
AC 2012-5005: ASSESSMENT OF STUDENT'S CONFIDENCE OF LEARNED KNOWLEDGE

Prof. Kyle B. Reed, University of South Florida

Kyle Reed is an Assistant Professor in the Department of Mechanical Engineering at the University of South Florida. He performs research on human-robot interaction, rehabilitation robotics, haptics, medical robotics, and engineering education. He received his B.S. in mechanical engineering from the University of Tennessee, Knoxville, in 2001, his master's (2004) and Ph.D. (2007), both in mechanical engineering, from Northwestern University and was a Postdoctoral Fellow at the Johns Hopkins University. More information on his research can be found at his research lab website: <http://reedlab.eng.usf.edu/>.

Assessment of Student's Confidence of Learned Knowledge

Abstract

An important skill for students is to recognize how well they actually know the answers to real-world questions. Students may get the right answers on quiz questions, but may not be sure of their answers and may get similar questions wrong. Confidence in one's answer or design is not typically important during undergraduate schooling, but is vital in a job and in graduate education where the confidence students have in their solutions is nearly as important as the solutions themselves. The method described here uses the concept of self-efficacy to help students learn by making them assess the confidence they have in their answer and also serves as a metric by which the instructor can determine which students need help in the class.

Senior students in an undergraduate level Mechanical Engineering semester-long course were instructed to indicate their confidence level on answers on each of five quizzes. The grade for each question was based on both their confidence level and whether it was right or wrong. The best case is to have the right answer and be confident in it (5 points) and the worst case is to be confident of a wrong answer (0 points). Lacking confidence in a right answer is beneficial, but not perfectly so, thus students get 4 points. Students get 2 points for admitting that they do not know or are not sure, which is the perfect answer if they do not know. Making up an answer is a very bad habit that is encouraged under typical grading schemes.

An anonymous survey shows that approximately 3/4 of students were supportive of this idea and many that disliked it stated that they would have preferred grading that included partial credit. The results show that the scoring method slightly biased the students towards selecting not-confident, which is desired for training students to analyze their answers in an educational setting. Approximately 3/4 of students were able to maximize their scores by accurately assessing the correctness of their answers. Students' ability to maximize their quiz answers had a relatively low correlation to their overall course grade ($R^2=0.38$), which indicates that students have a good knowledge of their abilities regardless of their mastery of the material.

Introduction and Background

An important aspect of student learning is having students understand how much they actually know. The confidence students have on their answers can be nearly as important as the answers themselves. Students may get the right answers on a quiz, but many of the students that got it right may not be sure of their answer and may have gotten it wrong if the question was revised slightly. If a bridge has a sign that said: "Weight limit might be 3 tons", one would be hesitant to drive across it, but if it says: "Weight limit 2 tons", most people would not hesitate to drive a sedan across it. Although confidence in one's answer is a vital part of the answer itself, confidence in one's answer is not typically emphasized during schooling. Nevertheless, it is vital when students enter the work force. The confidence-based grading method described in this paper gives students practice assessing the confidence in their answer and also serves as a metric the instructor can use to evaluate how well the students think they know the material.

The method is based on grading the student's answers according to both correctness of the answer as well as the student's selection of "confident" or "not-confident". The points are awarded as shown in Table 1. The worst case is to have a student be confident of a wrong answer, so they get no points. Obviously, the best case is to have the right answer and be confident in it, so they get 5 points. Having a right answer, but lacking confidence, is beneficial, but

Table 1: Points awarded for each combination of correctness and confidence

	confident	not-confident
right	5	4
wrong	0	2

not perfectly so. This would be similar to asking a colleague to double check your work, which would hopefully increase the confidence in that work, but still requires the use of someone else's time, thus they get 4 points. The final case is wrong and not-confident in the answer. In some ways, this is similar to simply saying: "I do not know", which is the perfect answer when a person truly does not know; guessing without stating that it is an educated guess is a bad habit and can have serious consequences. However, this practice of guessing often is encouraged in typical grading schemes. In the case where the student clearly states that they are not sure, the student gets 2 points for recognizing that they do not know and that they should ask for help.

