

2

The DUET Experiment

An online course evaluation experiment called DUET was conducted at Duke University during the 1998–1999 academic year. This experiment possessed several unique features, including the solicitation of survey responses from students both before and after they had received their final course grades and monitoring of students as they viewed mean grades and course evaluations of courses taught in the past.

DEBATE OF THE GRADING-REFORM PROPOSAL among members of the Duke faculty spawned numerous questions regarding the role of grades at the university. Issues of particular concern included the effects that grades had on student course selection decisions and student responses to end-of-term course evaluations, and the extent to which GPA reflected a fair summary of students' academic performance. Although the achievement index proposal was defeated, many faculty members were not satisfied that these underlying issues had been adequately addressed. A short time later, I was asked by Professor John Richards, chair of the Academic Affairs Committee, and Robert Thompson, dean of Trinity College of Arts and Sciences, to chair a committee to continue an examination of these issues. Other individuals asked to join the committee included Professor Daniel Gauthier, from physics; Mary Nijhout, associate dean of Trinity College; Benjamin Kennedy, vice president of academic affairs in the Duke student government; and Jeff Horowitz, cofounder and a coadministrator of a student website called Devilnet. In addition to collecting information regarding the effects of grading practices on undergraduate education, the committee was charged with studying the feasibility of collecting end-of-term course evaluations electronically over the web.

Shortly after its formation, the members of the committee agreed to proceed with the construction of a website that would have as its primary purpose the collection and dissemination of course evaluation data. As a secondary objective, the website would serve as an experimental platform to collect data useful for answering some of the questions that had been posed during the grading-reform debate.

After several months of meetings in which items to be included on the course evaluation form were negotiated among committee members, followed by a perfunctory review of the experimental

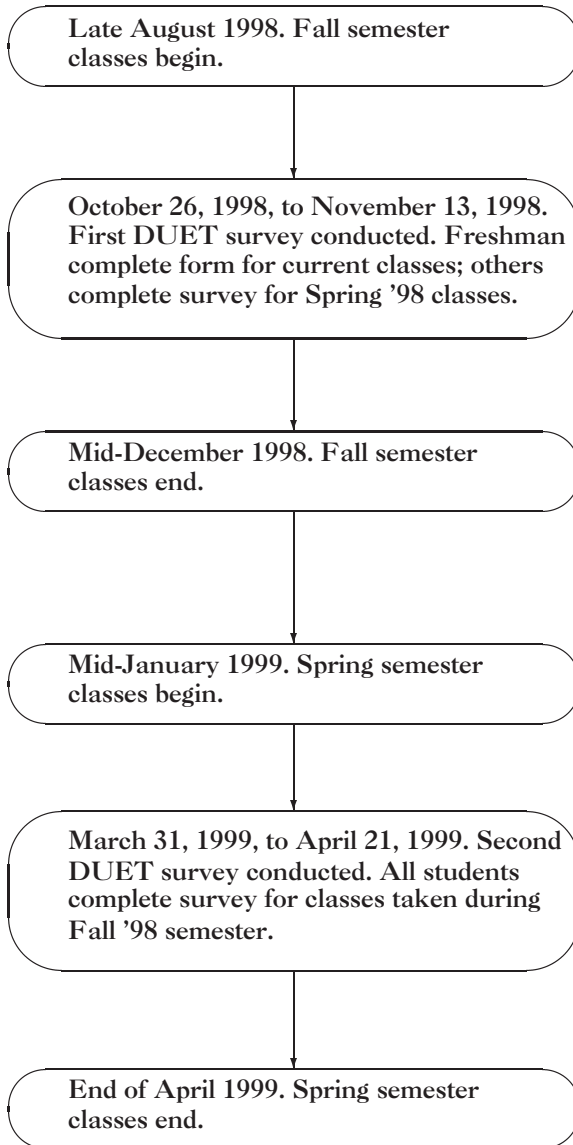
design by a human subjects protocol committee and numerous meetings with the Academic Affairs Committee to gain final approval to proceed with the experiment, the DUET (Duke Undergraduates Evaluate Teaching) website was launched in the fall of 1998. It was scheduled to run for three years, through the spring of 2001, but for reasons detailed later in this chapter, it was terminated after the spring survey period in 1999. The basic operation of the website is described below.

First, to publicize the DUET website and to encourage student participation, Dean Thompson agreed to send personalized letters to all full-time Duke undergraduates requesting their participation in an experiment designed to study the feasibility of collecting course evaluation data over the web. Included in this letter were individual student access codes that permitted students to enter the website. Ideally, student passwords from the university computing system would have been used in place of these codes, but it proved technically impossible to transfer these passwords to Webslingerz Inc., the company that administered the site, in time for the launch of the survey. Students were also informed that they would be able to review both course evaluation data entered by other students and mean course grades of courses taught in the past after they had completed their surveys.

The DUET website was activated for two three-week periods, the first beginning the week prior to fall registration in 1998 and the second a week prior to spring registration in 1999. Both periods fell approximately ten weeks into their respective semesters. The survey was conducted immediately before and during registration as an incentive for students to participate; by completing the DUET survey prior to registration, students could view course evaluation data and grade data for courses they planned to take the following semester. With the exception of first-year students in the fall of 1998, all students completed the survey for courses they had taken the previous semester. As part of the experimental design, first-year students were asked to complete the survey for courses they were currently taking. A timeline for the DUET experiment is depicted in Figure 1.

Upon entering the DUET website, students were initially confronted with text informing them that course evaluation data col-

FIGURE 1
DUET timeline.



lected on the site would be used as part of a study investigating the feasibility of collecting course evaluation on the web. They were also told that their responses would not be accessible to either their instructors, other members of the faculty, other students, or university administrators. After indicating their consent to participate in the study, items on the survey were presented to students in groups of 5–7 questions. Each course they had taken the previous semester (or, for freshmen in the fall, that they were currently taking) appeared simultaneously as a row in a table, while item responses were listed in columns next to each course. Item text appeared above the courses, and students were required to click a button in response to each item for every course before the next set of items appeared. With only one or two exceptions, all items included a “Not Applicable” response. After completing the 38 items on the survey, students were invited to view course evaluation data collected from other students, course mean grades for all classes that had been taught in the past five years, and their adjusted GPAs. They were also encouraged to express free-form opinions on grades and grade adjustments. Survey items and possible responses are listed below.

DUET SURVEY ITEMS

1. In what format was this course primarily taught?
 - 1) Lecture 2) Seminar 3) Project
 - 4) Combination of Lecture, Seminar, and Project
 - 5) Skills course 6) Not Applicable
2. Did you take this course to satisfy a
 - 1) Major requirement 2) Distributional requirement (not major requirement) 3) Elective
3. How would you rate your interest in the subject matter of this course prior to enrolling in it?
 - 1) Very Low 2) Low 3) Moderate 4) High 5) Very High
 - 6) Not Applicable
4. What proportion of the class sessions did you find challenging?

