

University of Scranton
Guide to the Student Course
Evaluation Survey

Center for Teaching & Learning
Excellence

Summer 2007

University of Scranton

Guide to the Student Course Evaluation Survey

Contents

1.	Introduction.....	4
2.	Student Ratings and Teaching Quality	4
3.	University Policy	5
4.	Center for Teaching and Learning Excellence.....	7
	Mission and Services	7
	Administration of the USCS.....	7
5.	Other Groups Responsible for the Course Evaluation Process	7
6.	Interpreting the Course Survey Summary.....	8
	Course Information	8
	Controlling for Initial Student Interest	8
	Extreme Student Responses	8
	Other Potential Biases	9
	Course Comparisons (Above Average, Average, Below Average)	9
	Progress on Objectives Items	9
	How Can You Improve Your Course Ratings?	10
7.	Evolution of University Course Evaluations	10
	Early History.....	10
	Current University of Scranton Course Survey Form	11
	Individual Items Changes.....	11
	Addition of Above and Below Average Categories	11
	Comparison Group Changes	11
	On-Line Course Evaluations	12

Changes in the Statistical Analysis	12
8. Questions about the Course Evaluation Ratings	13
9. Further Reading	16
APPENDIX A	19
Sample Course Survey Form	20
APPENDIX B	25
Course Evaluation Comparison Analysis	26

1. Introduction

The University of Scranton Course Survey (USCS) is the official form for student evaluation of courses. Course evaluations are administered by the Center for Teaching and Learning Excellence (CTLE). This guide will provide information on the history, evolution, use, and interpretation of the USCS. Information is also provided to help faculty use the course evaluation as a tool to enhance the quality of teaching. To assist in the interpretation of official student evaluations, an explanation of the USCS summary, a history of the development of the USCS, and a list of commonly asked questions regarding the student evaluation procedure are included. For faculty interested in learning more about student ratings, suggestions for further reading are provided. Appendix A provides examples of the evaluation form and feedback information. Appendix B provides a detailed description of the statistical procedures used in the analysis of USCS data.

2. Student Ratings and Teaching Quality

A review of the literature on the course evaluation process suggests that student course evaluations are the most reliable and valid measure of teacher instructional quality presently available. For example, self-ratings of teaching performance are reliable but not valid measurements.

Student course evaluations reflect quality of instruction in at least two ways:

- 1) Reliable and face valid measures of student satisfaction with the instructor, course and goal attainment; student satisfaction is important in its own right.
- 2) Highly reliable and moderately valid index of variables that reflect instructional quality (Marsh, 1987; Marsh & Roche, 1997). For example, studies have shown that scores on standardized final examinations are in general moderately correlated to student ratings. Also, student evaluations are moderately related to alumni ratings, the ratings of trained observers, instructor self-ratings, and measures of student motivation.

One extraneous factor that may impact course evaluations is initial student motivation or interest in the content of the course. This factor is controlled for in our

Course Survey Summaries. However, other factors such as the interaction of student characteristics and course characteristics are not taken into account. Some other factors that might seem to be important but have only a minor or indeterminate effect are class size and discipline (e.g., humanities, social sciences, and physical sciences).

The categories used in the USCS include (below average, average, and above average) are based on a comparative process, not on an absolute scale of good to bad teaching quality. A course is categorized in relation to other courses at the University using standard behavioral science analyses. Discriminations based on extreme categories, (e.g. below average and above average) are least likely to be inaccurate.

3. University Policy

Administration of the USCS form is mandatory for part-time faculty each term and first year full-time faculty during fall and spring semesters. In accord with the University Senate recommendation of November 9, 1984, administration is mandatory for all others on a rotating semester/year basis (spring and fall one year, followed by a non-mandatory spring and fall). In practice, this has evolved so that evaluations are mandatory in both semesters of odd calendar years (e.g., 2003, 2005, 2007).

During the mandatory semesters, the results of the student evaluations of each faculty member are made available to the faculty member, the Provost's Office and to the respective dean who then gives that third copy to the respective department chair. A copy of the course evaluation summaries are placed in the faculty member's file that is housed in the Provost office. During non-mandatory semesters, the results of the student evaluations are available only to the faculty member who opted to be evaluated, with the exception for first year and part-time faculty who follow the above distribution. Intersession and summer session courses are non-mandatory for full-time faculty; however, they are mandatory for part-time faculty.

The Course Survey Summaries are used by the individual faculty, department chairs, and deans to help the faculty enhance their teaching skills. The Summaries are also used by the department chairs, deans, Provost, and Board on Rank and Tenure for making personnel decisions (see the Faculty Handbook citation below). Student comments from the evaluation form are only provided to the individual faculty member.

