

Self-Assessment in an Interteaching Setting:

Can Students Grade Their Own Work?

Bonnie Reed

California State University, Chico

Abstract

Previous research has shown that courses taught in an interteaching style have been highly effective in increasing students' participation and achievement. Other research has shown that students' self-assessment of their work has produced similar results. We explored the effects of combining the two methods in an undergraduate psychology class expecting that achievement and learning would be exponential, in addition to increasing metacognition. We also expected to see high consistency and accuracy in students' grading abilities. We found, in agreement with past research, that students tended to slightly over-estimate their grades when self-assessing, but that they enjoyed the immediate feedback and the ability to see their own mistakes. We did not notice an improvement in metacognition, as their ability to predict their scores remained the same, which was high to start with. We also did not find a significant improvement in learning due to self-assessment as measured by exam scores, although there may have been other learning-enhancing facets to the course that drowned this effect. We conclude by discussing possible limitations and future research for improving the combination of these methods in an attempt to enhance learning and encoding in college courses.

Self-Assessment in an Interteaching Setting:

Can Students Grade Their Own Work?

Student self-assessment is commonly taught and utilized in educational settings, due to the many benefits that come from its use. It is not as often, however, that we see this method utilized in higher education (Butler & Lee, 2010; Kahrama, 2014; Walser, 2009). Researchers report that this is likely due to two major concerns: reliability (Boud & Falchikov, 1989; Burke, 1969; Davis & Rand, 1980; Logan, 2015; Strong, Davis, & Hawkes, 2003; Zoller & Ben-Chaim, 1988) and the perceived increased time and work an instructor would have to invest into implementing self-assessment (Boud & Falchikov, 1989; Matsuno, 2009). The purpose of our study is to find solutions to these potential downfalls to self-assessment, as well as to show that, with these solutions, self-assessment can be a valuable tool in higher education.

Self-assessment appears to have many benefits at any grade level. Most importantly, self-assessment has been shown to improve learning and student performance. For example, research in this area has found a significant increase in quality of work, as well as increased objectivity regarding the self-assessment procedure over time and with practice. These same studies have also shown to improve students' skill in the areas of class work, exam scores, course grades, and overall achievement (Arter, Spandel, Culham, & Pollard, 1994; Johnson & Gelfand, 2013; Rolheiser & Ross, 2001; Ross, 2006; Ross, Hogaboam-Gray, & Rolheiser, 2002a; Ross, Hogaboam-Gray, & Rolheiser, 2002b; Ross, Rolheiser, & Hogaboam-Gray, 1999; Schunk, 1996; Strong, Davis, & Hawks, 2003; Sung, Chang, Chiou, & Hou, 2005). Data from these studies also indicate that the students who typically performed the poorest in their classes showed the most improvement in class performance, self-efficacy, and course achievement when utilizing self-

assessment as compared to traditional instructor-grading (Arter, Spandel, Culham, & Pollard, 1994; Rolheiser & Ross, 2001; Ross, Rolheiser, & Hogaboam-Gray, 1999).

Besides the multiple learning benefits of self-assessment, metacognitive capabilities also appear to improve with this method. Self-assessment has helped students in higher education university courses to monitor their own progress, has allowed for reflection of their work and of the material, has given students more opportunities to give feedback to the instructor, and has increased interaction and communication between students as well as with the instructor (Burke, 1969; Orsmond, Merry, & Callaghan, 2004; Rolheiser & Ross, 2001; Thomas, Marin, & Pleasants, 2011; Walser, 2009; Zoller & Ben-Chaim, 1988). Edwards (2007) found that students who participated in self-assessment in a college-level social statistics class reported that they enjoyed being able to identify their own mistakes and receive immediate feedback on their work. Taking the responsibility to self-grade can also reduce student anxiety, as well as student-instructor conflict (Edwards), and students' general tendency for skepticism of their course grade validity decreased (Rolheiser & Ross, 2001). Additionally, self-assessment allows students to understand the grading process more comprehensively, as well as gain a deeper learning and understanding of the material itself (Dart & Clarke, 1991; Logan, 2015; Sadler & Good, 2006; Zively, 1967).

Self-assessment does not only have academic benefits; it also supports more general psychological benefits. In higher education courses, students can expect to see individual growth and an improved internal locus of control with the use of self-assessment, (Dart & Clarke, 1991; Logan, 2015; Sadler & Good, 2006). These benefits are not limited only to the classroom, as self-assessment can pave the way to help create life-long learners (Logan, 2015). Studies have shown that student self-assessment tends to increase students' motivation to learn and

participate, improve student self-efficacy, build their confidence in their own abilities, and create an increased sense of responsibility for their own learning (Arter, Spandel, Culham, & Pollard, 1994; Rolheiser & Ross, 2001; Ross, 2006; Ross, Hogaboam-Gray, & Rolheiser, 2002a; Ross, Hogaboam-Gray, & Rolheiser, 2002b; Ross, Rolheiser, & Hogaboam-Gray, 1999; Schunk, 1996; Strong, Davis, & Hawks, 2003).

Given all the benefits of self-assessment, it is fortunate that students embrace this method. Edwards found that with the implementation of self-assessment in his social statistics class, his students gave an “overwhelmingly favorable” (2007, p. 73) response to this method of assessment. He reports that his students saw value in recognizing their mistakes immediately, reviewing their work, and that they learned the material better due to repeated exposure. Students indicated that it was positively reinforcing and that they enjoyed the self-assessment method over the traditional style of instructor grading (Edwards, 2007). In general, students responded positively to the notion and practice of self-assessment, by use of questionnaires and feedback, adapted well to its use, and reported that having this added clarity of expectations for assignments was more reasonable (Davis & Rand, 1980; Edwards, 2007; Rolheiser & Ross, 2001; Strong, Davis, & Hawks, 2003).