The purpose of this method is to encourage students to evaluate their understanding of the material. Since students are inherently motivated to try to get as many points as possible and the confidence is an important part of their grade, the students are motivated to put some serious thought into their choice. Although confidence is not usually a clear binary choice, this method aims to encourage students to seriously think about their knowledge of the subject. Some students may never check the box marked "not-confident" if they think this method is pointless or if they are always confident in their work. Some students may always check the "not-confident" box, so they are guaranteed a minimum score of 40%, but they are also limiting their maximum score to 80%. Ideally, students will think about how sure they are of each individual answer and respond accordingly with different confidence markings for different questions.

Related research¹² has studied the confidence that students have in their answers, but did so in a largely passive way such that 25% of answers had a neutral rating. The student's grades were also weighted by a maximum of only 5% based on an 11 point scale of their confidence ranging from -5 (very confident it is wrong) to +5 (very confident it is right). Petr's study found that "A" students were better at evaluating their answers than all other students, students generally disliked the method, and that students tended to be optimistic regarding the correctness of their answer. The study described in this paper refines Petr's method by making the students' confidence an integral part of their grade, which encourages students to take the time to think about both their answer and a meta-analysis of their confidence. The grading scheme shown in Table 1 directly relates their answer of the question and their confidence to their grade in a straightforward and easy implementation that students can understand.

In the educational setting, a balance needs to be made between the ultimate goal of having an accurate assessment of their answers and having a supportive and realistic learning environment. The scores assigned to each of the four categories listed in Table 1 are chosen to be most

educational to students. Since students are not typically familiar with this concept and the intention is to help them learn to admit when they are unsure, the weighting encourages students to select “not-confident”. This method reinforces the idea that stating an answer that one is not certain of is the best alternative to having a correct answer as long as it is stated with the condition of not being confident. The score of 2 for wrong and not-confident combined with the other scores was chosen to slant the choice toward selecting not-confident as an educational tool. For example, assuming one knows that there is a 50% chance of getting the answer right, it is clearly advantageous to select not-confident since you get 2 points if it is wrong and only lose 1 point if it is right. Thus, these specific weights are chosen to emphasize one’s confidence in the answer. As will be demonstrated in the results section, the choice of weights for the four categories of answers is important for encouraging students to benefit from this grading method.

From the student perspective, the confidence-based evaluation has elements of the classical form of game theory⁹ where each player (i.e., the student) desires to obtain the highest grade possible. Instead of having another player involved, the student is playing against their belief in their answer. Once they have chosen their answer, the answer is either right or wrong, but they only have partial information regarding the correctness of their answer^{2,3}. The game theory aspect comes from them assessing their answer and choosing whether they are confident or not-confident. This act of assessing their grade makes them further think about the problem in a new way. Because students desire to maximize their grades, one way to analyze this method is based on their ability to obtain the highest score on each question.

The idea of confidence has been studied in several contexts within the field of psychology and is typically referred to as self-efficacy^{1,8}, which is a general belief in one’s own abilities or competence. Parsons et al.¹¹ found that first year engineering students who were better mathematically qualified were generally more confident and successful in mathematics. A study examining the perceived competence of self compared to others showed that individuals will try harder to solve a given problem when they see someone they perceive as similarly competent solve the same problem⁴.

A study by Fantz et al.⁷ found that students who had hobbies related to engineering and students who had pre-engineering classes had significantly higher self-efficacy measures than students without these interests or extra classes in first year students. A survey of first year engineering students’ self-efficacy beliefs found that students’ motivation to succeed in the course and their understanding of the material were ranked as the most influential factors that would contribute to their success in the course¹⁰. Ponton et al.¹³ suggest that professors can enhance a student’s self-efficacy by developing skills, peer interaction, encouraging students, and explaining coping strategies, all of which are important for practicing engineers.

