Four Thousand Reports Three Ways

Stephanie R. Thompson, Rochester Institute of Technology, Rochester, NY

ABSTRACT

How do you go about generating over four thousand PDF reports in up to three different versions? When a large, southern research university decided to add up to five optional questions per class section and up to five more questions at the prefix level to their core set of fifteen questions on the student evaluation of faculty survey, it seemed like a project that would never be completed. If the additional questions weren't enough, the reports were being revamped at the same time to improve their appearance for delivery on the web. Each report had a tabular section and two customized box and whisker plots. Thanks to ODBC / SAS Access, PROC SQL, macro, DATA Step programming, PROC GPLOT, goptions, and ODS it all came together. This paper summarizes how each SAS® component was used and contributed to the completion of the project.

INTRODUCTION

Student evaluations of faculty are nothing new on a college campus. Most of the time the questions are set in advance and the students fill out paper or electronic forms. When one large research university decided to update its old survey, they wanted to make the instrument more flexible for faculty and departments and easier for students for fill out. The old instrument had thirty-four questions with selected questions grouped into five scales. It was cumbersome to fill out on the original paper form and not much easier on the web.

The new instrument, designed by a subcommittee of the faculty senate, consisted of nine core questions and six questions for categorization (the core survey). The new fifteen question survey would be administered for all sections of each course. There are over four thousand sections each semester at this university. To complicate matters even more, each course prefix (e.g., ENGL, PSYC, and MATH) could add up to five questions entered by the department chair for that group of courses. Individual professors could add up to five questions of their own for each course section that they taught to provide more specific feedback. Each student signing into the web could see a survey ranging from fifteen to twenty five questions.

The aggregated results of the core survey were to be made publically available on the web if class size or participation met the reporting threshold. Results were presented in tabular form and graphically. The results of the core survey and any prefix-based questions were to be made available to deans, department chairs, and professors. The results of the section-based questions (which include any prefix-based questions) were only to be made available to the professor teaching the section. Therefore, as many as twelve thousand different reports may need to be generated each semester. In addition to the new question flexibility, the format of the report was being changed for better web readability, the addition of color, and to track NA responses which were a new response choice on the redesigned survey.

GETTING THE DATA

To make processing easier, macro variables were set up in the beginning of the program. These macro variables contained the full name of the term to process, parsed sections of the term name such as year and semester, titles, footnotes, as well as the name of the directory to store the completed reports. Having all of this information up front made it easy to find and eliminated having to search through the program each term to find the values that needed to be changed.

The data from the online survey was stored on an Information Technology (IT) department server in a Microsoft SQL Server database. Since SAS does not have a SAS/ACCESS® SQL Server product for the Microsoft Windows operating system, SAS/ACCESS to Open Database Connectivity (ODBC) was used to create the connection needed on the PC. Once the ODBC connection was established, PROC SQL was used to bring the raw survey data into SAS. Five separate tables were used to hold the data in the original database. Tables contained student respondent data, responses, question details, question categories, and response code details. An example of the SQL code used is shown below:

```
proc sql;
    connect to odbc(dsn=&mdsn uid=&usr pwd=&pwd);
    create table questcat as
    select question_category_id, question_category,
```

```
question_category_sequence, question_category_instruction
from connection to odbc (select * from sirs_prod.dbo.question_category)
where question_category_id ne 3;
disconnect from odbc;
quit;
```

In the above example, the question categories are being pulled. A question_category_id of "3" was for open ended questions which were not part of these reports. Another item to note is that the name of the ODBC connection, database id, and database password are stored in macro variables for security as well as to make the programming easier (the IT department assigned a very creative password and username that is prone to mistyping).

SUMMARIZING THE DATA

Now that the data were in SAS, they needed to be combined, analyzed and summarized for reporting. Before this could be done some re-coding was necessary. The database design recorded a zero for a response of "N/A." Including the zeroes would artificially deflate the scores since a response of "1" was "Strongly Disagree." Allowing an "N/A" response was new to this iteration of the reporting so validating what was stored for that response was an important step in ensuring the values would be reported correctly. All response values of zero were recoded to missing.

