8-2013

Analysis of the Role of Homework in Predicting and Improving Exam Performance

Charles E. Galyon
shouga@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Charles E. Galyon entitled "Analysis of the Role of Homework in Predicting and Improving Exam Performance." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in School Psychology.

Robert L. Williams, Major Professor

We have read this dissertation and recommend its acceptance:

Chris H. Skinner, Brian E. Wilhoit, David F. Cihak

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Analysis of the Role of Homework in Predicting and Improving Exam Performance

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Charles E. Galyon

August 2013

Acknowledgements

I would like to thank Dr. Robert Williams for his insightful and patient guidance, which played a critical role in the completion of this dissertation and my career as a graduate student. I appreciate the high standards to which he held me, which contributed to my growth as a professional and an individual. He always pushed me to strive for my best and never settled for anything less. His genuine interest in seeing his students succeed will always be an inspiration to me.

I would also like to thank Dr. Thomas Brinthaupt, my advisor and mentor while I was in the MTSU psychology graduate program. Through him I learned what I consider my most valuable lesson as a researcher, that one can achieve more through collaboration in research than individually. I also appreciated his ongoing support while I completed my doctorate at UTK. He continues to serve as a role model of what I wish to achieve as an instructor and a researcher.

I also wish to thank Drs. Chris Skinner, David Cihak, and Brian Wilhoit for their support, suggestions, and constructive criticism as I completed my dissertation. Their input with respect to methodology, analytical techniques, and future research was valuable and contributed substantially to the quality and completion of this dissertation. They are all extraordinary researchers and I consider myself privileged to have had the opportunity to work with them.

I would also like to thank my friends and colleagues Carolyn Blondin, John Parkhurst, Dr. Lisa McCleary, Dr. James Sternberg, and Cora Taylor for their support and input as I grappled with the complex and challenging process of completing my dissertation. They were always available to provide encouragement and helpful

suggestions when I needed them. They are exemplary colleagues and it has been my fortune to know them. Additionally, I would like to thank Kyle Voils, Bethany Forbes, and Taotao Long for their help in developing and validating the methodology used to evaluate homework in my dissertation.

And finally, I would like to thank my wife, Anna Galyon. Her love and support were unwavering throughout this process. I greatly appreciate her patience during a time when I was arguably less mentally and emotionally available. I could not ask for a better life companion.

Abstract

Homework is one of many factors thought to improve students' academic performance, given that homework provides a means for students not only to master course content, but also to develop valuable study habits, improve their time management, and learn to work independently. Unfortunately, college students commit considerably less time to homework than is conventionally thought necessary, and their answers to homework questions frequently indicate an erroneous and/or incomplete understanding of the course material. The current study examined relationships between potential predictors of and trends in exam performance in a large undergraduate educational psychology course. The relationship between homework completion, homework accuracy, and exam performance was examined, as well as a potential methodology to improve the accuracy and thoroughness of students' homework.

The first study evaluated data collected over the past seven years to identify patterns of exam performance, critical thinking, and GPA among students across years in school (N = 3,591). The results showed a distinctive pattern of exam performance across units in the course and significant differences in critical thinking and exam performance between students at the beginning and end of their undergraduate careers. The results also supported a relationship between critical thinking, GPA, and exam performance. The second study (N = 167) evaluated the relationships between critical thinking, participation in class discussion, the accuracy of student homework responses, and exam performance, and examined a methodology for evaluating student homework responses. The results indicated a significant relationship between homework accuracy and exam performance, in some cases proving to be a stronger relationship than between critical thinking and

exam performance. The results of the third study (*N* = 71) showed that course credit contingent on the accuracy of students' homework answers increased both accuracy and thoroughness of homework. Improved accuracy of homework contributed to improvement in exam scores overall, and broke a historical pattern of decreasing exam scores in the most difficult units in the course. Although other factors, such as a critical thinking, GPA, and year in school, also significantly predicted exam performance, they did not interact with the homework contingencies in changing scores on homework or exams.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction and Literature Review

Over the past decade a considerable body of research has been conducted within the context of an undergraduate educational psychology course (Educational Psychology 210) to identify factors predicting or affecting student performance. This dissertation is a collection of three studies examining and furthering this line of research. The first study analyzed data from previous research to identify factors that have been shown to influence student performance and explored patterns of student performance over a period of seven years (2004 to 2011). The second study evaluated student homework accuracy and its possible relationship to exam performance during the spring semester in 2010. The third study applied experimental control to improve student homework accuracy and examined the corresponding changes in exam performance during the fall semester of 2011.

Multiple-choice Exams as a Measure of Student Success

Multiple-choice exams are commonly used to assess student performance and content-mastery in large entry-level undergraduate courses (Holtzman, 2008; Walker & Thompson, 2001). Instructors can use these exams to accurately predict competency as measured by professional success or summative coursework evaluations in a subject area (Stepp, Schrock, & Coscarelli, 1996). In addition to being appropriate and valid measures of student performance, multiple-choice exams also frequently demonstrate better reliability than open-ended exams due to both a larger number of questions and the fact that even when students guess, their guesses may reflect some knowledge of course

content (Bridgeman, 1992; Hassmen & Hunt, 1994). In some instances, multiple-choice exams may actually measure the same knowledge as open-ended exams (Bridgeman). However, as this effect is not consistently observed, the structure and content of the exams is more likely responsible for overlapping measures of knowledge (Becker & Johnston, 1999; Harris & Kerby, 1997). Lastly, multiple-choice exams also can be constructed in a way that develops superior strategies for learning information and improves student understanding and mastery of that information (Parmenter, 2009).

Another advantage to multiple-choice exams is the apparent objectivity of the exam. While teachers and students may not always agree on the clarity of wording in an item, multiple-choice exams are generally less likely to produce controversy regarding the adequacy of an exam answer than are open-ended exams. Furthermore, multiple-choice exams have demonstrated the capacity to highlight and precisely identify gaps in students' knowledge and reasoning (Wallace & Williams, 2003). A number of skills are involved with correctly selecting an answer to an item on a multiple-choice exam, including recalling information related to an item's alternatives, identifying incorrect information included in those alternatives, and selecting the most supportable and factually correct answer. Due to this collection of skills, a student's answer to a multiple-choice question may assess a student's knowledge and reasoning more precisely than a comparable open-ended essay question.

While there are certain pedagogical benefits to the use of multiple-choice exams as discussed above, another important and likely common reason for using this type of exam in large courses is efficiency in grading. Developing multiple-choice exams that yield all of the benefits above can be very labor intensive, but scoring them can be

accomplished quickly and automatically with the use of test scanners. Ideally each student's answer sheet is scored immediately after the student completes the exam and is returned to the student with a tabulation of the total number of correct answers and clear identification of the correct answers for incorrect responses (Hautau et al., 2006; Krohn et al., 2008). In addition to providing clear and immediate feedback, many introductory students prefer multiple-choice exams over essay exams. It may be much easier for them to evaluate alternative choices on a multiple-choice exam item than to construct correct answers on essay exams (Struyven, Dochy, & Janssens, 2005).

The ability to predict exam performance may provide a framework for developing interventions to improve exam performance. A number of variables have proven predictive of student exam performance, including critical thinking, generic vocabulary, and pre-course knowledge (Turner &Williams, 2007; Williams & Worth, 2002). Some activities have also shown potential for improving student exam performance, including out-of-class exercises such as taking practice exams similar to the actual exams (Holtzman, 2008; Oliver & Williams, 2005) and completing short writing assignments related to exam items (Hautau et al., 2006; Krohn et al., 2008). While identifying predictors of exam performance does not directly address student difficulties, it appears to be an important first step in developing effective interventions or strategies for improving student success.

Student Predictors of Multiple-choice Exam Performance

Doubtless a source of frustration for instructors is that some students will continue to struggle in a class despite a variety of instructional strategies and proposed study techniques. Similarly, other students appear to succeed without any additional help. This

leads one to reason that students enter a course with certain cognitive and behavioral characteristics that help or hinder their success. Thus, efforts to aid failing students must focus on developing a more complete understanding of which characteristics are likely to lead to success and how to remediate the ones that do not.

Critical thinking. Critical thinking should be an excellent predictor of performance on multiple-choice exams that have been designed to assess critical analysis of the subject matter. Critical thinking in the context of academics is a measure of a student's ability to analyze presented information and derive the most defensible conclusion based on the evidence available. It requires students to discern between supportable conclusions and those that lack evidence, and to identify potential biases in information sources. Having strong critical thinking ability should enable a student to determine which information in alternative answers for an exam item is the most accurate and supported.

Previous research has demonstrated that critical thinking is a consistent and moderate predictor of exam performance, with correlations ranging from .30 to .40 between critical thinking and exam performance (Wallace & Williams, 2003; Williams, Oliver, & Stockdale, 2004; Williams & Worth, 2002). Williams and Worth, however, have shown that students' work habits can make a greater difference in exam performance than critical thinking skills alone. Some students with low critical thinking skills have performed relatively well (Williams & Stockdale, 2003). Such students have generally demonstrated superior note-taking skills and made significant gains in critical thinking skills during the course. However, even these diligent students tended to perform more

poorly than students with high critical thinking skills. Thus, critical thinking is likely to exert some influence on student exam performance regardless of academic interventions.

Grade point average (GPA). Grade point average (GPA) is among the most used predictors of performance in a wide array of courses and in academic programs as a whole (Daniels et al., 2009). Although GPA is widely accepted as an indicator of general academic performance, the behaviors contributing to GPA need more specific delineation. A student's GPA is likely a result of both good study habits and good thinking skills.

Likely contributors to GPA include academically-oriented behaviors reflecting high interest in schoolwork, effort to earn high grades, and active attempts to independently master subject matter (Corno & Mandinach, 1983; Pressley, Borkowski, & Schneider, 1987; Sivan, 1986; Zimmerman & Schunk, 1989). More specifically, students with high GPA are noted for their consistent class attendance, participation in class discussion, timely completion of assignments, and contact with the instructor outside of class (Bender, 1997). High GPA is further characterized by good time management, concentration on targeted subject matter, and effective test-taking strategies (Everson, Weinstein, & Laitusis, 2000).

Academic outcomes, as measured by GPA, are significantly related to autonomous learning strategies and complex thinking (Pizzolato, Hicklen, Brown, & Chaudhari, 2009). Additionally, GPA has been related to students' personal development, impulse control, and other positive characteristics (e.g., organization, responsibility, and initiative) (Demoulin & Walsh, 2002; Kirby, Winston, & Santiesteban, 2005; Zimmerman & Schunk, 1989). GPA has also been associated with

the strategic learning abilities measured on the *Learning and Study Strategies Inventory* (LASSI), including time management, concentration, test-taking strategies, selection of main ideas, self-assessment, and study aids (Everson, Weinstein, & Laitusis, 2000).

Participation in class discussion. Because participation in class discussion is highly valued by most college instructors (Weaver & Qi, 2005), one would expect it to also be significantly linked to exam performance. Ideally, when students ask and answer course-relevant questions in class, they are improving their own understanding of the material. Foster, McCleary, and Williams (2010) found a significant relationship between class participation and exam performance. Students who exhibited a pattern of low participation tended to make poorer quality comments than those who participated more frequently. Furthermore, the low-participants tended to perform worse on course exams than high-participants. Conceivably, students may have been reluctant to participate due to a lack of content mastery (directly tied to exam performance). Overall, when class discussions emphasize issues to be addressed on course exams, participation in class discussion is more predictive of exam performance (Harton, Richardson, Barreras, Rockloff, & Latane, 2002).

Completion of homework assignments. The purpose of homework assignments is usually to help students practice and master course concepts, perhaps in ways they would not independently use (Epstein & Van Voorhis, 2001; Gajria & Salend, 1995). Homework, in turn, should improve performance on course exams. However, homework can also serve other purposes and help develop skills beyond mastery of course content. Homework assignments can help students learn to work independently (Bursuck et al., 1999), develop time-management skills (Cooper, 2001), and ultimately build good study

habits (Salend & Gajria, 1995). Although homework appears more effective for secondary than younger students, a relatively limited amount of evidence is available to support the effectiveness of homework in promoting the academic achievement of college students (Cooper, Robinson, & Patall, 2006).

Research on the efficacy of homework assignments has primarily been conducted with students having learning disabilities. Students receiving special-education services often experience more difficulty completing homework assignments than other students. Some possible reasons for this difficulty are poor organizational skills, inattention, a tendency to procrastinate, and academic skills deficits (Bryan, Burstein, & Bryna, 2001; Polloway, Epstein, & Foley, 1992). These factors can impact academic performance and the benefits of homework in ways that may not affect the general-education population. It is therefore necessary to evaluate homework with a broader population in order to acquire a better understanding of its efficacy with general-education students.

Despite the potential benefits of homework for improving academic performance, relatively little research has been conducted with students at the college level. One potentially discouraging finding regarding homework with college students is that students often spend much less time on homework than conventionally thought appropriate (e.g., two hours out-of-class for every hour in class) (Young, 2002). Indeed, some surveys have shown that 63% of college freshmen devote less than 15 hours per week to class preparation, with almost 20% focusing on homework for only 1 to 5 hours per week. This pattern remains fairly consistent from undergraduate college entry to graduation.

While the amount of time students devote to homework is certainly important, the design of the homework assignment is similarly critical to helping students get the most out of homework. Assignments that demand more active involvement from students, such as answering questions over material the student has just read, should help students develop greater mastery of material (Hautau et al., 2006; Turner et al., 2006). On the other hand, passive assignments, such as reading assigned content, are less likely to aid the student with later recall. Furthermore, assignments that specifically deal with concepts addressed on the exams should help students perform better on those exams.

Chapter II

Synthesis of Previous Research in Educational Psychology 210

The following collection of studies conducted over the past decade have identified and evaluated factors hypothesized to influence student performance on course exams. A variety of topics have been addressed within the context of these studies, including critical thinking, pre-course vocabulary, pre-course content knowledge, daily writing assignments, participation in class discussion, participation in cooperative learning groups, note-taking behaviors, and academic self-efficacy. The results of these studies have provided considerable insight regarding factors influencing student performance and strategies instructors can use to promote student success. Following is a summary of the results of these previous research endeavors.

Critical thinking. Critical thinking has been a common focus in the research conducted in the course targeted in the current studies. Initially identified as a potential explanation for student exam performance in 2002 (Williams & Worth, 2002), critical thinking was shown to explain a significant amount of variance in student exam scores. However, this first study evaluated academic behaviors (notetaking and attendance) in combination with critical thinking scores and found that while critical thinking was significantly correlated with performance, notetaking was a stronger predictor of exam performance.

A follow-up to the Williams and Worth (2002) study was a study conducted by Williams and Stockdale (2003), which attempted to determine how some students who perform poorly on the critical thinking test still perform well in the course, and what behaviors separate them from students with low critical thinking who perform poorly in

the course. This study was conducted across two semesters and employed two measures of critical thinking: the *California Critical Thinking Skills Test* (CCTST, Facione & Facione, 1994) and the *Watson-Glaser Critical Thinking Appraisal Form-S*, or WGCTA-S (1994). The results indicated that students who performed poorly on a critical thinking test, yet performed relatively well on course exams, exhibited superior notetaking skills and higher rates of attendance than the students who performed poorly on course exams. Nevertheless, students who performed well on the critical thinking test tended to perform well on the course exams regardless of their academic behaviors and also exhibited less variation in course exam performance in general.

The next study in this line of research attempted to determine if domain-specific critical thinking (e.g., psychological critical thinking) might predict course performance and could be significantly improved through course instruction (Williams, Oliver, Allin, Winn, & Booher, 2003). This study was the first in this line of research that attempted to actually improve students' critical thinking abilities and the first study to examine domain-specific critical thinking. Domain-specific critical thinking was evaluated by a test developed by Lawson (1999) and was assessed at the beginning and end of the course. This pre-test, post-test design enabled evaluation of improvement that might result from the inclusion of critical thinking tasks in practice exams in the course.

The results indicated the measure of domain-specific critical thinking was a significant predictor of course success at the beginning and end of the course (a better predictor at the end of the course), and perhaps more importantly, that domain-specific critical thinking could be improved within the span of a single course (a significant increase in mean scores from 16.45 to 18.81). Another important finding from this study

was that the low-performing students did not exhibit a significant increase in domain-specific critical thinking. Conceivably, low-performing students engage less in course learning activities and thus would be expected to benefit less from the course's potential to improve domain-specific critical thinking. This study, while illuminating, was not sufficient to establish a precise cause-effect relationship, but provides clear evidence of a possible relationship between promotion of domain-specific critical thinking and improvement in course performance.

A direct follow-up to the Williams et al. study (2003) was a study by Williams, Oliver and Stockdale (2004), which compared domain-specific critical thinking with generic critical thinking as both predictors of exam performance and outcome variables. Students were again administered the domain-specific critical thinking test developed by Lawson (1999), as well as the WGCTA-S. Both critical thinking tests were given at the beginning and end of the semester to evaluate potential changes in critical thinking scores. The results indicated that domain-specific critical thinking was both a better predictor of course performance and was also more amenable to change over the duration of the semester than the WGCTA-S (a generic measure of critical thinking). In fact, generic critical thinking did not significantly improve. Low performing students again demonstrated negligible improvement in their domain-specific critical thinking scores as well as in their generic critical thinking. As noted by the authors, this study highlighted the difficulty in improving the critical thinking ability (domain-specific or generic) of students who perform poorly in a course. It also showed that domain-specific critical thinking is a better predictor of performance within its corresponding subject area than generic critical thinking.

Critical thinking appears well established as a predictor of exam performance. It consistently and moderately correlates with exam performance, and domain-specific critical thinking is amenable to change over the duration of a course. This improvement, however, appears limited to students who perform well in the course, which raises the question of what factors might be contributing to poor performance and limited improvement in critical thinking. Conceivably, students engagement in learning activities offered in a course should promote better performance in the course. However, it is also possible that other characteristics, behavioral or cognitive, prevent low-performing students from benefiting from the learning activities and also from performing well in the course. Additional studies have attempted to address these possibilities.

Cooperative learning groups. Cooperative learning activities present an opportunity for students to work together collaboratively in a group, ideally with the goal of improving the performance of all the group's members. Stockdale and Williams (2004) evaluated the effectiveness of cooperative learning activities in an undergraduate educational psychology course by comparing student performance during a cooperative learning activity to performance before and after involvement in the cooperative learning activity. Students were assigned to groups based on their exam scores in the preceding unit and placed in heterogeneous groups (e.g., high-, medium-, and low-performing students mixed) such that each group had the same average exam score from the preceding unit. A reward contingency was put in place, which granted students bonus points if the group's average exam score increased by a certain amount over the preceding unit. By comparing exam scores in the cooperative learning unit with the

preceding unit, Stockdale and Williams found that the cooperative learning unit produced significantly higher exam scores than the preceding unit. Further analysis revealed that this effect occurred more for low-achieving students than high-achieving students, whose scores actually decreased in some cases. Students' performance was also significantly higher during the unit with the cooperative learning activity than in the following unit, which did not include cooperative learning groups.

A follow-up study by Williams, Carroll, and Hautau (2005) examined the effects of individual vs. group accountability on the effectiveness of cooperative learning groups. Over the course of three semesters, different reward contingencies were put in place. These contingencies included a "group-only" contingency, in which students received bonus points for improvement in the group's average exam score; an "integrated group/individual" contingency, in which students received bonus points if they improved both their own and the group's average exam score; and a "differentiated group plus individual" contingency, in which students received bonus points for meeting each of the two requirements. Students at different performance levels (e.g., low-achieving and high-achieving students) were also evaluated separately to determine if the contingencies affected them differently.

The results suggested that, overall, the three contingencies similarly affected exam performance. However, more detailed analysis revealed that while the low- and middle-achieving students showed comparable degrees of improvement in each of the three contingencies, the high-achieving students actually decreased in performance under the group-only contingency. This study highlighted the importance of evaluating the

differential effects of interventions on students based on their performance (exam scores or prior GPA).

One final study, conducted by Carroll, Williams, and Hautau (2006) attempted to further adjust the reward contingencies such that they were related or unrelated to each other. In one semester students could receive individual bonus credit without receiving the group credit (and vice-versa) for improving exam performance; in the other semester students had to qualify for individual credit before they could receive group credit. The results indicated that making the bonus contingencies dependent on each other (related) was generally more effective than providing them independently. Being able to earn individual credit without regard for the group may have resulted in the student focusing more on his/her own performance.

Academic behaviors. A variety of additional academic behaviors or characteristics were implicated in the preceding studies. Efforts to tease out the importance of academic behaviors, such as notetaking, attendance, practice exams, self-perception, pre-course knowledge, generic vocabulary, and reading comprehension, were generally included as part of another study and were evaluated in comparison with other factors, such as critical thinking.

Notetaking. The study conducted by Williams and Worth (2002) was the first conducted in this educational psychology course to evaluate a set of academic behaviors as possible predictors of student exam performance. In addition to critical thinking, attendance (as evaluated by a daily sign-in roster) and notetaking were examined for their possible roles in student performance. The measure of notetaking involved in this early study shares some characteristics with the homework evaluations conducted in Studies 2

and 3 of the current dissertation. Specifically, student notetaking was to be performed within a Study Guide provided to students. The completed Study Guide was collected prior to each exam and evaluated along three dimensions: completeness (the number of questions answered), length (the amount of space used), and accuracy (evaluation of three randomly selected questions). This study was conducted over a period of two semesters.

While the notetaking assignments bore considerable similarity to the homework assignments used in Studies 2 and 3 of this dissertation, the method of evaluation differed in several key aspects. The notetaking assignments were a combination of in-class and out-of-class assignments, thus differing from the homework in Studies 2 and 3 (which were exclusively out-of-class assignments). The length measure in the Williams and Worth (2002) study required a subjective evaluation of the amount of space used and ranged from 0 to 3, as opposed to being an objective line count as in Study 3. And lastly, the accuracy measure was a holistic subjective evaluation (with inter-rater agreement) in the Williams and Worth study, while Studies 2 and 3 employ a potentially more objective tally of inclusion of pre-identified concepts in the students' answers.

The results in the Williams and Worth study (2002) provided evidence that notetaking was significantly correlated with exam performance ($r = 0.49$). Additionally, the in-class notetaking, out-of-class notetaking, and attendance were all significantly correlated with one another ($r = 0.38$ to 0.72). However, critical thinking was not significantly correlated with any of these behaviors. One shortcoming to this study was that it only assessed statistically significant relationships between these behaviors and student performance, but did not manipulate the variables to determine causality.

A similar study by Williams and Eggert (2002) examined the relationship between notes taken during class discussion and out-of-class notes pertaining to course readings. Notes were again recorded in the course study guide provided to students. The items in the study guide were equally distributed between the course readings and the class discussion topics. Students were required to fill out the study guide and hand it in for evaluation at the end of the semester. The answers were then evaluated along three measures: the completeness of their notes (the number of questions the students attempted), quantity (the number of words included in student answers), and accuracy. Accuracy was evaluated by graduate assistants, who rated the answers on a scale of 0 to 5. The raters achieved inter-rater agreement in the mid-to-high 0.90s. However, only 20% of the student questions were evaluated for accuracy or quantity.

The results indicated that notes taken during class discussion were a better predictor of exam performance than notes taken over the readings, and the accuracy of the notes was a better predictor than either completeness or quantity. The authors suggested that perhaps cognitive abilities (such as processing previously presented information while attending to ongoing class discussion) required to effectively take notes during a class discussion increases the potency of notetaking as a predictor of exam performance. Unfortunately, the method used to evaluate the accuracy of student answers in their notes was not precisely operationalized .