According to the Faculty Handbook, Appendix II, section A.1.:

The Board [on Rank and Tenure] will assess teaching quality on the basis of the

following items:

- The evaluations of departmental members and deans, who should specify the basis of their judgments
- Official student evaluations
- Written testimony from students and others
- Any material submitted by the instructor such as syllabi and samples of assignments and examinations.

According to University Senate Bill S-12/1980-81:

PROLOGUE

- A major goal of the University community at large is to achieve and maintain the highest possible quality in the educational process.
- Excellence in teaching is certainly an important aspect of reaching this goal.
- Assuming that excellence in teaching implies a continuous developmental process, faculty members should be concerned with improving and monitoring the quality of their teaching.
- The Administration of the University has a responsibility to evaluate the quality of teaching at the University.
- The Board on Rank and Tenure has a prescribed responsibility of evaluating the performance of individual faculty members.
- The opinions of students should be solicited in a systematic way as a part of a well prescribed evaluation process.

SPECIFIC GOALS RELATING TO STUDENT EVALUATIONS OF TEACHING

- 1) Student views should be given and received in a spirit of honest and constructive communication between faculty and their students.
- 2) Students should be cognizant of the basic purposes and importance of evaluations.
- 3) Student views should be recognized as an important factor but only one of a number of sources of information.
- 4) Student views should be obtained and presented in such a way as to maximize

their validity.

4. Center for Teaching and Learning Excellence

MISSION AND SERVICES

The Mission of the Center for Teaching & Learning Excellence is to promote a culture that supports the Scholarship of Teaching and Learning. The CTLE is responsible for the administration, interpretation, development, and use of the USCS. Results from the USCS can be used to help faculty develop and/or enhance their teaching skills. If faculty would like more information about the CTLE, please call ext. 4038.

ADMINISTRATION OF THE USCS

Student rating of instruction takes place each semester. Evaluations are mandatory every semester (Fall and Spring) for all first year full-time faculty and every term (Fall, Intersession, Spring, Summer sessions) for all part-time faculty. Faculty receive an email announcing the opening of the Online Course Evaluation System and are asked to rate their objectives. Faculty can also include up to twelve (12) additional questions.

Full-time faculty members wishing to be evaluated in a non-mandatory semester receive instructions via mail to opt into the system.

Faculty members should encourage student to complete their course evaluations. Students complete the evaluations online after receiving an email announcing the opening of the student portion of the Online Course Evaluation System. Any problems, questions, or requests for assistance in this process should be directed to the CTLE at ext. 4038 or at oce@scranton.edu.

5. Other Groups Responsible for the Course Evaluation Process

The CTLE works with a technical advisory group, known as the “Course Evaluation Committee” (CEC), to make recommendations regarding technical aspects of the evaluation tool, its administration, and the reporting of the course evaluation results. In addition, the Academic Support Committee (ASC) of the Faculty Senate is responsible for making recommendations to the full Senate regarding the course evaluations.

6. Interpreting the Course Survey Summary

COURSE INFORMATION

An example of the Course Survey Summary is provided in Appendix A. The number of students responding appears on the summary with the percent responding out of the total number enrolled. As a rule of thumb, if less than 70% of the students responded, there are serious concerns about whether the ratings should be considered representative of the class as a whole.

Results tend to be unstable when fewer than ten students respond. In this case, the reaction of one or two additional students could have a disproportionate influence on the mean rating. Also, when class size is fewer than ten students, ratings tend to be higher. Classes with less than three students responding are excluded from the analysis process.

CONTROLLING FOR INITIAL STUDENT INTEREST

Initial student interest in course content has been found, here as well as at other colleges, to be directly related to satisfaction with the instructor, the class as a whole, and goal attainment. Yet, initial interest in course content is not under the control of the instructor and so is considered a serious biasing factor in student evaluations of courses. This initial interest is measured on the U. of S. Course Survey with the item, "Before enrolling, I really wanted to take this course REGARDLESS of who taught it."

To control for this bias, a standard statistical procedure called regression residual analysis is used (Tabachnick & Fidell, 2007; Appendix B). In this procedure, differences between actual student ratings and predicted ratings based on initial interest are substituted for the original score. This is done across all courses in a mandatory course evaluation semester. Courses in non-mandatory semesters are combined with courses from a mandatory semester (e.g., Intersession 2007 with Spring 2007).