Self-efficacy can be difficult to measure since it is one’s perception about one’s own abilities. To assist in measuring self-efficacy, Carberry et al.⁵ validated an instrument based on 36 questions to measure self-efficacy in engineering design tasks. Their measure does not apply as well to task-specific concepts and they state that further study is needed on how self-efficacy relates to cognitive learning outcomes in engineering education.

In contrast to much of the literature that has focused on student's confidence in general terms, such as their ability to succeed in a course or on a complex design project, this paper focuses on self-efficacy in the face of specific problems that are directly graded based on their ability to assess their own solution. This method is focused on answers that are right or wrong, but the end of this paper discusses options for integrating this method with partial credit grading schemes.

Methods

The goal of this study is to understand how well students are able to assess whether they have answered questions correctly. By design, this study does not have a control group, but is rather examining the fundamental ability of students to understand and assess their own abilities when answering specific questions on quizzes given in class. The long-term goal of this research is to help students increase both their ability to correctly assess their understanding and also increase their overall confidence in the material. By simply making students aware of the concept of confidence, the desire is that they will make better engineering judgements in their future jobs, but a followup study is necessary to track their performance and opinions.

The confidence-based grading was integrated into every quiz in the senior-level engineering Mechanical Controls course, except for the first quiz. The first quiz was intentionally excluded to get the students accustomed to the all-or-nothing grading (i.e., no partial credit). This first quiz was used to make the point that there are clearly right and wrong answers and if you do not get the right answer, it is wrong. Either the sedan will make it across the bridge or it will not. The remainder of the quizzes had some problems with a checkbox to indicate that they are "not-confident" in their answer. By default (not marking the box), they are confident, so if they want to ignore this method, they can do so and still take the quiz all or nothing, just like the first quiz of the course. The problems on quizzes with the "not-confident" checkbox are scored out of five total points as shown in Table 1.

The 5 quizzes included the following 13 problems with the quiz number indicated as Q#: (1, Q2) block diagram reduction, (2, Q3) Laplace Transforms, (3, Q3) Final Value Theorem, (4, Q3) block diagram reduction, (5, Q4) determining the order of a system from a Bode plot, (6, Q4) system response from a step input, (7, Q4), determining system parameters from a transfer function, (8, Q4) finding system parameters from a Bode plot, (9-11, Q5) determining gain and phase margins from a bode plot, (12, Q6) binary logic truth table, and (13, Q6) ladder logic multiple choice question.

This study included 91 senior undergraduate students from two separate semesters of the same course taught by the author. The quizzes were graded by the author and/or a TA and each student verified the grading of their own quizzes. The quizzes were changed slightly between the two semesters to reduce the chance of "sharing" from students who already took the course. On average, 9.2% of the students were absent from taking the quizzes, so the following data includes a total of 1074 problems. This study was approved by the University of South Florida's Institutional Review Board.

Results

For brevity, the four categories will be abbreviated as:

- R:C (right and confident),
- R:NC (right and not-confident),
- W:C (wrong and confident), and
- W:NC (wrong and not-confident).

Table 2: Percentage of quiz questions answered in each category

	confident (C)	not-confident (NC)
right (R)	51.8%	10.3%
wrong (W)	12.4%	25.5%

The breakdown of the four categories based on correctness and confidence is shown in Table 2 for all problems measured using this method throughout the semester. The best case, confident in a right answer, accounted for over half the answers and the worst case, confident in a wrong answer, only accounted for about 1/8 of the answers. As shown in Figure 1 confidence generally rose throughout the semester. Other than confidence, there are not any significant learning effects over the semester, but this is likely caused by the delay associated with grading in which quiz two was returned after quiz three was given and so on throughout the course. One interesting anomaly is that the last quiz had a noticeably larger group of answers in the R:NC category than the other 4 quizzes. One of the questions on this quiz was a multiple choice question where the right answer was particularly similar to a wrong answer (all other questions were open-ended questions). Although 83% of the students selected the right answers, it is likely that the similarity of two answers made them question their selection and, thus, increased the number of students that selected not-confident.

One analysis measure is based on the percentage of correct answers amongst not-confident assessments vs. the percentage of correct answers amongst confident assessments. In the cases where an answer was marked not-confident, only 29% of the students got the right answer as calculated by $R:NC / (R:NC + W:NC)$. Students knew they were not likely to get the right answer,

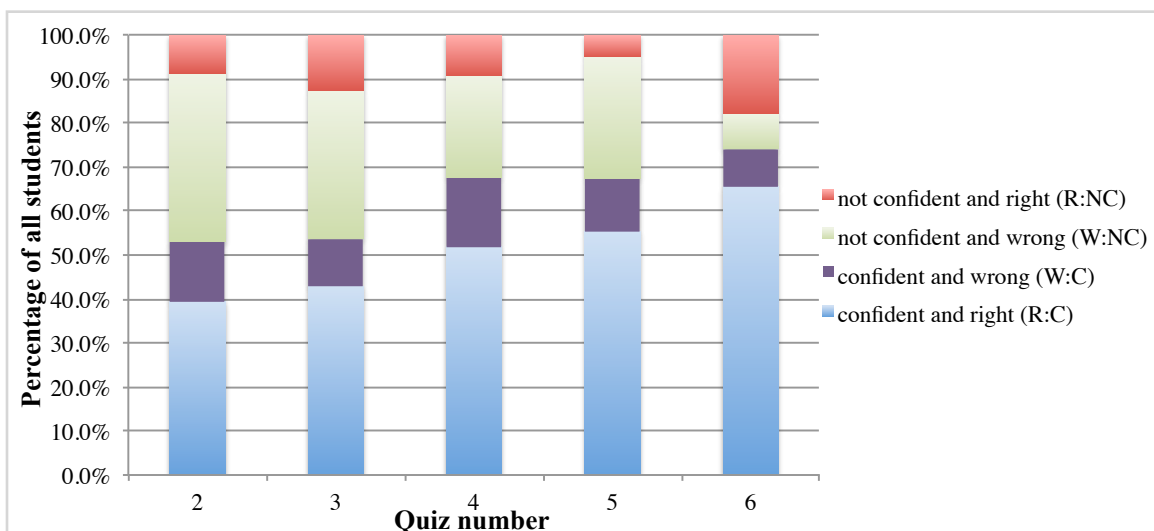


Figure 1: The percentage of students marking each of the four categories. Note that quiz one did not use the confident/not-confident scoring and, thus, is not reported.

so they correctly marked not-confident. In the cases where an answer was marked confident, 81% got the right answer as calculated by $R:C/(R:C + W:C)$. In general, this indicates that students were able to assess their performance reasonably well on each problem.

At first glance, it appears that the students were selecting conservative levels of confidence since less were in the W:C than the R:NC category. Only 19% of answers that were confident got the wrong answer, which is less than the 29% of not-confident answers that were right; note that both of these groups incorrectly assessed their answer. As discussed in the intro, the weighting of the grades encourages students to select not-confident since they are likely to get a higher score if they are equally torn between two answers. With the weights as they are, the students would have to be right at least 66.7%[†] of the time to make always selecting confident result in a better outcome than always selecting not-confident. Since they were only right 62.1% ($R:C + R:NC$) of the time, they have correctly developed an inclination toward not-confident, but continue to select confident in appropriate instances.