Practice exams. Following up on the Williams and Worth (2002) and Williams and Eggert (2002) studies, Oliver and Williams (2005) evaluated how contingencies related to completion and accuracy on practice exams affected actual exam performance. This study is, in some ways, the antecedent for Study 3 of this dissertation. Practice

exams, a shortened multiple choice, take-home exam, constitute a type of homework. The practice exams grant students an opportunity to work through a number of questions that are formatted and worded in a manner similar to those found on the actual exams. Students were advised that they would receive practice exam credit for either the number of items they completed (the completion contingency) or the number of items they answered correctly (the accuracy contingency). The results showed that students performed better on the practice exams under the accuracy than the completion contingency. Importantly, the results also showed that this improvement extended to the actual exams, where student scores increased significantly under the accuracy contingency.

Daily writing activities. Similar to the studies on notetaking (a form of written homework), Turner et al. (2006) assigned students brief daily writing activities to be completed at the beginning of class meetings in an effort to improve exam performance. In one semester the students did not engage in these writing assignments, while in another semester the writing assignments were required. The results indicated that students performed better on exam items that were related to the writing assignments, but the improvement did not generalize to unrelated exam items.

In a similar study, Hautau et al. (2006) linked different reward contingencies to daily writing assignments. Students could receive credit for all of their writing in each unit, credit for one randomly selected writing activity in each unit, or no credit for any writing in each unit. The results showed that the daily credit contingency produced higher exam scores than the random or no credit contingencies, and similarly affected writing scores (i.e., better writing when receiving daily credit). The writing scores were

significantly correlated with exam scores under all of the conditions, ranging from r = 0.66 to 0.78. Inasmuch as daily writing assignments bear similarity to homework assignments, the results of the study strongly support the notion that homework is a viable means of improving student performance. Because the Hautau et al. study did not feature changing conditions (students remained in the same contingency for all units), further research is necessary to establish a causal relationship.

Self-perception. A study conducted by Williams and Clark (2004) examined students' perception of their own ability, the amount of effort they put forth in studying for exams, and the degree to which they felt the instructors made a contribution to their understanding of exam material. Using a 12-item self-report survey, students rated on a scale of 1 to 3 ("low" to "high") how much each of these variables factored into their performance. Students indicated they believed their effort had the greatest impact on their exam scores, but their ratings of their own ability and the teacher's input actually correlated more strongly with exam performance. The high-achieving students appeared to have the greatest confidence in their own abilities and higher regard for teacher contributions to their exam performance. Unfortunately, only 28% of the low-achieving students actually completed the ratings (in contrast to 82% of the high-achieving students).

The results of the Williams and Clark (2004) study suggest that students may not have a clear understanding of what is actually influencing their performance, or are unable to clearly articulate the contributing factors. Another possibility is that students will tend to indicate they put forth considerable effort, regardless of the amount of time they actually spent preparing for an exam, which would explain the low correlation of

effort with exam performance. In this study, students completed the surveys immediately after receiving their exam results, which suggests that student perceptions of their ability (i.e., self-efficacy) may have been adversely affected by the feedback they received. In contrast, a study conducted by Galyon et al. (2012) found that academic self-efficacy, when evaluated prior to taking course exams, was mildly predictive of student performance. Students who performed poorly on the exams in the Williams and Clark study were also more inclined to rate the teacher's contribution to their performance more poorly (i.e., an external locus of control).

Pre-course knowledge. Another factor thought to possibly impact student performance is pre-course knowledge of both course-relevant content (i.e., psychological knowledge) and vocabulary. Many students have indicated they felt the language used on the exams was a significant factor in their performance. Turner and Williams (2007) tested this possibility by comparing pre-course vocabulary knowledge, pre-course content knowledge, and critical thinking as predictors of student exam performance. Students were administered a pre-course vocabulary test at the beginning of the course, and then re-administered the vocabulary test at the end of the course to evaluate potential improvement in their vocabulary. The results indicated that all three predictors were significantly related to exam performance, though pre-course vocabulary was the strongest predictor, followed by pre-course knowledge and critical thinking, respectively. Students who performed well in the course also made significant gains in vocabulary knowledge ($p < 0.001$). This study strongly suggests that the knowledge students possess as they enter the course will significantly affect their success in the course, perhaps even more so than other cognitive abilities such as critical thinking.

Framework for the Current Study

This dissertation is a combination of three studies involving an undergraduate educational psychology course. The course is offered at a large Southeastern university and is required for entry into the teacher-education program. The first study includes an analysis of trends over a period of seven years (Spring of 2004 through Spring of 2011) and identified patterns of exam performance and critical thinking between students at different grade levels, students with different critical thinking abilities, and students with different GPAs. This analysis is necessary to clearly understand past trends regarding exam performance (i.e., typical changes in exam scores from unit to unit).

The second study has already been completed and the data analyzed (Galyon, Blondin, Forbes, & Williams, in press). It examined a combination of potential predictors of exam performance in a recent semester (Spring 2010) including critical thinking, homework completion, and participation in class discussion. The results of this study substantiated a significant relationship between homework completion and exam performance. While critical thinking significantly correlated with both homework completion and participation in class discussion, homework completion and participation in class discussion did not significantly correlate with each other. Principally, while the results suggested a significant relationship between homework completion and exam performance, a lack of experimental manipulation prevented establishment of a causal relationship.

The third study (conducted in Fall 2011) is a continuation of research on the role of homework in improving exam performance**.** Whereas the second study attempted to determine if there is a significant relationship between homework accuracy and exam

performance, the third study was intended to determine if this relationship could be manipulated to improve student performance, and if a reward contingency can be used to significantly improve homework accuracy. To that end, the third study compared the effects of credit contingencies (linked to accuracy versus completion of homework) on both accuracy and completion of homework, as well as the indirect effects of the credit contingencies on exam performance.

Chapter III

General Method

Overview

In the following studies I examined data collected over the past seven years and analyzed apparent trends in student exam performance. I proceeded to evaluate homework accuracy as a predictor of exam performance, and finally evaluated the efficacy of targeting homework in improving exam performance. The latter two studies collected information on student homework accuracy and student exam scores along with critical thinking, an established predictor of student exam performance (Wallace & Williams, 2003; Williams, Oliver, & Stockdale, 2004; Williams & Worth, 2002). These studies attempted to determine what effect homework accuracy has on student exam performance and the efficacy of targeting homework performance in attempting to improve exam scores.

Participants

Participants in all three studies were drawn from an undergraduate educational psychology course required for entry into a teacher-education program at a large, Southeastern state university. The course was comprised of five units addressing different facets of human development. Unlike many undergraduate courses, the targeted course was discussion-based rather than lecture-based. Given the comparatively demanding nature of the course, undergraduate students at the freshman level were encouraged to consider withdrawing from the course and re-enrolling at a later date. Nevertheless, some freshmen remained enrolled in the course despite this advice.

The students ranged from freshmen to graduates, but the majority tended to be sophomores (34.7%) and juniors (23.4%). Most of the students were female (66.9%). More detailed demographic information about the sample can be found in Table 1. Also the majority of the students were Caucasian, though a diversity of students has been observed, including African-American, Asian, and Hispanic students. Nonetheless, demographic information on student ethnicities was not available.

Measures

All three studies collected background information from students indicating their gender, prior GPA, expected grade in the course, hours of employment outside school, number of courses in which they were enrolled, and year in school (e.g., "sophomore"). All three studies also included data on student exam performance and critical thinking ability, as measured by the *Watson-Glaser Critical Thinking Appraisal Form-S* (WGCTA-S, Watson & Glaser, 1994). Additionally, the second and third studies included the collection of information on participation in class discussion and completion of homework assignments. Information on GPA, gender, year in school, critical thinking, and exam scores was recorded with some degree of consistency over the past seven years.

Assessment of exam performance. To evaluate performance in the course, students completed five 50-item multiple-choice exams each semester. Answers were recorded on scan sheets, which were given to the instructor upon completion. The instructor was able to immediately grade the tests using a test-scanning machine. This arrangement provided students with immediate feedback on their performance and indicated correct answers for missed items. This system was efficient and objective, and students indicated they appreciated receiving their grade in this expeditious manner.

Previous research has been conducted on the exam items to evaluate the type of knowledge required to answer items correctly. The researchers found that items could be categorized as either direct recall (26%) or comprehension (58%), with inter-rater agreement of 73% (Wallace & Williams, 2003). Some items failed to achieve a high level of categorical agreement amongst raters (< 66% inter-rater agreement) and were thus categorized as "mixed" items. In another study, internal consistency of student responses to exam items across units was reported as 0.87 (Turner et al., 2006). Total exam scores were evaluated as a measure of student performance, rather than just exam items corresponding to homework questions. Though evaluating exam items corresponding directly to selected homework questions would likely be more sensitive to the relationship between the two, the entire exam may better represent student mastery of the course content as a whole and is consistent with assessment of exam performance prior to Studies 2 and 3.

Critical thinking. As has been suggested previously, students are presumed to enter the course with certain cognitive and behavioral characteristics. Critical thinking was evaluated as a measure of pre-course cognitive ability. It was assessed at the beginning of the course using the *Watson-Glaser Critical Thinking Appraisal Form-S* (WGCTA-S, Watson & Glaser, 1994). The WGCTA-S is a shortened version of the assessment and is intended primarily for adults. This test measures several components of critical thinking including inferences, deduction, assumption recognition, and interpretation and evaluation of arguments (El Hassan & Madhum, 2007). The test includes 40 items with two to five answer options, though only one correct response for each item.

There are a number of psychometric characteristics that make the WGCTA-S an excellent measure of critical thinking ability. It achieves a degree of cultural fairness by providing all information necessary within the test materials for answering the questions. It has been evaluated for internal consistency ($r_{alpha} = 0.81$) and test-retest reliability ($r = 0.81$, $p < 0.001$) (Watson & Glaser, 1994). The test also has been evaluated multiple times for criterion-related validity and has demonstrated a significant relationship with other measures of academic outcomes (Hildebrandt & Lucas, 1980; Hurov, 1987; Steward & Al Abdulla, 1989; Wilson & Wagner, 1981). Lastly, the WGCTA-S has proven to be one of the most consistent predictors of exam performance in the course targeted in the present set of studies (Wallace & Williams, 2003; Williams et al., 2004; Williams & Worth, 2002). Students were provided 30 minutes at the beginning of the semester to complete the WGCTA-S.

Chapter IV

Study 1

Overview

Study 1 was conducted to review and synthesize previous data regarding performance in the undergraduate educational psychology course used in Studies 2 and 3 and identify trends in exam performance over the past seven years. The data included students' gender, year in school, GPA, critical thinking score (as evaluated by the WGCTA-S), exam scores, and quiz scores. For purposes of this study, exam scores were evaluated across different course units and as a whole between year in school, GPA, and critical thinking scores. As essay quiz scores are considered to reflect a more subjective evaluation of student answers, they were not included in this evaluation. Study 1 included 3,591 participants, the majority of whom were female (67%). Though the majority of the students were sophomores (34.7%) and juniors (23.4%), the sample also included freshmen (6.2%), seniors (10.2%), and graduate students (5.7%). Approximately 19.9% of the sample either did not report a grade (19.2%) or indicated a classification other than freshmen through graduate (e.g., "non-degree seeking student").

Retrospective Analysis of Exam Scores and Critical Thinking

The content of the exams has changed somewhat over the years, but the course material has remained largely consistent and changes in the exam questions have been comparatively minimal. One significant change in the course structure was a reversal of the third and fourth units (Unit C and Unit D). Prior to 2006, these two units were switched in their sequence such that what is presently considered Unit C, was previously Unit D and vice-versa. To correct for this reversed sequence, the exam scores for these

two units prior to 2006 were reversed in the database prior to analysis. After that correction, the sequence of course units was consistent across all years and the content of the course was largely the same across years. Therefore, the data were grouped together across all years and evaluated as a single sample.

Between units exam performance. Analysis of variance (ANOVA) in the scores across units indicates regularly occurring changes between units, $F(4, 15617) = 95.572$, $p < 0.001$. As shown in Table 2, post-hoc comparisons reveal Unit B scores are significantly lower than all other units ($p < 0.001$). Additionally, Unit C scores are higher than all other units ($p < 0.001$ for all). No other significant differences were found between units.

Year in school, critical thinking, and exam performance. Further analysis examined the total exam scores and critical thinking ability of students at different years in school. An ANOVA indicated significant differences in the critical thinking ability of students at different years in school, $F(6, 2055) = 3.83$, $p = 0.001$. More specifically, post-hoc analyses (Table 3) showed graduate students scored significantly higher on critical thinking than Freshmen ($p = 0.004$), Sophomores ($p = 0.032$), and Juniors ($p = 0.018$) students. Additionally, Seniors tended to score significantly higher than Freshman students ($p = 0.026$) on critical thinking.

Significant differences were also revealed in the total exam scores of students across grade levels, $F(6, 2458) = 14.27$, $p < 0.001$. Post-hoc analyses (Table 4) indicated that graduate students scored significantly higher on exams than students at all other grade levels ($p = 0.001$ to $p = 0.004$) and seniors scored significantly higher than students at lower grade levels ($p = 0.001$ to $p = 0.008$). These differences persisted after

controlling for critical thinking scores. No significant differences were observed between students at other grade levels.

GPA, critical thinking, and exam performance. I then evaluated the predictive potential of GPA with respect to exam performance and a possible interaction between GPA and critical thinking ability in predicting exam performance. The sample was divided by GPA into high (GPA >= 3.00) or low (GPA < 3.00) groups. Similarly, students were categorized as having a high critical thinking ability (>=34, corresponding to the 75th percentile or higher) or low critical thinking ability (<=24, corresponding to the 5th percentile) based on a national ranking of college graduates. An ANOVA indicated that students with a high GPA were significantly more likely to have higher critical thinking scores, $F(1, 1536) = 20.95$, $p < 0.001$, and higher exam scores, $F(1, 1834) = 213.26$, $p < 0.001$, than students with a low GPA. Furthermore, students with high critical thinking ability were more likely to have both higher exam scores, $F(1, 1213) = 253.24$, $p < 0.001$, and higher GPAs, $F(1, 871) = 52.25$, $p < 0.001$, than students with low critical thinking ability.

Although these results were also evaluated for possible interaction effects between critical thinking and GPA, no interaction was indicated, $F(1, 726) = 3.38$, $p = 0.066$. Furthermore, after controlling for critical thinking as a covariate, there was still a significant difference between the high and low GPA groups on exam performance, $F(1, 727) = 65.97$, $p < 0.001$.

Finally, there appeared to be no significant differences between high and low critical thinkers with respect to changes in GPA as students progressed through their academic years. Both groups appeared equally likely to improve as they advanced across

academic years, $F(5, 829) = 0.297$, $p = 0.915$. One thing that should be noted in the interpretation of these results, however, is that college students are already selected ostensibly from the higher-performing students in secondary education. They represent a population of restricted variability with respect to GPA and possibly with regard to critical thinking ability.

Discussion

The retrospective analysis revealed a number of interesting findings relevant to the subsequent studies in this dissertation and other future studies. First, there was clearly a pattern of differential exam performance across units in the course. This pattern largely held true across the past seven years. With remarkable consistency, students performed worse on average during Unit B than in any other unit of the course. This tendency may be explained by the difficulty of the content in Unit B (e.g., brain functioning and cognitive psychology), the more demanding vocabulary (i.e., large number of potentially unfamiliar terms) and the smaller likelihood that students had a substantial amount of pre-course knowledge on these topics. Previous studies have established the significance of pre-course knowledge and vocabulary on exam performance within the course (Turner & Williams, 2007), which may provide an explanation for the results of this analysis.

Similarly, students almost always performed better in Unit C. The nature of the content in Unit C may be more familiar as many students have had the opportunity to work in cooperative groups and have first-hand experience with social interactions. Thus, vocabulary and pre-course knowledge may have been less of a factor in this unit. However, a notable characteristic of Unit C was also the cooperative learning groups

exclusively used in that unit. As demonstrated in a series of studies, cooperative learning groups can substantially improve student performance (Carroll, Williams, & Hautau, 2006; Stockdale & Williams, 2004; Williams, Carroll, & Hautau, 2005). The credit contingency used in Unit C (i.e., group credit and individual accountability) has also proven to be the most effective contingency to use with cooperative learning groups (Carroll, Williams, & Hautau; Williams, Carroll, & Hautau).

A number of interesting patterns emerged with respect to year in school and student performance. Graduate students scored significantly higher than students at other grade levels on their critical thinking test and exam performance. This pattern is perhaps not surprising, as students who have completed an undergraduate education are expected to demonstrate a cumulative increase in their critical thinking ability (Brabeck, 1983; King et al. 1990; McMillan, 1987; Mines, King, Hood, & Wood, 1990; Williams, Oliver, Allin, Winn, & Booher, 2003; Williams, Oliver, & Stockdale, 2004) and hopefully have developed more effective study habits and test-taking skills.

Following this same pattern, college seniors typically score higher on the critical thinking test than freshmen. However, no significant differences were obtained at other grade levels (e.g., juniors scoring higher than freshmen, or seniors scoring higher than sophomores). This pattern is similar to previous research findings, which showed no significant differences between undergraduate levels when critical thinking was not specifically targeted (Browne, Haas, Vogt, & West, 1977; Worth, 2000). When made the focus of a course, substantial gains in critical thinking have been observed within a comparatively short period of time (Allegretti & Frederick, 1995; Tsui, 1998). These results suggest that, in many instances, a near-completion of an undergraduate education

may be required for the difference in critical thinking ability to become significant. Commensurate with their higher critical thinking scores, seniors also scored higher than students at all other grade levels on their exams. A variety of other academic behaviors could also account for some of these differences, such as time management, organization, or development of effective study habits.

Another result from the analysis demonstrated significant differences in the critical thinking and exam scores of students with high GPAs. Students with high GPAs (3.00 or higher) scored significantly higher on critical thinking and exam scores than their low GPA (less than 3.00) counterparts. Similarly, students with high critical thinking scores (34 or higher) obtained significantly higher exam scores and higher GPAs than students with low critical thinking (<= 24). Initially this difference might suggest that because critical thinking and GPA were clearly related to each other, GPA might confound the relationship between exam performance and critical thinking (or vice-versa). However, additional analysis showed no significant interaction after controlling for critical thinking as a covariate (i.e., significant differences in exam scores were still observed for the two groups of GPA).

Previous researchers have demonstrated that students with high critical thinking ability tend to perform better on course exams regardless of their academic history (as represented by GPA) (Williams & Worth, 2002). Also, students who have effective study habits generally perform well on course exams regardless of their critical thinking ability (Williams & Stockdale, 2003). Inasmuch as GPA reflects a student's academic history, it follows that their exam performance would continue to be high. The results of the current study appear to support these findings.

The final set of analyses investigated the possibility that students with different levels of critical thinking might exhibit differential trends in their performance as they progressed through school (i.e., that students with lower critical thinking may exhibit decreasing performance in response to the increasing demands of their undergraduate education). To evaluate this possibility, I examined GPA trends across years in school for different levels of critical thinking (high and low). No significant interaction effects were observed, suggesting that both groups were equally likely to improve (or decline) as they advanced through their undergraduate studies.

Chapter V

Study 2

Overview

Study 2 was conducted to investigate a hypothetical relationship between the accuracy of student homework answers and subsequent exam performance. Specifically, I sought to determine if the accuracy of student answers to homework questions (aka "homework accuracy") accounted for a significant amount of variance in exam scores and how homework accuracy compared to other known predictors of exam performance (i.e., critical thinking and participation in class discussion). Two independent raters scored the students' homework assignments for accuracy at the end of the semester using the methodology described below. The results of the homework evaluation were then correlated with exam performance and compared to other predictors of exam performance, including critical thinking scores and participation in class discussion. All predictor variables were then entered into a regression equation to determine the best combination of variables to predict exam scores.

Method

Participants. The participants included 167 students enrolled in three sections of an undergraduate educational psychology course during Spring, 2010. The participants included a ratio of approximately 3:1 females to males, which is historically typical for this course. The student sample spanned freshmen to graduate students, though the majority of the sample consisted of sophomore and junior students. Specifically, the sample consisted of freshmen (9.8%), sophomore (51.2%), junior (23.8%), senior (11%),

and graduate students (3.7%). More information on the student sample is available in Table 5. On average, students entered the course with a self-reported GPA of 3.19.

Measures. Measures included critical thinking scores as assessed by the *Watson-Glaser Critical Thinking Appraisal Form-S* (WGCTA-S, Watson & Glaser, 1994), self-reported frequency of participation in class discussion, a measure of homework completion based on accuracy and thoroughness of student responses, and measures of exam performance.

Critical thinking. Critical thinking was evaluated at the beginning of the semester using the same methodology outlined previously. Students were given 30 minutes to complete the WGCTA-S at the beginning of the semester. Despite being a measure of generic, rather than subject-specific critical thinking ability, this relatively short test has proven to be a consistent predictor of exam performance (Williams, Oliver, & Stockdale, 2004; Williams & Stockdale, 2003; Williams & Worth, 2002).

Recording participation in class discussion. Student participation in class discussion was measured by classroom observation. Two observers (graduate teaching assistants) conducted observations on four days during each of the five units in the course (a total of 20 days of observation). The observers recorded the number of voluntary, content-related student comments. A comment was considered "voluntary" if the instructor did not request the student to make a comment, and the student's comment was relevant to the current class discussion. Observer records were evaluated for inter-rater agreement by comparing the records of each observer during each unit and dividing the smaller number of recorded comments by the larger number. Substantial inter-rater agreement was achieved (an average of 97.69%).

In an effort to increase the number of comments students made in class discussion, students could earn credit for making one to two comments per day during selected units in the course. This credit was based on students' self-reported records of their participation. Students recorded their comments and questions on record cards, which were submitted at the end of class on the four discussion days in each of the five units. Students were not informed for which days or even in which units participation credit would be awarded until the end of the course. Students were simply informed that the days and units for participation credit would be randomly selected at the end of the course.

Homework question selection and rubric creation. For each of 5 units in the course, 10 homework questions were selected for evaluation. Three graduate teaching assistants who had previously taught the course and were familiar with the content independently evaluated these questions. The raters found these questions illustrated the strongest correspondence with exam items and the largest number of course concepts included in the official instructor answer. The number of questions from which the 10 target questions were selected varied from 41 to 55 per course unit. The selection process was based on two criteria: the number of exam questions corresponding to a particular homework question and the number of unit concepts included within the official answer.

Three independent raters (the afore-mentioned graduate teaching assistants) identified the number of exam questions corresponding to each homework question. Correspondence was established when any of the concepts in the official answer to a question were also included in an exam question. The raters then tallied the number of exam questions that corresponded with homework questions to produce a correspondence

index. Agreement between raters on the correspondence index was evaluated across five units and produced an average of 84% inter-rater agreement. The homework items with the highest correspondence index were selected for inclusion first, and then selection proceeded to those with the lowest correspondence index. This arrangement is illustrated in the following sample chart:

Homework Question	Corresponding Exam Questions	Correspondence Index
1	1, 3, 47, 48	4
2	15, 25, 36	3
3	18, 32	2

In addition to the correspondence index, raters independently identified concepts that should be reflected in a complete answer. Concepts included information presented in the course through lecture, assigned readings, or official notes in the course text. Each rater independently produced a list of concepts that reflected a maximally complete answer. These concept lists were evaluated for inter-rater agreement. Average inter-rater agreement on concept identification across 50 questions was 78%. The more inclusive list of concepts identified by a rater was retained and tallied to produce a maximum concepts score for a homework question.