A rough indication of a course's initial student interest is found under Initial Student Interest and Interest Level. These Interest Levels (Very Low, Low, Moderate, High, and Very High) represent five 20% categories: the lowest 20% on initial interest, the next highest 20% group, and on up to the highest 20% group.

EXTREME STUDENT RESPONSES

Many faculty worry that one or two disgruntled students can bias the overall ratings (skew the data). To deal with this realistic concern, a standard transformation is used to reduce such negative skewing (Tabachnick & Fidell, 2007, page 89; Appendix B).

OTHER POTENTIAL BIASES

Another potential biasing factor is class size. This is not a problem at the University since class sizes are all relatively small compared to universities where it has been found to be an important issue.

The effects of other potential biasing factors such as instructor gender are much smaller than the effect of initial student interest and/or are inconsistent. However, if consistent and important biases are identified, they could be controlled for in the same way that initial student interest is.

COURSE COMPARISONS (ABOVE AVERAGE, AVERAGE, BELOW AVERAGE)

Interpreting the results of course evaluations, like the interpretation of any set of data, can be difficult and potentially inaccurate or even biased. To reduce these interpretation problems, standard data analysis procedures have been established in the social/behavioral sciences, including the use of inferential statistics. This procedure compares the results for a sample (e.g., one class) with a comparison value (e.g., the results for all courses in a mandatory semester) and determines whether the sample is significantly above or below this comparison value based on within sample variation, the difference between the sample and the comparison value, and sample size. This is the procedure (specifically the "t-test") that is used to produce the Below Average, Average, and Above Average designations under Comparison Category on the Course Summary sheet.

In the spring of 2007 (a mandatory semester), 14% of the 1246 courses evaluated were classified Below Average on instructor rating while 25% were Above Average (i.e., 61% were Average). For Course rating and the Progress on Objectives composite, the percentages were 10% - 14%, and 7% - 15%, respectively.

PROGRESS ON OBJECTIVES ITEMS

One of the major differences between the USCS and other evaluation forms is the opportunity for faculty to identify their course objectives and to be evaluated on their success in fulfilling them. In the current process, if an instructor identifies objective # 1 as essential or important, the student ratings on that item are compared to other courses for which objective # 1 was essential or important. Please note that a distinction between important and essential is not made here.

In calculating the aggregate for progress on objectives, however, individual course objectives are weighted. That is, in averaging the ratings across the objectives, only those marked important or essential are used in the calculation and objectives chosen as essential are given twice as much weight as those that are chosen as important.

HOW CAN YOU IMPROVE YOUR COURSE RATINGS?

Many faculty members believe that the best way to identify what is right or wrong with a course from the students' perspective, is to consider their comments. That is why comments for each of the items and overall are emphasized on the evaluation survey.

7. Evolution of University Course Evaluations

EARLY HISTORY

A statement of goals and procedures regarding student course evaluations was prepared by the University Senate's Faculty Affairs Committee and presented to the University Senate in the fall of 1973 with subsequent additions made in the spring of 1975. An initial course evaluation was conducted in the fall of 1976. The questionnaire was modified and the revision was presented to the Senate by the Academic Vice President and recommended by that body for implementation. The revised form was then used in the next administration in the fall of 1977.

To determine preferences regarding course evaluation items, an item evaluation survey involving 530 students and 140 faculty was conducted in the Intersession and Spring of 1979. This information was used to again modify the questionnaire prior to its administration in the fall of 1980.

During the same academic year, an Ad Hoc University Senate committee reviewed and revised the statement of the "Goals and Use of Student Evaluations of Teaching." The Senate also recommended a new administration procedure whereby faculty members would administer the questionnaires in one another's classes, however, when this procedure was implemented in the fall of 1983, it was found to be very cumbersome. As a result, in the spring of 1984 the Senate recommended that the new procedure be abandoned and the old method of using student administrators be reinstated. On November 9, 1984, the University Senate recommended the adoption of the rotating semester evaluation schedule which determines mandatory semesters.

In the 1985-86 academic year, an Ad Hoc University Senate Committee reviewed evaluation forms that used national normed comparisons and recommended Kansas State University's IDEA form. This questionnaire was adopted in the fall of 1986. In 1991 a number of changes were made in the form and data processing (including in-house analysis) because 1. Scranton ratings were almost identical to the national comparisons used by Kansas State, 2. there were problems with Kansas State's individual items and 3. the comparison process for Progress on Objectives was flawed.

CURRENT UNIVERSITY OF SCRANTON COURSE SURVEY FORM

The current in-house form and process for student evaluation of courses is a modified version of the IDEA Form and process developed at Kansas State University (Cashin & Sixbury, 1992) and the methods items from the Student Evaluation of Educational Quality (SEEQ) form developed by Marsh (1987).