- 1) less than 35% 2) 35-65% 3) 65-85% 4) 85-95%
 - 5) More than 95% 6) Not Applicable
5. What proportion of the class sessions did you find relevant to course objectives?
 - 1) less than 35% 2) 35-65% 3) 65-85% 4) 85-95%
 - 5) More than 95% 6) Not Applicable
 6. What proportion of the reading and writing assignments did you find challenging?
 - 1) less than 35% 2) 35-65% 3) 65-85% 4) 85-95%
 - 5) More than 95% 6) Not Applicable
 7. What proportion of the reading and writing assignments did you find relevant to the course material?
 - 1) less than 35% 2) 35-65% 3) 65-85% 4) 85-95%
 - 5) More than 95% 6) Not Applicable
 8. Do you know what the goals of the course were?
 - 1) No 2) Yes 3) Not Applicable
 9. Approximately how many hours per week did you spend in this course attending class meetings, discussion sessions, and labs?
 - 1) Less than 2 hours 2) 2-3 hours 3) 3-5 hours
 - 4) 5-8 hours 5) More than 8 hours 6) Not Applicable
 10. Approximately how many hours per week did you spend outside of class completing reading and written assignments, studying for exams, and completing projects, etc.?
 - 1) Less than 2 hours 2) 2-3 hours 3) 3-5 hours
 - 4) 5-8 hours 5) More than 8 hours 6) Not Applicable
 11. What proportion of the reading assignments did you complete?
 - 1) less than 30% 2) 30-50% 3) 50-75% 4) 75-90%
 - 5) More than 90% 6) Not Applicable
 12. What proportion of the written assignments did you complete?
 - 1) less than 50% 2) 50-75% 3) 75-90% 4) More than 90%
 - 5) Not Applicable
 13. What proportion of the classes did you attend?
 - 1) less than 30% 2) 30-50% 3) 50-75% 4) 75-90%
 - 5) More than 90% 6) Not Applicable

14. How difficult was the material taught in this course compared to other courses that you've taken at Duke?
1) Not difficult 2) Less difficult than average
3) Average 4) More difficult than average
5) Very difficult 6) Not Applicable
15. How effective was the instructor in encouraging students to ask questions and express their viewpoints?
1) Very poor 2) Poor 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
16. Did this class have a TA (teaching assistant)?
1) Yes 2) No 3) Not Applicable
17. How important was the TA's role in the course compared to that of the primary instructor(s)?
1) Not important
2) Less important than the primary instructor(s)
3) Equally important as primary instructor(s)
4) More important than the primary instructor(s)
5) Not Applicable
18. How would you rate the TA of this course?
1) Very bad 2) Bad 3) Fair 4) Good 5) Excellent
6) Not Applicable
19. How effective were exams, quizzes, and written assignments at measuring your knowledge of course material?
1) Very bad 2) Bad 3) Fair 4) Good 5) Excellent
6) Not Applicable
20. How was this class graded?
1) Very leniently 2) More leniently than average
3) About average 4) More severely than average
5) Very severely 6) Not Applicable
21. How valuable was feedback on examinations and graded materials?
1) Very poor 2) Poor 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
22. What grade do you expect (or did you get) in this class?
1) A+
2) A
3) A-
4) B+

- 5) B
 - 6) B-
 - 7) C+
 - 8) C
 - 9) C-
 - 10) D+/D/D-
 - 11) F
 - 12) Not Applicable
23. On a scale from 1 to 5, with 1 being "completely unaware" and 5 being "completely aware", how aware were you of how this course would be graded when you enrolled in it?
- 1) 1
 - 2) 2
 - 3) 3
 - 4) 4
 - 5) 5
 - 6) Not Applicable
24. How much did your knowledge of how the course would be graded positively affect your decision to enroll in it?
- 1) No effect or negative effect
 - 2) Slight effect
 - 3) Moderate effect
 - 4) Significant effect
 - 5) Very significant effect
 - 6) Not Applicable
25. How well did you learn and understand the course material?
- 1) Very poorly
 - 2) Poorly
 - 3) Fair
 - 4) Well
 - 5) Very Well
 - 6) Not Applicable
26. How much did you learn in this course compared to all courses that you have taken at Duke?
- 1) Much less than average
 - 2) Less than average
 - 3) Average
 - 4) More than average
 - 5) Much more than average
 - 6) Not Applicable
27. How would you rate the instructor(s) knowledge of course material?
- 1) Very bad
 - 2) Bad
 - 3) Fair
 - 4) Good
 - 5) Very Good
 - 6) Excellent
 - 7) Not Applicable
28. How easy was it to meet with the instructor outside of class?
- 1) Very difficult
 - 2) Difficult
 - 3) Not Hard
 - 4) Easy
 - 5) Very Easy
 - 6) Not Applicable
29. How would you rate the organization of the instructor(s) in this course?

- 1) Very poor 2) Poor 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
30. How good was the instructor at relating course material to current research in the field?
1) Very bad 2) Bad 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
31. To what extent did this instructor demand critical or original thinking?
1) Never 2) Seldom 3) Sometimes 4) Often 5) Always
6) Not Applicable
32. The instructor's concern for the progress of individual students was
1) Very Poor 2) Poor 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
33. How would you rate this instructor's enthusiasm in teaching this course?
1) Very bad 2) Bad 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
34. How good was the instructor(s) at communicating course material?
1) Very bad 2) Bad 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
35. How does this instructor(s) compare to all instructors that you have had at Duke?
1) Very bad 2) Bad 3) Fair 4) Good 5) Very Good
6) Excellent 7) Not Applicable
36. How would you rate your interest in the subject matter covered in this course now?
1) Very Low 2) Low 3) Moderate 4) High 5) Very High
6) Not Applicable
37. Would you recommend this course to others?
1) Yes 2) No 3) Not Applicable
38. Would you take another course from this instructor?
1) Yes 2) No 3) Not Applicable

Several aspects of the DUET website should be emphasized here since they play a prominent role in later analyses. First, all course-evaluation data collected during the experiment were linked

to respondents. Because complete student transcripts were also available from the university registrar, this meant that student grades and other background variables could be used in conjunction with the evaluation data to provide a clearer indication of the meaning of student responses to the survey. Some of the background variables that were available for this purpose included student major, courses taken and grades received, SAT math and verbal scores, high school GPA, gender, and ethnic group.

Second, and unbeknownst to students, a record was kept of every student query into the course-evaluation and mean-grade databases. That is, every time a student viewed either a mean course grade or a histogram summary of data entered by other students, the date, time, course, and instructor of that course, along with the querying student's ID, were all appended to a second database. These data were later used to investigate the effect that knowledge of mean grades and course evaluations had on students' course enrollment decisions.

A final experimental aspect of the DUET website involves the survey design for first-year students. First-year students completed the survey for their fall courses twice, once before completing their fall courses and receiving their final grades, and once after. Because one DUET survey item asked students what grade they received or expected to receive in their courses, the responses collected from freshman at the two time points provide an ideal mechanism for investigating the influence that expected and received student grades have on student evaluations of teaching. Analyses based on these data are presented in the last sections of Chapter 4.

A number of factors that affected student participation in the survey are also worth mentioning here. The first concerns a technical problem that surfaced shortly after the DUET website was launched. This problem involved limits on the volume of traffic that the server hosting the site could handle and resulted in students being locked out of the site—often after already completing a substantial portion of their survey—whenever traffic volume exceeded the server's capacity. The host was eventually replaced by a server that could handle more connections, but in the interim many students' surveys were lost, and many students decided that participation in the study was not worth the hassle.

A similar problem centered on the use of non-Duke passwords for system access. As mentioned previously, it proved to be administratively impossible for students to use their Duke computing account passwords to access the site. Instead, students had to use passwords that were included in the mailing from Dean Thompson soliciting their participation in the experiment. Unfortunately, many students lost this correspondence before the study began, and so were unable to access the site without first requesting their password from the site administrator.

As serious as the impact on survey response from these technical glitches was, faculty opposition to the site soon proved to be far more threatening. Faculty opposition to the experiment was caused by several factors, the most prominent being a perception held by many professors that posting student evaluations of their courses on the web represented a violation of their privacy. And to be truthful, I must confess that had the experiment been initiated by someone else, I would have had similar concerns given the academic climate that then prevailed at Duke. During that period, professors had almost complete control over their course evaluations, at least as to whether they were released to students and other faculty members. In fact, student efforts to publish course synopses were regularly hampered by instructors who refused even to allow representatives of the Duke student government from visiting their offices to compile summary statistics from their evaluations.

Other members of the faculty were suspicious of any use of electronic media to collect information. Such concerns ranged from healthy skepticism over the confidentiality of student responses to something approaching paranoia. For example, one professor circulated an email in which she stated that the on-line course evaluation system was “far too close to techniques of surveillance for my comfort,” and likened the DUET website to intelligence gathering activities of “the US military (and other military units around the world),” who were tasked “to design, refine, and implement research and development efforts that will give them superiority over information.”¹

¹Personal email communication.