Evaluation of student homework responses. Student answers to the selected questions were evaluated for the number of distinct, accurate concepts included. Students submitted their homework answers in a typed document at the beginning of the class

period in which those questions were scheduled. Accuracy was evaluated by comparing the concepts in the students' answers to the concepts identified by the raters in the official answer. The number of pre-identified concepts included in the student's answer was tallied, producing a Homework Accuracy and Completeness score. Inter-rater reliability for this score was evaluated by two raters across all five course units and reflected an average Pearson reliability coefficient of $r = 0.89$ (coefficients ranged from $r = 0.83$ to $r = 0.93$).

It was improbable that a student would achieve 100% on their Homework Accuracy score, as the number of concepts identified in the official answer was quite extensive. A higher score, therefore, simply indicated a more complete and accurate answer than a lower score. Each correct concept included in the student's answer counted as one point. Absence of an answer was not recorded in the database. A score of 0 indicated an answer was present, but entirely incorrect.

Analysis and Results

Analysis of exam performance predictors in Spring 2010. Though there were three sections of the course, all analyses were conducted with all sections combined into one sample. A procedure recommended by Soper (2011) was used to calculate the necessary sample size to maintain adequate statistical power. Statistical conventions demand a sample size of at least 48 for an expected effect size of 0.25, which corresponds to $R^2 = 0.20$ (Cohen, 1989; Kraemer & Theimann, 1987; Soper, 2011). As the sample included in the present study was 167 students, it was deemed more than sufficient to maintain statistical power. In the interest of being thorough, I conducted the analyses for all units combined, as well as for each unit separately.

Differences in means on target variables by academic classifications. One concern with respect to the diversity of the population was that a wide range of students was represented, ranging from freshman to graduate level. It was thus important to determine if academic classification made a difference in scores on the variables included in the study. Generally, I found that graduate students attained higher raw-score means than undergraduates. The graduate student mean critical thinking score, for example, was 32.67 (approximately the 60th percentile), whereas the mean scores for undergraduates ranged from 25.06 to 27.71 (approximately the 10th to 25th percentiles respectively). However, these differences were not always statistically significant as shown below. Descriptive statistics for all variables by year in school are available in Table 6.

I ran a MANOVA to determine if there were significant differences between academic levels on the mean scores of the target variables. The results broadly indicated near-significant differences between academic levels, Wilks' $\lambda(16, 455.84) = 1.663$, $p = 0.051$. Subsequent univariate ANOVAs indicated significant differences between academic levels for average participation in class discussion, $F(4, 157) = 3.020$, $p = 0.020$. Additional post-hoc analyses were conducted to examine the specific differences between academic levels (Table 7). Graduate students participated more frequently in class discussion than underclassmen; however, the graduate student sample was very small ($n = 6$), making it difficult draw any substantive conclusions from this difference. No significant differences were noted between students at the various undergraduate levels. On the basis of these results, I proceeded to perform a regression analysis with the sample in its entirety rather than evaluating differences at each academic level separately.

Variable distributions. To provide context for the correlational and regression analyses used in determining the relationships between predictor variables and exam scores, I examined the nature of the distribution for each predictor variable and exam performance (see Table 8 and Figures 1 - 4). Because the sample distribution was relatively homogeneous on all variables across year in school and in some instances the majority of the scores were located in the lower range of possible values (positively skewed), the results of correlational and regression analyses were likely to be somewhat tempered.

Correlations between predictor variables and exam performance. Analysis of the combined data included multiple correlations and multiple stepwise regressions. The analysis first determined that in general critical thinking and composite homework accuracy were significantly correlated (r = 0.396, p < .001). However, participation in class discussion was not significantly related to either critical thinking (r = 0.118) or homework accuracy (r = 0.182).

Additional analyses were conducted for each unit as well (see Table 9). Critical thinking and homework accuracy were significantly correlated in Units A through C, but not in Units D and E. With respect to participation in class discussion, critical thinking was modestly, but significantly correlated with class participation in the first three units of the course (r = 0.175 to 0.196). Homework accuracy was also significantly correlated with class participation, but only in Unit A (r = 0.185) and Unit C (r = 0.185).

Partial correlations between predictor variables and exam performance. Another series of correlations were computed to evaluate the relationship between each predictor variable and exam performance, while controlling for the other predictor

variables (e.g., relationship between critical thinking and exam performance after accounting for the relationship between critical thinking and both homework completion and participation in class discussion).

Critical thinking and homework accuracy generally correlated significantly with exam performance ($p < 0.001$), and overall the two correlations were comparable for the first three units of the course (see Table 10). However, while the correlations in the last two units remained significant for critical thinking and exam performance ($r = 0.408$ to 0.423), the correlation between homework accuracy and exam performance did not remain significant ($r = 0.132$ to 0.135).

The partial correlations between class participation and exam scores were also significant for three units in the course ($r = 0.195$ to 0.261, Units C through E), though significantly weaker than the relationship between critical thinking and exam scores for the first and last units of the course. Furthermore, in the first unit of the course homework accuracy exhibited a significantly stronger relationship with exam performance than did participation in class discussion.

The final stage of the analysis (see Table 11) consisted of a series of stepwise multiple regressions to determine which combination of variables best-predicted exam performance. The results of the regression analysis suggest that homework accuracy was the single best predictor of exam performance for the first three units of the course, though a combination of homework accuracy and critical thinking yielded the best predictive formula for exam performance ($\beta_{CT} = 0.383$, $\beta_{HW} = 0.311$). Similarly, critical thinking was the best singular predictor of exam performance in the last two units, but a combination of critical thinking and class participation accounted for significantly more

variance in exam scores than did critical thinking alone. Though critical thinking proved to be the most consistent predictor of exam performance (consistently correlated across all units), a combination of critical thinking and homework accuracy explained more variance in exam scores (32%) than did critical thinking alone (25%). The only unit in which all three predictor variables contributed significantly to predicting exam performance was Unit C, which places a heavy emphasis on cooperative learning and may lend itself better to participation in class discussion.

To better explain the amount each predictor variable contributed to exam scores, I computed the proportionate variance each variable contributed to the *Adjusted* R^2. Because partial correlations represent the relationship between a predictor variable and the targeted outcome variable after accounting for inter-correlations between predictor variables, their squared values do not necessarily combine to equal the total percentage of variance explained (typically represented as *Adj.* R^2). Therefore the data in Figure 5 depict the portion of the *Adj.* R^2 that can be attributed to each predictor variable's partial correlation with exam performance. If x represents a predictor variable, r_x is the partial correlation between a predictor variable and exam performance, and i represents the set of all predictor variables, then the proportionate variance explained was calculated using the following formula: $\mathbf{(r_x / \Sigma r_i) * Adj.}\ \boldsymbol{R^2}$.

Discussion

The purpose of this study was to extend a line of research on predicting exam performance and identifying possible routes for improving exam performance. In this study I developed a method for evaluating student homework for accuracy and thoroughness and used it to compare the predictive potential of homework to other

previously-investigated predictors (critical thinking and participation in class discussion). The homework evaluation methodology was designed to promote a systematic and objective approach for evaluating homework accuracy and thoroughness in contrast to previously holistic and more subjective approaches (e.g., reading a student's answer and rating it based only on opinion). Nevertheless, the methodology had some shortcomings in that it still required considerable time and raters had to be well-informed in the subject matter to interpret the correctness of student answers.

Overall, homework accuracy showed promise as a predictor of exam performance and thus as an intervention target to improve student success. In fact, in some instances homework accuracy proved to be a more significant predictor than critical thinking, a well-established predictor of exam performance, and should be more amenable to change through intervention than critical thinking.

Closer inspection of the findings. Homework accuracy proved to be a significant predictor of exam performance in the first three units of the course. In contrast, critical thinking was a significant predictor across all course units. One possible explanation for this result is that students may be more likely to focus on homework during the beginning of a course while they are becoming acclimated to the course structure. Such a focus may then fade as students become more comfortable with the course. An alternative explanation is that factual information was more strongly emphasized in the first three units, while reasoning was a stronger focus in the last two units. An opportunity to rehearse information for later recall may partially explain the difference in the predictive potential of homework and critical thinking.

Another consideration is the effect that the type of homework assignment and

type of exam may have on the relationship between these two variables. In the current study, the homework assignments were brief written answers, whereas the exams were strictly multiple-choice. Conceivably, short essay homework assignments would better predict performance on a topographically similar essay exam, or multiple-choice homework assignments would better predict performance on a multiple-choice exam. Indeed, previous research on the effect of manipulating accuracy and completion contingencies on a multiple-choice homework assignment demonstrated a significant effect on the multiple-choice exam as well (Oliver & Williams, 2005). The correlations between these multiple-choice homework assignments and the exams were generally equal to or greater than those obtained in the current study between homework and exam performance. There may be some value in increasing the similarity between the homework task and the exam. However, it can be difficult to construct a multiple-choice homework assignment that fully encompasses all of the major course concepts.

The results from this study also confirmed previous findings regarding the relationship between critical thinking and exam performance. Where previous research has demonstrated a consistent relationship between critical thinking and exam performance (Wallace & Williams, 2003; Williams & Stockdale, 2003; Williams & Worth, 2002; Worth & Williams, 2001), the current study served to extend this finding by comparing critical thinking to homework accuracy and participation in class discussion as predictors of exam performance. However, despite the considerable ability of critical thinking to predict exam performance, it has been found generally less suitable as an outcome variable given the difficulty in effecting a substantive improvement in critical thinking in a single course (Williams et al., 2004).

In contrast to research on critical thinking and course performance, relatively little research has addressed the relationship between participation in class discussion and exam performance. The results of the current study indicate that though sometimes a significant predictor of exam performance, participation in class discussion is generally not a strong predictor of exam performance. It is possible to increase the likelihood students will participate in class discussion by providing reward contingencies in the form of credit for quantity of comments, or increase the quality of student participation by providing credit contingent on the quality of comments (Foster et al., 2009; Krohn et al., 2010). However, it appears that some students will refuse to participate in class discussion regardless of the credit contingencies in place.

Limitations. The greatest limitation to the current study was undoubtedly the inability to establish a cause-effect relationship between homework accuracy and exam performance. While the results certainly suggested a significant relationship between homework accuracy and subsequent exam performance, neither variable in this relationship was manipulated. Two events that followed submission of student homework further mitigated the strength of the results in this study. First, within the context of class discussion, homework items were specifically introduced as a topic of conversation. Although many students' homework answers were far from being complete or accurate, they were able to clarify their misunderstanding or address an incomplete understanding of the homework questions during class discussion.

A second confound was that students received a complete set of instructor answers to all of the homework questions at the end of each unit but prior to the exam. Thus, students had a second method of correcting their understanding of course concepts

even if they failed to participate in or attend to class discussion. Indeed, many students anecdotally reported they found the instructor answers useful in clarifying their understanding of course concepts represented in homework assignments.

A third limitation to the current study was the nature of the sample. As mentioned in the analysis, the critical thinking and homework scores featured distributions that were heavily skewed toward relatively low scores, which likely diminished some of their potential to predict exam scores. Similarly, the sample population was drawn from an undergraduate university class, which included some selection criteria (e.g., GPA and standardized test scores) that further restricted the range of student abilities (e.g., cognitive abilities and study habits). If the study were replicated with a broader range of abilities and characteristics (e.g., in primary and secondary education classrooms), both the strength and the generalizability of the findings might have been strengthened.

There are also some concerns regarding the procedure used to collect data in the current study. Class participation was demonstrably the weakest of the predictor variables, though when combined with critical thinking, participation significantly improved the ability to predict exam performance in Units D and E. Participation also improved the ability to predict exam performance in Unit C (the cooperative learning unit), when combined with critical thinking and homework. The ability of participation in class discussion to predict exam performance may have been partially limited by the use of a strictly quantitative measure (e.g., how many comments a student made) instead of a qualitative component (e.g., how useful the comments were in advancing discussion). A qualitative measure may have been a better predictor of exam performance, as it could tap into the depth of a student's understanding of the course

concepts. Certainly, a well-informed discussion could be of more benefit to students than a discussion replete with unfounded opinions.

Multiple-choice tests can differ greatly with respect to their content. Tests can emphasize factual recall or critical thinking to different degrees. Many tests likely emphasize the former over the latter and may benefit minimally from strong critical thinking ability. On the other hand, tests that heavily emphasize a student's ability to reason through relevant information to arrive at the most defensible conclusion to an item likely would benefit substantially from critical thinking ability. Accordingly, the degree to which critical thinking predicts exam performance will vary according to the nature of the exam. Therefore, the generalizability of the findings from this study, with respect to critical thinking ability, is limited to the degree to which critical thinking itself is emphasized on course exams.

Components of this study would benefit from being replicated across a variety of college courses, student populations, and pre-college grade levels (e.g., elementary and secondary education). The student sample, which was mainly comprised of Caucasian, female, college sophomores and juniors, may differ from other populations at different grade levels (e.g., college senior seminar or graduate-level courses, or middle or high school students), in different areas of study (e.g., mathematics or biology), or with different demographics (e.g., minority cultures). While rehearsal of information, which is ostensibly a primary feature of homework activities, should be beneficial in a variety of subject areas, its impact on student performance may be quite different in various subject areas or with students at different grade levels. Similarly, the nature and benefits of

participation in class discussion could be substantially different at other grade levels in which students may be more (or less) reliant on teacher instruction.

Chapter VI

Study 3

Overview

Whereas Study 2 substantiated a relationship between the accuracy of student homework answers and exam performance, Study 3 endeavored to manipulate contingencies that would improve accuracy of homework answers to produce corresponding changes in exam performance. Thus, Study 3 had two primary hypotheses: 1) to determine if a reinforcement contingency for accuracy could affect accurate completion of student homework answers, and 2) to determine if changing the accuracy of student homework answers would lead to a change in exam performance. To test these hypotheses, I evaluated student homework assignments for accuracy each day in three separate sections of the course. Accuracy and completion contingencies were applied in consecutive units in three evening sections of the course.

Students were informed that their homework grade would be determined by a set of evaluation criteria specific to each condition. The criteria included accuracy of student responses (compared to concepts indicated in a rubric) and completion of homework (reflecting only the number of questions for which students provided an answer). Students received feedback each day during the Accuracy Contingency. The feedback included both the number of points earned, as well as the answer represented by the rubric. Feedback linked to the Completion Contingency was given only at the end of each unit.

Method

Participants. The total sample, including the students who were not participating in the homework contingencies (Control Group), was 237 students drawn from six sections of an undergraduate teacher education course in Fall of 2011. Participants in the homework contingencies included 71 students in three sections of the course. An additional 166 students from the remaining three sections of the course did not participate in the homework contingencies but rather constituted one large control group for comparison with the sections participating in the homework contingencies. The student grade classifications included Freshmen (2.1%), Sophomores (47.3%), Juniors (29.5%), Seniors (13.9%), and Graduate Students (3.4%). Approximately 3.8% of the students did not report their grade classification or were non-degree students. More information on the sample is available in Table 12. The average self-reported GPA was 3.24 upon student entry into the course.

Measures. Many of the same measures used in Study 2 were employed in this study as well, including critical thinking as assessed by the *Watson-Glaser Critical Thinking Appraisal Form-S* (WGCTA-S) (1996) and measures of homework accuracy homework length, and exam performance. Evaluation of critical thinking was conducted in the same way as in Study 2. However, modifications were made in the evaluation of exam performance and student homework answers as described below.

Change in exam performance. Study 1 revealed a pattern in exam scores across units that was used as a point of comparison for exam scores achieved in the current study. The typical pattern of variability in exam scores across course units could blur changes in exam scores resulting from the homework contingencies. For example,

students have historically decreased their exam scores from Unit A to Unit B. However, if a student shifted from the one homework contingency in Unit A to a different homework contingency in Unit B, the exam score in Unit B may be indistinguishable from the exam score in Unit A (e.g., both 37/50). This pattern would contrast with the typical decrease in exam scores in the transition from Unit A to Unit B (i.e., approximately a 2.5 point drop). Thus, the homework contingency in Unit B may have significantly increased the exam mean in Unit B compared to a typical Unit B exam mean, even though the mean remained the same as Unit A.

It is possible to determine if the contingencies were having the above-described effect by evaluating changes in exam scores from unit to unit (i.e., if a student's exam score increased or decreased from one unit to the next). This comparison would better account for variability between units as it only compares changes in exam scores within students and would more precisely evaluate the intended effect of the contingency on improving exam scores. To achieve this analysis, I subtracted the exam score of one unit from the exam score in the subsequent unit (e.g., Unit B score - Unit A score = Change in Exam Score).

Evaluation of accuracy of student homework responses. Study 3 used the same general methodology to evaluate the accuracy of student homework answers as was used in Study 2. However, before the study began, the rubrics used for evaluating student answers were re-examined. In Study 2, I found that a high degree of precision in identifying concepts included in an ideal answer produced a large number of criteria for evaluating answers to each question. Thus, many more concepts were identified for possible inclusion than were likely to be reflected in student answers. This discrepancy

greatly diminished the possibility that students would acquire scores indicating 100% correct answers and conversely resulted in comparatively few concepts being fully and correctly identified in student answers (the average percent correct across all units in Study 2 was 41%, the minimum was 7%, and the maximum was 67%). Because evaluations of accuracy in student answers formed the basis of student homework grades in some contingency units in Study 3, I decided that the rubrics needed to be less rigorous to increase the likelihood that students would receive higher grades.

In preparing the new rubrics, the raters (a graduate teaching assistant, an undergraduate research assistant, and I) focused more on identifying broad concepts that might indicate comprehension of the material. Rather than identifying specific concepts (e.g., “Students may be less likely to wear seatbelts” and “Students may be more likely to drive quickly”), the raters instead used broader equivalents (e.g., “Students may engage in more reckless behaviors”). On the one hand, this broadening of concept delineation enabled a rater to be more lenient in grading student responses (i.e., giving students credit for a larger variety of answers that reflect the same concept) and reduced the maximum number of points available for a question (e.g., from two points to one point in the example above). A reduced number of maximum points could increase the likelihood of students receiving a higher percentage of credit for their answer to a homework question. On the other hand, this procedure also permitted more room for interpretation from the raters and thus necessitated additional inter-rater evaluations beyond those conducted in Study 2. The rubrics used for evaluation in Study 3 are available in Appendices J through N.

An undergraduate research assistant trained in using the rubric and I conducted inter-rater evaluations in each unit. We used rubrics to independently evaluate approximately two homework answers for the same 20% of the students each day. Inter-rater correlations were computed using Pearson-correlation coefficients. Inter-rater agreement was generally quite adequate across units (Overall $r = 0.82$; Unit A $r = 0.81$, Unit B $r = 0.86$, Unit C $r = 0.71$, Unit D $r = 0.77$, Unit E $r = 0.93$).

Student instructions and feedback on homework performance. At the beginning of each unit students were informed of the point contingency in effect via an announcement in class, on the course website, and through email (see Appendixes C through G for these contingency announcements). The exact sequence of the contingencies across units varied between the three sections participating in the study. The sequence of contingencies can be seen in the flowchart in Appendix I.

The design of the announcements was changed from Unit A to the subsequent units. In Unit A, a broad announcement providing a description of both the Accuracy and Completion Contingencies was made available to students (see Appendix C). Students were then informed which contingency was in effect for the unit and no further announcements were made. Performance feedback was given at the end of the unit only. At the end of Unit A several students expressed confusion regarding the contingency. On all subsequent units, the announcement was simplified (see Appendixes D through G). The announcement detailed the contingency in effect, the criteria for earning points, and in the case of the Accuracy Contingency, a table showing the number of points earned for each percentage-correct category. Student confusion appeared to be significantly reduced.

Thereafter, performance feedback was provided to students each day via email during the Accuracy Contingency. The email followed a standardized format to ensure all students received the same information. The only differences between students with respect to the feedback email were the % correct and number of points received. All emails included the question selected for evaluation and the answer from the rubric. This enabled students to receive feedback on their performance and compare their answer to the answer in the rubric. The feedback form letter is included in Appendix H. In contrast to the timing of feedback under the accuracy condition, feedback to students in the completion contingency was given only at the end of each unit and indicated the number of points they received for completing their homework.

At the beginning of the semester and again at the beginning of all units using the Accuracy Contingency, students were informed that one question would be randomly selected for evaluation each day. In actuality, a set of questions had already been selected for evaluation in each unit as in Study 2. However, the full set of questions was always greater than the number of days in the unit (> 4) and often included at least 2 questions per day. When there were at least two questions in a day, answers to both questions were evaluated, but only one question was selected as the basis for student grades on that day. My reason for having two questions available on some days was to generate more usable data for that day, which would provide a broader base for assessing student homework. However, some days had only one question that met the criteria necessary for evaluation (i.e., number of concepts identified and exam item correspondence). All students received a grade based on the same question across all sections applying the Accuracy Contingency in that unit. The selected question featured

the highest number of clearly discrete, factual concepts to limit variability in interpretation of student answers by the raters and generally produced the highest overall scores for students (thus students received the more favorable scenario for their grades). Though students received a grade based on only one of the two questions each day, scores for both questions were recorded in the database for later analysis.

Evaluation of student questions was similar to the methodology outlined in Study 2 with one notable exception: the percent of concepts correctly identified was divided by the number of possible concepts and then rounded to the nearest 10%. This percentage was included in the feedback to students. In the interest of providing more precise analysis, the ratio of correct concepts to the number of possible concepts was also recorded in the database (values could range from 0 to 1.00), along with the raw number of concepts the student identified. These data were available only for students participating in the homework contingency conditions (i.e., homework data were not available for students in the control/comparison group).

Evaluation of length of student homework responses. The length of student homework answers was calculated by performing a line count on answers for approximately 10 selected questions per unit (Unit C included 11 questions) for all students participating in the homework contingency conditions (i.e., no homework data were available for students in the control group). A line constituted any words or word fragments. A trained graduate assistant counted the number of lines for each of the selected questions for each student and recorded the result on a record sheet. These data were then transferred to the SPSS database.

Establishment of GPA and critical thinking groups. Using criteria previously established by Williams and Stockdale (2003), I identified three levels of critical thinking according to percentiles for college graduates in the WGCTA-S manual (1996): low critical thinking scores <= 24 (less than the 5^{th} percentile), middle critical thinking scores between 24 and 34 (between 5^{th} and 75^{th} percentiles), and high critical thinking scores >= 34 (greater than the 75^{th} percentile). To evaluate differences in GPA levels, the sample was divided into high, GPA >= 3.00, and low, GPA < 3.00.

Analysis and Results

Effect of contingencies on homework. Though accuracy of homework answers was the primary target of the contingencies, I also expected students to produce more complete answers, resulting in greater length of homework answers. Therefore, analysis of the effect of the homework contingencies on homework completion included two dimensions: accuracy and length of answers.

***Effect of contingencies on homework accuracy*.** The first hypothesis stated that setting a credit contingency for accurate completion of homework would increase the accuracy of student answers. The mean accuracy score for each day was computed across all students participating in the contingency conditions and, if more than one question was available, across both questions. The mean accuracy score for each day was compared between the Accuracy and the Completion Contingencies across all sections participating in the contingency conditions. The results indicated significant differences between the effects of the contingency conditions across all units for homework accuracy, $t(308) = 9.539$, $p < 0.001$, with students under the Accuracy Contingency scoring significantly higher on homework accuracy than students under the Completion

Contingency (M_{Acc} = 0.569, M_{Comp} = 0.395). Changes in homework accuracy scores between contingencies across all participating sections (i.e., 3:40, 5:05, and 6:30 TR sections) separately and combined are shown in Figures 6 through 9. See Table 13 for means between homework contingency groups. The means in Table 13 included all participants in the contingency sample, as opposed to only participants for whom all data were available.