INDIVIDUAL ITEMS CHANGES

The initial student interest item was changed in 1991 to "Before enrolling, "I really wanted to take this course REGARDLESS of who taught it" because the original IDEA item was contaminated by instructor interest (see section 6 and Prave & Baril, 1993). Kansas State made the same change later. In addition, changes were made to the Methods items based on an analysis of the results of the spring and fall 1989 administration. Finally, the "Developed mathematical/quantitative skills" objective was added.

In the spring of 2007, the form was extensively revised based on problems arising from the introduction of on-line course evaluations (see On-Line Course Evaluations section below).

ADDITION OF ABOVE AND BELOW AVERAGE CATEGORIES

In the spring of 1993, the Course Evaluation Committee surveyed deans and members of the Board on Rank and Tenure about their interpretation of the forms. The results indicated that there was considerable variability in how the ratings are interpreted. The normative data from 1989 were analyzed using four statistical tests to evaluate the utility of inferential statistical analyses which is the standard in the behavioral sciences in reporting survey data. As a result of the survey and these analyses, the Faculty Senate accepted the recommendation to report simply above average, average, and below average categories and discontinue reporting percentiles, means and standard deviations. The changes were motivated by their concern about over-interpretation of the data when there might not be meaningful differences among individual faculty members. They were also concerned with unfairly placing an individual in a category that would be to the faculty member's disadvantage. This process was implemented in the spring of 1994 (See section 6 and Changes in the Statistical Analysis section).

COMPARISON GROUP CHANGES

In 2004, the comparison year was updated from 1989 to 2003 because of changes in overall ratings over the years. Over time, a larger and larger proportion of

courses were falling into the above average category. It was likely that the majority of courses would be above average at some point. To avoid this, the comparison group norms became based on the data from the previous mandatory evaluation year (2003, 2005, etc.) as recommended by the Faculty Senate on December 5, 2003. The effect of this change was that the percentages in the categories were almost the same as they were in 1989.

In the spring of 2007, the comparison group process was replaced by a much more sophisticated and valid system involving regression analysis (see Changes in the Statistical Analysis section below).

ON-LINE COURSE EVALUATIONS

In 2004, the Faculty Senate concurred with the recommendation of the CEC that an online course evaluation process be developed and tested in the Intersession of 2005. This recommendation followed an extensive evaluation of the current status of online evaluations across the country and particularly the online implementations at Wellesley, Bates and Yale where a high response rate was obtained through the use of restricted grade access. The test was successful and the new system was implemented in the spring of 2005 (Faculty Senate recommendations of October 8, 2004, and February 11, 2005).

However, an extensive analysis of the CE results for the spring and fall 2005 indicated that the online system seemed to decrease student attention to the CE items. As a result, the Faculty Senate recommended that the evaluation form be shorted including the modification of the methods items based on the Marsh's (1987) SEEQ form (Senate recommendations of October 13, 2006, and February 9, 2007). The Objectives items were also changed based on a faculty poll conducted by the CEC in the fall of 2004. The new form was piloted in the Intersession of 2007 and implemented in the spring of 2007.

CHANGES IN THE STATISTICAL ANALYSIS

In 2007, the CTLE and System Software Resources (SSR) recommended that the programming for the Below and Above Average comparative analysis needed to be simplified. As a result an intensive evaluation of the whole inferential statistical analysis was conducted. This led to the implementation of a much simpler and more valid analysis that was reprogrammed by SSR. This analysis retained the basic inferential statistical test (the t-test) but used a standard transformation to control for the inordinate effect of a few highly negative ratings. More importantly, instead of the crude system of using three normative groups to control for initial student interest, a more valid regression residual analysis was implemented.

8. Questions about the Course Evaluation Ratings

I. What do student course evaluations really measure?

In general, they measure student satisfaction with the course and instructor, and opinions about course processes and outcomes. Student satisfaction is important in its own right. But opinions about process and outcome need to be evaluated against some standard, such as an independent measure of amount learned. Fortunately, many studies have shown that student ratings are moderately related to both amount learned in the class, instructor self-ratings, alumni ratings, and student motivation.

II. Are evaluations by peers more useful than evaluations by students?

There is a divergence of opinion regarding the usefulness of peer evaluation for faculty teaching effectiveness. There seems to be some evidence that classroom visitation by an instructor's peers may not be accurate in identifying teaching effectiveness. As Marsh (1987) states in his review of the literature, "these findings... clearly indicate the use of peer evaluations of university teaching for personnel decisions is unwarranted" (p.296). Nonetheless, many colleges and universities use peer evaluations for rank and tenure decisions.

