

Technical and Logistical Support for Course Assessment

by

Simone N. Pugh

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Electrical Engineering

and Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

January 17, 2003

Copyright 2003 Simone N. Pugh. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and
distribute publicly paper and electronic copies of this thesis
and to grant others the right to do so.

Author _____

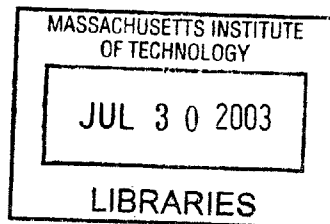
Department of Electrical Engineering and Computer Science
January 17, 2003

Certified by _____

Albert R. Meyer
Thesis Supervisor

Accepted by _____

Arthur C. Smith
Chairman, Department Committee on Graduate Theses



BARKER

Technical and Logistical Support for Course Assessment
by
Simone N. Pugh

Submitted to the
Department of Electrical Engineering and Computer Science

January 17, 2003

In Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science in Electrical Engineering
and Master of Engineering in Electrical Engineering and Computer Science

Abstract

Course assessment is a valuable tool for improving teaching and learning in a class. As part of the course assessment project being started for 6.042 and several other EECS classes, students took diagnostic entry exams and end-of-term surveys in which they were asked to rate themselves on various objectives of the class. I examined this data along with the students' grades, and performed calculations such as finding correlation coefficients to determine what areas of the course might need improvement. After analyzing four semesters of 6.042, it was discovered that the correlation coefficients were not as high as expected. However, the results are still useful in assessing the course, and with several improvements to the current assessment procedure, future results may be more clear.

Thesis Supervisor: Albert R. Meyer
Title: Hitachi America Professor of Engineering

Table of Contents

Abstract	2
Table of Contents	3
Table of Figures	4
Table of Tables	5
Introduction	6
Background	6
Procedure	9
6.042 Fall 2000 Term	11
6.042 Spring 2001 Term	11
6.042 Fall 2001 Term	13
6.042 Spring 2002 Term	13
Results	13
Correlation Coefficients of Self-Assessment Surveys	13
Comparing Self-Assessment Surveys to Grades	16
Diagnostic Exams	21
Discussion	22
Discussion of Procedure	22
• Problem Weightings Are Not Important	22
• Survey Medium: Paper vs. Computer	23
• The 1-5 Rating Scale	23
• Converting the 1-5 Rating Scale to a Percentage Scale	24
Correlation Coefficients of Self-Assessment Surveys	24
• Reasons for Low Correlation Coefficients	24
• Effect of the Rating Scale of Correlations	24
• Students' Attitude Toward the Survey	25
• Differences in Students' and Instructors' Perception of Class	25
Comparing Self-Assessment Surveys to Grades	26
• Graphs Comparing Self-Assessment to Class Grade	26
• Graphs of Self-Assessment by Grade Level	26
Diagnostic Exams	27
• Diagnostic Exams Did Not Always Cover Relevant Material	27
• Errors in Grading Diagnostic Exams Invalidate Results	27
• Correlation Coefficients Related to Diagnostic Exams Were Not Useful	27
Conclusions and Recommendations	28
• Low Correlation Coefficients Should Be Expected	28
• List of Outcomes Should Not Change Between Semesters	29
• Improving the Rating Scale	29
• Suggestion for Giving Midterm Surveys	29
• Improvements to the Diagnostic Exam	29
• Improved Response Rate Needed	30
References	31
Appendix A: 6.042 Course Objectives and Outcomes	32
Appendix B: 6.042 Exit Self-Assessment Survey	34

Table of Figures

Figure 1. Example of Weak Correlation: 6.042 Outcome 1, Spring 2002.	14
Figure 2. Example of Strong Correlation: 6.042 Outcome 3, Spring 2002.	15
Figure 3. 6.042 Average Grades and Self-Assessments by Outcome.	16
Figure 4. Average Grades and Self-Assessments by Outcome, 6.042 Fall 2000.	17
Figure 5. Average Grades and Self-Assessments by Outcome, 6.042 Spring 2001.	17
Figure 6. Average Grades and Self-Assessments by Outcome, 6.042 Fall 2001.	18
Figure 7. Average Grades and Self-Assessments by Outcome, 6.042 Spring 2002.	18
Figure 8. Self-Assessment Averages by Grade Range, 6.042 Fall 2000.	19
Figure 9. Self-Assessment Averages by Grade Range, 6.042 Spring 2001.	20
Figure 10. Self-Assessment Averages by Grade Range, 6.042 Fall 2001.	20
Figure 11. Self-Assessment Averages by Grade Range, 6.042 Spring 2002.	21

Table of Tables

Table 1. Correlation Coefficients between Grade and Self-Assessment for 6.042, by Outcome.	13
Table 2. Overall Correlation Coefficients by 6.042 Term.	15
Table 3. Data for 6.042 Spring 2001 Entry and Exit Exams.	21
Table 4. Correlation Coefficients between Entry Exam and Final Grade, 6.042 Spring 2002.	22

Introduction

As part of a system to assess teaching and learning at MIT, some classes in the Electrical Engineering and Computer Science department have begun to gather and analyze data such as class grades, student self-assessment surveys, and diagnostic entry exams. We need a standard system to analyze this data, and we need to review the effectiveness of the current system and see how it can be improved. In my project, I will review the current assessment system being used by several EECS classes. I will carry out parts of this system by performing analyses of the data gathered from various classes, such as determining the correlations of entry exam scores with class grades and self-assessments with class grades, and will determine what analyses are useful and can provide important information for teachers. By examining these results, I will make conclusions about what data should be gathered and which are the optimum assessment methods. Finally, possible improvements to the current system will be suggested.

Background

The primary purpose of developing an assessment system is to provide feedback to students and teachers in order to improve the quality of learning and teaching. Outcomes-based assessment is a system in which the teacher creates a list of outcomes or objectives that represent the crucial concepts students should learn throughout the course [1]. This allows teachers to define and formalize what they are trying to teach and what course goals they consider important to learn. Once these goals are specified and made public, it makes it easier to define what students are and are not learning. Teachers can see how well students are mastering the work all throughout the course, not just at the end. The students can also self-assess their own learning to discern what concepts they need help with. Students' awareness of their own learning progress is increased; they can better understand what they have learned in a class, and will be able to specifically communicate to others what the class was about and what they came out of it knowing [1]. Assessment can also be used to evaluate teaching effectiveness, and to create or revise a curriculum and determine the effects of curriculum changes.

An assessment system is different from looking only at students' final class grades because assessment can take into account the students' understanding of many individual aspects of the course instead of the course as a whole. Teachers can use it to see what students came into the class already knowing and what they remember after the course has concluded. However, grades on course work can still be used to assess teaching and learning by mapping the questions on homework and exams to the list of course outcomes. This system of "embedded assessment"

[1] is currently being used to some extent in a few classes in the EECS department at MIT: 6.001, 6.002, 6.003, 6.004, 6.012, 6.034, 6.046, 6.042, 6.045, 6.101, 6.111, and 6.170. As well using outcomes assessment, some classes have developed other procedures and collect data about different things, such as attendance, prerequisites, study habits, etc.

There are also many other colleges that have their own assessment programs, gathering data on various things from quality of student life to teaching effectiveness. At the Georgia Institute of Technology, students are surveyed on a wide variety of things. There is a freshman demographics survey, a survey asking about student involvement, aspirations, and cognitive effect and development, and a senior survey focusing on knowledge and skills gained, student opinions of faculty and advising, and the factors that influenced personal and educational gains. These are just a few of the things Georgia Tech assesses through their program [2]. Rose-Hulman gives an entry survey focusing on students' educational and social backgrounds, as well as a Mechanics Baseline test covering the basic concepts of mechanics that students are expected to know before entering the program. They also administer a senior survey asking students about the curriculum, their experiences, and their views on teaching [5]. North Carolina State University employs student outcomes assessment to survey students' retention of material, graduation rates, satisfaction with their academic programs, and skills learned. Students do course evaluations and are also asked about their goals, involvement in activities, previous employment, and similar things. Like many other schools, North Carolina State University surveys new and graduating students, but they also survey second term sophomores, alumni, and employers of alumni to get a thorough idea of the effects of their education [4]. These are just examples of a few of the many assessment programs in use at other colleges.

Out of the schools I looked into, MIT and North Carolina State University are the only ones that currently give students a self-assessment type survey on which they are asked to rate themselves on certain outcomes. The NCSU surveys, however, were given to sophomores and graduating seniors and asked students about the effectiveness of their program as a whole in helping them achieve the outcomes. The self-assessment surveys were not, as far as I could tell, given to students after they completed a certain class. At NCSU, the survey results were used to tell instructors which areas of the curriculum were effective and which needed improvement [4].