Our discussion so far has focused on the benefits of self-assessment to students. However, the benefits are not limited to students but apply to instructors as well. Ideally, self-assessment by students would save instructors time and energy, allowing them to be more available to the students and the grading of other material. When students grade a portion of their own assignments, instructors often find themselves with more time to spend on the other various duties that are required of them (Rolheiser & Ross, 2001; Sadler & Good, 2006). The numerous benefits of self-assessment and the positive feedback given by the students has been shown to

promote growth in instructors' confidence in their own skills, making them more apt to utilize a greater variety of assessment and teaching techniques in the classroom (Rolheiser & Ross, 2001). In turn, these effects create an atmosphere that increases instructor efficacy, thereby promoting student learning (Rolheiser & Ross, 2001).

Despite the benefits discussed thus far, one reason why instructors do not use self-assessment is a concern about students' ability to grade their work in a reliable and consistent manner (Boud & Falchikov, 1989; Burke, 1969; Davis & Rand, 1980; Logan, 2015; Strong et al., 2003; Zoller & Ben-Chaim, 1988). However, encouraging results have been found supporting the idea that students are certainly capable of grading with accuracy and fairness (Goodrich, 1930; Logan, 2015; Lopez-Pastor, Fernandez-Balboa, Santos-Pastor, & Aranda, 2012; Ross, 2006; Sadler & Good, 2006; Sung, Chang, Chiou, & Hou, 2005; Zively, 1967). Findings also indicate that disagreements between student and instructor scores were very low, (Davis & Rand, 1980; Lopez, et al., 2012) and that students, with proper instruction, became more accurate and objective in their grading ability over time (Butler & Lee, 2010; Zively, 1967). However, the accuracy of self-assessment, the decrease in grade disagreements, and the increase in academic performance are most likely to occur when the instructor shares in the responsibility of grading; if the students expect the instructor to check their assessments they are more likely to perform evaluations with high validity. Researchers suggest that teaching the students how to properly grade their work, as well as using a rubric to establish guidelines for grading, will ensure the most accuracy, highest validity, and highest level of consistency (Andrade & Boulay, 2003; Rolheiser & Ross, 2001; Zively, 1967). Furthermore, Matsuno (2009) found that, compared with self-assessment and instructor's assessments, peer-rating was the most consistent. This indicates that an ideal level of consistency, validity, and confidence can be reached in employing this

method in higher education if students are taught how to properly self-grade, have other students to work with to whom they were equally accountable, and are provided with a model rubric for grading from which to guide their assessments. If this practice were to become an accepted standard, thereby offering more options for the students who may not fare well with traditional methods, students and instructors both would become more practiced and efficient in employing self-assessment. With all of its benefits, it is widely agreed that self-assessment is an important tool for academic performance and improvement (Hotard, 2010).

Considering that peer-assessment can be highly valid and consistent, and also noting that student accountability helps to encourage grading candor and fairness, we propose combining the self-assessment approach with a method of collaborative instruction known as interteaching. We herald this new method with the idea that interteaching will also aid in embedding the class material by repeated exposure, along with the notion that teaching aids in learning. Multiple studies have shown that various benefits come from combining interteaching with traditional lecture format as compared to the classic style of teaching, with interteaching producing student learning above and beyond what lectures can achieve (Arntzen & Hoiium, 2010; Saville, Lambert, & Robertson, 2011; Saville, Pope, Truelove, & Williams, 2012; Saville, Zinn, & Elliot, 2005; Querol, Rosales, & Soldner, 2015). Researchers in this field seek to add more class options for the students who do not always fare as well in their classes as some of the other higher-achieving students, considering that poor-performing students showed the most significant improvement in their course achievement when participating in interteaching compared to traditional lecture (Saville, et al., 2012). Some of the many benefits of interteaching include reciprocal peer tutoring, cooperative learning, increase in academic engagement and student collaboration, and improved satisfaction in the classroom (Querol et al., 2005). Interteaching, a

form of collaborative learning, has its roots in behavior analytic principals (Arntzen & Hoium, 2010; Saville et al., 2005; Saville, et al. 2011; Querol et al., 2015) and is reported to be user-friendly for both instructors and students (Saville et al., 2011; Querol et al., 2015). Dating back to the 1950s, Skinner (1953) suggested behavioral approaches to education, due to the ineffectiveness of aversive consequences used with traditional approaches to classroom instruction. He reports that positive reinforcement techniques can often be quite difficult and time-consuming to employ, and so these are not often used; interteaching may be an effective solution to these issues (Skinner, 1999). Additionally, Kuh (2008) suggests that collaborative learning encourages learners to learn how to work and solve problems with other students from various backgrounds and with differing life experiences, and how to hone their own understanding by listening seriously to the insights of others. Hence, considering the purported downfall of the American education system due to ineffective teaching methods (Saville, et al., 2011), interteaching appears to be a highly beneficial and effective alternative to traditional classroom lecture.

Our goal is to present data on the effectiveness and validity of an approach to student self-assessment that will demonstrate students' ability to grade their own assignments. We introduced the method of self-assessment into a class that was conducted in an interteaching style, where the students come to class prepared to teach each other the material, having done the reading and taken notes before class. The two methods combined offer great promise of profound, active learning on the part of the students, encouraging their independence and self-accountability. This research may also offer a potential solution to creating accountability among students within their groups, thereby decreasing their propensity for grade-inflation. We plan to utilize both methods of teaching in an undergraduate psychology course.

Method

Participants

Data were collected over the course of the Spring 2015 and Fall 2015 semesters. Students who were enrolled in an undergraduate psychology course on Brain, Mind, and Behavior, taught at a 4-year college in Northern California, participated in this study. Enrollment in the Spring semester was 49 students (28 males and 21 females), and in the Fall 43 students (19 males and 24 females). Students were informed that the grading of the interteaching assignments was part of a study investigating the possible benefits of self-grading, and that their grades would serve as data.

Materials

This course utilized a mid-semester evaluation, 12 preparation guides, and 12 answer keys. The mid-semester evaluation included 21 questions asking students to give feedback about the implementation of both interteaching and self-grading in the course, with questions relating to their experiences and preferences of teaching method. Questions 1-17 were answered using a 5-point Likert scale, with 1 meaning “disagree” and 5 meaning “agree”. Questions 18-21 were open-ended questions that allowed the students to formulate their own feedback. See Appendix A.

The preparation guides included essay-style questions relating to the material in the textbook, each one ranging from five to eight questions each. These are the weekly assignments that students are to complete at home and bring to class to collaborate with their group for interteaching. Preparation guides were posted on the course website and contained a mixture of comprehension and application questions. See Appendix B.

The answer keys had for each question 2 parts: a model answer with key elements underlined, and a point breakdown for how to score each of these elements. Each preparation guide had a corresponding answer key. See Appendix C.

Procedure

Author Ms. Reed was a Teaching Assistant for this class while Dr. van den Berg was the course instructor for this class.

The course consisted of 12 sessions using the interteaching method. For each interteaching session there was a preparation guide. For the odd-numbered preparation guides the teacher graded the interteaching assignments, while during the even-numbered preparation guides the students graded their own assignments. When students entered the classroom on interteaching days, they pulled a number from a box containing three sets of each number. Students with the same number formed a group for that day, ensuring that groups were as random as possible and varied each session. Self-grading was done immediately following completion of the assignment with their group. To grade their assignments, the professor provided the answer key, which students returned once grading was complete. To clearly identify grading, students graded their own work during class time using a green pen provided by the instructor. Afterwards, the professor, together with one or more teaching assistants, checked their grading. To provide additional and clearly visible feedback, any changes made to the students' grading were indicated with a red pen. If necessary, the professor adjusted the students' grading. The adjusted grade was recorded for class points. Students' own grades and the professor's grades were compared to investigate student-rigor. Additionally, the students also filled out an "Interteaching Report" at the end of all 12 interteaching sessions, which asked them to provide feedback about the covered material and assignment, as well as to predict the numeric score they

would expect to receive before grading began. This estimate was used to measure the students' metacognitive abilities.

Results

To explore students' rigor, we used a repeated-measures ANOVA to compare the students' self-given grades to the final professor-given grades. Results indicated that students consistently grade their work higher than the professor, Spring semester; $F(1, 29) = 19.62, p < .05$, Fall semester; $F(1, 25) = 51.83, p < .05$. However, using a 10-point scale for grading, differences were statistically significant yet minimal; for Spring semester the students' self-given mean scores ($M = 9.63, SD = 0.33$) were on average 2% higher than the professor-given scores ($M = 9.43, SD = 0.23$), and for Fall semester the students' self-given mean scores ($M = 10.02, SD = 0.24$) were an average of 3% higher than the professor-given scores ($M = 9.69, SD = 0.35$).

The above analysis compared scores given by students and the professor on individual assignments, showing a statistically significant, though negligible, difference. We can also look at the impact of these differences on students' overall interteaching grades. To do so, we compared final interteaching scores over all twelve sessions based on either the students' self-graded scores or the professor's scores. In this course the lowest interteaching score was dropped, with the final score presented as a percentage score out of the possible 110 points for the remaining eleven assignments. There was a significant difference in the scores for the Spring semester; $t(47) = -3.27, p = .002$, and no significant difference in the Fall semester; $t(42) = 1.56, ns$, indicating that students' self-given final interteaching grades were higher than the professor-given final interteaching grades for Spring, and similar in the Fall. However, the difference in means between self-given interteaching grades for Spring ($M = 87.24, SD = 10.7$) and professor-given interteaching grades ($M = 86.45, SD = 10.5$) was 0.8%, suggesting that the difference is

not substantial. For the Fall semester, the (non-significant) difference in means between self-given grades ($M = 85.60$, $SD = 13.54$) and professor-given grades ($M = 85.01$, $SD = 13.22$) was 0.6%, further confirming the insubstantial overall difference.

Given that the self-grading resulted in a significant, yet negligible, difference on overall interteaching scores, we also compared its effect on the overall course grades. Given that the overall interteaching score was only 28% of the overall course grade (the remainder accounted for by exams and other assignments), the impact on final course grades will likely be even less influential. To investigate this, a paired-samples t-test was conducted. We compared the difference between the students' overall course grades, contrasting what they would have received if they had been awarded their self-given grades to what they actually received in the course. A significant difference was found in the Spring semester, $t(47) = -3.32$, $p = .002$, and no significant difference in the Fall semester, $t(42) = -1.56$, ns. The difference in means between the self-given course grades ($M = 87.37$, $SD = 5.97$) and the professor-given course grades ($M = 87.09$, $SD = 5.95$) in the Spring was 0.28%, implying that the students' overall course grade was not largely affected by the difference. This was confirmed by non-significant results in the Fall; professor-given course grades ($M = 84.04$, $SD = 10.29$), self-given course grades ($M = 84.22$, $SD = 10.32$).

In order to examine the metacognition of the students over the course of the semester, we used a repeated-measures ANOVA to compare what the students predicted they would earn to the grade they received. We found a significant difference; Spring semester; $F(1, 5) = 28.61$, $p < .05$, Fall semester; $F(4.6, 59.6) = 7.12$, $p < .05$. Results suggest that the students' metacognition did not improve over the semester; students consistently predicted higher grades ($M = 9.84$, $SD =$

0.11) than they received ($M = 9.36$, $SD = 0.25$) in the Spring, as well as in the Fall; predicted ($M = 9.87$, $SD = 0.19$), received ($M = 9.38$, $SD = 0.17$).

Additionally, we explored the potential of a learning benefit in self-grading due to deeper encoding and re-exposure, and prompt feedback to the material. Comparison of the items on the exams that were directly from self-graded interteaching material ($M = 86.45$, $SD = 10.5$) to those that were solely from professor-graded interteaching material ($M = 86.45$, $SD = 10.5$) revealed no significant difference in the Spring, $t(47) = .013$, ns, and a significant negative difference in the Fall; $t(41) = 2.74$ $p = .009$, self-graded ($M = 80.78$, $SD = 12.31$), professor-graded ($M = 84.25$, $SD = 9.25$). This suggests that self-grading techniques will not necessarily always have an impact on deeper learning and understanding of material. This is further examined in the discussion.

Lastly, we analyzed the data from the mid-semester evaluation questionnaire that the students completed. A factor analysis was conducted to identify possible groupings of questions and two meaningful factors were found. One of these factors included several questions related to the students' evaluation of interteaching (questions 1, 2, 8, and 9), while the other factor related to their evaluation of self-grading (questions 11 through 16). Instead of analyzing these questions separately, we averaged them to create two new variables corresponding to "Interteaching" and "Self-Grading". We found that on the "Interteaching" questions, students reported this teaching method to be beneficial and enjoyable ($M = 4.72$, $SD = .37$). Likewise, on the "Self-Grading" questions they reported that self-grading was helpful and enjoyable ($M = 4.06$, $SD = .76$). We also asked the students which combination of learning methods they preferred for this course. As shown in Figure 1, 52% preferred interteaching with self-grading, 24% preferred interteaching alone, 12% preferred traditional lectures, and 12% had no

preference. These results suggest that interteaching is a positive learning tool for the students and that self-grading is helpful and generally accepted by the students. It appears as though combining interteaching and self-grading can be a useful, beneficial experience for most students and is typically preferred to the traditional method of lecturing.

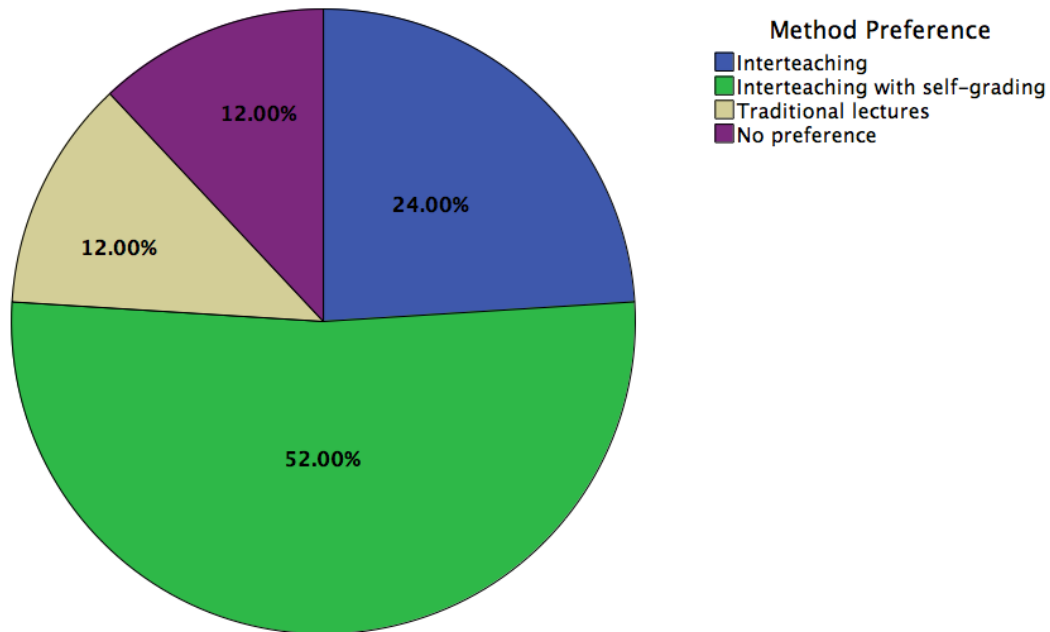


Figure 1. Frequency of students' choices for preferred "method of teaching" for this course

Discussion

The purpose of this study was to implement the self-assessment method in a course that uses interteaching in an attempt to examine three questions: whether students could accurately grade their own assignments, whether it would improve students' ability to assess the quality of their own work, and whether greater learning resulted in higher exam scores due to utilizing self-assessment. The design of this course relied on a student-centered teaching method and offered a

pedagogical technique for students to be exposed to the material multiple times, potentially resulting in deeper encoding and greater metacognition. In this study we were interested in the contributions self-assessment could make to this teaching method, while possibly reducing the amount of time spent by the instructor on grading.

Two semesters of testing out self-assessment yielded interesting data. In response to our inquiry regarding reliability and accuracy on the part of the students, the scores that the students gave themselves were slightly higher on individual assignments, which accordingly resulted in a slightly higher overall interteaching grade, and a higher overall course grade. This difference was consistent and significant, however, the actual scores were only 2% higher on individual assignments, 0.7% higher for the overall interteaching grade, and 0.28% higher on their overall course grade. Regarding the expected improvements in metacognition, the students did not show increased accuracy in predicting the scores of their own work over the course of the semester. Similarly, we found that the self-assessed material did not show better scores on exam questions. One semester even showed lower scores on the exam questions that related to the self-assessed material (80.8% for exam material based on self-assessed interteaching material and 84.3% for exam material based on professor-graded interteaching material).

The biggest question in this study was the accuracy with which students would grade their own work, and we found that students were mostly accurate and competent in their ability to assess their own work, with a slight tendency to overestimate. This was a small, though significant, difference. We suspect that a possible reason for this small, yet significant, difference was that the nature of the interteaching setting created a ceiling effect, resulting in a restriction of range. During interteaching meetings, the students are not tested on information that was already covered earlier in class. Instead, this was their first exposure to the material. They were able to

work on the assignment with their peers, using the notes that they made at home and brought to class. Because it was only their first exposure, the answer key that we provided as a rubric for self-assessing tends to be a little more lenient than for an exam. Also, the group work, as well as the opportunity to ask the professor for help and clarification, results in relatively high interteaching scores (an average of 9.43/10 and 9.69/10 for Spring and Fall, respectively). We believe that the significance that was found was due to the consistency with which students either overestimated or agreed with the professor, and only rarely underestimated their score. Specifically, in the Spring semester, students estimated higher than the professor 34.3% of the time, agreed with the professor 52.1% of the time, and underestimated only 13.6% of the time. In the Fall semester, students estimated higher than the professor 37.6% of the time, agreed with the professor 59.8% of the time, and underestimated only 2.56% of the time. Most likely, students tend to err on the optimistic side of self-scoring, considering that the assignment grades do ultimately affect their course grades and final GPAs. This combination of a ceiling effect, together with the students' tendency to grade themselves optimistically, resulted in a small, yet consistent, difference. It is important to keep the small effect size in mind, as well as, the high frequency with which students agreed with the instructor's grade, pointing at a high degree of accuracy in students' self-assessment.

These findings agree with prior research. For example, Goodrich (1930) found that his students were 98% accurate in marking their own papers. Meanwhile, other researchers have also found that students had a propensity to either slightly over-estimate their grades (Boud, et al., 1989; Burke, 1969; Logan, 2015; Strong, et al., 2003; Zoller, et al., 1988) or to assign grades that were very comparable to those of the professor (Davis, et al., 1980; Sadler et al., 2006; Sung, et al., 2005). In the current project the students were informed that the teacher was also going to

grade their work. It is possible that, in addition to the accountability built in to the interteaching groups, this system of double checking the grades (first graded by the group, then re-graded by the professor) assisted in producing more accurate results. As indicated by our results, students' inclination to over-estimate their grades appeared to decrease with shared responsibility and multiple points of accountability, which replicates past studies (Ross, et al., 1999; Zively, 1967).

Not only did we ask students to assess their own work based on a grading rubric, we also asked them to guess what scores they believed they would receive before assessing their work. We expected students to become more accurate in their score predictions over time, showing improved metacognition in estimating the quality of their own work. This did not happen, however, which may be because they were quite close in their predictions from the beginning of the semester. The students were given detailed answer keys and clear instructions how to grade. This may have resulted in accurate predictions and scores from the start, since they likely felt confident with the tools they were given, in addition to the trend of very close expectations and scores they were already experiencing. Likewise, the combination of the ceiling effect mentioned earlier and the students' tendency to overestimate their scores likely resulted in them being so accurate with little room for improvement.

It was a surprise that self-assessment did not result in better performance on corresponding exam questions. Self-assessment may have been only one small component in a course with many other learning-enhancing aspects that drowned out the effects of self-assessment. Interteaching has been shown to increase grades dramatically (Hughes, Sullivan, & Mosley, 1985; Ross, 2006; Sadler, et al., 2006; Schunk, 1996; Sparks, 1991), possibly as a result of the multiple exposures to the material this method provides. Students first read and take thorough notes, next discuss and collaboratively write their answers, and then either self-assess

the assignment or hand it in for the professor to grade. The following day in class the students attend a lecture about that same material. Finally, they study for the exam and attend a review session. Given the many opportunities to work with the material, it is not surprising that the small amount of time spent grading their interteaching answers did not single-handedly influence their exam scores. We consider the small, negative effect seen in the Fall semester a fluke and are currently collecting data to confirm that this is, in fact, the case.

One limitation of this particular study concerning the issue of improved exam scores is that the number of exam questions that directly related to the items on the interteaching preparation guides was limited. A large number of exam items came from material that was only covered in lecture. This made it more difficult to assess the effects of self-assessment on learning as measured by exam scores. In the future, having an equal number of exam questions from both lectures and interteaching would give the self-assessment method a fair chance to show its learning-enhancing abilities.

Conclusion

From its earliest roots, psychology has provided some of the bases for the most effective pedagogical methods (Skinner, 1953). Considering the state and rating of America's academic system (Saville, et al., 2011), it stands to reason that teachers, instructors, and professors alike would want to give students the most they can from their courses and overall schooling experiences. We conclude that, with some adjustments for limitations, combining two highly effective methods will allow students and instructors to reach their highest potentials in learning and teaching (Sadler, et al., 2006; Saville, et al., 2005, Saville, et al., 2011; Saville et al, 2012; Ross, 2006; Ross, et al., 2002; Querol, et al., 2015).

We found that students can, in fact, competently grade and assess their own work, with only a small, negligible over-estimation. It is important to realize that this small discrepancy is likely thanks to the double-checking system that was mentioned earlier, and from all the tools that allow for students' high confidence levels regarding their abilities. There are no obvious benefits for metacognition or learning as shown by exams scores, although by overcoming our previously mentioned limitations, this could potentially change. Students did enjoy the prompt feedback they received by doing the self-assessments in the interteaching course, and most even reported to prefer the self-assessment method over traditional professor-grading. Students also reported that they mostly prefer the combined style of self-assessment and interteaching together. In sum, this is a useful method that has been shown by past research (Sadler, et al., 2006; Ross, 2006; Ross, et al., 2002) to encourage learning and learning skills. After all, with the high level of accuracy that has been shown by the students, this method is likely to save professors time and energy. Students will also reap the many previously documented benefits we have discussed, along with the enjoyment they will experience from such an interactive combination of effective teaching styles.

References

- Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21-34.
- Andrade, H. L., Du, Y. & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(2), 3-13.
- Arntzen, E., & Hoium, K. (2010). On the effectiveness of interteaching. *The Behavior Analyst Today, 11*(3), 155-160.
- Arter, J., Spandel, V., Culham, R., & Pollard, J. (1994, April). *The impact of training students to be self-assessors of writing*. Paper presented at the annual meeting of American Educational Research Association, New Orleans.
- Boud, D. (1988). Moving towards autonomy, in Boud, D.J. (Ed.) *Developing Student Autonomy in Learning*, Second edition. London: Kogan Page, 17-39.
- Boud, D. (1995). Assessment and learning: contradictory or complementary? *Assessment for Learning in Higher Education*, London: Kogan Page, 35-48.
- Boud, D. (1991). *Implementing Student Self-Assessment*. HERDSA Green Guide. Second Edition. Sydney: HERDSA.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529-549.
- Burke, R. J. (1969). Some preliminary data on the use of self-evaluations and peer ratings in assigning university course grades. *The Journal of Educational Research, 62*(10), 444-448.

- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31.
- Dart, B. C., & Clarke, J. A. (1991). Helping students become better learners: A case study in teacher education. *Higher Education*, 22(3), 317-335.
- Davis, J. K., & Rand, D. C. (1980). Self-grading versus instructor grading. *The Journal of Educational Research*, 73(4), 207-211.
- Edwards, N. M. (2007). Student self-grading in social statistics. *College Teaching*, 55(2), 72-76.
- El-Koumy, A. (2010). *Student self-assessment in higher education: Alone or plus*. Paper presented at the CPLA Conference Lebanese American University, Lebanon.
- Fox, S., & Dinur, Y. (1998). Validity of self-assessment: A field evaluation. *Personnel Psychology*, 41, 581-592.
- Goodrich, T. V. (1930). Can pupils mark their own papers accurately? *The Journal of Educational Research*, 21(4), 255-261.
- Hillocks, G. (1986). *Research on written composition; New directions for teaching*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.
- Hotard D. J. (2010). *The effects of self-assessments on student learning of mathematics* (Unpublished master's thesis). Louisiana State University, Louisiana.
- Hughes, B., Sullivan, H., & Mosley, M. (1985). External evaluation, task difficulty, and continuing motivation. *Journal of Educational Research*, 78(4), 210-215.
- Johnson, C. S., Gelfand, S. (2013, July). Self-assessment and writing quality. *Academic Research International*, 4.4, 571-580.

- Kahraman, N. (2014, July). Investigating the relationship between self-assessment and self-efficacy of pre-service science teachers. *International Journal of Education and Research*, 2(7), 77-90.
- Logan, B. (2015). Reviewing the value of self-assessments: Do they matter in the classroom? *Research in Higher Education Journal*, 29, 11.
- Lopez-Pastor, V. M., Fernandez-Balboa, J. M., Santos-Pastor, M. L., & Aranda, A. F. (2012). Students' self-grading, professor's grading and negotiated final grading at three university programmes: Analysis of reliability and grade difference ranges and tendencies. *Assessment and Evaluation in Higher Education*, 37(4), 453-464.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 075-100.
- Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, 41(3), 273-290.
- Querol, B. I., Rosales, R., & Soldner, J. L. (2015). A comprehensive review of interteaching and its impact on student learning and satisfaction. *Scholarship of Teaching and Learning in Psychology*, 1(4), 390-411.
- Rolheiser, C., & Ross, J. A. (2001). *Student self-evaluation: What research says and what practice shows*. Retrieved from http://www.cdl.org/resource-library/articles/self_eval.php.
- Ross, J. A. (2006). The reliability, validity and utility of self-assessment. *Practical Assessment, Research and Evaluation*, 11(10), 1-13.

- Ross, J. A., Hogaboam-Gray, A. & Rolheiser, C. (2002a). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment*, 8(1), 43-58.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002b). Self-evaluation in grade 11 mathematics: Effects on achievement and student beliefs about ability. In D. McDougall (Ed.), *OISE Papers on Mathematical Education*. Toronto: University of Toronto, Canada.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray., A. (1999). Effect of self-evaluation on narrative writing. *Assessing Writing*, 6(1), 107-132.
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.
- Saville, B. K., Lambert, T., & Robertson, S. (2011). Interteaching: Bringing behavioral education into the 21st century. *The Psychological Record*, 61, 153-166.
- Saville, B. K., Pope, D., Truelove, J., & Williams, J. (2012). The relation between GPA and exam performance during interteaching and lecture. *The Behavior Analyst Today*, 13(3&4), 27-31.
- Saville, B. K., Zinn, T. E., & Elliot, M. P. (2005). Interteaching versus traditional methods of instruction: A preliminary analysis. *Teaching of Psychology*, 32(3), 161-163.
- Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, 33(2), 359-382.
- Skinner, B. F. (1953). *Science and human behavior*. New York: The Free Press.
- Skinner, B. F. (1999). The science of learning and the art of teaching. In *Cumulative record, definitive edition* (pp. 179-191). Cambridge, MA: B.F. Skinner Foundation. (Original work published 1954).