III. Are student ratings biased by lenient grading?

Grades are weakly to moderately related to student ratings. This is to be expected given that student evaluations are related to achievement and that grades reflect achievement. In addition, a number of studies have shown that the relationship between grades and ratings is not the result of satisfaction or dissatisfaction with a grade. However, this issue continues to be controversial.

IV. Are courses with heavy workloads penalized on student evaluations?

Not in general. Most studies find a weak positive relationship between workload and student ratings: that is the higher the workload, the higher the ratings.

V. Does class size affect student ratings?

Student ratings in small classes tend to be slightly higher. However, the use of inferential statistics and initial student interest categories minimize the effect. Our data demonstrates that the percent of faculty falling into each

category is approximately the same for all class size. The biggest difference at the University of Scranton is between courses with fewer than ten students and all others. Even this difference is very small. At the other extreme, there is virtually no difference among classes ranging from 10 to 60 students, the effective range of class size at the University.

VI. Do student characteristics--such as GPA, age, and gender--effect the evaluations?

In general, the current evidence suggests not. The most important student characteristic is initial interest in the course content, particularly for overall course rating and goal attainment. This is why the effect of initial student interest is controlled for in our analysis.

VII. Would it be better simply to use departmental comparisons?

The problem is that usually there are too few courses and instructors in a department to make these comparisons stable and generalizable. Also, there are still likely to be substantial differences in initial student interest from course to course even within a department which a departmental comparison does not control. In fact, the net effect of basing comparisons on initial student motivation and allowing the individual instructor to designate the importance of course objectives is to get a comparison that is more authentically "local" than departmental comparisons.

VIII. Are course objectives manipulated to misrepresent course goals in the interest of improving the ratings?

The ratings for each chosen goal are only compared to the ratings from courses with the same goal (identified as essential or important). The results for the chosen goals, therefore, are not distorted. However, the failure of the faculty member to identify legitimate course goals as essential or important will result in distortion of ratings. The departments are encouraged to standardize their objectives for multi-sections.

X. Would percentile rankings provide a better indication of your teaching performance?

While percentile rankings provide comparative data, their use can be subjective and can lead to over interpretation of differences in teaching when for all practical purposes those differences might not be meaningful.

XI. Are student's written comments more valuable than statistical summaries of rating scales?

Many faculty value student comments when considering changes in their courses; however, the statistical summaries are particularly useful and important for faculty and administrators in obtaining an overall perspective on student perceptions of individual classes: the first step in deciding about possible changes.

XII. Do faculty characteristics such as age, gender and rank effect the evaluations?

According to Erickson (1984), "senior" professors start out with a slight advantage over "apprentice" professors. Age and years of teaching experience, however, are not, in general, correlated with student ratings (Cashin, 1995). Although no global relationship between gender and student ratings exists, Cashin (1995) indicated that in a few studies male students rated male professors higher, and female students rated female professors higher.

XIII. Should means be provided as they used to be?

No, for at least two reasons. First, means like percentile ranks are often over interpreted because of their apparent precision. This is one of the major problems we tried to avoid with the present system. Second, a simple mean is an inappropriate measure of central tendency for student ratings. This is because ratings tend to be highly skewed; it is common to have one or two extreme ratings which will have an undue impact on the mean. This is why a transformation for skew is used in analyzing the data from the USCS. In addition, the control for initial student interest also precludes the use of simple means.

XIV. Why is it that sometimes what seem to be very high student ratings are designated as either average or even below average?

The research literature has clearly shown that student course evaluations suffer from a very substantial leniency bias or the Lake Woebegone effect; everybody is above average. When the mid-point on the student rating scale is designated "average," the mean ratings are almost always well above that point when logically that should not be the case. Also, courses that are high on initial student interest are most likely to show this "inconsistency" between relative and absolute rating, because the comparison values tend

to be very high for these classes, as they should be. But even in these instances, the interpretation of the Above and Below Average designations is the same as always since the ratings are adjusted for initial student interest.

9. Further Reading

The literature on student ratings of instruction is vast. There are thousands of research articles published on all aspects related to this topic. As might be expected, the number of articles and books making recommendations on the topic is even larger. For faculty interested in further reading on this subject, a brief bibliography on student rating is provided. References within the works cited here may be consulted for more in-depth study.