In addition to these complaints, which were, to some extent, anticipated, there was also substantial and unexpected opposition to the DUET experiment from the Department of Mathematics. Some of this resistance likely stemmed from the privacy concerns mentioned above, with more generated by a rather vocal member of the mathematics department who had received particularly bad ratings for an introductory calculus course he had taught the previous semester. But most seemed to arise from a fear among many in the math department that their course enrollments would suffer if mean grades of previous courses continued to be displayed on the web. They apparently anticipated one of the major conclusions of the study, though they were likely naive in thinking that they were not already suffering from exactly this phenomenon. Further opposition from the math department was ignited when another member of the department was asked by a student to explain why the class mean grade in his class was an entire letter grade lower than the mean class grades awarded in three other sections of the same course. As it happens, the mean GPAs and SAT scores for students in all sections of this course were nearly identical, so the student may have had a point. In any case, the senior professor in question was not amused by the inquiry.

In April of 1999 I received a letter written on behalf of five departments (Asian and African Languages, Art and Art History, Germanic Languages and Literature, Romance Studies, and Slavic Languages and Literatures) and 28 faculty members (21 from Mathematics) demanding that their courses and departments be excluded from the DUET website. In this letter they threatened me with unspecified “further action” if their request was not satisfied. Copies of this letter were sent to the provost, two deans, and several other administrative officers. I suspect that even greater pressure was applied to members of the administration to end the study, and shortly thereafter, Dean Thompson succumbed to the inevitable and asked me to terminate the experiment after the spring 1999 collection period. Several weeks later, the Arts and Sciences Council considered a motion to postpone the experiment until issues regarding the use of faculty members as possible subjects in the experiment and privacy concerns had been addressed. A vote on

the motion was postponed until the following fall, by which time the plug had already been pulled.

For many of the analyses planned using data collected during the DUET experiment, the early termination of the experiment was not particularly problematic. During the two semesters that the DUET website was active, 11,521 complete course evaluations were recorded, with each evaluation consisting of 38 item responses for a single course. Of the 6,471 eligible full-time, degree-seeking students who matriculated at Duke University after 1995, 1,894 students (29%) participated in the experiment. Of these, approximately one-half participated in both the fall and spring surveys. Participation during the following year would likely not have increased greatly, and response patterns for different student groups would probably also have remained about the same. There was, however, one analysis that was severely affected by the experiment's early termination: an analysis of the effects of student grade expectations on course enrollments. For that analysis, data that would have been collected during the third semester of the experiment were crucial for assessing the impact of sample selection effects. To overcome the absence of these data, I decided to conduct a follow-on email survey of selected study participants during the 1999–2000 academic year. Details of this follow-on survey and its role in subsequent analyses are described in Chapter 6.

Histogram summaries of student responses collected during the experiment are displayed in Figure 2.

APPENDIX: ISSUES OF NONRESPONSE

Because only 29% of eligible students participated in the DUET experiment, it is important to consider the effects of “nonresponse.” Survey nonresponse occurs when a subset of a study population fails to participate in a survey. If students’ decisions to participate in the DUET survey were strongly related to the way they (would have) responded to the survey, then the generalizability of study conclusions based on survey data collected only from those students who did participate is severely compromised. On the other hand, if the “response mechanism” is uncorrelated or

FIGURE 2

Histogram summary of DUET responses. For clarity, “Not Applicable” responses were excluded from the plots.

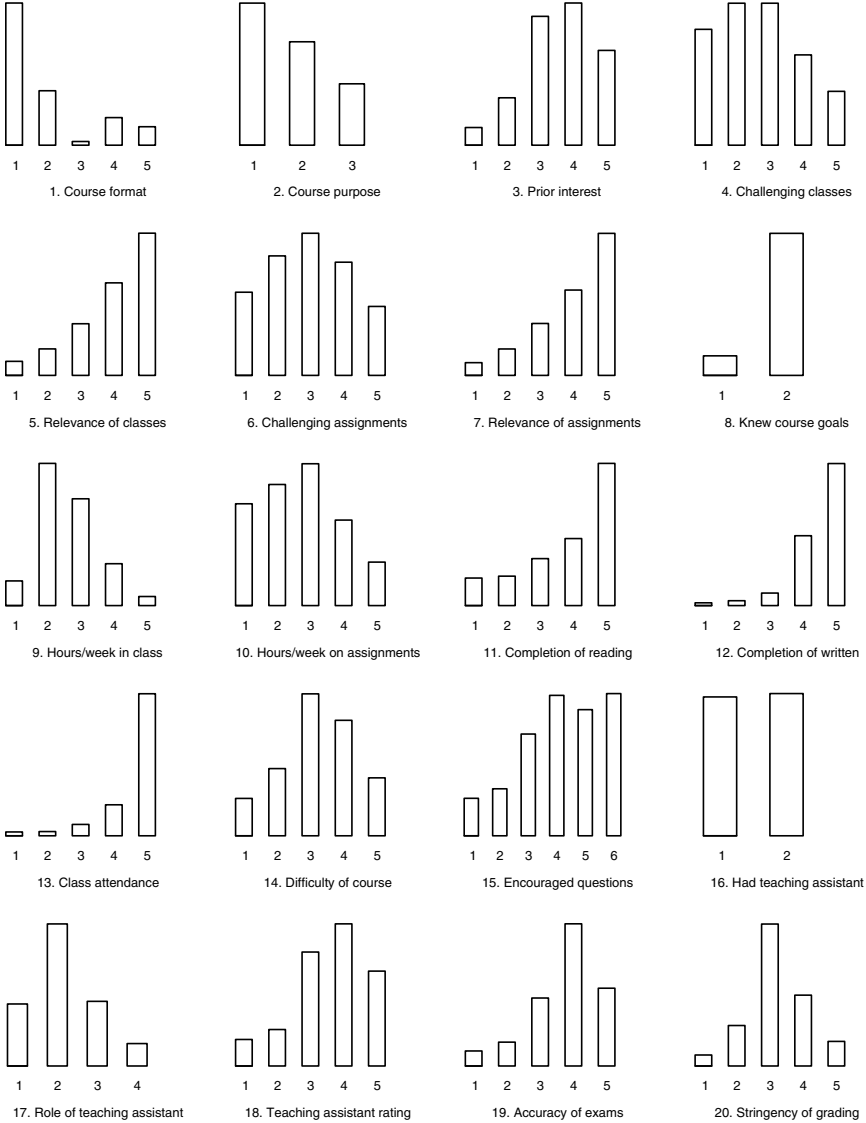
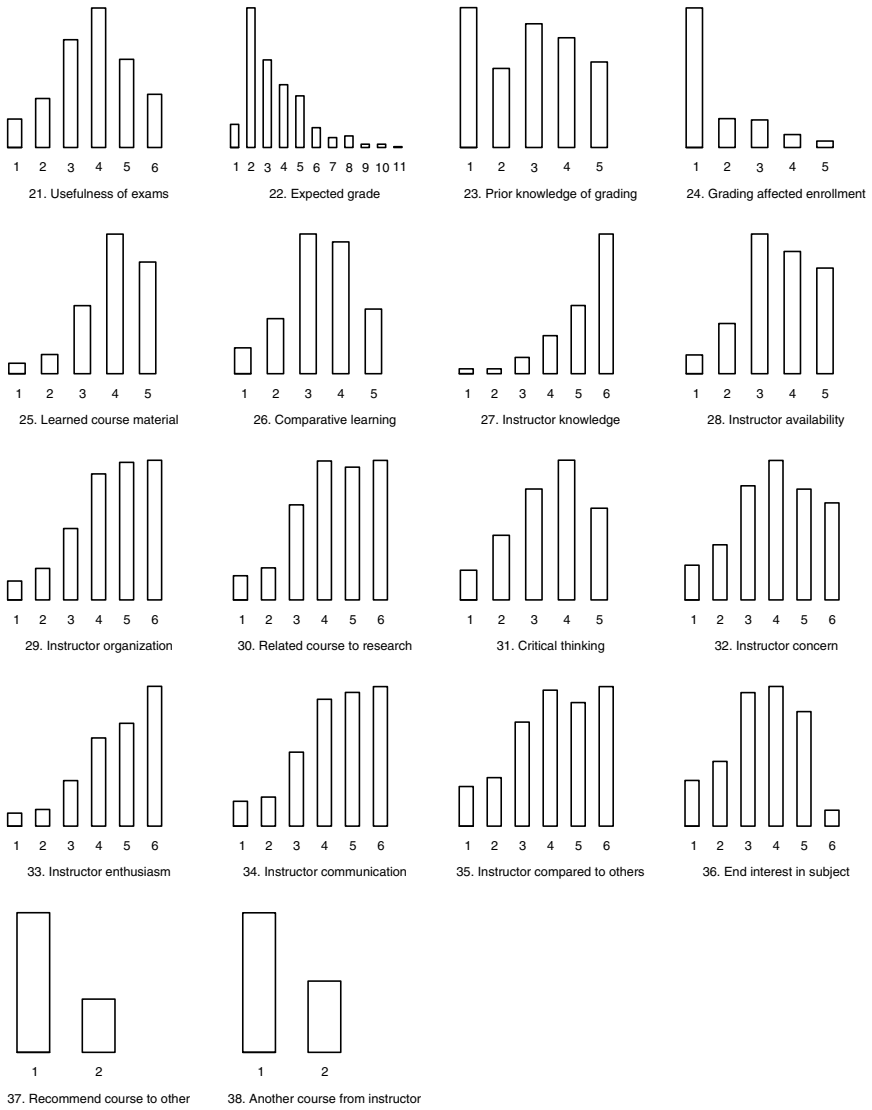


