Improving Metacognition in the Classroom

Jacquelynn Jordan

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Faculty Excellence, United States Military Academy, West Point, NY, 2014.

Abstract

This study seeks to understand how metacognition may impact students during exams throughout a semester and is interested in determining whether or not metacognitive training can improve performance across this timeframe. To explore this question, two hypotheses were studied: (a) that metacognition can be improved across a semester through guided reflection, and (b) that high performers will underestimate their performance but be the most accurate, middle performers will be accurate but will overestimate their scores and low performers will be inaccurate and will grossly overestimate their scores. The results show mixed support for the first hypothesis with a non-significant general trend toward metacognitive growth in the experimental group versus the control group. The second hypothesis was generally supported as the high performance group’s accuracy and confidence were significantly more accurate and lower than in the low performance group. The data shows insignificant data however for the difference in accuracy between the low and average performers, while the low performers showed significantly higher confidence than did the average performers. The results suggest that with more research on larger populations with more strict controls, it may be possible to show an increase in metacognition over the course of the semester. This study shows the importance of this research because higher metacognition is correlated with higher performance.

Keywords: Metacognition, students, performance
Improving Metacognition in the Classroom

Throughout my first year of teaching, I was struck by the observation that many of my students were surprised, and generally not pleasantly so, by their performance on exams. After the exams, students attended additional instruction and lamented that they had spent several hours studying and just didn’t understand what had gone wrong - they even left the test pretty confident. I would then pose fairly basic questions about the material and they were unable to answer correctly – they were not good at judging what they knew, even after getting their results back. I began thinking about what I teach on metacognition and realized that a lack of metacognition seemed to be the real issue in students’ failure to perform to their expectations on exams.

Literature Review

Educators and researchers alike are constantly looking for ways to better educate future generations even while increased reliance on standardized tests are limiting knowledge and creative thinking (Kuhn & Dear, 2004). College instructors often assume that students walking into classrooms know how to learn, though many students have never actually been taught how to do so effectively. Metacognition may be the key to creating more successful learners. Lin (2001) defines metacognition as “the ability to understand and monitor one’s own thought and assumptions and the implications of one’s activities” (p. 23). This definition involves self-direction, self-regulation and motivation, all of which are supported by other research as key pieces of metacognition (Ganz & Ganz, 1990; Cleary & Zimmerman, 2004).

Further research shows that there are large differences in metacognitive skills between high and low performers (Bol & Hacker, 2001; Hacker, Bol, Horgan & Rakow, 2000). High performance is correlated with high metacognition, whereas low performers are not aware of
what they do not know (Bol & Hacker, 2001; Hacker, et al., 2000; Lin, 2001; Roll, Aleven, McLaren & Koedinger, 2011). Furthermore, students who have poor content knowledge lack the ability to understand that their strategies for learning are ineffective and often blame their poor performance on external factors (Hacker, et al., 2000; Cleary & Zimmerman, 2004). What is particularly troubling about low metacognition is that it lowers self-efficacy and undermines motivation, ultimately leading to lower values placed on learning (Ganz & Ganz, 1990; Cleary & Zimmerman, 2004).

If metacognition is so important in learning then something must be done to help students develop this skill. Having students merely reflect on their performance does not increase metacognition due to the passive nature of reflection, students must be guided (Nietfeld, Cao & Osborne, 2005). Metacognitive skills can be developed and must be trained (Lin, 2001; Goos, Galbraith & Renshaw, 2002; Cleary & Simmerman, 2004; Kuhn & Dear, 2004; Boekaerts & Corno, 2005; Dun, Lo, Mulvenon & Sutcliffe, 2012).

The factors that impact students’ motivation to learn though are not only impacted by learning oriented goals, and educators and researchers must be careful to consider the social nature of learning in developing metacognition (Goos, et al., 2002; Kuhn & Dear, 2004; Boekaerts & Corno, 2005). By effectively capitalizing on the relationships within the learning environment to develop metacognition, students can become more self-regulated, self-directed, and motivated learners (Goos, et al., 2002; Kuhn & Dear, 2004; Boekaerts & Corno, 2005).

How then should metacognitive training be conducted in the classroom? Bol and Hacker (2001) found that practice tests are not effective because students do not use the tests effectively and become overconfident in their knowledge, thus performing more poorly on the exams. Cleary and Zimmerman (2004) assert that students must set self-directed goals and learn how to
use effective learning strategies from models before using them independently and gaining feedback. Kuhn and Dear (2004) also highlight the role of feedback through evaluation but believe in supporting and encouraging students in the use of prompted reflection.

In an attempt to aid students in their mastery of understanding study skills early in their collegiate careers, I conducted a study to see whether or not metacognition is tied to performance. I hypothesized that better metacognition correlates with higher performance on tests, that metacognition can be improved across a semester through guided reflection, and that high performers will underestimate their performance but be the most accurate, middle performers will be accurate but will overestimate their scores and low performers will be inaccurate and will grossly overestimate their scores.

**Method**

**Participants**

Participants were 67 undergraduate freshmen and 2 undergraduate sophomores enrolled in a required general psychology for leaders course during the fall term at the United States Military Academy at West Point. Of the 69 participants, 7 were female, 62 were male; the school has a female population of roughly 17%, so women are underrepresented in the sample (10%). All students of the class of 2017 were equally likely to be placed into the sections participating in this study, therefore the simple random sampling conducted allows the results to have the least bias and the most generalizability.

The participants were then randomly assigned to either the control or experimental group. In order to ensure that the control and experimental groups were approximately equal in their academic ability prior to the treatment being conducted, their performance on the first test in the course was compared. The control group averaged an 82% (8.06% standard deviation) while the
experimental group averaged an 81% (12.54% standard deviation). The difference between the groups was not significant \((p = .709)\), meaning that the two groups were evenly paired based on academic ability.

**Materials**

On the final lesson before each exam, beginning with exam two, the experimental group was given a worksheet for guided reflection. On this sheet each student evaluated their knowledge of each performance objective covered over the block of instruction to be tested. In so doing, they were able to begin to see the gaps in their knowledge.

Performance was measured over five exams throughout the semester. The first test consisted of 25 multiple-choice items and a total of 50 points worth of short answer items, equating to ten application questions; this test served as a baseline and no treatment was received prior to this exam. The second test was similar to the first in that it consisted of 25 multiple-choice items and a total of 60 points worth of short answer questions, equating to eight application questions. The third and fourth exams each consisted of 50 multiple choice items. The final exam consisted of 125 multiple-choice items. All items were either adapted from the test bank which accompanies the textbook, or created by the instructors teaching the course.

Metacognition was tested through asking students to predict and post-dict their performance on each test on a sheet of paper located on the door of the testing classroom. They entered their estimates by participant number in an attempt to reduce social desirability bias.

**Procedure**

The four sections of the general psychology for leaders course were taught by the same instructor. Two randomly assigned sections were assigned to the treatment condition involving metacognitive training; the other two sections received no training.
The instruction for the four sections was equivalent with the exception of the metacognition training. The students in the treatment sections were administered a guided reflection study sheet on which they were prompted to think about each key concept and how well they knew the material for each performance objective. They were given no more than ten minutes to complete the reflection. After the students responded to the reflection, they were allowed to retain the form to use as a study tool.

The students in the control group received instructor-led review of the content for the same amount of time, but did not receive a guided reflection sheet. An outline of major concepts was presented on a power point slide to guide the review.

Before distributing the five exams, students were told that their instructor was conducting research on students’ ability to predict and postdict their exam performance. As students were entering each of the exams, they predicted their performance on a sheet of paper outside the room. Upon finishing the exam, the students post-dicted their scores on a sheet of paper outside the room.

**Results**

The results are reported according to my hypotheses: (a) that metacognition can be improved across a semester through guided reflection, and (b) that high performers will underestimate their performance but be the most accurate, middle performers will be accurate but will overestimate their scores and low performers will be inaccurate and will grossly overestimate their scores.

**Metacognition can be improved across a semester**

Due to missing data across the semester, only a subset of the data was able to be used to measure improvement across the semester. This limited data included only 32 students’
predictions, test scores, and postdictions across the semester. This sample showed an average prediction of 6.8% higher than performance for the experimental group and 9.54% higher than performance for the control group. The average postdiction for the experimental group was 6.4% higher than performance and 7.79% higher than performance for the control group. Compared to the entire sample’s prediction (average 11% higher) and postdictions (average 10.55% higher), it is possible that the subset analyzed here is a different type of performer.

The guided reflection used in this study appeared to have some impact on students’ metacognition throughout the course of the semester as can be seen in the data represented in charts 2.1 and 2.2. The results however, did not prove to be significant (p = .415 and .277 for predictions and postdictions respectively). Therefore, the hypothesis that metacognition can be improved across a semester is not supported by the current research.

Prior to each exam in the course, the students are afforded the opportunity to meet with their peers in a group review session facilitated by the instructor. These review sessions are entirely voluntary but are held during the dropped class prior to the exam, so all students can attend if they choose. While not necessarily indicative of their knowledge of their own preparation for the exams, it is interesting to note that the students in the experimental group
attended the group review sessions at a rate of 71.9% while those in the control group only attended at a rate of 62.5%. The difference in the number of students attending additional instruction is not significant, $p = .12$, but the information helps paint a picture about students’ willingness to put forth extra effort into the course.

At the end of the course, students are afforded the opportunity to provide feedback about the course. The control and experimental groups were equally likely to strongly agree or agree with the statements: “the class objectives were reflected in the exams” and “the graded requirements allowed me to demonstrate how well I had mastered the course material”. The control group, however, was more likely to strongly agree (33%) with the statement, “you expected to get a higher grade in this course than you have now” than the experimental group (14%). The difference in the responses were statistically significant ($p = 0.015$), showing that the experimental group had more realistic expectations by the end of the course for their performance in the course.

**Performers, Accuracy and Confidence**

In order to examine accuracy and confidence of predictions across performance groups, the overall course grades were considered for the students. Those students who performed in the above the average for the exam are considered high performers, those who scored within one standard deviation below the average, average performers, and those who scored below one standard deviation below the average are the low performers. The descriptive statistics show that high performers tend to be more accurate in estimating their performance both in pre- and postdictions when compared to the medium and low performers. The results showed statistically significant differences for predictions between the low and high group, $p = .04$. For postdictions,
the results were significant between both the low and high groups and the medium and high groups, \( p = .019 \) and \( .04 \) respectively.

Based on the data above, the low performers are clearly far more inaccurate in their estimates of their performance both before and after taking the exam than are the high performers. However, their results are not significantly different that the estimates of the average performers.

Additionally, when looking at the confidence of the performance groups, the high performers routinely underestimated their performance both before and after taking their exam, and they were the only ones to do so. The gross overestimation of performance by the low performers, was significantly different than the estimates by the average performers (prediction \( p = .04 \), postdiction \( p = .03 \)) and the estimates by the high performers (prediction \( p = .02 \), postdiction \( p = .018 \)). The overestimation by the average performers was also significantly different than the underestimates of the high performers (prediction \( p = .009 \), postdiction \( p = .011 \)).
Chart 2.3. The confidence of performance on the tests prior to the exam across the semester.

Chart 2.4. The confidence of performance on the tests after the exam across the semester.

Discussion

The current literature about metacognition looks into how accurate metacognition increases the performance of students as well as ways to increase metacognition. Clearly
IMPROVING METACOGNITION

accurate metacognition is an important tool for learners, however, there is not a good definition as to what “accurate” metacognition is. In the sample tested here for improvement over time, students predicted and post-dicted their scores within 10%, or within a letter grade beginning on the first test where metacognition was measured. Improvement across time was not significant, but this could potentially be the case because 10% is accurate to begin with and this is a result of a ceiling effect.

Based on the current research, the low performers had an average pre- and post-diction which differed by 19.89% from their scores. The average performers scores were more accurate with a 8.12% difference and the high performers were most accurate with a 4.4% difference in accuracy. Based on this data, it would be reasonable to judge “good accuracy” somewhere below 8%. “Good” should not equate with “average” and therefore an accurate rating should be what you would get from someone who does better at estimating performance than the average learner. If this study were to be replicated, we could perhaps get closer to understanding what good accuracy is.

Threats to Validity

Some threats to validity occurred in this study which may have limited the ability to precisely assess students’ metacognition and the improvement of it across time. In an attempt to ensure anonymity for the students, the students wrote their pre- and post-dictions on a spreadsheet taped to the front door of their testing room. They did not identify themselves by name, but by participant number, which allowed some protection from judgment which one would imagine would increase their likelihood of entering their true expectations for their performance. What was not anticipated was the time each student would take standing in front of the piece of paper. What appeared to be happening was that students were assessing the
anticipated scores of other students prior to entering their own pre- and post-dictions. By having the scores publicly available, social desirability bias potentially occurred. If most of the students who entered their scores prior to a participant thought they would perform well, perhaps that swayed the prediction of a participant high; or the opposite could have occurred.

While the guided review was only presented in class to the experimental group, there is the possibility of contamination as I encouraged my students to use each other to gain a better understanding of the material. If students from the control group studied with students from the experimental group, it is possible that they gained the potential benefits of the handout. I do not suspect that this was a big issue in the study, however, because no student from the control group ever asked me for the handout or questioned why they did not receive it.

Another potential threat in this study was the lack of rigidity in applying the experimental condition. The course is not lecture style in nature and the open dialog with the students, on occasion, made the lesson run long, meaning valuable time was cut from both the control and experimental review or guided reflection. The insignificant results then, may have resulted because of improper application of the method. If this is better controlled for, it is likely that the trend of the results as shown in this study would become significant.

**Conclusions and Future Study**

When taken in combination with the other research performed in this area, it is important for instructors to understand that increasing metacognition is important to student performance. Based on how inaccurate some participants were, it is clear that even college professors cannot make the assumption that students have had any sort of metacognitive training prior to college and spending some time on this throughout a semester shows promise for increasing the ability, self-efficacy and motivation of our students. It is important for more research to be conducted in
this area with stronger controls in order to show whether or not metacognitive accuracy can be improved over the course of a semester and which tools can be used effectively to do so.

As an instructor, I hope that my students take something of value away from my course no matter their performance on graded events. It is my hope that through this project, some students’ awareness about their own knowledge and understanding was improved. If this is the case, the study was beneficial.