- Sparks, G. E. (1991). *The effect of self-evaluation on musical achievement, attentiveness and attitudes of elementary school instrument students*. Unpublished doctoral dissertation, Louisiana State University and Agriculture and Mechanical College.
- Strong, B., Davis, M., & Hawks, V. (2003). Self-grading in large general education classes. *College Teaching*, 52(2), 52-57.
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self-and peer-assessment system. *Computers and Education*, 45(2), 187-202.
- Thomas, G., Martin, D., & Pleasants, K. (2011). Using self- and peer-assessment to enhance students' future learning in higher education. *Journal of University Teaching & Learning Practice*, 8(1), 1-17.
- Walser, T. M. (2009). An action research study of student self-assessment in higher education. *Innovative Higher Education*, 34, 299-306. doi:10.1007/s10755-009-9116-1
- Zively, S. (1967). A cautious approach to student grading. *The English Journal*, 56(9), 1321-1322.
- Zoller, U., & Ben-Chaim, D. (1988). Student self-assessment in HOCS science examinations: Is there a problem? *Journal of Science Education and Technology*, 7(2), 135-147.

Appendix A

Mid-Semester Evaluations

PSYC 321

Instructor: Martin van den Berg

1. The preparation guides help me to come to class prepared
a. Not at all b. c. d. e. Yes, very
2. Interteaching encourages me to keep up with the material of the course
a. No, disagree b. c. d. e. Yes, agree
3. Discussing the prep guides with fellow students helps me to learn the material
a. No, disagree b. c. d. e. Yes, agree
4. I believe that, with minimal guidance, I can learn the material in the book by myself
a. No, disagree b. c. d. e. Yes, agree
5. Comparing my own preparation habits to that of my discussion partners is useful
a. Not at all b. c. d. e. Very
6. The fear of being criticized by my discussion partner encourages me to prepare well
a. Not at all b. c. d. e. Very
7. I feel that I can ask the instructor during class about material that is unclear on the prep guides
a. Never b. c. d. e. Always
8. Getting feedback on my understanding of each prep guide allows me to gauge my progress with the material and in the course
a. Not at all b. c. d. e. Very much
9. The prep guides are a good indication of the material I need to know for the exam
a. Not at all b. c. d. e. Highly
10. The in-class discussion allows me to better set my own pace than a lecture would
a. No, disagree b. c. d. e. Yes, agree
11. Grading our interteaching answers ourselves is a valuable tool in learning the class material.
a. Not at all b. c. d. e. Very much
12. When I grade my own answers with my group we are able to give ourselves an honest and accurate score.
a. No, disagree b. c. d. e. Yes, agree

13. I retained more information from the material using the self-grading method versus the assignments that were graded by the instructor.

- a. No, disagree b. c. d. e. Yes, agree

14. The immediate feedback during the self-grading sessions was more helpful than having to wait for the feedback from the professor's grading.

- a. No, disagree b. c. d. e. Yes, agree

15. I enjoyed using the self-grading method on our assignments.

- a. Not at all b. c. d. e. Very much

16. The self-grading should be used:

- a. Less often b. c. d. e. More often

30. Imagine that in a next life you would take this class again. Which format would you prefer?

- A. I would prefer interteaching
- B. I would prefer interteaching with the self-grading method
- C. I would prefer traditional lectures
- D. I have no preference

17. What should be changed about this course?

18. What should stay the same about this course?

19. What is your opinion about grading your own assignments?

20. Please write down any additional comments you may have.

Appendix B

Prep Guide 11 – Answer Key

1. The Cartesian point of view comes from René Descartes, a 17th century French philosopher who was interested in the mind. He made a distinction between the physical body and the non-physical mind or “spirit”. This non-physical mind was considered to be outside the observable physical world.

The neuroscience point of view, on the other hand, sees mind or consciousness as a direct result of physical brain processes that can be studied like any other physical phenomenon. The neuroscience point of view is therefore most suitable to scientific study as it allows for observation and measurement, whereas the Cartesian view does not.

How to score:

Correct description of Cartesian view	½ point
Correct description of neuroscience view	½ point
Statement about which one is most suitable for scientific study	½ point

2. Some reasons why the prefrontal cortex may be involved in consciousness all center on the fact that this part of the brain is responsible for highly sophisticated functions, such as:
- The motor area, including Broca’s area for language can be found there
 - Lots of activity in the prefrontal cortex during heavy cognitive work
 - It processes and interprets the input from the limbic system
 - The frontal lobe developed most recently in evolution
 - The frontal lobe is the largest part of the human cortex
- To Carter, all of these suggest that the frontal lobe performs very complex functions, so possibly also consciousness.

How to score:

List at least two reasons	½ point each
---------------------------	--------------

3. People with ADD have a lot of activity in the back of the brain, and less in the front. These people are unable to control their own attention but are controlled more by outside stimuli. People with schizophrenic catatonia also have most activity in the back of the brain and little in front. These people can be completely inactive for hours on end, sitting or lying completely still. They are unable to initiate new behavior. Both of these suggest that you need activity in the front of the brain to control your thoughts and your behavior.

How to score:

Description of ADD brain activity pattern	½ point
Description of schizophrenia brain activity pattern	½ point
Explanation why both of these suggest something about free will	½ point

4. Four processes, their brain regions, and their role in consciousness
- a. Self-awareness

- i. Takes place in the frontal cortex
 - ii. The brain not only processes information, but also monitors itself while doing so. This self-monitoring or self-control makes it possible to tell the difference between external information and self-generated information.
- b. Working memory
 - i. Takes place in the dorsolateral prefrontal cortex
 - ii. This area allows for monitoring one's own performance, learning from mistakes, and choosing the best strategy for a specific task
- c. Orbito-frontal cortex
 - i. This part is responsible for planning and sticking with a plan
 - ii. It inhibits other urges that may come up while performing a task. For example, it will inhibit the urge to check Facebook while studying for an exam. It allows us to stay on track
- d. (Emotional) meaning
 - i. Takes place in the ventromedial cortex
 - ii. This area adds emotional significance to our ideas. This is especially important for decision-making. To make a decision we often need to feel more strongly about one option than another. This area gives us that strong feeling.

