


Spring 2010

# The Effect of Individual or Group Guidelines on the Calibration Accuracy of High School Biology Students

Camilla C. Walck  
*Old Dominion University*

Follow this and additional works at: [https://digitalcommons.odu.edu/teachinglearning\\_etds](https://digitalcommons.odu.edu/teachinglearning_etds)

 Part of the [Science and Mathematics Education Commons](#), and the [Secondary Education Commons](#)

---

## Recommended Citation

Walck, Camilla C.. "The Effect of Individual or Group Guidelines on the Calibration Accuracy of High School Biology Students" (2010). Doctor of Philosophy (PhD), dissertation, Teaching and Learning, Old Dominion University, DOI: 10.25777/twc5-x684 [https://digitalcommons.odu.edu/teachinglearning\\_etds/31](https://digitalcommons.odu.edu/teachinglearning_etds/31)

This Dissertation is brought to you for free and open access by the Teaching & Learning at ODU Digital Commons. It has been accepted for inclusion in Teaching & Learning Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

THE EFFECT OF INDIVIDUAL OR GROUP GUIDELINES ON THE  
CALIBRATION ACCURACY OF HIGH SCHOOL BIOLOGY STUDENTS

by

Camilla C. Walck  
B.S. May 1994, Old Dominion University  
M.S. December 1999, Old Dominion University

A Dissertation Submitted to the Graduate Faculty of  
Old Dominion University in Partial fulfillment of the  
Requirements for the Degree of

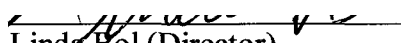
DOCTOR OF PHILOSOPHY

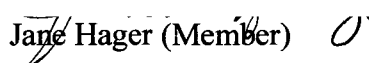
CURRICULUM AND INSTRUCTION

OLD DOMINION UNIVERSITY

May 2010

Approved by:

  
Linda Bol (Director)

  
Jane Hager (Member)

  
Sueanne McKinney (Member)

## ABSTRACT

### THE EFFECT OF INDIVIDUAL OR GROUP GUIDELINES ON THE CALIBRATION ACCURACY OF HIGH SCHOOL BIOLOGY STUDENTS

Camilla C. Walck  
Old Dominion University, 2010  
Director: Dr. Linda Bol

The effect of individual or group guidelines on the calibration accuracy of high school biology students was investigated. The study was conducted with 102 International Baccalaureate Middle Years Program biology students in a public school setting. The study was carried out over three testing occasions. Students worked in group or individual settings with and without calibration guidelines. Four intact classes were randomly assigned to one of four conditions: groups calibrating without guidelines; groups calibrating with guidelines; individuals calibrating without guidelines; individuals calibrating with guidelines. The students participated in the calibration activities one block before they actually took each of the three tests. On the day of each test, immediately before taking the test, each student made predictions as to what they thought they would score on the test. Immediately after taking the test each student made postdictions on what they thought they scored on the test. Calibration accuracy was determined by calculating the difference between prediction and postdiction scores and the actual test score achieved. The results indicated that students who calibrated in groups showed trends of more accurate calibration predictions. Although one testing intervention showed significant results for postdiction accuracy, the other two testing interventions showed varied results. Students who calibrated in groups achieved higher

scores on tests than did students who calibrated individually. In addition, guidelines were shown to be a significant factor in increasing achievement for students who calibrated individually. For students calibrating in groups guidelines had little impact. The results support the need for more research in metacognition and calibration techniques in order to improve student academic success.

Co-Directors of Advisory Committee:

Dr. Jane Hager

Dr. Sueanne McKinney

This thesis is dedicated to my husband, Kevin Walck, for all his patience and understanding of the many hours required to complete the process of earning a doctoral degree. I look forward to many hours spent making up for the time lost. Thank you for all your help, encouragement, and housecleaning. I love you.

## ACKNOWLEDGMENTS

There are many people who have contributed to the successful completion of this dissertation. First of all, I would like to thank my committee members for their endless hours of hard work and editing. Dr. Bol, Dr. Hager and Dr. McKinney are shining examples of what great educators should be. All of them have motivated me and kept me focused on my goal of finishing this journey. Old Dominion University is a better place because they are a part of what the university has to offer.

I especially would like to thank Dr. Bol for being the reason I have completed this dissertation. Her guidance, advice, and endless encouragement were vital to my success. I am truly a better person and writer due to her influence and guidance. I don't think she is fully aware of just how influential and important she is to so many students. Thank you Dr. Bol for truly caring about your job and the success of the students you serve.

Secondly, I would like to thank my family and friends for all their encouragement and understanding for all the events that I had to miss to complete this journey. My children, Danielle, Natalie, and Blaine, and my husband, Kevin, were and will always be my greatest source of pride and inspiration. I thank each of you for being the reasons I love my life and want to get the most out of it. I love all of you.

Finally, I would like to thank Dr. Deni Trombino for her encouragement and support which were vital to the completion of this dissertation. Thank you for participating in this journey with me. You are a true friend.

## TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vii
LIST OF FIGURES.....	ix
Chapter	
I. INTRODUCTION.....	1
SELF REGULATION AND CALIBRATION.....	1
METACOGNITION AND SELF-REGULATION IN CALIBRATION.....	3
NEED FOR RESEARCH IN CALIBRATION.....	7
DESIGN AND OVERVIEW OF STUDY.....	10
SUMMARY AND OVERVIEW OF SUBSEQUENT CHAPTERS...	11
II. LITERATURE REVIEW.....	12
INTRODUCTION.....	12
METACOGNITION, SELF REGULATION, AND CALIBRATION DEFINED.....	13
STUDIES INVESTIGATING KNOWLEDGE OF COGNITION.....	17
CLASSROOM STUDIES IN CALIBRATION.....	22
SUMMARY OF STUDIES.....	32
RATIONALE FOR STUDY.....	33
RESEARCH QUESTIONS AND HYPOTHESES.....	35
SUMMARY.....	36
III. METHODOLOGY.....	37
INTRODUCTION.....	37
PARTICIPANTS.....	38
DESIGN.....	39
MEASURES.....	40
PROCEDURES.....	41
SUMMARY.....	43
IV. RESULTS.....	44
INTRODUCTION.....	44
OVERALL DESCRIPTIVE FINDINGS.....	46
STATISTICS FOR INDIVIDUAL TEST RESULTS.....	48
STUDENT RESPONSES TO GUIDELINES.....	65
CONFIDENCE LEVELS.....	65
FACTORS INFLUENCING UNDERSTANDING.....	70

AREAS OF STRENGTH IN UNDERSTANDING.....	73
AREAS OF WEAKNESSES IN UNDERSTANDING.....	76
V. DISCUSSION.....	79
INTRODUCTION.....	79
ACHIEVEMENT.....	79
CALIBRATION ACCURACY.....	79
EFFECTS OF GUIDELINES AND SETTING.....	81
QUALITATIVE RESULTS.....	88
LIMITATIONS.....	92
DIRECTIONS FOR FUTURE RESEARCH.....	94
IMPLICATIONS FOR PRACTICE.....	96
SUMMARY AND CONCLUSIONS.....	97
REFERENCES.....	100
APPENDICES	
A. CALIBRATION GUIDELINES.....	107
B. PREDICTION AND POSTDICTION SHEETS.....	109
C. PARENT APPROVAL LETTER.....	110
VITA.....	111



## LIST OF TABLES

Table	Page
1. Overall Descriptive Statistics for Actual Score, Prediction Accuracy, and Postdiction Accuracy.....	47
2. Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test performance For Test 1.....	48
3. Follow-up ANOVA results for Treatments on Calibration Accuracy and Test Performance for Test 1.....	49
4. Means and Standard Deviations for Main Effect for Setting on Achievement for Test 1.....	50
5. Means and Standard Deviations for the Interaction of Setting and Guidelines on Achievement for Test 1.....	51
6. Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test Performance for Test 2.....	53
7. Follow-up ANOVA Results for Treatments on Calibration Accuracy and Test Performance for Test 2.....	54
8. Means and Standard Deviations for Main Effect of Setting on Achievement for Test 2.....	55
9. Means and Standard Deviations for Main Effect of Guidelines on Achievement for Test 2.....	56
10. Means and Standard Deviations for Main Effect of Guidelines on Prediction and Postdiction Accuracy for Test 2.....	57
11. Means and Standard Deviations for Significant Interactions of Setting and Guidelines on achievement for Test 2.....	58
12. Means and Standard Deviations for the Effect of Setting and Guidelines on Postdiction Accuracy for Test 2.....	60
13. Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test Performance for Test 3.....	62
14. Follow-up ANOVA Results for Treatments on Calibration Accuracy and Test Performance for Test 3.....	63

15. Descriptive Statistics for the Effect of Setting on Achievement for Test 3.....	64
16. Descriptive Statistics for the Effect of Setting on Prediction Accuracy for Test 3.....	64
17. Confidence Levels for Multiple-Choice and Short Answer Questions for Test 2.....	67
18. Confidence Levels for Multiple-Choice and Short Answer Questions for Test 3.....	69
19. Response Categories for What Made the Students More or Less Confident in Their Understanding of the Test Material.....	70
20. Quotes From Response to What Made the Students More or Less Confident in Their Understanding of the Test Material.....	72
21. Response Categories for Strength in Understanding of the Test Material.....	74
22. Quotes From Response to Areas of Strengths in Understanding of the Test Material.....	75
23. Response Categories for Weaknesses in Understanding of the Test Material.....	76
24. Quotes from Response to Areas of Weakness in Understanding of the Test Material.....	78

## LIST OF GRAPHS

Graph	Page
1. Interaction for Achievement (Test Score) for Test 1.....	52
2. Interaction for Achievement (Test Score) for Test 2.....	59
3. Interaction for Postdiction Accuracy for Test 2.....	61

# The Effect of Individual or Group Guidelines on the Calibration Accuracy of High School Biology Students

## CHAPTER I

### INTRODUCTION

This chapter begins with an overview of the constructs of self-regulation and calibration. It presents current definitions for metacognition and self-regulation as they relate to calibration. The use of group settings and guided practice in calibration is discussed and supported. The need for research in calibration will be presented and followed by the research questions for this study. Finally, a brief overview of the design is proposed.

#### Self-regulation and Calibration

The ability to self-regulate one's learning is vital to success in all academic endeavors. Self-regulation uses information from past performance to adjust future performance, and accurate self-evaluations are valued as a guide in regulating behavior in order to accomplish future goals (Radhakrishnan, Arrow, & Sniezek, 1996). Many students lack the ability to estimate their own level of understanding and often this leads to repeated experiences of failure. In fact, students are often dismayed by how poorly they have performed on an assessment for which they believed they were well prepared (Hacker & Bol, 2001). Students who can accurately assess their level of knowledge are in a better position to intensify or redirect their studying for a test, provide self-guidance during reading for better comprehension, or generate self-feedback indicating that a new skill is being properly acquired (Hacker, Bol, & Keener, 2008).

Unfortunately, self-regulation of learning is rarely encouraged in the classroom. Many students lack the metacognitive skills that are necessary to regulate learning and

make adjustments to their learning techniques as the learning process unfolds. These skills are important for students to develop in order to regulate their own learning and accurately calibrate the level of knowledge they have acquired. Well-developed skills in metacognition—awareness of one’s cognitive processes, cognitive strengths and weaknesses, and self-regulation are important for successful academic functioning (Klassen, 2002). Accurate calibration of learning is vital in order to make the needed adjustments to improve the accuracy of understanding of the level of knowledge obtained.

### *Calibration*

In order to be successful in academic pursuits one must be able to evaluate his or her level of understanding of the material being studied. By being aware of the level of understanding of material students can determine how well they are prepared for success on an evaluation of that material. The accuracy of this understanding can be assessed through calibration investigations. Calibration has been defined as the accuracy with which students can predict their own performance (Hacker, Bol & Bahbahani, 2008). Calibration accuracy has been used in studies to evaluate many curricular areas, including reading comprehension. Readers whose predictions and performance are highly correlated are considered to have good calibration of comprehension, whereas readers whose predictions and performance are minimally correlated are considered to have poor metacomprehension (Hacker, Dunlosky, & Graesser, 1998). Other studies in calibration have used the difference between predicted test scores and actual test scores to evaluate calibration accuracy.

When students gain the ability to calibrate their knowledge level it can facilitate

improved academic achievement. Garavalia & Gredier (2002) found that students who were accurate grade predictors earned the highest average grade for the course. In addition, they found that grade differences between these students and the comparison group, who were inaccurate predictors, were statistically significant. This is supported by Bol & Hacker (2001) who showed that high-achieving students were more accurate in their calibrations than low-achieving students. High-achieving students may earn high marks because they have developed accurate calibration skills. If this is true, then low-achieving students could improve their performance by developing more accurate calibration techniques. In order to improve calibration skills students need to be exposed to and practice self-regulating techniques.

### Metacognition and Self-Regulation in Calibration

#### *Metacognition*

Metacognition is a term coined by educational psychologists to describe the various aspects of how a learner processes new knowledge with an explicit understanding and recognition that continual learning is taking place (Orange, 1999). In essence, cognition is the awareness of ones' thought processes, and metacognition is the monitoring of these thought processes. Awareness of metacognition allows students to effectively monitor the acquisition of new knowledge. Researchers are convinced that metacognitive beliefs, decisions, and actions are important, but are quite often overlooked as determinants of success or failure in a wide variety of activities (Garofalo & Lester, 1985).

#### *Self-Regulation*

Self-regulation involves the willingness and ability to effectively manage or direct

one's learning using appropriate strategies and attitudes that help sustain goal-directed behaviors and to ask for assistance when necessary (Orange, 1999). Self-regulation is vital to calibration accuracy because it allows for the ongoing assessment of the progress that is being made towards a goal. The self-regulated process will end with the student being aware of how much knowledge he or she has gained. Calibration accuracy can be used to determine an individual's level of awareness of learned knowledge. Hence, self-regulation can improve calibration accuracy and improved calibration accuracy can result in improved academic performance. According to Zimmerman (2002), "Self-efficacy beliefs have been found to be sensitive to subtle changes in students' performance context, to interact with self-regulated learning processes, and to mediate students' academic achievement" (p. 82).

Teaching students how to self-regulate should be a part of their educational experience. A major goal of education should be to equip students with the intellectual tools, self-beliefs, and self-regulatory capabilities to educate themselves throughout their lifetime (Bandura, 1993). Research has shown that students who are better at calibrating their own level of learning are more successful academically. Unless the instructional environment creates and sustains an appropriate structure for practicing study techniques, it may be particularly difficult to change epistemological stances that undergrid what the student classifies as productive self-regulation (Pintrich, Marx, & Boyle, 1993). There are several methods that can be used to promote metacognition and improve calibration skills. Teachers need to be made aware of metacognitive processes and how they can be improved through classroom instruction.

### *Calibration in Group Settings*

Group settings provide an ideal situation for fostering metacognitive skills, especially when students are guided towards the development of these skills. Just as teachers should model metacognition, social interaction among students could also be used to cultivate metacognitive capacity. When working in groups' students gain the benefit of hearing how others address and solve problems. Group experiences can be used to guide the students in their individual development of metacognitive skills. If students are encouraged and guided to think critically together, then their spoken reasoning will ideally make these cognitive tools more readily available to them (Martinez, 2006). Teachers who recognize the importance of peers to the learning process encourage and offer opportunities for personal responses and collaborative interactions (Wiseman, 2003). Independent study lacks the dynamically responsive scaffolding and guidance that can be made available when learning proceeds in the context of social interaction (Winne, 1995). Student interactions provide opportunities for metacognitive development as they discuss the material and share their processes of learning new material.

Many researchers and practitioners are now convinced that by promoting metacognitive processes during instruction, more durable and transferable learning can be achieved. Tutors, learning assistants, and teachers, for their part can become the student's "metacognitive conscience" by asking questions of the student in order to develop his or her awareness and analytical processes (Taylor, 1999). Having students conduct metacognitive activities in collaborative settings can develop metacognitive skills. Group activities are easily incorporated into the classroom and not only benefit the student metacognitively, but allow the student to learn from his or her peers.



Group work with metacognitive processes such as calibration should be incorporated into the classroom setting in order for students to have opportunities to enhance their own understanding of material. Exposure to calibration practice in settings where students can analyze their own calibration techniques as compared to that of others allows the student to make needed adjustments in his own calibration techniques. Providing guiding questions in order to help focus the group on development of calibration accuracy can enhance group review activities. It is important that the questions the students are asking about their level of knowledge are focused on the metacognitive process.

#### *Guided Practice in Calibration in Group Settings*

Metacognitive skills can be further enhanced by guidance from the teacher in the form of verbal or written strategies that help maintain the focus of the collaborative activity on calibration. Students construct strategies from experience but also can be guided by teachers and peers to discover and control the development of effective learning tactics (Paris & Newman, 1990). Peers may bring new insight to the discussion that can help the development of individual calibration skills.