Once the recoding was done, the data were combined to make a single file for creating statistics. After some testing, the sort – merge approach turned out to run faster than PROC SQL. In addition to just putting the files together, new sections flagged as "COMBINED" were created to generate statistics across instructors for sections that were taught by more than one instructor.

Since the "N/A" response option was new, it was decided to include a count of scaled responses and a count of "N/A" responses to assist in evaluating the mean and standard deviation of the Likert scale responses. If a mean for a question was reported as a "5" for strongly agree, it would be useful to know if one or forty students in a fifty student section responded to the question.

Each section was analyzed and reported by itself. In addition, sections of the same course were combined and analyzed, other sections of the same level (lower, upper, or graduate) with the same prefix were aggregated and reported as well as other courses of the same level within the same college. This was to provide a comparison of each section to somewhat similar courses in the same department and college. It was intended to serve as relative measure of faculty performance (e.g., is the faculty member who taught ENGL1010001 perceived by students as more effective than other lower-level English professors).

All of the statistics were calculated with PROC UNIVARIATE. The procedure was run each time for the various combinations of statistics that were needed. Each time a data set was created that was eventually combined into a single data set used for reporting. Once the final reporting dataset was generated, a data set containing a distinct list of sections and instructors was created to speed up reporting. The observation number was copied to the variable "order" to generate a numeric counter within the data set.

TABULAR REPORTS

Before the reports could be generated, the directory for the reports needed to be created on the PC and the default directory needed to be established. The X command was used to submit a command to the PC's operating system without leaving the SAS session. The code to do this is below:

```
** ^DIRTREE^ -- DCL statement to create directory for current semester output **;
x "mkdir &outroot.\&term.";

*move to that dir;
x "cd &outroot.\&term.";
```

Setting up a new directory each semester prevented old reports form being overwritten. The one drawback to only using the code above is that an "exit" command in the DOS window must be issued for the program to continue. Using the noxwait and noxsync options eliminate the need for the exit command.

Now that the directory is created, a macro loop was set up to iterate for each report that needs to be created. Additional macro variables were created to be used in the reports as shown next:

```
proc sql noprint;
   select count(*)
      into :numsect
      from seclist;
quit;
%put ******** Number of sections &numsect;
* create loop;
%macro reports;
%do i = 1 %to &numsect;
* create macro variables to use in titles etc.;
proc sql noprint;
   select sect_id, instruct
      into :sid, :iuuid
      from seclist
where order = &i;
quit;
* create macro variables to use in PDF name;
data null;
set seclist;
call symput('iuuid_pdf', cats(instruct));
where order = &i;
run;
* data for current report;
proc sql;
create table temp as
select *
from corebase5
where sect_id = "&sid" and instruct = "&iuuid"
order by question_sequence;
quit;
```

The sections of the report showing the statistics and counts of responses were generated with PROC REPORT. In order to wrap the text of the questions in the report and fit everything on a single page in portrait format, the FLOW and WIDTH= options were used on the DEFINE statement. This was one the easiest sections to code. PROC REPORT created almost all of what was needed with very little additional coding or use of options.

GRAPHS

Generating the graphs was another story. Creating the box and whisker plots from scratch was not desirable. In order to get the code to create the base of each graph (prefix and college representations), the data set used for reporting was saved as a permanent dataset. The data was bought into SAS Enterprise Guide®. Changes could be made interactively to the appearance of the graphs until the desired look was achieved. Since Enterprise Guide generated the basic code, it was simply copied into the report program. Superfluous code generated by Enterprise Guide was subsequently deleted. The only real editing necessary was to make sure the data set references were correct for the main program.