To further evaluate the effectiveness of the conditions on improving homework accuracy, I compared homework accuracy within the framework of the Percent Exceeding the Median (PEM) method (Ma, 2006). According to Ma, this method of analysis is less sensitive to outliers than the percentage of non-overlapping data method (Scruggs, Mastropieri, & Casto, 1987) and can be useful in evaluating the effectiveness of treatment conditions in reversal designs. I used established conventions for evaluating effectiveness of treatment conditions, in which a treatment is considered "highly effective" when PEM is 90% or higher, "moderately effectively" when PEM is 70% to 89%, "mildly effective" when PEM is 50% to 69%, and "ineffective" when PEM is below 50% (Ma). The Accuracy Contingency was moderately effective (75% to 88% PEM) in promoting accuracy above the median of student accuracy scores in the Completion Contingency. Median lines used for evaluation of PEM are displayed in Figures 7 through 9.

Effect of contingencies on homework length. It was anticipated that students would also produce longer answers in the Accuracy condition, as they sought to include all relevant concepts in their answers. The mean length score for each day was computed across all students participating in contingency sections. When multiple questions were

included in a single day, the average length was calculated for that day. The mean length score for each day was compared for all sections participating in the Accuracy and Completion Contingencies. The results indicated significant differences between conditions across all units for homework length, $t(287.96) = 6.974$, $p < 0.001$, with students in the Accuracy Contingency scoring significantly higher on homework length ($M_{\text{Acc}} = 5.25$, $M_{\text{Comp}} = 3.63$). Changes in homework answer length between contingencies across all participating sections (i.e., 3:40, 5:05, and 6:30 sections) combined and separately are shown in Figures 10 through 13**.** See Table 13 for means between homework contingency groups. The means in Table 13 included all participants in the sample, as opposed to only participants for whom all data were available.

To evaluate the effectiveness of the conditions on increasing length of homework answers, I assessed the unit median scores for each section using PEM. In comparison to the Completion Contingency, the Accuracy Contingency appeared to generally be highly effective, as average PEM for each section of the course ranged from 83% to 100%. Median lines used for evaluation of PEM are displayed in Figures 11 through 13. Thus, the contingencies significantly influenced the length and accuracy of student homework answers, with students in the Accuracy Contingency providing longer and more accurate answers.

Effect of contingencies on exam performance. The second hypothesis stated that improving homework accuracy would indirectly improve exam scores. Significant differences on exam scores have historically indicated a clear pattern of changes in exam scores from one unit to the next. An analysis was conducted to compare the mean exam scores in the current study to those in Study 1 to determine if they followed the same

pattern of change between units. The analysis included students in the treatment group (those participating in the homework contingencies), as well as students in a non-participating, control group. The treatment group can be further separated into two sets: students transitioning from the Accuracy Contingency in one unit to the Completion Contingency in the following unit, and the reverse.

Mean exam scores also were compared across three categories (Accuracy Contingency, Completion Contingency, and Control Group) (see Table 13 for means across categories). ANOVA results indicated no significant differences in exam scores across the three contingency conditions, $F(2, 1141) = 0.474$, $p = 0.623$. However, students who scored high on the exams (scores of A or B) demonstrated significantly higher homework accuracy, $F(1, 211) = 5.215$, $p = 0.023$, and homework length, $F(1, 212) = 7.659$, $p = 0.006$, than students who scored low on exams (scores of D or *F*). See Table 14 for more information on the differences between high-performing and low-performing exam groups.

The categories (Accuracy, Completion, and Control Group) were then compared on the mean change in exam score across units using a factorial ANOVA. (See Table 15 and Figure 14 for mean change in exam scores across units.) The results indicated significant differences in the mean change in exam score across units, $F(3, 894) = 52.066$, $p < 0.001$, and contingency conditions, $F(2, 894) = 6.198$, $p = 0.002$. However, there were no significant interaction effects between contingency conditions and units, $F(6, 894) = 0.383$, $p = 0.890$. Pairwise comparisons (see Table 13) showed that students transitioning to the Accuracy Contingency exhibited significantly greater improvement in exam scores than students transitioning to the Completion Contingency and students in

the Control group ($p = 0.012$ and $p = 0.002$ respectively). Students in the Completion Contingency were statistically indistinguishable from students in the Control group ($p = 1.000$). When students transitioned from the Completion to the Accuracy Contingency, they demonstrated a mean change in exam score of +1.851, while students exhibited a mean change in exam score of -0.206 when they transitioned from the Accuracy Contingency to the Completion Contingency. Students in the Control Group showed a mean change in exam score of -0.033 as they progressed through units. These means differ slightly from those in Table 13, which included all participants in the sample, as opposed to only participants having complete data.

The results indicated that the mean raw scores of students in the transition from the Accuracy to the Completion Contingency and students in the Control Group followed the historical pattern of exam scores. However, when students transitioned from the Completion to the Accuracy Contingency, their mean raw scores did not follow the historical pattern. In the Accuracy to Completion transition in the Contingency Groups, as well as in the Control Group, mean exam scores decreased from Unit A to B and from Unit C to D, increased from Unit B to C, and showed almost no change from Unit D to E. In the Completion to Accuracy transition in the Contingency Groups, mean exam scores showed almost no change from Unit A to B and from Unit C to D, and increased from Unit B to C and from Unit D to E. In other words, unit transitions that produced decreases in exam means in the other two groups instead sustained exam scores at the same level in the Completion to Accuracy transition. See Table 16 for more information on the mean change in exam scores between groups and units.

Relationship of grade level, critical thinking, and GPA to targeted dependent variables. It is possible that variables, such as grade level, critical thinking, and GPA, which were predictive of exam scores in Study 2, could also have affected students' responses to the dependent variables (i.e., homework accuracy, homework length, and exam scores) in Study 3. In that case, the impact of the homework contingencies on the dependent variables was examined in relationship to the impact of the designated predictor variables. To evaluate possible interaction effects between the predictor variables (grade level classification, critical thinking level, and GPA level) on the dependent variables, a factorial MANOVA was first used to determine if any or all of the predictor variables were related to homework accuracy, homework length, exam performance, and change in exam performance.

The results indicated an interaction effect for grade-level classification and GPA, Wilks' $\lambda(12, 426.26) = 2.099$, $p = 0.016$ (or Wilks' $\lambda(4, 85) = 4.924$, $p = 0.001$). Follow-up ANOVAs indicated significant interactions between year in school and GPA for exam performance, $F(1, 88) = 14.702$, $p < 0.001$. Pairwise comparisons showed that upper-classmen (juniors, seniors, and graduate students) scored significantly higher than lower-classmen (freshmen and sophomores) on exam scores in the low GPA group (M_{Exam} = 40.5 and 27.3 respectively), but there were no significant differences between upper- and lower-classmen in the high GPA group (M_{Exam} = 39.7 and 40.4 respectively). No other interaction effects were significant. See Tables 17 and 18 for comparisons of means between year in school and GPA respectively.

Inasmuch as critical thinking was not involved in the interaction effect, I examined the possibility of a main effect resulting from critical thinking. Indeed, a

significant main effect was found for critical thinking, Wilks' $\lambda(4, 85) = 5.596$, $p < 0.001$. Follow-up analyses indicated significant differences in the mean scores between critical thinking groups for homework accuracy, $F(1, 88) = 9.146$, $p = 0.003$, homework length, $F(1, 88) = 8.122$, $p = 0.005$, and exam performance, $F(1, 88) = 6.636$, $p = 0.012$. Pairwise comparisons indicated that students with high critical thinking scores (>= 34) achieved higher exam scores ($p < 0.001$) than students with low critical thinking scores, but lower homework accuracy ($p = 0.045$). Despite significant differences in the multivariate analysis, pairwise comparisons revealed no significant differences in the length of homework answers between critical thinking levels, $p = 0.069$. See Table 19 for comparisons of means across critical thinking levels.

Moderation of significant treatment effects. Because year in school, GPA, and critical thinking were implicated (by interaction or main effects) as possibly contributing to one or more of the dependent variables, the next step in the analysis was to include each of these variables with the contingency conditions to determine possible interactions between the treatment conditions and the predictor variables on one or more of the dependent measures.

A factorial ANOVA was conducted to determine if the homework contingencies affected student exam performance differently between combinations of the GPA levels and year in school (upper-classmen and lower-classmen). The results indicated no significant interaction effects between homework contingencies, GPA level, and year in school, $F(1, 122) = 0.016$, $p = 0.899$. A factorial MANOVA was conducted to evaluate possible interaction effects between homework contingencies and critical thinking on homework accuracy, homework length, exam scores, and change in exam scores. The

results indicated no significant interactions, Wilks' $\lambda(4, 111) = 0.781$, $p = 0.540$. These analyses indicate that none of the predictor variables moderated the effect of the contingency conditions on any of the dependent variables (i.e., homework accuracy, homework completion, exam scores, and improvement patterns in exam scores across units).

Discussion

Improvement of homework accuracy. The first major hypothesis related to whether homework accuracy and length could be improved via reward contingencies. The results broadly supported this possibility. More specifically, students increased both the accuracy and length of their homework answers in response to the Accuracy Contingency. PEM analysis indicated that the Accuracy Contingency was moderately more effective than the Completion Contingency in improving the accuracy of students' answers and highly more effective at increasing the length of students' answers. It may be that in their efforts to include as many concepts as possible within their answers, students produced substantially longer answers. However, these longer answers were not consistently more accurate than shorter answers.

Improvement of exam performance. The second hypothesis stated that improving student homework accuracy would indirectly improve exam performance. In examining the exam scores, I used data from both of the contingency conditions, as well as from the sections of the course not participating in the study (the Control group). Broad comparisons of the three groups on their exam scores revealed no significant differences. The exam scores of students participating in the contingency conditions were largely indistinguishable from those who were not participating. However, as noted

in the analysis, this type of evaluation is confounded by a consistent pattern of changes in exam scores between units. To account for this pattern of change in exam scores in the current study, I calculated and evaluated the change in exam scores between units.

Analyzing changes in exam scores showed that students in the Accuracy Contingency demonstrated a significantly larger (and positive) change in exam scores than students in the Completion Contingency and Control Group. Overall, students transitioning from the Completion to Accuracy Contingency showed an average increase in exam scores of nearly 2 points, whereas students transitioning in the reverse direction showed an average decrease in exam scores of about 0.2 points. Students in the Control group evidence a pattern of change in exam means similar to that in the transition from the Accuracy to the Completion Contingency.

Moderation effects. Additional analyses investigated possible interaction effects between previously identified predictor variables (GPA, critical thinking, and year in school) and the dependent variables (homework accuracy, homework length, exam scores, and change in exam scores). The results indicated both an interaction effect for GPA and year in school with respect to exam scores, and a main effect for critical thinking with respect to homework accuracy, homework length, and exam scores. Follow-up analyses showed that though these predictor variables impacted the dependent variables either through interaction or main effects, there were no significant interactions between the predictor variables and the homework contingencies with respect to effects on homework and exam performance. The homework contingencies appeared to be equally effective in improving students' homework accuracy, homework length, and exam performance across combinations of critical thinking, GPA, and year in school.

Limitations. The first limitation necessarily has to do with the sample used in the current study. The sample was a comparatively small convenience sample, which tempered the strength of the findings. Though exam data were available for the non-participating sections, homework data were not. Thus, collective analyses of interaction effects and main effects of the independent variables (critical thinking, GPA, and year in school) were restricted to the sections of the course participating in the homework contingency conditions.

Generalizability of the findings. Accompanying concerns with the sample size is a necessary caution regarding generalizability of the findings. While certainly many courses include homework and exams, not all exams are multiple-choice and even those that are multiple-choice may be constructed differently from those used in the current study. It is unknown if the quality of homework answers would affect performance on essay exams in the same manner they apparently affected performance on multiple-choice exams. Inasmuch as the homework assignments in the present study were brief written answers (rather than multiple-choice), the assignments might actually better predict performance on essay exams than on multiple-choice exams. Different courses likely emphasize critical thinking to a greater or lesser degree, which necessarily will influence the impact of critical thinking on both the homework assignments and exam performance. To the extent that there are similarities between the structure of the course used in the current study and other courses, the findings here may be applicable in other courses.

Presentation of the contingencies. The way in which the homework contingencies were presented was adjusted after the first unit of the study. It was

discovered early in the study that these contingencies needed to be presented to students in very explicit terms. Initially, students were to be presented with both homework contingencies at the beginning of each unit and were informed of which contingency was presently active. Students expressed confusion and uncertainty regarding the contingencies during Unit A. For this reason, the results pertaining to Unit A were interpreted with caution, and adjustments were made in the presentation of the homework contingencies in later units (i.e., students were presented with only the currently active contingency on a daily basis).

Differences between units. Analysis of patterns of change in homework accuracy and homework length also indicate that the units are not equally difficult. Specifically, students tended to produce homework answers with significantly lower accuracy scores during Unit A, though Unit A also differed from subsequent units in how the homework contingencies were presented to the students. Students also tended to produce homework answers in Unit D with greater accuracy than in Unit C. Similarly, students tended to produce longer answers in Unit D than in Units A and B. One possible explanation for this pattern is that during Unit C the students worked in groups and tended to share answers or rely more on their group members for learning material, whereas in Unit D students presumably worked independently again and needed to be more self-reliant. Perhaps students also acquired new strategies for completing their homework while in a group, or the effects of reinforcement for higher exam performance during Unit C carried over and motivated students to perform better in Unit D.

Summary of findings. Broadly speaking, the results support the hypotheses that reward contingencies can significantly affect the quality of student homework answers,

which in turn may significantly affect the pattern of student exam performance. Interestingly, the effect of the contingencies on student homework answers and subsequent exam performance appeared to be largely unaffected by previously identified predictor variables such as critical thinking, GPA, and year in school. This finding was unexpected inasmuch as previous studies have indicated that students engage in qualitatively different academic behaviors depending on their critical thinking ability (Williams & Stockdale, 2003). Students with low critical thinking ability who achieve high exam scores necessarily engage in more academically rigorous behaviors than students with low critical thinking who perform poorly. For example, high-performing low critical thinkers are much more accurate and diligent note takers than low-performing low critical thinkers. In the current study, the former students represent such a small sample that it would be difficult to statistically compare the quality of their homework with that of their low-performing, low critical thinking peers.

It is encouraging to find that the quality of students' homework answers can be impacted by credit contingencies. Furthermore, homework quality can indirectly impact the pattern of exam performance. There are limitations to these findings, however, that must be considered before generalizing the results more broadly. The sample sizes were small, limiting the depth of analysis possible (e.g., between different critical thinking and GPA levels), and the study was conducted within the context of a single course, limiting the generalizability of the findings to other types of courses. Both of these shortcomings could be addressed in future studies by expanding the sample population to other courses and grade levels (e.g., primary and secondary education).

Chapter VII

Dissertation Discussion

Broadly, this dissertation explored a collection of variables hypothesized to influence student performance on major multiple-choice exams. Exam scores are generally regarded as a valid measure of student success, because they often reflect student mastery of course concepts and are predictive of actual performance beyond the course. A variety of factors have been hypothesized to predict exam performance, including critical thinking, participation in class discussion, prior academic performance (GPA), pre-course vocabulary, pre-course knowledge, and homework completion. A wealth of research has been conducted in the targeted undergraduate course to evaluate many of these factors, the results of which informed the current series of studies.

The first study evaluated data collected over the last seven years. The analysis of these data demonstrated a pattern of exam performance across units within the course that suggests features inherent in the course structure and sequence. The difficulty or controversial nature of the material in some units, vocabulary demands, and cooperative learning activities likely influenced exam performance. Similarities in the pattern of exam performance in Study 3 suggested these same factors may have continued to influence exam performance, and thus were accounted for by comparing the pattern in exam performance between selected units with the historical pattern of scores between those units.

A notable finding from the retrospective analysis in Study 1 was that students with a combination of higher GPA and higher critical thinking ability were likely to have higher exam scores than students with low scores on both variables. However, being

high on just one dimension might be enough to maximize exam scores. That is, students with high GPAs, regardless of their critical thinking ability, were more likely to demonstrate high exam performance than students with low GPAs. Previous research has already provided an explanation for students with low critical thinking ability and high GPAs who achieved high exam scores; these students often demonstrated superior study habits (Williams & Stockdale, 2003).

Students with high critical thinking ability were also more likely than students with low critical thinking to achieve high exam performance, regardless of their GPAs. This pattern also suggests the possibility that a student may have a low GPA, despite high critical thinking and high exam performance. This possibility is likely explained by the fact that course grades often reflect a combination of exam scores and homework assignments. If students with high critical thinking ability do not devote a significant amount of time to their homework assignments, then they may have a relatively low GPA regardless of their exam performance. This particular combination was found in a small number of students (n = 4, 2% of the sample) in Study 3.

Results from Study 2 and Study 3 supported the hypothesis that homework performance relates to exam performance. The results from Study 2 showed that homework completion was significantly, but moderately related to exam performance. When compared with other, established predictors of exam performance (e.g., critical thinking and participation in class discussion), homework completion proved to be more potent at times, but less consistent than critical thinking, in predicting exam performance. However, homework performance (both completion and accuracy) was assumed to be more amenable to change than critical thinking, a possibility that Study 3 investigated.

Study 2 found that there was a significant, but modest correlation between homework completion and exam performance. Study 3, on the other hand, attempted to manipulate this relationship to improve exam performance. The Accuracy Contingency, in which student homework grades were based on the accuracy of their answers, proved moderately to highly effective at improving the accuracy and length of homework answers when compared to the Completion Contingency (in which homework grades were based on the number of questions answered).

Subsequently, changes in exam scores were evaluated between the treatment groups (Accuracy Contingency and Completion Contingency) and a non-participating Control Group. The exam scores in the Completion Contingency were more similar to those of the non-participating Control Group than they were to scores under the Accuracy Contingency. The Accuracy Contingency produced significantly higher changes in exam scores between units, effectively breaking a pattern of changes in exam scores between units observed over the past seven years (e.g., significant decreases in exam scores from Unit A to Unit B and from Unit C to Unit D).

Limitations

The nature of the sample for Study 3, a relatively small convenience sample, was a major impediment to the generalization of the results to other college classes. In the case of Study 1, the sample size was very large, but data were present only for a handful of variables and some variables of interest were not available, such as participation in class discussion, homework completion, and study skills. By continuing to collect information on variables of interest over a period of several years, researchers may extend the pattern of results reported in the current dissertation. Examples of future

research areas might include how study skills change during undergraduate education and how these changes may be different for students across critical thinking levels, genders, or ethnicity (a variable for which no data were available in the current dissertation). Such information could help educators better understand how best to help certain types of students succeed and possibly improve retention and matriculation.

In Studies 2 and 3, the sample was adequate for evaluation of variables that potentially predicted exam scores, but generalizability of the findings was limited because the sample was restricted to a specific course in an undergraduate program. Plus, students within a teacher education program may possess different characteristics (e.g., critical thinking, academic history) from students in other programs (e.g., biology, engineering, or medical school). It would be helpful to replicate this study across different fields of study and with students in primary or secondary education, which should be more representative of the general population (due to an absence of admissions standards) than are higher-education samples. Indeed, as students at the primary or secondary levels likely represent a wider range of critical thinking ability and academic behaviors, it may be helpful for educators in primary and secondary education to understand the relationship between homework assignments and exam performance to better help struggling students. Study 3 was conducted over the course of one semester. Replicating it over multiple semesters would help to increase the sample size and may provide more insight into how the contingencies affect students with different cognitive and behavioral characteristics.

Another potential limitation was the differences in the design of the homework and the exams. The structure of the exams and homework assignments was not

topographically the same; homework assignments were short-essay answers, while exams were multiple-choice. It would be informative to examine the relationship between homework completion and exams when the two are topographically similar. The potential effectiveness of multiple-choice homework to improve exam performance has already been demonstrated by Oliver and Williams (2005). However, it remains to be seen if matching the design of homework assignments to exams is more effective than dissimilar homework assignments and exams.

The greatest practical limitation was arguably the labor-intensive evaluation of homework accuracy and the process of providing daily feedback on homework accuracy. These procedures required a considerable amount of time and the studies would not have been possible without the aid of other researchers in evaluating inter-rater reliability of the accuracy assessment. It is thus unlikely that an instructor would wish to undertake an evaluation process like the one used here. Inasmuch as the exams were multiple-choice, and there may be some potential benefit to matching task topography between homework and exams, the use of multiple-choice homework assignments could greatly reduce the amount of time needed to evaluate homework accuracy. In fact, there is already some evidence that multiple-choice homework assignments (in the form of practice exams) can significantly improve student exam performance (Oliver & Williams, 2005).

Conclusion and Implications for Future Research

Students and instructors often search for explanations for poor performance on exams—instructors because they wish to know how they can help their students perform better and students because they wish to improve their grades. Methodically reviewing students' answers on missed exam items will usually reveal an incomplete or inaccurate

understanding of the course concept. However, that procedure does not appear to help students correct their study habits on future exams. In fact, I have found students likely to repeat the same types of mistakes on subsequent exams. Therefore, I can only assume that the true explanation for poor performance is more complex than difficulty with a particular exam format or the wording of specific items on the exam. The difficulties students experience likely run the full spectrum of variables identified thus far and possibly include others yet to be identified.

The collection of studies described in this dissertation supports the notion that both cognitive and behavioral characteristics set high-achieving students apart from low-achieving students. Cognitive characteristics, such as high critical thinking, can overcome some of the behavioral characteristics that undermine mastery of course concepts independently of participating in class discussion or completing homework assignments. However, such students may not develop effective academic behaviors that would serve them well in the future (e.g., time management or personal organization).

There is good evidence that instructors can influence the amount of time and effort students invest in their homework assignments, producing answers that are both longer and more accurate in most cases. However, high critical thinkers' answers show some tendency to be shorter than those of low critical thinkers. Even if students have failed to develop effective academic behaviors, it may be possible to improve them through reward contingencies. Furthermore, these improvements appear to significantly improve performance on exams over highly difficult content. It would be helpful to replicate this study with larger samples to permit more detailed analysis of the effects of homework on exam performance for students with different combinations of cognitive

and behavioral characteristics (e.g., critical thinking and GPA). It is also unknown if these findings generalize across subject areas or into primary and secondary education. While I suspect that the effect may be even more pronounced in primary and secondary education, where there is a greater degree of variation in student characteristics, this speculation remains to be verified. Undoubtedly, if effective academic behaviors are beneficial to an individual in the long-term, improving them early in a student's educational career would be ideal.

References

Allegretti, C.L., & Frederick, J.N. (1995). A model for thinking critically about ethical issues. *Teaching of Psychology, 22,* 46–48.

Becker, W. E., & Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *The Economic Record, 75*, 348 – 357.

Bender, D. S. (1997). *Effects of Study Skills Programs on the Academic Behaviors of College Students.* Paper presented at the Annual Meeting of the Eastern Educational Research Association.

Brabeck, M. M. (1983). Critical thinking skills and reflective judgment development: Redefining the aims of higher education. *Journal of Applied Developmental Psychology, 4,* 23-34.