In the summer of 2000, Professor Meyer and several other MIT professors wanted to develop a plan for ongoing assessments of their teaching. Their procedure began with writing a list of outcomes and objectives for each of their courses; this is given to the students at the start of the semester to show them what they are expected to learn throughout the course. The objectives and outcomes list used in the last three terms of 6.042 is shown in Appendix A. At the beginning

of the course, the students took a non-credit diagnostic entry exam to provide the teachers with an idea of what the students' backgrounds were and how effective the prerequisites had been in teaching background material. Throughout the course, each of the homework, quiz, and final exam problems for the class were mapped to one or more of the outcomes and students' problem-by-problem grades were recorded. This allowed teachers to attain students' scores for each outcome and see what outcomes were or were not learned well in general. At the end of the semester the students took a self-assessment survey that asked them, among other things, to rate how well they thought they understood each of the outcomes. The survey given to students in 6.042 is shown in Appendix B. Their responses on the survey were then correlated with the students' grades for each outcome. Following this, the teachers could then meet to discuss the process and suggest changes and improvements.

Aside from 6.042, several other EECS classes at MIT have developed their own similar assessment systems. In 6.001, students were surveyed about topics such as how useful they found various aspects of the class (text book, lectures, homework, etc.), what their preferred form of lecture presentation was, how many hours they spent working on different assignments and in class, and how they felt about online lectures and problem sets. The instructors made charts and graphs showing the averages of the students' responses on the different questions. Students were also asked to rate themselves on the different outcomes and objectives of the class. These ratings were separated and averaged for certain grade levels (students whose grade was 60%-70%, 70%-80%, 80%-90%, and 90%+) to see how the average self-assessments varied by grade level. It was also recorded how many students rated themselves a 5, 4, 3, 2, and 1 for each objective and outcome. Surveys in 6.004 and 6.034 asked about the same types of things, and the results that were compiled had similar graphs and charts. The 6.034 results also showed bar graphs of attendance and suggested hours to spend on certain areas of the class. Overall, the procedure being used in these three classes is quite similar to the one we are developing for 6.042.

A variety of issues must be considered when creating an assessment system. Aside from the obvious question of what information to gather, one has to decide when and how often to assess. Students can be tested or surveyed at the beginning and middle of a course and all throughout the course if the teacher thinks the information provided by those surveys will be useful. For instance, one thing MIT is currently doing in some courses is administering a diagnostic entry exam to see how much of the course material and prerequisites students come into the course knowing. This could be useful in determining how well the prerequisites have prepared students for the class. Another possible time to survey students is after graduation or after the completion of a course. Some schools do this to measure students' retention of the

material they learned or the effectiveness of a course in preparing students for future courses or employment. A survey of this type was tried one semester in 6.042; an exit exam which contained the same questions as the entry exam was given at the end of the semester to measure how much students learned through the course. Some things to be considered in giving surveys are how to motivate students to take them, whether giving them to students is feasible or not, whether enough students will take them for the results to be reliable, and how the results can be used to improve the course and increase students' retention. By studying these issues and looking at how useful this type of survey has been in other schools, we can determine if it is something that should be tried.

Another issue that has come up is what medium to use to give assessment surveys to students. Some assessment surveys have been done over the computer and some have been given on paper. There are advantages and disadvantages to each method, but we need to find out which is more reliable and which gets more responses. Another thing to consider is whether to use the students' grades on homework and exams to assess their learning, or whether to use a separate test or survey. Issues with course work such as the possibility of collaboration between students could distort the data if this method is used, but separate assessment tests probably would not be as thorough or as easy to administer. Finally, we need to decide which results are reliable and which provide enough information to be used.

Procedure

The materials needed to start with for a particular course were a list of students homework and exam grades broken down by problem, copies of all the homework and exam problems, a list of the course objectives and outcomes, and the results of the entry or exit surveys given to students. The homework and exam problems were then mapped to one or more of the course outcomes, and this and the other data were entered into an Excel spreadsheet. Weights for the homework and exam problems were assigned based on how much that particular course dictated each to be worth. Then, by multiplying the students' grades on homework and exam problems with the appropriate weight and adding them up by outcome, the students' overall grade for each outcome was calculated. This grade was compared with their scores or responses on entry and exit surveys by calculating the correlation coefficient between the sets of data, for each outcome and overall. Scatter plots including linear best-fit lines were also created to show the relationship between the two sets of data [6] [7] [8].

After obtaining the correlation coefficients, I went back to see what effect the weighting system had on the results. In the Spring 2002 term of 6.042, for instance, the twelve homework

sets were worth 30% of the students' final grade combined, the two exams 25% combined, and the final exam 20%. However, most homework sets had up to eight problems each, whereas the exams had three or four and the final exam fifteen. This placed heavy emphasis on the exams and the final exam, so that the students' score on exam questions nearly dominated their score on a particular outcome. Other items were included in the students' grade in the course, such as tutorial problems and readings, but were not included in this analysis. These items were things that were either done more as a group instead of individually, did not relate to a specific outcome, or were not a good indicator of the students' knowledge. I experimented with different weightings, such as having every problem (homework, exam, and final exam) worth the same amount, or the exam problems worth twice as much as homework problems and final exam problems worth twice as much as regular exam problems. For each weighting system, I recorded the correlation coefficients and noted whether they changed significantly. The weighting system I eventually decided to use was one in which the homework questions were worth one point, the exam questions three points, and the final exam questions five points.

For the Fall 2000 and Spring 2002 terms of 6.042, the rating system students were asked to use to rate themselves in the end-of-the-year self-assessment surveys was misunderstood by a number of students. A 1-5 scale was used in which 1 corresponded to an excellent rating and 5 corresponded to a poor rating. This scale is reversed compared to the scale students seem to be accustomed to using on similar surveys and evaluations, leading around 10% of the students who answered the survey to misunderstand the scale and use 5 to represent an excellent rating and 1 a poor rating. To determine how large an effect this had on the results, I first had to sort through the surveys and find the ones that appeared to have been reversed. If the student's comments about the class suggested that he understood the material very well, but he rated himself poorly (4's or 5's) for most outcomes, this suggested that the student accidentally reversed the scale. For these students, I switched the scale back to what it was supposed to be and then compared the resulting correlation coefficients to the original ones. If a student made no comments and there was nothing else to compare his ratings against, nothing could be done about possible reversed ratings so the student's responses were not changed.

Based on a graphs made by the 6.001 staff in the analysis of similar 6.001 surveys, I also made bar graphs for each semester showing the average self-assessment for each outcome for different grade levels. I divided the students up by their total grade on each outcome into groups whose grade was below 60%, 60-75%, 75-85%, and over 85%. I decided on those groups because they usually led to similar numbers of students falling into each level. I felt these graphs were a little easier to understand than a correlation coefficient because it shows whether the self-

assessments were distributed as they should be (with the above 85% group having the highest and the below 60% group the lowest), or whether there was a significant amount of confusion in the students' ratings.

So far, data has been gathered and analyzed for four terms of 6.042: Fall 2000, Spring 2001, Fall 2001, and Spring 2002. The analysis done for each term varied slightly depending on what surveys the instructors gave the students.

6.042 Fall 2000 Term

In this term of 6.042, a mandatory self-assessment survey was given on paper at the end of the year. It was not mandatory for students to write their names on their survey. Out of the 88 students in the class for whom problem-by-problem grades were available, 65 completed the survey and included their name on it. There were actually a total of 98 students in the class, but problem-by-problem grades were not recorded for the ten students in one particular tutorial section. Those students were left out of the analysis I did.

First I looked at all the exams and homework problems and mapped them to one of the nine course outcomes. For the 64 students who returned surveys with their names on them, I found the correlation coefficients between the students' overall (homework, exams, and final exam) grade in the course and their self-assessment, broken down by outcome. The students' grades were computed as a percentage of the total number of points possible for each outcome. I also made a scatter plot of these values for each outcome. This was one of the terms in which the rating scale was "backward," so I changed the ratings for four students who appeared to have reversed the scale and then found the new correlation coefficients. To incorporate all the students into the analysis, I calculated the average final grade and average self-assessment of all the students for each outcome and found the correlation coefficient between those two sets of data. Finally, I created a bar graph of the students' average final grades and their average self-assessments for each outcome. This allows us to see which outcomes students did and did not do well on, and which outcomes had the largest discrepancies between the students' grades and self-assessments.

6.042 Spring 2001 Term

Along with the end-of-the-year self-assessment survey, entry and exit diagnostic exams were given in this term. The diagnostic exams were mandatory and names were collected for all the students. The exams given at the beginning of the term and at the end were identical; the objective was to see how much students came into the course knowing and how much the course

helped them improve. The self-assessment survey, given on paper this term, was not mandatory and was only completed by 34 out of the 139 students in the class.

The first sets of correlation coefficients I looked at were between the self-assessments and the final course grades, the entry diagnostic scores and the grades, the exit diagnostic scores and the grades, and the entry diagnostic scores and the change in scores between the entry and exit diagnostic exams. These were all broken down by outcome. In terms of the self-assessment survey, I did the same analysis as for the Fall 2000 term, except that for the Spring 2001 term the rating scale used was different so there was no confusion about it as there was for the previous term. I started using the same procedure for the analysis involving the entry and exit diagnostics, but breaking these scores down by outcome did not seem appropriate. The diagnostic only covered five of the eight outcomes, with some questions not really pertaining to any of the outcomes, and some outcomes having only one or two questions related to them.