From a game theory perspective¹⁴, students can, and should, think of this method as a maximization problem. Their goal is to maximize their score where they have some amount of uncertainty in their answers. Of all the problems graded using this method, 77.3% of the answers were selected such that their grade was maximized on that specific problem. Figure 2 shows the actual scores with bars that represent the maximum and minimum of scores possible assuming their answer does not change and only their confidence response is changed. The maximum achievable score for each student was calculated by (1).

$$S_{max} = \begin{cases} R:C = 5 & \text{if answer is right} \\ W:NC = 2 & \text{if answer is wrong} \end{cases} \quad (1)$$

The minimum achievable score was calculated by (2).

$$S_{min} = \begin{cases} R:NC = 4 & \text{if answer is right} \\ W:C = 0 & \text{if answer is wrong} \end{cases} \quad (2)$$

Since student's assessments of their confidence are not always perfect, it is not expected that the students would perfectly select the value that optimizes their answer. In addition, some of the subjective answers from the survey, discussed below, suggest that students may be risk averse and prone to selecting not-confident more often than they optimally should. Future studies could further evaluate risk aversion by conducting personality profiles of the students.

According to the grading scheme, students using this method will score the highest if they select confident when they have at least a 2 out of 3 chance of getting the answer right. In engineering practice, this is a low number to aim for since an answer that is marked as confident must have a much higher likelihood of being right. To teach students to avoid the detrimental W:C category, the score associated with W:NC should be increased from 2 to 3. However, increasing the score will likely encourage a significant increase in the not-confident category, which may hinder the

[†] The percentage of time the students need to have the right answer to obtain a higher grade by stating confident is calculated by solving this equation for α : $R:C*\alpha + W:C*(1-\alpha) = R:NC*\alpha + W:NC*(1-\alpha)$, where the weights are given in Table 1.

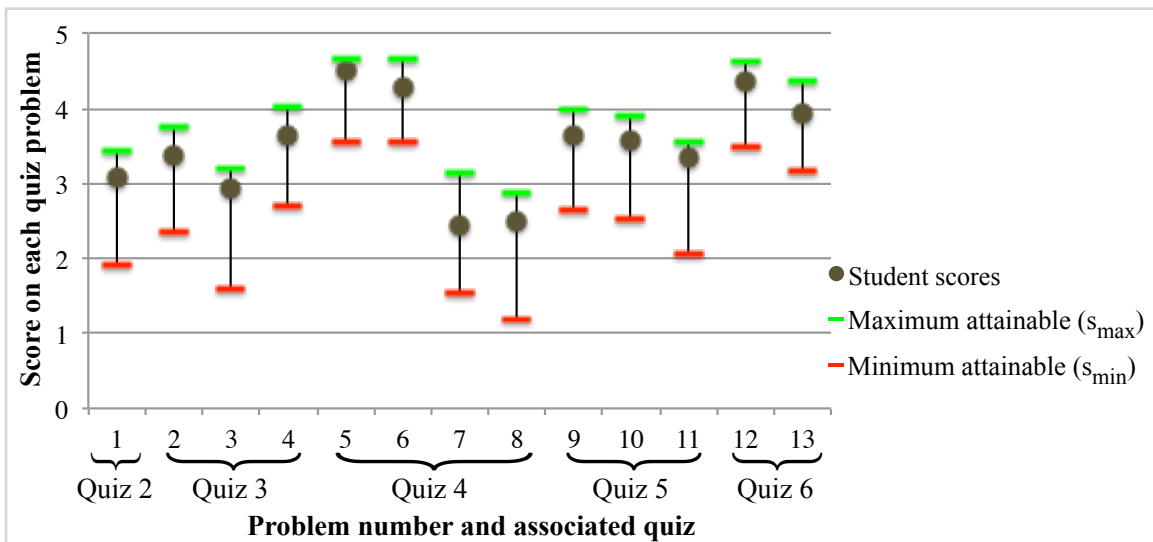


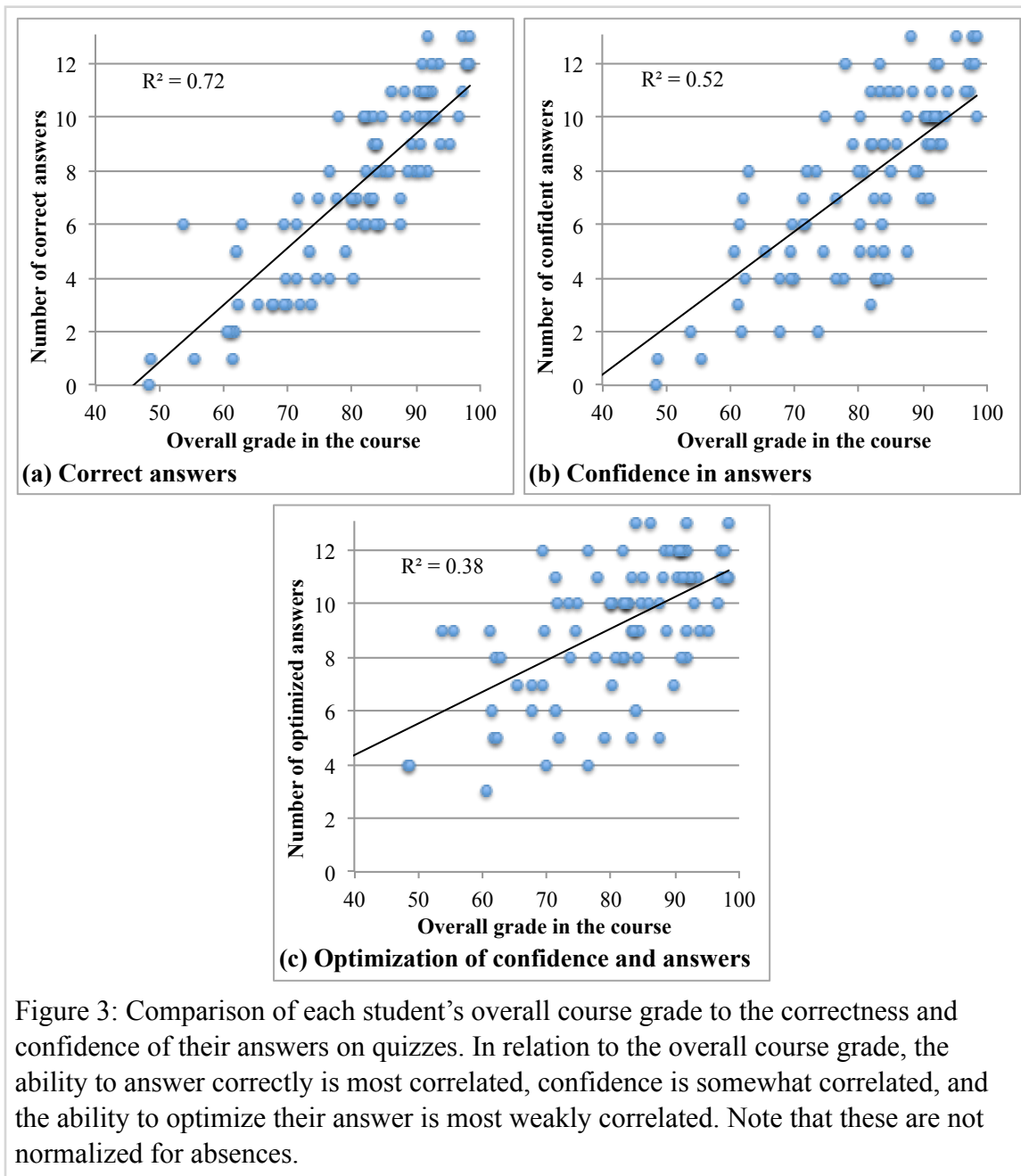
Figure 2: The average student scores on each problem shown with the range of scores that are possible given that students' answers are not changed, only their confidence is changed. The bars represent the scores assuming students perfectly (top, green) or imperfectly (bottom, red) assessed whether their problem was right or wrong.

learning process by discouraging and frustrating students⁶. Students in a class are learning, so they are inherently not going to be fully confident in all of their answers. Future versions of this method could be tested using different weights to assess how the students adapt given different weights.

One may think that harder problems are harder to assess than easier problems, but the data does not support this. There was no correlation between the number of optimized answers and the number of right answers for each of the 13 problems ($R^2=0.06$). This lack of correlation is beneficial since it implies that this method is likely to scale up to more difficult questions.

The overall grade in the class is correlated to several measures of the quizzes. As shown in Figure 3(a), one correlation is the course grade to the ability to answer a quiz question correctly ($R^2=0.72$), which makes sense since the rest of the course was focused on solving problems in one of several formats, such as homework assignments, exams, and lab reports and matches the results described by Petr¹². Another correlation is that the students who performed better in the course were generally able to better assess the correctness of their answers, however the relationship has some variability as shown in Figure 3(b). As students learn and do well on assignments/quizzes, their confidence goes up, which likely encourages learning, whereas the relationship can detrimentally occur in the opposite direction¹¹.

Another way to examine the students' performance is to determine how well they optimize their confidence assessment, i.e., how frequently do they obtain the highest number of points given the correctness of their answer. Figure 3(c) shows this in relation to their overall grade. This has a low correlation ($R^2=0.38$), which suggests that students have the ability to optimize each



problem regardless of their overall grade in the class. In other words, the results suggest that students accurately evaluate their own performance regardless of whether they get the right or wrong answers. This is desirable in the real-world where, regardless of how they perform, if they can assess their own performance, they can relate useful knowledge to colleagues. Common grading schemes do not encourage this meta-analysis of one's answer, thus many students do not get practice in this skill and tend to just present their solution as being viable even if they know it is not good.

This grading method also provides an additional metric to determine which students need help in the class. Clearly those that are getting the wrong answer need to learn the material better, but this group is easily identified based on traditional measures. Another group of students that need help are those that think they know the material, but do not – this is the group that fails to optimize their answers and are shown in the bottom of Figure 3(c). This group has typically been hard to identify and are unlikely to improve on their own since they think they have a good handle on the material. Identifying these students early on in the semester and reconciling their perceptions about their answers and the actual answer will help them build a better intuition about the material.

At the end of the semester, students filled out an anonymous survey regarding the class as a whole with several open-ended questions specifically asking about the confidence-based grading method. The survey was conducted in class with an 83% response rate. One question asked if the students liked or disliked the method: 74% of the students liked the method and 26% of them disliked the method. Many that did not like the confidence-based grading stated they would prefer partial credit since they can show that they are on the right path and they believe their grade would have been higher. Two open-ended questions asked about their thoughts on the method and whether it helped them understand the material better. Below are several representative statements from the responses:

- “As a student I find it annoying, but since I had an internship last summer I can appreciate the principle behind it.”
- “We as graduating engineers should not only know the material taught, but also know how well we know it.”
- “It’s definitely a double edged sword. However, it has gotten me to look at each question with a new perspective.”
- “The ‘not-confident’ answer made me second guess myself at times but was an indicator that I didn’t fully grasp some of the subject matter.”
- “I am too nervous in making a mental error and get zero points when saying I am confident.”

Future Work

This method of incorporating students’ confidence as discussed is well suited for questions that do not have any partial credit, but requires some adjustments to make it work with tests where partial credit is appropriate. Although there are likely many ways to extend this method, below are three possible ways to integrate the confidence-based grading method with partial credit.

- 1) Divide each question on the test into simpler pieces with “milestones” such that students who are not sure can mark the answer to certain portions as not-confident, then put down a default answer and continue on using the default answer. This does increase the effort required in grading as each problem may have two possible starting points and, thus, two answers, but it does allow each student to demonstrate the areas in which they are proficient.
- 2) Grade all-or-nothing if a student marks confident such that the grade is 120% of its original value if correct and 0% if wrong, or grade as normal with partial credit if not-confident. This

method would make grading easier since confident answers are all or nothing, but is likely to significantly increase the grades of the students that are doing very well in the course and will do little for the majority of students. A student would have to be very confident in an answer to risk this much on stating that they are confident.