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement, 29,* 253 – 271.

Browne, M. N., Haas, P. F., Vogt, K. E., & West, J. S. (1977). Design and implementation of an evaluation procedure for an innovative undergraduate program. *College Student Journal, 11,* 1-10.

Bryan, T., Burstein, K., & Bryan, J. (2001). Students with learning disabilities: Homework problems and promising practices. *Educational Psychologist, 36*, 167-180.

Bursuck, W. D., Harniss, M. K., Epstein, M. H., Polloway, E. A., Jayanthi, M., & Wissinger, L. M. (1999). Solving communication problems about homework: Recommendations of special education teachers. *Learning Disabilities Research and Practice, 14*, 149-159.

Carroll, E., Williams, R. L., & Hautau, B. (2006). Cooperative learning contingencies: Unrelated versus related individual and group contingencies. Journal of Behavioral Education, 15, 191 – 202.

Cohen, J. (1989). *Statistical power analysis for the behavioural sciences. 2nd Ed.* Hillsdale, NJ: Erlbaum.

Cooper, H. (2001). Homework for all – in moderation. *Educational Leadership, 58,* 34-38.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research 1987-2003. *Review of Educational Research, 76*, 1-62.

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, *18,* 88-108.

Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology, 101,* 948 – 963.

Demoulin, D. F., & Walsh, R. J. (2002). Comparing the personal development test scores against GPA, grade, and age of typical high school students. *Journal of Instructional Psychology*, *29*, 22 – 24.

El Hassan, K., & Madhum, G. (2007). Validating the Watson Glaser critical thinking appraisal. *Higher Education, 54*, 361-383.

Epstein, J. L., & Van Voorhis, F. L. (2001). More than minutes: Teachers' roles in designing homework. *Educational Psychologist*, *36*, 181-194.

Everson, H. T., Weinstein, C. E., & Laitusis, V. (2000). *Strategic learning abilities as a predictor of academic achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Facione, P. A., & Facione, N. C. (1994). *The California Critical Thinking Skills Test: Test manual*. Millbrae, CA: The California Academic Press.

Foster, L. N., Krohn, K. R., McCleary, D. F., Aspiranti, K. B., Nalls, M. L., Quillivan, C. Q., Taylor, C. M., & Williams, R. L. (2009). Increasing low-responding students' participation in class discussion. *Journal of Behavioral Education, 18*, 173 – 188.

Foster, L. N., McCleary, D. F., & Williams, R. L. (2010, May). *Reliability and predictive validity of a new qualitative measure of participating in class discussion.* Paper presented at the meeting of the Association of Behavior Analysis International, San Antonio, TX.

Gajria, M., & Salend, S. J. (1995). Homework practices of students with and without learning disabilities: A comparison. *Journal of Learning Disabilities, 28*, 291-297.

Galyon, C. E., Blondin, C. A., Forbes, B. E., & Williams, R. L. (in press). Does homework matter? A comparison of homework with established predictors of exam performance in large college classes. *Journal on Excellence in College Teaching*.

Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L., & Williams, R. L. (2012). The relationship of academic self-efficacy to academic engagement and performance. *Social Psychology of Education.*

Harris, R. B., & Kerby, W. C. (1997). Statewide performance assessment as a complement to multiple-choice testing in high school economics. *Journal of Economic Education*, *28*, 122 – 134.

Hassmen, P., & Hunt, D. P. (1994). Human self-assessment in multiple-choice testing. *Journal of Educational Measurement, 31*, 149 – 160.

Harton, H. C., Richardson, D. S., Barreras, R. E., Rockloff, M. J., & Latane, B. (2002). Focused interactive learning: A tool for active class discussion. *Teaching of Psychology, 29,* 10 – 15.

Hautau, B., Turner, H. C., Carroll, E., Jaspers, K., Krohn, K., Parker, M., & Williams, R. L. (2006). Differential daily writing conditions and performance on major multiple-choice exams. *Journal of Behavioral Education, 15*(3), 170-180.

Hildebrandt, S., & Lucas, J. A. (1980). *Follow-up of students who majored and are majoring in legal technology.* Research report, William Raney Harper College, Pallatine, IL.

Holtzman, M. (2008). Demystifying application-based multiple-choice questions. *College Teaching*, *56*(2), 114-120.

Hurov, J. T. (1987). *A study of the relationship between reading, computational, and critical thinking skills and academic success in fundamentals of chemistry*. (ERIC Document Reproduction Service No. ED 286569).

King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13 (2),* 167-186.

Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences, 15,* 213-222.

Kraemer, H. C., & Theimann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

Krohn, K. R., Asiranti, K. B., Foster, L. N., McCleary, D. *F*., Taylor, C. M., Nalls, M. L., Quillivan, C. C., & Williams, R. L. (2010). Effects of self-recording and contingent credit on balancing participation across students. *Journal of Behavioral Education, 19,* 134-155.

Krohn, K. R., Parker, M. R., Foster, L. N., Aspiranti, K. B., McCleary, D. *F*., & Williams, R. L. (2008). Effects of writing-related contingencies on both quality of writing and multiple-choice exam performance in large college courses. *Behavior Analyst Today, 9*(3, 4), 184-195.

Lawson, T.J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology, 26,* 207–209.

Ma, H. (2006). An alternative method for quantitative synthesis of single-subject researches. *Behavior Modification, 30*, 598 – 617.

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education, 26,* 3-27.

Mines, R. A., King, P. M., Hood, A. B., & Wood, P. K. (1990). Stages of intellectual development and associated critical thinking skills in college students. *Journal of College Student Development, 31,* 538-547.

Oliver, R., & Williams, R. L. (2005). Direct and indirect effects of completion versus accuracy contingencies on practice-exams and actual-exam performance. *Journal of Behavioral Education, 14*, 141 – 152.

Parmenter, D. A. (2009). Essay versus multiple-choice: Student preferences and the underlying rationale with implications for test construction. *Academy of Educational Leadership Journal, 13,* 57 – 71.

Pizzolato, J. E., Hicklen, S. T., Brown, E. L., & Chaudhari, P. (2009). Student development, student learning: Examining the relation between epistemologic development and learning. *Journal of College Student Development, 50,* 475 – 490.

Polloway, E. A., Epstein, M. H., & Foley, R. (1992). A comparison of the homework problems of students with learning disabilities and non-handicapped students. *Learning Disabilities: Research and Practice, 7*, 203-209.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. *Annals of Child Development*, *4*, 89-129.

Salend, S. J., & Gajria, M. (1995). Increasing the homework completion rates of students with mild disabilities. *Remedial & Special Education, 16*, 271-279.

Scruggs, T. E. , Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, *8*(*2*), 24-33.

Sivan, E. (1986). Motivation in social constructivist theory. *Educational Psychologist*, *21*, 209-233.

Soper, D. S. (2011) "The Free Statistics Calculators Website", *Online Software*, http://www.danielsoper.com/statcalc/.

Stepp, S. L., Schrock, S. A., & Coscarelli, W. C. (1996). Can a multiple-choice exam discriminate between masters and nonmasters in instructional design? *Performance Improvement Quarterly, 9*(1), 23-35.

Steward, R. J., & Al Abdulla, Y. (1989). *An examination of the relationship between critical thinking and academic success on a university campus*. (Eric Document Reproduction Service No. ED 318936).

Stockdale, S. L., & Williams, R. L., (2004). Cooperative learning groups at the college level: Differential effects on high, average, and low exam performers. *Journal of Behavioral Education,* 13(1), 37-50.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education. *Assessment and Evaluation in Higher Education, 30,* 325-341.

Tsui, L. (1998, November). *A review of research on critical thinking.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Miami, FL.

Turner, H. C., Bliss, S. L., Hautau, B., Carroll, E., Jaspers, K. E., & Williams, R. L. (2006). Brief daily writing activities and performance on major multiple-choice exams. *Journal of General Education, 55,* 221 – 246.

Turner, H. C., & Williams, R. L. (2007). Vocabulary development and performance on multiple-choice exams in large entry-level courses. *Journal of College Reading and Learning, 37*(2), 64 – 81.

Walker, D. M., & Thompson, J. S. (2001). A note on multiple choice exams, with respect to students' risk preference and confidence. *Assessment & Evaluation in Higher Education, 26*, 261 – 267.

Wallace, M. A., & Williams, R. L. (2003). Multiple-choice exams: Explanations for student choices. *Teaching of Psychology, 30,* 136-138.

Watson, G. B., & Glaser, E. M. (1994). *Watson-Glaser critical thinking appraisal Form S manual.* San Antonio, TX: Harcourt Brace.

Weaver, R. R., & Qi, J. (2005). Classroom organization and participation: College students' perceptions. *The Journal of Higher Education, 76,* 570-601.

Williams, R. L., Carroll, E., & Hautau, B., (2005). Individual accountability in cooperative learning groups at the college level: Differential effects on high, average, and low exam performers. *Journal of Behavioral Education, 14(3)*, 167 – 188.

Williams, R. L., & Clark, L. (2004). College students' ratings of student effort, student ability and teacher input as correlates of student performance on multiple-choice exams. *Educational Research, 46(3)*, 229 – 239.

Williams, R. L., & Eggert, A. C. (2002). Notetaking in college classes: Student patterns and instructional strategies. (2002). *Journal of General Education, 51,* 173-199.

Williams, R. L., Oliver, R., Allin, J. L., Winn, B., & Booher, C. S., (2003). Psychological critical thinking as a course predictor and outcome variable. *Teaching of Psychology*, *30(3)*, 220 – 223.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, *53*, 37-58.

Williams, R.L., & Stockdale, S. (2003). High performing students with low critical thinking skills. *Journal of General Education, 52,* 200-226.

Williams, R. L., & Worth, S. L. (2002). Thinking skills and work habits: Contributors to course performance. *The Journal of General Education*, *51(3)*, 200-227.

Wilson, D. G., & Wagner, E. E. (1981). The Watson-Glaser Critical Thinking Appraisal as predictor of performance in a critical thinking course. *Educational and Psychological Measurement, 41*, 1319 – 1322.

Worth, S. L. (2000). *Critical thinking ability as a predictor of success in a large undergraduate course.* Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Worth, S. L., & Williams, R. L. (2001). The relationship of critical thinking to success in college. *Inquiry: Critical Thinking across the Disciplines, 21,* 5-16.

Young, J. R. (2002). Homework? What homework? *Chronicle of Higher Education, 49*(15), A35-A37.

Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement: Theory, research, and practice.* New York: Springer-Verlag.

Appendices

Appendix A: Tables

Table 1

Study 1 Demographic Information from 2004-2011

Demographic variable	*n*	Percent
Female	2402	66.9%
Male	892	24.8%
Not reported	297	8.3%
Freshman	221	6.2%
Sophomore	1245	34.7%
Junior	841	23.4%
Senior	367	10.2%
Graduate	205	5.7%
Other	22	0.6%
Not reported	688	19.2%

Table 2

Study 1 Comparison of Mean Exam Scores between Units from 2004 to 2011

Unit	*n*	Mean	*SD*
A	3197	39.42_{a}	5.436
B	3208	37.68_{b}	6.695
C	3104	40.51_{c}	5.225
D	3080	38.91_{d}	6.062
E	3033	39.24_{ad}	6.250

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 3

Study 1 Comparison of Mean Critical Thinking Scores between Grade Levels

Year in school	*n*	Mean	*SD*
Freshman	740	24.70_{a}	5.864
Sophomore	4545	25.62_{ab}	5.767
Junior	2910	25.47_{ab}	6.004
Senior	1350	26.66_{bc}	5.972
Graduate	725	27.30_{c}	6.504

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 4

Study 1 Comparison of Mean Exam Scores between Grade Levels

Year in school	*n*	Mean	*SD*
Freshman	1036	38.63_a	5.939
Sophomore	5406	39.03_a	5.795
Junior	3550	38.86_a	5.943
Senior	1538	40.37_b	5.743
Graduate	910	42.16_c	5.491

Note. Different subscripts indicate significant differences at least at $p <$ 0.05 level.

Table 5

Study 2 Demographic Information

Demographic variable	*n*	Percent
Female	122	73.1%
Male	43	25.7%
Not reported	2	1.2%
Freshman	16	9.6%
Sophomore	84	50.3%
Junior	39	23.4%
Senior	18	10.8%
Graduate	6	3.6%
Other	1	.6%
Not reported	3	1.8%

Table 6

Study 2 Descriptive Statistics by Year in School

Grade level	GPA	Critical thinking	Exam scores	Participation	Homework accuracy
Freshman	*M* = 3.44	*M* = 27.27	*M* = 39.46	*M* = 5.86	*M* = 20.07
	SD = .51	*SD* = 5.01	*SD* = 4.47	*SD* = 3.47	*SD* = 4.64
	n = 15	*n* = 15	*n*= 16	*n* = 16	*n* = 16
Sophomore	*M* = 3.16	*M* = 26.98	*M* = 39.64	*M* = 5.37	*M* = 17.79
	SD = .46	*SD* = 6.19	*SD* = 4.29	*SD* = 2.74	*SD* = 4.58
	n = 82	*n* = 82	*n* = 84	*n* = 83	*n* = 80
Junior	*M* = 3.09	*M* = 26.28	*M* = 39.72	*M* = 4.62	*M* = 17.81
	SD = .46	*SD* = 6.60	*SD* = 4.40	*SD* = 2.94	*SD* = 3.84
	n = 38	*n* = 39	*n* = 39	*n* = 39	*n* = 39
Senior	*M* = 3.13	*M* = 25.06	*M* = 39.30	*M* = 4.88	*M* = 18.78
	SD = .40	*SD* = 7.17	*SD* = 3.66	*SD* = 3.35	*SD* = 4.05
	n = 18	*n* = 18	*n* = 18	*n* = 18	*n* = 18
Graduate	*M* = 3.69	*M* = 32.67	*M* = 45.03	*M* = 8.63	*M* = 21.90
	SD = .23	*SD* = 4.93	*SD* = 3.62	*SD* = 1.53	*SD* = 6.30
	n = 5	*n* = 6	*n* = 6	*n* = 6	*n* = 6

Note. Students in the “Other” and “Not reported” categories were excluded from this chart due to absence of data.

Table 7

Study 2 Comparison of Predictor Variable Means by Year in School

Year in school	Critical thinking	Homework accuracy	Participation	Exam scores
Freshman	*M* = 27.27,	*M* = 20.07,	$M = 5.86_{ab}$,	$M = 39.46_{ab}$,
	SD = 5.01	*SD* = 4.64	*SD* = 3.47	*SD* = 4.47
	n = 15	*n* = 16	*n* = 16	*n* = 16
Sophomore	*M* = 26.98,	*M* = 17.79,	$M = 5.37_{ab}$,	$M = 39.64_{a}$,
	SD = 6.19	*SD* = 4.58	*SD* = 2.74	*SD* = 4.29
	n = 82	*n* = 80	*n* = 83	*n* = 84
Junior	*M* = 26.28,	*M* = 17.81,	$M = 4.62_{a}$,	$M = 39.72_{a}$,
	SD = 6.60	*SD* = 3.84	*SD* = 2.94	*SD* = 4.40
	n = 39	*n* = 39	*n* = 39	*n* = 39
Senior	*M* = 25.06,	*M* = 18.78,	$M = 4.88_{ab}$,	$M = 39.30_{a}$,
	SD = 7.17	*SD* = 4.05	*SD* = 3.35	*SD* = 3.67
	n = 18	*n* = 18	*n* = 18	*n* = 18
Graduate	*M* = 32.67,	*M* = 21.90,	$M = 8.63_{b}$,	$M = 45.03_{b}$,
	SD = 4.93	*SD* = 6.30	*SD* = 1.53	*SD* = 3.62
	n = 6	*n* = 6	*n* = 6	*n* = 6

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 8

Study 2 Descriptive Statistics for All Variables across All Units

Unit	Critical thinking	Homework accuracy	Participation	Exam scores
Unit A	*M* = 26.81	*M* = 14.58	*M* = 5.12	*M* = 39.54
	SD = 6.30	*SD* = 4.51	*SD* = 3.91	*SD* = 4.95
	n = 162	*n* = 141	*n* = 163	*n* = 166
Unit B	*M* = 26.81	*M* = 16.53	*M* = 5.57	*M* = 37.94
	SD = 6.30	*SD* = 5.88	*SD* = 3.47	*SD* = 6.91
	n = 162	*n* = 126	*n* = 165	*n* = 165
Unit C	*M* = 26.81	*M* = 21.27	*M* = 5.60	*M* = 40.27
	SD = 6.30	*SD* = 6.90	*SD* = 3.30	*SD* = 4.72
	n = 162	*n* = 142	*n* = 164	*n* = 166
Unit D	*M* = 26.81	*M* = 20.90	*M* = 5.17	*M* = 40.49
	SD = 6.30	*SD* = 5.95	*SD* = 3.41	*SD* = 5.89
	n = 162	*n* = 135	*n* = 162	*n* = 164
Unit E	*M* = 26.81	*M* = 19.39	*M* = 5.21	*M* = 41.08
	SD = 6.30	*SD* = 5.66	*SD* = 3.13	*SD* = 5.35
	n = 162	*n* = 132	*n* = 112	*n* = 164
Unit total	*M* = 26.81	*M* = 96.67	*M* = 27.30	*M* = 199.07
	SD = 6.30	*SD* = 22.08	*SD* = 14.86	*SD* = 22.42
	n = 162	*n* = 82	n = 109	n = 162

Table 9

Study 2 Correlations between Predictor Variables

Unit	Variables	Homework accuracy		Participation	
Unit A	Critical thinking	$r = .201,^{*} p = .018$	$n = 139$	$r = .175,^{*} p = .027$	$n = 160$
	Homework accuracy	-	-	$r = .185,^{*} p = .030$	$n = 138$
Unit B	Critical thinking	$r = .282,^{*} p = .002$	$n = 123$	$r = .176,^{*} p = .026$	$n = 161$
	Homework accuracy	-	-	$r = .137, p = .125$	$n = 126$
Unit C	Critical thinking	$r = .221,^{*} p = .009$	$n = 139$	$r = .196,^{*} p = .013$	$n = 160$
	Homework accuracy	-	-	$r = .185,^{*} p = .028$	$n = 141$
Unit D	Critical thinking	$r = .137, p = .118$	$n = 132$	$r = .125, p = .118$	$n = 158$
	Homework accuracy	-	-	$r = .137, p = .112$	$n = 135$
Unit E	Critical thinking	$r = .007, p = .937$	$n = 129$	$r = .139, p = .149$	$n = 109$
	Homework accuracy	-	-	$r = 0.025, p = 0.807$	$n = 102$
Total	Critical thinking	$r = .396,^{*} p < .001$	$n = 81$	$r = .118, p = .228$	$n = 107$
	Homework accuracy	-	-	$r = .182, p = .152$	$n = 63$

Note. * indicates significance at least at $p < 0.05$ level.

Table 10

Study 2 Partial Correlations between Exam Performance and Predictor Variables

Units	Critical thinking		Homework accuracy			Participation	*n*
Unit A	.277**	=	.339***	>[ct]	>[hw]	.046	136
Unit B	.278**	=	.303**	=	=	.174	123
Unit C	.200*	=	.331***	=	=	.195*	138
Unit D	.408***	>	.132	=	=	.261**	132
Unit E	.423***	>	.135	>	=	.205*	99
Cross-unit	.396**	=	.330**	=	=	.209	62

Note. [ct] = Critical thinking. [hw] = Homework. * indicates significance at $p < 0.05$ level. ** indicates significance at $p < 0.01$ level. *** indicates significance at $p < 0.001$ level.

Table 11

Study 2 Stepwise Regression Analysis of Predictor Variables and Exam Performance (n = 62 to 138)

Models	Unit A (*n* = 136)	Unit B (*n* = 123)	Unit C (*n* = 138)	Unit D (*n* = 132)	Unit E (*n* = 99)	Unit total (*n* = 62)
Model 1	HW β = .382	HW β = .371	HW β = .384	CT β = .432	CT β = .440	CT β = .509
Adj. r^2	.139	.131	.142	.181	.185	.246
F	22.857*	19.354*	23.588*	29.895*	23.245*	20.953*
Model 2	HW β = .327	HW β = .296	HW β = .338	CT β = .394	CT β = .414	CT β = .383
	CT β = .261	CT β = .269	CT β = .221	PT β = .238	PT β = .186	HW β = .311
Adj. r^2	.199	.191	.183	.231	.211	.317
F	11.014*	10.029*	7.816*	9.423*	4.206*	7.222*
Model 3			HW β = .317			
			CT β = .187			
			PT β = .181			
Adj. r^2			.208			
F			5.323*			

Note. "CT" = Critical thinking, "HW" = Homework accuracy, and "PT" = Participation. * indicates significance at $p < 0.05$ level.

Table 12

Study 3 Demographic Information

Demographic variable	*n*	Percent
Female	168	70.9%
Male	66	27.8%
Not reported	3	1.3%
Freshman	5	2.1%
Sophomore	112	47.3%
Junior	70	29.5%
Senior	33	13.9%
Graduate	8	3.4%
Other	2	.8%
Not reported	7	3.0%

Table 13

Study 3 Comparisons of Means between Homework Contingency Groups

Homework contingency	Exam scores	Change in exam scores	Homework accuracy[a]	Homework length[a]
Control group	$M = 38.32_a$	$M = -.035_a$	--	--
	$SD = 5.70$	$SD = 5.11$	--	--
	$n = 163$	$n = 163$	--	--
Completion	$M = 38.05_a$	$M = -.205_a$	$M = .395_a$	$M = 3.77_a$
	$SD = 5.78$	$SD = 5.71$	$SD = .148$	$SD = 1.51$
	$n = 36$	$n = 32$	$n = 33$	$n = 36$
Accuracy	$M = 39.40_a$	$M = 1.85_b$	$M = .567_b$	$M = 5.36_b$
	$SD = 5.39$	$SD = 5.39$	$SD = .129$	$SD = 1.87$
	$n = 30$	$n = 32$	$n = 29$	$n = 30$

Note. Means were computed for all participants in the sample and thus differ slightly from means computed during analyses that excluded participants for whom not all data were present. Different subscripts indicate significant differences at least at $p < 0.05$ level. [a] Homework accuracy and homework length data were not recorded for the Control Group.

Table 14

Study 3 Homework Completion Means between Exam Performance Levels

Exam performance	*n*	Homework accuracy	Homework length
High (A or B)	153	$.494_a$	4.88_a
Low (D or *F*)	60	$.430_b$	4.05_b

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 15

Study 3 Comparisons of Means between Units

Unit	Exam scores	Change in exam scores[a]	Homework accuracy[b]	Homework length[b]
Unit A	$M = 37.84_a$	--	$M = .361_a$	$M = 3.78_a$
	$SD = 5.44$	--	$SD = .104$	$SD = 1.57$
	$n = 236$	--	$n = 61$	$n = 71$
Unit B	$M = 35.40_b$	$M = -2.49_a$	$M = .512_{bc}$	$M = 4.33_{ab}$
	$SD = 7.05$	$SD = 5.80$	$SD = .175$	$SD = 1.81$
	$n = 233$	$n = 233$	$n = 63$	$n = 69$
Unit C	$M = 40.96_c$	$M = 5.41_b$	$M = .449_b$	$M = 4.68_{ab}$
	$SD = 4.95$	$SD = 5.85$	$SD = .192$	$SD = 2.04$
	$n = 228$	$n = 228$	$n = 62$	$n = 65$
Unit D	$M = 38.65_a$	$M = -2.45_a$	$M = .560_c$	$M = 4.99_b$
	$SD = 5.65$	$SD = 5.13$	$SD = .208$	$SD = 2.25$
	$n = 223$	$n = 222$	$n = 62$	$n = 63$
Unit E	$M = 38.98_a$	$M = .35_c$	$M = .494_{bc}$	$M = 4.73_b$
	$SD = 5.79$	$SD = 4.56$	$SD = .153$	$SD = 1.85$
	$n = 224$	$n = 223$	$n = 62$	$n = 62$
Unit total	$M = 38.35$	$M = .21$	$M = .476$	$M = 4.48$
	$SD = 6.09$	$SD = 6.26$	$SD = .182$	$SD = 1.94$
	$n = 1144$	$n = 906$	$n = 310$	$n = 330$

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

[a] Mean change in exam score is reported between the indicated unit and the unit preceding it, thus mean change in exam scores is not reported for Unit A. [b] Sample sizes for homework accuracy and homework length include only the Treatment Groups.