FIGURE 2 (continued)



only moderately correlated with student response patterns, then nonresponse does not seriously limit the scope of study conclusions. Unfortunately, examining differences between the way study participants did respond to the survey and the way nonparticipants may have responded to the survey is difficult, owing to the absence of data from the latter group.

Students' decisions to abstain from participation in the DUET experiment may be attributed to several causes, including time constraints experienced by students during the survey periods, unavailability of a convenient access port to the DUET website, privacy concerns, objections to the perceived purpose of the study, and, of course, student apathy. As mentioned above, technical difficulties at the website and problems associated with the use of student access codes may also have contributed to the low response rate. For instance, problems with the database server caused an almost constant stream of service interruptions during the first three days of the initial survey period. As a result of these system failures, numerous student responses to the survey were lost, and it is difficult to estimate the number of students who may have permanently dropped out of the experiment due to frustration encountered in accessing the site during the initial solicitation.

In addition to early technical problems with the server, the fact that students could not use their usual computer passwords to access the site also reduced participation. Over one hundred requests for access codes were received by the webmaster during the two survey periods. Although access codes were provided to all students from whom such requests were received, subsequent participation rates among these students—along with an unknown number of other students who gave up without requesting access codes—were, without doubt, substantially reduced.

Although these technical problems undoubtedly affected survey participation, it is unlikely that they introduced significant response biases. However, other causes of nonresponse may have been less benign. For example, women were less likely to participate in the survey than men were, and African-American students were less likely to participate than Caucasians. To better understand the

effects of these and other sources of nonresponse, the relationships between an individual's propensity to participate in the survey and an individual's demographic attributes were examined.

Several demographic classifications that might plausibly be associated with an individual's decision to participate in the survey were extracted from data provided by the Office of the University Registrar. These classifications included gender, ethnicity (Asian, African-American, Hispanic, Native American, White, and Other), college major and school, academic year, and GPA.

As mentioned above, gender was a highly significant predictor of survey participation. One-third of eligible males (33%) participated in the study, while only 25% of females did. So too, were differences associated with ethnic groups. Participation among students who identified themselves as African-American was 14%, while for students who identified themselves as Native American participation was 25%. Participation rates for other ethnic categories were 38% for Asian students, 26% for Hispanic students, and 29% for White students. Students who declined to place themselves into one of these categories participated with 35% probability. Clearly, if mechanisms responsible for differences in response rates among different gender and ethnic categories are also linked to differences in response patterns to the DUET items, then adjustments would be necessary to correct for the disproportionate numbers of students from each group who participated in the experiment.

To evaluate the impact of these demographic classifications on response patterns in the survey data, histogram estimates of the probabilities that each gender and ethnic group responded to each item were constructed. These histograms are displayed in Figures 3 and 4.

Figure 3 shows that item responses did not vary substantively with gender. The greatest proportional differences between the responses of women and men occurred on the first two items, which queried the format and reason for taking a course. These differences are probably attributable to tendencies for women and men to major in different academic fields, and the concomitant differences in the types and numbers of courses required in these different

FIGURE 3

Comparison of DUET responses for women and men. For each response category, women's responses are depicted in the left bar and men's responses are depicted in the right bar.

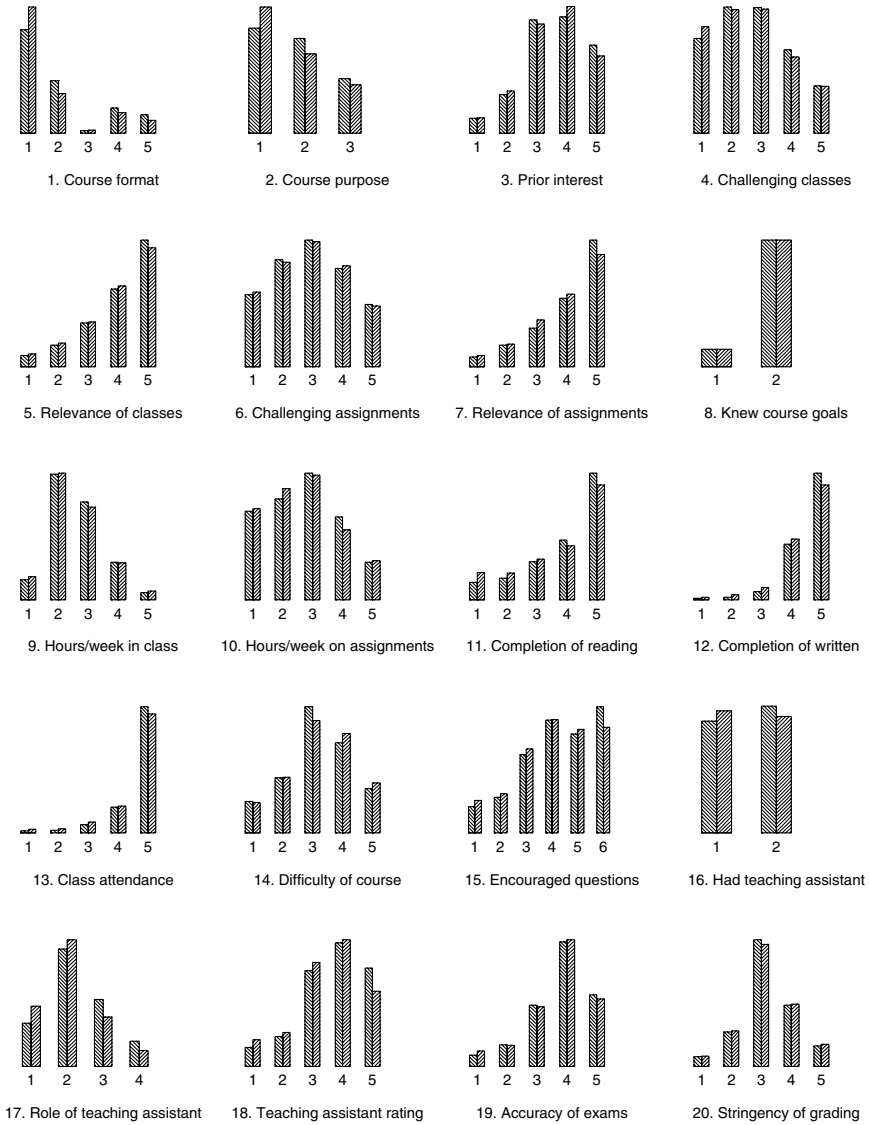


FIGURE 3 (continued)