Research on student ratings of instruction, including treatment of such questions as reliability of ratings, effect of class size, relationships between ratings and grades, and a myriad of other such matters, is summarized in a number of places. A very convenient, readable, six-page summary of research on student ratings is provided by Cashin (1988). Centra (1975) offers an important discussion of the unreliability of peer evaluations. At present, the most complete technical summary of this body of research on course evaluation is provided by Marsh (1987; Marsh & Roche, 1997).

There are several books on a variety of topics related to evaluation of faculty at the college level. These sources cover student ratings of instruction but also other methods of evaluating teaching effectiveness, as well as methods for evaluating scholarship and service. In treating student ratings, the sources provide useful summaries of contemporary practices at various institutions and discussions of the pros and cons of using student ratings. Three such sources are Miller (1987), especially Chapter 3 ("Evaluating Teaching: The Role of Student Ratings"); Millman (1981), especially Chapter 8 ("Student Ratings of Instruction" by L.M. Aleamoni); and Centra (1979), especially Chapter 2 ("Uses and Limitations of Student Ratings").

The reader interested in further study of student ratings would be well served to start with at least one of the summaries of research cited above and at least one of the books. All of these works are available in the Center for Teaching & Learning Excellence.

Bibliography

- Baril, G.L., Sebastianelli, R., Cannon, J.T. (1996). Using inferential statistics to improve the integration of course evaluations. In. R.J. Ebert and L. Foranz (Eds.), *1996 Proceedings: Decision Sciences Institute* (pp. 248-50). Decision Sciences Institute.
- Cashin, W.E. (1995). *Student ratings of teaching: The research revisited*. (IDEA Paper No.32). Manhattan, KS: Kansas State University, Center for Faculty Evaluation & Development.
- Cashin, W.E. (1988). *Student ratings of teaching: A summary of the research (IDEA Paper No.20)*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation & Development.
- Cashin, W.E., & Sixbury, B.M. (1992). *Description of database for the IDEA diagnostic form. (IDEA Technical Report No. 7)*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Centra, J.A. (1975). Colleagues as raters of classroom instruction. *Journal of Higher Education*, 46, 327-337. Centra, J.A. (1979). *Determining faculty effectiveness: Assessing teaching, research, and service for personnel decisions and improvement*. San Francisco, CA: Jossey-Bass.
- Erickson, S.C. (1984). *The essence of good teaching*. San Francisco, CA: Jossey- Bass.
- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Miller, R.I. (1987). *Evaluating faculty for promotion and tenure*. San Francisco, CA: Jossey-Bass.
- Millman, J. (Ed.) (1981). *Handbook of teacher evaluation*. Beverly Hill, CA: Sage Publications/National Council on Measurement in Education.
- Prave, R.S., & Baril, G.L. (1993). Instructor ratings: Controlling for bias from Initial student interest. *Journal of Education in Business*, 68, 362-366.

Tabachnick, B. S. & Fidell, L. S. (2007) *Using Multivariate Statistics* (5th ed.). New York, NY:
Pearson Education, Inc.

APPENDIX A

SAMPLE COURSE SURVEY FORM

Instructor: **John Doe**

Course: **PTRC 210-1; Statistics in Management Science (20139)**

Click the button which indicates your opinion:

1	2	3	4	5	NR
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

INITIAL INTEREST IN THE COURSE

RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

1. Strongly Disagree 2. Disagree 3. Neutral 4. Agree 5. Strongly Agree NR. No Response

1. Before enrolling, I really wanted to take this course REGARDLESS of who taught it.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

INSTRUCTIONAL METHODS: The Instructor...

RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

1. Strongly Disagree 2. Disagree 3. Neutral 4. Agree 5. Strongly Agree NR. No Response

1. Was enthusiastic about teaching the class.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

2. Made students feel welcome in seeking help/advice.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

3. Used evaluation methods which were fair and appropriate.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

4. Provided clear and well organized class materials/presentations.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

5. Provided context for course material.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

6. Encouraged students to participate in class.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

7. Assigned readings/texts that were appropriate.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please Comment:

Instructor: **John Doe**

Course: **PTRC 210-1; Statistics in Management Science (20139)**

PROGRESS ON COURSE Objectives: I....

RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

1. Strongly Disagree 2. Disagree 3. Neutral 4. Agree 5. Strongly Agree NR. No Response

	1	2	3	4	5	NR
1. Gained factual knowledge (terminology, classifications, methods, trends).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Learned fundamental principles, generalizations, or theories.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Learned to apply course material to improve thinking, problem-solving or decision making.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Developed specific skills needed by professionals in this field.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Developed creative capacities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Developed a sense of personal responsibility (self-reliance, self-discipline).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Gained a broader understanding and appreciation of intellectual-cultural activity (music, science, literature).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Developed skill in expressing myself orally or in writing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Discovered the implications of the course for understanding myself (interests, talents, values).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Developed mathematical/quantitative skills.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Acquired skills in working with others as a member of a team.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Gained an understanding of different cultures and groups.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Instructor: **John Doe**

Course: **PTRC 210-1; Statistics in Management Science (20139)**

OVERALL RATINGS

RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

1. Strongly Disagree 2. Disagree 3. Neutral 4. Agree 5. Strongly Agree NR. No Response

1. Overall, I rate this instructor an excellent teacher.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Please Comment:						
<input type="text"/>						
2. Overall, I rate this course as excellent.	1	2	3	4	5	NR
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Please Comment:						
<input type="text"/>						

WORKLOAD (using the scale provided)

Average number of **hours per week** I spent outside of class on work for this course:

0 or 1 2 or 3 4 or 5 6 or 7 8 or more NR

Please Comment:

Instructor: **John Doe**

Course: **PTRC 210-1; Statistics in Management Science** (20139)

EXPECTED GRADE

What is your expected grade in this class?

GENERAL COMMENTS

1. What would you suggest to improve this course?

2. What aspects were good and should not be changed?

Instructor: **John Doe**

Course: **PTRC 210-1; Statistics in Management Science** (20139)

Adapted from The IDEA Center's Student Rating Form: (C) 1998 The IDEA Center, Inc.
Permission for use granted by The IDEA Center, Inc.
All rights reserved by the IDEA Center, Inc.

SAMPLE COURSE SURVEY SUMMARY AND COUNTS

8/21/2019

Survey Summary

University of Scranton

Fall [REDACTED] Course Summary for

CRN: [REDACTED]

Instructor: [REDACTED] Course: BUS [REDACTED] Sec. 5; Students Responding: 16 out of 17 (94%) The questions below use the following scale: 1=Strongly Disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly Agree; NR=No Response
--

Interpreting the Course Survey Summary

INSTRUCTOR							
Based on:	Comparison Level *	1	2	3	4	5	NR
Overall, I rate this instructor an excellent teacher. Click here for comments	Above Average	0	0	0	0	16	0
COURSE							
Based on:	Comparison Level	1	2	3	4	5	NR
Overall, I rate this course as excellent. Click here for comments	Above Average	0	0	0	4	12	0
PROGRESS on OBJECTIVES	Comparison Level	1	2	3	4	5	NR
Composite of course goals identified as important(i) or essential(e):	Average						
Gained factual knowledge (terminology, classifications, methods, trends).	Average	0	0	0	6	10	0
Learned fundamental principles, generalizations, or theories.	Above Average	0	0	0	4	12	0
Learned to apply course material to improve thinking, problem-solving or decision making.	Average	0	0	1	4	11	0
Developed skill in expressing myself orally or in writing.	Average	0	0	1	3	12	0
Discovered the implications of the course for understanding myself (interests, talents, values).	Average	0	0	3	4	8	1
INSTRUCTIONAL METHODS	Comparison Level	1	2	3	4	5	NR
Was enthusiastic about teaching the class. Click here for comments	Above Average	0	0	0	1	15	0
Made students feel welcome in seeking help/advice. Click here for comments	Above Average	0	0	0	1	15	0
Used evaluation methods which were fair and appropriate. Click here for comments	Above Average	0	0	0	3	13	0
Provided clear and well organized class materials/presentations. Click here for comments	Average	0	0	1	4	11	0
Provided context for course material. Click here for comments	Above Average	0	0	0	4	12	0
Encouraged students to participate in class. Click here for comments	Above Average	0	0	0	2	14	0
Assigned readings/texts that were appropriate. Click here for comments	Average	1	0	0	3	12	0
INITIAL INTEREST IN THE COURSE	Interest Level	1	2	3	4	5	NR
Before enrolling, I really wanted to take this course REGARDLESS of who taught it.	Moderate	0	1	4	4	5	2

WORKLOAD											Comparison Level		0-1	2-3	4-5	6-7	8+	NR
Based on:																		
Average number of hours per week I spent outside of class on work for this course: Click here for comments											Below Average	1	12	3	0	0	0	
EXPECTED COURSE GRADE																		
A	A-	B+	B	B-	C+	C	C-	D+	D	F	PF/SU	NR	Mean	StDev				
8	6	0	1	1	0	0	0	0	0	0	0	0	3.729	.389				
GENERAL COMMENTS																		
1. What would you suggest to improve this course? 2. What aspects were good and should not be changed? Click here for comments																		
ALL COMMENTS																		
Click here for ALL comments																		

* The Comparison Levels are determined by a statistical analysis of all courses controlling for the effect of Initial Student Interest (see [Interpreting the Course Survey Summary](#)).