Table 16

Study 3 Mean Change in Exam Scores between Contingencies by Unit

Unit	Unit A to B	Unit B to C	Unit C to D	Unit D to E	Total
Control group	-3.000_a	5.454_a	-2.790_a	$.204_a$	$-.033_a$
Completion	-2.857_a	4.609_a	-2.184_a	$-.391_a$	$-.206_a$
Accuracy	$-.583_b$	7.000_b	$-.409_b$	1.395_b	1.851_b

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 17

Study 3 Comparisons of Means between Years in School

Unit	Exam scores[a]	Change in exam scores[a]	Homework accuracy[b]	Homework length[b]
Freshman	*M* =35.56	*M* = .550	*M* = .454	*M* = 5.53
	SD = 4.98	*SD* = 1.33	*SD* = NA	*SD* = NA
	n = 5	*n* = 5	*n* = 1	*n* = 1
Sophomore	*M* = 38.51	*M* = .263	*M* = .462	*M* = 4.30
	SD = 4.39	*SD* = 1.73	*SD* = .085	*SD* = 1.57
	n = 106	*n* = 112	*n* = 15	*n* = 15
Junior	*M* = 38.19	*M* = -.036	*M* = .490	*M* = 4.50
	SD = 4.92	*SD* = 2.21	*SD* = .098	*SD* = 1.47
	n = 65	*n* = 69	*n* = 25	*n* = 24
Senior	*M* = 39.37	*M* = -.065	*M* = .441	*M* = 4.98
	SD = 4.33	*SD* = 1.47	*SD* = .129	*SD* = 1.13
	n = 31	*n* = 31	*n* = 12	*n* = 11
Graduate	*M* = 42.70	*M* = .469	*M* = .524	*M* = 5.05
	SD = 3.60	*SD* = 1.00	*SD* = .081	*SD* = 1.11
	n = 8	*n* = 8	*n* = 5	*n* = 5
Total	*M* = 38.62	*M* = .139	*M* = .477	*M* = 4.62
	SD = 7.07	*SD* = 1.82	*SD* = .101	*SD* = 1.38
	n = 217	*n* = 227	*n* = 59	*n* = 57

Note. [a] Sample sizes for exam scores and change in exam scores include the Control Group and Treatment Groups. [b] Sample sizes for homework accuracy and homework length include only the Treatment Groups.

Table 18

Study 3 Comparisons of Means between GPA Levels

GPA	Exam scores	Change in exam scores	Homework accuracy	Homework length
High	$M = 39.30_a$	$M = .306_a$	$M = .506_a$	$M = 4.66_a$
	$SD = 4.43$	$SD = 1.40$	$SD = .085$	$SD = 1.34$
	$n = 148$	$n = 152$	$n = 33$	$n = 33$
Low	$M = 36.54_b$	$M = .034_a$	$M = .425_b$	$M = 4.31_a$
	$SD = 4.12$	$SD = 2.14$	$SD = .121$	$SD = 1.19$
	$n = 48$	$n = 52$	$n = 15$	$n = 13$

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Table 19

Study 3 Comparisons of Means between Critical Thinking Levels

Critical thinking	Exam scores	Change in exam scores	Homework accuracy	Homework length
High	M = 41.98_a	M = $.000_a$	M = $.434_a$	M = 4.16_a
	SD = 4.17	SD = 1.38	SD = .119	SD = 1.36
	n = 24	n = 25	n = 12	n = 11
Middle	M = 38.79_{ab}	M = $.203_a$	M = $.487_b$	M = 4.77_b
	SD = 4.55	SD = 1.40	SD = .103	SD = 1.33
	n = 113	n =116	n = 31	n = 30
Low	M = 37.20_b	M = $.141_a$	M = $.478_b$	M = 4.69_{ab}
	SD = 4.42	SD = 2.31	SD = .087	SD = 1.48
	n = 85	n = 92	n = 20	n = 19

Note. Different subscripts indicate significant differences at least at $p < 0.05$ level.

Appendix B: Figures

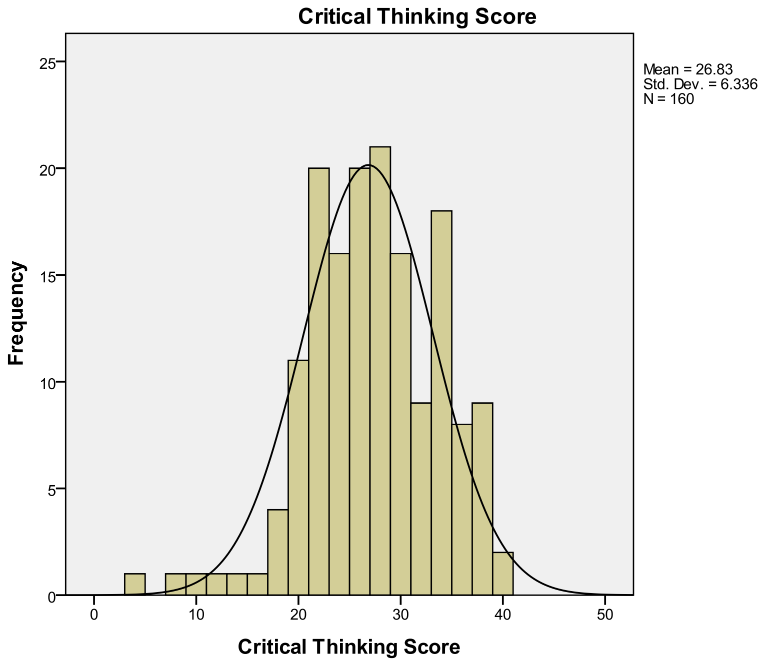


Figure 1. Study 2 Distribution of Critical Thinking Scores

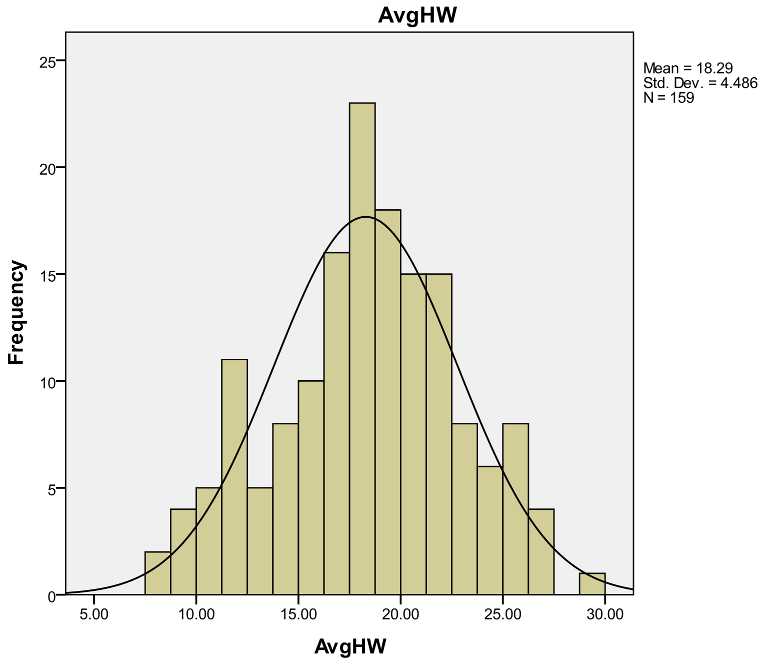


Figure 2. Study 2 Distribution of Homework Completion Scores

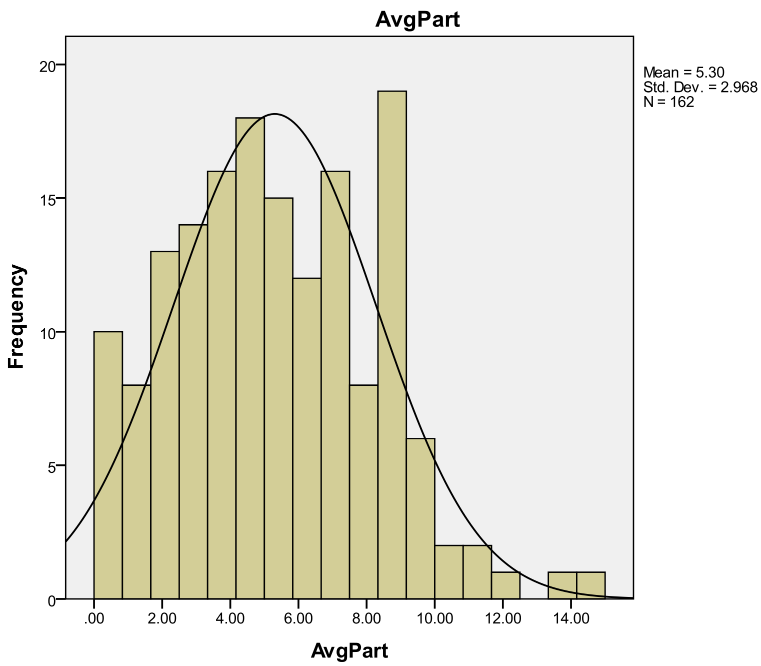


Figure 3. Study 2 Distribution of Participation Scores

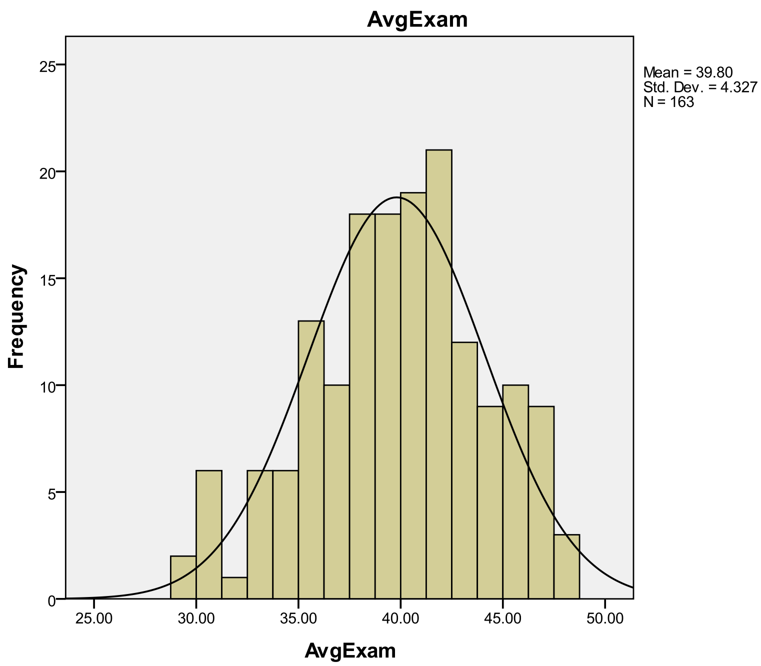


Figure 4. Study 2 Distribution of Exam Scores

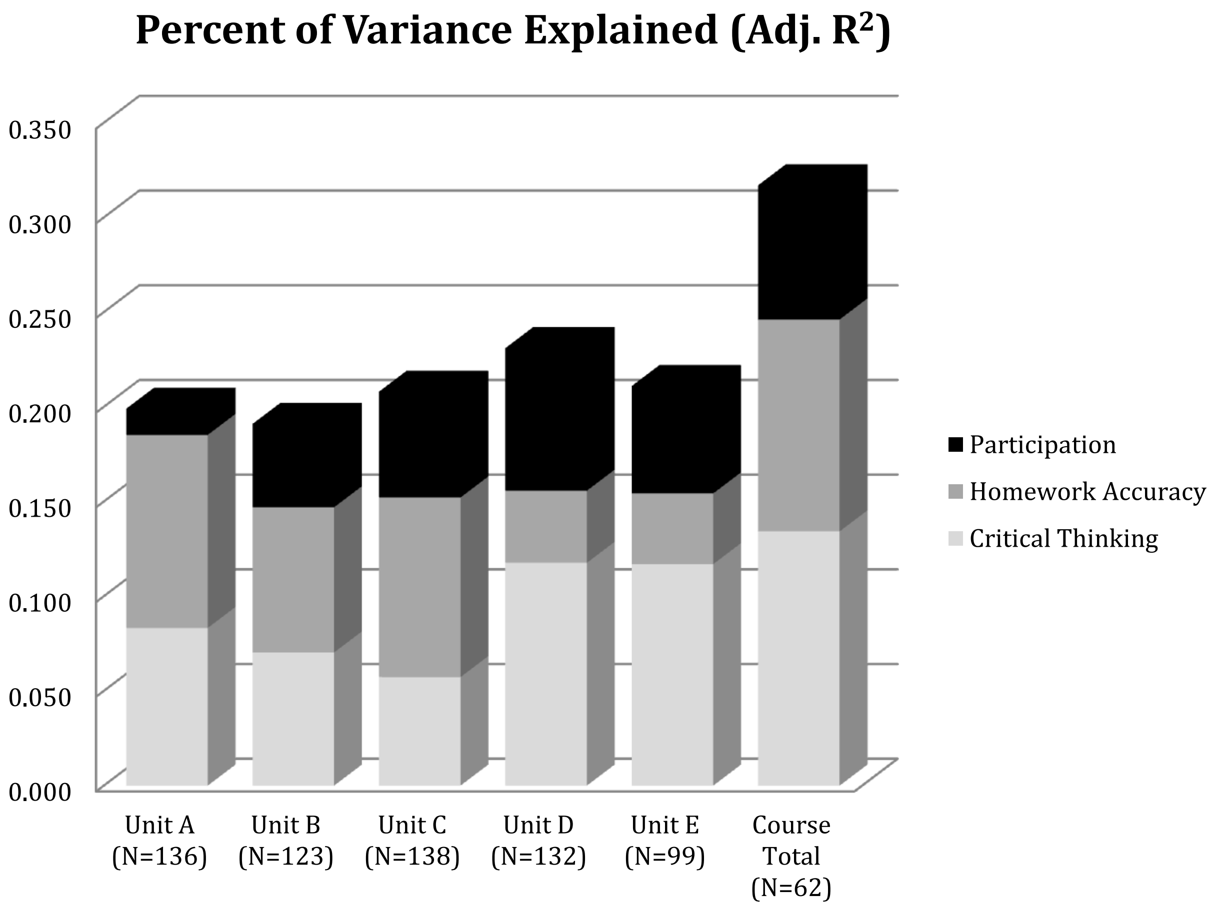


Figure 5. Study 2 Proportionate Percent of Variance Explained in Exam Scores

Note. The Y-axis represents as a decimal the percentage of variance in exam scores explained by the regression equation. The contribution of each variable is represented as a stacked bar.

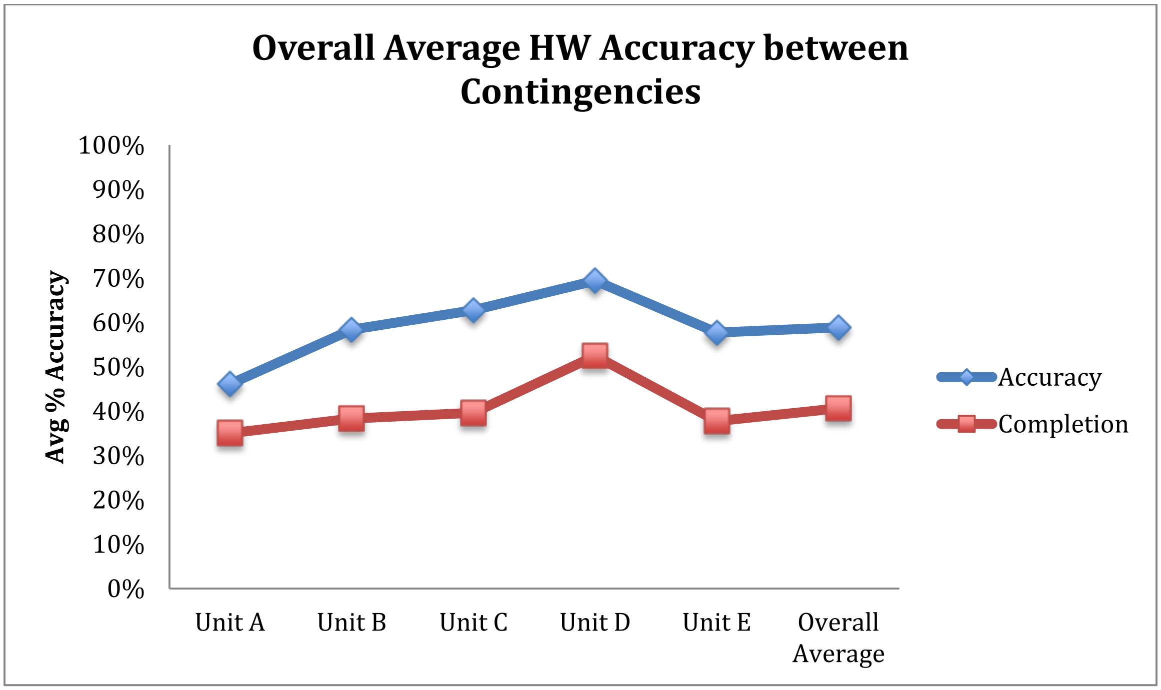


Figure 6. Study 3 Average Homework Accuracy between Homework Contingencies

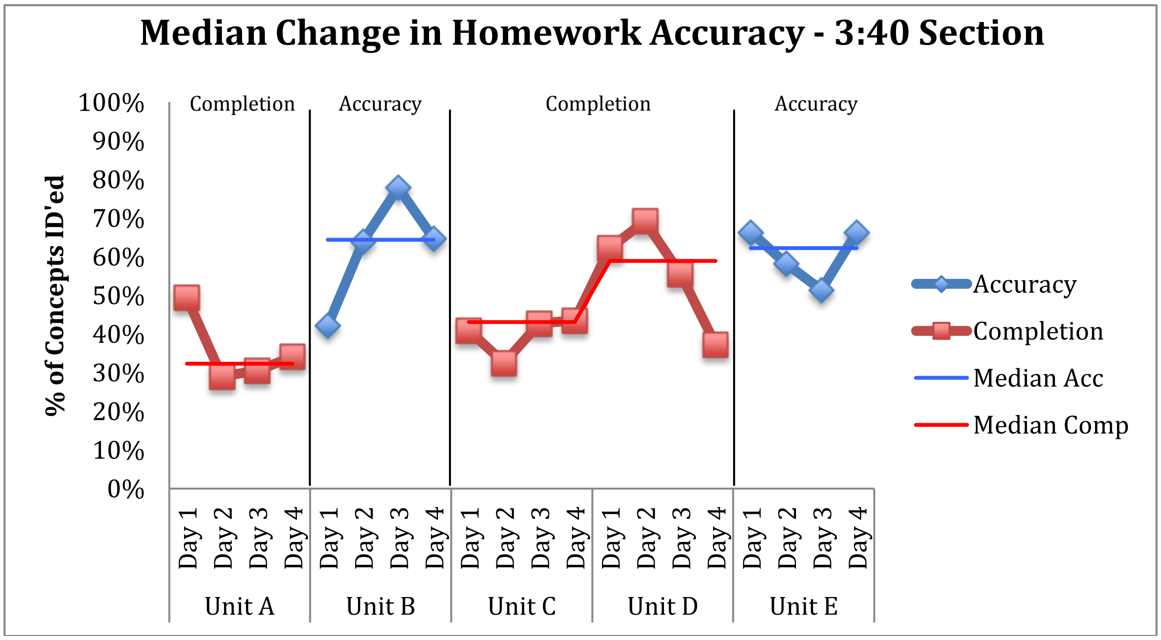


Figure 7. Study 3 Homework Accuracy between Homework Contingencies in 3:40 Section

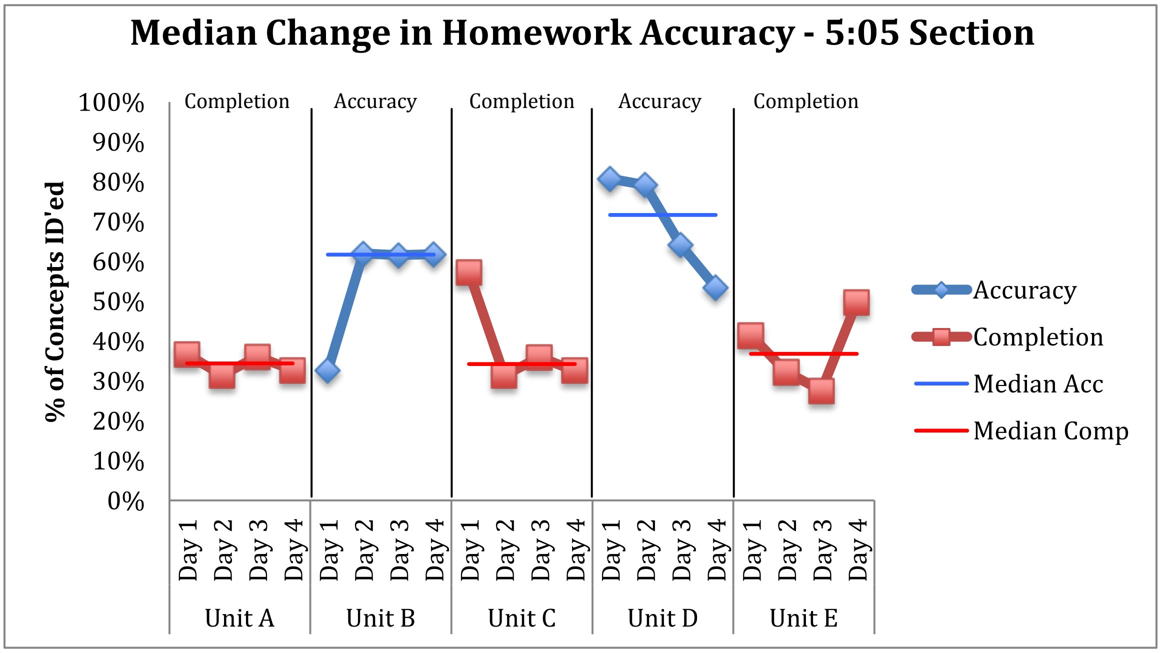


Figure 8. Study 3 Homework Accuracy between Homework Contingencies in 5:05 Section