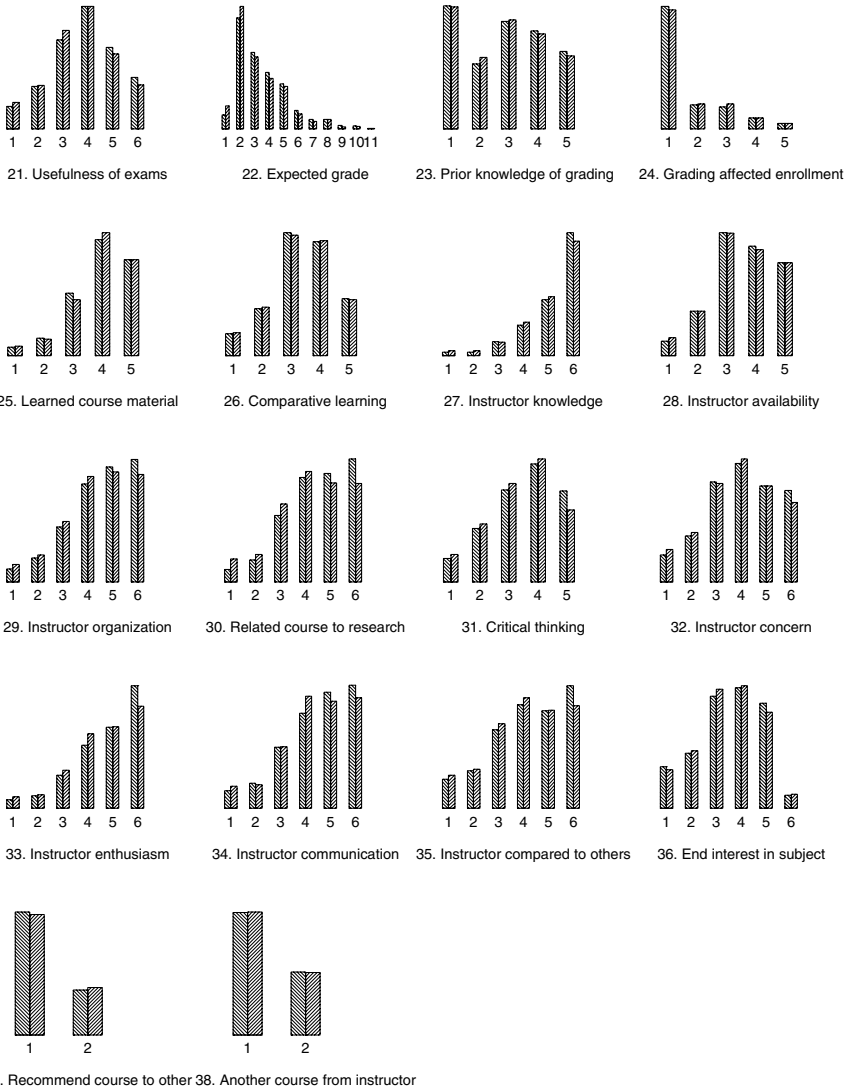


FIGURE 4

Comparison of DUET responses by ethnic group. For each response category and item, the order of the vertical bars is White, Asian, African-American, Hispanic, Native American, and Unknown.

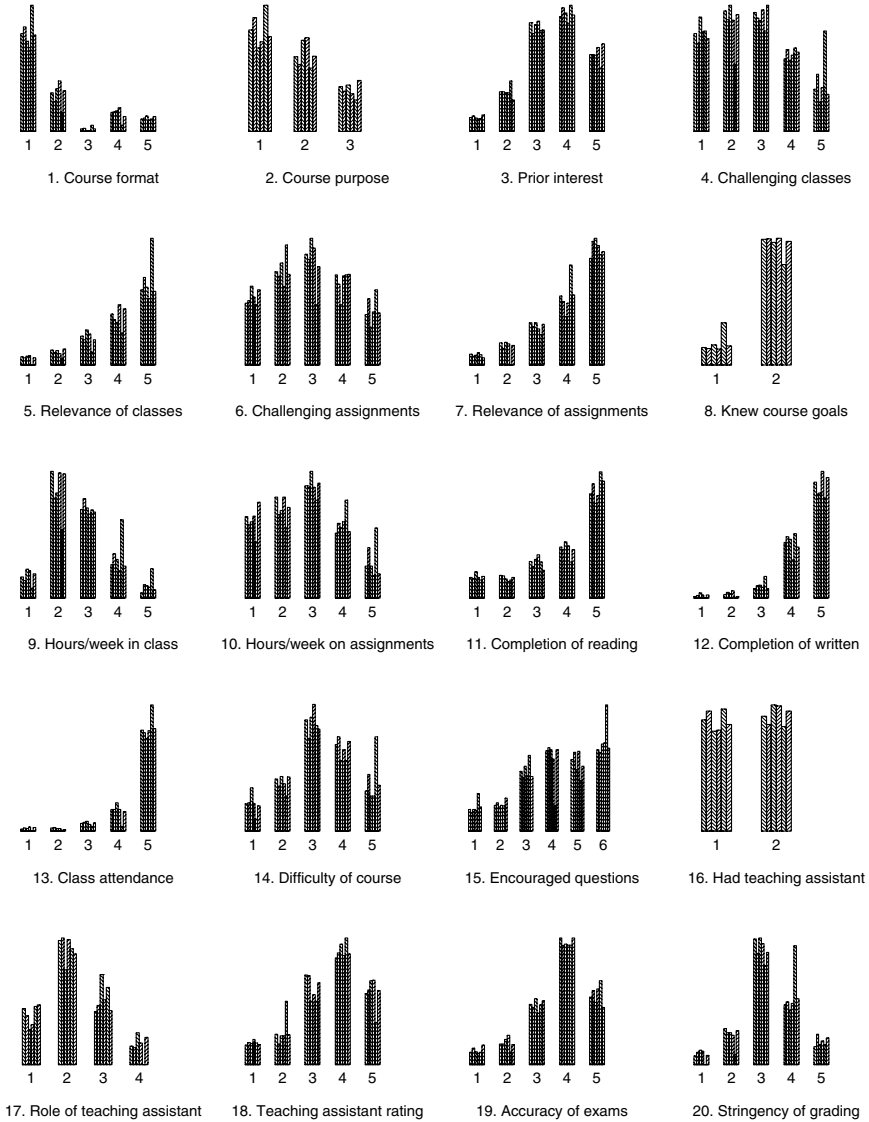
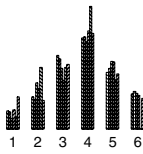
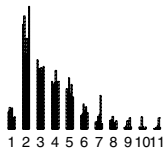


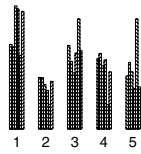
FIGURE 4 (continued)