Instead of breaking down the scores by outcome, I looked at the students' scores on each question on the diagnostic. To see how much the course helped students learn the material and improve, I found the correlation coefficients between the entry diagnostic score on each question and the change in score between the entry and exit diagnostics. The change in score was calculated as the amount a particular student's score increased or decreased divided by their score on the entry exam. So a student who scored an 80% on the entry exam and a 75% on the exit exam would have a score change of $-5/80$. The correlation coefficients were expected to be negative, suggesting that students who scored low on the entry diagnostic would improve their score more than students with high entry diagnostic scores. The hope was that, after taking the course, all the students' exit diagnostic scores would be shifted toward the high end of the scale. I also looked at the mean and variance of the students' score change to see how much they improved after taking the course. Another correlation coefficient I looked at was between the entry and exit diagnostic scores, to see if people who came into the course knowing more still knew more at the end of the course.

The diagnostic exam given in that term was not as thorough as some of those given in other terms. One question that arose was how good of a measure of students' knowledge the diagnostic was. To test this, I looked at the correlation coefficient between the students' total final exam scores and their total exit diagnostic scores. A low correlation coefficient there would suggest that the diagnostic was not very accurate in measuring students' knowledge of the material considered important in the course.

6.042 Fall 2001 Term

The analysis performed for this term was almost identical to that from Fall 2000. In the Fall 2001 term, there were 97 students in the class and 41 answered the optional self-assessment survey. The survey was given on the web through the 6.042 web tutor.

6.042 Spring 2002 Term

In this term, a self-assessment survey and an entry diagnostic survey were given to the students. I found the correlation coefficients between the final grades and the diagnostic scores, and between the final grades and the self-assessment ratings. These were all broken down by outcome. 121 students took the class, and self-assessment surveys were returned by 66 students who included their name and 87 students in all. The self-assessment survey was given on the web through a web form, and the diagnostic was given through the 6.042 web tutor. I also found the overall correlation coefficients not broken down by outcome between the self-assessments and the course grades.

Results

Correlation Coefficients of Self-Assessment Surveys

Table 1. Correlation Coefficients between Grade and Self-Assessment for 6.042, by Outcome.

	Fall 2000	Spring 2001	Fall 2001	Spring 2002	average
1. sets, relations, functions, integers	0.127455	n/a	0.241062	0.065964	0.13094
2. fallacious reasoning	0.310449	0.106389	0.063896	0.071778	0.151231
3. proofs	0.258562	0.241472	0.329096	0.51341	0.35143
4. graph theory	0.278946	0.358539	0.442968	0.360218	0.350766
5. invariants, well-founded ordering	0.172882	0.302488	0.325979	0.171039	0.224154
6. closed form and asymptotic expressions	0.509664	0.119139	0.19245	0.305173	0.316557
7. number of possible outcomes	0.32937	0.420405	0.355605	0.482346	0.398628
8. probabilities	0.426466	0.311085	0.376224	0.327327	0.36566
9. deviation from the mean	0.255577	0.321576	0.202012	0.2727	0.261295
number of named responses returned	65/98	34/139	41/97	66/121	

Table 1 shows the correlation coefficients between the students' final grade in 6.042 and their self-assessment for each of the nine outcomes. The values shown were found by listing each student's grade along with their self-assessment rating, and then finding the correlation coefficient between the two for each outcome in each term. In the Spring 2001 term, there were only eight outcomes; outcome 1 did not exist. The fifth column shows the average correlation

coefficient of the four terms, weighted by how many students there was data for in each term. A “high” correlation coefficient is generally considered to be .5 or above [7] [8]; values this high are only seen twice in the table, and most are significantly lower than this. Outcomes 8 and 9 have the highest correlation coefficients on average, and outcomes 1 and 2 have the lowest. The following two graphs, Figures 1 and 2, are examples of the scatter plots made of self-assessments versus grades, showing all the individual points. Figure 1 is an example of an outcome for which there was only a weak correlation between self-assessments and grades (for outcome 1 in the Spring 2002 term) and Figure 2 shows a stronger correlation (outcome 3 in the Spring 2002 term) [8].

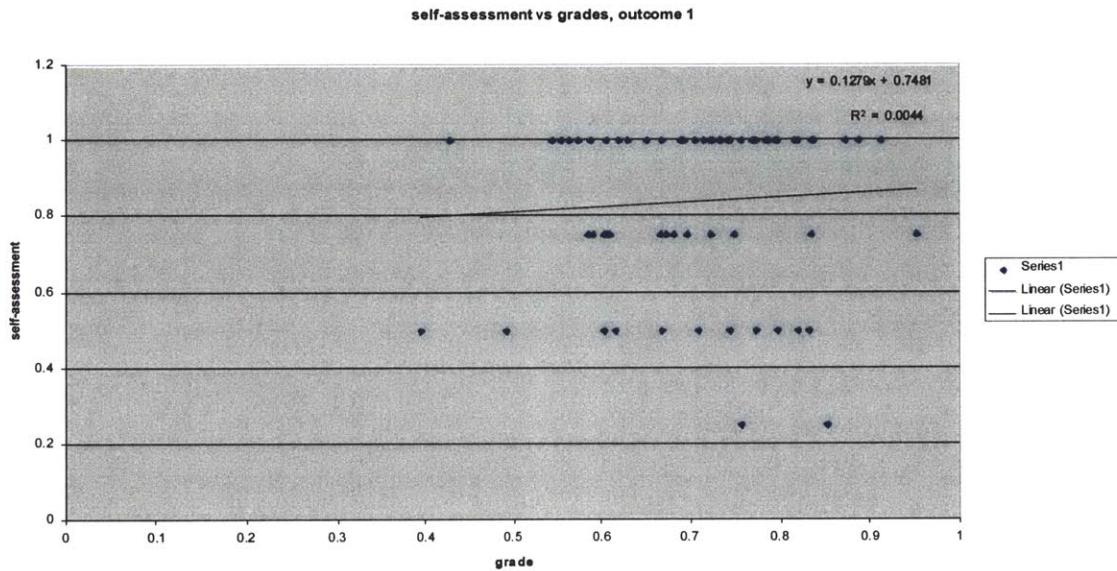


Figure 1. Example of Weak Correlation: 6.042 Outcome 1, Spring 2002.

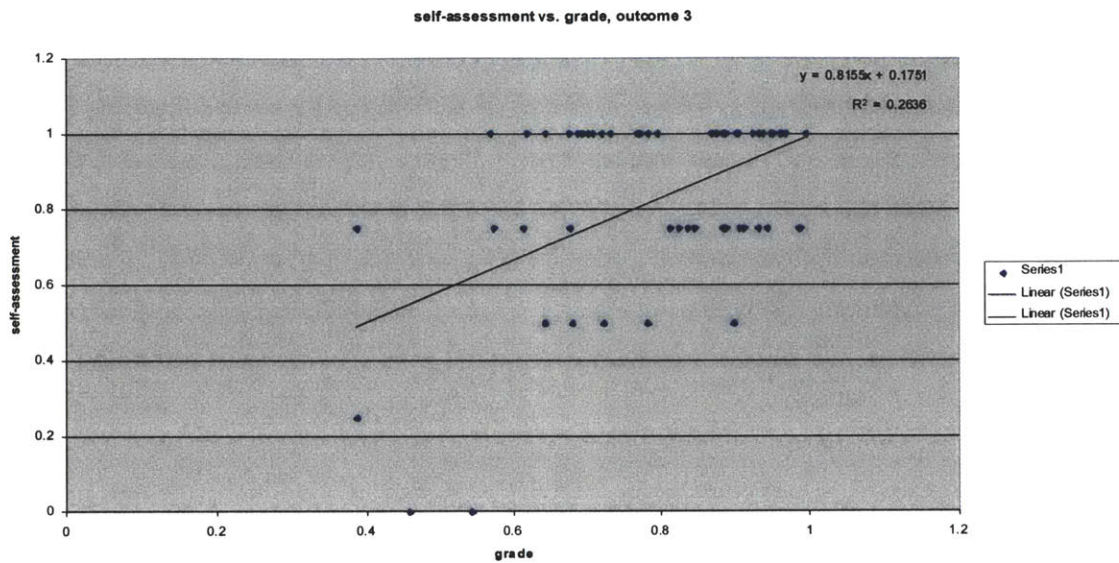


Figure 2. Example of Strong Correlation: 6.042 Outcome 3, Spring 2002.

Often the self-assessment surveys were not mandatory and as a result were not taken by all the students. Similarly, the surveys often did not require the students to include their names so a number of anonymous responses were handed in. In the previous table, this data was ignored because the calculations required that students complete a named survey. To be able to include the grades and survey responses of all the students, all their grades for a particular outcome were averaged and all their self-assessments for a particular outcome were averaged. This was done for all nine outcomes. Then the correlation coefficient between the average grade and the average self-assessment was found. The correlation coefficients are shown in Table 2.

Table 2. Overall Correlation Coefficients by 6.042 Term.