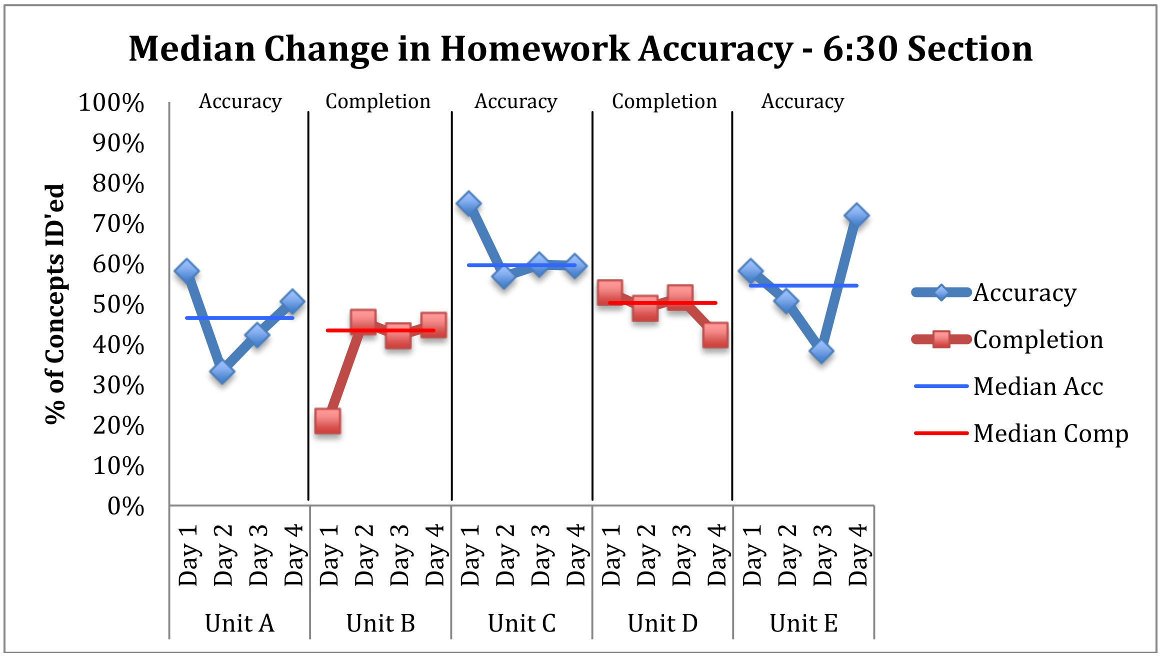


Figure 9. Study 3 Homework Accuracy between Homework Contingencies in 6:30 Section

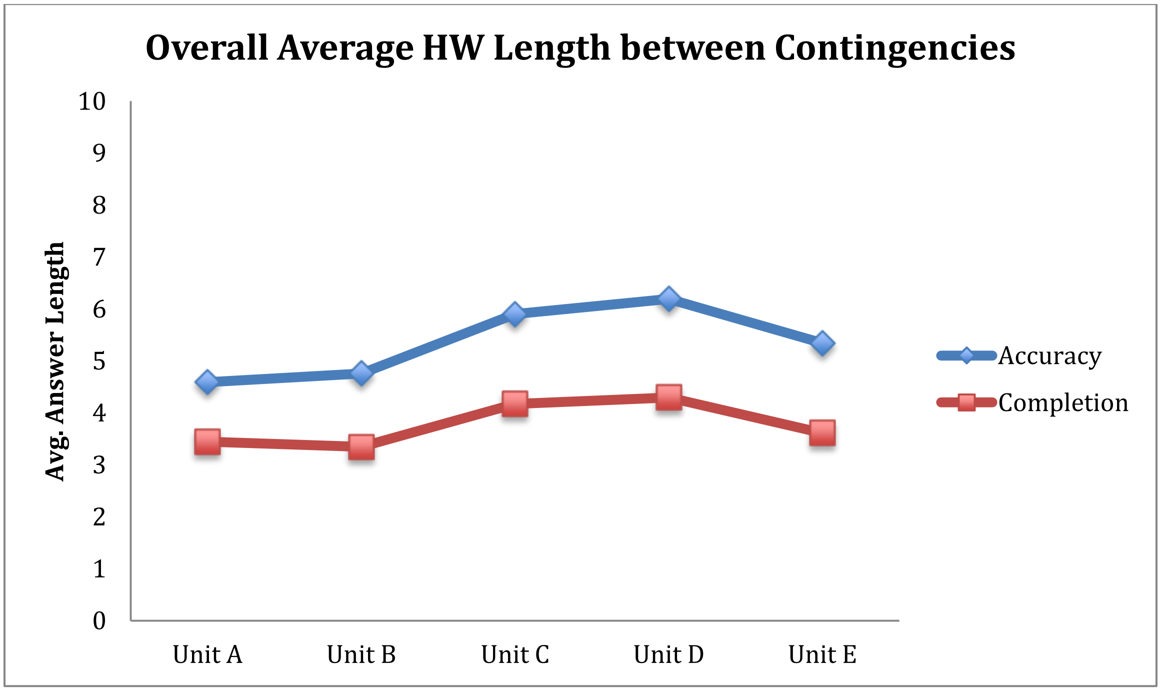


Figure 10. Study 3 Average Homework Length between Homework Contingencies

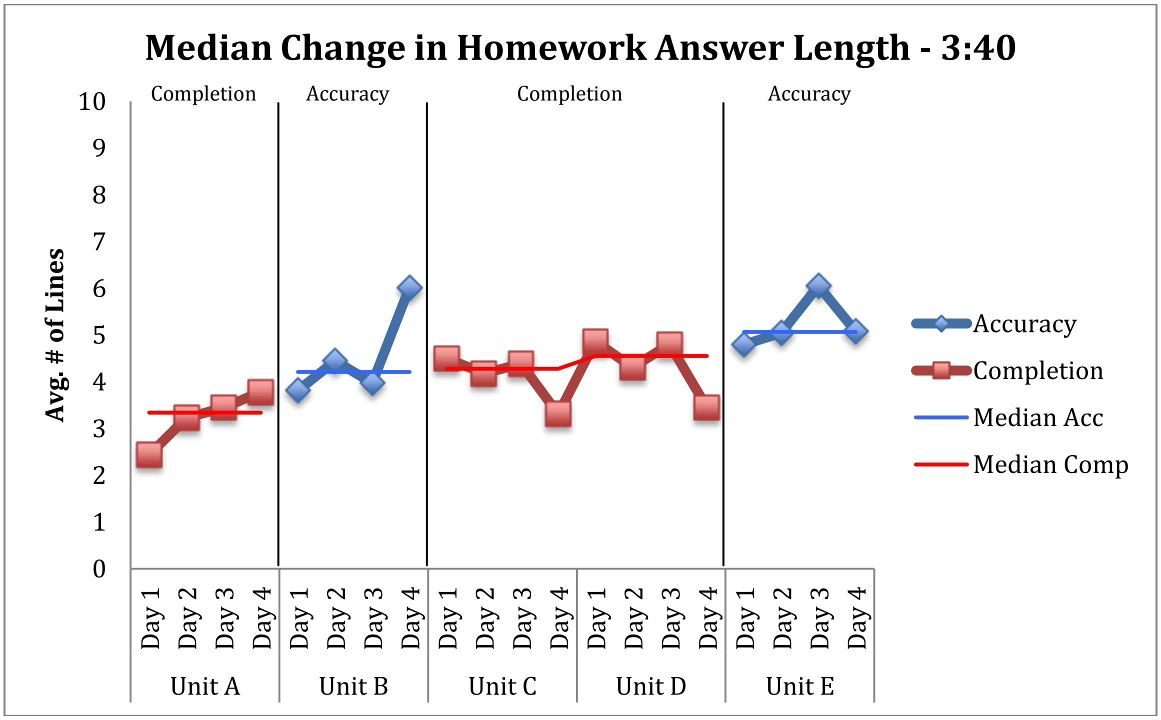


Figure 11. Study 3 Homework Length between Homework Contingencies in 3:40 Section

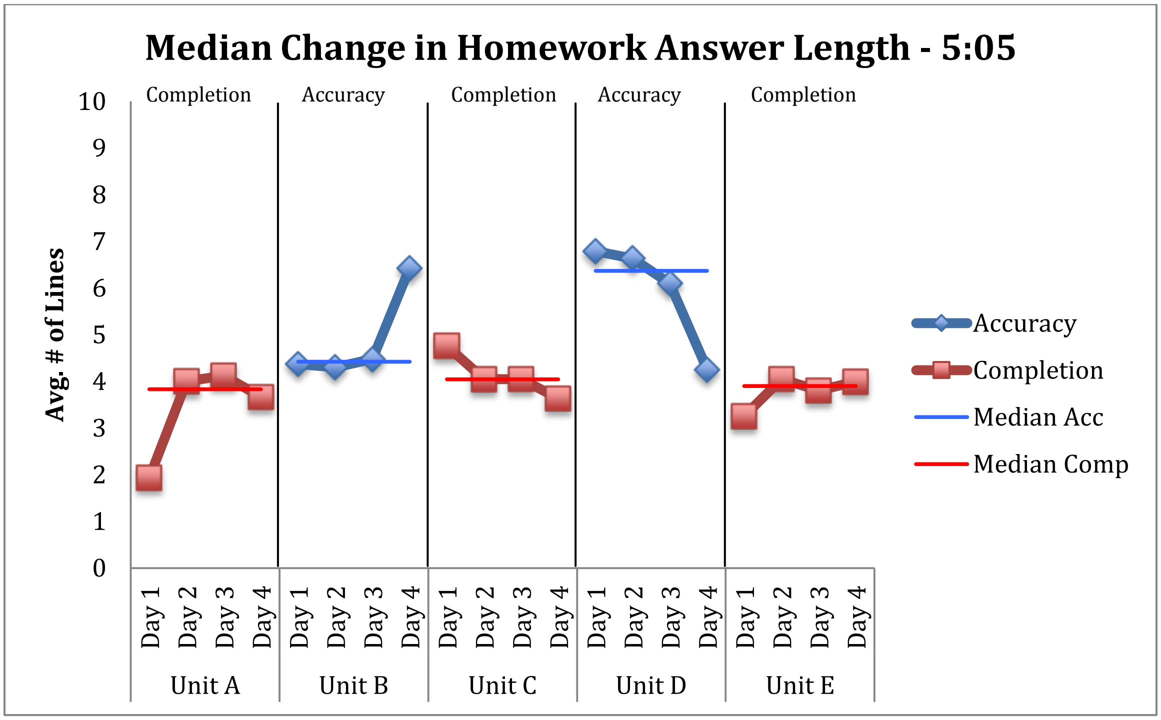


Figure 12. Study 3 Homework Length between Homework Contingencies in 5:05 Section

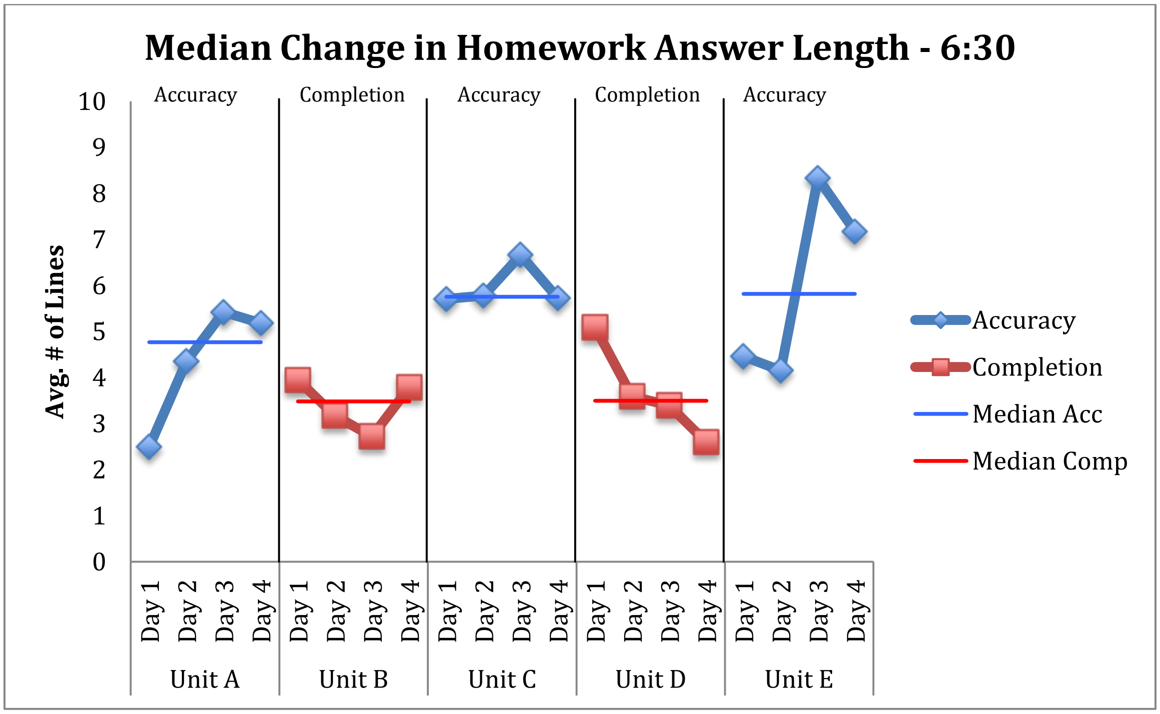


Figure 13. Study 3 Homework Length between Homework Contingencies in 6:30 Section

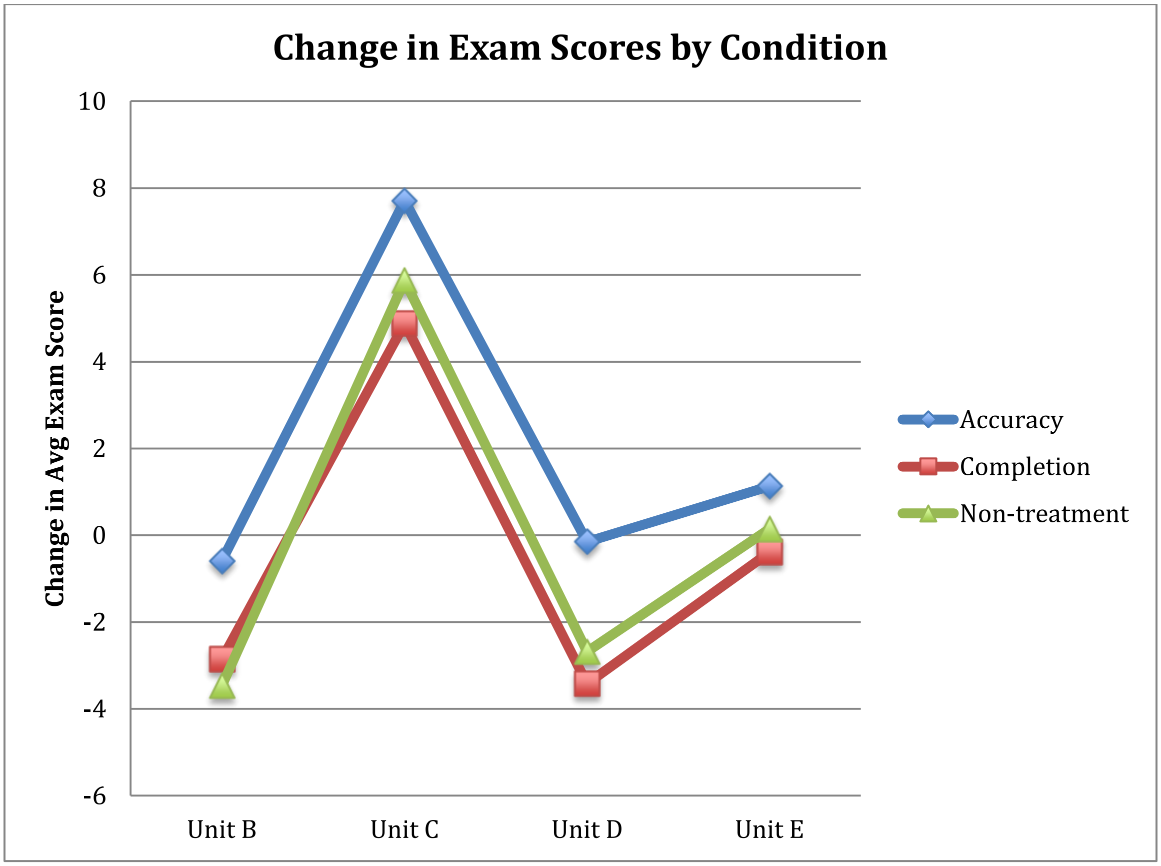


Figure 14. Study 3 Average Change in Exam Scores between Homework Contingencies

Note. Data are plotted by the change in contingency (e.g., Completion to Accuracy = "Accuracy")

Appendix C: Original Homework Contingency Instructions (Unit A only)

Homework Study Instructions

General Syllabus Description

Credit for Answers to Discussion Questions: Typed answers to the discussion questions will be collected each day for the first four days of each unit. In some units, credit will be based on the number of discussion questions you answer each day. However, in other units (which will be indicated at the beginning of those units), credit will be based on the accuracy of your answer to one or more discussion questions (randomly selected by the instructor each day). During these units the instructor will evaluate answers to the selected questions for accuracy and assign credit based on the accuracy of the answer to each question.

Unit Instructions

Accuracy Units: During these units discussion question answers will be evaluated for accuracy. Your credit will depend upon how accurate your answer is to one or more randomly selected discussion questions each day.

Completion Units: During these units discussion questions will be checked for completion. Your credit will depend on how many of the discussion questions you answer.

Appendix D: Revised Homework Completion Contingency In-class Announcement (Unit B onward)

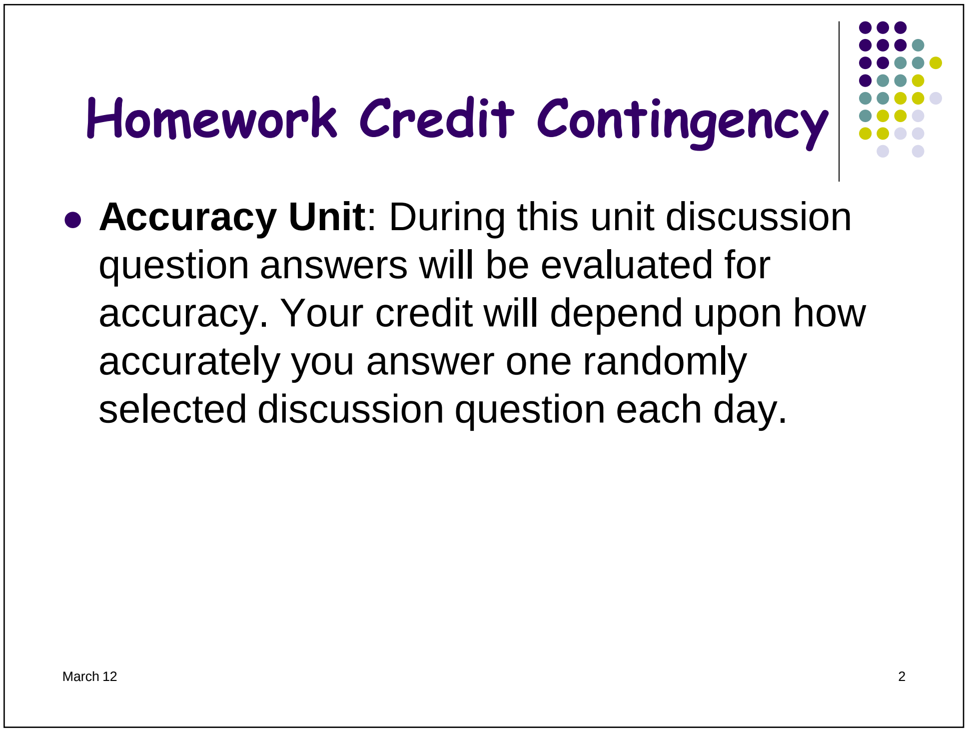
Homework Credit Contingency

- **Completion Unit**: During this unit discussion questions will be checked for completion. Your credit will depend on how many of the discussion questions you answer each day.

March 12

1

Appendix E: Revised Homework Accuracy Contingency In-class Announcement (Unit B onward)



Appendix F: Revised Course Website Completion Contingency Announcement (Unit B onward)

Hello Everyone,
I wanted to let you know that for Unit # your homework grade will be based on completion of homework questions. Each day we will look at how many of the homework questions you completed; the more of them for which you have an answer, the higher your points for that day. However, we will not be evaluating the accuracy of your answers to the homework questions. Therefore, you should be certain to provide an answer for all of your homework questions to maximize your homework credit.

Charles

Appendix G. Revised Course Website and Email Accuracy Contingency Announcement

(Unit B onward)

Hello Everyone,

As you know we are randomly selecting one question each day during this unit to determine your homework grade. Your homework grade is based on the number of concepts you correctly identified in your answer. This number is then compared to the number of concepts identified in the official answer. The formula is as follows:

# of concepts identified in your answer / # of concepts in the official answer = percentage of concepts correctly identified
The percentage of concepts correctly identified is rounded to the nearest 10% (e.g., 15% becomes 20%, 12% becomes 10%).

A score of >90% yields 5 points
70-90% yields 4 points
50-70% yields 3 points
30-50% yields 2 points
10-30% yields 1 point
Less than 10% yields 0 points.

You will receive an email stating the number of points you got for your homework each day during this unit. Included in this email will be the question that was selected and the list of concepts that should be included.

Charles

Appendix H: Homework Accuracy Contingency Feedback Email (Unit B onward)

Hello,

As you know we are randomly selecting one question each day during this unit to determine your homework grade. Your homework grade is based on the number of concepts you correctly identified in your answer. This number is then compared to the number of concepts identified in the official answer.

Your grade for the selected homework question for Day 1 is 0%, which gives you 0 out of 5 possible points. We recommend you review the official answer below and compare it to your own to gain a better understanding of how you can maximize your grade on subsequent days.