The use of guidelines during calibration can enhance learning processes by guiding the student through the metacognitive process of evaluating his or her learning of material. Increasing the student's self-awareness can help the student associate behaviors or successful (or unsuccessful) learning outcomes and aid in the accomplishment of the learning goal (Smith, 2001). The key is to help focus the student on thinking about the learning process and his or her personal goals to increase motivation (Talbot, 1997).

It is important for teachers to mediate group work to ensure the focus is on the

learning process. Metacognitive skills and knowledge can be acquired, and so, the argument goes, students can “learn how to learn”. Providing guidelines for the process of calibration prior to assessment allows the student to not only focus his cognitive processes on the task at hand, but helps in the development of metacognitive skills that are vital to the 21<sup>st</sup> century learner. Students must have the opportunity to practice and so must be placed in situations that require metacognition. If students are encouraged and guided to think critically together, then their spoken reasoning will ideally make their cognitive skills available to one another (Martinez, 2006).

An illustrative study highlights how group work can promote self-reflection and deeper understanding. Cantrell (2002) examined the content of small-group discourse and found that they provided opportunities to reflect further on readings, to clarify understandings, and to share insights from their own experiences. Cantrell also found that in many exchanges between and among the participants in the study, construction of knowledge occurred through deeper comprehension, clarification, and identification of important points. Thomas, Bol, Warkentin, Wilson, Strange and Rohwer (1993) found that one important role teachers play is in prompting student engagement in productive, demand-responsive study activities.

#### Need for Research in Calibration

Metacognitive skills have become more important in education as local, state, and national assessments have become the standard for measuring student ability. Student performance on high-stakes tests has an impact on educational placements, grade promotion, academic major, college admissions, graduation, and entry into various professions (Hacker, Bol, & Keener, 2008). Previous research has focused on calibration

ability as related to success on these high-stakes tests, but has failed to evaluate calibration practice as related to academic success in high school courses. Since calibration ability has been shown to be related to academic success this skill should be developed early in the educational experience.

Research suggests that metacognitive skills can be taught and can subsequently improve academic achievement (Hartley, 2001). Nickerson (1988) stated that there is abiding conviction among many educators that the development of thinking should be a primary goal of education. Although this belief has been prevalent for many decades, few teachers are aware of the need to foster metacognitive skills or the methods that may help them develop these skills in their students. Dahl (2004) states that:

“To help pupils with the metacognitive process it is necessary that the teachers are educated to be able to discuss the learning process and strategies with pupils. The development of the pupils metacognition will help the pupil at any level.”  
(pp. 153)

Whether calibration and other metacomprehension strategies can be improved with instruction remains a question that has not yet been definitively answered (Bol & Hacker, 2001). Previous research has focused mainly on college level investigations into calibration and has failed to adequately address calibration at the high school level. This research helps to fill that gap, and provides the added benefit of studying calibration in a classroom context. Hacker, Bol, & Keener, (2008) argue for the need to go outside the laboratory into more ecologically valid environmental situations in order to effectively evaluate calibration techniques.

Research on calibration in group settings is also lacking. Group settings provide

students the time to reflect on their learning in situations where individual reflections can be enhanced by group discussions. Orange (1999) found that using peer models to teach self-reflection was effective. By working in group settings and observing both successful and unsuccessful peers students may have become more aware of their own academic shortcomings and may have become more willing to modify their own behavior (Orange). Group interactions provide opportunities for students to seek help from their peers in self-regulatory processes. Research has shown that students who effectively monitor their overall use of self-regulation strategies seek help more often from peers, teachers, and parents and learn more than students who do not seek help (Zimmerman, 2008). By allowing instructional time for group review of material prior to testing, teachers allow opportunities for students to seek help who may otherwise not have done so. The success of group interactions on academic performance can foster continued use of help-seeking strategies that result in higher self-regulation skills.

The use of guidelines in group settings offers unique opportunities to improve calibration skills. Guidelines have the potential to focus the student on the metacognitive process and to help the student develop a pathway to the successful calibration of knowledge. Many students lack the ability to successfully reflect on their level of knowledge and need to be guided through the process in order to develop this skill. The use of guidelines in group settings has the added benefit of allowing the student to hear how others calibrate their level of knowledge.

This study will focus on the use of group interactions and calibration guidelines to foster the development of successful calibration skills. This study is unique in that it combines group investigations into calibration with the use of guidelines to foster

metacognitive skills. Previous research is lacking in studies that look at the interactions between these two variables.

### *Research Questions*

The research questions for this dissertation focus on the effects of calibration practice in either group or individual settings and with or without guidelines on calibration accuracy and achievement of high school biology students. In addition, written responses to guided questions from the group calibrating with guidelines will be collected. This will offer more insight into the effectiveness of the use of guidelines in the collaborative process of calibration as it is proceeding in the group settings. More specifically, the following research questions will be addressed:

1. Does receiving guidelines during calibration practice improve calibration accuracy and achievement for high school biology students?
2. Is calibration practice in groups more effective than individual practice in improving calibration accuracy and achievement for high school biology students?
3. How do guidelines and learning settings (group vs. individual) interact to affect calibration accuracy and achievement?
4. What do students write in response to guided questions designed to improve calibration?

### Design and Overview of the Study

A quasi-experimental research on the effects of calibration practice in either group or individual settings and with or without guidelines on calibration accuracy and achievement in a high school biology course was conducted. A fully crossed factorial

design was employed. Four intact biology classes were involved in the study; two classes participated in group calibration, with one class receiving group calibration guidelines and one class calibrating without guidelines; two classes participated in individual calibration, with one class receiving individual calibration guidelines and one class calibrating individually without guidelines.

The data collected consists of predictions and postdictions for three different testing occasions. In addition, qualitative data was collected in the form of responses to the calibration questions from the class that participated in group calibration with guidelines and the class that participated in individual calibration with guidelines. Quantitative data was analyzed and reported using multivariate analysis of variance (MANOVA) and factorial analysis of variance (ANOVA). Qualitative data consisting of responses to guiding questions was analyzed via content analysis.

### Summary and Overview of Subsequent Chapters

Chapter I has provided a rationale and the accompanying research questions that were addressed in this study. Metacognition and calibration were briefly defined and will be more fully explored in Chapter II. Chapter II investigates the current definitions attributed to metacognition, self-regulation, and calibration. It summarizes important findings in these areas as related to education, and compares findings from previous empirical research in metacognition, self-regulation, and calibration. Emphasis is placed on calibration studies. The need for more research into calibration at the high school level, specifically in science, is supported. The hypotheses for the research questions are addressed in Chapter II. Chapter III further and more completely outlines the methodology that was used for this research.

## CHAPTER II

### Literature Review

#### Introduction

There exists a substantial amount of research that investigates the use of metacognition, self-regulation, and calibration in educational settings. However, there is great variation as to how these constructs are operationally defined and delivered in the classroom setting. In addition, a discrepancy exists between those studies conducted in laboratory settings and those studies conducted in traditional classroom settings. The majority of previous research has focused on metacognition in non-traditional classroom settings.

This chapter provides a brief overview of metacognition, self-regulation, and calibration. Previous studies in these areas are outlined, and an overview of studies in these areas are presented in order to empirically investigate these constructs. Emphasis will be placed on research in calibration studies. Studies investigating student knowledge of cognition are presented and followed by studies specifically focused on calibration.

Previous classroom studies in calibration without interventions and previous classroom studies in calibration with interventions are presented and discussed. Calibration interventions are reviewed including studies that investigate the use of incentives and reflections, practice tests, group work, and peer interactions on calibration accuracy. Studies investigating achievement level and calibration accuracy are also presented. The need for more research into calibration at the high school level, specifically in science, is supported. Chapter II presents the research questions addressed in this research, and the proposed hypotheses for them. This chapter ends with a brief

overview of Chapter III.

### Metacognition, Self-regulation, and Calibration Defined

It is difficult to state a clear definition of metacognition, self-regulation, and calibration. In other words, metacognition, self-regulation, and calibration are all terms that help in defining each other. An overview of each of these cognitive domains is necessary in order to understand each individually.

#### *Metacognition*

Piaget referred to the process of “reflexive abstraction” as a mechanism for extracting, reorganizing, and consolidating knowledge (Garofalo & Lester, 1985). This definition could easily be used to describe metacognition as well. Garofalo & Lister argue that it is difficult to separate what is metacognitive from what is cognitive. Metacognition experiences are defined by Hacker, Dunlosky, & Graesser (1998) as being concerned with one’s awareness of his or her cognitive or affective processes and whether progress is being made toward the goal of a current process. In other words, metacognition is the ability of students to think about their level of knowledge attainment as they are investigating new information. Without adequate and appropriate cognitive processing it is impossible to successfully engage in metacognition. To distinguish between cognition and metacognition Nelson and Narens (1990) offered the following distinctions:

- (1) Mental processes are split into two or more specifically interrelated levels, a cognitive level and a metacognitive level;
- (2) the metacognitive level contains a dynamic model of the cognitive level; and



- (3) there are two dominance relations called control and monitoring, which are defined in terms of the direction of flow of information between the metacognitive and cognitive levels.

As outlined above, it is easy to see that these processes are interrelated and cannot exist independently of each other.

Grimes (2002) defines metacognition as a term coined by educational psychologists to describe the various aspects of how a learner processes new knowledge with an explicit understanding and recognition that learning is taking place. With metacognition the learner is not only aware he is learning, but is aware of how that learning is proceeding. Grimes summarizes the process as one that involves the abilities to appraise and manage the internal aspects of learning. Hence, for students to be successful in metacognition, they must be continually analyzing the effectiveness of their monitoring of cognitive strategies and not just be engaging in the use of these strategies.

According to Martinez (2006) the metacognitive process is the monitoring and control of thought. He identifies three major categories of metacognition: metamemory and metacomprehension (the understanding of one's own knowledge state), problem solving (the pursuit of a goal when the path to the goal is uncertain), and critical thinking (evaluation ideas for their quality- especially judging whether or not they make sense). It is clear that all learning involves metacognition. One way of viewing the relationship between cognition and metacognition is that cognition is involved in doing, whereas metacognition is involved in choosing and planning what to do and monitoring what is being done (Garofalo & Lester, 1985).

### *Self-regulation*

The process of self-regulation has been defined in various ways, but they all refer to the ability to regulate one's learning process. According to Zimmerman (1986), "Self-regulated learning is the degree to which students are metacognitively, motivationally, and behaviorally active participants in their own learning process" (p. 308). Orange (1999) summarized self-regulation as the willingness and ability to effectively manage or direct one's learning using appropriate strategies and attributes that help sustain goal-directed behaviors and to seek assistance when necessary. Self-regulation would not be possible without metacognition. One must be able to monitor his or her level of understanding in order to be successful in the self-regulation of that learning. According to Butler and Winne (1995)

"In academic contexts, self-regulation is a style of engaging with tasks in which students exercise a suite of powerful skills: setting goals for upgrading knowledge; deliberating about strategies to select those that balance progress toward goals against unwanted costs; and, as steps are taken and the task evolves, monitoring the accumulating effects of their engagement." (pp. 245)

As self-regulated learners engage in academic tasks, they draw on their knowledge and beliefs to construct an interpretation of a task's properties and requirements (Butler & Winne, 1995). Once these properties and requirements are decided upon, to be successful in completing them the learner must continue to evaluate his level of understanding. In essence, the learner cannot be successful in self-regulation without engaging in metacognitive behaviors as well. Metacognitive behaviors regulate the learning process that leads to self-regulation of knowledge retention as the learning

process unfolds. Self-regulation of knowledge refers to the degree to which individuals actively participate metacognitively, motivationally, and behaviorally in their individual learning processes (Filho & Yuzawa, 2001). A major component of self-regulation is the ability to calibrate one's level of learning.

### *Calibration*

According to Horgan (1990) calibration can be defined as “the accuracy with which students can predict their own performance”. More specifically, calibration has been defined as a measure of the degree to which a person's judged ratings of performance correspond to his or her actual performance (Hacker, Bol & Bahbahani, 2008). This ability is linked directly to metacognitive processes and can be explored by assessing an individual's belief of how well they think they will perform on a task both before (prediction) and immediately after (postdiction) completing the task. These cognitive evaluations are often measured as the correlation of test performance and predicted performance before studying, during studying, after studying, or even after a test itself (Green & Azevedo, 2007). Individuals who show little variation between predictions and postdictions and the actual score obtained are considered to be better calibrators.

Although there are several significant contributors to calibration accuracy, the underlying psychological process reflected in calibration entails a person's monitoring of what he or she knows about a specified topic or skill and evaluating the extent of that knowledge in comparison to some criterion task, such as an examination (Hacker, et al., 2008). Accurate calibration as to the level of knowledge obtained allows individuals to successfully plan effective study strategies focusing on areas of need. The metacognitive

skill of predicting performance on specific topics leads to appropriate focus in preparing for tests (Westley, 2008). Although there are several significant contributors to calibration accuracy, the underlying psychological process reflected in calibration entails a person's monitoring of what he or she knows about a specific task, such as an examination (Hacker, Bol & Keener, 2008).

It is clear that calibration, along with self-regulation, cannot be accomplished without involving the metacognitive processes. Therefore, it is important that students are not only aware of their own metacognitive processes, but understand how to use these processes to improve their self-regulation and calibration of the level of knowledge obtained. Monitoring the awareness of knowledge obtained is a skill that needs to be developed during the educational experience. Incorporating interventions that promote metacognitive processes that focus on the development of calibration skills can enhance metacognitive processes in the classroom.

#### Studies Investigating Knowledge of Cognition

Knowledge of cognition means that one has relative stable information about one's cognitive processes (Dahl, 2004). An accurate and ongoing assessment of one's learning level is an important factor in the ability to achieve academic success. One could argue that a condition for being able to develop one's metacognition is that when one reflects upon one's learning history, one can remember the general processes of what one does and how it works (Dahl). In other words, it is important to analyze how new material has been learned in the past in order to experience successful learning in the future. The knowledge of cognition is what drives the analysis of the metacognitive process. Students who lack the ability to analyze their learning metacognitively are

missing a vital component of successful learning strategies.

In an attempt to better understand how knowledge of one's cognition may play into one's ability to accurately calibrate, Carvalho & Yuzawa (2001) conducted a study on the role knowledge of cognition and regulation of cognition play into calibration accuracy. Their study involved 77 college students who were given a knowledge of cognition checklist to assess their level of awareness of their own knowledge, and knowledge of monitoring their own knowledge. Based on their responses the participants were divided into two groups. One group consisted of those with high levels of knowledge of their cognitive processes and the other group consisted of those with low levels of knowledge of their cognitive processes. Six multiple-choice tests to measure their accuracy of confidence judgments (calibration) were given. Prior to each test they were asked to rate the accuracy of their test performance based on how many test questions they thought they would answer correctly. After all six tests were completed, the participants were asked to rate their overall test performance on the combined six tests in order to get a measure of global metacognitive regulation. The accuracy of their judgments was used to divide the participants into two new groups. One group consisted of students with high-accuracy in predicting test performance and the other group consisted of low-accuracy in predicting test performance. The participants in the high level of knowledge of cognitive processes scored significantly higher than the group of participants in the low level of knowledge of cognitive processes. In addition, high regulators were more accurate in global prediction accuracy than low regulators. Knowledge of cognition was a good predictor of performance and level of confidence, in the same way that regulation of cognition was a good predictor of performance and

global accuracy (Carvalho & Yuzawa). It seems that both knowledge of cognition and regulation of cognition must be present in order to experience ongoing accuracy in calibration. Many students lack the metacognitive awareness of the process of calibration and are missing a vital component of the learning process.

A qualitative study on knowledge of cognition was conducted by Dahl (2004) on ten high-achieving high school mathematics students. Personal interviews were carried out in order to analyze the students' levels of metacognitive awareness. Although there was variation within the level of metacognitive awareness reported by the students, all the students reported some level of the use of metacognition as they solved mathematical problems. The student responses suggest that a combination of cognitive and non-cognitive factors are integrated into the learning process in order to achieve successful learning. Perhaps the awareness of the cognitive involvement in learning is related to the high level of success exhibited by the students.

In order to investigate age as related to cognitive abilities Justice and Dornan (2001) compared metacognitive differences between traditional and non-traditional (25 years and older) age college students. The participants were tested for cognitive functioning in order to assess their level of cognitive monitoring and self-evaluation of cognitive ability. Older (non-traditional) students reported more frequent use of cognitive study strategies including selection of cognitive task and active selection of a processing strategy. It appears that metacognitive strategies may become more sophisticated with age.

#### *Achievement Level and Calibration Accuracy*

Several studies have found achievement level to be linked to calibration accuracy.

In general, high achievers have been found to be more accurate calibrators than are low achievers. Nietfeld and Schraw (2002) conducted a study of 93 undergraduate psychology students. They divided the students into three different groups consisting of low, medium, and high achievers. The students answered 24 test questions and made confidence judgments based on how confident they were in their answers for each question. The difference between their confidence score and actual score was calculated and used to determine calibration accuracy. It was found that students who possessed higher levels of knowledge (high achievers) of material made more accurate confidence judgments than those who possess lower levels of knowledge (low and medium achievers). This is supported by Hacker, Bol, Horgan & Rakow (2000) who found that low achievers were inaccurate in their calibration of knowledge, and could not accurately predict or postdict their scores.