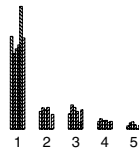
21. Usefulness of exams



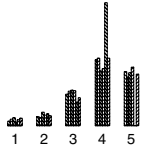
22. Expected grade



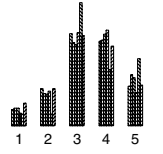
23. Prior knowledge of grading



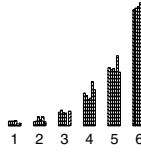
24. Grading affected enrollment



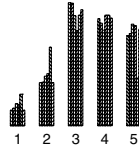
25. Learned course material



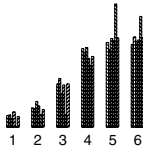
26. Comparative learning



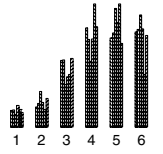
27. Instructor knowledge



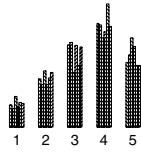
28. Instructor availability



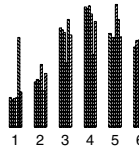
29. Instructor organization



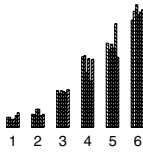
30. Related course to research



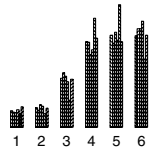
31. Critical thinking



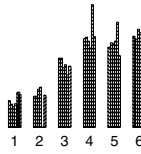
32. Instructor concern



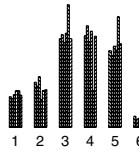
33. Instructor enthusiasm



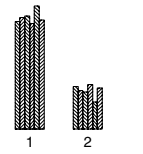
34. Instructor communication



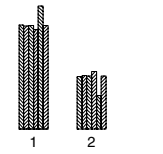
35. Instructor compared to others



36. End interest in subject



37. Recommend course to other



38. Another course from instructor

fields. Aside from these items, it is evident from Figure 3 that females were slightly more likely to find courses and assignments relevant, to complete written and reading assignments, and to give slightly higher ratings on items that evaluated specific teaching traits. Whether these differences can be attributed to gender differences or differences in the types of courses taken is not immediately clear.

A more quantitative summary of the effects of gender nonresponse on the distribution of observed responses can be constructed by comparing the proportion of responses actually collected in each item category to the estimated proportion of responses that would have been collected had women and men responded to the survey with equal probability. Using demographic data obtained from the registrar's office to make this correction, the probability that the "same" response would be drawn from both the observed and "nonresponse-corrected" distributions exceeded 98% for all items on the survey.² In practical terms, then, differences between the response rates of women and men probably had no substantive impact on survey conclusions.

Differences in the response patterns between ethnic groups are more pronounced than they are for gender. The most discrepant cells in Figure 4 correspond to respondents who classified themselves as Native Americans (the fifth vertical bar in each response category). However, because only 20 individuals identified themselves in this category, these deviations may simply represent random fluctuations in responses obtained from this group. Also, because this group represented only a small proportion of the eligible student population, corrections to response totals to account for these differences would not have a significant effect on survey totals. Similarly, the comparatively low response rates obtained from African-American students, though disappointing, does not constitute a significant threat to the veracity of survey conclusions, since only 8.3% of eligible students fell into this category. Furthermore, response patterns obtained from African-American students did

²Technically, the value of 98% was calculated by subtracting the total variation distance between the distributions from 1.0. It thus represents the probability that the same value would be drawn from both distributions in γ -coupling.

not differ greatly from responses obtained from other non-Native American ethnic categories.

The quantitative statistic cited above for gender differences extends also to differences in response probabilities attributable to ethnic groups. In this case, the probability that the same response would be drawn from the distribution actually observed for each item as from the distribution corrected for differences in response probabilities between ethnic groups exceeds 99% for every survey item. As was the case for gender, nonresponse patterns attributable to ethnic group do not appear problematic.

In addition to gender and ethnicity, variation in response patterns and response rates according to educational factors like student GPA, academic division of students' majors, and academic year were examined. Figure 5 displays the proportion of responses received from students who had GPAs less than 3.3 (left bar) against responses from students who had GPAs greater than 3.3. Once again, the response patterns do not vary substantially across the different demographic groups. In this case, the only notable differences between the responses collected from students having high GPAs and low GPAs were observed for item 11 (proportion of reading assignments completed), item 13 (class attendance), and item 22 (expected grade). For each of these items, differences between the distributions of responses from the two groups lean in the expected direction. The response rates for the two groups were 32% for students who had GPAs less than 3.3, and 27% for students with GPAs greater than 3.3.

The probability that equivalent responses would be drawn from the observed distribution and the distribution corrected for this GPA categorization exceeds 97% for all items except item 22, which probed the grade that students expected or had received in a course. Differences in the response patterns for this item are, of course, expected. Effects of nonresponse on study conclusions attributable to student GPA are discussed at appropriate points in the sequel.

Figure 6 depicts histogram estimates of responses by academic division. Academic divisions represented in this plot are, from left to right, Social Sciences, Engineering, Humanities, and Natural Sciences. According to this figure, response patterns for these groups

FIGURE 5

Comparison of DUET responses by GPA. For each response category and item, the vertical bars on the left represent responses obtained from students whose GPAs were less than 3.3, while the bars on the right depict responses obtained from students having GPAs greater than 3.3.

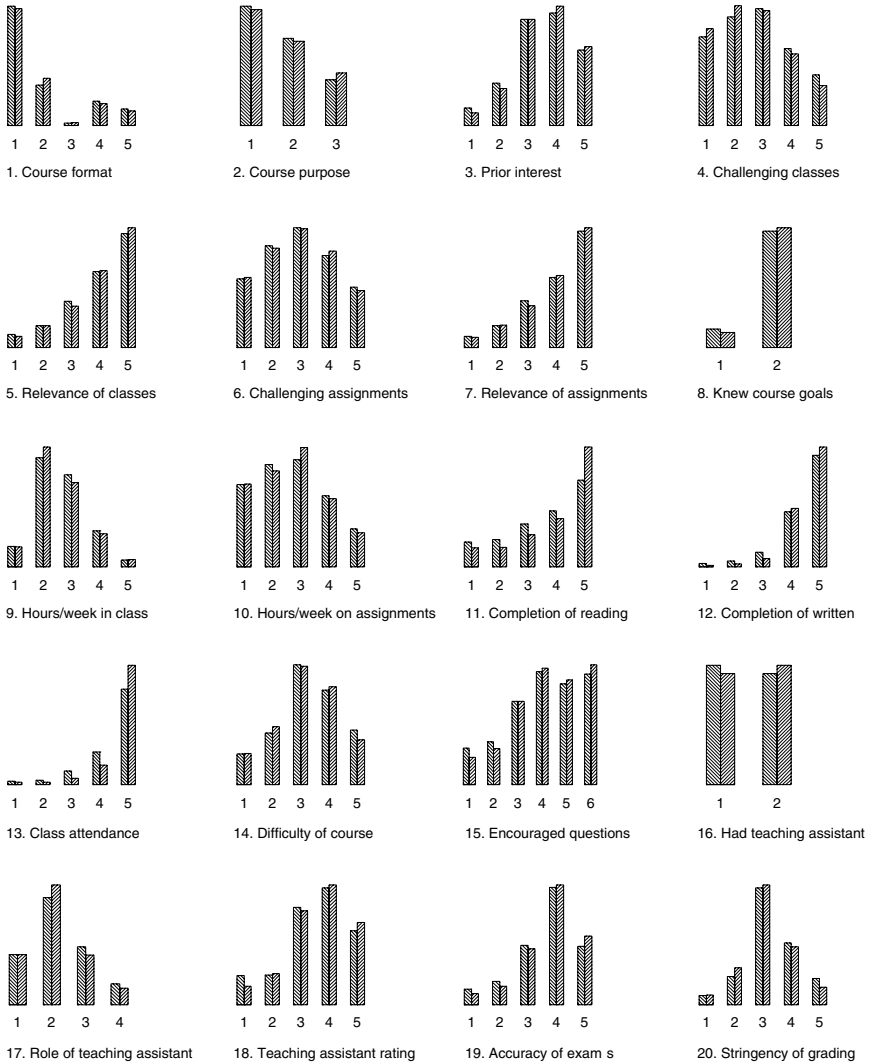


FIGURE 5 (continued)



FIGURE 6

Comparison of DUET responses by academic division. For each response category and item, the order of the vertical bars is Social Sciences, Engineering, Humanities, and Natural Sciences.