Term	Correlation Coefficient
Fall 2000	-0.45512
Spring 2001	0.415378
Fall 2001	0.11374
Spring 2002	0.751483

Comparing Self-Assessments to Grades

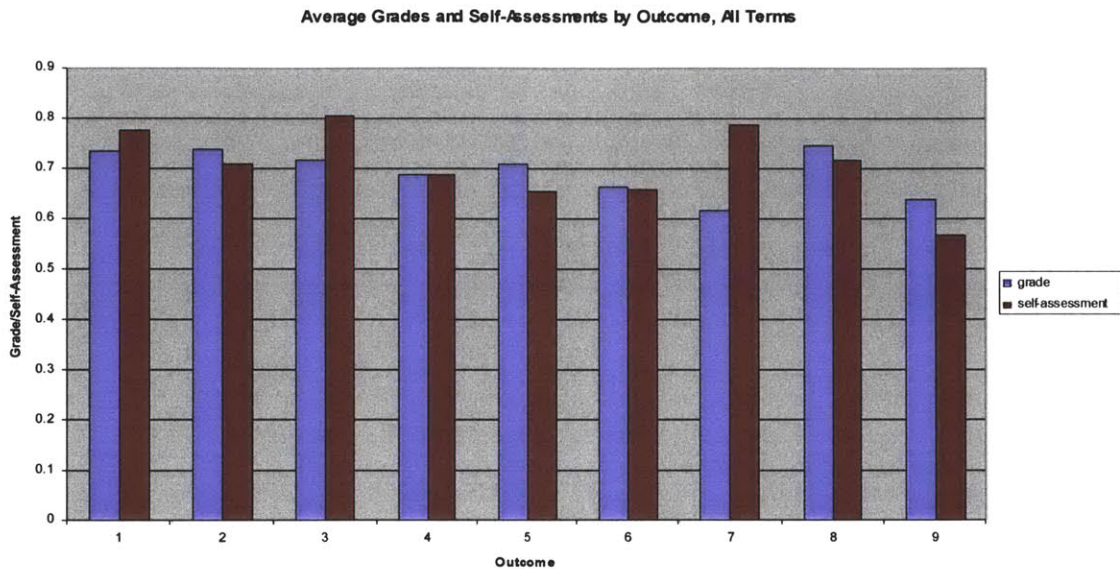


Figure 3. 6.042 Average Grades and Self-Assessments by Outcome.

Figure 3 compares the students' grade with their self-assessment for each outcome for the four terms combined. The chart shows that students are almost equally likely to overrate themselves as they are to underrate themselves. Outcomes 3 and 7 have the largest discrepancy between the grade and the self-assessment. Overall, the average difference between the grade percentage and the self-assessment percentage is only 1.3%. However, this calculation does not take into effect the fact that the difference can be either positive or negative. Looking at the absolute value of the differences in percentages, the average becomes 5.4%. Still, because this is an average of all four terms, it is more meaningful to look at the results for each term separately. This is shown in Figures 4-7.

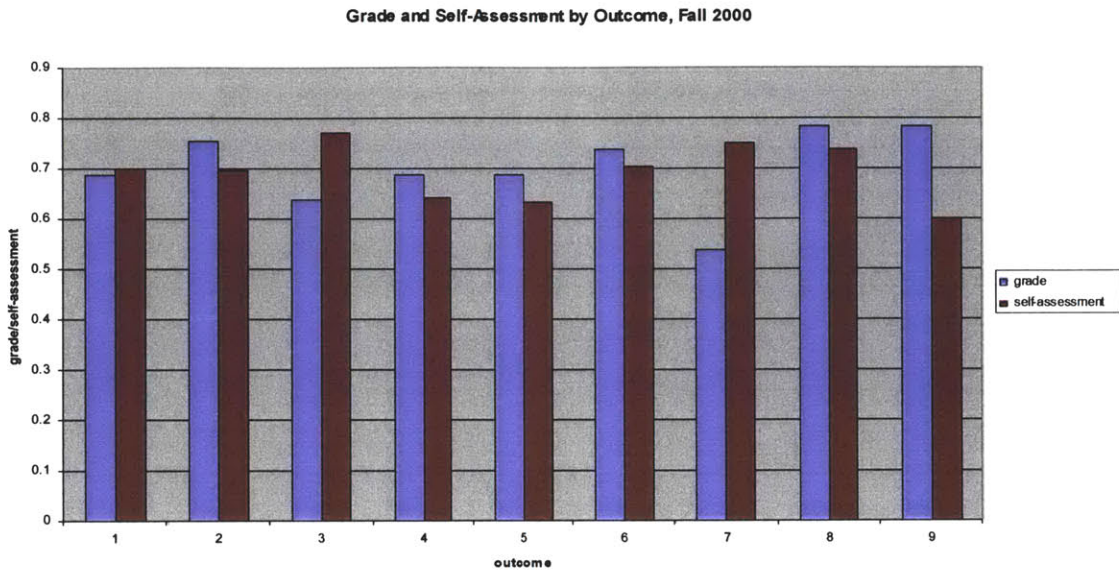


Figure 4. Average Grades and Self-Assessments by Outcome, 6.042 Fall 2000.

For the Fall 2000 term, the average of the absolute value of the differences is 8.7%. Outcomes 7 and 9 show the largest discrepancy.

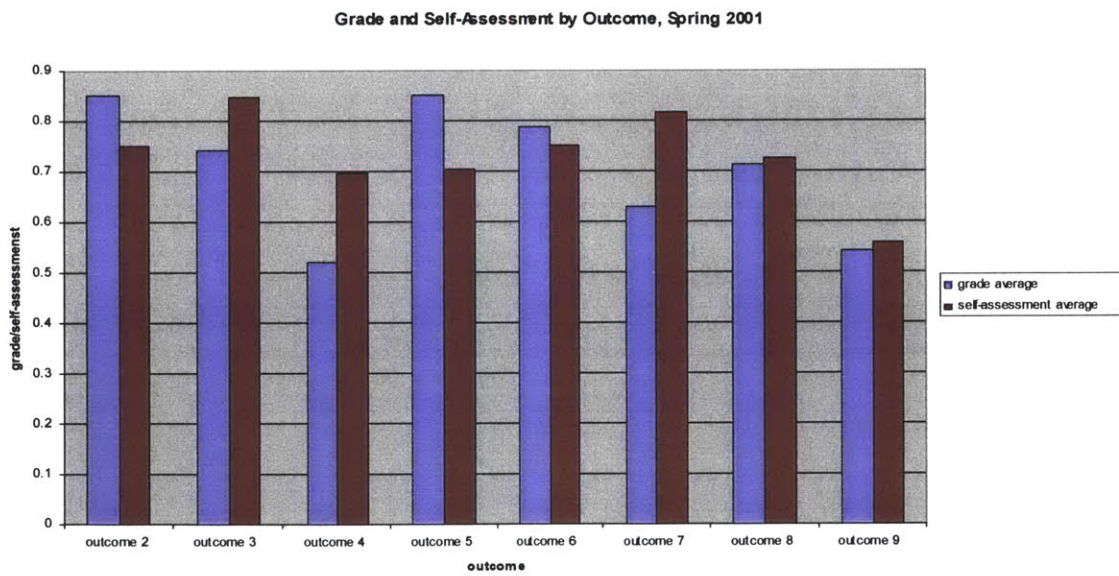


Figure 5. Average Grades and Self-Assessments by Outcome, 6.042 Spring 2001.

In Spring 2001, the average difference is 9.7%, with the largest discrepancy occurring in

outcomes 4 and 7.

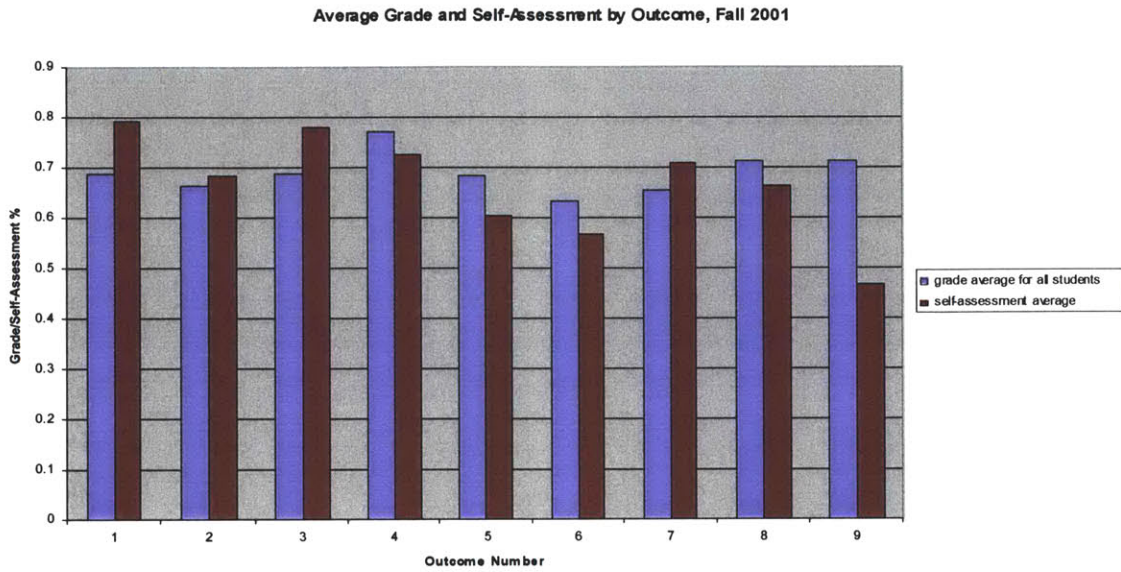


Figure 6. Average Grades and Self-Assessments by Outcome, 6.042 Fall 2001.