- 3) Grade as normal with partial credit, but if not-confident, the grade is scaled between 40% and 80% of the normal grade, so zero points would equate to 40% and 100% would equate to 80% with a linear transition between. This would have a minor increase in the effort required to grade, but provides a fair benefit for many students.

The ultimate goal of this confidence-based testing method is intended to encourage students to more adequately convey their confidence in their solution when on the job. The study performed here has only examined their performance on classroom quizzes. To evaluate the transfer of knowledge, follow up studies need to be performed to see how this method affects their job performance. Additionally, followup studies could correlate students' ability to optimize their answer to their risk aversion based on a personality profile. It is likely that some students are risk averse and will tend to select not-confident more often than they should. Future studies should additionally compare the confidence metrics to other demographic data, such as GPA, age, gender, culture, transfer status, class (e.g., junior, senior), SAT scores, credit hours taken, work experience, future career plans (e.g., industry, grad school), etc.

This study shows that students are reasonably good at correctly assessing their answers, but future studies should evaluate how this method affects their learning and understanding of the material. Whether or not they learn the material better, this method provides them additional opportunities to practice assessing their own abilities, which is a practical skill that is often overlooked in engineering education.

Acknowledgments

The author thanks Gigi Yuen-Reed, Ismet Handzic, and Samuel McAmis for their insights and feedback.

References

1. Bandura, A. (1977), 'Self-efficacy: toward a unifying theory of behavioral change.', *Psychological review* 84(2), 191.
2. Battigalli, P. & Bonanno, G. (1999), 'Recent results on belief, knowledge and the epistemic foundations of game theory', *Research in Economics* 53(2), 149--225.
3. Battigalli, P. & Siniscalchi, M. (2002), 'Strong belief and forward induction reasoning', *Journal of Economic Theory* 106(2), 356--391.
4. Brown, I. & Inouye, D. (1978), 'Learned helplessness through modeling: The role of perceived similarity in competence.', *Journal of Personality and Social Psychology* 36(8), 900.
5. Carberry, A.; Lee, H. & Ohland, M. (2010), 'Measuring engineering design self-efficacy', *Journal of Engineering Education* 99(1), 71--79.

6. Davies, T. (2000), 'Confidence! Its role in the creative teaching and learning of design and technology', .
7. Fantz, T.; Siller, T. & Demiranda, M. (2011), 'Pre-Collegiate Factors Influencing the Self-Efficacy of Engineering Students', *Journal of Engineering Education* 100(3), 604--623.
8. Gecas, V. (1989), 'The social psychology of self-efficacy', *Annual review of sociology*, 291--316.
9. Gibbons, R. (1992), *A primer in game theory*, FT Prentice Hall.
10. Hutchison, M.; Follman, D.; Sumpter, M. & Bodner, G. (2006), 'Factors influencing the self-efficacy beliefs of first-year engineering students', *Journal of Engineering Education - Washington* 95(1), 39.
11. Parsons, S.; Croft, T. & Harrison, M. (2009), 'Does students' confidence in their ability in mathematics matter?', *Teaching Mathematics and its Applications* 28(2), 53--68.
12. Petr, D. (2000), Measuring (and enhancing?) student confidence with confidence scores, in 'Frontiers in Education Conference, 2000. FIE 2000. 30th Annual', pp. T4B--1.
13. Ponton, M.; Edmister, J.; Ukeiley, L. & Seiner, J. (2001), 'Understanding the role of self-efficacy in engineering education', *Journal of Engineering Education - Washington* 90(2), 247--252.
14. Rasmusen, E. (2007), *Games and information: An introduction to game theory*, Wiley-blackwell.