The next step was to add a point for each question on each graph that represented where the particular section being reported fell. One would think this would be simple and to many it is. However, this turned into a two day effort. Replicating the original report program's approach written in Adobe PostScript was not feasible. The solution turned out to be very simple and was the addition of a small amount of code. The following lines were added to the goptions statement to create the point for the section data:

```
SYMBOL2
CO=BLUE
VALUE=dot
HEIGHT=0.5
```

Now that the symbol for the section was defined, it needed to be incorporated into PROC GPLOT. The code below, using the overlay option, puts the section point on the box and whisker plot that it creates.

```
PROC GPLOT DATA=tograph NOCACHE;
PLOT sec_mean*Quest plotmean*quest/ overlay
    VAXIS=AXIS1
    HAXIS=AXIS2;
where crs_div = "&cdiv" and coll = "&college";
run;
```

Only the nine core questions were graphed. The text of each question was paraphrased on the graphs due to space limitations. It would not have made sense to plot the prefix-based or section-based questions since there is not much of a comparison that can be made. No comparison, actually, in the case of section-based questions.

ODS

All of the reports were generated as PDF files. Since the reports were no longer printed and distributed, it seemed like a good opportunity to add some color. The style "sete" was created from the "BarrettsBlue" template provided by SAS. The style provided continuity between the tabular reports and graphs. In previous reporting terms, the PostScript files were provided to the IT department and they converted them to PDF files. It seemed a little inefficient to convert the files when SAS could create a PDF initially. The original program was run on a DEC Alpha VAX that was never upgraded past SAS 6.2. Now that the SAS program created the PDF, the IT department could work on determining which reports would be publically viewable and posting the results through the various portal channels.

To protect student privacy, only reports that had a total enrollment of five or more students were published by default. If a course had 4 or fewer students registered, at least two students needed to respond to the survey to be published. A text-based index file was created for the IT department to flag which reports could be made public and which could not.

```
data _null_;
set sirs2;
show = 'Y';
if actual_enrollment le 4 then do;
   if avgresp eq 1 then show = 'N';
end:
* set pageref;
pageref=compress("&outroot.&term.\20&yr.&sem._"||sect_id||"_"||instruct||".pdf",'
if length(crs_pref) = 4 then do;
   test = substr(pageref, 35, length(instruct));
   if instruct eq test then do;
         file "&outroot.&term.\index.txt";
         put @1 sect id @13 coll $1. @15 instruct $30. @45 pageref $100. @146
show @147 " ";
   end;
end;
else if length(crs_pref) = 3 then do;;
   test = substr(pageref, 34, length(instruct));
   if instruct eq test then do;
         file "&outroot.&term.\index.txt";
```

```
put @1 sect_id @13 coll $1. @15 instruct $30. @45 pageref $100. @146
show @147 " ";
   end:
end;
else if length(crs_pref) = 2 then do;;
   test = substr(pageref, 33, length(instruct));
   if instruct eq test then do;
         file "&outroot.&term.\index.txt";
         put @1 sect id @13 coll $1. @15 instruct $30. @45 pageref $100. @146
show @147 " ";
   end;
end;
run;
```

There were three different course prefix lengths and the conditional logic made sure that the correct instructor was associated with their reports only. The differing lengths were initially leading to mismatched reports. The instructor id was used as the key by IT to ensure that each instructor only saw their own section-based reports.

CONCLUSIONS

The entire process outlined above was repeated three separate times to generate the distinct sets of reports and index files that were needed: public (viewable by all on the web), prefix-based (for deans and chairs), and section-based (for section instructors only). The process was repeated multiple times since the database design did not lend itself to easily ordering the questions in a single run. The program was built to be efficient and different approaches such as comparing PROC SQL and the DATA step merge were done to increase speed.

Overall, this programming approach was significantly slower than the earlier version but it was much more flexible. Earlier versions had the verbatim of each question hard coded into the program. Although this made laying out the report very precise, it did not allow for differences between reports unless each report was separately coded. The general look of the report was now easier to read due to the portrait orientation and the color enhanced the look. After the first round of producing the reports, the programs were being moved from a PC to a Windows Server with SAS installed. Testing on this platform almost tripled the speed at which the program ran. Some additional set up was required to allow the server to properly communicate with the IT server and to remotely run the SAS session from a PC, but it was not an extensive effort.

Later versions of SAS provided a more flexible programming environment to produce the varied reports necessary to meet the requirements of the university to allow students to provide feedback on faculty. ODS, macro, and enhancements to the graphical capabilities of SAS truly transformed the reports. In time, the programs will be honed for more efficiency and to further tweak the look of the reports for presentation on the web. Generating over twelve thousand reports was an initial programming challenge that led to enhanced learning of parts of SAS than had not been used much before.