The average difference in Fall 2001 is 8.4%, and outcomes 1 and 9 show the largest discrepancy.

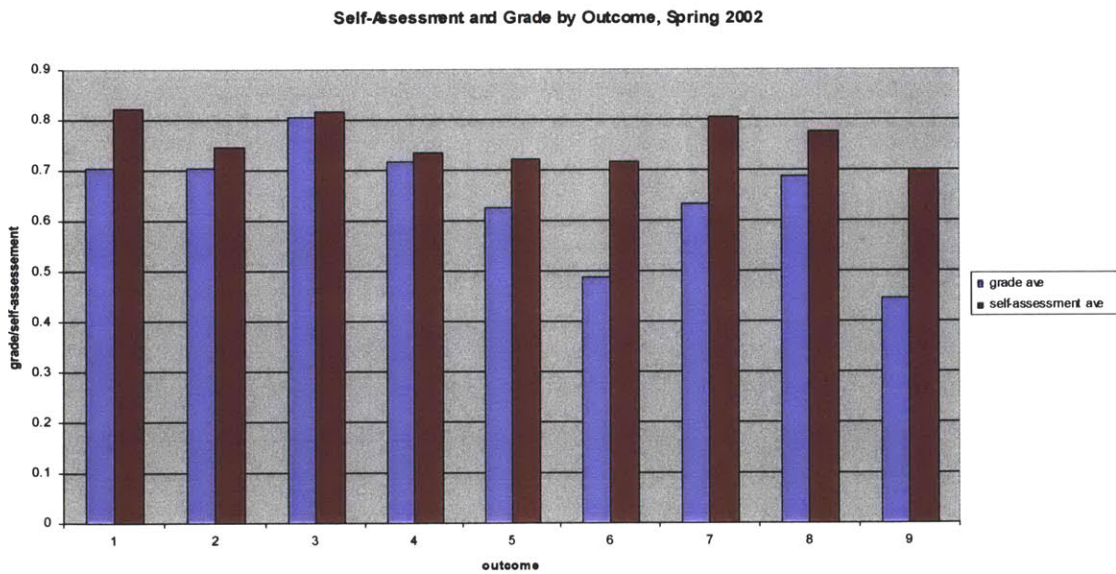


Figure 7. Average Grades and Self-Assessments by Outcome, 6.042 Spring 2002.

In Spring 2002 the average difference is 11.4%, with outcomes 6 and 9 having the largest discrepancy. Another thing to note for this term is that the students overrated themselves on all the outcomes.

The following four graphs, Figures 8-11, show the average self-assessment for each outcome, broken into groups by grade. The grades were divided into four groups: under 60%, 60-75%, 75-85%, an above 85%. The overall average self-assessment for each outcome is also shown. For graphs in which one column is missing, such as the “under 60%” group for outcome 2 in Fall 2000, there were no students whose grade fell into that percentage range.

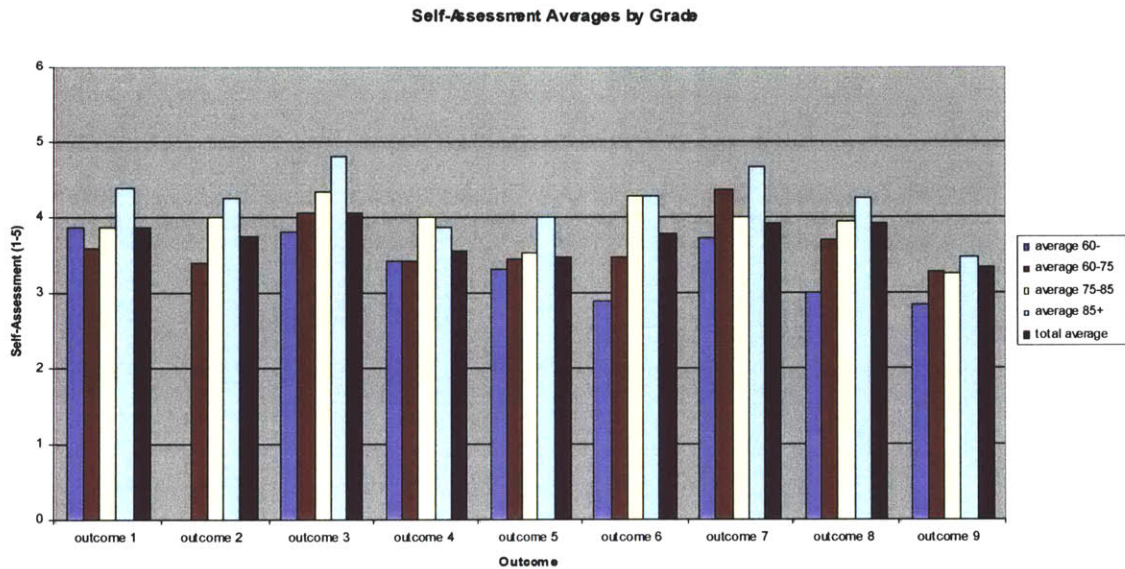


Figure 8. Self-Assessment Averages by Grade Range, 6.042 Fall 2000.

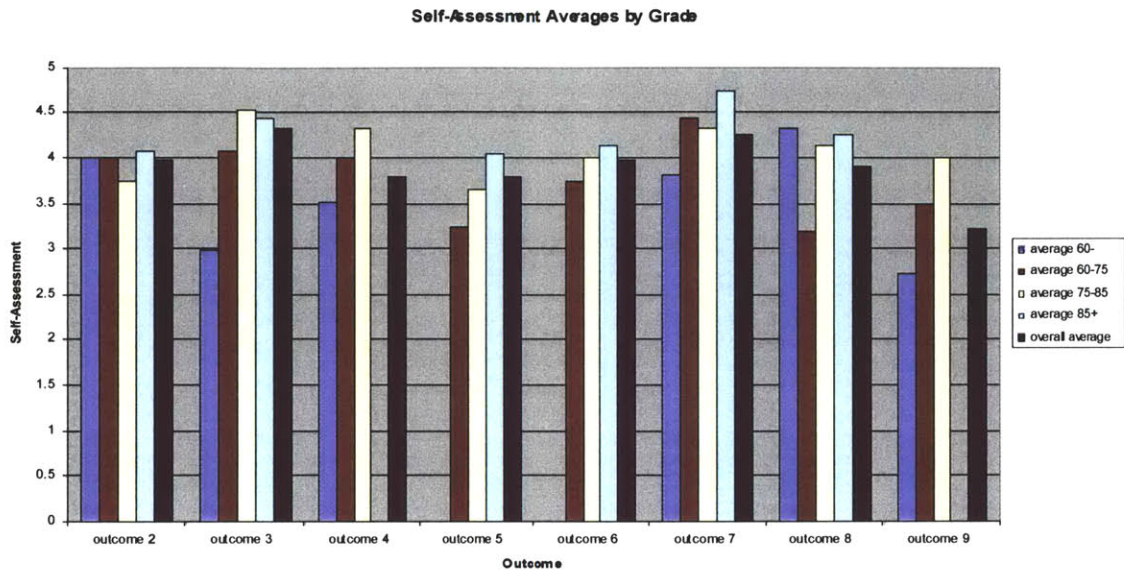


Figure 9. Self-Assessment Averages by Grade Range, 6.042 Spring 2001.

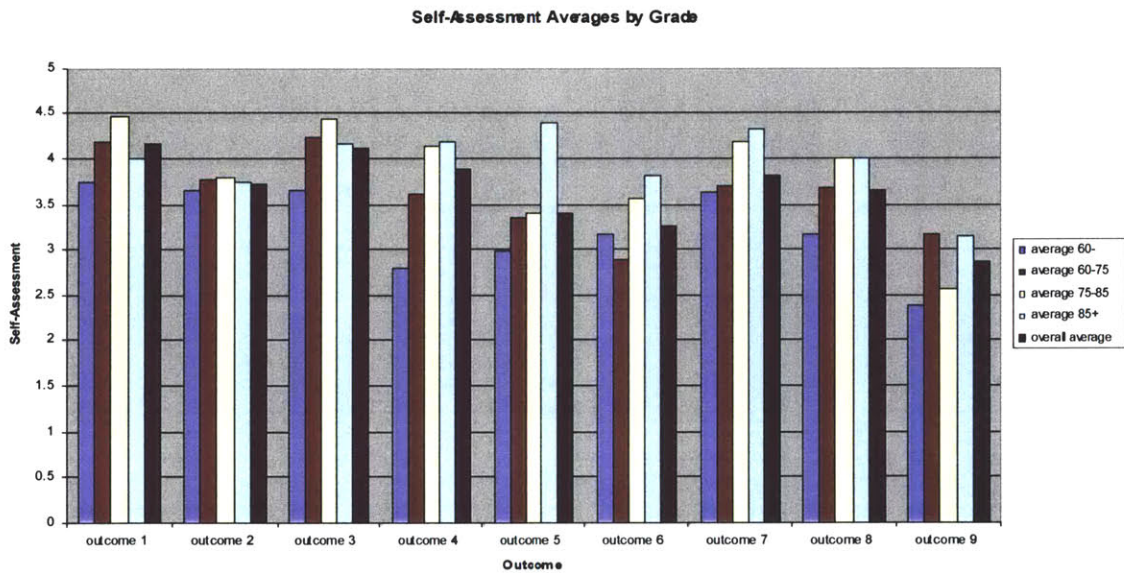


Figure 10. Self-Assessment Averages by Grade Range, 6.042 Fall 2001.