[RETURN](#)

[Printable Version](#)

APPENDIX B

COURSE EVALUATION COMPARISON ANALYSIS

The purpose of the course evaluation comparison analysis (see Interpreting the Course Survey Summary in the University of Scranton **Guide** to the Student Course Evaluation Survey, Summer 2007) is to determine whether the student ratings of individual courses are "significantly" (not likely to be due to chance) below or above the average. Such significance testing is a standard (in fact, near universal) practice in the social/behavioral sciences (see Glass & Hopkins, 1996).

However, before this process can be done, the data need to be transformed to reduce the high negative skew that is typical of course evaluations (i.e., negative ratings by a few students often have an inordinate impact on the class averages) (see Marsh & Roche, 1997). This too is a common practice when analyzing survey data in the social/behavioral sciences (see Glass & Hopkins, 1996; Tabachnick & Fidell, 2007).

Another problem common to course evaluations is rating bias based on initial student interest in the content of course (ISI) (see Marsh & Roche, 1997). The most sophisticated, yet simplest way to eliminate this problem is to use "regression residuals", differences between the actual mean rating and the predicted rating based on ISI (see Kerlinger & Pedhazur, 1973, pgs. 415-418 for a general discussion of residuals with examples; Glass & Hopkins, 1996, pgs. 159, 167-170 for a discussion of residuals and partial correlations; Cohen & Cohen, 1983, pgs. 213-23, for a discussion of using residuals in controlling for the effect of pre-scores and in partial correlations).

Specific steps in determining the Below Average, Average, and Below Average categories:

1. Reduce the negative skew of the individual ratings by reversing the scores (1→5, 2→4, 3→3, 4→2, 5→1), doing a log 10 transformation, and then re-reversing the scores as recommended by Tabachnick & Fidell (2007, pg. 89). Since this always produces the same values, this whole transformation does not need to be done. Instead, the scores can simply be recoded directly to 1→4.30, 2→4.40, 3→4.52, 4→4.70, 5→5.00.
2. Calculate the means, standard deviation (using n-1 for estimated population values), and number rating for each item for each course.
3. Calculate the standard deviation of these means (n-1).
4. Calculate the mean of these means.
5. Calculate the standard scores for all these means (Mean - the Mean of the Means divided by the standard deviation of the means).

6. Calculate the Pearson correlation between the item means (e.g., instructor rating means) and the initial student interest (ISI) means (the sum of the cross-products of the standard score means for each item and the ISI standard scores divided by the number of means, i.e. courses, minus 1).
7. Calculate the predicted values for the item means based on the ISI means (the standard score for each item mean times the correlation times the standard deviation of the item means plus the mean of the item means).
8. Calculate the difference between the actual mean and the predicted value for each item mean. This difference is called the residual.
9. These residuals represent the original values with the effect of initial student interest eliminated and are used in place of the original values.
10. The mean of the residuals will always be zero since the sum of the differences between the original scores and the predicted scores will always be zero.
11. Calculate the standard deviation of the residuals (using $n-1$).
12. Calculate the t-value for each of these residuals by subtracting the mean of the residuals from the residual (since the residuals mean is zero, you do not need to actually do a subtraction) and dividing by the quotient of the standard deviation ($n-1$) of the individual course scores divided by the square root of the number of students rating in each class (see Glass & Hopkins, 1996, pgs. 270-1, for a discussion of single-sample, two-tailed t-tests).
13. Compare each t-value with the "critical" t-value (which is the t-value that has a probability of .025, i.e., the "statistically significant" t-value) from a t-table for each class size minus 1 (the degrees of freedom).
14. T-values at or above the critical value are considered significantly "Below Average" or "Above Average" depending on whether the course ratings are below or above the average residual mean of 0. T-values below the critical value are not statistically significant and therefore are referred to as "Average."
15. The Progress on Objectives Composite comparison category is determined by a weighted mean of the individual objectives t-values (the weighting is based on instructor ratings of the importance for each objective).

- Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd.ed.). Hillsdale, New Jersey: Lawrence Erlbaum.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology* (3rd ed.). Boston: Allyn and Bacon.
- Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart & Winston.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Tabachnick, B. S. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). New York, NY: Pearson Education