Glance, external reports, Factbook, program review data, student & alumni surveys, ad-hoc surveys, SPRAAC support, retention analysis.

• OIE will continue to monitor mission effectiveness.

To prevent a future Focused Visit…

• The Dean’s Council has suggested that an annual progress report be solicited from all programs to ensure that they are carrying out some portion of their assessment plan every year.

• Programs will be encouraged to adopt an assessment cycle of no more than 2½ years. For academic programs, the assessment section of the program review will serve as one assessment report. A second report analyzing data and, when appropriate, suggesting program improvements should be submitted midway through the five year program review cycle.