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Stephanie R. Thompson Rochester Institute of Technology Nine Lomb Memorial Drive Rochester, NY 14623-5603 Work Phone: (585)475-7237 Email: Stephanie.Thompson@rit.edu

Web: http://finweb.rit.edu/irps/

ATTACHMENTS

Example of previous report (2 pages)

Example of new report (4 pages)

THE UNIVERSITY OF MEMPHIS

Student Instructional Rating System (SIRS) Evaluation Services Report

INSTRUCTOR INFORMATION	NUMBER OF STUDENTS / RESPONSE RATE	RESPONSE KEY

Name: College:	THOMPSON, STEPHANIE Fogelman Coll. Business	section Enrollment: Valid Responses:	15		(Unless otherwise specified)
Division:	Upper Division Course	Section:	8	53.3%	1 = Strongly Agree
Prefix:	ECON	Course:	8	53.3%	2 = Agree
Course:	4510	Prefix:	90	39.8%	3 = Neither
Section:	001	College:	1,735	34.8%	4 = Disagree
Semester:	Fall-06	University:	10,012	38.8%	5 = Strongly Disagree

0 Spurious Responses Removed Upper Division Course : ECON 4510 001

	SECTION		***COURSE***		***PREFIX***		***COLLEGE***	
	Mean STD DEV		Mean STD DEV		Mean STD DEV		Mean	STD DEV
1 The true to the contract of	1.13	0.35	1.13	0.35	1.47	0.89	1.68	0.94
 Instructor was enthusiastic when presenting course material. Instructor seemed to be interested in teaching. 	1.13	0.35	1.13	0.35	1.47	0.89	1.58	0.94
3. Instructor used helpful examples or personal experiences	1.00	0.00	1.00	0.00	1.54	0.80	1.71	0.87
4. Instructor seemed concerned with whether students learned	0.88	0.35	0.88	0.35	1.59	0.93	1.82	1.08
5. Students were interested in learning the course material.	1.13	0.35	1.13	0.35	1.69	0.93	1.88	1.02
6. Students were generally attentive in class.	1.25	0.46	1.25	0.46	1.54	0.77	1.69	0.84
	1.38	0.52	1.38	0.52	1.44	0.77	1.81	0.95
7. Students felt this course challenged them intellectually. 8. Students have become more competentdue to this course.	1.25	0.46	1.25	0.46	1.58	0.89	1.87	1.03
9. Instructor encouraged students to express opinions.	1.13	0.35	1.13	0.35	1.59	0.83	1.84	1.03
9. Instructor encouraged students to express opinions. 10. Instructor appeared receptive toothers' viewpoints. 11. Students had an opportunity to ask questions.	1.00	0.53	1.00	0.53	1.72	1.02	1.92	1.07
11. Students had an opportunity to ask questions. 12. Instructor generally stimulated class discussions.	1.00	0.00	1.00	0.00	1.41	0.69	1.56	0.76
12. Instructor generally stimulated class discussions.	1.13	0.35	1.13	0.35	1.64	0.99	1.96	1.12
13. Instructor did NOT cover too much material.	1.63	0.74	1.63	0.74	2.49	1.20	2.56	1.19
14. The Course materials were NOT presented too fast.	1.63	0.74	1.63	0.74	2.48	1.22	2.44	1.16
13. Instructor did NOT cover too much material. 14. The Course materials were NOT presented too fast. 15. The Homework assignments were NOT too time consuming.	1.63	0.74	1.63	0.74	2.30	1.12	2.28	1.12
16. The Reading assignments were NOT too difficult 17. Instructor related the concepts in a systematic manner.	1.63	0.74	1.63	0.74	2.39	1.22	2.29	1.09
17. Instructor related the concepts in a systematic manner.	1.25	0.46	1.25	0.46	1.78	0.93	1.83	0.88
18. The Course was well organized.	1.50	0.76	1.50	0.76	1.87	1.01	1.82	0.93
19. The Course materials were presented in logical content units.	1.13	0.35	1.13	0.35	1.76	0.90	1.74	0.82
20. The instructor effectively incorporated technology.	1.00	0.00	1.00	0.00	1.96	1.25	1.98	1.04
21. The Course directions were adequately outlined.	1.13	0.35	1.13	0.35	1.76	1.00	1.72	0.84
22. The Course contributed significantly to students' objectives.	1.13	0.35	1.13	0.35	1.72	0.98	2.01	1.10
23. Learning enhanced by intro. of technology in class.	1.25	0.46	1.25	0.46	2.22	1.28	2.25	1.14
24. The Percentage of course students felt they have learned. (1 = 90%+ ; 2 = 80%; 3 = 70%; 4 = 60%; 5 = 0% to 60%)	1.50	0.76	1.50	0.76	2.00	0.98	2.26	1.15
25. Instructor assessed how well students mastered the course.	1.13	0.35	1.13	0.35	1.99	1.15	2.03	1.11
26. The Course objectives were reflected in the exams.		0.35	1.13	0.35	1.88	1.03	1.86	1.00
27. Instructor appeared thoroughly competent in this subject.		0.00	1.00	0.00	1.41	0.87	1.49	0.81
28. Instructor was an effective teacher.		0.35	1.13	0.35	1.68	1.06	1.87	1.16
<pre>29. Are you confident that these ratings will be taken seriously? (1 = yes; 2 = no)</pre>		0.00	1.00	0.00	1.08	0.27	1.11	0.34
30. Was this form administered fairly & correctly? (1 = yes; 2 = no)		0.00	1.00	0.00	1.00	0.00	1.00	0.12
31. Was this a required course for you? (1 = yes; 2 = no)		0.35	1.88	0.35	1.34	0.48	1.09	0.30
32. Are you a major in this area? (1 = yes; 2 = no) 33. Expected Grade (1 = A; 2 = B; 3 = C; 4 = D; 5 = F)	1.63	0.52	1.63	0.52	1.36	0.48	1.41	0.51
	1.25	1.04	1.25	1.04	1.73	0.75	1.86	0.80
34. Overall GPA (1 = 0.0 to 2.2; 2 = 2.3 to 2.5; 3 = 2.6 to 2.9; 4 = 3.0 to 3.3; 5 = 3.4 to 4.0)	4.00	1.41	4.00	1.41	4.00	1.08	3.70	1.10
SCALE A. INSTRUCTOR INVOLVEMENT (ITEMS 1-4)	1.00	0.13	1.00	0.13	1.51	0.80	1.70	0.86
SCALE B. STUDENT INTEREST (ITEMS 5-8)	1.25	0.35	1.25	0.35	1.56	0.67	1.81	0.79
SCALE C. STUDENT - INSTRUCTOR INTERACTION (ITEMS 9-12)	1.06	0.29	1.06	0.29	1.59	0.74	1.82	0.88
SCALE D. COURSE DEMANDS (ITEMS 13-16)	1.63	0.71	1.63	0.71	2.41	1.06	2.39	0.96
SCALE E. COURSE ORGANIZATION (ITEMS 17-20)		0.36	1.22	0.36	1.84	0.88	1.84	0.78

THE UNIVERSITY OF MEMPHIS Student Instructional Rating System (SIRS) Evaluation Services Report

scores)

sections'

STEPHANIE d from all

excluded

(Spurious responses

Instructor:

THOMPSON,

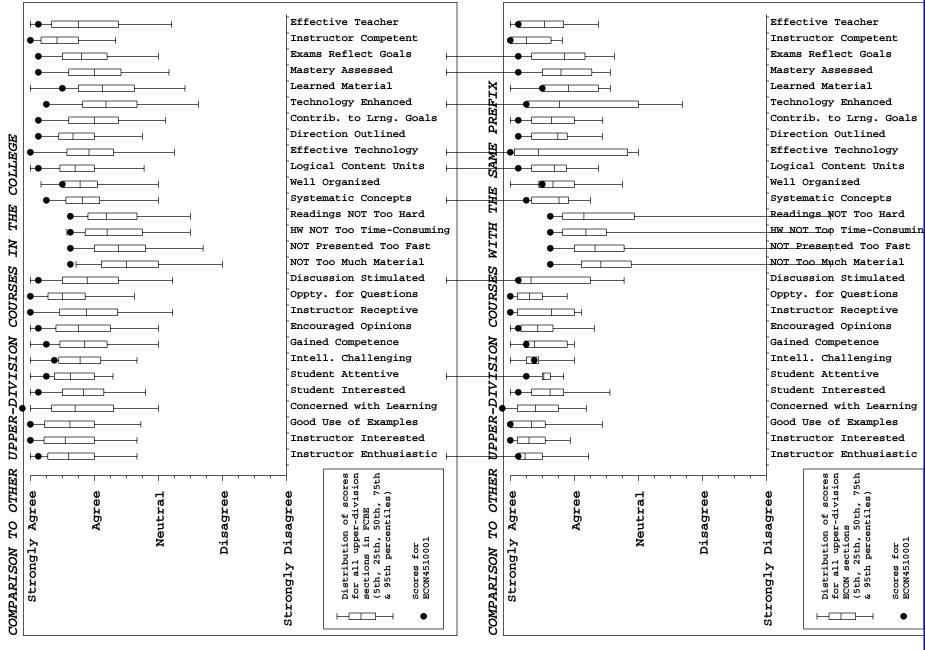
Reponses

SIRS

Distributions of ; ECON4510001

Percentile

Section:



U/F Student Evaluation or Teaching Effectiveness (SETE) ECON4510001 Taught by Thompson, Stephanie 6 Students Enrolled Upper Division Course

Statistical Summary

#	Question	Section Mean	Section Std Dev	Course Mean	Course Std Dev	Prefix Mean	Prefix Std Dev	College Mean	College Std Dev
1	I have become more competent because of this course.	4.75	0.50	4.75	0.50	4.18	1.11	4.23	1.01
2	I had an opportunity to ask questions in or outside of class.	4.75	0.50	4.75	0.50	4.70	0.50	4.50	0.74
3	The course was well organized.	4.25	0.96	4.25	0.96	4.20	1.15	4.26	0.97
4	The tests or other evaluation methods adequately assessed how well I learned the course material.	4.75	0.50	4.75	0.50	4.07	1.18	4.10	1.13
5	The instructor was enthusiastic when presenting course material.	5.00	0.00	5.00	0.00	4.41	0.97	4.38	0.91
6	The instructor was interested in teaching.	4.75	0.50	4.75	0.50	4.48	0.87	4.45	0.86
7	The instructor was concerned with whether the students learned the materials.	4.75	0.50	4.75	0.50	4.29	1.14	4.26	1.02
8	The instructor was knowledgeable about the subject.	4.75	0.50	4.75	0.50	4.68	0.66	4.65	0.64
9	In general, the instructor was an effective teacher.	4.75	0.50	4.75	0.50	4.32	1.08	4.24	1.09
10	Do you have confidence that these ratings will be taken seriously? (1=Yes 2=No)	1.25	0.50	1.25	0.50	1.09	0.29	1.13	0.34
11	Was this form administered fairly and correctly? (1=Yes 2=No)	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.05
12	Was this a required course for you? (1=Yes 2=No)	2.00	0.00	2.00	0.00	1.38	0.49	1.10	0.31
13	Are you a major in the area in which this course is being taught? (1=Yes 2=No)	1.50	0.58	1.50	0.58	1.15	0.36	1.37	0.48
14	What grade do you expect to receive in this course? (1=A 2=B 3=C 4=D 5=F)	1.25	0.50	1.25	0.50	1.83	0.70	1.77	0.75
15	What is your overall GPA? (1) 2.2 or less (2) 2.3-2.5 (3) 2.6-2.9 (4) 3.0-3.3 (5) 3.4-4.0	4.50	0.58	4.50	0.58	3.86	1.22	3.76	1.03

U/F Student Evaluation or Teaching Enectiveness (SETE) ECON4510001 Taught by Thompson, Stephanie

