Research shows that low-achieving students are less able to accurately assess their own weaknesses. As a result, many might fail to see the need to explore the subject matter more deeply, in order to improve their conceptual understanding and procedural fluency. This study investigates undergraduate mathematics students' self-assessment behaviors. Students from a broad range of courses at three universities were asked to predict their expected grades on assignments, and these predictions were compared with the grades assessed by their instructors. They were also asked to justify their self-assessments if they did not give themselves full points. Preliminary results showed that students overall overestimate their grades. There was a significant difference between expected and actual grades. As test scores increased, the difference increased from negative to positive. Students in the B-range (between 80-89%) were the most accurate predictors.

Key words: [Self-assessment, Undergraduate Mathematics Teaching, Metacognition]

Studies suggest that higher-ability students also have better metacognitive skills (Chi et al, 1989; Recker & Pirolli, 1992; Shute & Gluck, 1996; Wood & Wood, 1999), and that low-achieving students are less accurate when assessing their own weaknesses (Langendyk, 2006). One implication of these findings is that many low-achievers may not see the need to explore the subject matter more deeply, in order to improve their conceptual understanding and procedural fluency. Students think that they are doing “just fine” even if their knowledge and performance are weak (Kruger & Dunning, 1999). This over-confidence could be a factor contributing to students’ lack of success (Langendyk, 2006). Students often do not know when they need help or what form of support is appropriate (Aleven & Koedinger, 2002), and low-achieving students need external support in order to link assessment to learning (Langendyk, 2006).

Undergraduate mathematics courses have one of the highest dropped, failed or withdrawn (DFW) rates (Gardner Institute, 2013). However, little research has investigated mathematics students’ self-assessment of performance, or the reasoning for their self-assessments. This study addresses these gaps in the literature by investigating the following research questions: 1) How accurately do undergraduate mathematics students self-assess their performance? 2) How do self-assessments of successful and unsuccessful performers compare? 3) What reasons do students give to justify their self-assessment of their performances? 4) How does students’ self-assessment accuracy affect their self-regulated learning behaviors?

At this point, we only have information pertaining to the first two research questions.

Literature Review

This study is based on the theoretical framework of “meta-ignorance,” also referred to as the Kruger-Dunning effect (Dunning & Kruger, 1999; Kruger & Dunning, 1999). This framework asserts that people’s ignorance is often invisible to them, because “lack of expertise and knowledge often hides in the realm of the “unknown unknowns” or is disguised by erroneous beliefs and background knowledge” (Dunning, 2011, p. 248).

Research shows that good students have better metamemory (a conscious awareness of ones own processes with respect to memory) accuracy than do poor students, and are better
able to predict what they know and do not know (Sinkavich, 1995). Dunning and Kruger (1999) found that when asked to rank their performances relative to peers, bottom-quartile students overestimated their performance, while top-quartile students underestimated. Self-assessments are more likely to be inaccurate on difficult tasks for which people lack requisite knowledge (Lichtenstein & Fischhoff, 1977). If the task is too difficult or the person is unskilled, there is a greater likelihood of overconfidence (Dunning & Kruger, 1999).

Dunning and Kruger (1999) assert that people unaware of their incompetence suffer a dual burden, as “not only do they reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the ability to realize it. Instead, they are left with the mistaken impression that they are doing just fine” (p. 1121). They further argue that such people are likely to get stuck and become unaware of their incompetence, because “the skills that engender competence in a particular domain are often the very same skills necessary to evaluate competence in that domain” (p. 1121). Incompetent individuals are unlikely to recognize correct judgment if they cannot produce correct judgment. In other words, they lack metacognition skills (Everson & Tobias, 1998).

Research shows that good students are more successful in the metacognitive task of evaluating their own performances, such as anticipating which test items they will get right or wrong (Austin, Gregory, & Galli, 2008; Sinkavich, 1995). A survey of 15-year old students across 34 countries showed that higher performance and more accurate self-perceptions of math skills are associated with each other (Chiu & Klassen, 2010). However, good students also underestimate their abilities when comparing themselves to peers (Dunning, 2011), a behavior known as the “false consensus effect” (Ross, Green, & House, 1977). Below-average students, on the other hand, falsely believe that they are above-average (Ferraro, 2010). When students are given their peers’ work to grade, good students improve their ranking accuracy much more than their weaker peers (Dunning, 2011).