Bol and Hacker (2001) also found significant interactions between item format and achievement group on calibration accuracy. High achieving students were more accurate calibrators than low achieving students. Although high achieving students showed little difference between their calibration on multiple-choice and essay items, the low achieving students were less accurate in their calibrations of multiple-choice items. This finding has important implications due to the fact that many high-stakes tests are of the multiple-choice format. Significant effects were not found for postdiction accuracy. This could be due to the information offered by the actual presentation of the exams. Students should be better at calibrating on tests that they have taken because they have specific information (the actual test questions) upon which to base their predictions.

Bol, Hacker, O'Shea, and Allen (2005) conducted research on the influence of

overt practice, achievement level, and explanatory style on calibration accuracy and performance. Participants took six on-line quizzes during the course and a final exam at the end of the course. The participants were divided into two groups in which one group made predictions and postdictions for each of the five quizzes (overt group) and the final exam, and the other group only made predictions and postdictions for the final exam (covert group). Students were also asked to fill out a questionnaire that indicated the degree to which various factors influenced the accuracy of their predictions and postdictions (Bol, et al., 2005).

The results showed that higher achieving students were more accurate than lower achieving students (but underconfident) in their predictions. The lower achieving students were also less accurate in their postdictions than were the higher achieving students. In addition, the low achievers were found to be overconfident in their judgments, and the higher achievers were underconfident in their postdictions. There was no statistically significant difference found between the overt practice group and the covert group for calibration accuracy on the final exam, or across the five quizzes.

Perhaps high achieving students have developed metacognitive skills that enhance their ability to analyze questions, particularly multiple-choice questions, which are lacking in low achieving students. The aspect of metacognition involving the ability to monitor and regulate the use of cognitive activities affects academic performance (Justice & Dornan, 2001). It may be that high achieves are so because they are better at self-regulating the level of knowledge attained and possess more accurate calibration skills. The underconfidence exhibited by high achieving students may be due to the fear of high predictions or postdictions ‘jinxing’ the actual score obtained (Hacker & Bol, 2004).



Low achievers may exhibit overconfidence in order to preserve their self-esteem. When information about the self is not positive, the motivation towards accuracy is at odds with self-enhancement needs (Radhakrishnan, Arrow, & Sniezek, 1996). More research is needed in this area in order to gain insight into the metacognitive processes high achievers possess.

### Classroom Studies in Calibration

Classroom studies in calibration have attempted to develop an understanding of how calibration accuracy can help with academic achievement. Previous studies have investigated calibration skills in general, and investigated the use of interventions in order to evaluate the effectiveness of these interventions on calibration accuracy. Research in calibration accuracy has supported the need for developing successful calibration skills.

#### *Descriptive Studies in Calibration*

Several studies have been conducted in order to investigate the role that calibration ability plays in academic success. Grimes (2002) conducted a study consisting of 253 college macroeconomics students. He investigated the ability of the students to accurately predict their exam score by having the participants make predictions about their exam score 48 hours before the exam, immediately prior to the exam, and immediately after the exam. These scores were compared to the actual scores obtained. In addition, he had the students complete an expectation of concepts survey to analyze the students' awareness of the scope of the learning required for the examination.

The results of this study were statistically significant for the degree of overconfidence revealed by the students' pretest and posttest performance. Large and positive differences between expectations and performances were found in all cases,

indicating a resounding degree of overconfidence by members of the class (Grimes). The degree of overconfidence was found to be associated with the degree of predictive calibration performance. The positive and significant regression coefficient on this measure of metacognitive functioning indicated that higher degrees of overconfidence were associated with lower degrees of predictive calibration prior to the test (Grimes). However, the degree of overconfidence was found to diminish between the 48-hour prediction, the prediction immediately prior to the test, and the postdiction immediately after the test.

Not surprisingly, the more inaccurately the students identified their expectations of the concepts to be covered on test, the less accurate were their predictive scores. Awareness of what will be tested is necessary to accurately measure ones' test score. Grimes believes that students need feedback on performance in order to successfully calibrate their level of understanding. He specifically states that:

“In classes that rely heavily on lectures, students are not actively involved and do not receive significant amounts of instructional feedback concerning the state of their understanding and mastery of material. Thus, with a relative lack of information concerning their ongoing learning, the metacognitive processes of typical principles students may lead to inaccurate conclusions.” (pp. 27)

Teachers need to incorporate metacognitive processes such as calibration into the instructional day in order for students to develop accurate calibration skills. Since lecture remains a large percentage of the method of content delivery, it must include an avenue for the development of accurate calibration skills. This can be done during lecture by including inquiry based questioning that focuses on the level of knowledge obtained by

the student as the lecture proceeds. Without feedback as to their level of understanding students may continue to inaccurately assess their level of knowledge attainment.

Riggs, Bol, Nunnery, and Dickerson (2009) conducted a study investigating correlations between calibration accuracy and achievement level. The study involved 77 middle school math students who made predictions and postdictions on a sixth grade Virginia Standards of Learning (SOL) test. A median split was used to categorize the students into either a higher achieving group or lower achieving group based on their test scores. The achievement groups were compared for prediction and postdiction accuracy in order to see if there was a significant difference between the calibration accuracy of high and low achieving students. In addition, the students responded to open-ended questions addressing factors that they considered to have contributed to the accuracy of their predictions and postdictions.

The results showed that the high achievers were more accurate and slightly overconfident in their predictions and postdictions. Lower achieving students were less accurate and exhibited higher levels of overconfidence. The majority of the students identified studying, self-evaluation, and prior test performance as the main factors that influenced their calibration accuracy.

Investigating the level of calibration accuracy should involve an investigation of how aware students are of their metacognitive processes. Schraw (1997) conducted a study involving 95 undergraduate college students on their ability to accurately calibrate their performance on test items. The following four tests were given: the lexical comparison test (word choice test); the Nelson-Denny reading comprehension test (general comprehension knowledge); the syllogistic reasoning test (selecting valid

conclusions); and a basic math test (computing simple probabilities). Prior to the test the participants completed The General Monitoring Strategies Checklist (GMSC) to assess their general monitoring knowledge and techniques. Individuals were assigned to groups (low, average, and high level monitoring) based on their scores on the GMSC test. On all four tests participants rated their confidence level for each test item based on how confident they were in their answer choice. Confidence levels were compared with performance scores for each group. The results showed that individuals who have access to metacognitive knowledge use this knowledge to make more accurate judgments of their performance. It was also found that the low-monitoring group was significantly more under-confident than the average and high-monitoring groups.

“Evidence of the effect of metacognition knowledge on performance assessment was provided by the finding of a positive correlation between confidence judgments and bias scores. Future research should be directed toward investigating the construction of metacognitive knowledge from a developmental process.” (pp. 144)

Metacognitive practices should be introduced early in the educational experience in order for students to become more aware of this important cognitive process that has potential to improve their academic success. Practice in calibration beginning early in the educational experience may allow the students to develop more accurate calibration skills that can lead to improved academic performance. In addition, research needs to focus on specific strategies that may improve calibration accuracy in order to identify successful strategies for fostering these skills.

*Calibration Studies With Interventions*

Since calibration accuracy has been linked to achievement level, it is important to investigate interventions that may be successful in improving calibration accuracy. Misjudgments and inaccuracies in student self-assessment may result in poor study habits and ultimately poor performance on classroom assignments and examinations (Grimes, 2002). Starting early in the educational experience students should be exposed to opportunities to practice calibration techniques and analyze their metacognitive strategies. In order to successfully expose students to methods that can increase their calibration accuracy, it is important to know what interactions are effective in improving calibration skills.

In an attempt to investigate the effects of incentives and reflection on calibration accuracy Hacker, Bol, and Bahbahani (2008) conducted a study with 137 college educational psychology students. Calibration accuracy of the students was compared based on four conditions:

- “ (a) students who were asked to reflect on explanations for their calibration judgments but were not provided with extrinsic incentives to improve accuracy;
- (b) students who were not asked to reflect on their explanations of their calibration judgments but were provided with extrinsic incentives to improve accuracy;
- (c) students who were asked to reflect on their explanations and provided with incentives to improve accuracy; and
- (d) students who were not asked to reflect on their explanations nor provided with extrinsic incentives to improve accuracy” (pp. 103)

As expected, higher achieving students were found to possess more accurate calibration

skills than were lower achieving students. The significant interaction revealed that the lower-achieving students who received extrinsic rewards experienced more accurate postdiction accuracy for open-ended questions. The students identified a relationship between their level of knowledge of the material and their calibration ability (Hacker, et al., 2008).

In order to see if exposure to practice tests could improve calibration skills Bol & Hacker (2001) conducted an investigation into the effects of practice tests verses traditional review on calibration accuracy and performance. The study consisted of 59 students enrolled in two identical research methods courses taught by the same instructor. One class was given a practice tests prior to the midterm and the final exam, and the other group was not given practice tests. One group took the practice tests and then discussed their responses with the instructor. The group that did not receive practice tests spent the same amount of time in instructor led review and discussion. Both groups were asked to predict what they thought they would get on the midterm and final exam immediately prior to and immediately after the administration of the exams. The predictions and postdictions were compared to the actual mid-term and exam scores for both multiple choice and essay questions.

Prediction accuracy results for the mid-term and final exam showed that students in the practice tests group were significantly less accurate in their performance on the multiple choice section of both the mid-term and final exam than those students who did not take the practice tests, but no significant difference was found for the essay items. The students who did not take practice tests scored higher on the multiple-choice items than those who took the practice tests.

The traditional review session may have enhanced calibration and performance because it provided a more coherent and comprehensive overview of the material rather than a more limited focus on the structure and specific content of the multiple-choice items (Bol & Hacker, 2001). The whole class discussions that may have developed as a result of the traditional review sessions may have fostered metacognitive analysis by the students as a group thus allowing them to better individually calibrate their level of knowledge of the test material.

Hacker, et al. (2000) investigated the effects of practice tests on calibration as well. They conducted a semester long study of the calibration accuracy of undergraduate psychology students who took practice tests prior to taking three exams given during the semester. The participants were told to use their performance on the practice tests to gauge how well they knew the exam material. Immediately prior to taking each exam the participants made predictions of how many questions they would get right and immediately after the exams they made postdictions of how many questions they believed they answered correctly for each of the three exams. After each of the first two exams the students were made aware of their accuracy level and told to evaluate what they could do differently in order to improve their calibration accuracy for the final exam. Hacker, et al., found that the high achieving students were more accurate grade predictors for both predictions and postdiction scores. The average students were more accurate for postdiction, but not for prediction scores. Low achievers were the most inaccurate of all the achievement groups, and could not accurately predict or postdict their scores.

One important finding from this research was the fact that the high achievers' prediction and postdiction evaluation skills improved over the three exams, whereas low

achievers' evaluation skills did not (Hacker, et al., 2000). There appears to be a gap in low achievers metacognitive development that hinders them from improving on calibration accuracy. Perhaps the high achievers are so because they possess and are aware of the metacognitive ability to successfully calibrate their level of knowledge. More research into ways to improve calibration is needed in order to help the low achieving student reach higher achievement levels.

Peer interactions may play a role in the development of self-regulating techniques. Peers bring new insight into the learning process as they discuss various methods of learning material. Students often place more value on information that they hear from their peers than they do information that they hear from teachers. Hence, peers offer an avenue of enhancing self-regulation that can be incorporated into the classroom.

In order to investigate the effects of peer interactions on self-regulation Orange (1999) investigated the effects of peer delivered self-regulation techniques on 63 college level psychology students ranging in age from 19 to 56. The students were divided into two groups with one group receiving self-regulating intervention in the form of a video outlining twelve steps to self-regulation and the other group receiving no intervention. Prior to viewing the 25-minute video the students gathered as a group and shared and discussed their academic problems. The video was produced using peer models to teach self-regulation and took the students step-by-step through self-regulation techniques.

A self-regulating instrument (SRI) was developed for use as a pretest and posttest of the level of self-regulation practiced by members of both groups. The questions focused on the students' perceptions of their self-regulating techniques including the ability to accurately calibrate their level of knowledge (including: assessing one's



progress or performance against a goal; using metacognitive strategies to assess one's performance and needed action; critically assessing behaviors, attitudes, and actions in terms of personal standards) (Orange, 1999). Both groups completed the SRI as a pretest and then again after the intervention as a posttest.

Orange found a statistically significant difference in the performance of the two groups on the SRI pretest and posttest, suggesting that the peer-models were successful in improving self-regulation. It is evident that there is a need for teaching self-regulation and calibration, and for more investigations into the effectiveness of using group work with peers for fostering these skills.

To better understand the influence of social interaction on cognition, Lundeberg and Moch (1995) investigated the effect of supplemental instruction aimed at encouraging students to "think aloud" as they calibrate their knowledge level as a group on their ability to accurately calibrate their own level of knowledge. The qualitative study involved nursing students enrolled in a two-semester health science course. The supplemental instruction was carried out after the regularly scheduled class meetings. The meetings were facilitated by graduate students trained as to how to encourage cognitive learning aspects including confirming the capacity for learning, calibrating learning, and connecting learning to academic success (Lundeberg & Moch). Probing questions were used to help the participants calibrate their level of knowledge and adjust their thinking towards the successful attainment of health science knowledge.

Results showed that the collaborative group discussions influenced cognitive reactions. One such reaction was a process of calibration in which students assessed their own and others' knowledge and advanced one another's thinking (Lundeberg & Moch).

The group as a whole seemed to facilitate each member's ability to accurately calibrate their individual level of obtained knowledge.

“Calibration occurred in conversations which included a great deal of probing, building on initial responses, and encouraging other students' explanations. The supplemental instruction leader both modeled the kind of discussion she expected and continually probed the students to challenge and stimulate further thinking.”  
(pp. 322).

In lecture classes there is rarely time for students to develop appropriate questions due to the limited time and the focus on writing down as much material as possible during the lecture process. By allowing time for students to calibrate their level of understanding, the student can develop questions that will foster his or her understanding of material. This process is so rarely used that most students find it difficult to proceed through material with cognitive awareness of their own level of learning.

In order to investigate the effects of self-regulatory study strategy training on reading achievement, Nelson and Manset-Williamson (2006) compared guided instruction in self-regulation with explicit instruction in self-regulation in order to see which method was more effective for improving the accuracy of calibrating reading skills. The study consisted of 21 students in grades 4 – 8 who were identified to be at least 2 years behind in their reading achievement levels. The participants were divided into two groups consisting of a guided reading group and an explicit reading group. The guided reading group participated in a method of self-regulation strategy training that modeled comprehension strategies without direct instruction of these strategies. The explicit reading group participated in a method of self-regulation strategy training that

included direct instruction as to how to self-regulate learning. Feedback about the correct strategy use and reading outcome was provided for the explicit reading group but not for the guided reading group.

In order to determine reading calibration ability the participants completed a reading self-efficacy measure in which they read passages and then rated their perceived ability to answer questions about the passages correctly. After making these predictions, they were asked specific questions about the passages in order to evaluate their calibration accuracy. Participants in the explicit reading group were more accurate in calibrating their level of reading comprehension skills than those in the guided reading group.

The participants were also presented with scenarios in which students had failed to accurately calibrate reading comprehension and asked to explain to the researcher what strategies the students in the scenarios may have failed to use correctly in their calibration of reading comprehension. Participants in the explicit reading group made greater gains in attributions to incorrect strategy usage when presented with reading failure scenarios than did participants in the guided reading intervention (Nelson & Manset-Williamson, 2006). These findings could be attributed to the fact that the explicit reading group had to monitor their own strategy use as part of their training.

“Compared to the more fluid and teacher-controlled instruction of the Guided Reading intervention, the Explicit Reading intervention was more rigorous, explicitly calling upon students - after explicit instruction, modeling, and practice - to take control of their strategy usage, set their own goals for reading, and monitor their strategy usage and understanding.” (pp. 226)

Walck, Bol, Hager, & Mckinney (2009) investigated the effects of setting (group or individual) and guidelines (with or without) on calibration accuracy and achievement. The results showed that students calibrating in groups were more accurate at predictions than were those students calibrating individually. In fact, those students calibrating in groups ( $M = 5.40$ ) were more than twice as accurate in their predictions as those students who were calibrating individually ( $M = 11.12$ ). In addition, students calibrating with guidelines were more accurate at predictions than those students who did not use guidelines. Low achieving students with guidelines were found to be more accurate at predictions than were low achieving students calibrating without guidelines. Students calibrating with guidelines in groups were much more accurate in their predictions than were those students calibrating in groups without guidelines. For students calibrating individually the guidelines made little difference in their calibration accuracy. Perhaps most importantly, guidelines and group calibration were linked to significantly higher achievement scores.

Similar to Riggs, et al. (2009), Walck, et al. (2009) also found that high achievers were more accurate at predictions than were low achievers. Lower achieving students who calibrated with guidelines were more accurate in their predictions than were lower achieving students who calibrated without guidelines. Higher achieving students showed little difference in their accuracy regardless of receiving guidelines or not. No significant differences were found for postdiction accuracy. This study suggests that the use of guidelines can be an effective method for increasing student metacognitive processes and therefore increasing calibration accuracy and achievement.

### Summary of Studies

The conclusions of the literature review suggest that research in calibration is lacking - and that this is especially true for the high school level. In addition, few studies in the knowledge of metacognition, or calibration of learning have been conducted. Those that have been conducted have shown that high achieving students are more aware of their metacognitive processes (Carvalho & Yuzawa, 2001; Dahl, 2004), but have failed to fully explain this phenomenon. The research suggest that there is still much to be learned about how to improve calibration accuracy and incorporate this metacognitive construct successfully into the educational experience.

Previous studies in calibration accuracy were primarily studies at the college level (Bol & Hacker, 2001; Bol, Hacker, O'Shea & Allen, 2005; Grimes, 2002; Hacker, Bol & Bahbahani, 2008; Hacker, Bol, Horgan & Rakow 2000; Lundeberg & Moch, 1995; Nietfeld & Schraw, 2002; Schraw, 1997). There is clearly a lack of studies at the high school level, and none could be found that look at the effect of group work and guided questions on calibration accuracy in high school biology classes. The proposed research will help to fill the gap in this area of calibration investigation.

### Rationale for Study

It is obvious that more research in calibration is needed in order to understand how this metacognitive skill can be used to help increase academic performance. Many students lack the ability to self-regulate their learning and accurately calibrate their understanding of material. This leads to continued misinterpretation of the level of knowledge obtained and the lack of development of effective study strategies. It is

important that educators recognize this deficit and incorporate activities that can help students develop accurate calibration skills into the curriculum. Previous research indicates that the degree of instructor support for the development of cognitive strategies affects the frequency of cognitive monitoring (Justice & Dornan, 2001).

In order to increase the level of cognitive modeling experienced by students, teachers need to be introduced to strategies that can help them address and improve metacognitive processes. Research into the effectiveness of strategies that can be used to enhance calibration skills is important in order to identify those strategies that are the most effective for improving calibration accuracy. The development of calibration accuracy has the potential to positively impact academic achievement.

Improving calibration skills could result in equal levels of improvement in students' self-efficacy beliefs. Students with high self-efficacy have been shown to be more successful in many areas of interactions. Self-efficacy perceptions influence choice of activity, task perseverance, level of effort expended, and ultimately, degree of success achieved (Klassen, 2002). Therefore, improving calibration skills can in turn improve ones' self-efficacy, and improving ones' self-efficacy can improve ones' calibration skill. Since both calibration and self-efficacy are related to self-regulation, improvements in both areas are linked to the improved ability to self-regulate. How individuals interpret the results of their performance attainments informs and alters their environments and their self-beliefs, which in turn inform and alter their subsequent performances (Pajares, 1996).

Teaching students how to self-regulate should be a part of their educational experience. Since tests remain the major measure of students abilities, educators should

expose students to all possible techniques that can help them achieve success on these tests. Effective test taking depends on two important skills: selecting correct responses to test questions and monitoring one's performance accurately (Schraw, 1997). A major goal of formal education should be to equip students with the intellectual tools, self-beliefs, and self-regulatory capabilities to educate themselves throughout their lifetime (Bandura, 1993). Calibration ability is an important tool that is too often overlooked in the educational experience.

Researchers are now convinced that metacognitive beliefs, decisions, and actions are important, but frequently overlooked, determinants of success or failure in a wide variety of activities (Garofalo & Lester, 1985). There is little question that studies into calibration skills need more focus in order to understand how to best incorporate effective metacognitive tools into students' repertoire of successful study strategies. There are few areas of interaction with one's environment in which metacognition is not involved – and the ability to calibrate one's level of understanding is vital if one is to accurately pursue success in many varied activities. Teaching calibration skills in high school and even earlier may allow students to be successful in any future endeavors as they develop skills to calibrate their level of understanding.

Previous studies in metacognition and calibration at the college level have supported the need for more practice in and exposure to techniques that can improve calibration skills earlier in the educational experience (Schraw, 1997; Bol & Hacker, 2001; Carvalho & Yuzawa, 2001; Grimes, 2002; Hacker, Bol, & Bahbahani, 2008). The development of accurate calibration skills prior to the college experience could help students experience academic success in their college courses. If college students fail to

accurately calibrate their level of knowledge, they are at greater risk of failing courses.

Calibration accuracy at the college level is of great importance due to the fact that many classes are passed or failed based on the scores achieved on a small number of assessments. Since one of the main missions of high school is to adequately prepare students for success in college, improving calibration skills will only help the student as he enters into college and later into the workforce. To be competent and motivated to “know how you know” puts one in charge of one’s knowing, of deciding what to believe and why, and of updating and revising those beliefs as one deems warranted (Kuhn, 1999).

Group calibration with the use of guidelines has the potential to improve calibration accuracy and achievement. When students are guided through the process of calibration in group settings they have the additional benefit of learning from their peers. Peers bring new insight to the metacognitive process and may help in the development of successful calibration skills.



## Research Questions and Hypotheses

Whether calibration and other metacomprehension strategies can be improved with instruction is a question that has not yet been definitively answered (Bol & Hacker, 2001). In an attempt to answer this question, this research investigated the effects of calibration practice in either group or individual settings and with or without guidelines on the calibration accuracy of a high school biology course on test scores. The following research questions were addressed:

1. Does receiving guidelines during calibration practice improve calibration accuracy and achievement for high school biology students?
2. Is calibration practice in groups more effective than individual practice in improving calibration accuracy and achievement for high school biology students?
3. How do guidelines and learning settings (group vs. individual) interact to affect calibration accuracy and achievement?
4. What do students write in response to guided questions designed to improve calibration?

It was hypothesized that the students who received guidelines for calibration skills would score higher on tests than those who did not receive guidelines. It was further hypothesized that those students who engaged in group calibration would score higher on tests than those who engaged in individual calibration. In addition, it was explored whether there would be an interaction between guidelines and learning settings on calibration accuracy and achievement.

## Summary

Calibration is the accuracy with which students can predict their own performance (Horgan, 1990). It can be investigated by having students predict what they will score on a test both immediately prior to and after the test is administered. These predictions can be compared to actual scores obtained to interpret the level of calibration accuracy. The examination of the use of group work and guidelines in the calibration process may prove to have the ability to enhance calibration skills. With enhanced calibration skills students can more successfully monitor their learning processes.

This chapter has provided an overview of metacognition, self-regulation, and calibration as they pertain to the educational setting. Metacognition is understood as regulation of cognition including the planning before one begins to solve a problem and the ongoing evaluation and control during the problem solving and learning (Dahl, 2004). This understanding can refer to ones' calibration of performance before (prediction) and after (postdiction) an academic assessment. Self-regulation of learning is vital for success and accurate calibration skills enhance the ability to successfully self-regulate learning. Research suggests that metacognitive skills can be taught and can subsequently improve academic achievement (Hartley, 2001).

Chapter II has supported the need for calibration skills and outlined how group work and guidelines may enhance calibration accuracy. When a discrepancy exists between current and desired performance, self-regulated learners seek feedback from outside sources such as peers' contributions in collaborative groups (Butler & Winne, 1995). Guidelines to facilitate the calibration process help to maintain the focus of the group conversation on the metacognitive process. This can in turn improve the student's

calibration accuracy. The importance of accuracy in calibration has been presented and defended. This chapter has shown that there is a lack of empirical research investigating the use of calibration in educational settings, specifically in high school science courses. Research into calibration in group settings is also lacking.

The methodology for investigating the research questions and hypotheses that were presented for this research are outlined in Chapter III. Chapter III provides specific details concerning the methods and conceptual framework, participants, data collection and analysis and timeline for the study.

## CHAPTER III

### METHODOLOGY

#### Introduction

This chapter outlines the methodology used to evaluate the effect of individual or group guidelines on the calibration accuracy of high school biology students. It begins with a detailed description of the participants followed by the design, measures, and procedures that were used to carry out the study. Protection of participants' privacy were addressed.

#### Participants

Participants in this study included high school students enrolled in biology classes at a suburban high school with a population of diverse ethnic backgrounds. The students were distributed between four MYP (International Baccalaureate Middle Years Program) biology classes that were all taught by the same instructor. The students enrolled in the classes were all 14-15 years of age and included 53 females and 49 males ( $N=102$ ). The ethnic diversity of the participants consists of 53% Caucasian, 26% Asian, 11% African American, 4% Indian, and 6% other.

This convenience sample was selected for this study from Princess Anne High School due to the fact that the researcher teaches at this school. The specific biology classes were purposefully selected in order to ensure that the same instructor would teach all classes in the same manner. However, the teacher was not the instructor for these classes.

### Design

This research employed a quasi-experimental factorial design. The independent variables were the type of calibration used (individual vs. group) and whether calibration guidelines were provided (group guidelines vs. individual guidelines). The dependent variables were calibration accuracy and achievement of the subjects. The four intact classes were randomly assigned (by use of a blind drawing) to one of four groups as shown below:

	Calibration Guidelines	No Calibration Guidelines
Group Calibration	MYP Biology – Class 1	MYP Biology – Class 2
Individual Calibration	MYP Biology – Class 3	MYP Biology – Class 4

Control variables included the use of the same teacher for instruction, the same unit of investigation, the same method of coverage of material presented, the same amount of time allotted for instruction and review, the same assessment methods, the same amount of time allotted for calibration, and the same classroom setting. Half of the students (classes 1 and 3) were given guidelines for calibration, and half of the students (classes 2 and 4) were not given guidelines for calibration.

Class 1 and class 3 were both given guidelines for calibration. Class 1 was given guidelines for calibration within groups. Class 3 was given guidelines for individual calibration. Both classes were provided with the calibration guidelines (see Appendix A) immediately prior to the time allotted for calibration. Both classes stopped their group

review at the mid-time and were told to reflect (for five minutes) on their understanding of the review material. Class 1 reflected as a group using the group calibration guidelines and Class 3 reflected individually using the individual calibration guidelines. Each group in class 1 selected a member to record the group responses to the guideline questions. Class 3 recorded responses to their individual guideline questions. Both groups then returned to the group review activity for the remainder of the time. Both groups were given 40 minutes to review the material, which made the mid-time 20 minutes into the review activity. The teacher collected the guideline question responses from each group in Class 1 and from each individual in Class 3.

Class 2 and class 4 calibrated their performance without guidelines. Class 2 calibrated their knowledge as a group, and class 4 calibrated their knowledge individually. Class 2 stopped their group review activity at the mid-time and were told to collectively reflect (for 5 minutes without guidelines) on their groups understanding of the material. Class 4 stopped their group review activity at the mid-time and were told to individually reflect (for 5 minutes without guidelines) on their individual understanding of the material. They then returned to the group review activity for the remained of the time. Both groups were given 40 minutes to review the material, which made the mid-time 20 minutes into the review activity.

### Measures

Comparing predicted and postdicted scores with the actual test scores and exam score achieved determined calibration accuracy. Predicted scores were subtracted from postdicted scores in order to determine each student's level of accuracy. Students were asked to predict how many points they will earn out of 100 possible points (1-100). (See

Appendix B) Predicted and postdicted scores were subtracted from actual scores in order to determine each students' level of accuracy. Absolute values, rather than signed differences were used in the analyses. For example if a student predicted that he or she will receive a 90 on the test but actually received a score of 82 the accuracy score would be 8. Therefore, lower scores represented better accuracy.

The student scores on the tests determined achievement level. The three tests consisted of both multiple choice and short answer questions. All questions came from previously released International Baccalaureate exams. Scores were compared across all four groups for each of the three tests as well as for the overall accuracy for the combined averages of all three tests.

Qualitative data collected from responses of the calibration group with guidelines and responses from the individuals calibrating with guidelines were analyzed via content analysis. This data gave insight into the effectiveness of the guided questions in facilitating metacognitive thought and calibration accuracy.

#### Procedure

Prior to the beginning of this study approval was obtained from the City of Virginia Beach Public Schools, and the principal of Princess Anne High School. Participants and their parents were informed about the nature of the research. A letter was sent home to all participants in order to obtain parent approval (see appendix C). The students were informed of the research and provided with explanations of calibration. Class 1 and 3 received calibration guidelines- class 1 as a group and class 3 individually. The calibration guidelines are found appendix A. Class 1 implemented the calibration guidelines during group review for each test. The group responses to the

guideline questions were collected by the teacher at the end of each review session. Class 3 implemented the calibration individually with the guidelines after group review for each test. Class 2 was asked to calibrate their level of understanding of the material as a group without guidelines and class 4 was asked to calibrate their level of understanding on an individual basis after group review without guidelines.

The groups rotated in terms of membership and the group membership consisted of mixed ability students. Even though there was little difference between the students' ability, they were assigned to groups based on previous test scores. The students were divided into three achievement categories: high achievers, average achievers, and low achievers. Two students from each achievement category were randomly selected to form the collaborative groups within each class. The random selection took place by a blind drawing from each achievement category. Approaches that rely on the grouping of high and low achievers have been shown to be effective in producing learning gains relative to more traditional forms of classroom instruction (Gabriele & Montecinos, 2001). The random assignment of the groups was repeated for each review activity. Each group consisted of five or six students. The calibration activities took place in the class period before the administration of the test.

On the day of the test each student individually predicted what he/she thought he/she would score on the test (based on 100 possible points). They recorded their prediction on the student calibration sheet (see Appendix B). They then took the test and immediately after taking the test they made a prediction on how they thought they did on the test. They recorded their predictions on the student prediction calibration sheet, which was attached to the back of each assessment item. The accuracy of each class was



calculated and compared in order to answer the research questions.

During the review activity, all groups were told to spend the allotted time (40 minutes) reviewing the material that was on the assessment. All lecture, laboratory investigations, and activities carried out during the unit of study were the same for all groups. The teacher monitored the groups during the review in the same manner, and the assessments were identical in content and structure.

### Summary

Chapter III has outlined the methodology that was used for this study. It has presented details of the participants, the procedures, and the measures that were used for the study. Chapter IV will present the data collected during the study and the results of the statistical analysis of the data. In addition, Chapter IV will present qualitative data collected from the students responses to the guiding questions (both group and individual) used during the calibration activities.

## CHAPTER IV

### RESULTS

#### Introduction

This chapter presents the results of the analyses used to evaluate the effectiveness of guidelines (with or without) and setting (group or individual) on the calibration accuracy and achievement level of high school biology students. It begins with a presentation of descriptive results for overall prediction and postdiction accuracy, and overall average test scores. Next the multivariant analyses of results for each of the three individual tests are presented. This will determine the effects of treatment on each of the dependent variables associated with each test. It follows with a description of the results for follow-up analyses of variance for each the three tests. In addition, descriptive statistics are reported to interpret the findings. Qualitative data is presented in order to analyze how students respond to guiding questions during calibration.

Actual test scores were compared with students' predictions and postdictions. Prediction and postdiction accuracy was determined by calculating the absolute difference between the students' predicted score and their actual score. Lower scores represented better accuracy since they deviated the least from the actual scores. Means and standard deviations were calculated for achievement, prediction accuracy, and postdiction accuracy. In order to investigate the effects of calibration practice and guidelines on calibration accuracy and achievement a factorial multivariant analysis (MANOVA) was run. The independent variables were guidelines (with or without) and setting (group or individual). Dependent variables included overall prediction accuracy, overall postdiction accuracy, and actual score achieved. Analyses for each of the three

tests for prediction accuracy, postdiction accuracy, and actual grade received were conducted. Follow-up factorial analyses of variability (ANOVA) were used in order to determine main effects of treatment and any interactions between treatments.

Due to teacher error in the delivery of the guided questions all qualitative data for test 1 was discarded. Guided questions used for this test included questions that were generated before being altered based on feedback from the dissertation committee. The responses to the modified guided questions were analyzed for both test 2 (genetics) and 3 (DNA and Technology). The guiding questions included two Likert-style close-ended questions with four choices for response (not at all confident; somewhat confident; confident; extremely confident) and three open-ended questions.

Responses to the data collected from responses to the Likert-type questions were compared across response categories and quantitative data was recorded for percent of students selecting each level of confidence for both group and individual calibration. Qualitative data collected from responses to the open-ended guiding questions for group and individual calibration conditions were analyzed via content analysis. Themes were developed for related responses. The researcher and another doctoral student trained in qualitative methods independently coded 20 percent of the data into related themes. The coding of the data can be considered reliable since the researcher and the doctoral student reached 96 percent agreement. Once the data was coded, patterns and relationships between responses were identified and reported. This data gives insight into the effectiveness of the guided questions in facilitating metacognitive thought and calibration accuracy.

### Overall Descriptive Results

In order to begin to examine prediction accuracy, postdiction accuracy, and overall achievement (actual score obtained) the means and standard deviations were calculated for each tests across all four groups and for overall tests results (see Table 1). The mean scores across all groups and tests were very similar ( $M = 90.5$  for test 1;  $M = 91.0$  for test 2;  $M = 90.7$  for test 3;  $M = 91.0$  for overall). It appears that all tests were fairly equal in difficulty when looking at combined group scores. On average, the students scored very well on these tests.

Prediction accuracy varied little on each test, and did not appear to improve across tests. Students' prediction accuracy was best on test 1 ( $M = 5.7$ ) and most inaccurate with test 2 ( $M = 6.8$ ). Prediction accuracy improved again with test 3 ( $M = 6.1$ ). Overall students were fairly accurate in their predictions with actual scores varying an average of about six points from their predicted scores.

Students' postdictions were slightly more accurate than their predictions. Postdiction accuracy for test 1 ( $M = 4.8$ ), test 2 ( $M = 4.1$ ), and test 3 ( $M = 4.6$ ) fluctuated little across the three tests. Students were most accurate in their postdictions for test 2 and the least accurate for test 1. On the average, students' actual scores deviated less than five points from their postdiction scores.

Table 1

*Overall Descriptive Statistics for Actual Score, Prediction Accuracy, and Postdiction Accuracy*

	Mean	SD
<u>Test 1 (n = 102)</u>		
Actual Score	90.5	6.2
Prediction Accuracy	5.7	5.4
Postdiction Accuracy	4.8	5.1
<u>Test 2 (n = 101)</u>		
Actual Score	91.0	6.9
Prediction Accuracy	6.8	4.1
Postdiction Accuracy	4.1	3.4
<u>Test 3 (n = 102)</u>		
Actual Score	91.0	6.3
Prediction Accuracy	6.1	4.7
Postdiction Accuracy	4.6	4.2
<u>Overall (n = 305)</u>		
Actual Score	90.7	6.5
Prediction Accuracy	6.1	4.7
Postdiction Accuracy	4.6	4.2

### Statistics for Individual Tests Results

#### *Results for Test 1*

The MANOVA results for the first test showed a significant main effect for setting, Wilks' Lambda =  $F(3,96) = 4.76, p < .00, \eta^2 = .129$ , and a significant interaction between setting and guidelines, Wilks' Lambda =  $F(3,96) = 5.45, p < .00, \eta^2 = .145$ . No significant main effect was found for guidelines, Wilks' Lambda =  $F(3,96) = 1.90, p < .14, \eta^2 = .056$ . The results for the MANOVA are presented in Table 2.

Table 2

*Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test Performance for Test 1*

	<i>F</i>	<i>df</i>	<i>Sig.</i>	$\eta^2$
Setting	4.76	3, 96	.00	.129
Guidelines	1.90	3, 96	.14	.056
Setting x Guidelines	5.45	3, 96	.00	.145

*N* = 102

Since the MANOVA showed a significant main effect for setting and a significant interaction between setting and guidelines, follow-up analyses focused on these findings. Follow-up ANOVA results revealed a significant main effect on the achievement test,

$F(1,98) = 8.65, p < .01, \eta^2 = .081$ . In addition a significant interaction was revealed for setting and guidelines on achievement,  $F(1,98) = 9.0, p < .03, \eta^2 = .084$ . The ANOVA results are presented in Table 3.

Table 3

*Follow-up ANOVA results for Treatments on Calibration Accuracy and Test Performance for Test 1*

Source	Dep. Var.	df	<i>F</i>	Sig.	$\eta^2$
Setting	Achievement	1, 98	8.65	.01	.081
	Pred. Accuracy	1, 98	.037	.85	.000
	Post. Accuracy	1, 98	1.08	.24	.014
Guidelines	Achievement	1, 98	5.16	.01	.061
	Pred. Accuracy	1, 98	2.31	.13	.023
	Post. Accuracy	1, 98	1.41	.24	.014
Setting x Guideline	Achievement	1, 98	9.00	.00	.084
	Pred. Accuracy	1, 98	.714	.40	.007
	Post. Accuracy	1, 98	1.14	.30	.011

$N = 102$

Those students who calibrated in groups scored higher on the test than students who calibrated individually. The mean test score for students calibrating in groups was 93.0 and for students calibrating individually 89.6, a difference of 3.4 points. The means and standard deviations for students calibrating in groups and students calibrating individually were analyzed and are presented in Table 4. This data suggests that group calibration increases achievement more so than does individual calibration as measured by test scores. However, the difference is not large.

Table 4

*Means and Standard Deviations for Main Effect for Setting on Achievement for Test 1*

	<i>n</i>	Mean	SD
Group	50	93.0	5.3
Individual	52	89.6	5.8

*N* = 102

Interactions between the effects of setting and guidelines on achievement level for students calibrating in groups and students calibrating individually are presented in Table 5. The mean test score for students calibrating in groups with guidelines was 92.0 and without guidelines was 93.0. The mean test score for students calibrating individually with guidelines was 92.0 and without guidelines was 87.0. The data indicates that



guidelines do not significantly impact achievement for students calibrating in groups but significantly improve achievement for students calibrating individually. Figure 1 further displays the interactions of setting and guidelines on achievement.

Table 5

*Means and Standard Deviations for the Interaction of Setting and Guidelines on Achievement for Test 1*

	Individual			Group		
	n	Mean	SD	n	Mean	SD
No Guidelines	25	87.0	5.3	28	93.0	5.9
Guidelines	27	92.0	5.2	22	92.0	4.5

$N = 102$

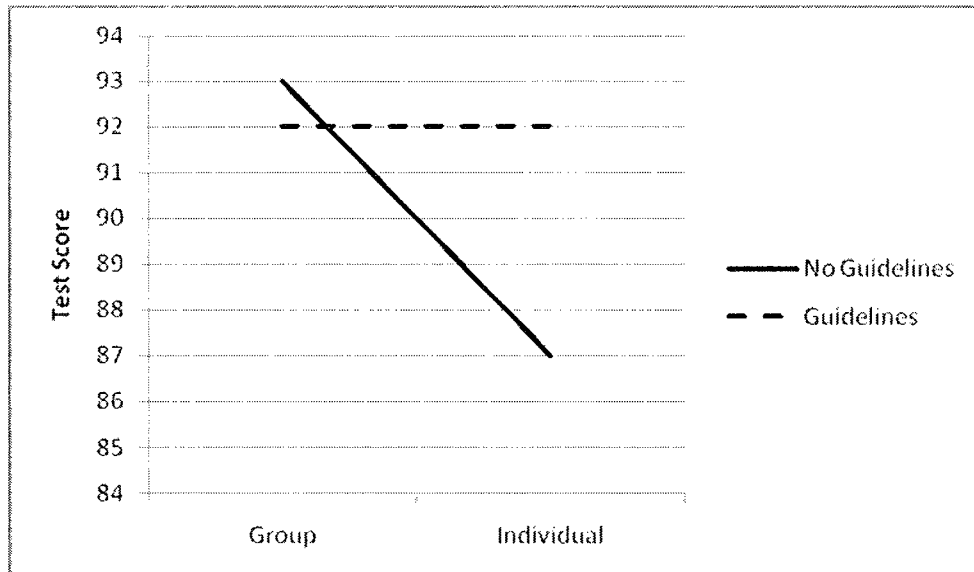


Figure 1. Interaction for Test 1 for achievement (test score).

### Results for Test 2

Test 2 revealed significant main effects for setting, Wilks' Lambda =  $F(3,98) = 6.04, p < .00, \eta^2 = .159$ , and for guidelines, Wilks' Lambda =  $F(3,98) = 3.43, p < .02, \eta^2 = .097$  (see Table 12). In addition, a significant interaction was seen for setting and guidelines, Wilks' Lambda =  $F(3,98) = 3.83, p < .01, \eta^2 = .107$ . The MANOVA results for test 2 are reported in Table 6.

Table 6

*Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test Performance for Test 2*

	<i>F</i>	df	Sig.	$\eta^2$
Setting	6.04	3, 96	.00	.159
Guidelines	3.43	3, 96	.02	.097
Setting x Guidelines	3.83	3, 96	.01	.107

*N* = 102

Follow-up ANOVA results showed that setting significantly impacted achievement,  $F(1,98) = 16.0, p < .00, \eta^2 = .140$ , but not prediction or postdiction accuracy. Guidelines were found to be significant for achievement,  $F(1,98) = 7.35, p < .01, \eta^2 = .070$ , prediction accuracy  $F(1,98) = 3.9, p < .05, \eta^2 = .038$ , and postdiction accuracy,  $F(1,98) = 4.57, p < .04, \eta^2 = .045$ . In addition, a significant interaction for setting and guidelines was seen for achievement,  $F(1,98) = 11.4, p < .00, \eta^2 = .104$ , and for postdiction accuracy,  $F(1,98) = 4.37, p < .04, \eta^2 = .043$ . Follow-up ANOVA results are presented in Table 7.

Table 7

*Follow-up ANOVA results for Treatments on Calibration Accuracy and Test Performance for Test 2*

Source	Dep. Var.	df	<i>F</i>	Sig.	$\eta^2$
Setting	Achievement	1, 98	16.0	.00	.140
	Pred. Accuracy	1, 98	1.48	.23	.015
	Post. Accuracy	1, 98	1.21	.27	.012
Guidelines	Achievement	1, 98	7.35	.01	.070
	Pred. Accuracy	1, 98	3.86	.05	.038
	Post. Accuracy	1, 98	4.56	.03	.045
Setting x Guideline	Achievement	1, 98	11.4	.00	.104
	Pred. Accuracy	1, 98	.223	.64	.002
	Post. Accuracy	1, 98	4.37	.04	.043

*N* = 102

Setting (group or individual) significantly impacted test scores. The mean test score for students calibrating in groups was 93.6 and for students calibrating individually 88.6, a difference of 5 points. It appears that group interactions during calibration activities increases test scores. The means and standard deviations for the main effect of setting are displayed in Table 8.

Table 8

*Means and Standard Deviations for Main Effect of Setting on Achievement for Test 2*

	<i>n</i>	Mean	SD
Group	50	93.6	5.3
Individual	52	88.6	7.8

*N* = 102

Guidelines were significant in increasing achievement, prediction, and postdiction accuracy. For students calibrating with guidelines the average test score was 92.7 and without guidelines 89.6, a difference of 4.1. It appears that the use of guidelines during calibration activities increases achievement. The means and standard deviations for the main effects of guidelines on achievement are presented in Table 9.

Table 9

*Means and Standard Deviations for Main Effect of Guidelines on Achievement for Test 2*

	<i>n</i>	Mean	SD
Guidelines	49	92.7	5.8
No Guidelines	53	89.6	7.9

*N* = 102

In addition, guidelines were also linked to better calibration accuracy. Prediction accuracy for students calibrating with guidelines was 5.6 and without guidelines was 7.1, a difference in accuracy of 1.5 absolute points. Postdiction accuracy for students calibrating with guidelines averaged 3.5 and without guidelines 5.0, a difference in accuracy of 1.5 absolute points. These results suggests that the use of guidelines can be effective not only in increasing test scores, but promoting metacognitive (calibration) skills as well. The means and standard deviations for prediction accuracy and postdiction accuracy are presented in Table 10.

Table 10

*Means and Standard Deviations for Main Effect of Guidelines on Prediction and Postdiction Accuracy for Test 2*

	Prediction Accuracy			Postdiction Accuracy	
	<i>n</i>	Mean	SD	Mean	SD
Guidelines	49	5.6	3.4	3.5	3.0
No Guidelines	53	7.1	4.2	5.0	3.5

*N* = 102

Interactions of setting and guidelines were significant for achievement and postdiction accuracy. The mean test score for students calibrating in groups with guidelines was 93.1 and without guidelines was 94.0, a difference of only .9 points. The mean test score for students calibrating individually with guidelines was 92.4 and without guidelines was 84.8, a difference of 7.6 points. The data indicates that guidelines significantly impacted achievement for students calibrating individually, but had little effect on students calibrating in groups. Interactions between setting and guidelines for achievement for students calibrating in groups and students calibrating individually are presented in Table 11. Figure 2 further illustrates this interaction.

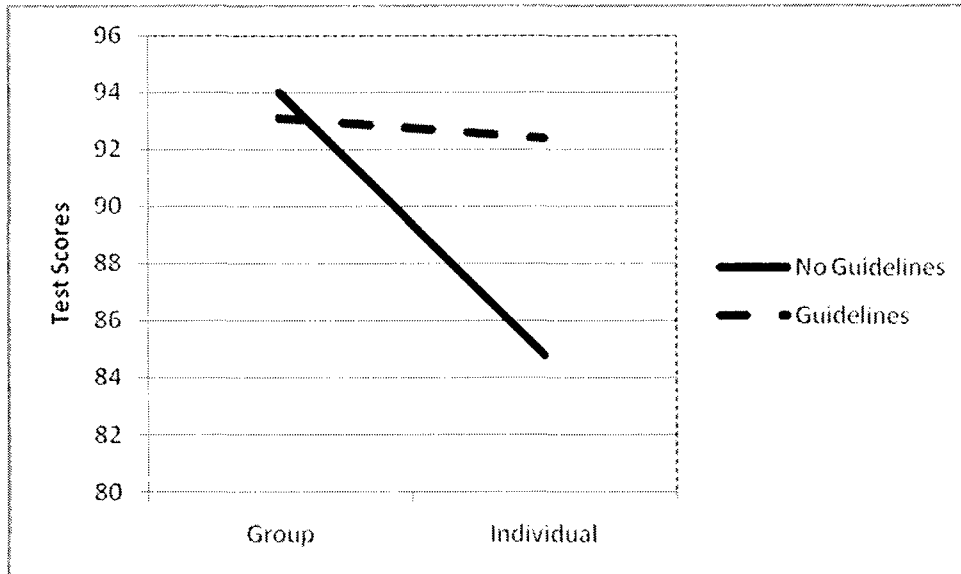
Table 11

*Means and Standard Deviations for Significant Interactions of Setting and Guidelines for Achievement on Test 2*

	Individual			Group		
	n	Mean	SD	n	Mean	SD
No Guidelines	25	84.8	5.8	28	94.0	5.5
Guidelines	27	92.4	7.4	22	93.1	5.8

$N = 102$





*Figure 2.* Interaction for achievement (test score) for test 2.

Significant findings also resulted for the interaction of setting and guidelines on postdiction accuracy. Postdiction accuracy for groups with guidelines averaged 3.7 and without guidelines 3.9, a difference in postdiction accuracy of only .2 absolute points. Postdiction accuracy for individuals with guidelines averaged 3.2 and without averaged 6.0, a difference in postdiction accuracy of 2.8 absolute points. Individuals benefitted from the guidelines for postdiction accuracy, but for groups using guidelines there were only slight increases in accuracy. Table 12 displays the means and standard deviations for the effect of setting and guidelines on postdiction accuracy. Figure 3 graphically depicts this interaction.

Table 12

*Means and Standard Deviations for the effect of Setting and Guidelines on Postdiction Accuracy for Test 2*

	Individual			Group		
	n	Mean	SD	n	Mean	SD
No Guidelines	25	6.0	4.0	28	3.9	2.9
Guidelines	27	3.2	2.9	22	3.7	3.1

$N = 102$

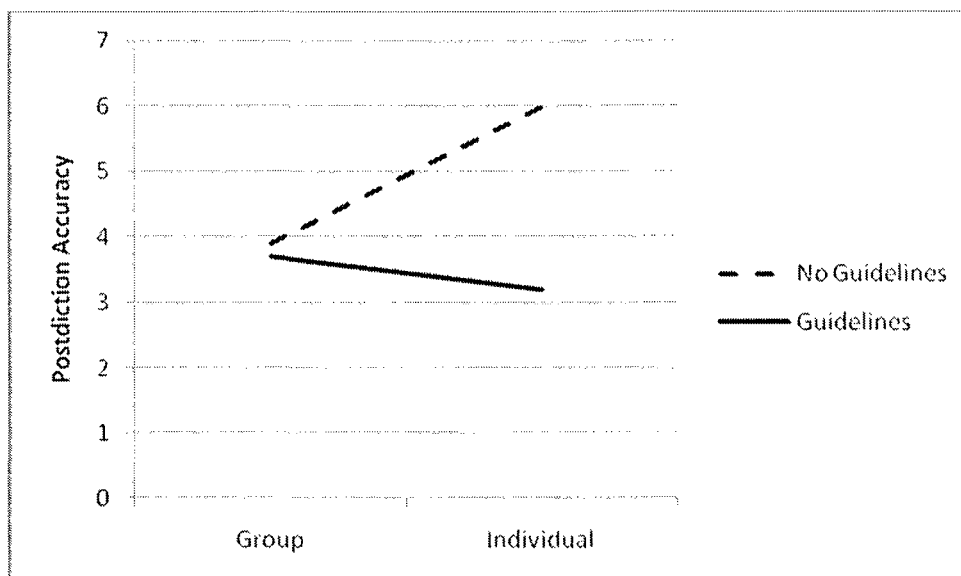


Figure 3. Interaction for postdiction accuracy for test 2.

### Results for Test 3

The only main effect revealed by the MANOVA was for setting, Wilks' Lambda =  $F(3,96) = 3.60, p < .02, \eta^2 = .101$ . The data revealed no main effect for guidelines,  $F(3,96) = 0.73, p < .12, \eta^2 = .101$ , and no overall interaction between setting and guidelines, Wilks' Lambda =  $F(3,96) = 1.98, p < .12, \eta^2 = .058$ . The MANOVA results are displayed in Table 13.

Table 13

*Results of MANOVA (Wilks' Lambda) for Treatments on Calibration Accuracy and Test Performance for Test 3*

	<i>F</i>	df	Sig.	$\eta^2$
Setting	3.60	3, 96	.02	.101
Guidelines	0.73	3, 96	.54	.022
Setting x Guidelines	1.98	3, 96	.12	.101

*N* = 102

The follow-up ANOVA results are displayed in Table 14. Follow-up analysis from ANOVA revealed that setting was significant for achievement,  $F(1,98) = 8.63$ ,  $p < .01$ ,  $\eta^2 = .081$ , and prediction accuracy,  $F(1,98) = 3.80$ ,  $p < .05$ ,  $\eta^2 = .037$ . This outcome supports the results from the first two MANOVA's, highlighting the important influence of setting on achievement. It should be noted that though the ANOVA's showed significant interactions between setting and guidelines on prediction accuracy, the omnibus MANOVA did not indicate a significant interaction.

Table 14

*Follow-up ANOVA results for Treatments on Calibration Accuracy and Test*

*Performance for Test 3*

Source	Dep. Var.	df	<i>F</i>	Sig.	$\eta^2$
Setting	Achievement	1, 98	8.63	.00	.081
	Pred. Accuracy	1, 98	3.80	.05	.037
	Post. Accuracy	1, 98	1.52	.22	.015
Guidelines	Achievement	1, 98	.045	.83	.000
	Pred. Accuracy	1, 98	2.03	.16	.020
	Post. Accuracy	1, 98	.938	.34	.009
Setting x Guidelines	Achievement	1, 98	3.71	.06	.037
	Pred. Accuracy	1, 98	4.04	.05	.040
	Post. Accuracy	1, 98	1.87	.18	.019

*N* = 102

The mean test score for students calibrating in groups was 92.5 and for students calibrating individually 89.2, a difference of 3.3 points (see Table 15). Following the pattern seen in both test 1 and test 2, this test also indicated that group calibration was effective in increasing achievement.

Table 15

*Descriptive Statistics for the Effect of Setting on Achievement for Test 3*

	n	Mean	SD
Group	50	92.5	5.8
Individual	52	89.2	7.4

$N = 102$

Prediction accuracy for students calibrating in groups was 5.0 and for students calibrating individually 6.7, a difference of 1.7 absolute points. Means and standard deviations for prediction accuracy are displayed in Table 16.

Table 16

*Descriptive Statistics for the Effect of Setting on Prediction Accuracy for Test 3*

	n	Mean	SD
Group	50	5.0	2.8
Individual	52	6.7	3.2

$N = 102$

## Qualitative Findings

Both close-ended and open-ended items were used to evaluate how students respond to guided questions during group and individual calibration. Students responded to the close-ended questions by choosing one of four options on a Likert-style scale (not at all confident; somewhat confident; confident; extremely confident). Only one response (individual) was excluded from the analysis due the inability to interpret the student's response.

Percentages of responses for each category were calculated and comparisons were made between group and individual responses for each of the two close-ended questions. Themes for each of the three close-ended questions were identified and used in coding the student responses. The analysis of the students' responses to the two close-ended and three open-ended questions for test 2 and test 3 are presented in the following sections. Due to the variation in responses between test 2 and test 3 the data is presented independently for each test.

### Confidence Level

#### *Test 2 - Genetics*

Students were asked how confident they were in their ability to answer short answer questions. With the exception of one student, all responses for both group and individual calibration were limited to somewhat confident and confident. Individuals were almost equally split between the two choices (somewhat confident 46%; confident 50%). Groups were equally split between the two categories. The data indicates that the confidence levels between students calibrating in groups and students calibrating

individually are similar.

Students reported more confidence in their ability to answer multiple-choice questions on this test. Individuals were not as confident as groups for answering these types of questions. Groups reported being extremely confident (33%) more than individuals (3%) reported being so.

The data indicates that students calibrating in groups and individually feel that they can answer both short answer and multiple-choice questions correctly. However, both groups and individuals reported stronger confidence for multiple-choice questions with groups showing higher confidence in multiple-choice questions. It is interesting to note that none of the students reported that they were not at all confident in their ability to answer either short answer or multiple-choice questions. Table 17 presents the confidence levels for multiple-choice and short answer questions for test 2.



Table 17

*Confidence Levels for Multiple-Choice and Short Answer Questions for Test 2*

	Short Answer		Multiple Choice	
	Individual (n = 26)	Group (n = 6)	Individual (n = 26)	Group (n = 6)
Not At All Confident	0%	0%	0%	0%
Somewhat Confident	46%	50%	12%	0%
Confident	50%	50%	85%	67%
Extremely Confident	1%	0%	3%	33%

*N* = 48

*Test 3 – DNA, Protein Synthesis, Genetic Engineering*

Students were asked how confident they were in their ability to answer short answer questions on the test material. Almost all students in both group and individual calibration interventions limited their responses to somewhat confident and confident. No group or individual student identified that they were not at all confident.

Students calibrating individually reported that they were confident (33%) or somewhat confident (59%) that they could answer short answer questions. Only two individuals reported that they were extremely confident (8%). Students calibrating in groups reported that they were confident (50%) or somewhat confident (50%) that they

could answer short answer questions. None of the students in calibrating in groups reported that they were extremely confident.

Similar to the data for test 2, students reported more confidence in their level of ability to answer multiple-choice questions. Groups and individuals were mainly split between confident (60% for both), and extremely confident (40% and 20% respectively). Individuals reported some responses for somewhat confidence (20%) while no groups were only somewhat confident. Again, no students reported lack of confidence (not at all confident) for their ability to answer short answer or multiple-choice questions for test 3. The level of extreme confidence reported for short answer questions was much higher for this test than for test 2. Table 18 displays the confidence levels for multiple-choice and short answer questions for test 3.

Table 18

*Confidence Levels for Multiple-Choice and Short Answer Questions for Test 3*

	Short Answer		Multiple Choice	
	Individual (n = 26)	Group (n = 6)	Individual (n = 26)	Group (n = 6)
Not At All Confident	0%	0%	0%	0%
Somewhat Confident	50%	0%	59%	20%
Confident	50%	60%	33%	60%
Extremely Confident	0%	40%	8%	20%

$N = 48$

### Factors Influencing Understanding

Several common themes emerged from responses to what made the students more or less confident about their understanding of the test material. Due to the similarities in responses between both test 2 and test 3, data was collapsed across the two testing interventions. Group and individual responses were also similar and collapsed as well. Table 19 displays response categories for factors influencing the student's confidence levels in their understanding of the test material.

Table 19

*Response Categories for What Made the Students More or Less Confident in Their Understanding of the Test Material*

Category	% of Responses
Studying	42
Content Covered	25
Reviewing/Practicing	15
Other	18

$N = 46$

The majority of the students in both group and individual settings reported that studying (42% of the responses) or the amount of material and content covered on the test

(25%) made them more confident in their understanding. One individual reported that “A large study guide is good and I need a lot of time to study”. This was echoed by the groups as well as evident from one group reporting “How much we studied and if we studied”. In addition, many students reported that having review packets and practicing in class (15%) were factors that made them more confident. For example one individual reported “Just paying attention in class and practicing problems makes me more confident”. The remaining comments (18%) did not fit into any of the existing categories and were listed under “other”. This category consisted of comments such as “No competent people could help me”, “Silly mistakes”, and “I don’t care”. Table 20 lists common responses to the what made the students more confident in their understanding.

Table 20

*Quotes From Responses to What Made the Students More or Less Confident in Their Understanding of the Test Material*

---

Studying

“I just need to study more, I haven’t reviewed all the content.”

“How much I studied and if I didn’t study.”

“A large study guide is good and we need a lot of time to study.”

“How hard I studied.”

Content

“Amount of material.”

“The material”

“So much material.”

Reviewing/Practicing

“The amount of practice with the material.”

“Just paying attention in class and practicing problems makes me more confident”

“The number of times we practiced in class and having a review sheet.”

“Practice.”

---

$N = 42$

### Areas of Strength in Understanding

Several common themes emerged for test 2 and test 3 in response strengths in understanding of the test material. Almost all responses for both tests focused on content material. Although the content was different for both tests (Test 1, Genetics; Test 2, DNA Technology) for the purpose of analysis all content specific data will be reported under “content”. In addition, group and individual responses were collapsed since similar responses were reported.

Student responses as to their areas of strength in understanding the test material were mainly mixed between focus on specific test material (70%) and type of test items (12%). However, both groups and individuals listed content as an area of strength. In fact, all six groups mentioned content as strength. Other responses were varied and included responses such as “paying attention”, “short answer”, and “visuals”. Response categories as to the areas of strength in the students understanding are found in Table 21. Table 22 follows and presents sample responses as to the areas of strength in the students understanding of the test material.

Table 21

*Response Categories for Strength in Understanding of the Test Material*

---

Category	% of Responses
Content	70
Type of Test Item	12
Visuals	6
Other	12

---

 $N = 42$



Table 22

*Quotes From Responses to Areas of Strengths in Understanding of the Test Material*

---

Content

“Punnett squares”

“Protein synthesis”

“Processes, especially labeling them.”

“Labeling DNA and we are good at protein synthesis.”

“I understand how DNA makes RNA.”

Type of Test Item

“Multiple choice.”

“Fill in the blank – with a word bank.”

“Drawing the processes.”

Visuals

“visuals”

“Pictures of things.”

---

*N* = 42

### Areas of Weaknesses in Understanding

The responses to this question were also similar for both tests, and between group and individual responses. Since this pattern was seen, data was collapsed across the two tests and for group and individual responses. The responses from students as to their areas of weakness in understanding for this test were split between specific details of the material on the test and how well they prepared for the test. Content (58%), type of test item (23%) and not studying (10%) emerged as the main themes for this question. The response categories are presented in Table 23.

Table 23

*Response Categories for Weakness in Understanding of the Test Material*

Category	% of Responses
Content	58%
Type of Test Item	23%
Not Studying	10%
Other	8%

$N = 102$

While students were reporting multiple-choice questions as an area of strength, they were reporting short answer questions as an area of weakness. This resonates with

the comments on what made them more or less confident in their knowledge. Students stated multiple-choice made them more confident and short answer made them less confident. Interestingly, both groups and individuals reported “none” once in response to weaknesses in understanding. Common responses to the what made the students less confident in their understanding are found in Table 24. Since the scores were high on both tests, it this would make sense. The results of the data analyses indicate that the students had few weaknesses.

Table 24

*Quotes From Responses to Areas of Weakness in Understanding of the Test Material*

---

Content

“Pedigrees”

“Complex Inheritance”

“Genetic engineering and DNA technology”

Type of Test Item

“Short Answer”

“Short answer questions”

Not Studying

“Not looking over the material.”

---

$N = 40$

## CHAPTER V

### DISCUSSION

#### Introduction

The purpose of the study was to evaluate the ability of guidelines and setting to improve the calibration accuracy (prediction and postdiction) and achievement level of high school biology students. This chapter begins with a discussion of the quantitative results related to the overall calibration accuracy and how guidelines, setting (group or individual) and their interactions affected calibration accuracy and achievement. In addition, qualitative data in response to the guiding questions (guidelines) are discussed. This chapter concludes with a discussion of the limitations, directions for future research, and practical applications of the findings from this study.

#### Achievement

Overall, the students scored very well on all three tests indicating that all three tests were similar in their level of difficulty. The grade distributions for each of the three tests were also similar. Most of the students received A's and B's on the tests, with only a few students receiving C's. This is not surprising since the students are all enrolled in an advanced program of study which requires high performance on entrance exams for acceptance. However, this did limit the ability of the researcher to analyze differences in achievement levels between low and high achieving students.

#### Calibration Accuracy

The descriptive statistics suggest that students were fairly accurate in their predictions and postdictions. This is not surprising since previous research supports the fact that high achieving students are generally more accurate at calibration (predictions

and postdictions) than are low achieving students. Studies have found achievement to be related to calibration accuracy. Maki, Shields, Wheeler, and Lowery (2005) investigated the effect of verbal ability on the calibration accuracy of college students. Their results support the ability of high performing students to more accurately calibrate performance. Bol and Hacker (2001) found that lower-achieving students were much less accurate in predicting and postdicting their achievement levels than were high-achieving students. In a similar study on calibration accuracy Walck, et al. (2009) also found high-achieving students to be more accurate in predicting and postdicting their test scores. However, other studies have observed inconsistencies in prediction and postdiction accuracy.

Garner (1990) found that college students exhibit inconsistencies in their metacognitive performance. Bol, Hacker, O'Shea, and Allen (2005) studied the effect of overt calibration practice on increasing calibration accuracy on quizzes and their results showed that overt calibration practice did not significantly impact calibration accuracy or exam performance. However, Bol and Hacker (2001) investigated the effect of practice tests on calibration accuracy and found that exposure to practice tests resulted in less accurate predictions and postdictions on multiple-choice questions, but significantly impacted the ability of students to accurately calibrate their performance on essay items.

Although postdictions were found to be more accurate than predictions for all testing situations, postdiction accuracy did not seem to improve across the testing occasions. Again, test two was the only testing occasion to show significant increases in postdiction accuracy. The increased accuracy of postdictions may be due to the exposure to the test (actually seeing the questions) that gives the students more focused material on which to base their postdictions. Sawyer, Graham, and Harris (1992) found similar

results in their study on the effects of strategy-development interventions. Their results showed that post-test self evaluation levels did not differ after strategy-development interventions. Walck, et al. also found postdictions to be more accurate than predictions in their study on the effects of guidelines on the calibration accuracy of high school biology students.

### Effects of Guidelines and Setting

#### *Guidelines*

The results for guidelines varied over the three testing occasions, with only test two showing significant main effects for this treatment. Main effects for test two for guidelines were seen for achievement, prediction accuracy, and postdiction accuracy. Those students who used guidelines for calibration achieved higher scores on this test than did those students who did not use guidelines. In addition, those students using guidelines were more significantly accurate in both their predictions and postdictions.

The results for test two suggest that the use of guidelines during calibration activities can be effective in increasing student achievement. The findings for this test are consistent with previous research (Frederiksen & White, 1997; White & Frederiksen, 1998) that found significant improvements in understanding of the subject matter and the inquiry process when reflective assessment prompts were used. Perhaps guidelines helped to focus the student on the metacognitive process by providing prompts to encourage productive reflection.

Davis (2003) found that students who engaged in productive reflection expanded their repertoire of ideas and identified weaknesses in their knowledge, and they were more ready and able to link and distinguish their ideas. Students who lack metacognitive

skills may have improved these skills by being guided through the calibration process. Nietfeld and Schraw (2002) found that strategy training improved performance and monitoring accuracy independent of general ability. The performance training is similar to the use of guidelines for calibration activities because the guidelines help train the student how to metacognitively reflect and evaluate his or her level of knowledge. This is supported by Carvalho and Yuzawa (2001) who investigated the effect of college students' level of knowledge of cognition on achievement. They found that students with high levels of knowledge of cognition scored significantly higher on tests than did students with low levels of knowledge of cognition. Perhaps the guidelines used in this study helped in training the students to focus on the metacognitive evaluation of their learning as well. "Thinking about one's thought – in contrast to simply engaging in it – opens up a whole new plane of cognitive operations that do not exist at a simple first-order level of cognition" (Kuhn, 1999, p. 18).

Students need more calibration practice with guidelines over more testing interventions in order to become more accurate in the calibration of their knowledge. Research has shown that metacognitive skills can be taught, but must be practiced in order to improve calibration accuracy (Hartley, 2004). One way of encouraging the development of self-regulation and interpretation is through questioning (Hacker, Dunlosky & Graesser, 1998). "Questioning opens up possibilities of meaning and thus what is meaningful passes into one's own thinking on the subject" (Gadamer, 1993, p. 375). Calibration guidelines can be one method teachers could use to help the students learn metacognitive skills in order to improve their calibration accuracy. Van Zee, and Minstrell (1997) investigated the use of questioning to guide physics students thinking



during discussions with the use of questions the researchers termed “reflective tosses”. These questions influence the students thinking by having the students evaluate what they know at different points during the discussion. The researchers found that the “reflective tosses” helped the students make their understanding of the material more clear as they monitored their own thought processes.

### *Setting*

Main effects were found for all three testing occasions for setting. In addition, setting was the only variable found to be significant for achievement across all three testing occasions. Those students who calibrated in groups consistently achieved higher scores on the tests. Research has shown that when students engage in cooperative activities learning achievement is higher. Lundeberg and Moch (1995) conducted a study investigating the effect of instructional strategy training on achievement. Students were offered the opportunity to attend supplemental instruction strategy training facilitated by their peers. Results showed that those students who attended the training achieved higher grades than those students who did not attend.

Previous research has shown cooperative learning can enhance achievement at all grade levels. Dekker, Elshout-Mohr, and Wood (2006) investigated third grade math students in cooperative learning groups and found that the interactions between the students evolved into a genuine collaboration and the opportunities for learning increased for all the children involved. Analysis of the conversations showed that the students corrected each other and helped each other not only identify mistakes but learn how to avoid them.

Cohn (1999) conducted a study investigating the effects of cooperative learning

groups on the achievement of college macroeconomics students and found that those students who participated in cooperative learning groups earned higher grades and reported greater interest in the subject matter than did those students who did not participate. Their findings are supported by Johnson, Johnson, and Smith (1998) who suggest that cooperative learning at the college level should facilitate improved academic performance.

It is evident that group work has the potential to enhance students' metacognitive processes and achievement levels. When students reflect and verbalize their learning process, there is potential for improving their learning if their metacognitive skills are further developed (Dahl, 2004). The collaborative interactions that develop in groups offer opportunities for students to evaluate their own level of understanding as well as their methods of calibrating their level of knowledge. Research confirms that learners are more effective in monitoring their knowledge levels when they receive and interpret externally provided feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kulhavy & Stock, 1989; Meyer, 1986).

Placing students in groups offers them the opportunity to seek help in a non-threatening environment. Many students may be too intimidated to seek help from peers on their own, but may do so when placed in group situations that focus on calibration of knowledge. This is supported by Orange (1999) who investigated the effect of using peer models to teach self-regulation on achievement level of students. He found that students who were taught self-regulation strategies by peers showed significant increases in test performance. Observing successful and unsuccessful peers may help the student become more aware of his or her own academic shortcomings and they may become more willing

to change their own behavior (Orange). The group setting provides the learner with feedback from external sources such as peers' contributions in collaborative groups that allows them access to criteria against which to evaluate their own level of knowledge (Butler & Winne, 1995).

Prediction accuracy was found to be significant for setting only for testing occasion three. However, it is important to note that students were fairly accurate in their predictions. Significant results were found for this testing occasion for students calibrating in groups. Students calibrating in groups were more accurate in predicting their level of knowledge than were those students calibrating individually. Students calibrating individually lack the social interaction that could help direct and stimulate calibration of knowledge. As students calibrate their level of knowledge in social settings they may develop more effective calibration techniques as they listen to the calibration processes practiced by their peers. According to Zimmerman (2008) self-regulated learning is deemed as important in social forms of learning, such as seeking help from peers. In addition, Zimmerman found that students who were high in their overall use of self-regulation strategies sought help more frequently from peers, teachers, and parents and learned more than students who did not seek help. Group interactions offer the student opportunities to evaluate and improve upon their own self-regulating strategies.

Postdiction accuracy for groups was not found to be significant for any of the three tests. This indicates that postdiction accuracy remains fairly stable over group interventions. As noted earlier, previous research has shown that postdictions are generally more accurate than predictions (Walck, et al., 2009; Hacker & Bol, 2001).

However, in a study conducted by Hacker, Bol, and Bahbahani (2008) lower- achieving students who received extrinsic rewards experienced more accurate postdictions. It is obvious that more research into methods of improving postdiction accuracy is needed.

### *Interactions*

Interactions between setting and guidelines were found to be significant for testing occasion one and two for achievement. It appears that the interaction of group work and guidelines improves achievement. For those students calibrating individually, guidelines were significant in increasing their achievement. For those students calibrating in groups, the guidelines were not found to be significant. Those students calibrating in groups may have been provided with feedback and questions from their peers that developed into their own guiding questions throughout the collaborative review process. In another study Walck, et al. (2009) found that students working in groups benefited from the use of guidelines. More studies investigating the effect of setting and guidelines on achievement and calibration accuracy may show that guidelines are beneficial to students regardless of the setting. Teachers who recognize the importance of peers to the learning process encourage and provide opportunities for personal responses and collaborative interactions (Wiseman, 2003).

A significant difference was found for test three for prediction accuracy between those students calibrating individually with guidelines and those students calibrating individually without guidelines. The results showed that those students calibrating individually with guidelines were significantly more accurate in their predictions than those students calibrating individually without guidelines. This is supported by Walck, et al. (2009) who also found that students calibrating individually with guidelines enhanced

achievement levels over those students who calibrated individually without guidelines.

The guidelines may have helped to focus the learners' attention on calibration of their knowledge by directing their thought processes. Without guidelines the student may have simply looked over the material to see if they knew the material and failed to metacognitively analyze (calibrate) their level of obtained knowledge. Thomas, Bol, Warkentin, Wilson, Strange, & Rohwer, (1993) found in their study that the teacher played an important role in prompting and impeding student engagement in productive study activities. Perhaps guidelines can serve the role of prompting for metacognitive activities.

The social interaction could have enhanced their thinking, because individuals learn to solve problems independently by first analyzing those problems with their peers (Lundeberg & Moch, 1995). In a similar study (Davis, 2003) investigating the effect of generic prompts (having students stop and reflect on their level of knowledge obtained) and directed prompts (providing hints for directing the students to stop and reflect on their level of knowledge obtained) on the understanding of material. Her study revealed that generic prompts were more effective in the development of understanding than were directed prompts. Davis suggested that the directed prompts may not have corresponded with the students' level of understanding. Her study suggests that just having students stop and reflect on their knowledge is important for improving knowledge retention.

As for postdiction accuracy, an interaction was seen only for test 2, and the accuracy fluctuated over the three tests. For test 2, individuals calibrating with guidelines were more accurate in postdiction accuracy than were individuals calibrating without guidelines. Groups working with guidelines showed little difference in postdiction

accuracy from those groups working without guidelines. Perhaps the guidelines helped focus the individuals on the calibration activity, resulting in more accurate postdictions. In general, students were more accurate in postdictions for both group and for guidelines and this may explain why no significant interaction was seen for test one and test three. Similar results were found by Sawyer, Graham, & Harris (1992) when they investigated the effect of self-regulatory strategy training on the writing success of middle school students. They found no difference in post-test scores among four different study-strategy interventions. It is obvious that more research is needed in order to fully address the effects of guidelines on postdictions.

### Qualitative Results

#### *Confidence Levels*

Responses to the Likert- style open-ended questions about how confident the students were in their understanding of the material (not at all confident; somewhat confident; confident; extremely confident) showed that overall the students felt confident that they knew the material. The average test scores for all three tests being very high supports this.

A slight difference in confidence levels was reported between group and individuals. It appears that group calibration produced higher levels of confidence than did individual calibration. Groups reported being confident or extremely confident more than did individuals. Perhaps the use of group calibration produced a greater understanding of the material than did individual calibration. The use of group work as a method of enhancing student understanding has been supported by Paris and Newman (1990). They argue that teachers and peers can help guide students to discover and

control effective learning tactics. Effective learning tactics can in turn increase confidence levels.

Previous research has shown that students exhibit overconfidence in general when it comes to calibrating their performance (Sink, Barnett & Hixon, 1991; Grimes, 2001; Hacker, et al., 2007; Nietfeld, et al., 2005). In addition, the level of overconfidence has been shown to be greater among lower scoring students (Hacker, et al., 2007). Previous studies have shown higher achievers to be a bit under-confident. Maki, et al. (2005) found in their study that students who exhibited high levels of verbal abilities were under-confident in judging past performance (postdictions), and students with lower levels of verbal abilities were overconfident in predicting future performance (predictions). Although this study could not address the overconfidence level found in lower achieving students due to the ceiling effect exhibited on the three testing occasions, the accuracy scores (signed not absolute differences) suggest that on the average students were a bit under-confident when making predictions and postdictions. The findings are indicative of previous findings that found high achievers to be under-confident in their calibrations.

Although both groups and individuals reported being confident, variation existed between the confidence levels reported for short answer and for multiple-choice questions. Responses as to the students' confidence levels in their ability to answer short answer questions were very similar for students working in groups or working individually. However, in response to confidence levels for the students' ability to answer multiple-choice questions both individuals and groups were more confident with this type of question. In addition, groups reported being extremely confident more so than did individuals. Students felt more prepared for questions where they had choices

for responses than they did for open-ended type questions. The fact that groups felt more confident in their abilities is supported by their higher achievement on the tests.

There were no groups or individuals that reported that they were not confident in their ability to answer either short answer or multiple-choice questions. One wonders if students who had no calibration intervention at all (group or individual) would report lower levels of confidence. Perhaps just having the students think metacognitively about their understanding improves their level of confidence. When students take the opportunity to reflect on their understanding of science, they identify areas where new ideas can be generated and connections and distinctions between their ideas can be made (Davis, 2003).

#### *Factors Affecting Confidence*

In response to what made the students more or less confident in their understanding of the material the overwhelming majority of responses focused on the amount of time they had spent studying. It appears that students place great value on studying material in order to be successful on tests. This could be detrimental when it comes to engaging students in classroom activities. Students may simply believe that they don't need to be 'engaged' in the lesson to be successful as long as they have time to study the material at a later time.

Most of the remaining responses for this study for both groups and individuals pertained to content covered on the test and the amount of reviewing and practicing. One group stated, "A large study guide is good and we need a lot of time to study." This was echoed by an individual who stated, "I just need to study more, I haven't reviewed all the content." Hacker, Bol and Bahbahani (2008) found similar results in their study.



Specifically, students in their study identified a relationship between how much they studied and their calibration accuracy. In a similar study investigating calibration accuracy, Hacker and Bol (2001) reported that the most frequently noted factor leading to prediction accuracy was based on how much the students had studied. This is further supported by Hacker, et al. (2000) who found that calibration accuracy was related to the number of hours that the students had spent studying.

#### *Areas of Strengths and Weaknesses in Understanding*

The responses for areas of strength for both tests were focused mainly on specific material that appeared on the test. Students also reported visuals as an area of strength in their understanding. The use of visuals can help enhance the learning process because many students are visual learners. According to Kieff (2005) artistic explorations of a topic can excite interest and involvement in study. When students are involved in the process, their knowledge is better retained.

The final theme that emerged as an area of strength for the students in both group and individual calibration was the type of questions asked. Students reported that multiple-choice questions were an area of strength in their understanding. This makes sense since they reported that they were less confident in their ability to answer short answer questions than they were for multiple-choice questions. It seems that when students have choices for answers to questions posed they feel more confident in their ability to select the right answer. Obviously, having choices allows the student to recognize the right answer instead of having to come up with the answer totally from memory. However, in a study conducted by Pressley, Ghatala, Woloshyn, and Pirie (1990) it was discovered that students were more accurate in their predictions on short

answer questions than they were for multiple-choice items. Perhaps, even though students in this study felt more confident in their ability to answer multiple-choice questions, they were more successful on the short-answer section. Unfortunately, the data to make this analysis was not available to the researcher. Future research needs to focus more on the comparison of confidence levels and achievement levels between types of questions addressed.

Similar to reports for areas of strength in understanding, the student's responses to areas of weaknesses in understand were mainly split between specific test material and how much they studied. The main weakness reported for test 2 on genetics was pedigrees, and for test 3 on DNA technology was genetic engineering. Students were also reporting short answer questions as an area of weakness. Short answer questions were reported earlier as a factor that made the students less confident in their level of knowledge and this is supported here as well. However, for this response both groups and individuals reported that they had no weaknesses. The high average scores that were seen across the three testing situations supported the fact that the students had few weaknesses.

### Limitations

The use of a convenience sample threatens the external validity of this study. However, this information should be generalizable to similar populations of students enrolled in honors biology classes. However, changes in calibration ability resulting from this study shed light as to the ability of guidelines and group work to improve student calibration accuracy for at least this population of students and perhaps others.

As with all quasi-experimental designs, selection bias is a potential threat to

internal validity. The study entails the use of intact classes randomly assigned to condition, rather than randomly assigning students to condition. However, the fact that the students were all tested prior to entrance into the MYP program helps to ensure that they are of similar academic ability prior to the implementation of the study. Prior achievement scores of the participants were checked and found to be similar, which helps with the assurance of equivalent groups.

Teacher effects also threaten the internal validity of the study. The teacher may not have explained the requirements or procedures for the research in the same manner for all four classes. In addition, the teacher may not have delivered the material to be tested in the same manner among the four classes. The researcher attempted to minimize these threats by outlining the importance and procedures of the study prior to its implementation. The researcher encouraged the teacher to ask questions for needed clarification and to summarize the research methods in order to make sure the teacher fully understood the methods. In addition, having the same teacher teach the four classes helped decrease internal threats due to possible differences in the delivery of the material to be tested.

Further threats warrant mention. The four classes were taught at different times of the day and the number of students in each of the classes was not equal. In addition, the students in the classes may not have honestly reported their predictions and postdictions for each of the three tests. In order to encourage honesty the researcher had the teacher assure the students that the predictions and postdictions were confidential and only seen by the teacher and the researcher. The researcher made sure that the teacher personally collected the predictions and postdictions as well as the group and individual

responses to the open-ended questions in order to reinforce assurance of confidentiality to students. Two individuals both trained in qualitative research coded the open-ended questions and inter-rater reliability was established at 96%.

The error in the administration of the wrong guiding questions for the first testing occasion is an additional limitation to this study. This made the qualitative analyses of the study limited to only two of the testing occasions, and limited the ability of the researcher to fully address the effectiveness of the guided questions. However, the information gained from the second and third testing occasions does offer some valuable insights into the student's cognitive processes.

Finally, the study may be threatened by a ceiling effect due to the high scores achieved across the testing situations. Many of the students achieved perfect (100%) or very close to perfect scores on the tests. This is why the researcher could not compare the calibration accuracy of high achieving students to that of low achieving students, although the literature suggests this is important. Previous studies have found that high achieving students are consistently more accurate in calibration of their knowledge (Barton & Hixon, 197; Hacker, et al, 2007; Neitfeld, et al., 2005;). Although this study could not fully evaluate the differences in calibration accuracy between high and low achieving students, this study did find significant differences for achievement for some of the testing interventions.

#### Directions for Future Research

Additional research is needed in order to fully understand the factors that improve calibration accuracy and achievement among students. This study indicates that group work and guidelines can be used to enhance calibration accuracy and achievement.

However, the results were varied as to the effectiveness of guidelines to enhance achievement and calibration accuracy. More studies in this area would help to shed light on the ability of guidelines to focus the students on the metacognitive process of calibration.

Few studies have been conducted combining group calibration activities with guiding questions. In fact, the only other investigation to do so was the pilot study for this dissertation (Walck, et al., 2009). Future investigations combining groups and guidelines need to be carried out over a longer period of time in order to evaluate the ability of the interventions to increase calibration accuracy and achievement over time. Cognitive changes occur over time, increasing a students' ability to make mature intellectual decisions (Orange, 1999). An expansion of this study may more fully answer the questions addressed.

In addition, few students have had experiences with calibration activities at all. Studies that introduce calibration earlier in the educational experience are needed in order to evaluate the effectiveness of practice in calibration for developing calibration accuracy early in the educational experience. These types of studies could support the need to incorporate metacognitive processes into the educational curriculum.

Research into the effectiveness of guidelines and group work on calibration accuracy and achievement needs to be investigated in regular education classes. This study investigated the ability of the interventions to improve the accuracy and achievement of honors students, but lacks ecological validity for other populations of students. Perhaps honors students are higher achieving because they have learned to metacognitively process their level of knowledge (calibrate) on their own. This is

supported by Hacker et al. (2000) who found that high achievers were more accurate at evaluating their performance both before and after exams. Investigations with regular education students could offer important information about how calibration skills can be developed and enhanced in this population.

It would be interesting to investigate the differences in achievement and calibration accuracy for groups and individuals on both multiple-choice and short answer questions after exposure to calibration practice. In a study conducted by Miller and Pajares (1997) it was found that students who took multiple-choice tests obtained higher scores than did students who took short-answer tests. Since students in this study reported higher levels of confidence in answering multiple-choice questions, perhaps calibration practice with guidelines can be used to help students become more confident in their ability to answer open-ended questions.

Although research has consistently found that group work enhances student performance, few studies on the effect of guided questions for calibration of knowledge have been conducted with group settings. Future research in this area is needed in order to more fully evaluate the effectiveness of guided questions during group calibration activities.

Finally, more qualitative studies addressing the data from responses to guiding questions are needed. The data gained from these types of studies could illuminate the thought processes of the students as they move through the calibration activities. It would be beneficial to understand how one's knowledge and beliefs influence one's decisions on when to take notes, what questions to ask, and when to ask questions (Garofalo & Lester, 1985). Perhaps audiotapes of the discussions that result as a product of the guided

questions could be obtained for more in-depth analysis of the calibration process.

### Implications for Practice

This study shows that the use of calibration with guidelines can have a positive effect on calibration accuracy and achievement level. It brings to light the need for educators to teach students metacognitive skills. “Because we assume that students can be taught or helped to become better self-regulated learners, educators should observe when students are engaging in misguided regulatory behavior because of their metacognitive deficit” (Carvalho & Yuzawa, 2001). In order to observe this it is important that students are given time in class to practice metacognitive skills such as calibration of their knowledge. By being able to successfully calibrate their level of knowledge students are directed toward appropriate study activities. Many students lack study skills and do not know how to metacognitively analyze their level of knowledge.

In the current educational climate of high-stakes testing, teachers rarely take the time to have students calibrate their level of knowledge. If studies can consistently support the positive effect of group calibration with guidelines on achievement and calibration accuracy, teachers may come to realize how these metacognitive activities can offer achievement gains on these high-stakes tests. According to Bain (2004) “the best teachers use metacognition, which is defined as thinking about thinking” (p. 152). In addition, previous research indicates that the level of instructional support for cognitive strategies affects the frequency of students cognitive monitoring (Curley, Estrin, Thomas, & Rohwer, 1987).

### Summary and Conclusions

This quasi-experimental study focused on the use of group interactions and

calibration guidelines to foster the development of successful calibration skills among high school students. The students in this study were all from a suburban school district and were enrolled in an academically advanced program. The interventions of group and guidelines were tested for their effect on academic achievement and calibration accuracy (predictions and postdictions).

The first research question addressed the ability of guidelines to improve calibration accuracy and achievement for high school biology students. The results showed that guidelines were significant for improving achievement when students were calibrating individually. However, when students were calibrating in groups the guidelines did not have a significant impact on their achievement. With reference to calibration accuracy, the use of guidelines was found to be significant for predictions for only one of the testing situations. As for postdictions, guidelines were also found to be significant for only one of the testing situations. It is obvious that more studies are needed with guidelines in order to definitively answer this question.

In response to the second research question of whether calibration practice in groups is more effective than individual practice in improving calibration accuracy and achievement for high school biology students, the results for this study were significant for group effect on achievement. Students working in groups achieved higher scores than did students calibrating individually. The results for prediction accuracy were found to be significant only for test 2, although trends in more accuracy prediction calibration were seen across all tests for students working in groups.

The effect of guidelines and learning setting (group vs. individual) to interact to affect calibration accuracy and achievement were addressed in the third research



question. Significant results were found for the interactive effect of the two variables on achievement. As for prediction accuracy, only one of the testing interventions (test 3) showed significance. However it is important to note that students were fairly accurate in their predictions across the three testing occasions. Postdiction accuracy was found to be significant for only one testing occasion as well, with only test 2 showing a significant interaction. Again, as stated for guidelines, more studies are needed to fully address this research question.

In addressing the question of what students would write in response to guided questions designed to improve calibration this research showed that responses from both group and individuals were consistently similar. The majority of students (group and individual) reported that they were confident in their ability to answer the questions on the test, but more confident with multiple-choice style questions than with short answer questions. Most of the students reported that they felt the amount of time spent studying impacted their confidence level on the tests. The material on the tests impacted their confidence level as well.

This study has shown the ability of group calibration to enhance student achievement. In addition, the use of guidelines and group work showed promise for increasing prediction and postdiction accuracy. In classrooms where time is limited, and great emphasis is placed on standardized testing scores, it is more important than ever to discover effective ways to increase student achievement. More research into how calibration activities impact these constructs may result in methods that can enhance student achievement while at the same time promoting student collaboration. The results of this study support the need for more research into the effects of group and guidelines

on calibration accuracy and achievement.

## REFERENCES

- Bain, K. (2004). *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117 – 131.
- Bangert-Drowns, R., Kulik, C., Kulik, J., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213 – 238.
- Bol, L., & Hacker, D. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*(2), 133 – 151.
- Bol, L., Hacker, D., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Science, 73*(4), 269 – 290.
- Butler, D., & Winne, H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245 – 281.
- Cantrell, S. (2002). Promoting talk: A framework for reading discussions in teacher education courses. *Journal of Adolescent and Adult Literacy, 45*(7), 642 – 657.
- Carvalho, F., & Yuzawa, M. (2001). The effects of social cues on confidence judgments mediated by knowledge and regulation of cognition. *The Journal of Experimental Science, 69*(4), 325 – 343.
- Cohn, C. (1999). Cooperative learning in a macroeconomics course. *College Teaching, 47*, 51 – 59.

- Curley, R., Estrin, E., Thomas, J., & Rohwer, W. (1997). Relationships between study activities and achievement as a function of grade level and course characteristics. *Contemporary Educational Psychology, 12*, 324 – 343.
- Dahl, B. (2004). Analyzing cognitive learning processes through group interventions of successful high school pupils: Development and use of a model. *Educational Studies in Mathematics, 56(2)*, 129 – 155.
- Davis, E. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences, 12(1)*, 91 – 142.
- Dekker, R., Elshout-Mohr, M., & Wood, T. (2006). How children regulate their own collaborative learning. *Educational Studies in Mathematics, 62(1)*, 57 – 79.
- Frederiksen, J., & White, B. (1997). Reflective assessment of students' research within an inquiry-based middle school science curriculum. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Gabriele, J., & Montecinos, C. (2001). Collaborating with a skilled peer: The influence of achievement goals and perceptions on partners' competence on the participation and learning of low-achieving students. *The Journal of Experimental Education, 69(2)*, 152 – 178.
- Gadamer, H. (1993). *Truth and Method* (2<sup>nd</sup> ed.). New York: Continuum.
- Garavalia, L., & Gredler, M. (2002). An exploratory study of academic goal setting, achievement calibration and self-regulated learning. *Journal of Instructional Psychology, 4(29)*, 221 – 231.

- Garner, R. (1990). When children and adults do not use learning strategies: Towards a theory of setting. *Review of Educational Research, 60*, 517 – 529.
- Garofalo, J., & Lester, F. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*(3), 163 – 176.
- Greene, J., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334 – 372.
- Grimes, P. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education, 33*(1), 15 – 30.
- Hacker, D. & Bol, L. (2004). Metacognitive theory: Considering the social-cognitive influences. In D.M. McNerny, & S. Van Etten (Eds.), *Big*
- Hacker, D., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*, 101 – 121.
- Hacker, D., Bol, L., Hogan, D., & Rakow, E. (2000). Test prediction and performance in classroom context. *Journal of Educational Psychology, 92*, 160 – 170.
- Hacker, D., Bol, L., & Keener, M. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Memory and Metacognition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hacker, D., Dunlosky, J., & Graesser, A. (1998). Self-regulated comprehension during normal reading. *Metacognition in Educational Theory and Practice*, Lawrence Erlbaum Associates, 165 – 191.

- Hartley, K. (2001). Learning strategies and hypermedia instruction. *Journal of Education Multimedia and Hypermedia*, 10(3), 285 – 293.
- Horgan, D. (1990). Competition, calibration, and motivation. *Teaching Thinking and Problem Solving* 12, 5 – 10.
- Johnson, D., Johnson, R., & Smith, K. (1998). *Active learning: Cooperation in the college classroom* (2<sup>nd</sup> ed.). Edina, MN: Interaction Book Co.
- Justice, E., & Dornan, T. (2001). Metacognition differences between traditional-age and nontraditional age college students. *Adult Education Quarterly*, 51, 236 – 249.
- Kieff, J. (2005). Creating an interest in learning science. *Childhood Education*, 81(4), 220 – 225.
- Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly*, 35(2), 88 – 97.
- Kuhn, K. (1999). A developmental model of critical thinking. *Educational Researcher*, 28, 16 – 25.
- Kulhavy, R., & Stock, W. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279 – 308.
- Lundeberg, M., Moch, S. (1995). Influence of social interaction on cognition: Connected learning in science. *Journal of Higher Education*, 66(3), 312 – 324.
- Mali, R., Shields, M., Wheeler, A., & Lowery, T. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 94(4), 723 – 731.

- Martinez, M. (2006). What is metacognition? Teachers intuitively recognize the importance of metacognition but may not be aware of its many dimensions. *Phi Delta Kappa*, 87(9), 696 – 699.
- Meyer, L. (1986). Strategies for correcting students' wrong responses. *Elementary School Journal*, 87, 227 – 241.
- Miller, M., & Pajares, F. (1997). Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *Journal of Experimental Education*, 65(3), 213 – 227.
- Nelson, J., & Manset-Williamson, G. (2006). The impact of explicit, self-regulatory reading comprehension strategy instruction on the reading-specific self-efficacy, attributions, and affect of students with reading disabilities. *Learning Disability Quarterly*, 29(3), 213 – 230.
- Nelson, T., & Narens, L., (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125 – 171.
- Nickerson, R., (1988). On improving thinking through instruction. *Review of Research in Education*, 15(3), 3 – 57.
- Nietfeld, J., & Schraw, G. (2002). The effect of knowledge and strategy explanation on monitoring accuracy. *Journal of Educational Research*, 95, 131 0 142.
- Orange, C. (1999). Using peer modeling to teach self-regulation. *The Journal of Experimental Science*, 68(1), 21 – 39.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543 – 578.

- Paris, S., & Newman, R. (1990). Developmental aspects of self-regulated learning. *Educational Psychologist, 25*(1), 100 – 116.
- Pintrich, P., Marx, R., & Boyle, R. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167 – 199.
- Pressley, M., Ghatala, E., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and don't realize it: Confidence in response to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly, 25*, 232 – 249.
- Radhakrishnan, P., Arrow, H., & Sniezek, J. (1996). Hoping, performing, learning, and predicting: Changes in the accuracy of self-evaluations of performance. *Human Performance, 9*(1), 23 – 32.
- Riggs, R., Bol, L., Nunnery, J., & Dickerson, D. (2009). *The calibration accuracy of middle school students in math classes*. Paper submitted for presentation.
- Sawyer, R., Graham, S., & Harris, K. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84*, 340 – 352.
- Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *The Journal of Experimental Education, 65*(2), 135 – 146.



- Sink, C., Barnett, J. & Hixon, J. (1991). Self regulated learning and achievement by middle school children. *Psychological Reports, 69*, 979 – 989.
- Smith, P. (2001). Understanding self-regulated learning and its implications for accounting educators and researchers. *Issues in Accounting Education, 16(4)*, 663 – 675.
- Talbot, G. (1997). Can self-regulation learning be taught to college students? ERIC ED409-289. U.S. Department of Education, Office of Educational Research and Improvement.
- Taylor, S. (1999). Better learning through better thinking: Developing students metacognitive abilities. *Journal of College Reading and Learning, 30(1)*, 34 – 39.
- Thomas, J., Bol, L., Warkentin, R., Wilson, M., Strage, A., & Rohwer, W. (1993). Interrelationships among students' study activities, self-concept of academic ability, and achievement as a function of characteristics of high-school biology courses. *Applied Cognitive Psychology, 7*, 499 – 532.
- Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of Learning Sciences, 6(2)*, 227 – 269.
- Walck, C., Bol, L., Hager, J., & McKinney, S. (2009). *Does student calibration practice increase student test scores in high school biology classes?* Paper submitted for presentation.
- Westley, D. (2008). Accuracy of student calibration on specific topics: Very good students vs. others. Proceedings of the Informing Science & IT Education Conference. Petroleum Institute, Abu Dhabi.

- White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3 – 118.
- Winne, P. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30, 187 – 200.
- Wiseman, A. (2003). Collaboration, initiation, and rejection: The social construction of stories in a kindergarten class. *The Reading Teacher*, 56(8), 802 – 807.
- Zimmerman, B. (1986). Development of self-regulated learning: Which are the key subprocesses? *Contemporary Educational Psychology*, 11, 307 – 313.
- Zimmerman, B. (2002). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82 – 91.
- Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166 – 18.

## APPENDIX A

## Calibration Guidelines

## Group Calibration

As you are reviewing the material for the test you will assess (once at the mid-time of your review – the teacher will let you know when this time has arrived) your groups understanding of the material by answering the following questions:

You will have a five minute time limit – the teacher will let you know when your time is up.

Rate your answer to the first three questions using the rating scale found below each question. Check the appropriate box that corresponds to your answer.

- How confident is your group that its members could correctly answer short answer questions on the test material?

Not at all confident	Somewhat confident	Confident	Extremely confident

- How confident is your group that its members could answer multiple-choice questions on the test material?

Not at all confident	Somewhat confident	Confident	Extremely confident

- What makes you more or less confident about your groups understanding of the test content?
- What are areas of strengths in your groups understanding on any of the test material?
- What are areas of weaknesses in your groups understanding on any of the test material?

### Individual Calibration

As you are reviewing the material for the test periodically (once at the mid-time of your review – the teacher will let you know when this time has arrived) assess your understanding of the material by answering the following questions:

You will have a five minute time limit – the teacher will let you know when your time is up.

Rate your answer to the first three questions using the rating scale found below each question. Check the appropriate box that corresponds to your answer.

- How confident are you in your ability to correctly answer short answer questions on the test material?

Not at all confident	Somewhat confident	Confident	Extremely confident

- How confident are you in your ability to correctly answer multiple-choice questions on the test material?

Not at all confident	Somewhat confident	Confident	Extremely confident

- What makes you more or less confident about your understanding of the test content?
- What are areas of strengths in your understanding of any of the test material?
- What are areas of weaknesses in your understanding of any of the test material?

## APPENDIX B

## Prediction and Postdiction Sheets

Prediction SheetStudent Calibration Sheet Test # \_\_\_\_\_

Student ID #: \_\_\_\_\_ Class: \_\_\_\_\_

*How many points do you think you will earn on the test?* \_\_\_\_\_ (1-100 points)Postdiction sheetStudent Calibration Sheet Test # \_\_\_\_\_

Student ID #: \_\_\_\_\_ Class: \_\_\_\_\_

*Now that you have taken the test, how many points do you think you earned?*

\_\_\_\_\_ (1-100 points)

## APPENDIX C

## Parent Approval Letter

To: Parents of MYP Biology students  
 From: Mrs. Walck / Ms. Durbin  
 Ref: Research Study

My name is Camilla Walck and I am a biology teacher at Princess Anne High School. I am currently enrolled in a doctoral program at Old Dominion University. In order to complete this course of study I am required to conduct a research investigation. I have chosen to investigate the abilities of students to calibrate their learning achievement. Calibration is an important metacognitive process in which students assess their level of understanding of curricular material. During the research investigation your students will periodically predict how well they are prepared for upcoming tests and then predict how well they think they performed immediately prior to taking each test.

Previous research has shown that calibration ability and student performance are directly related. It is my hope that this research will shed light on the importance of student calibration and help your child to improve in his/her self-assessment abilities.

Individual student data will be confidential – seen only by the researcher and the student. In order for your child to be part of this research, simply sign the consent line below. Participation is optional, however I feel it will be an important learning process. If you have any questions please feel free to call me at 749-8065 or email me at [camilla.walck@vbschools.com](mailto:camilla.walck@vbschools.com).

Thank you in advance for your support of this important project. As always, it is a pleasure to work with such wonderful students.

Camilla C. Walck

-----  
 I \_\_\_\_\_ (guardian name) give permission for  
 \_\_\_\_\_ (student name) to participate in this calibration  
 research project.

OR

I \_\_\_\_\_ (guardian name) **do not** give permission for  
 my student \_\_\_\_\_ (student name)

## VITA

Camilla Crockett Walck, NBCT

8508A Atlantic Avenue  
 Virginia Beach, Virginia 23451  
 Email: [camill.walck@vbschools.com](mailto:camill.walck@vbschools.com)  
 Cell Phone: 757-749-8065 Home Phone: 757-496-3604

## EDUCATION:

- May 8, 2010 Expected date of completion  
 Ph. D. in Education Curriculum and Instruction  
 Old Dominion University, Norfolk, Virginia  
 Dissertation Topic: The Effect of Individual or Group Guidelines  
 On the Calibration Accuracy of High School Biology Students
- 1994 M. S. in Secondary Education, Biology, Old Dominion University,  
 Norfolk, Virginia  
 Concentrations: Biology; Teaching with Psychodrama  
 Project: The Effects of Psychodrama on Student Retention in  
 High School Biology Classes
- 1985 B.S. in Biology with Endorsement in Education, Old Dominion  
 University, Norfolk, Virginia  
 Senior Project: A Case Study of the Progression and Treatment of  
 Lymphoma in Hodgkin's Patient Joanna Turbeville

## EXPERIENCE:

- 1995 –present International Baccalaureate Biology Teacher and Department Chair,  
 Princess Anne High School, City of Virginia Beach Public Schools,  
 Virginia Beach, Virginia
- 2003 – present Adjunct instructor, Methods of Teaching Elementary School Science and  
 Secondary Science Methods, Virginia Wesleyan College, Norfolk,  
 Virginia

## PUBLICATIONS / PRESENTATIONS:

- Walck, C. (2003). *Motivating the Unmotivated*. A Tapestry of Knowledge, Virginia  
 Beach Public Schools Publication: VA
- Walck, C., Bol, L., Hager, J., & McKinney, S. (2009). *Does student calibration practice  
 increase student test scores in high school biology classes?* Paper submitted for  
 presentation.