At this point it should be clear that Carter has her very own definition of consciousness. All of these regions and functions are involved in “free will”, the ability to control our own thoughts, actions, and decisions. Free will is an important and interesting topic in its own right, but this is NOT the usual definition of consciousness, which Rama will give us in Chapter 12.

How to score:

Correct description of each of the four functions ½ point each

5. Mary Knight had a strong belief that she was pregnant, but she was not really pregnant. However, this belief caused her body to produce all the symptoms of pregnancy. She had a large belly, swollen breasts, mottled nipples, but there was no fetal heartbeat and her belly button was an “innie” instead of an “outie”. The fact that a mental belief can influence the physical body clearly shows the strong relationship between body and mind.

How to score:

Brief description of Mary Knight's symptoms ½ point
 Explanation of how it shows the influence of mind on body ½ point

6. First of all, Rama thinks we should take “scientific anomalies” seriously. There are many examples of phenomena that science cannot explain. For example:
- a. Hypnosis can cure warts. But how can the suggestions of a hypnotist make a wart (usually caused by a viral infection) go away?
 - b. Some people have “learned allergies”. These people get allergic reactions when they see something they think should induce an attack. For example, a plastic rose (that does not

have the physical molecules that would normally create a response) can still trigger an asthma attack

- c. Food aversion is another example. After getting sick from a certain kind of food once, seeing that food again or only thinking about it (without actually taking it into the body) can lead to the same illness and even infections as the time that food was really ingested. Rama finds it a shame that these interesting phenomena are done away as simple “anomalies” and even “new age nonsense”. He claims that, when there are enough of these anomalies, we should conclude that our current scientific paradigm does not work anymore and that instead we should make a radical change in how we do science. The scientific paradigm includes all the currently accepted research methods, research questions, and theories. When these don’t provide the desired results anymore a complete change may be needed and such a big change is called a “paradigm shift”. One example of such a big change was the Copernican revolution, when scientists agreed that the Sun and not the Earth was the center of our solar system. This completely changed everybody’s worldview and can therefore deservedly be called a paradigm shift. Rama suggests that we should continue to use our Western scientific methods, but that we should use these to investigate Eastern questions that directly look into the relationship between body and mind.

How to score:

Some statement about the need for a paradigm shift	½ point
Some indication of the use of Western methods to study Eastern questions	½ point

Appendix C

PSYC 321 – Brain, Mind, and Behavior

PREPARATION GUIDE 11Interteaching for: **Tuesday, December 1**Based on: **Carter, Chapter 8, (pg. 180-202) & Rama, Chapter 11, (pg. 212-226)**

Instructions: Read the questions below. Then read the assigned material while taking notes that will allow you to answer these questions in class. To make your interteach during class most productive it is best to:

- Take notes but do not answer the questions yet. Instead, make sure that your notes will allow you to discuss the best possible answer with your partner and then write up your answer in class.
- As always, use your own words. Copying sentences from the book is not a useful way for learning this material.
- Prepare appropriately. Three well-prepared students can really help each other in understanding and mastering this material.

1. The last chapter in Carter's book is about consciousness. Unfortunately, not only is it difficult to study consciousness, this difficulty is greatly increased by a disagreement about what exactly consciousness is. Carter gives two schools of thought, the Cartesian and the neuroscience points of view. How does each of these views regard consciousness, and which one is most suitable for scientific study?

2. Carter suggests that the frontal lobe is the seat of consciousness, in particular the prefrontal cortex. Early in the chapter she gives some evidence for this. What evidence is there to back up this suggestion?

3. One possible argument for the role of the frontal lobes in consciousness comes from the fact that the frontal lobes seem to play a role in taking initiative, voluntary control, or even the philosophically difficult term of "free will". The pattern of brain activity seen in people with ADD and schizophrenic catatonia support this idea. For both people with ADD and people with schizophrenic catatonia, describe what type of brain activity they have, and how this supports the idea that the frontal lobe is responsible for free will.

4. Carter tries to get a handle on consciousness by examining the processes that are needed for conscious thought. For the following terms, briefly describe their role in conscious thought: a. Self-awareness b. Dorsolateral prefrontal cortex / working memory c. Orbito-frontal cortex d. Meaning / ventromedial cortex

5. Rama discusses the mind-body relationship in Chapter 11. He uses the case of Mary Knight to illustrate the complex relationship between the mind and the body. How does this case show that the mind can influence the body?

6. From Carter's chapter you may have gotten the idea that a biological explanation of consciousness is still very far away. Such an explanation would have to make clear how the body and brain could create a conscious feeling of awareness of oneself and the environment. In other words, it should explain the mind-body relationship. According to Rama, what is probably needed (in terms of research directions) if we ever hope to understand that relationship? PSYC

Please bring this to class and complete AFTER your participation in interteaching.

Interteach Report

Date: _____

Participants: _____

Assigned Readings _____

Duration of Interteach: _____ Sufficient Time Provided Yes No

Quality of Session (on a scale of 1 = poor to 10 = excellent): _____

Reason for giving this particular rating: _____

What score (out of 10) do you think you got on this assignment: _____

Topics that gave difficulty, and the nature of the difficulty (e.g., Q4 / Text was confusing, Q6 / Question was ambiguous):

Topics you would like clarified in class:

What parts did you find most interesting and would you like to hear more about?

If a question were to be omitted, which should it be?

Other comments or suggestions? Please give feedback.