Dunning and Kruger (1999) found that students are willing to rate themselves more negatively if they are equipped with intellectual resources. They showed that even a short 20-minute lesson to solve a certain task improved students’ logical reasoning and self-assessment accuracy. Research also shows a link between self-assessment and learning if students use their knowledge to formulate new strategies and learning goals (Boud & Falchikov, 2006). However, making people aware of their limitations does not necessarily induce them to overcome their limitations (Prasad et al., 2009). Many of them are unwilling to anticipate their incompetence even if they receive feedback on their work (Hacker, Bol, Horgan, & Rakow, 2000; Ferraro, 2010). Such unwillingness may be caused by self-esteem, self-defensiveness, or the difficulty they experience when trying to improve (Sheldon et al., 2014).

Little research has investigated undergraduate mathematics students’ self-assessment behaviors, how they justify their self-assessments, and whether their self-assessment accuracy impacts their self-regulated learning behaviors. We hope that the results of this study would inform future mathematics instruction and the design of professional development activities for instructors to help students become better learners.

Methodology

Four faculty researchers collected data in their respective universities: a private university in central Georgia, and two public universities in north and southwest Georgia. Data were collected from 229 students in a broad range of undergraduate courses taught by the researchers: introduction to mathematical modeling, college algebra, elementary statistics, calculus I, II, and III, differential equations, and mathematical probability and statistics.
Students in these courses were given an initial survey asking their self-reported readiness to take the course and their expected end-of-semester grades. Students were asked to write their expected grades for all in-class quizzes and exams, which were graded based on the instructors’ grading rubrics. Students were also asked to justify their self-assessment if they did not give themselves full points in those problems. Since they were asked to write their expected grades at the bottom of each assignment, they were able to compare the two scores after the assignments were returned to them.

Students’ self-evaluation of their performances was also measured through a short survey after each exam. Toward the end of the semester, a purposeful sample of students (those who consistently overestimated, underestimated, or made almost accurate predictions of their scores) was asked to voluntarily participate in a few semi-structured interviews. They were reminded of the differences between the two scores and asked to explain their perceived reasons for the inaccuracy (or accuracy) of their assessment. This report is based on data from the spring 2016 semester, but does not include qualitative data from the interviews. Data will also be collected in the fall 2016 semester and a comprehensive study of the combined data will be made.

Based on existing research and our personal experiences, we hypothesized that top performers would be more accurate predictors of their scores, bottom performers would overestimate themselves and be less accurate predictors, and students would become better predictors as the semester progressed.

**Results**

This paper reports preliminary results from quantitative analysis pertaining to only the first two research questions. We have used both descriptive and inferential statistics. An independent t-test was used to determine the statistical significance of the average difference between the students’ expected grades and the grades assigned by the instructor. Pearson’s correlation test was used to determine the significance of the relationship between the students’ predicted and actual grades. We found that students overall overestimated their scores. The test shows a statistically significant difference between their expected and actual grades ($t (1799)= -6.89$, $p < 0.01$), with a mean difference of 5.85 on a 100-point scale. For A students (scoring 90-100%), the difference was still statistically significant ($t (381)= 7.84$, $p < 0.01$), but they underestimated their performances by an average of 5.22 points on the 100-point scale. B students (scoring 80-89%) slightly overestimated themselves, but the difference between predicted and actual grades was not statistically significant. The mean difference was only 1.42 points.

For C students only (70-79%), the difference was statistically significant ($t (150)= -6.58$, $p < 0.01$). They overestimated themselves on average by 8.43 points. Students in the D range and below (69% and less) were the most miscalibrated, overestimating their performance ($t (500) = -14.4$, $p < .001$) by 20.25 points on a 100-point scale. The line graph (see Figure 1) shows how the difference between instructors’ average grade and students’ expected grade changes in relation to the average instructors’ grade. This shows that bottom performers tend to overestimate themselves, and top performers tend to underestimate. The graph suggests that students in the B-range (80-89%) were more accurate predictors of their performances.

