GIVING CREDIT ENTICES MORE STUDENTS TO CHECK THEIR WORK, BUT…

David W. Petr¹

Abstract – This paper focuses on the effects of giving students credit for checking their answers. Two semesters of a beginning circuit analysis course are compared. In both semesters, answers on weekly quizzes were graded “all-or-nothing” in hopes that this would encourage the students to check their work carefully. A previous study of the first semester showed that this was insufficient inducement, so in the second semester, students were required to check their work by allocating quiz points specifically for checking. This resulted in a dramatic increase in the percentage of students who checked their answers on quizzes, but other important goals were not achieved. Specifically, quiz scores were not substantially higher in the second semester, and there were mixed results regarding checking answers on exams. This study also highlights the importance of conducting effective (vs. ineffective) checks and corroborates previous evidence that time pressure is not the primary barrier to checking answers.

Index Terms – assessment, checking answers, student confidence, motivation

INTRODUCTION

This is the third in a series of papers exploring ways in which to motivate students to become proficient in checking their answers. Engineering students need to develop this habit in order to become effective practicing engineers, yet it seems that very few acquire this important skill. Engineering faculty seem to agree that students do not develop this skill, yet we have tended to “look elsewhere for the problem rather than critically evaluating the processes that we have control over within our own classrooms.” The preceding is from a discussion [1] of the motivating role of the engineering professor, including the importance of “mastery experiences” to enhance students’ sense of self-efficacy, that is, self-judgment concerning capability. Answer checking has the potential to be such a mastery experience. However, this activity is seldom seen as inherently interesting or enjoyable, so that the intrinsic motivation level [2] is low. The problem then becomes one of finding appropriate and effective extrinsic motivation.

In the first paper of this series [3], a confidence scoring methodology was devised and applied to exams. Students were encouraged to rate their confidence in their answers, with the intent of motivating students to check their work (so that they would be reasonably confident that each answer was either correct or incorrect). Although this study provided some important information about how students view their work, it seemed to fail in its primary purposes in two ways. First, although students were given motivation (through exam points) to think about whether or not their answers were correct, it did not measure whether they actually performed any checking. Second, a surprising number of students chose to say, in essence, "I have no idea if my answer is correct or not."

The second experiment [4] used a more frequent assessment method (weekly quizzes) and gave what was hoped to be a sufficiently strong grade motivation: all-or-nothing scoring of answers. Although this study indicated a definite positive correlation between checking answers and obtaining good scores, and evidence of this was given to the students each week, it also revealed a surprisingly strong resistance or apathy to checking on the part of many students. One of the conclusions of that paper was that, if we want students to establish cross-checking as a habit, "it may be necessary (as with other aspects of learning) to actually require that students check their work and base part of the score directly on the cross-checking demonstration." This paper investigates the effectiveness of such a technique.

EXPERIMENT BACKGROUND

This paper focuses on two beginning circuit analysis courses that were similar in many respects. Such a course presents an ideal opportunity to introduce (or reinforce) the practice of checking answers because students tend to be toward the beginning of their engineering education and because the material offers various means and ample opportunities for checking answers.

Each class (Fall 2000 and Spring 2002) comprised a mix of second-year electrical and computer engineering majors in addition to other students (primarily architectural engineering) of various levels. Specifically, the mix of electrical/computer engineering plus other students was 47+21 in Fall 2000 and 33+28 in Spring 2002.

Weekly Quiz Procedures

In each class, weekly quizzes were given which counted for 20% of the semester grade. Problems were carefully selected to be very straightforward and to ensure that answers could be checked in some way. Methods of checking included consistency checks (e.g., from a set of node voltage answers, find branch currents with Ohm’s law

¹ David W. Petr, University of Kansas, Electrical Engineering and Computer Science, 1013 Learned Hall, Lawrence, KS 66045 petr@eeecs.ku.edu

0-7803-7961-6/03/$17.00 © 2003 IEEE

33rd ASEE/IEEE Frontiers in Education Conference

November 5-8, 2003, Boulder, CO

T3A-8
and then check their consistency using Kirchoff’s current law, checking power conservation, working backwards (e.g., from a voltage or current answer to a source value), working the problem with an alternate method, etc. For most of the quizzes, the problem in Spring 2002 was the same as the Fall 2000 problem except for specific numerical values. The primary difference between the semesters was in how the quiz problems were graded.

In Fall 2000 [4], for all but the first quiz reviewing prerequisite topics, each of the one or two numerical answers was graded “all-or-nothing,” receiving either full credit if it was completely correct (including units) or zero credit otherwise. This grading method was intended to encourage careful checking, since it placed such emphasis on getting the correct answer. However, as mentioned in the Introduction, students showed considerable resistance or apathy to checking their answers.

In the course evaluations at the end of the Fall 2000 semester, students commented that it was not fair to receive zero credit for failing to include a unit on an otherwise correct answer. In response to this complaint, in Spring 2002 the all-or-nothing grading was retained but modified so that it considered only the numerical value of the answer, with or without proper unit. These all-or-nothing numerical answer scores accounted for only eight of the ten points on each quiz, and a ninth point was awarded for including correct unit(s), regardless of whether numerical answers were correct.

The tenth point on each quiz was awarded rather generously if the student performed some sort of check on at least one of the numerical answers. If the student showed evidence of attempting to check an answer, even if the check was weak or done incorrectly, the tenth checking point was awarded. In this way, explicit credit (10% of each quiz score, for a total of 2% of the course grade) was awarded for checking answers.

Exam Procedures

Explicit credit was not awarded for checking answers on exams in either semester. Rather, exams were used to determine if answer checking carried over into situations in which less emphasis was placed on it.

Three exams were given during the course of each semester. In Fall 2000, the first two were 50-minute, in-class exams, but in Spring 2002 the first two exams were given during special evening class periods to reduce time pressure. In both semesters, the third exam was a 3-hour final exam that was comprehensive, but focused on material in the final third of the course. In all exams, problems with a range of difficulty were included, and partial credit was awarded for correct set-up and approach.

Another small difference between the semesters was that all exams in Spring 2002 has specifically-labeled space encouraging the students to check their work, whereas in Fall 2000 ample space was provided for checking but was not specifically labeled.

To investigate the influence of time pressure on answer-checking, exams that were turned in at least 10 minutes prior to the end of the exam period were considered to be turned in early and recorded as such. Furthermore, when the exams were graded by the instructor, a record was kept of attempted cross-checking on a per-student, per-problem basis.

Other Checking Activities

In both semesters, the instructor made a concerted effort to emphasize the importance and practice of cross-checking answers. At the conclusion of nearly every example worked out during lectures, we would discuss how the answer could be checked. In Fall 2000, the average scores of students who did and did not check their quiz answers was announced to the class each week, clearly showing a positive correlation between cross-checking and good scores. In Spring 2002, homework problems specifically aimed at cross-checking (so-called Verification Problems) were assigned from the text [5].

Quiz Results

Frequency of Checking

Figure 1 clearly shows that there was a dramatic increase in the percentage of students who performed a check on their answers. Due to slight changes in the pace of material covered, there was a quiz on Laplace circuit analysis in Spring of 2002 but not in Fall 2000, and there was a quiz on thevenin impedance in Fall 2000 but not in Spring 2002. Also, the first three quizzes in Fall 2000 were not analyzed for checking, hence the lack of data for them. Excluding the first four quizzes of each semester, the average checking percentage for Spring 2002 was 70.6% compared to 15.6% for Fall 2000. The variation in checking frequency throughout the Spring 2002 probably reflects the relative difficulty of determining a check for the various types of problem.
Cross Checking and Quiz Scores

As in Fall 2000, there was a significant positive correlation between checking and good scores in Spring 2002. Table I summarizes the results for the two semesters, excluding the first three quizzes. Also, the scores for Spring 2002 exclude the point for checking to avoid biasing the difference in score between checkers and non-checkers. We can see that the checking differential is similar between the two semesters, with a somewhat smaller checking differential in Spring 2002, when students were given credit for performing a check.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Percent Checkers</th>
<th>Average Score</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2000</td>
<td>15.6%</td>
<td>8.5</td>
<td>5.3</td>
</tr>
<tr>
<td>Spring 2002</td>
<td>7.06%</td>
<td>6.9</td>
<td>3.9</td>
</tr>
</tbody>
</table>

One would hope that the substantially increased frequency of answer-checking in Spring 2002 would result in higher quiz scores compared to Fall 2000. Our comparison will exclude the first review quiz in each semester, since it was of a different character. This comparison is problematic because of the different scoring methods used in the two semesters. The following are three possible ways to adjust the scores from Spring 2002 so that they can be meaningfully compared with Fall 2000 scores. In each case, AdjScore refers to the adjusted score that can be compared with Fall 2000 scores. NumScore refers to the score for the numerical answers (8 points possible) in Spring 2002, and UnitScore refers to the score (0 or 1) in Spring 2002 for supplying correct units.

1. \[ \text{AdjScore} = 10\times(\text{NumScore}/8) \]
2. \[ \text{AdjScore} = 10\times[(\text{NumScore} \times \text{UnitScore})/8] \]
3. \[ \text{AdjScore} = 10\times[(\text{NumScore} + \text{UnitScore})/9] \]

None of these methods is perfect. Method 1 compares only the numerical answer scores for the two semesters, but ignores the effect of failing to provide proper units in Fall 2000 (a zero score for the answer). Method 2 attempts to emulate this effect by “zeroing out” the numerical score if correct units were not supplied in Spring 2002. Method 3 simply adds the numerical answer score (8 points possible) and the unit score (1 point possible) and uses a 9 point basis. Method 2 is perhaps the closest approximation to the Fall 2000 scores, but it could be argued that students were less careful with units in Spring 2002 since the penalty for not supplying units was considerably less.

Table II compares the average quiz score from Fall 2000 with each of these three adjusted score averages from Spring 2002. Regardless of adjustment method used, there is certainly not a dramatic improvement in the Spring 2002 scores, despite the dramatically increased frequency of checking! This suggests that many of the checks were ineffective, that is, not accomplishing their primary purpose. This will be discussed further in a subsequent section.

**TABLE II**

<table>
<thead>
<tr>
<th>Semester (Method)</th>
<th>Average Quiz Score (Last 11 Quizzes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2000</td>
<td>6.03</td>
</tr>
<tr>
<td>Spring 2002 (1)</td>
<td>6.55</td>
</tr>
<tr>
<td>Spring 2002 (2)</td>
<td>6.04</td>
</tr>
<tr>
<td>Spring 2002 (3)</td>
<td>6.82</td>
</tr>
</tbody>
</table>

**EXAM RESULTS**

We now turn our attention to results for the three exams.

**Frequency of Checking**

Figure 2 shows the distribution (for each exam) of the number of exam answers that students checked in Spring 2002. It is clear from this chart that checking decreased with each exam (percent who checked no answers grows and percent who checked more than 3 answers shrinks, both in dramatic fashion). This may be due to any number of factors, alone or in combination. These include increasing difficulty of exam problems, increasing difficulty of determining a checking procedure, and decreasing enthusiasm for checking. For the final exam in particular, students may care less about their score (and hence about checking their answers) since course grade is largely determined for most students by the time they take the final exam. (In Spring 2002, the final exam weight was 25%, the same as the other two exams.)

A comparison of the two semesters is also interesting. Figure 3 compares the semesters in terms of the number of students who checked at least one exam answer, for each exam. We see that checking frequency decreased as the
semester progressed in both semesters. There does appear to be a substantial improvement in the number of students who checked at least one Exam 1 answer in Spring 2002, but the improvement diminishes in the second exam and disappears altogether in the third exam.

![FIGURE 3](image)

**FIGURE 3**

Exam Checking Comparison Between Semesters, By Exam

Figure 4 is another semester comparison, but this time in terms of distribution of number of problems checked, averaged over all exams. We again see some improvement for the Spring 2002 semester, particularly in the number of students who checked many (more than 3) answers. However, we must keep in mind that relatively more time was allowed for Spring 2002 exams and that the Spring 2002 exams encouraged answer checking more explicitly by providing specifically labeled space for it.

![FIGURE 4](image)

**FIGURE 4**

Exam Checking Comparison Between Semesters, By Number of Answers Checked

Checking and Student Scores

Figure 5 shows that there is a small, but consistent, difference between the exam scores of students who checked at least one answer and the scores of those who did not check any. This is consistent with corresponding data from Fall 2002 [2], but the score differences are smaller for Spring 2002. Also repeated for reference in Figure 5 are the data on frequency of checking.

![FIGURE 5](image)

**FIGURE 5**

Student Score Comparison: Spring 2002

The previous discussion has classified students as “checkers” or “non-checkers” based on whether or not they checked at least one answer on a given exam. It is also worthwhile to compare the scores for answers that were checked with the scores of answers that were not checked. We do this for each exam in Figures 6 and 7 for the two semesters. In each semester, we see similar trends in these figures for number of answers checked as we saw for number of students who checked at least one answer (Figure 3). As summarized in Table III, there appear to be relatively more answers checked in Spring 2002, but the score difference between checked and unchecked answers is smaller. This suggests that perhaps the checks being done in Spring 2002 were less effective, even if they were being done more frequently.

![FIGURE 6](image)

**FIGURE 6**

Answer Checking Analysis: Fall 2000

Check any. This consistent with corresponding data from Fall 2002 [2], but the score differences are smaller for Spring 2002. Also repeated for reference in Figure 5 are the data on frequency of checking.
do more harm than good, in that it can appear to validate an incorrect answer.

Table IV shows the results obtained for Exam 1 in Spring 2002. First, note that although many of the answers (30%) were checked on this exam, many of those checks (45%) were ineffective ones. Apparently, a number of the students who were checking their answers had not learned how to do so effectively. Not surprisingly, we see that the average score of answers with an effective check is indeed substantially higher than the average score of other problems. However, note that the average score of problems with an ineffective check is actually lower than the average score of answers with no check. This underscores the importance of teaching students how to perform effective checks and provides another possible explanation for the decreasing number of students who checked answers on subsequent exams: their experience with answer checking on Exam 1 was not positive, so they abandoned it on subsequent exams.

Effective vs. Ineffective Checks

Data collected for Exam 1 in Spring 2002 sheds some light on the topic of effective vs. ineffective checking. On that exam, in addition to recording whether or not a student provided a check for a given answer, a judgment was made as to whether or not the check was an effective one. An effective check was defined as one for which the checking procedure was fundamentally correct. That is, an effective check would be likely to (though not guaranteed to) expose an error in an answer. An example of an effective check for a nodal analysis problem would be to use the node voltage answers to find branch currents and then write at least one correct Kirchhoff’s current law (KCL) equation using these currents. Such a check may not reveal an error in the node voltages, depending on which KCL equation was checked. An ineffective check, in contrast, was defined as one in which the checking procedure was fundamentally flawed, that is, would not reveal an error. An example of an ineffective check would be the above procedure, but with an error in the calculation of current values or an incorrectly written KCL equation. In particular, a student might check his/her node voltage answers using the same incorrectly written KCL equations that were used to find the answers in the first place. This illustrates that an ineffective check may

Time Pressure

Finally, we present results that corroborate the conclusion in [2] that time pressure is not the primary barrier to answer-checking. We again use the definition that a student has checked answers on an exam if he/she has checked at least one answer on the exam. We first note that significantly more students turned in exams early in Spring 2002 (47% vs. 23% in Fall 2000), probably because relatively more time was allowed for the exams in that semester (see Exam Procedures). Table V shows that in both semesters, over half of the students who turned in exams early did not check any of their answers.

<table>
<thead>
<tr>
<th>Type of Check</th>
<th>Number of Answers</th>
<th>Avg Score of Answers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Check</td>
<td>136</td>
<td>96.8</td>
</tr>
<tr>
<td>Ineffective Check</td>
<td>113</td>
<td>80.2</td>
</tr>
<tr>
<td>No Check</td>
<td>591</td>
<td>84.5</td>
</tr>
</tbody>
</table>

Time Pressure and Checking

<table>
<thead>
<tr>
<th>Student Population (Turned In Exam Early or Did Not Turn In Early)</th>
<th>Percent Who Checked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned In Early</td>
<td>Fall 2000</td>
</tr>
<tr>
<td></td>
<td>Spring 2002</td>
</tr>
<tr>
<td>Did Not Turn In Early</td>
<td>Fall 2000</td>
</tr>
<tr>
<td></td>
<td>Spring 2002</td>
</tr>
</tbody>
</table>
SUMMARY

This study has compared student checking and scores for two similar classes, with the primary difference being that students in Spring 2002 were required to check their answers on weekly quizzes by allocating points specifically for checking, whereas students in Fall 2000 were not given any quiz credit specifically for checking their work. Giving credit specifically for answer-checking did have the desired effect of enticing more students to check their work. Unfortunately, other desired objectives were not achieved. Specifically, comparing Spring 2002 against Fall 2000,

- Quiz scores were not substantially higher.
- Checking on exams, where no credit was given specifically for checking, was more frequent for the first exam, but the improvement diminished to zero by the end of the semester.
- The exam score differential between checking and not checking decreased.

This study also highlighted the importance of performing effective checks: exam scores for answers that were ineffectively checked were actually lower than for answers that were not checked at all. Bad experiences with checking (for example, performing an ineffective check that appears to validate an incorrect answer) may lead to student discouragement with the entire concept of answer-checking, possibly leading to abandonment of the technique. This is a possible contributing factor in the declining frequency of exam checking in both semesters.

Finally, it was again demonstrated that time pressure cannot be primary impediment to answer-checking, since over half of the students who turned exams in early did not check any answers.

CONCLUSIONS AND FUTURE WORK

This study points to a relatively easy way to get students to check their work: give credit for it. But since credit was awarded in this experiment only for performing a check, with no evaluation of the quality of the check, we perhaps should not be surprised that many students did not seem to learn how to perform an effective check. This is an example of the limitations of extrinsic motivation [2]. Future efforts should be focused towards helping students learn how to become both habitual and proficient in checking their answers. That is, we should be helping them learn how to perform effective checks on their answers, in addition to helping them to develop answer-checking as a habit.

Several areas of further investigation are warranted. We still do not have much direct information about why students are resistant/apathetic towards checking; perhaps we should ask them! Also, more can be done in training students to become proficient answer-checkers. Ideas in this area include:

- Giving students more information (e.g., during lectures) about the types of checks that can be performed and what distinguishes an effective check from an ineffective one.
- Attempting to provide self-endorsed extrinsic motivation [2] by impressing on students the importance of checking answers in professional practice (for example, by reviewing well-publicized "disasters" that resulted from failure to check answers).
- Assigning homework problems that specifically require students to check (and correct) someone else’s work.
- Assigning homework problems that require students to evaluate the effectiveness of a check that someone else has done.
- Allocating points on quizzes and exams based on the effectiveness (not just the existence) of a check.

Of course, these methods should be evaluated for their effectiveness in achieving the goal of transforming students into habitual, effective answer-checkers.

ACKNOWLEDGMENT

The author is grateful to student assistant Ashley Stone for help in processing and formatting some of the data presented in this paper.

REFERENCES