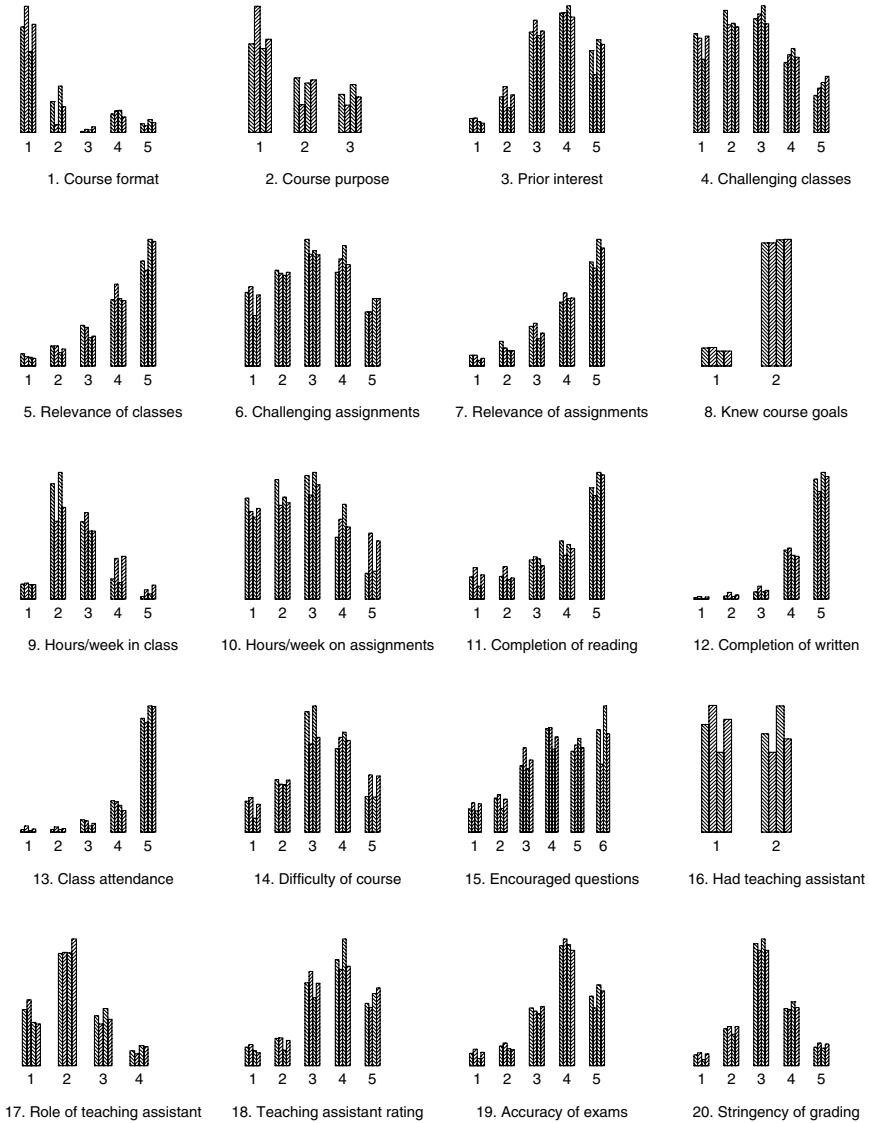
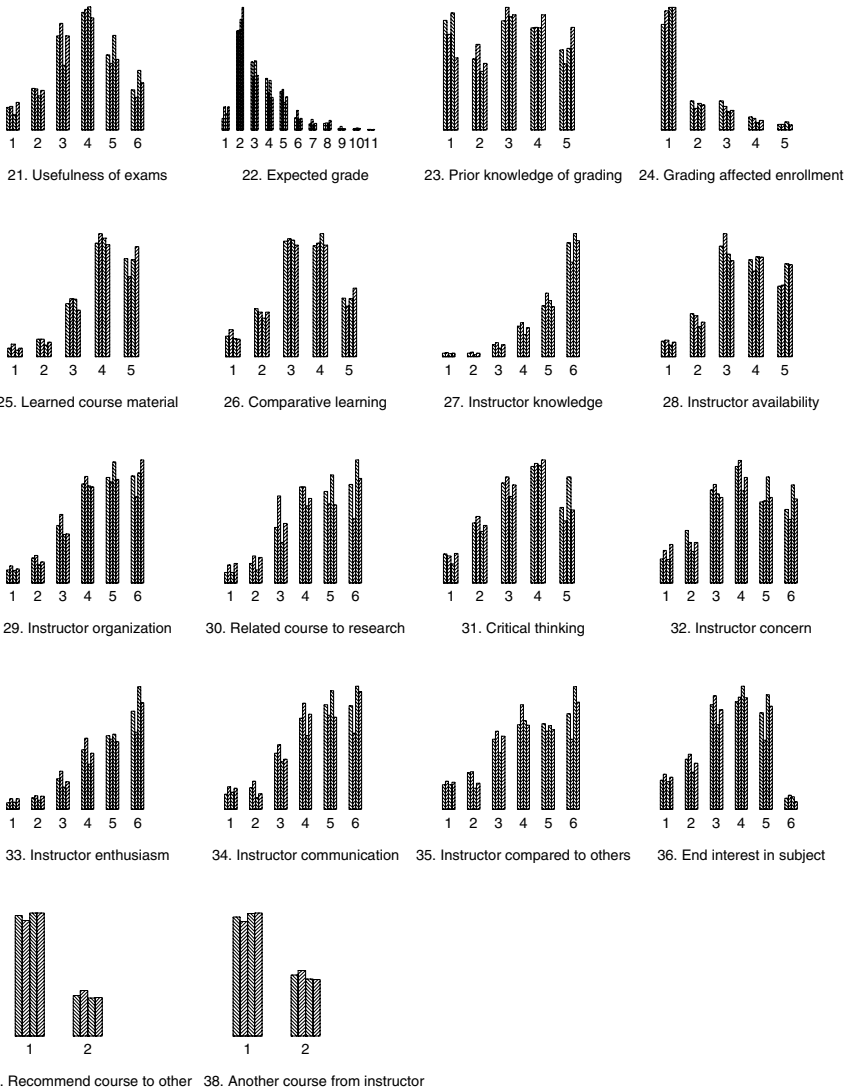


FIGURE 6 (continued)



are less consistent than were the patterns observed for groupings based on gender and ethnicity.

Engineering students tended to give higher responses on the items that probed hours/week on assignments (item 10), but lower responses to items probing prior subject matter interest (item 3), the extent to which instructors encouraged questions (item 15), usefulness of exams (item 21), prior knowledge of course grading policies (item 23), instructor knowledge (item 27), instructor organization (item 29), instructor's ability to relate course material to current research (item 30), instructor enthusiasm (item 33), and instructor communication (item 34). The response rates for engineers were also unusually high—39%—compared to the other groups, although they represented only a relatively small proportion of students eligible for the survey (15%, compared to 21% for humanities, 23% for natural and mathematical sciences, and 42% for social sciences). Because of the relatively large differences between the response patterns of students from the School of Engineering and those from Arts and Sciences, and their overrepresentation in the study sample, several later analyses were conducted without data collected from engineering students. In particular, several of the analyses conducted in Chapter 5 are limited to responses received only from Arts and Sciences students. Differences in the response patterns of social science, humanities, and natural science students were less significant.

Despite the differences between engineering students and non-engineering students, the observed distribution of responses and the distribution of responses corrected for the differential participation rates of students from different divisions were again quite similar. In this case, the probability of drawing the same response from both distributions exceeded 97% for all survey items.

Student responses broken out by academic year are depicted in Figure 7. Except for item 23, which probed student knowledge of course grading policies prior to enrollment, the response patterns by year group are also surprisingly similar. In the case of item 23, first-year students' responses were much higher in the first category, corresponding to little or no prior knowledge of grading policy. Because freshmen participants enrolled for their first-semester courses prior

FIGURE 7

Comparison of DUET responses by academic year. The order of the response proportions are, from left to right, seniors, juniors, sophomores, and freshmen.

