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.

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. Alsle, studiel92 reW*n06 156

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:

A major goal of the University community at large is to achieve and maintain the highest possible quality in the educational process.

their validity.

The Mission of the Center for Teaching & Learning Excellence is to promote a

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.

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

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.

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).

The initial student interest item ×ÁÓ ÃÈÁÎÇÅÄ ÉÎ ρωωρ ÔÏ Ó"ÅÆÏÒÅ ÅÎÒÏÌÌÉÎÇ Ó) really ×ÁÎÔÅÄ ÔÏ ÔÁËÅ ÔÈÉÓ ÃÏÕÒÓÅ 2%'!2$,%33 ÏÆ ×ÈÏ ÔÁÕÇÈÔ ÉÔ ÂÅÃÁÕÓÅ ÔÈÅ 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).

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).

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).

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.

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.

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.

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.

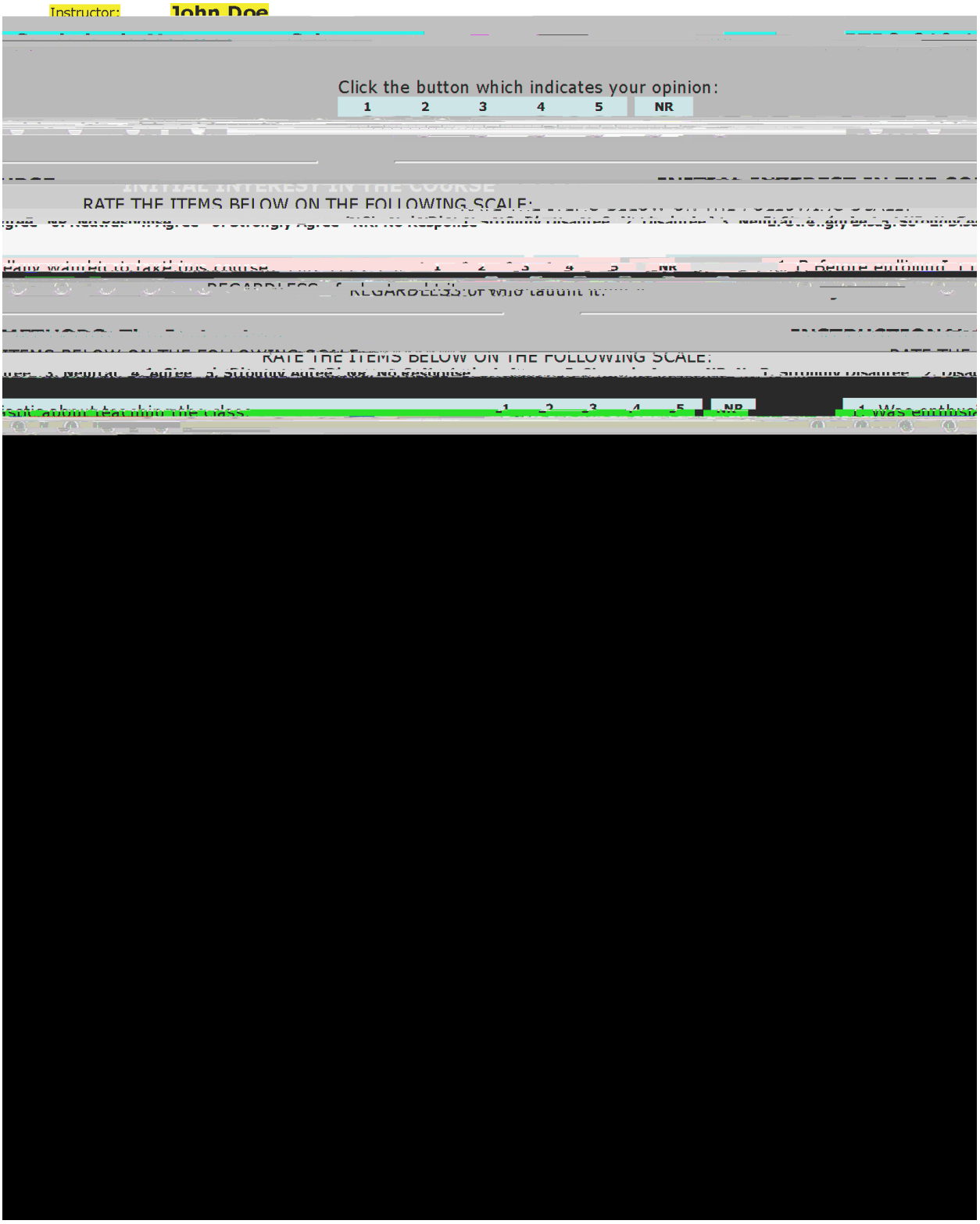
Many faculty value student comments when considering changes in their courses; however, the statistical summaries are particularly useful and important for faculty and administrators in obtaining an overall

to be very

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.), (pp. 248-50). Decision Sciences Institute.