Preliminary test results did not support our initial hypothesis that top performers are more accurate predictors, as B students actually proved to be the most accurate. But the results supported the hypothesis that bottom performers are less accurate predictors than others.

The regression line in Figure 2 shows the relation between average grades vs. average differences. The two variables are strongly correlated, $r (89) = 0.67$, $p < 0.01$. This shows that
as the test scores increase, the difference between the instructor’s grade and students’ predicted grade increases from negative to positive.

![Figure 1. Average Instructors’ Grades vs. Average Differences](image)

We also looked at how the differences between predicted and actual grades changed as the semester progressed. Students overestimated their performance by 5.58 points on the first quiz, then underestimated by a mere 0.74 points on the second. Interestingly, they then overestimated by 5.7 and 6.3 points on the third and the last assignments (quizzes or tests). This did not validate our initial assumption that students would be more accurate predictors of their performances as the semester progressed.

![Figure 2. Average instructors' grades vs. average differences](image)
Discussion

We had little information about college math students’ self-assessment behaviors prior to this study. The results show that they are likely to be more miscalibrated when they either perform well or do not perform well, as also found in existing research (Kruger & Dunning, 1999; Langendyk, 2006). Both top (A students) and bottom performers (C and below) were inaccurate predictors. Top performers slightly underestimated their scores, while bottom performers overestimated by a huge margin. More interestingly, B-range students turned out to be the most accurate predictors, as the difference between their expected and actual grades was not statistically significant. This was in agreement with previous findings that students in between the top and bottom performers are more accurate predictors (Langendyk, 2006). We do not know why this group was able to better predict their performances, but we will have better insight once we finish analyzing our qualitative data.

Existing research shows that top performers in general underestimate their performance when comparing themselves to their peers (Dunning, 2011). Our study showed that top-achievers in college math courses still underestimate themselves, even when they are not comparing themselves with their peers. Since we haven’t analyzed our qualitative data yet, we do not know why this group of students underestimate themselves. Either they are not completely confident that their solutions are correct, or they think they have not met instructors’ expectations. The false consensus effect (Ross, Green, & House, 1977) in this group of students might be encouraging them to always work harder.

Lack of knowledge may prevent poor performers from knowing what they have done wrong (Dunning, 2011; Kruger & Dunning, 1999). If they had the ability to recognize right or wrong solutions, they probably would have been able to better predict their scores. Since they did not know that their solutions were incorrect, they might have been overly optimistic about their performances because they were hoping for a passing grade. Because of their mistaken belief that they know the subject matter, they might not even realize that they need to work harder to gain deeper understanding of the subject matter and gain procedural fluency.

We did not find studies that investigated college math students’ self-assessment behaviors as we did in this study. One implication of our findings is that both low- and high-achievers need instructor support. Low-achievers need help figuring out conceptual and procedural gaps in their knowledge. They often do not know when they need help and what kind of support they need (Aleven & Koedinger, 2002). High-achievers, on the other hand, need help figuring out that what they know is actually correct. Preliminary findings show that they know the content but still lack confidence in their knowledge, as evident from their predicted scores. Since the students in between these two groups know, in general, whether they did right or wrong, it shows that these in-between students (mostly B students) have potential to improve. We will have better understanding about students’ self-assessment behaviors once we finish analyzing their justifications for their self-assessments.

Results from this study showed that our initial assumption that students would become more accurate predictors of their performances as the semester progresses was not necessarily true. Even though they made almost accurate predictions of their scores in the second assignment (a quiz or a test), they then overestimated their performances on the next two quizzes or tests. Seeing the difference between their expected and actual grades on the first assignment might have made them more cautious about expecting higher scores on the second assignment. But why the students then became overly optimistic on the next assignments is unknown. We need to analyze interviews and collect more data to make any conclusions about this student behavior. This can be a good topic for further investigation.
References


Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159-183.