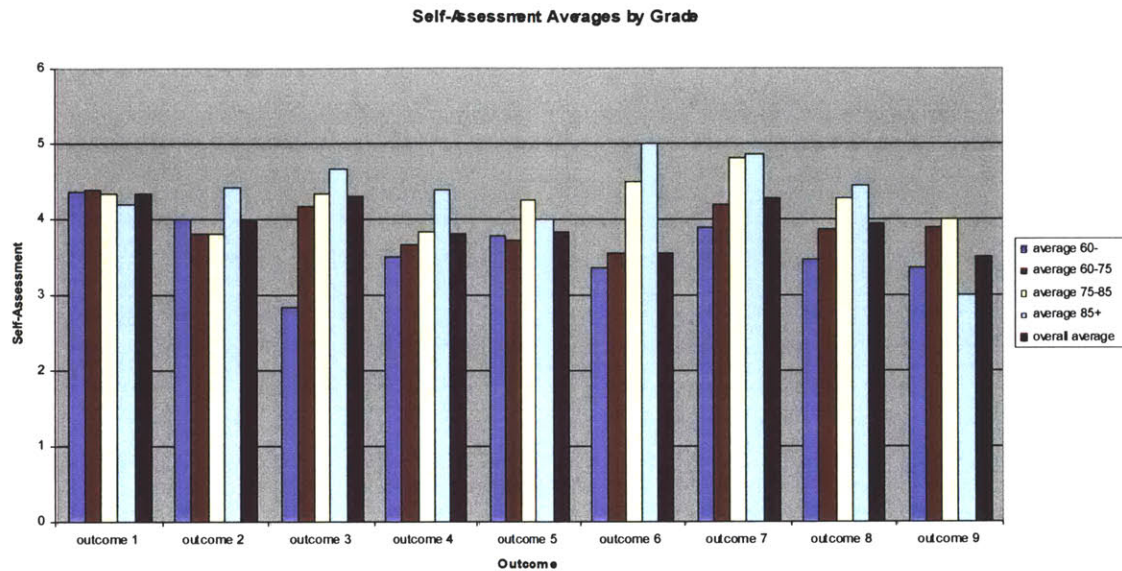


Figure 11. Self-Assessment Averages by Grade Range, 6.042 Spring 2002.

Diagnostic Exams

For the Spring 2001 and Spring 2002 terms, a diagnostic exam was given to students at the beginning of the term to test prerequisites and see how much of the course material they already knew. In Spring 2001, the same exam was given at the end of the term as well as a way of testing how much the course helped students improve. To see how good of an indicator of student knowledge the exam was, it was compared with the final exam given that term. The correlation coefficient between students' scores on the exit exam and their scores on the final exam was .573. Another value looked at for this term was the change in score between the entry and exit exam. The average percent improvement, as well as the correlation coefficient between the entry exam and the change in score, was found for several particular problems. The correlation coefficient between the scores on the entry and exit exam was also calculated for the same problems. This is shown in Table 3.

Table 3. Data for 6.042 Spring 2001 Entry and Exit Exams.

	problem 3	problem 4	problem 5	problem 6	problem 7	problem 8	overall
entry-change correlation	-0.8231	-0.67571	-0.91634	-0.65389	-0.90021	-0.63056	n/a
entry-exit correlation	0.119551	0.019524	0.132585	0.271712	0.132092	0.347959	n/a
average percent improvement	15%	-22.80%	24.60%	33.50%	20.20%	7.40%	9%

I also found the overall average improvement from entrance to exit exam: 8.9%. The variance of the score changes was 2.6%. For this calculation, I looked at the absolute percentage change between the entrance and exit exam. If the score change was calculated as a percent of the student's grade on the entrance exam, the average improvement was 23.5% with a variance of 17.9%.

Table 4 shows the correlation coefficients between the students' scores on the entry exam in Spring 2002 and their final grade in the course, for the eight different outcomes covered on the entry exam.

Table 4. Correlation Coefficients between Entry Exam and Final Grade, 6.042 Spring 2002.

	correlation coefficient
Outcome1	0.25944675
Outcome2	0.23183858
Outcome3	0.17686495
Outcome4	0.399723923
Outcome5	0.148419853
Outcome6	0.098802544
Outcome7	0.443048738
Outcome8	0.283836902

Discussion

Discussion of Procedure

- Problem Weightings are Not Important

One aspect of my procedure worth mentioning is way I decided to weight the homework, exams, and final exam in determining students' overall grade for each outcome. As mentioned in the Procedure section, using the same weights as were used in the class (homework 30%, exams 25%, and final exam 20%) placed too large an emphasis on exam and final exam problems, particularly for semesters in which there were only two exams each containing four or five problems. I made each homework problem worth 1 point, each exam problem 3 points, and each final exam problem 5 points. I decided that the final exam should be worth the most, since it tests what students leave the course knowing. Similarly, regular exams should be worth more than homework problems, since students might collaborate on homework. However, I found that changing the weighting system around did not have a large effect on the correlation coefficients, or even on students' grades. Often, after I changed the weighting, a student's grade for a

particular outcome would still remain the same, or within 1% or 2% of its previous value. The most grades ever changed by was about 9%, but this was uncommon and generally only happened for students who skipped a problem set or several questions on an exam. My conclusion is that what weighting is used is not terribly important, and the results will be similar as long as one component of the grade does not heavily outweigh the others.

- Survey Medium: Paper vs. Computer

Another thing to consider is whether the surveys should be given on paper or on the computer. In terms of named response rate, the survey medium does not seem to matter; there were low and high response rates for surveys given both on paper and on the internet. The one thing that makes web forms a more appropriate method of administering surveys is the reduced probability of errors. When the surveys are given on paper, the results eventually have to be typed into a computer to be analyzed. This introduces an opportunity for typos and other such errors. In addition, paper surveys seem to be much easier lost than those given on the computer. The diagnostic entrance surveys for both the Fall 2000 and Fall 2001 terms were misplaced and as of yet have not been found, whereas none of the results from the surveys and exams given through the internet were lost. One disadvantage to web surveys is that it is more difficult to make them mandatory. The instructor could require a handout to be turned in, say, at the final exam, but it would be harder to have students complete a mandatory web form, especially since mandatory surveys cannot require students to include their name. However, as the person who analyzed much of the 6.042 data, I feel that web-given surveys save time and are less likely to result in mistakes.

- The 1-5 Rating Scale

The 1-5 scale that students were told to use to rate themselves caused a few problems. Because a 1-5 scale in which 1 represents the highest rating is opposite from the one some students are accustomed to using (in which 5 represents the highest rating and 1 the lowest), about 10% of the students misunderstood the scale. For surveys on which the student's comments made it obvious that they had accidentally reversed the scale, I changed their scale back to what it should have been to see how big of a difference this made in the correlation coefficients. Surveys for which it was not clear whether the ratings had been reversed were left alone. One way to help fix all these problems is to have students rate themselves on the typical A-F scale, or on a percentage scale, that is more common to everyone and not as open to individual interpretation.

- Converting the 1-5 Rating Scale to a Percentage Scale

The calculations I made in which self-assessments are compared to grades on the same percentage scale were made by converting the 1-5 self-assessment scale into a 0%-100% percentage scale in which a 1 represented a 100%, 2 a 75%, 3 a 50%, 4 a 25%, and 5 a 0%. This was done so that the grades and self-assessments had the same range of possible values and could be compared and viewed on the same graph. However, perhaps this was not the appropriate way to convert the 1-5 rating scale. It is unlikely that a student would understand 0% of the material, so perhaps the low end of the scale should have been, say, 25% or some other non-zero percentage. This is another area in which having an A-F rating scale or a percentage scale on the self-assessment survey could make the results more clear.

Correlation Coefficients of Self-Assessment Surveys

- Reasons for Low Correlation Coefficients from Self-Assessment Surveys

The correlation coefficient between students' self-assessments and final grades for 6.042 were low on average. The highest correlation coefficient seen for any outcome was 0.51, but the average over all outcomes and all terms was only 0.28 which indicates a fairly weak correlation. A "strong" positive correlation coefficient is usually considered to be 0.5 and above.