Cashin, W.E. (1995). (IDEA Paper No.32). Manhattan, KS: Kansas State University, Center for Faculty Evaluation & Development.

Cashin, W.E. (1988).


Instructor:
John Doe
Click the button which indicates your opinion:
1 2 3 4 5 NR
RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:
RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

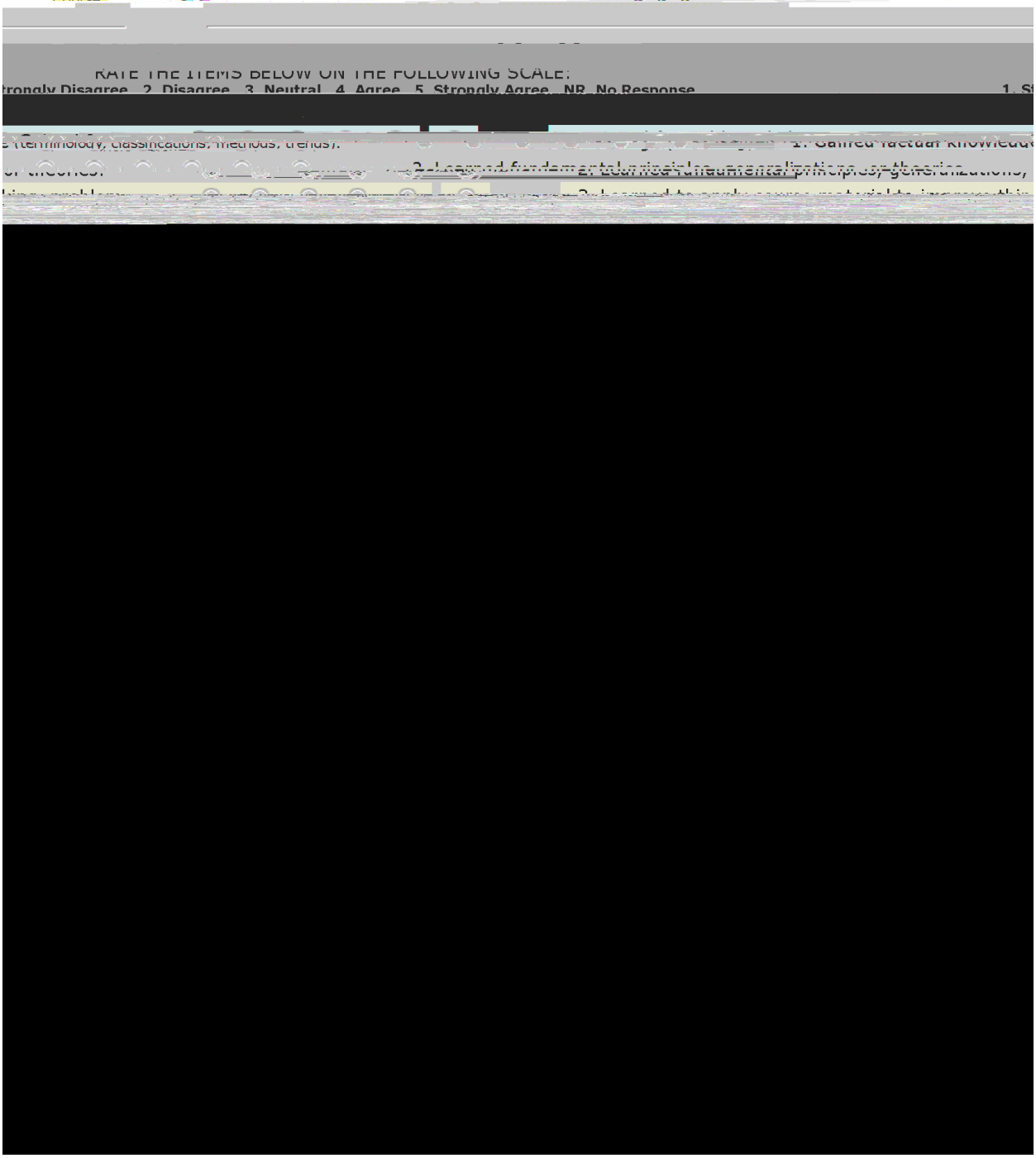
Please Comment:

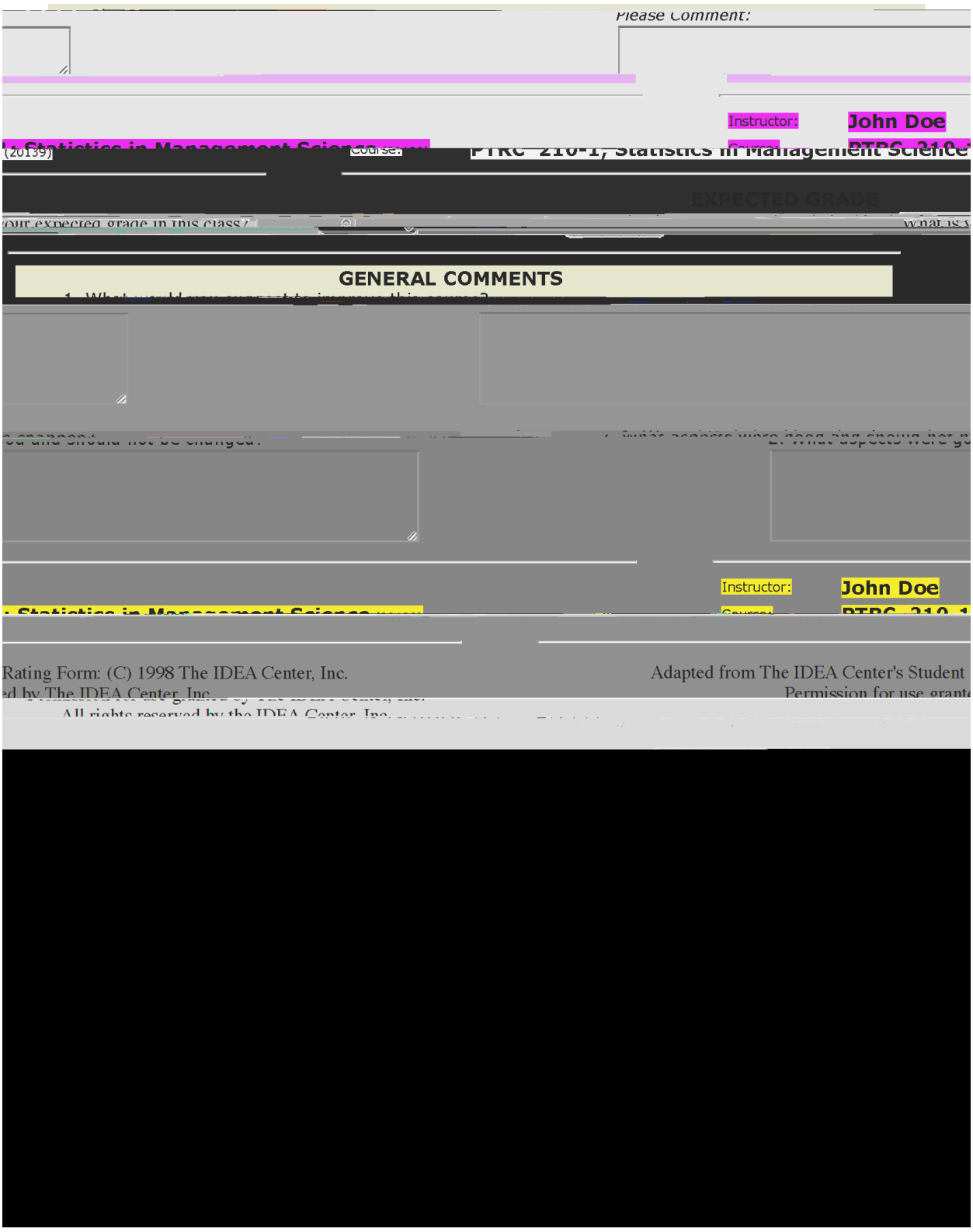
Instructor: **John Doe**

Course: **PTPC 210-1: Statistics in Management Science** (20120)

RATE THE ITEMS BELOW ON THE FOLLOWING SCALE:

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

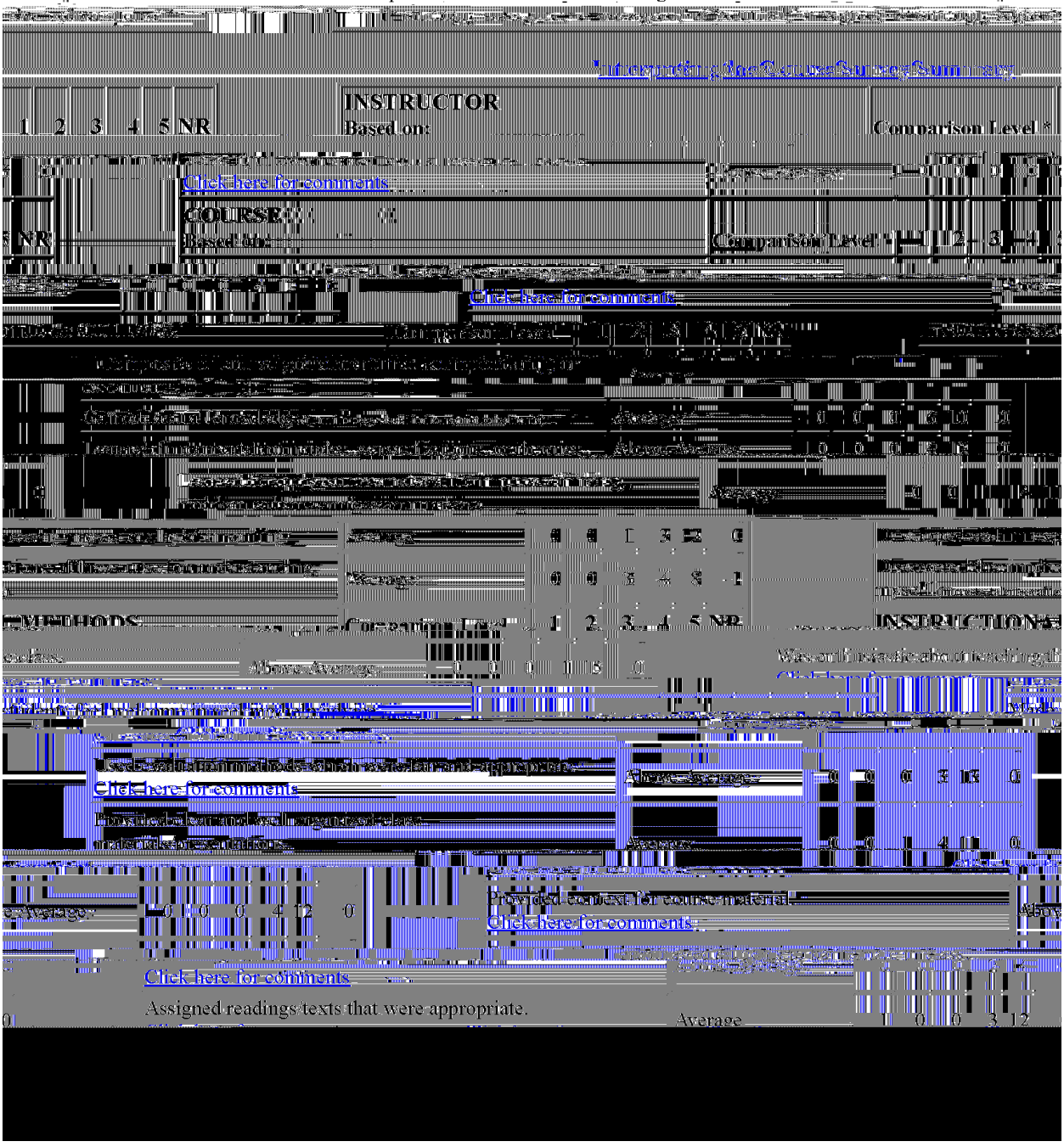



Please Comment:
Instructor:
John Doe
(20139)
Course:
PTRC 210-1, Statistics in Management Science
EXPECTED GRADE
GENERAL COMMENTS
Instructor:
John Doe
Rating Form: (C) 1998 The IDEA Center, Inc.
Adapted from The IDEA Center's Student

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:



KLOAD Comparison Level 0-1 1-3 3-5 5-7 7+ NR WOR

Average number of hours per week I spent outside of class on

EXPECTED COURSE GRADE

A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F	PASS	NR	Mean	SDev
8	6	0	1	1	0	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 of this course do you think should not be changed?

Click here for comments

ALL COMMENTS

Click here for ALL comments

Summary

Initial Student Interest (see Interpreting the Course Survey

RETURN Printable Version

The purpose of the course evaluation comparison analysis (see Interpreting the Course Survey Summary in the University of Scranton 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-

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.

Cohen, J., & Cohen, P. (1983).
(2nd.ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Glass, G. V., &