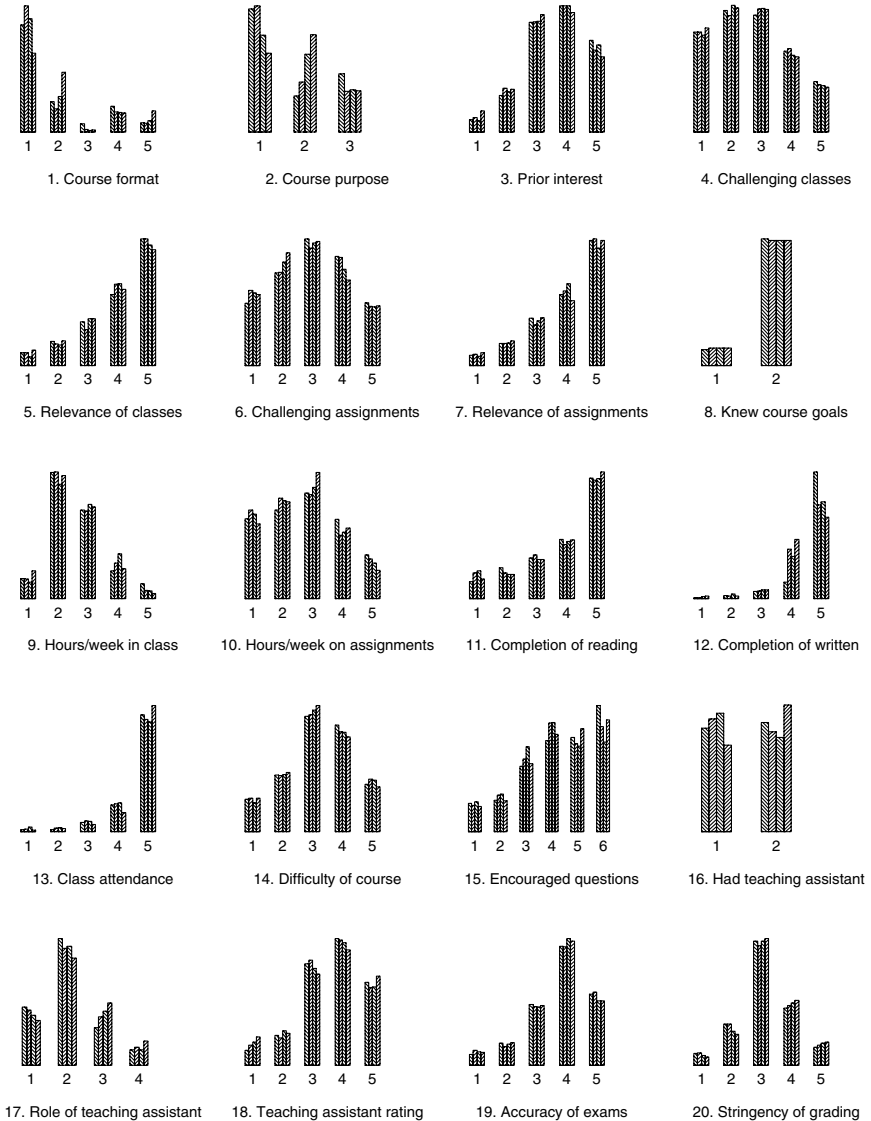
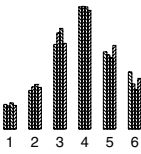
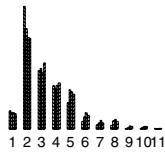


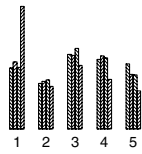
FIGURE 7 (continued)



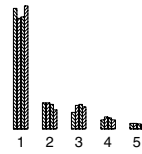
21. Usefulness of exams



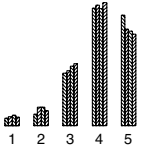
22. Expected grade



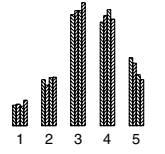
23. Prior knowledge of grading



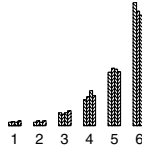
24. Grading affected enrollment



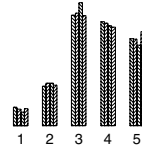
25. Learned course material



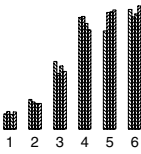
26. Comparative learning



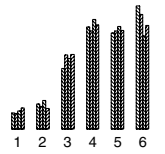
27. Instructor knowledge



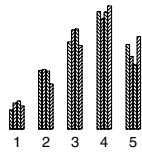
28. Instructor availability



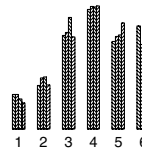
29. Instructor organization



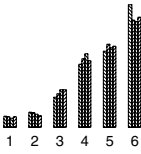
30. Related course to research



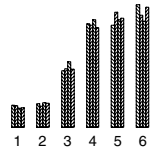
31. Critical thinking



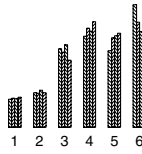
32. Instructor concern



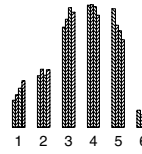
33. Instructor enthusiasm



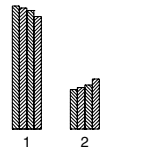
34. Instructor communication



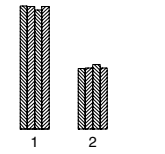
35. Instructor compared to others



36. End interest in subject



37. Recommend course to other



38. Another course from instructor

to arriving at Duke, the discrepancy for first-year students on item 23 is expected.

Despite the similarity of response patterns by year, it is important to note that the probability that a student participated in the experiment did vary significantly with academic year. Only 15% of fourth-year students completed the DUET survey, while 27% of third-year students did. Among second- and first-year students, participation rates were 39% and 35%, respectively.

Differences in response rates between year groups can probably be attributed to “intellectual dropout.” As a class of students moves closer to graduation, it seems that an increasing proportion of students lose interest in academics. Typically, these students either turn their attention to life after college or focus on nonacademic activities. The decreasing trend in the participation rates of students by academic year may simply be a consequence of increasing apathy. Conversely, the similarity of responses obtained from student participants from different academic years can probably be attributed to the fact that those students who did choose to participate in the survey were the same students who also chose to remain academically engaged.

Nonresponse attributable to student apathy or intellectual dropout does present a potential threat to the generalizability of survey conclusions. Because nonacademically oriented students were probably underrepresented in the experiment, and because there is no clear measure of students’ academic motivation in records provided by the registrar, corrections for this response mechanism cannot be made. Intuitively, however, it seems reasonable to assume that less academically oriented students are probably swayed more by grading considerations when choosing their classes and rating their instructors than are more serious students. Thus, in terms of impact on study conclusions, this source of response bias likely means that conclusions based on the DUET data are *conservative* in the sense that the measured effects of grades on student course selection and student evaluations of teaching would have been more pronounced if the sample had been more representative of the student body at large.

Of course, the conservative effect of student apathy and intellectual dropout on nonresponse must be balanced against the possibility that those students most interested in selecting courses according to grade distributions were overrepresented in the survey sample. Although the letter that solicited student participation on the survey was worded so as to emphasize the pedagogical benefits of collecting teacher–course evaluation data on-line, the fact that mean course grades could be viewed on the DUET website was also prominently featured. Whether this effect was more important than the opposing effect of student apathy and intellectual dropout is an open question. However, even if grades had absolutely no influence on the behavior of every student who chose not to participate in the survey, the impact of grades on students as measured by the DUET data is still highly significant; study conclusions based on the DUET data unquestionably apply to nearly 30% of all Duke undergraduates, and nearly 40% of first- and second-year students.

Finally, the plots and quantitative summaries discussed in this appendix suggest that nonresponse associated with known demographic variables can largely be ignored. Effects of other sources of nonresponse are more difficult to assess, and arguments for both positive and negative biases on study conclusions are plausible. However, until compelling arguments for biases in either direction can be substantiated, it seems reasonable to analyze survey responses as they were collected, and to regard conclusions based on these analyses as being approximately representative of conclusions that would have been drawn using a less self-selected survey sample.