- Effect of the Rating Scale on Correlations

There are many possible reasons for why the correlation coefficients are so low. Students were asked to rate their understanding of each outcome on a scale of 1-5, corresponding to that outcome being understood excellently, reasonably, adequately, somewhat, or not at all, respectively. However, the students might each have had their own interpretation of the rating scale and the meaning of terms like "reasonably" and "adequately." Or, some students might have rated themselves absolutely while others compared their performance to the rest of the class. Some could have considered a rating of 3 an average understanding of the material, but others might have given themselves this rating only if they felt their understanding was quite lacking. Similarly, students may have exaggerated the ratings they gave themselves. A student who was a little confused about a certain topic might have rated themselves a 5 on that outcome out of frustration even though their actual grade could have been closer to the 2-3 level. There was also the problem of the reversed scale mentioned earlier. While the difference in correlation coefficients caused by this misunderstanding was not large, it was noticeable; if there were even more students who reversed the rating, the correlation coefficients would be lowered even more.

- Students' Attitude Toward the Survey

Another possible source of low correlation coefficients is that students did not take the survey entirely seriously. For instance, in looking at the survey responses, I noticed one or two students each semester who rated themselves the same number on all or most of the outcomes, such as all 3's or all 5's. This suggests they might have rated themselves on their overall understanding of the course and did not take the time to think about their understanding broken down by outcome. Since the survey was given near the end of the semester, if not after the class had already ended, students were probably focusing on final exams and other such things and did not want to spend much time analyzing their understanding of each outcome. This is especially true if the survey was given after the class had ended, since then the students' understanding of the material may not matter to them any more. Having an optional survey may also have caused some students to take it less seriously. Plus, since the survey was often optional and sometimes anonymous, the named response rate varied only from 26% to 66% of the class size. Increasing the sample size would lead to more reliable results. One suggestion is to make it mandatory for the students to take the survey, and give the survey during the last week of class or have it due at the final exam. That way, when students reply to the survey, they will hopefully be studying for the final exam and will already be thinking about what concepts of the class they do and do not know.

- Differences in Students' and Instructors' Perception of Class

One question to ask is what the importance of the survey is in the first place. Instructors can see what outcomes the students need help on simply by looking at the students' grades; no self-assessment survey is needed for that. The self-assessment survey only lets instructors know how the students think they did. However, as the results of these surveys showed, what the students think it means to understand the material in the course can be quite different from what the instructors think it means. For example, only a basic, general knowledge of a concept may be expected of students, so while the students' grades on the questions asked of them are high, the students' may still feel like they do not understand the concept. This works the other way around as well. Or, what the students think they know about a particular outcome might not be reflected on exams and homework. Another possibility is that a student might gain a full understanding of the material only after he is tested on it, or might forget a concept that was covered early in the semester. This does not necessarily say anything about how well the students learned the material, but merely that how students rate themselves is different than how instructors grade students.

Comparing Self-Assessments to Grades

- **Graphs Comparing Self-Assessment to Class Grade**

By converting the self-assessments to a percentage (with 1 representing 100%, 2 representing 75%, etc) we can compare the grades to the self-assessments with bar graphs. Looking at the results in this way, we can see which outcomes had the largest discrepancy between the grades and self-assessments. For example, outcome 9 (deviation from the mean) was the outcome that resulted in the largest discrepancy for three out of the four terms of 6.042 looked at. In two terms, students highly underrated themselves and in one term they highly overrated themselves on this outcome. Most likely this discrepancy occurred because it was the last topic covered in the class; there were much fewer homework and exam problems on this topic as compared to the other outcomes. Additionally, exam questions on this topic tended to occur at the end of the exam, and students often ran out of time before even attempting to answer the question. Outcome 7 also tended to have a large discrepancy, but since it was covered closer to the middle of the semester, the reason for this is unknown. For all terms and all outcomes, the absolute difference between the students' average grade and average self-assessment was 9.2%. Another thing to note on the bar graphs is that students tended to overrate and underrate themselves equally as often, with the exception of the Spring 2002 term in which they overrated themselves on all outcomes.

- **Graphs of Average Self-Assessment by Grade**

The bar graphs showing the average self-assessment by grade can often tell us more than the correlation coefficient about how students rate themselves. For instance, the correlation coefficient between self-assessments and grades for outcome 3 in Fall 2000 is only 0.26, but from the third set of columns in Figure 8 it can be seen that the ratings follow the trend that they should. For outcome 4 in the same semester, the columns show a bit of confusion (lower grades not always resulting in lower self-assessment ratings) even though the correlation coefficient for this outcome is slightly higher, 0.28. The groups I divided the grades into (under 60%, 60-75%, 75-85%, and above 85% on a particular outcome) was rather arbitrary; this division resulted in approximately even distributions of students most of the time. There were a few instances when there were very few or no students in a particular division and many in another, but this did not happen often. The graphs would change a little if the breakpoints in grades were different. It could alternately be done by dividing the students up into those who got A's, B's, C's, D's and F's.

Diagnostic Exams

- **Diagnostic Exams Did Not Always Cover Relevant Material**

The other data looked at was the grades on the diagnostic exam given in the Spring 2001 and Spring 2002 terms. The Spring 2002 entry exam covered outcomes 1-8, but the exam in 2001 only covered 5 of the outcomes and was not a very thorough exam. To see how well the exam tested what instructors considered important concepts, it was compared to the final exam given that term. The correlation coefficient between students' grade on the exit diagnostic and their grade on the final exam was calculated to be 0.57—a moderately strong but nowhere near perfect correlation. This suggests that the diagnostic exam was probably not testing the same things as the final exam. Some of what was covered on the diagnostic were questions dealing primarily with prerequisites for the course, so it may have served as a better entry exam than exit exam. The questions were also much simpler and shorter than questions appropriate for the final exam, so the diagnostic may very well have been testing different things than the final exam.

- **Errors in Grading Diagnostic Exams Invalidate Results**

I also looked at students' scores on specific questions on the Spring 2001 diagnostic exams to see how much, if any, they improved their knowledge after taking the course. On average, the improvement was only 9%, but the data used in this calculation was questionable: the way the exam was graded at the beginning of the term differed from the grading at the end of the term. Questions with multiple parts were counted as one question on the entry exam, while on the exit exam each sub-part was counted as a separate question. Because of this, students' scores could actually go down between the entry and exit exam even if they answered more of the question correctly. Things like correlation coefficients and percentage improvements associated with the diagnostic exams for this term are therefore not completely reliable because of this mistake in grading. By looking instead at the students actual responses on the exams, instructors might get a better idea of what exact concepts students did and did not improve on, but this information is difficult to attain solely from the students' exam scores.

- **Correlation Coefficients Related to Diagnostic Exams Were Not Useful**

For the Spring 2001 term, another correlation coefficient examined was between the scores on the entry exam and the change in scores from the entry exam to the exit exam. The hope was that there would be a strong negative correlation, with correlation coefficients close to -1, suggesting that taking the class helped students who scored low on the entry exam improve their

scores a lot. The correlation coefficients were in fact close to -1 on average, but this is difficult to interpret because students who scored low initially have much more room to improve than students with high entry exam scores. The problem still exists when the improvement is calculated as a percentage of the student's score on the entry exam. Because of this, I believe the correlation coefficient is not a useful thing to look at in this case, and that examining just the improvements from entry to exit exam (when the exams are uniformly graded) would be more informative.

The correlation coefficient between the entry exam scores and the final grade in the class was found for the Spring 2002 term, to see if students who came into the class knowing more did better overall. The average correlation coefficient was 0.25, a weak positive correlation. This could be considered a good result, suggesting that students who know more in the beginning do not have an extreme advantage in the class. However, there could also be other unrelated reasons for the correlation coefficient being what it is.

Conclusions and Recommendations

- **Low Correlation Coefficients Should Be Expected**

I believe that the current system has promise, though the results are not as clear as expected. Originally we planned to look only at the correlation coefficients. Those turned out to be difficult to interpret since most fall into the range of 0.1 to 0.4 instead of filling the entire range of -1 to 1, or even 0 to 1. This initially made it seem like all the correlations were weak, but now we know not to expect many correlation coefficients much higher than 0.5. The results may never be any more clear than this because of the differences between individual students and between students and teachers, as discussed earlier in the Discussion section. However, we can still see which outcomes had relatively high correlation coefficients and which had low ones, which could indicate which topics in the class were understood well and which were not. It is also useful to look at bar graphs like those in the Results section to see the discrepancies between self-assessments and grades, and to see which outcomes students thought they did not understand. This is another indicator of which topics might need to be clarified in class.

- **List of Outcomes Should Not Change Between Semesters**

A few things could be done to eliminate some of the uncertainty in the results. First of all, between the four semesters of 6.042 that I analyzed, the outcomes varied slightly: in Fall 2000 they were listed in a different order, and in Spring 2001 there was no outcome 1. To ensure that results from previous years can be accurately compared with current ones, the same list of

outcomes and the same survey should be used each semester (unless the topics covered in the class change).

- Improving the Rating Scale

The same rating scale should also be used each semester. A 1-5 scale was always used, but in Fall 2000 and Spring 2002, 1 was a high rating and 5 a low rating; in Spring and Fall 2001, 5 was a high rating and 1 a low one. Instead of changing the rating scale each term, the instructors should pick the one that seems to work best and stick with it. Since students seem accustomed to a scale in which a higher number means a higher rating, the scale used in Spring and Fall 2001 is the best to use. A different scale, such as an A-F scale, might be more familiar to students and therefore individual students might interpret the five different ratings in similar ways. But since the 1-5 scale has been used for several years now, it might be best to continue using it.