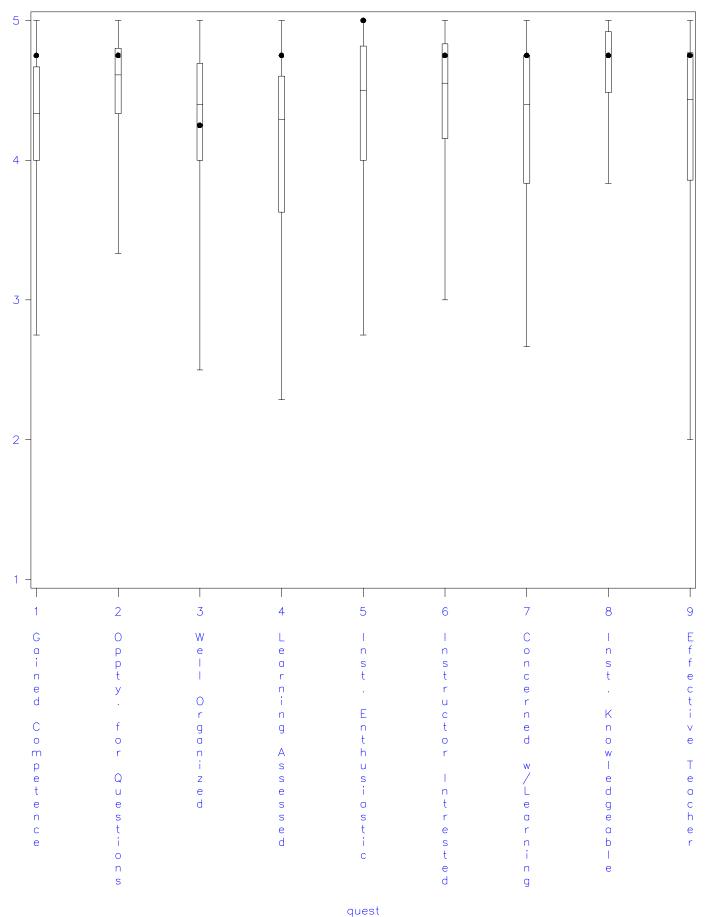
6 Students Enrolled Upper Division Course

Response Summary

		Section	Section	Course	Course	Prefix	Prefix	College	College
#	Question	Responses	N/A	Responses	N/A	Responses	N/A	Responses	N/A
1	I have become more competent because of this course.	4	0	4	0	55	0	1400	4
2	I had an opportunity to ask questions in or outside of class.	4	0	4	0	56	0	1404	2
3	The course was well organized.	4	0	4	0	56	0	1401	3
4	The tests or other evaluation methods adequately assessed how well I learned the course material.	4	0	4	0	54	2	1399	7
5	The instructor was enthusiastic when presenting course material.	4	0	4	0	56	0	1402	4
6	The instructor was interested in teaching.	4	0	4	0	56	0	1402	4
7	The instructor was concerned with whether the students learned the materials.	4	0	4	0	56	0	1400	5
8	The instructor was knowledgeable about the subject.	4	0	4	0	56	0	1400	3
9	In general, the instructor was an effective teacher.	4	0	4	0	56	0	1394	2
10	Do you have confidence that these ratings will be taken seriously? (1=Yes 2=No)	4	0	4	0	55	1	1363	40
11	Was this form administered fairly and correctly? (1=Yes 2=No)	4	0	4	0	56	0	1387	16
12	Was this a required course for you? (1=Yes 2=No)	4	0	4	0	56	0	1399	1
13	Are you a major in the area in which this course is being taught? (1=Yes 2=No)	4	0	4	0	55	1	1394	5
14	What grade do you expect to receive in this course? (1=A 2=B 3=C 4=D 5=F)	4	0	4	0	53	2	1381	21
15	What is your overall GPA? (1) 2.2 or less (2) 2.3-2.5 (3) 2.6-2.9 (4) 3.0-3.3 (5) 3.4-4.0	4	0	4	0	51	2	1373	22

07F Student Evaluation of Teaching Effectiveness (SETE)

College Mean Distribution Comparison to ECON4510001



07F Student Evaluation of Teaching Effectiveness (SETE)

Prefix Mean Distribution Comparison to ECON4510001