1. How should the distinction between concrete and formal operational reasoning influence the types of learning experiences that teachers provide students? (*p*. 2)

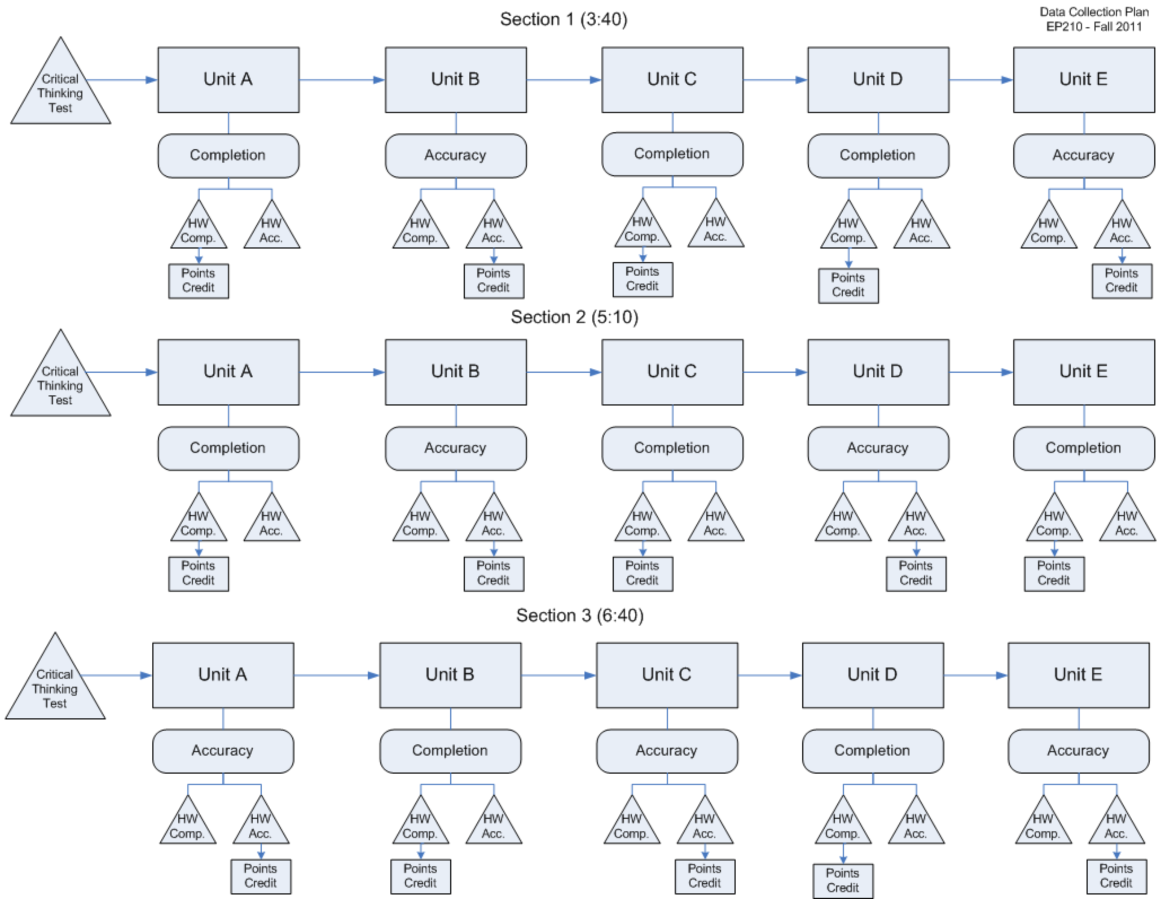
Concrete operational thought mostly deals with tangibles and relationships between observable events; formal operational thought deals with abstract notions and thus may be more suitable for some educational topics (such as theoretical topics, or advanced mathematics)

Exam Questions 5, 6, 7, 10, 13, 20

Concepts to be included = 3
* concrete operational deals with tangibles/observable events (or concrete operational more appropriate for hands-on activities/observable facts)
* formal operational deals with abstract concepts
* formal operational more appropriate for theoretical topics/advanced math

Sincerely,
Charles Galyon

Appendix I: Data Collection Flowchart



Appendix J: Unit A Homework Evaluation Rubric

Unit A Discussion Questions Rubric
1. What factors likely account for the differences in exercise patterns for boys and girls? (p. 2) • The influence of social norms • It is more socially acceptable for boys to be aggressive and competitive • Girls prefer non-competitive physical activity
2. Given that girls experience more stress reduction from exercise than do boys, why don't more girls exercise regularly? (p. 3) • Girls may have fewer role models who exercise • Females may receive approval for other activities • Other activities diminish time for exercising • In order to experience stress reduction girls must actually exercise first • Girls may not attribute stress reduction to exercise
3. Evaluate the validity of the claim that schools could be the single most influential institution in society in promoting healthy and productive living. (p. 4, Slide 5) • A large number of children attend school • Health and physical education can be effective • Children spend more time per day at school than any other institution • Some parents may have more influence over their children • Many parents spend little time with their children or have poor health practices • If teachers model and teach good health practices, there could be far reaching effects • Many educators have not bought into the notion of improving societal health through what they teach and model.
4. To what extent does Slide 14 indicate that our nation is on the right track in reducing drug use among high school seniors? (Slide 14) • High school seniors use of all three drug categories has decreased for all three patterns of use consistently from 1999 to 2008 • The tread has been especially favorable with respect to smoking rates • This pattern suggests that our nation is on the right track to reducing drug use • Because nicotine is a gateway drug for using other drugs, reduced smoking is likely to lead to a reduction in the use of other drugs
5. Why is self-directed quitting typically a more successful way of giving up

smoking than cold-turkey quitting? What assumptions underlie the two approaches? (pp. 10, 11, Slide 20)

- Self-Directed involves systematically developing a plan to stop smoking
- Self-Directed assumes quitting should be supported by environmental changes
- Some examples of environmental changes (include removing smoking cues, spending less time in smoking situations, spending more time with non-smokers, asking smoking friends to refrain from offering you cigarettes)
- Environmental changes are followed by a target date in which smoking is no longer permitted
- Cold turkey operates on the assumptions that quitting can occur smoker has enough will power
- Both approaches involve total cessation from smoking
- Cold-turkey involves no environmental changes

6. What are the major similarities and differences between the dietary plans highlighted in this unit (original food pyramid, redesigned food pyramid, Atkins food pyramid, and my pyramid)? (Slides 22-25)

- The original food pyramid and the redesigned food pyramid emphasize whole grains
- The original food pyramid and the redesigned food pyramid emphasize high levels of vegetables and fruit
- Original food pyramid does not distinguish between complex and refined carbs
- The redesigned food pyramid minimizes white and refined carbs
- Redesigned food pyramid minimizes red meat
- Redesigned food pyramid emphasizes vegetable oils
- The Atkins food pyramid emphasizes increasing proteins
- Atkins food pyramid minimizes carbs
- The redesigned food pyramid appears to be the most efficacious in facilitating long-term health
- The redesigned pyramid allows for alcohol use in moderation
- The redesigned pyramid includes exercise
- My pyramid is individualized

7. What criteria should be used in evaluating the efficacy of a special dieting plan? (p. 12)

- Criteria should include whether the plan incorporates essential nutrients
- Criteria should include whether the plan is balanced in regard to whole grains, vegetables, omega-3 fatty acids, fruit, protein, dairy products, and

complex carbs (need 1) • Criteria should include whether the plan has low levels of saturated and trans fat • Criteria should include whether the plan has high levels of fiber • Whether it can be maintained over the long-run • Drastic changes in diet could be detrimental to health
8. What are the principal differences in the recommended approaches to quitting smoking and altering one's food intake to lose weight? (pp. 10, 12, 13) • Most successful prognosis for stopping smoking is to keep trying to quit • May need to use a variety of strategies to finally quit smoking • Trying a variety of diets to lose weight may result in decreased metabolism • Decreased metabolism makes it harder to lose weight in the future • Quitting smoking and losing weight are similar in that both are easier to accomplish when one exercises
9. What are the most important pros and cons of abstinence-only versus abstinence-plus sexuality education? (p. 15) • Abstinence-only pro: if followed, students are guaranteed to be safe from unhealthy sexual behaviors • Abstinence-only con: The total-abstinence expectation of abstinence-only sexuality education may be unrealistic for many students • Abstinence-only con: provides little or no information regarding safe sex • Abstinence-plus pro: if students do engage in sexual activity, they will obtain the knowledge of contraceptives to keep them safe • There is little evidence that discussing contraceptives encourages sexual activity
10. What moral issues, if any, should be considered in sexuality education? (p. 15) • Could discuss whether premarital sex violates the well-being of your partner • Results of unwanted pregnancy • Increases the risk of sexually transmitted disease • Violates one's personal/religiously-based ethics about premarital or extra-marital sex

Appendix K: Unit B Homework Evaluation Rubric

Unit B Discussion Questions Rubric
1. How should the distinction between concrete and formal operational reasoning influence the types of learning experiences that teachers provide students? (p. 2) • Concrete operational deals with tangibles/observable event (or concrete operational more appropriate for hands-on activities/observable facts) • Formal operational deals with abstract concepts • Formal operational more appropriate for theoretical topics/advanced math
2. According to Piaget, what experiences facilitate the natural development of conservation? How is conservation related to schemes and operations? (p. 3, Slide 6 "Conservation Tasks") • Experiences that challenge the child's understanding (of conservation) • Working with more advanced children can help advance the child • Conservation requires consideration of multiple dimensions at once (operations) • Example of conservation (e.g., volume = height and width both)
3. What is the principal distinction between what IQ tests and achievement tests measure? Which provides more useful information about a child's cognitive development in school? (p. 6) • IQ tests measure generic achievement, memory, analytical skills, and expected/possible academic success • Achievement tests measure academic achievement in specific areas (math/reading...) • Achievement tests more useful for identifying actual academic deficits/success/strengths/skills
4. To what extent are IQ tests helpful to educators in serving the intellectual needs of children? How could curriculum-based assessment (CBA) be more useful than IQ tests in determining how to promote children's academic development? (p. 8) • IQ tests indicate academic potential • IQ tests do not identify specific areas of difficulty • IQ tests often used to determine eligibility for special education • CBA identifies specific areas of difficulty • CBA is more useful in determining how to help the child (provides some direction)

5. What are the principal factors that have sustained the use of IQ tests in American education? (p. 8) • IQ tests considered the domain of psychologists, who preserve the use of them • IQ tests are the most highly valued measure of intellectual ability (in society) with a long history • IQ scores provide something teachers can use to explain poor performance • IQ tests predict future academic success/performance
6. Which has greater academic value—determining students' IQ scores or their creativity scores? Why? (p. 8) • IQ tests reflect convergent thinking, used in academic tests • Creativity measures divergent thinking, which helps in forming variety of answers • Ideally both are used • Ability to generate multiple ideas is more useful in some settings
7. Which has the greatest potential for fairly and effectively assessing a child's cognitive potential, IQ tests, creativity tests, or critical thinking tests? • Critical thinking tests provide all the info needed • Critical thinking tests are the most culturally fair • Critical thinking tests use both divergent and convergent thinking • IQ tests are good at predicting academic success • IQ tests rely on information acquired in prior experiences • IQ tests can be culturally biased • Creativity tests rely on prior experiences • Creativity tests can be culturally biased • Critical thinking tests may be the best measure of cognitive potential
8. Why is direct instruction among the most recommended approaches for remediating the deficits associated with identified learning disabilities? (p. 12) • Direct instruction identifies and focuses on specific skills deficits • Direct instruction is faster and more efficient • Direct instruction has been proven to be effective with learning disabilities (can discuss the converse with regards to holistic instruction)
9. What are the major differences between the whole language and phonics approaches for promoting reading skills? (p. 15) • Whole language does not provide immediate and/or corrective feedback

• Whole language does not target word-attack skills (the converse stated with regards to the phonics approach can also count) • Whole language assumes reading/writing is natural/naturally acquired and develops in response to the environment • Phonics assumes skills must be explicitly taught
10. Explain the respective roles of task analysis, curriculum-based assessment, drill and practice, criterion-referenced evaluation, and feedback in direct instruction? (pp. 14-15) • All (task analysis, CBA, drill and practice, criterion-reference eval, and feedback) are applied in direct instruction • Direct instruction targets specific skills • CBA may use task analysis (to ID specific skills) • Feedback in direct instruction is immediate and corrective • Student progress measured by set criteria (criterion-referenced) (conversely student progress not measured by norm-referenced/other students) • Corrections are provided with a reminder of the rule to be applied • Learned skills are practiced with drill and practice to develop automaticity

Appendix L: Unit C Homework Evaluation Rubric

Unit C Discussion Questions Rubric
1. What are the principal similarities and differences between Slavin's and Kohn's models of cooperative learning? (p. 1, Slide 3) • Both emphasize student diversity • Slavin focuses on conventional school work; Kohn focuses on creative projects • Less emphasis on competition in Kohn's approach (than Slavin's) • In Slavin's approach the teacher structures the assignment; Kohn's approach allows students to structure the assignment • Slavin uses extrinsic reinforcement, Kohn emphasizes intrinsic reinforcement
2. In what ways could cooperative learning be beneficial or detrimental to the academic development of high-achieving students? (p. 2) • High-achieving students may develop deeper mastery (by answering questions and teaching) • High-achieving students may become aware of areas in which they do not have a complete understanding • High-achieving students may not cover as much material
3. Explain how the combination of individual and group-reward contingencies would facilitate performance more than either individual or group contingencies separately. (p. 4) • With only individual contingencies, students may seek information but not share any • With only group contingencies, students may rely on the highest performers • (With group contingencies, group members become reliant on each other)
4. What would be the pros and cons in CWPT of dividing the class on a random or ranking basis? (p. 5) • By using a ranking basis, increases the probability that low-performing students get needed help from high-performing students • Having high- and low-performers paired may contribute to social skills • When roles are rotated, high-performers may not gain as much • Students may become aware of rankings; random basis decreases this • In random assignment, low-performers may end up working together

5.	Do the logistics of CWPT legitimately qualify as peer tutoring or is there a more accurate label for this process? (p. 6) • Peer tutoring is thought of as a high-performer helping a low-performer • In CWPT, students switch roles • Process is more like studying in pairs or reviewing
6.	Do gender differences in the development of social skills generally favor boys or girls? Why would boys' attachments with their mothers be stronger than girls' attachment with their fathers? (p. 8) • Girls have more positive interpersonal relationships; boys tend to have more conflict and competition • Girls are more relationally aggressive, boys are more physically aggressive • Children are usually more attached to mothers as the primary care giver • Parents usually set the tone for a parent-child relationship • (Girls tend to form attachments through talk, something fathers may not engage in as much)
7.	Why do the natural consequences of bad social behavior not consistently increase constructive social skills and diminish bad social conduct as children get older? (p. 9) • Children may not always receive negative consequences • May have bad role models • Relational aggression increases as children get older • Children may receive positive consequences for bad behavior
8.	What appear to be the social characteristics of students who commit violent acts at school? How can schools prevent these social characteristics from turning into violent actions? (p. 10) • Bullied • Isolated from mainstream • Easy access to guns at home • Personnel need to monitor bullies • Identify victims • Threats should be taken seriously • Teachers should model respectful behavior • Arrange for victims to work with supportive students
9.	What are some likely contributors to an authoritarian parenting style? (p. 12) • High need for control • Excessive concern for conformity to parental standards • Parents were authoritarian

• Feel respect for authority is important • Militaristic or fundamentalist backgrounds • Living in a dangerous environment
10. Compare authoritarian and authoritative parenting with respect to parental involvement with the child. (pp. 6, 13) • Both are very involved in children's lives • Authoritarian involved in a controlling manner • Authoritative involved in promoting independence
11. Explain how parenting styles differentially affect students' school performance? (p. 13) • Authoritarian children have low confidence and little initiative • Authoritarian approach negatively associated with grades • Authoritative children have high confidence and independence • Authoritative approach positively associated with grades • Indulgent children tend to be impulsive and disobedient • Indulgent approach negatively associated with grades • Uninvolved children have low frustration tolerance and little self-control • Uninvolved approach negatively associated with grades

Appendix M: Unit D Homework Evaluation Rubric

Unit D Discussion Questions Rubric
1. Compare the nature and predictive potential of the different self-concept models. (p. 1) • Nomothetic is very general • Nomothetic is not as predictive (due to lack of specificity) • Hierarchical is more specific • Hierarchical is the best predictor • Taxonomic indicates all self-concepts are separate • Compensatory indicates self-concepts are negatively related to each other (inverse; strength in one means weakness) • Research indicates that self-concepts are positively related to each other
2. How are the notions of locus of control and self-efficacy alike and how are they different? (pp. 2-3, Slide 4 "Locus of Control and Self-Efficacy") • Both involve perception of personal control • Locus of control relates to control over outcome events (affect the outcome) • Locus of control can be external or internal • Self-efficacy relates to control over specific actions (ability to perform an action) • Self-efficacy can be high or low
3. Why are students with physical disabilities more likely to be accepted by peers than students with cognitive or behavioral disorders? (p. 2) • Physical disabilities can be seen • Students more inclined to help students with physical disabilities • Behavioral/cognitive disabilities are harder to understand • Assumption that students could behave/do better if they wanted
4. Contrast the ways high and low achievers account for success and failure experiences. (p. 4, slide 5 "High Achievers") • High achievers attribute success and failure to self (internal locus of control) • Attribute success to ability and effort • Attribute failure to inadequate or misplaced effort • Low achievers attribute success to luck (external locus) • Low achievers attribute failure to lack of ability (internal locus) • Causal attributions of high achievers contribute to effort, low achievers are unlikely to continue trying

5. Describe transitions in causal attributions from kindergarten to the high school years. (p. 4) • Young children do not distinguish between effort and ability • Young children believe hard work means high ability • Older children distinguish between effort and ability • Older children believe hard work means low ability
6. Contrast behavioristic and humanistic analyses of the relationship between behavior and feelings? Which model offers the greater potential for enhancing both behavior and feelings? (pp. 4 & 7) • Behaviorists argue behaviors contribute to feelings • If a person acts a certain way, they will feel that way • Humanists believe that feelings contribute to behaviors • If a person feels a certain way, they will act consistently
7. How are positive and negative reinforcement alike and how are they different? (pp. 4 & 5) • Both increase a behavior and are presented after the behavior • Positive involves presenting something to the child • Negative involves taking something away
8. What are the different ways that extrinsic reinforcement can affect intrinsic reinforcement? (p. 5, slide 9—"Beneficial Extrinsic Reward Conditions") • Extrinsic can undermine intrinsic when intrinsic is high • Extrinsic can improve intrinsic when intrinsic is low • Extrinsic is more effective with unexpected rewards • Extrinsic is more effective with rewards linked to the target behavior • Extrinsic is more effective with social rewards • Extrinsic is more effective when linked to quality of performance
9. Explain the similarities and differences between punishment, extinction, and response cost. (p. 6) • All three weaken a target behavior • Punishment usually presents something following the behavior • Extinction usually withholds reinforcement following the behavior (e.g., ignoring the behavior) • Response cost usually takes away something already given (e.g., earned privileges or points for a grade on a paper)
10. Compare the behavioral and humanistic positions on educational goals. (pp.

6-7)

- Behaviorists do not hold specific goals for education
- Goal setting should be consistent with the needs of the student
- Behavioral goals defined in operational terms
- Humanistic goals include positive feelings, enhancing self-concept, and satisfying human needs

11. Explain how the behavioral approach can be used to achieve humanistic goals. (p. 7)

- Behaviorists provide a framework to achieve any goal
- Humanistic teacher could set a goal for (e.g., higher self-concept) for the classroom
- Operationally define the behaviors that indicate the goal has been met
- Provide reinforcement (rewards) for engaging in behaviors consistent with the goal

12. Contrast optimists' assumptions of good and bad events with the explanations pessimists would likely advance for good and bad events. (p. 9)

- Optimists believe good events are result of own actions
- Optimists believe good things will continue to happen
- Optimists believe good events will improve their lives
- Pessimists believe good events are result of luck
- Pessimists believe good events are unlikely/infrequent
- Optimists believe bad events are rare and have little impact
- Pessimists believe bad events are the norm and they are responsible for them

13. Evaluate the effectiveness and efficiency of using brainwave biofeedback to alter cognitive and behavioral characteristics associated with ADHD. (p. 12, Slides 27-28)

- As many as 40% of children can go off medication permanently
- Helps with attention, independent study and learning
- Requires as many as 40 one-hour sessions; expensive in time and money
- Adjusts brainwave activity to lead to more permanent reduction of symptoms

14. What is the most judicious use of medication to treat ADHD tendencies? (p. 12, slide 20)

- Highly effective for children with ADHD (60-80% benefit)
- Probably best used in combination with behavior modification
- Lowest effective dose should be used

• Use of medication may allow for behavior modification to be more effective because child will have more sustained attention?
15. Explain the difference between rate of suicide as a comparative cause of death for adolescents/young adults versus older adults. (p. 13, Slide 30) • Suicide as a comparative cause of death reaches its peak in adolescence and early adulthood • Older individuals commit suicide more frequently, but also die from a variety of other causes (e.g., heart disease and cancer) • Younger individuals are much less likely to die, thus suicide is comparatively high (ranks 2nd or 3rd highest)

Appendix N: Unit E Homework Evaluation Rubric

Unit E Discussion Questions Rubric
1. Explain how an incentive for higher scores and perceived surveillance affect the probability of cheating on the Circles Test. How are cheating and lying distinguished on the Circles Test? (p. 3) • Incentive for higher scores increases cheating and lying • Perceived surveillance decreases cheating and lying • Report of 4 or more correct indicates likely lying or cheating • Determine lying by quickly collecting test answers and comparing them against student reports
2. How are Kohlberg's clinical interview and Rest's Defining Issues Test alike and different? (p. 3) • Both assess moral reasoning • Both use hypothetical moral dilemmas • Clinical interview is harder to administer • Answers must be interpreted in clinical interview • Options representing moral levels provided in DIT • DIT measures principal moral reasoning, equivalent to post-conventional
3. Explain the relationship between moral reasoning and moral conduct. (p. 1) • Both are indicators of moral development • Conduct is overt behavior • Reasoning is the "why" • Generally high level of reasoning associated with conduct • But behavior is often situation specific
4. Why have the instances of cheating in school nearly doubled in the last 30 years? (p. 5) • Standards of academics have increased • More education is now needed to achieve success • Honesty is not modeled as consistently • Consequences of cheating may be less severe • Parents not as inclined to support school punishment
5. Why would high-GPA students be less likely to observe cheating than low-GPA students but more likely than low-GPA students to confront cheaters? (p. 5) • High GPA students more focused on taking the exams • Low achievers more inclined to look around • High achievers more likely to confront cheaters because they value high

performance • High achievers work hard for their grades and expect others to do the same
6. At international, national, and personal levels, how can humankind satisfy current needs for natural resources without undermining the habitability of the earth for future generations? (p. 8) • Present consumption of resources is beyond sustainable levels • Purchasing only necessities and environmentally friendly goods • Using more economical transportation • Get independently verifiable information about environmental issues • Develop industries that meet environmental and economic needs
7. What is the current status of global warming and what are the prospects for global warming in the 21st Century? What do you see as the most compelling arguments for or against the reality of global warming? (pp. 8-9, Slide 14—Summer Arctic Sea) • Global temperatures rose about one degree in 20th century • May have contributed to rising ocean levels and extreme weather • Rising temperatures linked to CO2 levels • Expected to rise 3.2 – 7.8 deg in 21st century • Temperatures in Arctic increasing twice as fast • Ice shelf is cracking • Opponents contend earth goes through temperature cycles
8. Explain how early childhood personality tendencies could predict adult political ideology. How could genetic predispositions and environmental influences factor into your explanation? (p. 10) • Adult ideology somewhat predicted by childhood personality • Being fearful as a child may lead to seeing world as threatening • No socioeconomic or educational differences were documented from parents in the study • One might speculate that parents differ in authoritative and authoritarian parenting practices • Authoritarian parenting tends to be associated with childhood timidity; authoritative with confidence
9. Identify ideological and psychological characteristics shared by religious fundamentalists (e.g., Christian, Jewish, Muslim, Hindu) worldwide. What do these characteristics suggest about the possibility of peace across cultures heavily dominated by fundamentalist ideology? (p. 12)

• Typically authoritarian and ethnocentric • Strong emphasis on suppressing dissent • All believe they have the one and only truth • Makes the prospect for peace very dim • Those outside the faith are regarded as demonic or devoid of moral values
10. What are the similarities and differences between blind and constructive patriotism? When one politician attacks another politician's patriotism, what is the likely form of patriotism embraced by the attacker and by the attackee? (p. 13) • Both show deep loyalty to their country • Blind patriots of unquestioning loyalty • Constructive patriots feel free to question their leaders decisions • Politicians most likely to emphasize blind patriotism when attacking another politician and attack constructive

Vita

Charles Edward Galyon was born in Knoxville, Tennessee. He attended the University of Tennessee, Knoxville where he acquired his Bachelor of Arts degree in Psychology in 2004. He continued his education at Middle Tennessee State University where he received a Masters of Arts in Cognitive Psychology in 2007. In 2008, Charles entered the School Psychology Ph.D. program at the University of Tennessee, Knoxville. He is due to graduate in 2013 upon completion of his internship.