- Suggestion for Giving Midterm Surveys

While giving surveys at the end of a semester is beneficial to future terms of a class, a midterm survey could be more useful to the current students. Instructors certainly can identify weaknesses by looking at homework and exam grades. Still, a survey similar to the one given at the end of the course would let instructors know what students think are the weak points of the class that need to be improved or better explained. Giving the survey midway through the course could allow instructors to respond to this feedback in time to make a difference to the current students, as opposed to long after the course concludes. The idea of collecting midterm course evaluations, or other similar methods of assessment, is in practice at several other schools [3] [4].

- Improvements to Diagnostic Exams

A few improvements to the diagnostic exams can also be suggested. First of all, since the exams were lost two of the semesters they were given on paper, it seems giving them through the web is more appropriate. Like the self-assessment survey, it would have to be given in such a way that students would be likely to take it seriously and think about their responses. Making it a homework assignment, or part of a homework assignment, might be one way to accomplish this. In terms of the content of the exams, the Spring 2002 exam seemed more relevant to the course than the one given in Spring 2001. If the instructor wants to look at correlation coefficients involving scores on the entry exam, the exam needs to be made to cover most of the outcomes of the course so it can be compared to students' final grades or self-assessment surveys. I also believe that the idea of giving the same entrance and exit exam, as was done in 6.042 in Spring

2001, has potential. When created to cover the concepts important to the class and when graded consistently, it would let instructors see exactly how much the class helped students learn in certain areas. The questions asked would have to be similar to those on homework and exams in the class, to make sure the exam is truly testing what instructors consider to be important.

- Improved Response Rate Needed

Looking at the number of students who returned self-assessment surveys and included their names, it is clear what a final improvement could be: a better response rate. Making the surveys mandatory is the simplest way to get more responses. In the Fall 2000 term of 6.042, when a mandatory self-assessment survey was given and required to be turned in at the final exam, the largest named response rate was seen; 65 out of the 98 students wrote their name on the survey. Including the survey as part of a homework assignment, or attaching it to a homework assignment, is another means of reminding students to turn it in. One possible way to get more responses is to require students to complete the survey and encourage them to include their name by letting them know how their information will be used and why writing their name is helpful. Students might be more willing to write their names on the survey and be more open and honest in their responses if they know that their information will be used to improve future semesters of the course.

References

1. Angelo, Thomas A. and Cross, K. P. Classroom Assessment Techniques, A Handbook for College Teachers, Second Edition, San Francisco: Jossey-Bass Inc. Publishers, 1993.
2. Georgia Institute of Technology, Office of Assessment, [Online document], Available HTTP: <http://www.assessment.gatech.edu/>
3. Gold, Bonnie and Keith, Sandra Z. and Marion, William A. Assessment Practices in Undergraduate Mathematics, United States: The Mathematical Association of America, 1999.
4. North Carolina State University, Planning and Assessment Homepage, [Online document], Available HTTP: <http://www2.acs.ncsu.edu/UPA/assmt/>
5. Rose-Hulman Institute of Technology, Institutional Research, Planning, and Assessment Homepage, [Online document], Available HTTP: <http://www.rose-hulman.edu/IRA/IRA/index.html>
6. Rummel, R.J. Understanding Correlation, [Online document], Available HTTP: <http://www.mega.nu:8080/ampp/rummel/uc.htm>
7. Smillie, K. W. An Introduction to Regression and Correlation, Great Britain: The Ryerson Press, 1966.
8. Smith, Peter J. Into Statistics, Second Edition, Singapore: Springer-Verlag Singapore Pte. Ltd. 1998.

Appendix A: 6.042 Course Objectives and Outcomes

The following is the list of course objectives and outcomes created for the last three terms of 6.042. It details the important concepts instructors feel students should understand by the end of the course. This was handed out to the students at the beginning of the semester, and the self-assessment survey given at the end of the course was based on these objectives and outcomes.

Course Objectives and Outcomes

Objectives

On completion of 6.042, students will be able to explain and apply the basic methods of discrete (non-continuous) mathematics in Computer Science. They will be able to use these methods in subsequent courses in the design and analysis of algorithms, computability theory, software engineering, and computer systems.

In particular, students will be able to

1. **reason mathematically about basic data types and structures** used in computer algorithms and systems; distinguish rigorous definitions and conclusions from merely plausible ones; synthesize elementary proofs, especially proofs by induction.
2. **model and analyze computational processes** using analytic and combinatorial methods.
3. **apply principles of discrete probability** to model and solve elementary problems of reliability and estimation.
4. **work in small teams** to accomplish all the objectives above.

Learning Outcomes

Students will be able to:

1. use **logical notation** to define and reason about **fundamental mathematical concepts** such as sets, relations, functions, and integers.
2. **evaluate elementary mathematical arguments** and identify fallacious *reasoning* (not just fallacious conclusions).
3. **synthesize induction hypotheses and simple induction proofs**.
4. **apply graph theory models** of data structures and state machines to solve problems of connectivity and constraint satisfaction, *e.g.*, scheduling.
5. apply the method of invariants and well-founded ordering to **prove correctness and termination of processes** and state machines.
6. **derive closed-form and asymptotic expressions** from series and recurrences for growth rates of processes.
7. **calculate numbers of possible outcomes** of elementary combinatorial processes such as permutations and combinations.
8. **calculate probabilities** and discrete distributions for simple combinatorial processes; calculate means and variances.

9. solve problems of estimation and error tolerance by **applying theorems on deviation from the mean.**
10. problem solve and **study in a small team** with fellow students.

Appendix B: 6.042 Exit Self-Assessment Survey

The following is the survey given to 6.042 students at the conclusion of the semester.

6.042 Assessment Survey

1 Instructions

The EECS Department is trying to develop a system to monitor and improve teaching effectiveness. Part of this effort involves formulating and checking an Educational Objectives and Outcomes statement like the one for 6.042 on the course web page. We would like your feedback on how well you feel these objectives and outcomes were achieved for you personally. We would like you to write your name on your survey submission to crosscheck self-assessments against grades; we think the cross-check will be helpful in improving the course. However, **writing your name is optional**. No one on the 6.042 staff will look at any survey results until after grades are assigned, so you need not be concerned that what you say in your self-assessment will impact your grade at all.

In the indicated space next to each item, please enter a digit from five (5) to one (1) where

5 means “this objective/outcome was **thoroughly** achieved for me personally.”

4 means “ : : : **reasonably well** : : : ”

3 means “ : : : **adequately** : : : ”

2 means “ : : : **somewhat** : : : ”

1 means “ : : : **not** : : : ”

2 Objectives

On completion of 6.042, students will be able to explain and apply the basic methods of discrete (non-continuous) mathematics in Computer Science. They will be able to use these methods in subsequent courses in the design and analysis of algorithms, computability theory, software engineering, and computer systems.

In particular, students will be able to

1. **reason mathematically about basic data types and structures** used in computer algorithms and systems; distinguish rigorous definitions and conclusions from merely plausible ones; synthesize elementary proofs, especially proofs by induction.

(enter 5 “thoroughly achieved, : : : , 1 “not achieved”)

2. **model and analyze computational processes** using analytic and combinatorial methods.

3. **apply principles of discrete probability** to model and solve elementary problems of reliability and estimation.

4. **work in small teams** to accomplish all the objectives above.

3 Learning Outcomes

Students will be able to:

1. use **logical notation** to define and reason about **fundamental mathematical concepts** such as sets, relations, functions, and integers.

(enter 5 “thoroughly achieved, : : : , 1 “not achieved”)

2. **evaluate elementary mathematical arguments** and identify fallacious reasoning (not just fallacious conclusions).
3. **synthesize induction hypotheses and simple induction proofs.**
4. **apply graph theory models** of data structures and state machines to solve problems of connectivity and constraint satisfaction, e.g., scheduling.
5. apply the method of invariants and well-founded ordering to **prove correctness and termination of processes** and state machines.
6. **derive closed-form and asymptotic expressions** from series and recurrences for growth rates of processes.
7. **calculate numbers of possible outcomes** of elementary combinatorial processes such as permutations and combinations.
8. **calculate probabilities** and discrete distributions for simple combinatorial processes; calculate means and variances.
9. solve problems of estimation and error tolerance by **applying theorems on deviation from the mean.**
10. problem-solve and **study in a small team** with fellow students.

4 Course Activities

Indicate using the same five (5) to one (1) scale:

- How helpful were the following materials and activities in

	understanding the big ideas	preparing for exams
course notes		
the Rosen text		
lectures		
lectures slides		
in-class problem solving sessions		
problem sets		
online problems on the reading		
discussion with staff during in-class sessions		
other discussion with staff, e.g., in office hours		
email from course staff		
collaboration with fellow students on psets		

- How helpful overall was the course in helping you achieve the Objectives and Outcomes?
- How interested would you be in having another class in the “lecture/studio” style of 6.042?
Circle one:
enthusiastic interested somewhat interested uninterested unwilling

5 Further Comments

We would be interested in hearing any other comments or suggestions you may have about the course: