Winter 2014

Global and Criteria Based Judgments of an Undergraduate Exit Writing Examination

Katrice Alexandria Hawthorne
Old Dominion University

GLOBAL AND CRITERIA BASED JUDGMENTS OF AN UNDERGRADUATE EXIT WRITING EXAMINATION

by

Katrice Alexandria Hawthorne
B.A. May 2001, Virginia Union University
M.A. May 2003, Virginia Commonwealth University

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

DOCTOR OF PHILOSOPHY

EDUCATION

OLD DOMINION UNIVERSITY
December 2014

Approved by:

Linda Bol (Director)

Shana Pribesh (Member)

Yonghee Suh (Member)

ABSTRACT

GLOBAL AND CRITERIA BASED JUDGMENTS OF AN UNDERGRADUATE EXIT WRITING EXAMINATION

Katrice Alexandria Hawthorne
Old Dominion University, 2014
Director: Dr. Linda Bol

The effect of a calibration strategy requiring students to predict and postdict their scores on a writing exam was investigated. The utility of rubric-referenced calibration and the interaction between achievement and self-efficacy on calibration accuracy were also explored. Five hundred ninety six undergraduate students enrolled in an urban, comprehensive, public university participated. Students were assigned to one of three calibration conditions: (1) a global condition (overall judgments only), (2) a global and criteria condition (a general rubric), or (3) a global and detailed criteria condition (a detailed rubric). Students in all three conditions provided global calibrations before and after the exam. Students also completed the Writing Self-Regulatory Efficacy Scale. Neither calibration condition alone nor self-efficacy alone was found to effect calibration accuracy. Calibration condition and SAT critical reading achievement were found to be significant for predictive accuracy in organization and development and analysis only. Calibration condition and global writing scores interacted to significantly effect prediction and postdiction accuracy in sentence structure, as well as prediction accuracy in grammar, diction, and mechanics. Higher achieving students in all three conditions were more accurate than lower achieving students. Additional research is needed to fully examine the relationships among calibration accuracy, achievement, self-efficacy and specific writing criteria.

Dedication

In loving memory of my grandmother, Patricia Diane Carson

ACKNOWLEDGMENTS

I am deeply indebted to my advisor Dr. Linda Bol for her fundamental role in my doctoral work. She has my wholehearted appreciation. I would like to thank my committee members, Dr. Shana Pribesh and Dr. Yonghee Suh for their time, feedback, and support.

I was fortunate to have friends, colleagues, teachers, and fellow students who encouraged me throughout. I sincerely thank them all. I am especially thankful to the students who participated in this study. I also owe a special thanks to Dr. Kimberly Matthews, who assisted with this study and is a constant source of support.

Lastly, I thank my family, specifically, my mother, Rose, who loves and believes in me unconditionally.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

The educational value of writing cannot be overstated. Writing is an essential part of thinking and learning. According to Lindemann (2001), writing is taught because it serves as a source of economic power, is a social necessity, and functions as a form of knowing. Since the 19th century, American colleges and universities have been grappling with students' deficiencies with writing. In 1874, Harvard College implemented writing entrance examinations amid concerns that a sizeable number of students, even those from the best schools, were incapable of writing effectively, and by 1897, all Harvard College students were required to take a course in composition (Connors, 1996). Producing students who are competent writers is a fundamental aim of higher education. Monroe (2003) contends that "effective writing is central to the work of higher education" (p. 4), and Tritelli (2003) referred to writing as the "fulcrum" of the undergraduate curriculum.

Since the mid-1800s when Horace Mann advocated that written examinations replace oral examinations (U.S. Congress, 1992), writing has long been considered an ideal method to gauge student learning. According to Covill (2012), the majority of higher education faculty believe that assigning writing is one of the best pedagogical practices across disciplines. The emergence of writing across the curriculum and writing in the disciplines programs throughout higher education emphasize the importance of writing throughout students' undergraduate careers (Monroe, 2003).

Writing is also important in the workplace, as correlations between effective writing and professional advancement have been found (Lindemann, 2001; National Commission on Writing, 2004). Results from a survey of 120 major American

corporations indicate that writing is a nearly universal professional skill required in industries from finance to manufacturing (National Commission on Writing, 2004). Yet, many students are unprepared to meet the writing demands required for success in the emerging 21st century workplace (Yancey, 2009). Many corporations view writing as a "threshold skill" for employment and promotion, and they are dismayed when college graduates "aren't even aware when things (e.g., singular/plural agreement, run-on sentences) are wrong" (National Commission on Writing, 2004, p.5). Employers expect clear, concise, effective writing from newly hired graduates, and newly hired graduates will find that writing poorly may jeopardize their success and hinder their professional advancement (Lindemann, 2001; National Commission on Writing, 2004). In addition, corporate leaders equate good writing with good thinking in the same vein as compositionists.

Writing is an essential and transferable skill that is needed in every discipline. Even though the need for effective writing is considered essential to success in school and in the professional sector (Zimmerman & Risemberg, 1997), the challenge to support students' writing development in practical and meaningful ways is evident (Bean, 2011; Covill, 2012). The National Commission on Writing in America's Schools and Colleges (2004) notes that writing is often shortchanged at every level of education because it is both time consuming for students to produce and for teachers to assess. As Sinclair Lewis said, "writing is just work – there's no secret. If you dictate or use a pen or type or write with your toes – it is still just work" (Lindemann, 2001, p. 10). Writing is a challenging task that requires not only discipline but also considerable self-regulation.

Self-Regulated Learning

Academic self-regulation is a process through which students marshal their cognitive abilities into academically useful skills (Zimmerman, 1998). Self-regulated learning is a complex, multifaceted process that has developed from an extensive body of research that encompasses various processes such as goal setting, self-evaluation, self-observation, and self-judgment. Self-regulated learning is defined as the "systematic use of metacognitive, motivational, and/or behavioral strategies" (Zimmerman, 1990). Self-regulation is important as it provides valuable information on how students master the learning process (Zimmerman, 1998).

Learners can be described as self-regulated to the degree that they are active participants in their own learning process (Zimmerman, 1989). Zimmerman (1989) maintains that students must, on the basis of self-efficacy perceptions, intentionally use self-regulated learning strategies to achieve academic goals. Perceptions about one's ability to organize, implement, and perform a task are forms of self-efficacy (Zimmerman, 1989). The strategies used to self-regulate learning include self-evaluation, organization, planning, self-monitoring, and reviewing (Zimmerman, 1990). The ability to reflect on and assess one's own thinking and behavior and to control the processes necessary to continuously make adjustments to complete a task are essential components of self-regulation and may significantly enhance student learning and achievement.

Characteristically, in academic settings, self-regulated learning emphasizes effective use of both cognitive and metacognitive skills to successfully aid in academic learning (Zimmerman, 1990). Self-regulated learning is a multidimensional process; it is "never an absolute state of functioning but rather varies in degree" (Zimmerman, 1989, p.

332). Self-regulated learners are never passive participants; rather, they are active, constructive participants in the learning process.

Zimmerman's model of self-regulation (1989) presumes that learners can monitor, control, and regulate their cognition, behavior, and environment through commitment to academic goals and effective strategy use. Learners' personal achievement orientation and self-efficacy perceptions are assumed to affect their ability to self-regulate. Learners' self-efficacy (i.e., their perceptions about their abilities) may mediate their use of self-regulated learning strategies. Conversely, strategy application may provide useful self-efficacy knowledge.

Most self-regulated learning theorists view learning as a cyclical, open-ended, triadic process. This view acknowledges that self-regulated learning is a complex process wherein learners actively and consistently employ cognitive and metacognitive strategies to achieve academic success. Thus learners must not only possess self-regulatory skills, but they must also be able to apply these skills "persistently in the face of difficulties, stressors, or competing attractions" (Zimmerman & Bandura, 1994, p. 846).

There is an extensive body of research that provides evidence of the role of students' use of self-regulated learning strategies and its relationship to academic achievement (Ferrari, Bouffard, & Rainville, 1998; MacArthur & Philippakos, 2013; Zimmerman & Bandura, 1994; Zimmerman & Risemberg, 1997). Self-regulated learners engage in effortful and sustained use of both cognitive and metacognitive skills in order to successfully complete academic tasks. Self-regulated learners set goals, organize, self-monitor, and self-evaluate during the learning process, and they distinguish themselves by the goals they set and the accuracy of their self-monitoring and attributions

(Zimmerman, 1990; Zimmerman, 1998). According to Zimmerman (2011), in self-regulated learning, abilities are transformed into task-related skills. The criteria that are used to examine effective writing can be delineated into tasks by criteria. For example, the ability to accurately find and correct grammatical errors is a task that is quite different from the ability to effectively organize a paragraph.

In self-regulated learning, learners may receive internal or external feedback. Internal feedback is a function of self-testing or self-monitoring and is fundamental to self-regulation. Learners must be accurate self-monitors in order for internal feedback to be effective. External feedback can be used to focus learners' attention on calibration and to help learners become better self-monitors and to make more accurate calibration judgments (Stone, 2000).

A number of self-regulated learning models exist; however, few have been researched empirically. The exceptions are Pintrich (2000), Winne and Hadwin (1998), and Zimmerman (2000). Motivation is highlighted in Pintrich's (2000) model. While Winne and Hadwin's (1998) model consists of a preparatory phase, a performance phase, and an evaluation phase. Task definition and goal setting are placed in distinct phases in Winne and Hadwin's (1998) model, and monitoring and control are prescribed in each phase.

Zimmerman and Campillo's (2003) cyclic phase model of self-regulated learning (Figure 1), which serves as the theoretical framework for this study, is useful in explaining students' efforts to learn and to become masters of their own learning processes. The model is cyclical, and the three phases - forethought, performance, and

self-reflection - include not only components of self-regulation, but also analogous characteristics of calibration.

The forethought or preparatory phase includes task analysis, goal setting, and strategic planning. In this phase, self-efficacy perceptions and various motivational beliefs influence learners' understandings about the task and the goals they set. The performance phase includes the selection of effective strategies and appropriate self-monitoring and self-instruction activities. The final phase, self-reflection, involves self-evaluation, attributions for successes and failures, and adaptions that can be used to improve future performance.

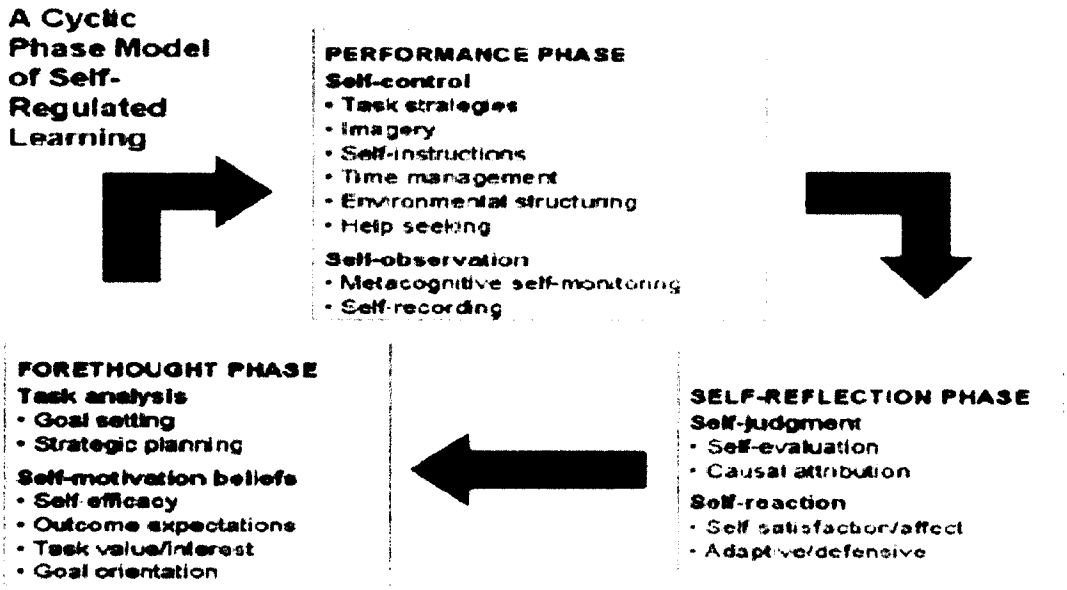


Figure 1. A cyclic phase model of self-regulated learning. From "Motivating self-regulated problem solvers," by B. J. Zimmerman & M. Campillo, 2003, in J.E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving*. New York: Cambridge University Press.

Calibration

Calibration, a measure of the relationship between confidence in one's performance and the accuracy of one's judgments, is an emergent issue in the field of

self-regulation (Stone, 2000; Zimmerman, 2011). Calibration is prominent in Zimmerman's cyclical model of self-regulated learning, as it is a key measure of the accuracy of leaners' self-monitoring (Zimmerman, 2008). According to Zimmerman (2008), calibration is an indicator of the accuracy of students' self-monitoring. Zimmerman and Campillo's (2003) cyclic phase model of self-regulated learning is appropriate for use in this study as it highlights both calibration (i.e., outcome expectations and self-judgments) and self-efficacy. In addition, writing is a cyclical process that includes planning (i.e., forethought), writing (i.e., performance), and revising (i.e., self-reflection).

Individual characteristics, specifically self-efficacy and self-regulation, overlap in the calibration literature (Bembenutty, 2009; Garavalia & Gredler, 2002; Stone, 2000). While confidence can be measured in self-regulated learning and in calibration, the measures, though related, may tap distinct aspects of confidence. In calibration, the learner is estimating his or her confidence on current knowledge either before or after a performance, which may affect self-regulation. Conversely, in self-regulated learning, confidence measures require the learner to estimate future performance or confidence in one's ability to learn or complete a future task (Bembenutty, 2009; Stone, 2000). According to Bembenutty (2009), calibration is an essential metacognitive process that directs achievement and regulates task completion.

The process of self-testing or self-monitoring in self-regulation is nearly identical to the process of calibration (Stone, 2000). In self-regulated learning, self-monitoring involves reflection and is an assessment of one's own progress (Garavalia & Gredler, 2002; Stone, 2000). Leaners' ability to self-test or self-monitor likely affects levels of

confidence (i.e., calibration) before and after a task. Postdictions, confidence judgments made after completion of a task, are generally more accurate. It is assumed that self-testing is enhanced after completion of a task, thus sparking more accurate confidence ratings. Stone (2000) suggests that calibration is related to depth of processing and that self-testing improves depth of processing, thereby enhancing calibration accuracy.

The methods of analyses used in self-regulation and calibration studies are important, as the time of the assessment and how the data are examined provide much needed information on the distinct aspects of each process that are tapped (Garavalia & Gredler, 2002; Schraw, Kuch, & Gutierrez, 2013; Stone, 2000). Much of the literature on calibration features general knowledge items, which are assessed using multiple-choice questions. Calibration may differ on essay questions, thus additional research and measures are needed to analyze both self-regulation and calibration in this context (Bol & Hacker, 2001; Stone, 2000).

Metacognition is assessed using a variety of methods, including calibration techniques, which include estimates and "indices of actual performance" (McCormick, Dimmitt, & Sullivan, 2012, p. 71). Calibration is fundamental to the concepts of metacognitive judgments, self-regulation, and self-efficacy beliefs (Alexander, 2013). Metacognitive judgments, estimates of learning, estimates of effort and time expenditures, and, perhaps most importantly, estimates of correctness (Desoete & Ozsoy, 2009), are an integral part of calibration. Moreover, calibration, "the degree of fit between a person's judgment of performance and his or her actual performance" (Bol & Hacker, 2012, p.1), is fundamental to both cognitive and metacognitive processes (Bembenutty, 2009).

Calibration is a crucial component of self-regulation, as effective self-regulation requires accurate self-assessment (Hacker, 1998). According to Butler and Winne (1995), "the most effective learners are self-regulating" (p. 245). Knowing whether learners can accurately calibrate should be of great concern because the ability to gauge one's performance accurately will likely affect subsequent effort and behavior. Self-regulatory writing strategies, such as "checking pronouns for referential suitability" (Zimmerman & Risemberg, 1997, p. 75), require self-monitoring, a judgment of understanding and a judgment of correctness, which translates to an ability to accurately calibrate what one knows and can do. Learners' ability to assess themselves or to accurately calibrate requires that they not only monitor their performance, but also that they self-regulate (Hacker, 1998). Inaccurate monitoring may cause learners to suspend studying before learning is complete (van Loon, de Bruin, van Gog, & van Merriënboer, 2013), which might also influence learners' self-efficacy.

Self-efficacy

Self-efficacy, an essential component of self-regulation, can be defined as an individual's belief that he or she is capable of performing a task (Bandura, 1977; Bandura, 1993). Self-efficacy is included in the forethought phase of Zimmerman's (2008) model, as it guides students' actions and influences their beliefs. Students' perceptions of performance can influence their learning experiences, as self-efficacy beliefs determine how individuals think, feel, behave, and motivate themselves (Bandura, 1994). Personal beliefs about one's efficacy affect the selection of goals, as self-efficacy perceptions influence the learning activities that students' participate in and the goals

they set for themselves (Bandura, 1993). A strong sense of self-efficacy fosters academic success and heightens and sustains effort in the face of difficulty (Bandura, 1994).

Perceived self-efficacy contributes to cognitive development and function through four major processes: *cognitive, motivational, affective*, and *selection* (Bandura, 1993). According to Bandura (1993), purposeful behavior is governed by "cognized goals" (p. 118). Self-efficacy beliefs affect cognitive processes in a number of ways, as students set learning and achievement goals based on their perceived efficacy. Motivational processes affect self-efficacy beliefs as they determine the goals students' set for themselves, the amount of effort they expend, their persistence in the face of difficulties, and their resilience to failure. The ability to regulate anxiety and to cope effectively by controlling stress and negative self-attributions is also essential to regulating self-efficacy. Affective processes, such as depression, control, anxiety, and one's means of coping, also are affected by self-efficacy beliefs.

High self-efficacy beliefs correlate with the use of effective metacognitive strategies. Self-efficacy beliefs often predict the level of effort learners will expend on a task and their motivation to complete the task (Bandura, 1977). Bandura (1977) suggests that efficacy expectations, specifically mastery expectations, influence performance, as repeated success on a task builds learners' efficacy perceptions. Self-regulated goal setting helps to develop self-efficacy, as knowledge of what one knows and what one seeks to know influences the learning activities in which one engages (Bandura, 1986).

Accurate appraisal of one's efficacy (i.e., judgments of capability) is valuable (Bandura, 1986), as inaccurate self-efficacy appraisals, rather than lack of capability or skill, can lead to adverse academic behaviors (Bandura, 1986; Pajares, Hartley, &

Valiante, 2001). Faulty judgments of self-efficacy may lead learners to avoid certain tasks and to give up more easily in the face of obstacles (Pajares et al., 2001). Learners often attribute their academic success or failure to a number of factors, including ability, effort, and task difficulty, and these attributions often affect learners' self-efficacy perceptions (Schunk, 1989). It must be noted that high self-efficacy alone will not increase academic achievement, especially if needed skills are lacking (Schunk, 1989).

Nonetheless, positive self-efficacy beliefs are essential for effective learning (Bandura, 1986), as they not only promote learning they also enhance motivation to self-regulate (Zimmerman, 1998). Students' beliefs in their efficacy to control or regulate their learning determine the mastery goals they set for themselves. Learners who attribute their success to their abilities feel more capable of performing well in the future (Schunk, 1989). Research examining self-efficacy and achievement in reading and writing indicate a predictive relationship between self-efficacy beliefs and writing achievement for college undergraduates (McCarthy, Meier, & Rinderer, 1985; Prat-Sala & Redford, 2012; Shell, Murphy, & Bruning, 1989).

While highly self-efficacious students will set more challenging goals, students with low self-efficacy beliefs may set unambitious goals despite possessing the requisite knowledge and skills. As such, learners who possess a strong sense of self-efficacy set challenging goals, and they persevere and adapt in the face of failure; however, if self-efficacy is low difficult tasks may be avoided (Bandura, 1994). A low sense of efficacy to exercise control over one's learning may lead to impaired functioning and academic failure. Thus simply having knowledge and skills is insufficient, as students must also

possess favorable self-efficacy beliefs about their ability to perform well as these beliefs determine whether they undertake increasingly challenging tasks.

The ability to judge one's own performance is a calibration process closely related to judgments of self-efficacy. Research suggests that self-efficacious learners are more accurate self-monitors and self-evaluators (Zimmerman, 1989; Zimmerman, 1998). According to Bembenutty (2009), learners' beliefs about their capabilities connect self-efficacy to self-regulation and calibration. Research suggests that highly calibrated learners generally have positive self-efficacy beliefs (Chen, 2003). Since writing is a complex and demanding task that requires active and intentional self-regulation (Ferrari, Bouffard, & Rainville, 1998; Graham & Harris, 1997; MacArthur & Philippakos, 2013; Zimmerman & Risemberg, 1997), rubrics may serve as helpful self-regulatory tools.

Rubrics and Writing Self-Efficacy

Rubrics can help students self-regulate and self-assess, thus serving as instructional and evaluative tools. Zimmerman and Bandura (1994) assert that self-regulation of one's own motivation and learning affects the self-management of writing activities. Self-regulated learning can be placed into three fundamental processes: *environmental*, *behavioral*, and *covert* or *personal*. Environmental processes refer to the physical or social setting in which writers write. Behavioral processes concern writer's self-regulation of overt behavioral activities, and covert or personal processes involve writers' self-regulation of cognition and attitudes associated with writing (Zimmerman & Risemberg, 1997). Research suggests that good writers are more metacognitively involved in writing, more active monitors of their writing and more aware of their audience than poor writers (Ferrari et al., 1998). McCarthy et al. (1985) suggest that

"accurate assessment of self-efficacy predicts writing performance" (p.469); however, they note that developmental or basic writers often vastly overrate or underrate their writing performance, and they advocate for additional research that includes evaluations of one's efficacy, one's writing ability, and an assessment of one's written work.

While instructors often use rubrics to evaluate student work, rubrics can serve dual purposes. Rubrics may be used as self-assessment instruments. They articulate expectations for assignments by detailing evaluation criteria and by describing levels of quality, which distinguish between good and poor responses. Empirical evidence of students' use of rubrics is limited; however, students have reported that rubrics help them by determining expectations, clarifying standards for performance, and by guiding their internal feedback about progress towards those standards (Andrade & Du, 2005; Reddy & Andrade, 2010). Consequently, rubrics have the potential to promote writing self-efficacy as well as self-regulatory behaviors, such as goal setting and self-assessment. In addition, calibrating both holistically (i.e., globally) and analytically (i.e., by criteria) will allow researchers to examine whether students' self-assessments are more accurate globally or by criteria.

Significance and Purpose

Students must be aware of their own writing skills in order to effectively monitor, control, and evaluate the progress of their thinking and writing. According to Zimmerman and Risemberg (1997), "high self-evaluative standards can help writers improve the quality of their prose" (p. 82). The purpose of this study is to examine the impact of calibration condition on calibration accuracy and writing self-regulatory efficacy. This

study is significant, as it will not only assess students' self-regulatory efficacy in writing but also their calibration accuracy on a writing exam.

While many studies focus on estimates of individual item correctness or total correctness, this study will measure students' estimates of their total scores and criterion scores. Zimmerman and Risemberg (1997) maintain that writing, despite being one of the most complex skills taught, is often poorly learned, and they hypothesize that self-assessments can be helpful in a broader range of writing tasks than revision alone. Thus calibration research that examines the accuracy of students' judgments regarding their own writing is needed. Hacker, Keener, and Kircher (2009) contend that investigating writing is a difficult task (p. 164). Both writing and calibration research are steadily shifting from theory to practice, thus research that provides a parallel effort to enhance understanding of both provides a promising line of inquiry.

This experimental study is designed to explore the relationships among calibration accuracy, self-efficacy, and writing achievement. The researcher randomly assigned sections of the Examination of Writing Competency to one of three conditions: (1) a global condition, (2) a global and criteria condition, and (3) a global and detailed criteria condition. All students estimated their performance globally (i.e., their total score from one to 20); however, depending on condition, some students rated their performance by criteria (i.e., their score on one of the four scoring criteria from one to five). For example, students predicted their global (i.e., total) score from one to 20, or they predicted their global score and their criterion score based on four writing categories scored from one to five.

In the global condition, students calibrated globally only. In the global and criteria condition, students calibrated globally and by criteria using a general rubric that included scores and levels of performance, but not performance descriptors. In the global and detailed criteria condition, students calibrated globally and by criteria using the EWC Scoring Rubric (Appendix A), which includes scores, levels of performance and detailed performance descriptors for each criteria. Only students in the global and detailed criteria condition had access to the full EWC Scoring Rubric to make their calibrations.

The following research questions were addressed:

1. Does calibration accuracy in writing differ by calibration condition?
2. Does self-efficacy in writing differ by calibration condition?
3. Does calibration accuracy differ by criteria?
4. Do calibration condition and achievement level interact to influence calibration accuracy?
5. Do calibration condition and self-efficacy level interact to influence calibration accuracy?

Summary and Overview

Chapter I highlights the importance of effective writing skills and outlines the theoretical concepts of self-regulation, calibration, and self-efficacy. Zimmerman's model of self-regulation serves as the framework for understanding this relationship. This research seeks to respond to an observed gap in the literature by examining not only the calibration accuracy and performance of university students on a writing exam, but also by investigating the relationship between writing self-efficacy, calibration, and writing achievement.

The following chapter details the relationship between self-regulated learning and writing achievement. The relationship between self-regulation and calibration is also described. Studies include research designed to illuminate the phenomenon of calibration and its relationship to achievement, self-regulation, and self-efficacy. Investigations into the validity of calibration measures will be presented, as will empirical studies that examine interventions intended to enhance calibration accuracy and achievement.

Chapter III provides a detailed account of the implementation of this experimental study. The selection process for the sample is discussed, and general characteristics of the participants are provided. The instrumentation of the study, including the validity and reliability of the tests and questionnaires used, is explored. Data analysis methods and study limitations are also addressed.

Examination of Writing Competency (EWC), Writing Self-Regulatory Efficacy, and calibration scores are reported in Chapter IV. Inferential statistics, specifically the results of MANOVAs and an ANOVA, are used to address the study's research questions.

Chapter V begins with a discussion of the study's findings in relation to previous literature. The relationship between calibration accuracy and performance is explored. In addition, the findings are linked to and contextualized within the context of existing research on the topic. The chapter concludes with a discussion of the implications of this study and directions for future research.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter provides a review of the research on self-regulation, self-efficacy and calibration. Zimmerman's model of self-regulated learning, which was described in the previous chapter serves as the study's theoretical framework, thus contextualizing the relationship between self-regulated learning and calibration accuracy. The relationship between self-regulated learning and writing achievement is discussed. Then, the role of self-efficacy in self-regulated learning is addressed, followed by an investigation of the relationship between self-efficacy and writing. Writing is a complex and demanding process that requires a number of self-regulated learning strategies, including but not limited to goal direction, planning, text production, and revision. Since, self-efficacy is a construct that differs by task, writing self-efficacy and rubric use are explored in detail. Next, a synopsis of empirical research focused on measuring calibration is presented. An overview of calibration studies that examine the relationship between calibration accuracy and achievement follows. The chapter concludes with an examination of the role of rubric-referenced self-assessment on writing achievement.

Self-Regulation and Writing

Proficient and struggling writers differ in the strategies they use for goal setting, planning, revising, self-monitoring, and self-evaluation – all essential elements of self regulation (MacArthur & Philippakos, 2013). Ferrari, Bouffard, and Rainville (1998) examined the writing and self-regulatory processes of 48 good and poor junior college student writers using direct observation. Students were randomly assigned to one of two conditions and provided 50 minutes to compose a comparative essay. Half of the students

were told that they would not be evaluated on surface-linguistic aspects of their text (e.g., spelling, grammar, and punctuation) and that they could ignore those aspects when writing. An observer noted the time students spent doing various activities while writing (e.g., planning, prewriting, and revising). These observations were used as evidence of students' self-regulation. A simple comparative writing essay was used to mediate the effects of prior knowledge by selecting a writing task with which all subjects had roughly equivalent knowledge and to evaluate students' writing performance. After students finished writing, they were asked to evaluate whether their essay had an introduction, a main body, and a conclusion.

The results indicated that poor writers devoted less time to planning and generating ideas, and poor writers began to write sooner than good writers. The results also indicated no effect of condition on students' use of self-regulatory strategies. There was also no effect of condition on the number of surface-linguistic errors in students' essays. In addition, while both good and poor writers were observed making changes to their essay, poor writers introduced more detrimental changes to their essay than good writers. Poor writers' attempts at self-regulation often caused them to introduce more errors into their text because they lacked sufficient syntactical knowledge, thus they were applying generally effective strategies in ways that adversely affected their performance. Good writers were more accurate self-evaluators than poor writers, and they were better able to evaluate whether their essays contained the needed organizational elements than poor writers. In fact, 52 percent of effective writers were perfectly accurate in evaluating the elements of their essays in comparison to poor writers (Ferrari et al., 1998). The authors contend that self-regulation may, at times, work against poor writers because they

lack the domain knowledge needed to not only organize their writing, but to ensure syntactical correctness.

While self-regulation is an important component of effective writing, good writers also possess considerable rhetorical and linguistic knowledge. Negretti (2012) found that in writing "knowing what is important to do does not always mean knowing how to do it, when, and why" (p. 160). Negretti's (2012) longitudinal study explored the metacognitive awareness and self-regulation of 17 community college students enrolled in a beginning composition course. Students were asked to maintain journals to reflect on the strategies they used to complete the course's various writing assignments. The results suggest that most students showed awareness of the self-regulatory processes needed to write academic essays. Students' journal entries highlighted their need to overcome writing challenges and to use specific writing techniques. After completion of essay assignments, journal entries reflected more awareness of the strategies needed to write effectively; however, students were not always aware of how to implement the strategies, when to implement the strategies, or why some strategies were more appropriate than others (Negretti, 2012). Both Ferrari et al. (1998) and Negretti's (2012) results suggest that students need greater domain knowledge and more knowledge of their own writing weaknesses in order to use self-regulatory strategies effectively.

MacArthur and Philippakos' (2013) study merged both specific writing strategy instruction with self-regulated learning strategy instruction, specifically training in self-monitoring, self-evaluation, self-instructions, goal setting, self-reinforcement, and time and environment management. The results indicated that self-regulated learning strategy

instruction increased developmental students' writing quality from pre-test to post-test (MacArthur & Philippakos, 2013).

While neither Negretti (2012) or MacArthur and Philippakos (2013), explicitly explored calibration accuracy, the results from both studies suggest that self-regulatory strategy use enables students to become better self-assessors of their writing performance. In Negretti's (2012) study, students, initially, judged their performance based on simply completing the required task and submitting it on time with little consideration of the rhetorical requirements of the writing assignment. As the course progressed and students became more aware of the rhetorical features of effective academic writing, they became more critical of their work and less optimistic about their performance, as the criteria they used to judge success became more varied and complex. The findings of both Negretti (2012) and MacArthur and Philippakos (2013) suggest that students must develop metacognitive awareness in order to select and use the self-regulatory strategies that will positively enhance their writing achievement. Negretti (2012) concludes that students' knowledge of which strategies to apply to which writing tasks stimulates their ability to effectively self-regulate, which is consistent with Zimmerman's (2000) model of self-regulatory development. The research on self-regulatory writing strategy use suggests that direct instruction in self-regulation provides students with a structured process to approach writing tasks (MacArthur & Philippakos, 2013; Negretti, 2012).

Self-Efficacy and Writing

Self-efficacy for writing describes writers' perceptions of their ability to accomplish designated writing goals (Zimmerman & Risemberg, 1997). Early research has consistently shown that writing self-efficacy beliefs and writing performance are

related (Shell et al., 1989; Zimmerman & Bandura, 1994). However, the research often defines self-efficacy in a number of ways. McCarthy, Meier, and Rinderer (1985) refer to writing self-efficacy as writers' evaluations of their general writing skills. Zimmerman and Bandura (1994) allude to students' writing self-regulatory efficacy, which they define as students' confidence to use self-regulatory strategies in addition to students' perceived self-efficacy in writing. Students' self-regulatory efficacy has also been associated with achievement in writing (Zimmerman, Bandura, & Martinez-Pons, 1992; Zimmerman & Risemberg, 1997).

This study uses Zimmerman & Risemberg's (1997) description of self-efficacy for writing, as writers' perceptions of their ability to accomplish designated writing goals. Research suggests that self-efficacy in writing contributes to writing performance in single and cross-domain analyses (Prat-Sala & Redford, 2012). Writing self-efficacy beliefs should be specific to the writing task, as self-efficacy in one domain of writing (e.g., creative writing) may be different from self-efficacy in another domain (e.g., essay exam writing). Positive writing self-efficacy beliefs should demand that students understand the components of the writing process. Students who believe that writing is a meaningful process and who have confidence in their writing skills are more likely to persist in the face of challenging writing tasks than students who lack confidence and view writing only as a school task (MacArthur & Philippakos, 2013).

To examine the relationship between writing self-efficacy and undergraduate students' writing performance, Prat-Sala and Redford (2012) studied both first year and second year undergraduate psychology students. The Self-Efficacy in Writing (SEW) scale was administered to 145 students. The writing performance of first-year students

was assessed using a 500-word essay, while the writing performance of second-year students was assessed using a 1,200-word essay. Writing performance was significantly related to the writing self-efficacy beliefs of both first year and second year students; however, the relationship between self-efficacy and writing were slightly stronger in second year students. Prat-Sala and Redford (2012) argue that second year students have had more opportunities to monitor and evaluate their writing self-efficacy, thus the increase in the relationship is expected as second year students have likely completed more writing tasks, and, consequently, have more sources of evidence on which to anchor their writing self-efficacy beliefs.

McCarthy et al. (1985) argued that strengthening students' efficacy expectations about their writing is an important step in improving students' writing performance. They maintained that effective writers are more self-directed and that they take active control of their writing. They hypothesized that students who evaluated themselves as capable and able to self-evaluate (i.e., they possess strong efficacy beliefs) would be better writers than students with weak efficacy beliefs. To investigate their hypothesis, 137 first-year students enrolled in a beginning writing course were asked to write in-class expository essays and to complete a self-assessment of writing survey. The results revealed that the most significant predictor of writing performance was students' self-efficacy beliefs. Students with a strong sense of efficacy wrote better essays than students with weak self-efficacy beliefs. McCarthy et al. (1985) suggest that writers with strong self-efficacy behave differently than those with weak self-efficacy; however, they maintain that additional research is needed to determine exactly what they do differently.

Students' self-efficacy for self-regulation correlates with writing achievement. According to Zimmerman and Bandura (1994), "writing presents special challenges to self-regulation" (p. 846). Writing tasks generally necessitate the use of extensive self-regulatory strategies, as these tasks are often completed alone and require sustained effort and repeated revision (Zimmerman & Kitsantas, 1999). Becoming an effective writer requires high levels of self-regulation, as the writing process consists of three main components: (1) planning and generating ideas, (2) translating ideas into text, and (3) revising and assessing what one has written. These components are cyclical and require the use of effective self-regulatory strategies, such as goal setting, strategic planning, environmental structuring, self-monitoring, and self-evaluation. Zimmerman and Risemberg (1997) contend, "writing self-regulation is a complex system of interdependent processes that are closely linked to an underlying sense of self-efficacy" (p. 95).

Using the Writing Self-Regulatory Efficacy Scale, a timed writing exam, and students' final grades, Zimmerman & Bandura (1994) examined the self-efficacy perceptions of 95 undergraduate students enrolled in a college writing course. Students were asked to estimate their final writing course grades from A to F and to estimate the certainty of their grade predictions on a scale from one (high uncertainty) to seven (high certainty). Students were not asked to estimate their performance on the timed writing exam. Using path analysis, Zimmerman and Bandura (1994) found that students' self-efficacy perceptions influenced their use of self-regulatory strategies and that their self-efficacy beliefs could also be used to forecast their writing achievement. Students rated their efficacy lowest for concentrating on writing when there were many distractions and

highest for locating and using appropriate reference sources to document important points. Students also indicated low efficacy for generating suitable outlines, composing engaging introductory paragraphs, and starting a writing project. The results indicated that students' perceived self-regulatory efficacy for writing influenced their academic achievement and the standards by which they self-evaluated, which were both linked to grade goals. Students with a high sense of personal efficacy set higher writing goals, which influenced the quality of the writing that they produced and with which they were content. The researchers expected verbal aptitude to contribute to writing achievement and self-efficacy; however, verbal aptitude was mediated by self-regulatory factors. The authors urge writing instructors to consider assessing students' self-regulatory efficacy for writing at the start of composition courses, as this information, in addition to indicating areas where students feel less than capable, can provide much needed information that can be used to provide students with self-regulatory strategy training that can enhance both teaching and learning.

While Zimmerman and Bandura's (1994) study serves a sort of framework for this study, this study assesses the accuracy of students' self-assessments, while Zimmerman and Bandura (1994) evaluated students self-satisfaction with end of course grades. Zimmerman and Bandura (1994) found that grades goals were correlated with final course grades; however, final course grades are based on a number of factors that while related to achievement, tell instructors and students very little about their ability to write effectively for a specific writing task. The ability of students to make accurate judgments regarding their performance on a particular writing task has the potential to

provide students with information on their performance that can help them improve their writing on a variety of writing tasks across genres and disciplines.

Measures of Calibration

Accurate self-evaluation, self-efficacy beliefs and self-regulation are critical components of calibration. Calibration is the extent to which confidence matches accuracy when measured across many judgments (Fischhoff, Slovic, & Lichtenstein, 1977). According to Alexander (2013), calibration, at its simplest, is the "degree to which individuals' judgments about their understanding, capability, competence, or preparedness corresponds to the understanding, capability, competence, or preparedness they actually manifest" (p. 1). Despite this straightforward definition, calibration has been operationalized in a number of different ways, and a variety of measures have been used to compute calibration across disciplines (Schraw, Kuch, & Gutierrez, 2013).

The measure of calibration selected for use is dependent upon the construct being measured and the outcomes of the study. Schraw (2009) identified three different classes of calibration judgments: (1) prospective (predictions), (2) concurrent, and (3) retrospective (postdictions). Predictions require an estimate of performance prior to the task; concurrent judgments require an estimate of confidence while performing the task, and postdictions require an estimate of performance after completion of the task. These judgments may be made locally (i.e., item-by-item) or globally (i.e., total performance) (Schraw, 2009).

The format used most often in calibration research requires that students answer a test item and then judge whether one's answer is correct or incorrect (Schraw et al., 2013). Schraw et al. (2013) noted that both continuous and dichotomous judgments have

been used in calibration studies. Continuous measures consist of individual judgments on a Likert or 100-point scale, while dichotomous judgments may use a 2 x 2 model in which accuracy of predictions and postdictions may be compared to actual performance. The 2 x 2 model is considered the best solution, as a two-factor solution using two separate scores is said to provide the best fit; however, no single measure has been found to account for all of the components associated with calibration (Schraw et al., 2013).

The choice of which measure to use in calibration research should be guided by the study's definition of and assumptions about monitoring and the measures used to assess and interpret the findings. Thus, an important distinction must be made between absolute accuracy and relative accuracy. According to Hacker, Bol, and Keener (2008), absolute accuracy and relative accuracy represent two very distinct aspects of monitoring and are measured in different ways. In fact, Schraw (2009) found at least six distinct types of measures used to calculate both absolute and relative accuracy. Both absolute and relative measures of accuracy are needed, and both are important for students to self-regulate their learning (Bol & Hacker, 2012), but it is important to note the differences between the two.

Absolute accuracy (i.e., calibration) is defined as an overall estimate that matches performance exactly, while relative accuracy (i.e., discrimination) is related to one's ability to estimate correct performance on one item relative to another (Bol & Hacker, 2012; Schraw, 2009). Absolute accuracy is ideal for investigating a treatment designed to enhance accuracy and improve student achievement, while relative accuracy measures are useful when investigating the consistency of confidence judgments across a set of items (Bol, Hacker, O'Shea, & Allen, 2005; Schraw, 2009). Absolute accuracy is

calculated using the deviation between a prediction and actual performance, whereby scores close to zero represent high accuracy, while scores close to one represent low accuracy (Schraw, 2009). According to Schraw (2009), relative accuracy "should be interpreted as a correlation" and measured using an index of association, such as the point-biserial correlation, gamma, or Pearson's *r* (p. 425). Simply put, absolute accuracy indicates whether a student can estimate actual overall test performance, whereas relative accuracy indicates whether a person can differentiate between the known and the unknown.

One of the issues that often arises in conducting calibration research concerns the question of which measure to use. Both absolute and relative accuracy refer to how well a student's judgment relates to performance; however, the two types of accuracy are statistically independent (Dunlosky & Thiede, 2013). The decision to use absolute versus relative measures of accuracy should ultimately be influenced by the goals and context of the study. Absolute accuracy measures are preferred for research that seeks to compare judgments with actual performance or to establish if accuracy differs as a result of an intervention. Conversely, relative accuracy measures are recommended in studies that seek to determine if students can discriminate between items they will perform well on versus items they will perform poorly on as well as in studies that seek to determine if students can make consistent judgments across items (Schraw, 2009).

There are a number of ways to calculate absolute accuracy. In an experimental study that sought to determine the influence of overt calibration practice, Bol et al. (2005) calculated the absolute difference between students' predictions and postdictions and their actual exam scores. Undergraduate students enrolled in an education course were

asked to both predict and postdict the percentage of questions they answered correctly on a final exam. The actual scores were then subtracted from the predicted scores; likewise actual scores were subtracted from the postdiction scores. Similarly, Bol, Riggs, Hacker, and Nunnery (2010), asked sixth grade students to predict and postdict their scores on a 50-item multiple-choice math test by estimating their raw scores (i.e., the number of items answered correctly). Absolute accuracy was calculated as the difference between students' calibrated judgments of their raw scores and their actual raw scores. For example, a student predicts that he or she will answer 25 out of 50 questions correctly; thus giving the student a prediction score of 25. The student then answers 25 questions correctly, thus their actual performance score is 25. To calculate absolute accuracy, which in this case is a score of zero, one would find the difference between the prediction score (25) and the actual score (25). Learners are perfectly calibrated to the extent that their predictions and/or postdictions of performance mirror their actual performance.

Relative accuracy is computed as a correlation coefficient. Confidence weighting (i.e., asking students to declare probabilities of correctness or to weigh items to reflect their confidence in choosing the correct answer) is a type of relative accuracy (Hattie, 2013). Dinsmore and Parkinson (2013) computed students' calibration accuracy using a rho coefficient after asking students to provide confidence ratings on a 100-mm line and by using magnitude scaling, wherein students indicate their confidence on an item by comparing it with performance on an anchor item. Higham (2013) used the plurality option, which involves multiple steps, to compute students' calibration on a multiple-choice exam. First, students chose their favorite answer and judged their confidence, and

then students excluded their least favorite answer, and judged their confidence that either the remaining items or their favorite answer was correct.

Calibration is not one-dimensional (Hattie, 2013), and its complexity justifies multiple methods of measurement. Schraw et al. (2013) examined the use of 10 different calibration measures and found that the 2-factor model produced the best indices; however, they indicated that two measures provide a better interpretation of the data than one measure alone, and they recommended the use of multiple measures to assess calibration. Dinsmore and Parkinson (2013) found no differences in students' calibration using the 100-mm line and magnitude scaling.

The literature is mixed as to which calibration measures are best and whether local or global measures are best; however, researchers (Dinsmore & Parkinson, 2013; Schraw, 2009; Schraw et al. 2013) agree that measurement has a direct effect on the calculation of calibration and that the measurement used must meet the assumptions of the data. In short, all calibration measures provide a quantitative appraisal of estimated performance versus actual performance.

Miscalibration

An examination of calibration measures is not complete without a discussion of bias scores. Bias scores can be considered a form of absolute accuracy, especially when the average judgment across items is subtracted from the average level of the target performance (Dunlosky & Thiede, 2013). Bias scores provide a measure of over- or underconfidence in judgment. The bias index represents the direction of the miscalibration, wherein a positive number represents overconfidence and a negative number reflects underconfidence (Schraw, 2009).

Calibration accuracy is important, yet research suggests that few people are perfectly calibrated (Bol & Hacker, 2012; Fischhoff et al., 1977; Kruger & Dunning, 1999; Lichtenstein & Fischhoff, 1977). Miscalibration can be defined as the extent to which confidence is reliably greater than or less than accuracy across judgments. According to Fischhoff et al. (1977), there are two aspects of knowledge: what one believes to be true and one's confidence in that belief. According to Lichtenstein and Fischhoff (1977), most people are only moderately well calibrated, and their confidence judgments suffer from a systematic bias, wherein individuals are overconfident. While level of expertise, may affect calibration accuracy; overall, Lichentenstein and Fischhoff found that "those who know more do not generally know more about how much they know (p. 179). Overconfidence is one example of miscalibration. Research suggests that people typically report confidence levels nearly 20 percent higher than their accuracy levels (Lichtenstein & Fischhoff, 1977).

Lichtenstein and Fischhoff (1977) and Fischhoff et al. (1977) conducted a series of experiments testing calibration accuracy and nearly all respondents were overconfident. However, the researchers found that respondents who were more knowledgeable were better calibrated, and that while all groups were apt to be overconfident, the more knowledgeable group displayed the least amount of overconfidence (Fischhoff et al., 1977; Lichtenstein & Fischhoff, 1977).

Underconfidence is another form of miscalibration. While most people are overconfident, typically, low achievers are far more inaccurate and overconfident than their higher achieving peers (Bol & Garner, 2011). High achievers, especially those who score above 80 percent correct, often become underconfident (Lichtenstein & Fischhoff,

1977), wherein they estimate scores or confidence ratings below their actual performance. Research also suggests that examinees are underconfident when answering easy items, yet overconfident when answering difficult test items (Lichtenstein & Fischhoff, 1977).

The premise that the lowest performing or least knowledgeable lack awareness of their deficits and are thus poorly calibrated and overconfident is also supported by recent research (Kruger & Dunning, 1999; Hacker, Bol, Horgan, & Rakow, 2000). Kruger and Dunning (1999) conducted a number of studies examining the miscalibration of low achievers by examining the logical reasoning and grammar skills of undergraduate students. In the logical reasoning study, 45 undergraduate students enrolled in an introductory psychology course completed a 20-item logical reasoning test and then made three ability estimates. Students provided percentile rankings comparing their logical reasoning abilities and their test scores to their classmates. Finally, the students estimated how many tests questions they answered correctly. Low achievers, students who scored in the bottom quartile, overestimated their logical reasoning ability, their performance relative to classmates, and the number of items answered correctly. The low achievers scored at the 12th percentile on average; however, they estimated that their scores ranked at the 62nd percentile. Conversely, high achievers scored at the 86th percentile, but they estimated that they scored at the 74th percentile. In the grammar study, 84 undergraduate students completed a test that assessed their ability to identify grammatically correct Standard English. Students estimated their ability by providing percentile rankings and raw estimates of the number of items answered correctly. Again, low achievers grossly overestimated their performance. Whereas low achievers scored in the 10th percentile on

average, they estimated their performance to be in the 61st percentile. High achievers underestimated their percentile ranking, scoring in the 89th percentile, but estimating their performance to be in the 70th percentile; however, high achievers did not underestimate their raw scores.

Kruger and Dunning (1999) contend that the lowest achievers are significantly stymied by miscalibration as they often continue to hold the mistaken impression that their performance is fine, even after being provided with evidence of their classmates' superior performance. Thus the miscalibration of low achievers originates from an error about the self, whereas the miscalibration of high achievers originates from an error about others, as high achievers often assume that their performance mirrors that of their classmates.

Miscalibration can have serious consequences for the least knowledgeable, as learners who are unaware of what they do not know will likely fail to critically evaluate their knowledge (Fischhoff et al., 1977) and to take the remedial steps necessary to improve their knowledge (Bol & Garner, 2011). Learning how to accurately calibrate is an important metacognitive skill and an essential self-regulatory process that should prove conducive to learning and academic achievement (Bembenutty, 2009; Bol & Garner, 2011).

Calibration and Achievement

Calibration studies provide an assessment of metacognitive judgments, as learners are commonly asked to make a prediction before the completion of an academic task and an evaluation after the completion of the task, which is then compared with actual performance on the task (Bol & Hacker, 2012). Since calibration accuracy has been

linked to academic achievement, it is important to investigate the extent to which calibration accuracy predicts achievement. Although an overconfidence bias has been found for both low and high achievers in grade school and in college over a variety of item formats, additional research is still needed to solidify the link between calibration and achievement. Particularly needed is research that examines the calibration strategies that enhance student achievement (Bol & Hacker, 2012).

Of the factors that most affect student learning, student expectations and self-reported grades are at the top of the list (Hattie, 2013). Achievement is associated with both prediction and postdiction calibration accuracy (Bol & Hacker, 2001). While much of the early research in calibration occurred in laboratory settings with contrived tasks (Fischhoff et al., 1977; Lichtenstein & Fischhoff, 1977), the need for research in more ecologically valid environments, including classroom contexts, is needed to determine the relationship between calibration and achievement. Early research (Fischhoff et al., 1977; Lichtenstein & Fischhoff, 1977) suggests that more knowledgeable subjects are generally more accurate, and recent research suggests that higher achieving students are generally better calibrated than lower achieving students (Bol & Hacker, 2001; Bol et al., 2010; Hacker et al., 2000). According to Hacker et al. (2000), improved calibration accuracy can help students understand their strengths and weaknesses, thus leading them to develop strategies to improve their performance and academic success.

Calibration accuracy and achievement over time

Several studies have investigated students' calibration accuracy over time. Hacker et al. (2000) examined calibration accuracy and test performance in two undergraduate educational psychology classes that emphasized self-assessment. The performance of the

99 participants was measured using scores on three multiple-choice exams administered over the course of 15-weeks. Students were asked to predict and postdict their performance on each exam by making an estimate of the percentage of items answered correctly. In general, postdictions were more accurate than predictions for all students. Lower performing students were grossly overconfident in their predictions, while higher performing students in the study demonstrated high predictive and postdictive accuracy. Students who scored over 80 percent showed slight underconfidence in their calibrations. Mean judgment differences of over 30 percentage points were found for students who scored below 50 percent, whereas for students who scored over 70 percent, their actual performance and their predicted and postdicted performance differed by less than eight percentage points. Thus, the lowest performing undergraduate students were the least accurate and most overconfident in their calibration judgments. The researchers posit that the lowest performing students "lack not only knowledge of the course content, but perhaps worse, [they] lack an awareness of their own knowledge deficits" (p. 168). As the study was conducted over the course of a semester, it was expected that students' calibration judgments would become more accurate as they gained experience with self-assessment. While high-performing students increased both their predictive and postdictive accuracy, lower performing students continued to show little predictive accuracy. Thus, it seems that lower performing students failed to take into account their previous test performance when making predictions.

Longitudinal studies of students' calibration accuracy provide helpful information about the sources of students' calibrations. Sjostrom and Marks (1994) investigated students' confidence ratings over the course of a semester. In their study, 90 students

enrolled in an introductory psychology course were asked to predict and postdict their performance on 12 multiple-choice tests. Instead of providing percentage correct estimates, students were asked to rate their confidence that they would pass the test or that they had passed the test on a scale ranging from one (not at all confident) to seven (extremely confident). The results indicated that confidence ratings were highest for high-achieving students and that high-achieving students were more accurate in their confidence ratings than low-achieving students. The researchers predicted that differences in confidence ratings between high-achieving students and middle- and low-achieving students would increase over the course of the semester; however, the findings did not support that hypothesis, as a general decrease in confidence was found as the semester progressed.

Generally, postdictions have been found to be more accurate than predictions; however, Sjostrom and Marks' (1994) results revealed that postdictions were only significantly more accurate than predictions on two of the 12 tests. The researchers also hypothesized that students' predictions would remain constant, and this assumption was supported by the findings and more recent research (Bol et al., 2008), as students seemed to disregard prior test performance when making confidence judgments across quizzes.

It might be assumed that students' calibration accuracy could be improved if students were asked to estimate their test scores across several tests. One might assume that this practice with calibration would help students to adopt self-regulatory test taking and studying behaviors that might improve performance. However, Valdez's (2013) investigation of students' absolute accuracy predictions and their performance on six multiple-choice quizzes and a final exam revealed that poorly calibrated students fail to

monitor, select, and apply appropriate test-taking strategies. Twenty-four students in an undergraduate language acquisition course were asked to rate each test item on the probability that the answer was correct by selecting one of four certainty estimates (e.g., 0 percent, 33 percent, 66 percent, and 100 percent). The findings indicated that students provided more accurate judgments on exams completed early in the semester than on exams completed later in the semester. These findings differ from Hacker et al. (2000) who found that students' predictive accuracy increased over the course of a semester, but they are similar to Sjostrom and Marks (1994) who found variations in students' calibration accuracy across tests. Valdez (2013) also found that high-achieving students provided more accurate judgments than low-achieving students. Improving the accuracy of students' calibrations over time has the potential to improve students' self-regulatory efficacy as it would provide an avenue through which students can self-monitor and track their progress.

Calibration accuracy and achievement across domains

Research suggests that those with knowledge in a particular domain are more likely to recognize their knowledge deficiencies than those without knowledge in that domain, and the least knowledgeable often have limited awareness of their knowledge deficits (Kruger & Dunning, 1999). Battistelli, Cadamuro, Farneti, and Versari (2009) investigated university students' ability to accurately self-evaluate performance across a number of domains. Sixty-five undergraduate education students were given three tests - one in arithmetic, one in formal reasoning, and one in linguistics – and asked to estimate the number of items correct, their performance relative to peers, and their general ability in that domain relative to peers. The students were divided into three performance

groups, low, average, and high, and the effect of group membership on the students' performance and calibrations was examined. The results indicated that low achievers overestimated their performance, while high achievers underestimated their performance. For high achievers, the estimated number of correct answers was always less than the actual number of correct answers; this difference was significant for the linguistics task. While low achievers overestimated the number of correct answers, their estimates indicated that they were aware that they had gotten few answers correct. The authors suggested that low achievers make performance attributions that are self-focused instead of task-focused; thus, low achievers overestimate their performance to safeguard their self-esteem. While some overconfidence is likely helpful as it may increase one's self-efficacy beliefs, overly optimistic appraisals of ability may be especially harmful for low-achieving students who may suspend studying early or apply ineffective test-taking behaviors.

Increasing one's domain knowledge expands individuals' insight into their performance, which enhances calibration accuracy. Those with greater domain knowledge slightly underestimate their performance, while those with less knowledge tend to overestimate their performance. Kruger and Dunning (1999) suggest that incompetence often robs people of the ability to realize that the strategies they have adopted to achieve success are flawed. Kruger and Dunning (1999) argue that the skills that lead to competence in a domain are the same skills needed to evaluate one's competence in that domain. Across four studies, Kruger and Dunning (1999) found that participants in the bottom quartile on three domain specific tests grossly overestimated their performance. The researchers attempted to enhance the calibration accuracy of low-

achieving students by improving their domain knowledge. Students were divided into two conditions: one group received a logical reasoning training packet and the other group did not. The researchers hypothesized that training in logical reasoning would increase the competence of the low-achievers and reduce their calibration error. Students in the bottom quartile who received the logical reasoning training became better calibrated in every way; no such increase was found in the calibration of bottom quartile students who did not receive the training. Increasing knowledge in the domain also increased calibration accuracy. It seems that some level of knowledge in a domain is needed for accurate self-assessment; thus, the best tool for increasing calibration is increasing knowledge.

Given the correlations between achievement and calibration, interventions that may be successful in improving calibration merit further investigation. Research has uncovered a number of variables that affect metacognitive control processes, but less empirical evidence is available regarding the variables that enhance metacognitive monitoring processes. Winne (2004) speculates that well-designed learning environments can help students determine the information that requires additional study and review. Bol and Hacker (2001) investigated the effects of practice tests and traditional review on student achievement and calibration accuracy in a graduate research methods course. The researchers randomly assigned one of two sections of the course to the treatment and control condition. Students in the treatment condition were provided with practice tests for exam review, while students in the control condition were provided with traditional test review. Students were asked to predict and postdict their performance on the midterm and final exams. Each exam contained 25 multiple-choice items and five short-

answer/essay items. Significant differences between high-achieving students and low-achieving students were found, with high-achieving students being consistently more accurate in their predictions and postdictions than low-achieving students. The results also indicated that traditional review was more effective at enhancing achievement and calibration accuracy than practice tests. The authors suggest that students in the practice test condition may have narrowly focused their studying on the content of the practice items. Winne (2004) contends that students often choose to re-study items that are "almost fully learned" instead of items that are "definitely not learned," as re-studying information that is "on the cusp of becoming" knowledge requires less effort (p. 482). Given Winne's assertion, practice tests may not be an effective strategy because students may not set appropriately rigorous study goals.

Low achievers may feel highly confident about incorrect responses because the answer was produced fluently, which may endanger metacognitive control processes (Desoete & Ozsoy, 2009). Hacker et al. (2000) and Bol and Garner (2011) suggest that low achievers anchor their calibrations on inaccurate, but optimistic, judgments of their ability. For the lowest performing students, learning how to accurately calibrate could prove a useful metacognitive and self-regulatory strategy, as it may prompt them to consider likely outcomes in advance and to develop more productive studying and test-taking habits.

A persistent finding within the study of confidence judgments concerns the hard-easy effect (Lichtenstein & Fischhoff, 1977). That is, individuals tend to show overconfidence for difficult tasks and underconfidence for easy tasks (Stone, 2000). Task difficulty is considered a component of what one knows; thus task difficulty should lead

to changes in confidence. Bol and Hacker's (2001) investigation of practice tests versus traditional review also includes an examination of whether calibration differed as a function of item format. The researchers found no difference in prediction and postdiction accuracy for high-achieving students; however, low-achieving students were more accurate in their predictions on essay items than on multiple-choice items. Bol and Hacker (2001) suggest that essays require deeper levels of processing, which may account for students' higher predictive accuracy on writing tasks, as deeper processing levels have been associated with better predictive accuracy. While the accuracy of low-achieving students increased on the essay items, overall low-achieving students were much less accurate and much more overconfident than high-achieving students who were more accurate, but somewhat underconfident, especially in their predictions. Stone (2000) suggests that calibration is more accurate when the tasks are challenging but achievable.

In addition to task difficulty and the hard-easy phenomenon, there are other measurement issues that should be considered when assessing calibration levels. Much of the research on calibration has dealt with general knowledge items and multiple-choice items; however, studies on other testing formats are emerging. As suggested by Bol and Hacker's (2001) findings, calibration may differ or writing tasks. Kruger and Dunning (1999) contend that "the skills that enable one to construct a grammatical sentence are the same skills necessary to recognize a grammatical sentence, and thus are the same skills necessary to determine if a grammatical mistake has been made" (p. 1121). Research that focuses on the syntactical aspects of writing in addition to development and analysis are

needed to determine if students are more accurate self-evaluators across all of the elements that generally encompass effective writing.

Rubric-Referenced Self-Assessment

Writing assignments are viewed as a best pedagogical practice; however, increasing class sizes require that students learn how to self-assess their own writing as instructors are often unable to provide extensive, individualized feedback and support to each student (Covill, 2012). Few calibration studies require students to make confidence judgments using rubrics; however, there is a growing body of research that examines the use of guidelines and self-reflection on students' metacognitive monitoring and self-evaluation processes (Hacker, Dunlosky, & Graesser, 1998). Guidelines offer students a way to self-evaluate and self-monitor in much the same way as rubrics do.

Bol, Hacker, Walck, and Nunnery (2012) examined the impact of students' use of guidelines on the calibration accuracy and achievement of 82 high school biology students. The guidelines asked students to self-assess their strengths and weaknesses in biology and were designed to motivate students to self-monitor their learning and content area knowledge. Students were asked to predict and postdict their scores on a biology test comprised of multiple-choice, short answer, and essay items. The students who received guidelines were more accurate in their calibrations than students who did not receive guidelines. Students who received guidelines also received the highest scores on the biology test. The authors suggest that the guidelines may have prompted students to engage in more focused and constructive studying (Bol et al., 2012).

Other studies have shown that guidelines in the form of metacognitive questioning promote domain specific metacognitive knowledge and achievement.

Kramarski and Mevarech (2003) compared middle school students who received or did not receive general and domain-specific metacognitive training in either individual or group settings. Metacognitive training consisted of a series of guiding questions that prompted students to consider the problem solving strategies they might use to answer complex math problems. The results revealed that students who received the metacognitive training outperformed those who did not. Moreover differences were found between the metacognitive and the non-metacognitive groups on domain-specific knowledge. Domain-specific metacognitive knowledge focuses on the unique features of a domain. These findings suggest a need to further investigate the distinctions between general and domain-specific metacognitive knowledge.

In composition, domain-specific metacognitive knowledge might include thinking about one's own thinking (e.g., reflection during the planning process), self-regulation (e.g., monitoring their writing to establish when writer's block occurs and how to overcome it), and self-efficacy beliefs (e.g., beliefs about writing that they bring to their work). In many classroom contexts, students have little explicit knowledge of the criteria needed for success in a particular domain (Hattie, 2013). Rubrics provide students with these much-needed criteria and are similar to guidelines and metacognitive questioning through their use of leveled performance descriptions. Writing rubrics have been adopted as one method of supporting student writing, as they help students self-assess their performance (Covill, 2012). While many calibration studies focus on estimates of individual item correctness or total correctness, studies that ask students to predict their performance by criterion might provide students with helpful information on their writing strengths and weaknesses. Essays do not permit the item-by-item analysis of multiple-

choice tests, and calibration by criterion may prove more constructive than global estimates. In addition, Schunk (2003) recommended students practice with criterion-referenced self-assessment in order to develop and sustain self-efficacy.

Rubric-referenced self-assessment might potentially cultivate students' self-regulatory skills and enhance their achievement. Hillocks' (1986) meta-analysis of secondary and postsecondary student writing revealed that using specific criteria to evaluate writing led students to revise more and resulted in better writing. Rubric use has been correlated with improvements in the quality of student writing (Andrade, 2001; Andrade, Wang, Du, & Akawi, 2009; Rawson, Quinlan, Cooper, Fewtrell, & Matlow, 2005), as rubrics provide specific performance standards to which students should aspire. Zimmerman and Risemberg (1997) suggest that perceptions of competence can be amplified by the successful use of self-regulatory processes. Rubrics have the potential to serve as effective self-regulatory tools as they may be used to promote thinking and learning.

Rubrics also have the potential to serve as working guides as the range of quality levels can serve as scaffolding for student development. Andrade and Du (2005) conducted a qualitative study of undergraduate students' perceptions of rubric-referenced self-assessment. Using focus groups, the researchers asked 14 teacher education students to discuss the ways in which they used rubrics. Students reported positive perceptions of rubric use. The students reported that rubrics helped them by providing direction and clear descriptions of instructors' performance expectations. Most students also reported using the rubric to plan an approach to writing assignments. Students indicated that they used the rubric to self-assess their writing both during and after writing, and many

students' attributed the use of rubrics to improvements in the quality of their writing. One student noted that rubrics are especially helpful in providing an objective lens through which to assess one's own writing and thus to make necessary revisions. Students' reported use of the rubrics is indicative of self-regulatory strategy use such as goal-setting, self-evaluation, and revision. Some students maintained that the use of rubrics helped them become better writers in other courses, suggesting that students use rubrics and the self-regulatory strategies that result from rubric use to develop an understanding of quality writing across disciplines and genres.

Rubrics provide students with the opportunity to think about their writing and to evaluate their own criteria for success against the standards outlined in the rubric. Andrade and Boulay (2003) examined the role of rubric-referenced self-assessment and hypothesized that self-assessment could support student learning and skill development. Using self-regulated learning and assessment of writing theories, the researchers examined whether a formal process of rubric-referenced self-assessment had a measureable effect on the writing of seventh and eighth grade students. Students were placed in two groups; one group received formal lessons in self- assessment while the other did not. Students in both groups wrote two essays and received the same rubrics that articulated the criteria for quality. It was predicted that students who received formal instruction in self-assessment would produce better writing than students who did not; however, the results suggested that there was no effect of treatment on essay scores. The analysis, however, did indicate a positive relationship between self-assessment and writing for girls, but no relationship with boys' writing when controlling for prior writing ability. The researchers suggest that rubric-referenced self-assessment may have a

positive relationship with girls' metacognitive processing, but they maintain that the intervention was insufficient to obtain a consistent, measurable effect. The researchers did not collect data on students' actual assessments of their essays, but students in the treatment condition were asked to provide evidence that their essays met the rubric's criteria by highlighting that information in their essays. The authors maintain that students were often surprised when they were unable to provide evidence, and they argue that the mechanisms associated with rubric-referenced self-assessment deserve further study despite their study's lack of statistically significant results.

Covill (2012) examined undergraduate students' use of a writing rubric and its effect on their writing quality and their self-efficacy for writing. Fifty-six students enrolled in a 200-level psychology course were separated into groups and provided with one of three self-assessment tools: a long rubric with eleven criteria, a short rubric with five criteria, and no rubric. Students were asked to use the assigned self-assessment instrument while writing a five-page essay that required analysis and application of course materials. Students in each condition were then asked to complete a self-efficacy in writing survey. Students were not asked to specifically rate their performance using the rubric nor were they asked to estimate their final score on the essay assignment; however, Covill (2012) argued that the self-assessment tool used influenced students' writing beliefs. While the results revealed no differences in the writing quality of students based on condition or rubric length, students who used the rubrics reported referring to the rubric throughout the writing process noting that they used the rubric to plan, draft, and revise their essay. Despite the lack of differences in students' performance on the writing task, students who used the long rubric indicated that the rubric helped them to write

better and that it made them more aware of what they needed to do to write an effective essay. Covill (2012) argued that a long rubric might influence students' thinking and writing practices more than a short rubric, as students who used the long rubric reported that the rubric heightened their awareness of the strategies needed to write an effective essay. Long rubric users also perceived the rubric as being potentially helpful for writing in general, and they believed that the requirement to self-assess was especially beneficial when they began their first and final drafts.

Covill (2012) argued that use of the rubrics provided students with more extensive processing abilities, as the rubrics relieved students of the need to store the relevant criteria required for effective writing in their working memory. It is worth noting that students who were not provided with a rubric did not perceive the self-assessment requirement as being worthwhile while writing their first or final drafts. While the rubric used affected neither students' writing quality or their self-efficacy for writing, the majority of students believed that being required to self-assess their writing made them set more specific goals for writing and caused them to work harder on the essay than they would have otherwise.

While rubrics were initially intended to solve "the problem of disagreement" between raters (Broad, 2003), student use of rubrics for self-assessment invites students to engage in the evaluation process and to grow as both writers and assessors. Rubrics have the potential to extend research on writer's self-regulation and self-efficacy, as the expectations articulated by rubric criteria can be used to channel students' effort and thereby enhance performance. Research suggests that writers with strong self-regulatory efficacy behave differently than writers with weak self-regulatory efficacy (McCarthy et

al., 1985). While some of those differences, obviously, involve factors such as control, planning, and rhetorical knowledge, rubrics provide criteria that help to determine the areas in which poor writers need to improve.

Andrade and Boulay (2003) indicate that it is not enough to simply provide students with rubrics, but rather students must be taught to self-assess using the rubrics if writing quality is to be enhanced. Much like calibration alone, rubric use alone does not significantly impact achievement; however, requiring students to self-assess seems to foster positive academic qualities that could increase student success over time. Rubrics make expectations clear and articulate standards for writing. As such, requiring that students calibrate using rubrics should help them to learn in a way that they cannot learn from holistic scores alone.

Summary and Hypotheses

Hacker, Keener, and Kircher (2009) posit that the act of writing itself is applied metacognition, yet calibration research that examines the accuracy of students' judgments regarding their own writing is limited. Calibration techniques often are used to assess metacognition in relation to learner comprehension and learner performance on multiple-choice tests (McCormick et al., 2012). While calibration studies have been conducted comparing multiple-choice items and short-answer/constructed response items (Bol & Hacker, 2012; Pallier et al., 2002), studies that focus specifically on academic essay writing are minimal. Self-regulation and self-efficacy have been linked to writing achievement (Covill, 2012; Negretti, 2012; Prat-Sala & Redford, 2012; Zimmerman & Bandura, 1994), and self-regulation and self-efficacy have been correlated with

calibration accuracy, thus calibration studies that focus specifically on academic essay writing are warranted.

This chapter provided an overview of self-regulation, self-efficacy, and calibration and highlighted how the processes are intertwined. Calibration is a metacognitive function that is closely linked with self-efficacy and achievement. Research, driven in part by the positive relationship between self-efficacy and writing achievement, has illuminated the characteristics of effective writers that differ from ineffective writers. For example, self-regulated and self-efficacious writers use rubrics to self-assess their writing. However, the accuracy of writers' self-assessments has yet to be fully investigated.

This review of the literature supports the need for accurate calibration and outlines how rubric-referenced self-assessment might enhance calibration accuracy. Calibration research that focuses specifically on writing is limited. Nonetheless writing is fundamental to teaching and learning in higher education, and calibration has the potential to raise students' awareness of the conventions within which they are expected to write by requiring that they examine their own writing. With enhanced calibration skills, students can more successfully monitor their writing processes and quite possibly become better writers.

The purpose of this study was to determine whether calibration condition effects calibration accuracy and whether self-efficacy and achievement interact with condition to influence calibration accuracy. In the present study, it was hypothesized that students' calibration accuracy would differ by calibration condition. It was also expected that self-efficacy would differ by calibration condition. It was also hypothesized that calibration

accuracy would differ by calibration criteria. It was hypothesized that calibration condition and achievement would interact to influence calibration accuracy. Finally it was hypothesized that calibration condition and self-efficacy would interact to influence calibration accuracy.

This study is significant as it merges self-assessment strategies (i.e., calibration and rubric-referenced) and engages students in the self-assessment process. Previous research has provided mixed results as to whether calibration practice increases calibration accuracy (Bol et al., 2010; Hacker et al., 2008). Given these mixed results, additional testing of strategies to increase calibration accuracy is needed. Using specific criteria to calibrate is one such strategy. In addition, self-assessment is important to improving one's writing. Accurate self-assessment offers students an opportunity to review and reflect upon the development of their writing and to identify goals and strategies to become more effective writers.

CHAPTER III

METHODOLOGY

This chapter describes the design, participants, procedure, data collection and data analysis techniques that comprise this study. The primary questions addressed concern whether calibration type and self-efficacy affect the calibration accuracy of college undergraduates or their performance on a writing exam. Students' writing self-regulatory efficacy and their judgments of performance both before and after the exam were examined to determine if students' self-efficacy beliefs differed based on calibration condition. Specifically, the following research questions were addressed:

1. Does calibration accuracy in writing differ by calibration condition?

   Hypothesis: Calibration accuracy in writing will differ by calibration condition.

2. Does self-efficacy in writing differ by calibration condition?

   Hypothesis: Self-efficacy in writing will differ by calibration condition.

3. Does calibration accuracy differ by criteria?

   Hypothesis: Calibration accuracy will differ by scoring criteria.

4. Do calibration condition and achievement level interact to influence calibration accuracy?

   Hypothesis: Calibration condition and achievement level will interact to influence calibration accuracy.

5. Do calibration condition and self-efficacy level interact to influence calibration accuracy?

   Hypothesis: Calibration condition and self-efficacy level will interact to influence calibration accuracy.

Research Design

This experimental study was conducted over the course of one academic semester. The study used a calibration strategy that required students to estimate their performance both before and after completion of an essay exam. Specifically, the intervention used consisted of a global and criterion based calibration strategy, which required students to predict and postdict their performance either globally or both globally and by criteria. The researcher randomly assigned sections of the Examination of Writing Competency to one of three conditions: (1) a global condition, (2) a global and criteria condition, and (3) a global and detailed criteria condition. All students estimated their performance globally (i.e., their total score from one to 20); however, depending on condition, some students rated their performance by criteria (i.e., their score on one of the four scoring criteria from one to five). For example, students predicted their global (i.e., total) score from one to 20, or they predicted their global score and their criterion score based on four writing categories scored from one to five. Table 1 outlines the calibration conditions.

Table 1
Calibration Conditions and Procedures

Calibration Condition	Before the exam	After the exam
Global	Global Score Prediction Writing Self-Regulatory Efficacy Scale	Global Score Postdiction
Global and Criteria	Global Score Prediction Criteria Scores Prediction Writing Self-Regulatory Efficacy Scale	Global Score Postdiction Criteria Scores Postdiction
Global and Detailed Criteria	Global Score Prediction Detailed Criteria Scores Prediction Writing Self-Regulatory Efficacy Scale	Global Score Postdiction Detailed Criteria Scores Postdiction

In addition to calibrating their scores for the exam, students completed the 25-item Writing Self-Regulatory Efficacy Scale before completing the exam. The independent variables used in the study were calibration condition, achievement level (i.e., low-achievers and high-achievers), and self-efficacy level (i.e., low self-efficacy and

high self-efficacy). SAT critical reading scores and global writing scores determined achievement level. Scores on the Writing Self-Regulatory Efficacy Scale determined self-efficacy level. The dependent variables were prediction and postdiction accuracy.

Participants

Urban Comprehensive Midsized University (UCMU) is a public, comprehensive, mid-sized university in the southeastern United States serving approximately 7,000 undergraduate and graduate students. Students registered to take the Examination of Writing Competency (EWC), an essay exam, offered in Spring 2014 were the target sample for this study. In 2012, the average SAT score of the entering first-year class was 885 points, and the mean high school grade point average was 2.84. The acceptance rate for the 2012 – 2013 academic year was 67 percent. The Carnegie Classification of the university is Master's Level 1, and the 2012 student to faculty ratio was 19:1 (Fact Book, 2012-2013).

All undergraduate students must successfully complete the EWC in order to graduate. Students are eligible to take the EWC after successful completion (i.e., final course grades of C or higher) of ENG 101: Communication Skills I and ENG 102: Communication Skills II or comparable transfer courses in writing and composition. Over 1,500 students take the EWC each academic year. All undergraduate students registered for the EWC offered in the Spring 2014 semester, approximately 750 students, were targeted for inclusion in the sample. While the goal was that all students registered for the EWC, (approximately 1,260), would complete the calibration forms and survey, it was understood that not all students would agree to participate. As there is no fee for registration and no penalty for withdrawing, not all students registered for the EWC

would take the exam as scheduled. Consequently, the targeted sample size for this study was 500 students.

The sample for this study consisted of 596 students. Approximately 73.40 percent of the 812 students who completed the EWC in Spring 2014 participated in this study. Over 62 percent of participants were female and nearly 32 percent were male. This sample is typical of the University's population as Fall 2013 enrollment indicates that 64.8 percent of students were female, and 35.2 percent were male (University Fact Sheet, 2013). Five percent of students did not indicate a gender on the demographic information form. The majority of students, 54.2 percent, who participated were seniors. Nearly 25 percent were juniors, 17.6 percent were sophomores, and 3.2 percent were freshmen. All five of the University's schools and colleges were represented in the sample. The majority of students were in the College of Liberal Arts (27.9 percent), the University's largest college. Students enrolled in the College of Science, Engineering, and Technology were the second highest represented group at 27.2 percent. The remaining students represented the School of Business (16.8 percent), the School of Education (8.6 percent) and the School of Social Work (6.9 percent). Descriptive statistics by students' academic major are provided in Appendix B.

An analysis of variance was conducted to determine if the groups differed in terms of achievement as measured by SAT critical reading scores. The groups did not differ statistically based on SAT critical reading scores, $F(2, 324) = 1.519, p < .05$.

Measures

To answer this study's research questions, three measures were used. Participants' calibrations, both predictions and postdictions of their writing scores, were used to

examine calibration accuracy. Scores on the Examination of Writing Competency (EWC) were used to measures participants' writing achievement. Finally, the Writing Self-Regulatory Efficacy Scale was used to measure participants' self-efficacy.

Calibration

The three test calibration forms (Appendices C, D, and E) used in this study require participants to estimate either their global or both their global and criteria scores in an effort to determine if detailed scoring guides (i.e., the EWC Rubric) increased monitoring, thus increasing accuracy. Students' predictions and postdictions on the calibration forms were used to calculate absolute accuracy. Absolute accuracy is the difference between the performance estimate and actual performance. For example, the researcher calculated absolute global prediction accuracy as the students' prediction minus actual performance:

Absolute Accuracy = Global Prediction – Actual Global Score

Participants recorded their predictions before the exam and their postdictions after completion of the exam on the appropriate test calibration form. The Global Test Calibration Form was used only in the global condition, and it asked students to estimate their total score from one to 20 both before and after the exam. The Global and Criteria Test Calibration Form was used only in the global and criteria condition, and it asked students to estimate their total score from one to 20 both before and after the exam and to estimate their score on four criteria from one to five both before and after the exam. The Global and Detailed Criteria Test Calibration Form was used only in the global and detailed criteria condition. The form asked students to estimate their total score from one to 20 both before and after the exam. The form also included the full EWC Rubric, and

students were asked to estimate their scores using the criteria and descriptors provided in the rubric from one to five both before and after the exam.

EWC Rubric

The EWC Rubric contains four criteria: (1) organization, (2) development and analysis (3), sentence structure, and (4) grammar, diction, and mechanics, scored on a scale from one to five, where one represents incompetence and five represents superior competence. Schraw (2009) indicates that multiple judgments (e.g., predictions, concurrent judgments, and postdictions) are recommended as test and task parameters, as well as the outcome measure used to evaluate judgments affect accuracy. In this study, both global and criteria judgments were explored because Schraw (2009) contends that outcome measures designed to compute absolute accuracy should be used when investigating the implementation of a monitoring treatment.

One key to rubric validity is the careful selection of criteria that match the concepts assessed. The criteria used in the EWC Rubric are also used in a number of both holistic and analytic scoring rubrics (Huot, 1990). Sadler (2009) suggests that there is a large pool of potentially valid criteria for use in the development of writing rubrics. Sadler's (1989) review identified 50 criteria used for assessing the quality of written composition. These criteria include, but are not limited to organization, development, depth of analysis, mechanics, paragraphing, punctuation, sentence structure, spelling, syntax, and vocabulary. Jeffery (2009) explored constructs of writing proficiency in direct large scale writing assessments through content analysis of rubrics and found that development, coherence, accuracy, and organization are common descriptors used to

define the features of good writing. These descriptors or their synonyms can be found in the EWC Rubric.

Examination of Writing Competency

The Examination of Writing Competency (EWC) is a three-hour proctored writing exam. The EWC consists of three writing prompts. Students must write a 500-600 word essay in response to one of the three writing prompts. The EWC is administered each term and serves as a graduation requirement for all undergraduate students. The EWC was administered to all participants. A retired EWC question form is included in Appendix F. The EWC Rubric, which was locally developed in 2005 by the EWC Coordinator and faculty from the Department of English and Foreign Languages and revised in 2009, was used to score the exams.

To determine EWC scores, two raters evaluate each exam using the EWC Rubric. If the two raters disagree on the pass/no pass rating, a third rater evaluates the essay. Students must receive scores of three or greater in each category in order to pass the exam. For the purposes of this study, the researcher, who trains raters and serves as the third rater when two raters disagree on the pass/fail rating, scored all 596 exams using the EWC Rubric. The researcher has a Master's degree in English with 18 graduate hours in composition and over five years experience teaching college composition. The actual scores used in this study are the scores provided by the researcher.

Since rater judgments of constructed response tasks are often subjective, the intra-class correlation (ICC), a measure of reliability, was used to assess inter-rater reliability on the EWC. ICC is one of the most conservative measures of interrater reliability, and it was deemed most appropriate for this study because it incorporates the magnitude of

disagreement between the raters, whereas Cohen's kappa quantifies reliability based on all-or nothing agreement. ICC is appropriate for studies with two or more raters and may be used when all subjects in a study are rated by multiple evaluators or when only a subset of the subjects are rated by multiple evaluators and the rest are rated by one evaluator (Hallgren, 2012). To calculate ICC, a second trained rater, a college instructor with a Ph.D. in Education scored 20 percent of the exams using the EWC Scoring Rubric. For total global scores, ICC was .81. For the first criteria, organization, ICC was .69. For the second criteria, development and analysis, ICC was .77. For the third criteria, sentence structure, ICC was .66. For the final criteria, grammar, diction, and mechanics, ICC was .82. According to Cicchetti (1994), ICC values between .60 and .74 indicate good reliability, and values between .75 and 1.0 indicate excellent reliability.

Writing Self-Regulatory Efficacy Scale

Zimmerman and Bandura (1994) found writing self-efficacy to be a meaningful construct that at times surpassed verbal ability in its predictions of writing outcomes. As such, participants in each of this study's conditions completed the Writing Self-Regulatory Efficacy Scale (Appendix E) before the exam. The scale, developed by Zimmerman and Bandura (1994), consists of 25 items that assess students' perceived abilities to execute strategic aspects of the writing process. Students rated their perceived efficacy for each item on a seven-point scale ranging from one (not well at all) to seven (very well). Cronbach's alpha yielded a reliability coefficient of .975 for the sample used in this study.

Procedure

The EWC is administered each term as a part of the normal assessment requirements of the university. The only difference in administration was that participants were asked to estimate their performance before and after the exam and to complete the Writing Self-Regulatory Efficacy Scale. Test sessions were randomly assigned to one of three conditions: (1) a global condition, (2) a global and criteria condition, and (3) a global and detailed criteria condition. The administration procedures are illustrated in Figure 2.

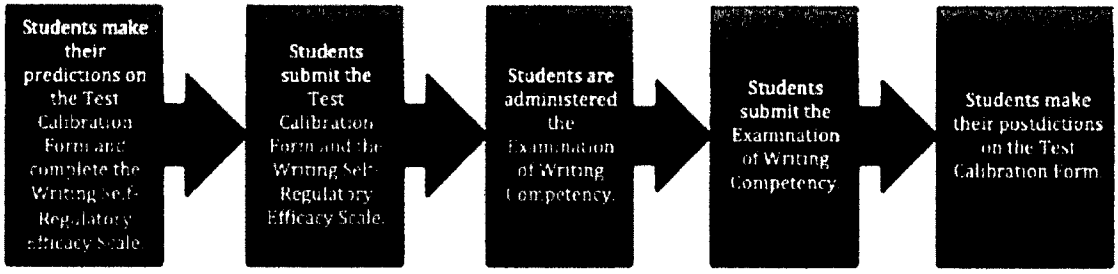


Figure 2. Administration Procedures

Students in the global condition predicted their total global performance from one (incompetence) to 20 (superior competence) on the appropriate test calibration form before the exam. Then students completed the Writing Self-Regulatory Efficacy Scale. Both measures were collected before the exam was distributed. After the exam was submitted, students then postdicted their total global performance from one (incompetence) to 20 (superior competence).

Before the exam, students in the global and criteria condition predicted their total global performance from one (incompetence) to 20 (superior competence) and their performance on each criteria from one (incompetence) to five (superior competence) on the appropriate test calibration form. Then students completed the Writing Self-

Regulatory Efficacy Scale. Both measures were collected before the exam was distributed. After the exam was submitted, students postdicted their total global performance from one (incompetence) to 20 (superior competence) and their performance on each criterion from one (incompetence) to five (superior competence).

Before the exam, students in the global and detailed criteria condition predicted their total global performance from one (incompetence) to 20 (superior competence) on the appropriate test calibration form. Students then predicted their performance on each criterion from one (incompetence) to five (superior competence) using the EWC Rubric. Students then completed the Writing Self-Regulatory Efficacy Scale. Both measures were collected before the exam was distributed. After the exam was submitted, students postdicted their total global performance from one (incompetence) to 20 (superior competence) and their performance on each criteria from one (incompetence) to five (superior competence) using the EWC Rubric.

Students did not have access to the rubric or the calibration forms while completing the exam. Students were provided with the notification letter (Appendix F), which details the study's purpose, the potential risks and benefits of participation, the voluntary nature of the study, and confidentiality assurances before the study.

Protections

Participants' performance estimates or calibrations on the rubric and their responses to the survey remained confidential. The researcher collected and stored hardcopies of the rubric and survey in a locked file cabinet after participants completed them. The EWC was scored and stored per the university's policy for student academic records. Hardcopy score reports were stored in a locked file cabinet by the researcher,

and electronic score reports were stored in a password-protected file for the duration of data collection and analysis. Notification letters were provided to students. While the normal administration procedures of the exam require students to complete a separate student demographic information form that includes students' names, students were asked to include only their student identification numbers on the exam, the calibration form, and the survey. Once data was collected student identification numbers were replaced with researcher selected identification numbers to protect participant confidentiality. Approval to conduct this study was granted by the Darden College of Education's Human Subjects Review Committee. The letter of approval is provided in Appendix G.

Data Analysis

Demographic data from the student information form, as well as test and survey results were analyzed descriptively and presented in tables. Prediction and postdiction accuracy was calculated as the difference between the students' predicted or postdicted scores and their actual scores. Bias scores were computed based on the direction of the calibration, wherein a positive number represents overconfidence and a negative number reflects underconfidence.

A multivariate analysis of variance (MANOVA) was used to determine the impact of calibration condition on calibration accuracy. To investigate the relationship between calibration and achievement, a median split was used to divide students into low- and high-achieving groups based on SAT critical reading scores and global writing scores. Participants who scored above the median were categorized as high achievers and those who scored below the median were categorized as low achievers. In order to

investigate the relationship between self-efficacy and calibration accuracy, an additional MANOVA was conducted. A median split was used to divide the participants into two groups based on self-efficacy; those who scored above the median on the Writing Self-Regulatory Efficacy Scale were classified as having high self-efficacy, and those who scored below were classified as having low self-efficacy. The dependent variables in the various analyses conducted included global prediction accuracy, global postdiction accuracy, and prediction and postdiction accuracy by criteria. The independent variables in the various analyses conducted included condition, achievement, as measured by SAT CR scores and EWC global scores, and self-efficacy, as measured by scores on the Writing Self-Regulatory Efficacy Scale.

Table 2 presents the variables and analyses used to answer each research question. The multivariate analysis of variance (MANOVA) was run to test for the linear composite or vector of the means between groups. The MANOVA is appropriate for this study's analyses as it maximizes the difference between the groups of the independent variables and tests for statistically significant differences between those groups. The assumptions that underlie the MANOVA were tested and satisfied. An analysis of variance (ANOVA) was also conducted and deemed appropriate for this study as this study's design measured the same dependent variable in three independent groups.

Table 2

Analysis by Research Question

Research Questions	**Independent Variable(s)**	**Dependent Variable(s)**	**Analysis**
Does calibration accuracy in writing differ by calibration condition?	Calibration condition	Global prediction accuracy Global postdiction accuracy	MANOVA
Does self-efficacy in writing differ by calibration condition?	Calibration Condition	Self-efficacy level	ANOVA
Does calibration accuracy differ by criteria?	Calibration Condition (criteria conditions only)	Prediction accuracy in organization, development & analysis, sentence structure, and grammar, diction & mechanics Postdiction accuracy in organization, development & analysis, sentence structure, and grammar, diction & mechanics	MANOVA
Do calibration condition and achievement level interact to influence calibration accuracy?	Calibration Condition Achievement Level (using median split SAT critical reading scores and EWC global scores)	Global prediction accuracy Global postdiction accuracy	Factorial MANOVA
Do calibration condition and self-efficacy level interact to influence calibration accuracy?	Calibration Condition Self-Efficacy Level (using median split Writing Self-Regulatory Efficacy Scale scores)	Global prediction accuracy Global postdiction accuracy	Factorial MANOVA

CHAPTER IV

RESULTS

The findings presented in this chapter are organized by research question. The results of the analyses used to evaluate the effect of calibration condition on the calibration accuracy of undergraduate students on the Examination of Writing Competency (EWC) are presented in this chapter. Self-efficacy was also analyzed to determine its effect on calibration accuracy and to establish if calibration condition in combination with self-efficacy interacts to influence calibration accuracy. A total of 596 students participated in this study, and 418 students calibrated both globally and by criteria before and after the exam. The remaining 178 students calibrated globally before and after the exam only. The students who calibrated globally only are referred to as the global condition. Students who calibrated both globally and by criteria are referred to as the global and criteria condition, while students who calibrated both globally and with detailed criteria (i.e., the EWC Rubric) are referred to as the global and detailed criteria condition. The results of the analyses used to answer the study's research questions follow.

Calibration Condition and Calibration Accuracy

The first research question sought to determine if calibration accuracy differed by condition. Absolute accuracy varied little across conditions (see Table 3). The EWC is scored on a 20-point scale. Students received a score ranging from one to five in four criteria. Thus for this task, absolute accuracy ranged from 0 to 20. Students' actual scores deviated about four points from their predictions and postdictions. Students' predictive and postdictive accuracy were best in the global condition and least accurate in

the global and criteria condition. Students' postdictions were slightly more accurate than their predictions. Considering the range of scores, the differences in students' actual scores and their predictive and postdictive accuracy could be considered quite large, as four points would represent 20 percent of the global score.

Table 3

Descriptive Statistics for Prediction and Postdiction Accuracy by Calibration Condition

	Global			Global & Criteria			Global & Detailed		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Global Prediction	175	4.17	3.39	227	4.56	3.32	176	4.31	2.78
Global Postdiction	175	3.84	3.49	227	4.41	3.42	171	4.17	3.09

Bias, the signed difference of the absolute accuracy calculation, indicates whether students were over or underconfident in their predictions and postdictions. A positive sign, reflects overconfidence; whereas, a negative sign reflects underconfidence. Across all conditions students were overconfident (see Table 4).

Table 4

Bias (Signed Accuracy) for Predictions and Postdictions

	Global Predictions			Global Postdictions		
	n	*M*	*SD*	*n*	*M*	*SD*
Bias	578	+ 4.37	3.19	563	+4.16	3.35

A one-way multivariate analysis of variance (MANOVA) was run to determine the effect of calibration condition on calibration accuracy. Two measures of calibration accuracy were assessed: prediction accuracy and postdiction accuracy. Preliminary

assumption checking revealed that the data were normally distributed, as assessed by the Shapiro-Wilk test ($p > .05$); however, there were a few univariate and multivariate outliers, as assessed by boxplot and Mahalanobis distance, respectively. Because of the large sample size and since the outliers were not extreme points, the outliers were not removed from the data or transformed because they were not viewed as materially affecting the results as the data were normally distributed. There were linear relationships, as assessed by scatterplots. There was no multicollinearity ($r = .784$) for global prediction and postdiction accuracy as represented by the Pearson correlation coefficient. For a MANOVA to provide valid results, the assumption that there is no multicollinearity must be tested. A relatively simple way to detect multicollinearity is by using the Pearson correlation coefficient. Mayers (2013) suggests there should be reasonable correlation between the dependent variables. Positive correlations should not exceed $r = .90$, and negative correlations should not exceed $r = -.40$ (Mayers, 2013). A more sophisticated method of detecting multicollinearity is the variance inflation factor (VIF). The values of VIF for the dependent variables were all below 3. Generally VIF scores between 4 and 10 indicate excessive or serious multicollinearity (O'Brien, 2007). There was homogeneity of variance-covariance matrices, as assessed by Box's M test ($p = .002$). A violation of this assumption would have resulted in a statistically significant p-value (i.e., $p < .001$). There was no effect of calibration condition on calibration accuracy for global predictions, $F(2, 551) = .485, p > .05, \eta^2 = .002$ or for global postdictions, $F(2, 551) = 1.940, p > .05, \eta^2 = .007$.

Table 5

MANOVA Results for Condition and Calibration Accuracy

	df	*F*	*P*	η^2
Global Prediction	2, 551	.485	.616	.002
Global Postdiction	2, 551	1.940	.145	.007

Calibration Condition and Writing Self-Efficacy

The second research question attempted to determine if self-efficacy in writing differed by condition. An ANOVA was conducted to determine if self-efficacy in writing differed when judgments were provided by condition. Self-efficacy in writing did not differ statistically by condition, $F(2, 587) = .113, p > .05, \eta^2 = .000$.

Self-efficacy as measured by the Writing Self-Regulatory Efficacy Scale differed little by condition (see Table 6). Self-efficacy scores ranged from zero to seven. Students in the global and criteria condition reported the highest self-efficacy, while students in the global condition reported the lowest self-efficacy. There was virtually no difference in mean self-efficacy scores by condition. All conditions scored very close to five, which suggests that most students rated themselves as having high self-efficacy.

Table 6

Mean Self-Efficacy Scores by Condition

	n	*M*	*SD*
Global	177	4.94	1.16
Global & Criteria	229	4.97	1.05
Global & Detailed Criteria	184	4.99	1.03

Calibration Accuracy by Criteria

The third research question examined calibration accuracy by criteria: (1) organization, (2) development and analysis, (3) sentence structure, and (4) grammar, diction, and mechanics. A MANOVA was run to determine the effect of calibration condition on calibration accuracy by criteria. Prediction and postdiction accuracy by criteria were assessed for students in the global and criteria condition and students in the global and detailed criteria condition. Preliminary assumption checking revealed a few univariate outliers as assessed by boxplot (see Appendix P); however, no multivariate outliers were revealed. The outliers were not considered excessive, and thus they were not removed from the data. There were linear relationships and no multicollinearity as assessed by the Pearson correlation coefficient (see Appendix Q). The difference in calibration accuracy was not statistically significant by criteria (see Table 7).

Table 7

MANOVA Results for the Effect of Calibration Condition on Calibration Accuracy by Criteria

	df	*F*	*p*	η^2
Prediction 1	1, 375	.002	.963	.000
Postdiction 1	1, 375	1.562	.212	.004
Prediction 2	1, 375	.879	.349	.002
Postdiction 2	1, 375	.328	.567	.001
Prediction 3	1, 375	2.798	.095	.007
Postdiction 3	1, 375	.909	.341	.002
Prediction 4	1, 375	.176	.675	.000
Postdiction 4	1, 375	2.268	.133	.006

While the results were not statistically significant by criteria, students in the global and detailed criteria condition were more accurate in their predictions across all four criteria than students in the global and criteria condition. On average, students' postdictions were more accurate than their predictions. In the global and detailed criteria condition, students' postdictions were less accurate than their predictions in one criterion (i.e., development and analysis). Students' calibrations all showed a positive bias, wherein students were overconfident. On average, students' criteria-based predictions and postdictions deviated from their actual criteria scores by a little over one point. Since the criteria scores range from one to five, students could be considered quite overconfident by criteria, as a score of one would represent 20 percent of the total criterion score.

Table 8

Descriptive Statistics for Prediction and Postdiction Accuracy by Criteria

	Global & Criteria		Global & Detailed	
	M	*SD*	*M*	*SD*
Organization				
Prediction	1.06	1.01	1.04	.88
Postdiction	1.09	.99	.97	.89
Development & Analysis				
Prediction	1.01	1.01	.91	.92
Postdiction	.98	1.04	.94	.96
Sentence Structure				
Prediction	1.12	.91	.96	.82
Postdiction	1.00	.96	.91	.89
Grammar, Diction, & Mechanics				
Prediction	1.08	.94	1.06	.84

Postdiction	1.12	.92	.98	.95

Impact of Calibration Condition and Achievement on Calibration Accuracy

The fourth research question sought to determine if calibration condition and achievement level interacted to influence calibration accuracy. Two measures of achievement, SAT critical reading (CR) scores and EWC global scores, were used and two separate factorial MANOVAs were conducted using each achievement measure. First, the results of the factorial MANOVA conducted to determine if calibration condition and SAT CR scores interacted to influence calibration accuracy are presented followed by the results of the factorial MANOVA conducted to determine if calibration condition and global EWC scores interacted to influence calibration accuracy.

Impact of Calibration Condition and SAT CR Scores on Calibration Accuracy

SAT CR scores differed little by condition. While students in the global and criteria condition had the highest SAT CR scores, the difference in scores by condition was not statistically significant. Table 9 provides SAT CR scores by calibration condition.

Table 9

Mean SAT CR Scores by Calibration Condition

Calibration Condition	*n*	SAT CR *M*
Global	94	422.98
Global & Criteria	137	419.56
Global & Detailed Criteria	94	435.11

A median split was used to categorize students as high or low achievers based on SAT CR scores. The split resulted in 170 low achievers and 136 high achievers.

Calibration condition and achievement level, as measured by SAT CR scores, did not interact to influence the accuracy of global calibration predictions, $F(2, 300) = 2.202, p > .05, \eta^2 = .014$ or global calibration postdictions, $F(2, 300) = .665, p > .05, \eta^2 = .004$.

Though the results were not statistically significant, an interesting finding emerged. Based on SAT CR scores, the actual EWC scores of low and high achievers differed by about one point. The mean EWC score of high achievers was 12.03, while the mean score of low achievers was 11.00. For the students in this study, the average SAT CR score was 425, which would place students in the 25th percentile nationally. Nationally, for the class of 2013, the average SAT CR score was 497 (College Board, 2014). Using SAT scores as a measure of achievement, the students in this study would generally be considered low achievers.

Table 10

Global Prediction and Postdiction Accuracy by Condition and Achievement (SAT CR)

		High Achievers (Score > 420)			Low Achievers (Score < 420)		
		n	*M*	*SD*	*n*	*M*	*SD*
Prediction							
	Global	37	3.73	3.26	54	4.33	3.16
	Global & Criteria	55	3.31	3.85	72	5.01	2.70
	Global & Detailed Criteria	44	3.98	2.59	44	3.93	2.72
Postdiction							
	Global	37	2.81	3.24	54	4.06	3.21
	Global & Criteria	55	3.84	3.78	72	4.94	2.74
	Global & Detailed Criteria	44	3.96	2.77	44	4.21	3.17

Impact of Calibration Condition and SAT CR Scores on Calibration Accuracy by Criteria

The impact of calibration condition and achievement on students' calibration accuracy by criteria were examined for the two criteria conditions. Calibration condition and achievement level, using SAT CR scores, were found to be significant for two criteria: (1) organization and (2) development and analysis.

Calibration condition and achievement level interacted to significantly effect prediction accuracy in organization, $F(1, 205) = 4.531, p < .05, \eta^2 = .022$. Using SAT CR scores, overall both high and low achievers in the criteria conditions were overconfident in their predictions and postdictions in organization.

Figure 3 illustrates the interaction of calibration condition and achievement level on prediction accuracy in organization for the two criteria conditions. In the global and criteria condition, there was a large difference in the calibration accuracy of high and low achievers, with low achievers being far less accurate than high achievers. However, in the global and detailed criteria condition, the difference in calibration accuracy between low achievers and high achievers was less extreme. In the global and detailed criteria condition, low achievers were, in fact, more accurate than high achievers. Thus, it appears that detailed criteria helped low achievers to be more accurate.

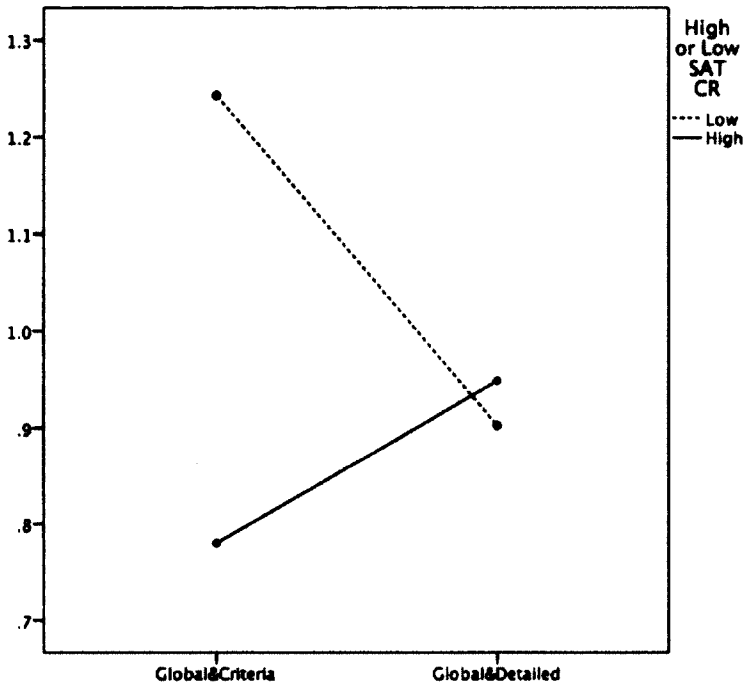


Figure 3. Interaction of calibration condition and achievement on prediction accuracy in organization

Calibration condition and achievement level also interacted to significantly effect prediction accuracy in development and analysis, $F(1, 205) = 3.917, p < .05, \eta^2 = .019$. Low achievers in the global and detailed criteria condition were the most accurate in their predictions ($M = .76$) in development and analysis, and low achievers in the global and criteria condition were the least accurate ($M = 1.14$).

Figure 4 illustrates the interaction of calibration condition and achievement level on prediction accuracy in development and analysis for the two criteria conditions. Again, in the global and criteria condition, there was a large difference in the calibration accuracy of high and low achievers. However, in the global and detailed criteria condition, low achievers were more accurate than high achievers. It appears that detailed criteria helped low achievers to be more accurate.

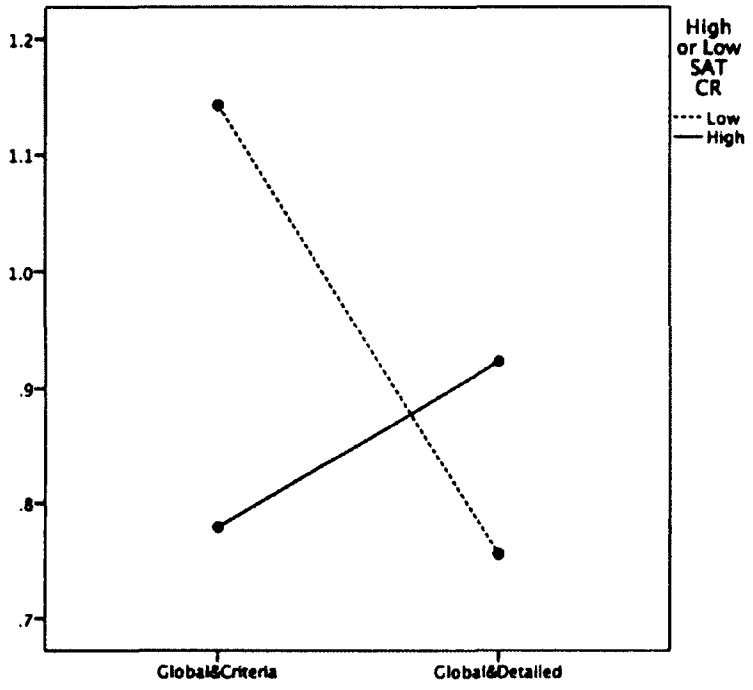


Figure 4. Interaction of calibration condition and achievement on prediction accuracy in development and analysis

Organization and development and analysis could be considered higher-level skills, and thus more cognitively complex, than the other two criteria. Prior achievement as measured by SAT critical reading scores seems to affect the more cognitive aspects of writing as illustrated by the interaction between calibration condition, achievement, and calibration accuracy in those areas. Detailed criteria seem to increase the prediction accuracy of low achievers in organization and development and analysis but not the calibration accuracy of high achievers.

Impact of Calibration Condition and EWC Scores on Calibration Accuracy

A MANOVA was conducted to determine if calibration condition and achievement level, using global EWC scores, interacted to influence calibration accuracy. Calibration condition and global EWC scores did not interact to influence the calibration

accuracy of global predictions $F(2, 355) = 1.442, p > .05, \eta^2 = .008$ or global postdictions, $F(2, 355) = 1.231, p > .05, \eta^2 = .007$.

Descriptive statistics for absolute accuracy are provided by achievement level using EWC global scores in Table 11. Both high achievers and low achievers were overconfident in their global predictions and postdictions. However, the global postdictions of both groups were more accurate than their global predictions. While high achievers were overconfident by less than two points, low achievers were overconfident by over five points.

Table 11

Global Prediction and Postdiction Accuracy by Achievement (EWC)

	High Achievers (Score > 12)		Low Achievers (Score < 12)	
	M	*SD*	*M*	*SD*
Global Prediction Accuracy	1.80	2.79	5.81	2.95
Global Postdiction Accuracy	1.79	2.96	5.68	3.02

Global EWC scores provided a greater contrast between low achievers and high achievers than SAT CR scores by calibration condition. While the results failed to reach significance, low achieving students in the global and criteria condition were less accurate in their predictions and postdictions than students in the other conditions by over six points. Low achievers in the global and criteria condition also made more accurate predictions than postdictions. Global predictive and postdictive accuracy are provided by condition in Table 12.

Table 12

Global Prediction and Postdiction Accuracy by Condition and Achievement (EWC)

		High Achievers (Score > 12)			Low Achievers (Score < 12)		
		n	*M*	*SD*	*n*	*M*	*SD*
Prediction							
	Global	41	1.46	2.88	80	5.66	3.22
	Global & Criteria	42	1.71	2.81	93	6.20	3.11
	Global & Detailed Criteria	27	2.41	2.58	78	5.51	2.34
Postdiction							
	Global	41	1.46	3.32	80	5.21	3.35
	Global & Criteria	42	1.86	2.88	93	6.24	3.05
	Global & Detailed Criteria	27	2.44	2.53	78	5.46	2.53

Impact of Calibration Condition and EWC Scores on Calibration Accuracy by Criteria

A factorial MANOVA was conducted to determine if calibration condition and achievement level, as measured by global EWC scores, interacted to influence calibration accuracy by criteria condition (i.e., global & criteria condition and global & detailed criteria condition). A median split was conducted using global EWC scores to divide students into high achievers ($n = 110$) and low achievers ($n = 251$). The mean EWC global score for high achievers was 14.38, while the mean EWC global score for low achievers was 9.41. Significant interactions were revealed for two criteria: (1) sentence structure and (2) grammar, diction, and mechanics.

Calibration condition and EWC scores interacted to influence prediction accuracy in sentence structure, $F(1, 237) = 9.225$, $p < .05$, $\eta^2 = .037$ and postdiction accuracy in

sentence structure, $F(1, 237) = 8.106, p < .05, \eta^2 = .033$. In sentence structure, high achievers in the global and criteria condition made the most accurate predictions (M = .53), and low achievers in the global and criteria condition made the least accurate predictions (M = 1.52). The sentence structure postdictions of all students were more accurate than their predictions. Figures 5 and 6, respectively, illustrate the interaction of criteria condition and EWC scores on students' predictive and postdictive accuracy in sentence structure.

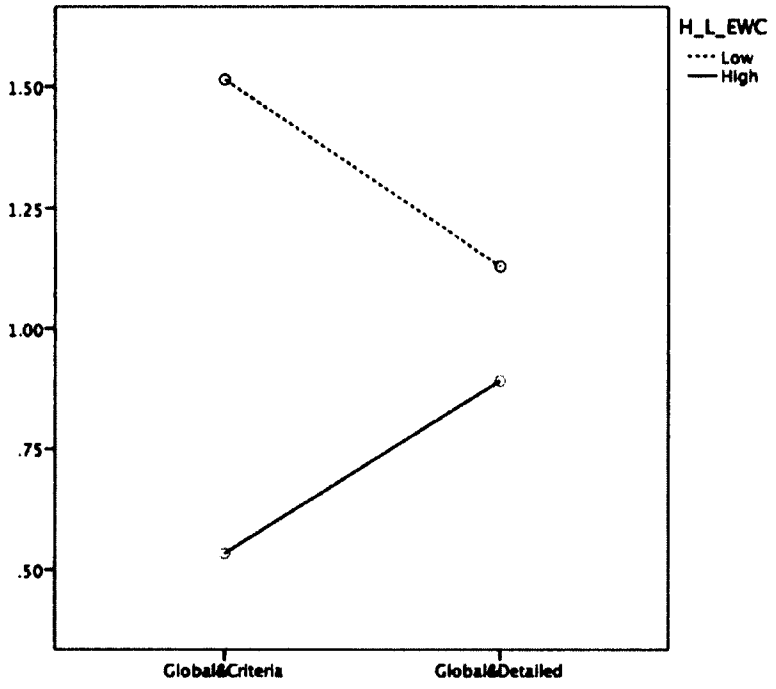


Figure 5. Interaction of calibration condition and achievement on prediction accuracy in sentence structure

In the global and criteria condition, there is a large difference in the prediction accuracy of low and high achievers in sentence structure. The difference in prediction accuracy between low and high achievers decreases in the global and detailed criteria condition. Detailed criteria appear to increase the calibration accuracy of low achievers; however, there appears to be little effect of detailed criteria on the accuracy of high

achievers. While high achievers in the global and detailed criteria condition were more accurate than low achiever, high achievers in the criteria condition were more accurate than high achievers in the global and detailed criteria condition.

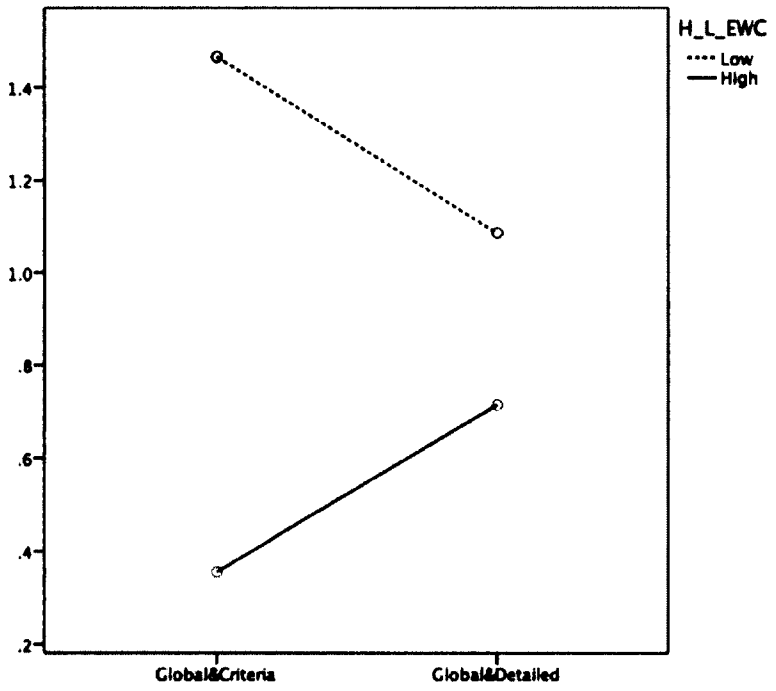


Figure 6. Interaction of calibration condition and achievement on postdiction accuracy in sentence structure

Figure 6 illustrates the interaction of calibration condition and achievement level on postdiction accuracy in sentence structure for the two criteria conditions. Again, in the global and criteria condition, there was a large difference in the calibration accuracy of high and low achievers. However, in the global and detailed criteria condition, the difference between high and low achievers seems to be mediated by the detailed criteria. While low achievers in the global and detailed criteria were overconfident and less accurate than high achievers, it appears that detailed criteria lessen the difference.

Calibration condition and EWC scores also interacted to significantly influence prediction accuracy in grammar, diction, and mechanics, $F(1, 237) = 4.353, p < .05, \eta^2 = .018$. In grammar, diction, and mechanics, high achievers in the global and criteria condition made the most accurate predictions ($M = .53$). Low achievers in the global and criteria condition made the least accurate predictions ($M = 1.51$). The predictions of low achievers in the global and detailed criteria condition were more accurate than their postdictions. Only high achievers in the global and detailed criteria condition made grammar, diction, and mechanics postdictions ($M = .75$) that were more accurate than their predictions ($M = .79$). Figure 7 illustrates the interaction of calibration condition and EWC scores on prediction accuracy in grammar, diction, and mechanics.

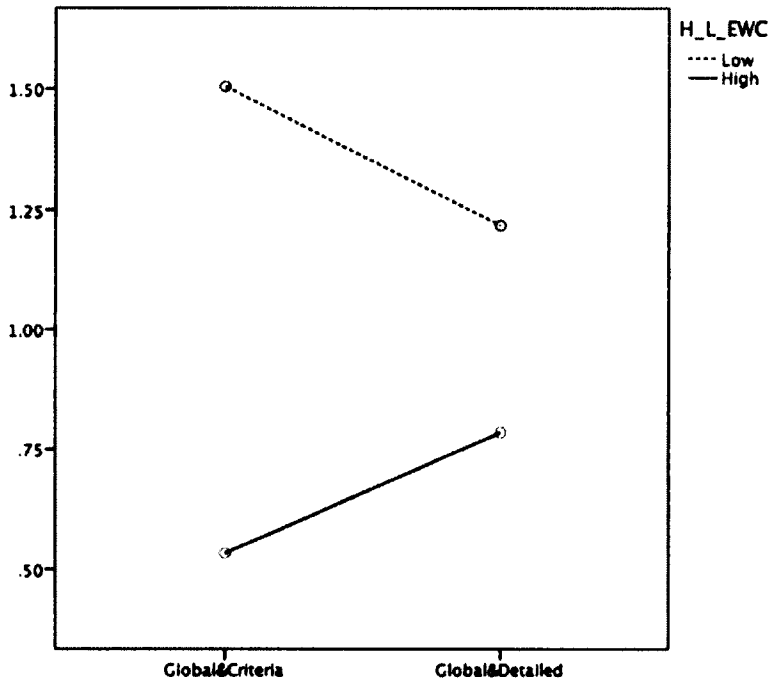


Figure 7. Interaction of calibration condition and achievement on prediction accuracy in grammar, diction, and mechanics

The same pattern emerges with low achievers in the global and detailed criteria condition, as the prediction accuracy of low achievers in the global and criteria condition seems to be aided by the use of detailed criteria. However, detailed criteria do not appear to increase the accuracy of high achievers, as high achievers in the global and criteria condition made the most accurate predictions.

Though statistically significant results were only revealed for two criteria, the descriptive statistics, which are provided in Appendix R, are revealing. When examining condition, criteria, and EWC achievement, high achievers in the global and criteria condition were the most accurate overall, and low achievers in the global and criteria condition were the least accurate. EWC achievement also appears to be linked to the surface aspects of writing, in contrast to SAT critical reading scores which were linked to the more complex aspects of writing. However, despite the achievement measured used, it seems that detailed criteria may help increase the calibration accuracy of low achievers.

Impact of Calibration Condition and Writing Self-Efficacy on Calibration Accuracy

To answer the final research question, a MANOVA was conducted to determine if calibration condition and self-efficacy level interacted to influence calibration accuracy. A median split was used to distinguish between students with low self-efficacy ($n = 280$) and students with high self-efficacy ($n = 268$). The actual mean EWC score of students with high self-efficacy ($M = 11.70$) differed less than a point from the mean EWC score of students with low self-efficacy ($M = 10.95$). Calibration condition and self-efficacy level were not found to have a significant effect on the calibration accuracy of global predictions, $F(1, 346 = .066, p > .05, \eta^2 = .000$, or global postdictions, $F(1, 346) = .762$, $p > .05, \eta^2 = .002$.

Overall, students with high self-efficacy were more overconfident in their global predictions than students with low self-efficacy. While the results are not statistically significant, those with low-self efficacy are often more accurate. High self-efficacy is often linked to overconfidence, and the majority of participants in this study would be categorized as having high self-efficacy. Table 13 provides the descriptive statistics of students' global calibration accuracy by self-efficacy level and condition, and Appendix S provides global prediction and postdiction accuracy by self-efficacy level.

Table 13

Global Prediction and Postdiction Accuracy by Calibration Condition and Self-Efficacy Level

		High Self-Efficacy (Score > 5)			**Low Self-Efficacy (Score < 5)**		
		n	***M***	***SD***	***n***	***M***	***SD***
Prediction							
	Global	88	5.14	3.03	84	3.26	3.45
	Global & Criteria	97	5.62	3.15	114	3.57	3.23
	Global & Detailed Criteria	83	5.31	2.31	82	3.34	2.74
Postdiction							
	Global	88	4.73	3.38	84	2.95	3.38
	Global & Criteria	97	5.57	3.18	114	3.62	3.32
	Global & Detailed Criteria	83	4.94	2.76	82	3.50	3.07

CHAPTER V

DISCUSSION

This study was an attempt to determine whether calibration condition and calibration by criteria influenced calibration accuracy and whether a relationship exists between calibration accuracy and achievement and calibration accuracy and self-efficacy. This chapter begins with a discussion of the influence of calibration condition on calibration accuracy. The effects of calibrating by criteria and a discussion of the relationship between calibration accuracy and achievement are provided. In addition, the influence of self-efficacy perceptions on calibration accuracy is considered. This chapter concludes by noting the limitations of the study and directions for future research.

Calibration Condition and Calibration Accuracy

The accuracy of both global and criteria-level calibration appraisals may help students differentiate between more and less reasonable calibration judgments. Previous research has explored the use of both calibration guidelines (Bol et al., 2012) and topical calibration strategies (Bol et al., 2010), which are similar to the calibration condition strategies used in this study. The calibration strategy used in this study required students to calibrate based on their membership in a global or global and criteria referenced condition. One of the hypotheses of the present study was that calibration accuracy would differ by calibration condition.

In contrast to the hypothesis, calibration accuracy varied little across conditions. While students' predictive and postdictive accuracy were best in the global condition (M = 4.17 and M = 3.84), and least accurate in the global and criteria condition (M = 4.56 and M = 4.42), respectively, the scores, at most, differed by a little over half a point,

which is not a statistically significant difference. Thus, the prediction that there would be a main effect of calibration condition on calibration accuracy was not supported.

One plausible explanation is that student's lack of domain knowledge may have led to the results. Overall, students in this study were low achieving. The mean EWC global score was 11.30, while a score of 12 is the minimum needed to pass. Early research (Fischhoff, Slovic, & Lichtenstein, 1977; Lichtenstein & Fischhoff, 1977) indicates that high achievers are generally more accurate than low achievers. In addition, recent research has demonstrated that higher achieving students are generally more accurate and better calibrated than lower achieving students (Bol & Hacker, 2011; Bol et al., 2010; Hacker et al., 2000). Kruger and Dunning (1999) suggest that lack of domain knowledge deprives learners of the ability to realize their weaknesses, as the skills needed to achieve competence are the same skills needed to evaluate one's competence. Domain knowledge is necessary for calibration accuracy, and research suggests that poor writers lack sufficient syntactical, rhetorical, and linguistic knowledge to recognize their errors (Ferrari et al., 1998; Negretti, 2012).

Calibration Condition and Writing Self-Efficacy

Students' confidence or self-efficacy in their writing skills has been linked to writing competence (Pajares et al., 2001; Prat-Sala & Redford, 2011). It was hypothesized that self-efficacy would differ by calibration condition. However, no effect of calibration condition on self-efficacy was found. Since accurate self-efficacy beliefs in writing require students to fully understand the components involved in not only the production, but also the evaluation, of the writing task, it was assumed that calibration condition would effect students' self-efficacy perceptions. Nevertheless, students' self-

efficacy differed little by condition (between .03 and .05 points), and overall students identified as highly self-efficacious despite their lack of achievement.

While self-efficacy is an important construct, calibration accuracy requires more than positive self-concept. Prat-Sala and Redford (2011) contend that writing self-efficacy may be specific to the writing task, as such it is likely that writing domain (e.g., creative writing versus essay writing) may influence self-efficacy more than calibration condition. However, Pajares and Miller (1997) found that students' self-efficacy judgments did not differ according to test format. While Pajares and Miller (1997) were examining students' math performance and self-efficacy using a multiple-choice test and open-ended performance tasks, the researchers found that students' self-efficacy judgments did not differ; although, student performance was worse on the open-ended tasks. Since this study did not seek to manipulate self-efficacy based on condition, it should be assumed that students with both high and low self-efficacy beliefs were equally distributed and that calibration condition alone is not an adequate intervention to increase self-efficacy beliefs. According to Schunk (1989), high self-efficacy alone will not increase achievement. Likewise, high-self efficacy alone may not increase calibration accuracy as the ability to accurately self-assess one's ability is needed to calibrate well (Alexander, 2013).

Calibration Accuracy and Calibration by Criteria

While there is a growing body of research on rubric-referenced self-assessment and its relationship to writing self-regulation and self-efficacy, few, if any, calibration studies require students to make confidence judgments on constructed response tasks using rubrics. Morozov (2011) and Covill (2012) examined rubric-reference self-

assessment, using rubrics with varying levels of detail and varying numbers of criteria. Researchers have also examined both the strengths and weaknesses of analytic and holistic rubrics (Carr, 2000; Huot, 1990; Jonsson & Svingby, 2007; Wiseman, 2012). For large-scale direct writing assessments, like the one in this study, holistic scoring is often used (Jonsson & Svingby, 2007). Holistic scoring is a global approach to evaluating a text that captures the qualities of writing using a single scale, while analytic scoring requires that raters evaluate judgments based on several domains representing the construct of writing (Wiseman, 2012). In terms of the reliability and validity of both types of rubrics, research has provided evidence that both holistic and analytic rubrics are reliable and valid; however, it is suggested that analytic rubrics provide more information about students' strengths and weaknesses (Jonsson & Svingby, 2007; Wiseman, 2012).

In this study, students were asked to make either global (holistic) or both global and criteria (analytic) judgments using the standards of the analytic scoring rubric used to assess student performance. Calibrating by criteria did not increase calibration accuracy. Sadler (2009) suggests that in order for students to improve their self-monitoring they need to be familiarized with evaluating quality holistically without being constrained by fixed criteria. This suggests that in order for students to calibrate accurately by criteria, they must understand how the criteria contribute to the work as a whole. Consequently, students might be better served by a holistic calibration procedure that provides them with the ability to monitor the quality of their writing in its entirety instead of a procedure that requires them to make discrete estimates based on specific criteria.

The research on rubric-referenced self-assessment has yet to evaluate students' actual judgments of performance. This study is significant because it attempts to close

that gap in the literature by requiring students to self-assess. Huot (1990) suggests that raters using holistic scales are most influenced by the content and organization of students' writing. Thus, students in the global condition may have been more accurate in their global predictions and postdictions because they based their global scores on their perceptions of the content and organization of their writing as a whole. While these results failed to reach significance, the effect of holistic versus analytic rubric use on the accuracy of students' self-assessment is worthy of additional study.

Students in the global and detailed criteria condition were more accurate in their predictions across all four criteria than students in the global and criteria condition; however, these results failed to reach significance, which suggests that rubric-referenced calibration does not increase students' calibration accuracy. Andrade (2001) surmised that simply distributing and explaining a rubric was associated with higher essay scores for eighth grade writers. However, this study's results align with Covill's (2012) findings that rubric-referenced self-assessment does not lead to better student writing and that students who were provided with more detailed criteria for evaluating their writing did not write better or worse than students provided with fewer criteria. The EWC Rubric used to evaluate student essays in this study is widely available to students and published in the examination's preparation guide. One possible reason why students in the criteria conditions did not calibrate more accurately than students in the global condition may be that all students may have reasonably good knowledge of the criteria, thus students may not have needed explicit criteria in order to accurately calibrate.

Conversely, Andrade and Du's (2005) qualitative study of undergraduate students' rubric use suggests that few students read rubrics in their entirety. Students

reported only reading the highest levels of the rubric, while others indicated that they might not read the rubric closely until they receive feedback from an instructor. Some students may perceive the rubric as a tool for satisfying a grader's demands rather than as a depiction of the criteria and standards of effective writing (Andrade & Du, 2005). Thus another plausible reason for students' failure to calibrate more accurately in the criteria conditions may be related to students' failure to absorb the rubric's criteria and to consider the performance level descriptions in each criteria. Additional explorations of students' rubric use and their misconceptions about the role of rubrics in their development as effective self-regulated learners are needed and may serve as promising lines of future research.

Interaction of Calibration Condition and Achievement on Calibration Accuracy

Various studies have established the relationship between achievement and calibration accuracy (Bol & Garner, 2011; Bol & Hacker, 2001; Hacker, Bol, & Keener, 2008; Kruger & Dunning, 1999; Valdez, 2013). In this study, students were split into two groups based on median SAT Critical Reading (CR) scores and characterized as either high or low achievers accordingly. As hypothesized, significant interactions were found between calibration condition, achievement, and calibration accuracy; however, the significant interactions were limited to two criteria.

Calibration condition and achievement level were found to significantly influence calibration accuracy for predictions in organization and development and analysis. While the effect sizes are small, they suggest that the effect for group differences in the MANOVA accounted for 22 percent and 19 percent of the variance in calibration accuracy, respectively. One possible explanation for the significant effect of condition

and achievement on predictive accuracy in organization and development and analysis is that these criteria require deeper levels of processing. Bol and Hacker (2001) suggest that deeper levels of processing have been associated with higher predictive accuracy, and they contend that essays, in general, require deeper levels of processing.

In contrast to the significant results revealed by criteria using SAT CR scores, using EWC global scores, significant interactions were revealed for predictive and postdictive accuracy in sentence structure. Significant interactions were also found between condition and EWC scores for predictive accuracy in grammar, diction, and mechanics. These results suggest that SAT CR scores might be more appropriate for complex criteria, while EWC scores are more appropriate for surface level processing.

Students' reported self-efficacy in rewriting wordy or confusing sentences correctly might provide some clues as to students' postdictive and predictive accuracy in sentence structure. Mean self-efficacy scores related to sentence structure were quite high, which suggests that students may have been able to more accurately gauge their performance. Students' reported self-efficacy in finding and correcting their grammatical errors might also indicate their ability to accurately predict their scores in that criterion. The low mean scores on this item suggest that students were aware that they could not accurately find and correct all of their grammatical errors. Battistelli, Cadamuro, Farneti and Versari (2009) found that while low achievers often overestimate their performance, their performance attributions indicate that they are aware that they have answered few questions correctly.

The mean scores of students classified as high and low achievers based on SAT CR scores, differed by 1.03 points, which suggests that these scores may be of little use

in categorizing students as high or low achievers in writing. Hilgers (1995) suggests that indirect measures of writing, like the SAT, do not serve as appropriate indicators of literacy or writing achievement. Thus, SAT CR scores may not have been the most appropriate measure for achievement as it is a multiple-choice test, and while it requires students to engage in tasks that are important for writing achievement, such as improving sentences and paragraphs and identifying sentence errors, it is not a direct assessment of writing.

The mean EWC scores of low and high achievers differed by nearly five points, which suggests that the EWC might be a better measure of writing achievement than SAT CR scores. As a direct writing assessment, EWC scores may be better able to differentiate between students' actual writing ability than SAT CR scores. However, neither measure of achievement resulted in statistically significant results in the same criteria. According to Hilgers (1995), multiple pieces of writing should be used to gauge students' writing ability and concomitantly their writing achievement. Thus, collecting student writing samples and averaging the scores across assignments might provide a better assessment of students' writing achievement.

Despite the proximity of mean SAT CR scores for low and high achievers, the descriptive statistics indicate that high achievers were better calibrated than low achievers regardless of the measure used for achievement. An analysis of global calibration accuracy and achievement using EWC global scores revealed that low achievers overestimated their global performance by over 5 points. While high achievers were also overconfident, their global calibration accuracy ranged from 1.46 to 2.41 point. The global calibration accuracy of low achievers ranged from 5.21 to 6.24 points.

High achievers were also more accurate by criteria than low achievers. These results mirror those obtained in other calibration studies that suggest that low achievers are far more overconfident and less calibrated than high achievers (Bol & Hacker, 2001; Bol et al., 2010; Hacker et al., 2000). Battistelli et al. (2009) contend that low achievers overestimate their performance to protect their self-esteem.

The significant interactions that were revealed among criteria condition, achievement, and calibration accuracy suggest that detailed criteria improves the calibration accuracy of low achievers. Detailed criteria also appeared to lessen the difference in calibration accuracy between low and high achievers. Thus these findings align with other studies that found treatment effects for the calibration accuracy of low achievers (Hacker et al., 2008). Bol and Hacker (2001) also found differential treatment effects for high and low achievers.

Overall, postdictions were more accurate than predictions. Previous research suggests that students' postdictions are often more accurate than their predictions, as postdictions are made after completion of the task (Bol & Hacker, 2001; Hacker, et al, 2000). In general, postdictions are more accurate than predictions because students have more information on which to base their judgments. Essentially, students are better able to self-evaluate what they know and do not know after testing.

Interaction of Calibration Condition and Self-Efficacy on Calibration Accuracy

Previous research has shown that self-efficacy is a predictor of achievement (Bandura, 1993; Schunk, 1989). More specifically, prior studies have affirmed a relationship between writing self-efficacy beliefs and writing performance (Pajares, 2003; McCarthy et al., 1985; Schunk, 2003; Shell, 1989). In this study, it was hypothesized that

calibration condition would effect self-efficacy and calibration accuracy. This hypothesis was not supported. In fact, students' self-efficacy differed very little by condition.

The findings of the current study are in contrast to the research that is available, which suggests a relationship between self-efficacy and calibration accuracy (Bembenutty, 2009). Bembenutty (2009) found that students with high self-efficacy are better calibrated. In the current study, students with high self-efficacy were less accurate and more overconfident than students with low self-efficacy. Students' achievement, in terms of EWC global scores, differed little for students with high and low-self efficacy. The mean EWC global scores of students with high self-efficacy was 11.70, while the mean scores of students with low self-efficacy was 10.95. Previous research suggests that self-efficacy is a predictor of performance (Chen, 2003); however, these findings do not support those conclusions.

Pajares (2003) suggests that writing self-efficacy contributes to the accurate prediction of writing outcomes independent of writing aptitude and previous writing achievement. This research followed the best practices in measuring self-efficacy (Bandura, 1997), specifically for measuring writing self-efficacy (Pajares, 2003). A valid and reliable instrument that assesses students' confidence to complete a writing task was used and the instrument was administered immediately before completion of the writing task. Self-efficacy is a contextual domain, and students' self-efficacy beliefs in writing may differ from their self-efficacy beliefs in another discipline. The research regarding the relationship between writing self-efficacy and calibration accuracy is extremely limited, thus additional research in this domain is warranted.

Limitations

One potential limitation of this study is the sample. This study was limited to students taking one writing exam at one public institution. Although students were randomly assigned group membership, additional implementation and testing at other universities is needed before the results can be generalized.

Although the current study revealed some interesting findings regarding the relationship between calibration accuracy, calibration condition, and achievement, another limitation was the categorization of student achievement based on students' SAT critical reading scores and/or EWC global scores. To start, SAT critical reading scores were not available for all participants. Additionally, SAT critical reading scores, while often used for placement in college composition courses, are derived from multiple-choice questions, while the EWC is a constructed response essay. Some calibration researchers have questioned using the instrument on which students calibrate as a measure of achievement, thus essays scores on a task that is similar to the EWC, might have served as a better source of student achievement. Ideally, student achievement in writing would be derived based on an analogous writing task.

Furthermore, this study tested writing in only one genre. For a different genre or a more or less complex writing task, students' self-efficacy perceptions might differ and their calibrations might be more or less accurate.

As with all self-report measures, the usefulness of the results depends upon the validity of participants' responses. The Writing-Self Regulatory Efficacy Scale and the calibration forms are self-report measures, thus their usefulness depends on the sincerity of students' responses. To encourage honesty, the researcher assured students that their

responses were confidential and that they would not affect their actual scores on the exam; however, there is always the possibility that students may not have honestly reported their predictions and postdictions or their responses to the Writing Self-Regulatory Efficacy Scale.

Finally, the calibration condition strategy may not have been robust enough to result in substantial gains in accuracy. The calibration condition strategy was a single, targeted treatment. A longer treatment might result in different results. Additionally, treatment fidelity is integral to both the interpretation and generalization of research findings. Students may not have closely attended to the criteria. Diffusion of treatment is also a concern, as students in all three of the calibration conditions may have had access to the EWC rubric before the exam. The EWC rubric is widely available to students for help in preparing for the exam. It is likely that all students in the study were familiar with the criteria. Students may have based their predictions and postdictions on their knowledge of the rubric criteria even though the rubric was only provided to students in the detailed criteria condition. Thus replication of the calibration condition strategy is needed in other studies.

Other limitations associated with the present study suggest more specific directions for future research. This study might be greatly enhanced by the opportunity for a longitudinal design, which tracks students' calibration accuracy over time.

Directions for Future Research

Additional research is needed in order to fully understand the conditions that enhance calibration accuracy. While a number of studies have been conducted on the use of rubric-referenced self-assessment to improve students' writing self-efficacy and self-

regulation (Andrade, 2001; Andrade & Boulay, 2003; Andrade, Wang, Du, & Akawi, 2009; Covill, 2012), little research is available that requires students to calibrate on essay tasks. Students must be able to discern weaknesses in their writing, and calibration seems an ideal method to assist students in improving their writing skills.

Introducing calibration accuracy into the composition classroom would provide students with calibration experience. This study required students to calibrate on an extemporaneous writing exam. Writing is a cyclical process, and students may need additional experience with calibration on various writing tasks in order to improve calibration on an exam like the EWC.

Positive relationships have been found between self-efficacy and writing achievement, and studies have shown that high self-efficacy is a predictor of high achievement (Zimmerman & Bandura, 1994). Andrade, Wang, Du, and Akawi (2009) also found positive relationships between long-term rubric use and writing self-efficacy. Future research should determine if self-efficacy interacts with rubric use, writing achievement, calibration condition, and calibration accuracy.

In this study, self-efficacy was not found to interact with calibration accuracy globally or by criteria. However, asking students to rate their self-efficacy by criteria might prove enlightening. Students may rate their efficacy by criteria differently than they rate their efficacy globally.

Implications

Successful calibration has been linked to appropriate self-regulatory behaviors and positive self-efficacy perceptions. Students' ability to metacognitively monitor their writing is important, as effective writing requires appropriate monitoring and control.

Students must learn to direct their focus appropriately (e.g., on specific criteria where there is a misalignment between their expectations and outcomes), in order to become effective writers and self-regulated learners.

This study's findings suggest that low achievers are often overconfident in their writing abilities. However, the significant interactions among calibration condition, achievement, and calibration accuracy suggest that detailed criteria help low achievers to make more accurate predictions. Targeted interventions that assist low achievers in recognizing their writing deficiencies are needed. Helping these students better align their calibration judgments may increase their self-efficacy perceptions and their writing achievement.

This study also has practical implications for use in the classroom. While peer review is common in the composition classroom, students should also be required to self-assess their own writing. Accurate self-assessment may increase the usefulness of peer review. If students are able to better determine the strengths and weaknesses in their own writing, they might provide more useful feedback to their peers.

Summary and Conclusions

This study focused on the use of calibration conditions to improve calibration accuracy. The study's participants were undergraduate students enrolled in a public institution. The interactions between academic achievement and self-efficacy were examined to determine their influence on calibration accuracy by calibration condition.

The first research question addressed the impact of calibration condition on calibration accuracy. Achievement and accuracy scores for students in the three calibration conditions were similar. The results revealed that calibration condition did not

affect calibration accuracy. Calibrating by criteria was not shown to improve calibration accuracy either. Many direct writing assessments, like the EWC, are scored holistically, thus holistic or global calibration may be more appropriate for these writing tasks.

Calibration condition was not found to effect self-efficacy. Additionally, self-efficacy was not found to influence calibration accuracy either globally or by criteria. Additional research is needed to clarify the relationship between writing self-efficacy and calibration accuracy.

In addressing the question of whether calibration condition and achievement level interacted to influence calibration accuracy, significant results were found for some measured variables. Calibration condition and SAT critical reading achievement were found to significantly effect predictive accuracy in organization and development and analysis only. Detailed criteria seemed to improve the predictive accuracy of low achievers.

Calibration condition and global writing scores were found to significantly effect prediction and postdiction accuracy in sentence structure as well as prediction accuracy in grammar, diction, and mechanics. The same pattern emerged wherein low achievers provided with detailed criteria made more accurate predictions. More studies are needed to fully examine the relationships among calibration accuracy, achievement, and specific writing criteria.

This study's findings suggest that calibration condition and achievement level are correlated with calibration accuracy in certain writing criteria. For higher level writing skills, like development and organization, SAT critical reading scores were found to

interact to influence calibration accuracy by condition, wherein low achievers were aided in making more accurate predictions in those areas using detailed criteria.

For surface level writing features, global writing scores were found to interact with calibration condition to influence calibration accuracy. Detailed criteria aided low achievers in making more accurate predictions and postdictions in sentence structure and more accurate predictions in grammar, diction, and mechanics. Additional research into how calibration activities impact writing self-concept and writing achievement is needed. The results of this study support the need for more research into the effects of rubric-referenced calibration.

REFERENCES

Alexander, P. A. (2013). Calibration: What it is and why it matters? *Learning and Instruction, 24*, 1-3. doi: 10.1016/j.learninstruc.2012.10.003

Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education, 4*(4). Retrieved from http://cie.asu.edu/volume4/number4/index.html

Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*, 21-34.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, & Evaluation, 10*(3), 1-11. Retrieved from http://pareonline.net/pdf/v10n3.pdf

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research, 102*, 287-302.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117-148.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior, Vol. 4*, (pp.71-81). New York: Academic Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman and Company.

Battistelli, P., Cadamuro, A., Farneti, A., & Versari, A. (2009). Do university students know how they perform? *International Electronic Journal of Elementary Education, 2*, 181-198.

Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom (2nd ed.).* San Francisco: Jossey-Bass.

Bembenutty, H. (2009). Three essential components of college teaching: Achievement calibration, self-efficacy and self-regulation. *College Student Journal, 43*, 562-570.

Bol, L., & Garner, J. K. (2011). Challenges in supporting self-regulation in distance education environments. *Journal of Computing in Higher Education, 23*, 104-123. doi: 10.1007/s12528-011-9046-7

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*, 133-151. doi: 10.2307/20152656

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology, 3*(1) 1-6. doi: 10.3389/fpsyg.2012.00229

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of over practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*, 269-290. doi: 10.2307/20157403

Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*, 280-287.

Bol, L., Riggs, R., Hacker, D. J., & Nunnery, J. (2010). The calibration accuracy of middle school students in math classes. *Journal of Research in Education, 21*, 81-96.

Broad, B. (2003). What we really value: Beyond rubrics in teaching and assessing writing. Logan, UT: Utah State University Press.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281. doi: 10.3102/00346543065003245

Carr, N. T. (2000). A comparison of the effects of analytic and holistic rating scale types in the context of composition tests. *Issues in Applied Linguistics, 11*, 207-241.

Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 73-92. doi: 10.1016/j.lindif.2003.08.003

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290. doi: 10.1037/1040-3590.6.4.284

Connors, R. J. (1996). The abolition debate in composition: A short history. In L. Z. Bloom, D. A. Daiker, & E. M. White (Eds.), *Composition in the twenty-first century: Crisis and change* (pp. 47-63). Carbondale, IL: Southern Illinois University Press.

Covill, A. E. (2012). College students' use of a writing rubric: Effect on quality of writing, self-efficacy, and writing practices. *Journal of Writing Assessment, 5*(1). Retrieved from http://www.journalofwritingassessment.org

Desoete, A., & Ozsoy, G. (2009). Introduction: Metacognition, more than the lognes monster? *International Electronic Journal of Elementary Education, 2*(1), 1-6.

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4-14. doi: 10.1016/j.learninstruc.2012.06.001

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58-61. doi: 10.1016/j.learninstruc.2012.05.002

Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing, 24*, 37-64. doi: 10.1177/0265532207071511

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science, 26*, 473-488.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Social Psychology: Human Perception and Performance, 3*, 552-564.

Garavalia, L. S., & Gredler, M. E. (2002). An exploratory study of academic goal setting, achievement, calibration, and self-regulated learning. *Journal of Instructional Psychology, 29*, 221-230.

Graham, S., & Harris, K. R. (1997). Self-regulation and writing: Where do we go from here? *Contemporary Educational Psychology, 22*, 102-114.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 1-24). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hacker, D. J., 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. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160-170. doi: 10.1037/0022-0663.92.1.160

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Memory and Metacognition,* (pp. 411-455). Mahwah, NJ: Lawrence Erlbaum Associates.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in Educational Theory and Practice*. Mahwah: NJ: Lawrence Erlbaum Associates.

Hacker, D. J., Keener, M. C., & Kircher, J. C. (2009). Writing is applied metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 154-172). New York: Routledge.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23-34.

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction, 24*, 62-66. doi: 10.1016/j.learninstruc.2012.05.009

Higham, P. A. (2013). Regulating accuracy on university tests with the plurality option. *Learning and Instruction, 24*, 26-36. doi: 10.1016/j.learninstruc.2012.08.001

Hilgers, T. (1995). Basic writing curricula and good assessment practices" Whene'er shall the twain meet. *Journal of Basic Writing, 14*(2), 69-74.

Hillocks, G. (1986). *Research on written composition: New directions for teaching.* Urbana, IL: National Institute of Education.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60*, 237-263.

Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing, 14*, 3-24. doi: 10.1016/j.asw.2008.12.002

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130-144. doi: 10.1016/j.edurev.2007.05.002

Kramarski, B., & Mevarech, Z. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310. doi: 10.2307/3699433

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134. doi: 10.1037/0022-3514.77.6.1121

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20* 159-283.

Lindemann, E. (2001). *A rhetoric for writing teachers*. Oxford: Oxford University Press.

MacArthur, C., & Philippakos, Z. A. (2013). Self-regulated strategy instruction in developmental writing: A design research project. *Community College Review, 41*, 176-195. doi: 10.1177/0091552113484580

Mayers, A. (2013). *Introduction to statistics and SPSS in psychology*. Pearson Education Limited

McCarthy, P., Meier, S., & Rinderer, R. (1985). Self-efficacy and writing: A different view of self-evaluation. *College Composition and Communication, 36*, 465-471.

McCormick, C. B., Dimmitt, C., & Sullivan, F. R. (2012). Metacognition, learning, and instruction. In I. B. Weiner, W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of Psychology (Vol. 7)*. (pp. 69-97). Hoboken, NJ: John Wiley & Sons, Inc.

Monroe, J. (2003). Writing and the disciplines. *Peer Review, 6*(1), 4-7.

Morozov, A. (2011). Student attitudes toward the assessment criteria in writing-intensive college courses. *Assessing Writing, 16*, 6-31. doi: 10.1016/j.asw.2010.09.001

National Commission on Writing. (2004). *Writing: A ticket to work...or a ticket out*. New York: College Board.

Negretti, R. (2012). Metacognition in student academic writing: A longitudinal study of metacognitive awareness and its relation to task perception, self-regulation, and evaluation of performance. *Written Communication, 29*, 142-179. doi: 10.1177/0741088312438529

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673-690. doi: 10.1007/s11135-006-9018-6

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology, 129*, 257-299.

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading and Writing Quarterly, 19*, 139-158. doi: 10.1080/10573560390143085

Pajares, F., Hartley, J. & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling & Development, 33*, 214-221.

Pajares, F. & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem-solving: Implications of using different forms of assessment. *Journal of Experimental Education, 65*, 213-228.

Pintrich, P.R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92-104. doi: 10.1006/ceps.1999.1017

Prat-Sala, M., & Redford, P. (2012). Writing essays: Does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate

students' performance in essay writing. *Educational Psychology, 32*, 9-20 doi: 10.1080/01443410.2011.621411.

Rawson, R. E., Quinlan, K. M., Cooper, B. J., Fewtrell, C., & Matlow, J. R. (2005). Writing skills development in the health professions. *Teaching and Learning in Medicine, 17*, 233-239.

Riggs, R. M. (2012). *Can practice calibrating by test topic improve public school students' calibration accuracy and performance on tests?* (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), Assessment, learning and judgement in higher education (45-63). Netherlands: Springer. doi: 10.1007/978-1-4020-8905-3_4

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 415-429). New York: Routledge.

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction, 24*, 48-57. doi: 10.1016/j.learninstruc.2012.08.007

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review, 1*, 173-208.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly, 19*, 159-172.

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*, 91-100. doi: 10.1037/0022-0663.81.1.91

Sjostrom, K. P., & Marks, A. (1994). Pretest and posttest confidence ratings in test performance by low-, medium-, and high-scoring students. *Teaching of Psychology, 21*(1), 12-16.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*, 437-475.

Tritelli, D. (2003). From the editor. *Peer Review, 6*(1), 3.

U.S. Congress, Office of Technology Assessment. (1992, February). *Testing American Schools: Asking the right questions*. (OTA-SET-519). Washington, DC: Government Printing Office. Retrieved from http://www.princeton.edu/~ota/disk1/1992/9236/9236.PDF

Valdez, A. (2013). Student metacognitive monitoring: Predicting test achievement from judgment accuracy. *International Journal of Higher Education, 2*, 141-146. doi: 10.5430/ijhe.v2n2p141

van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction, 24*, 15-25. doi: 10.1016/j.learninstruc.2012.08.005

Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research, 41*(6), 466-488. doi: 10.1016/j.ijer.2005.08.012

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum Associates.

Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *Iranian Journal of Language Testing, 2,* 59-92.

Yancey, K. B. (2009). *Writing in the 21st century: A report from the National Council of Teachers of English.* Urbana, IL: National Council of Teachers of English.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*, 329-339. doi:10.1037/0022-0663.81.3.329

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford Publications.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*, 166-183. doi: 10.2307/30069464

Zimmerman, B. J. (2011, December). Barry Zimmerman discusses self-regulated learning processes. *Science Watch*. Retrieved from http://archive.sciencewatch.com/dr/erf/2011/11decerf/11decerfZimm/

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal, 31*, 845-862. doi: 10.2307/1163397

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*, 663-676. doi: 10.2307/1163261

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The nature of problem solving* (pp. 233-258). New York: Cambridge University Press.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome to self-regulatory goals. *Journal of Educational Psychology, 91*, 241-250.

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22,* 73-101.

Appendix A

Examination of Writing Competency Rubric

EWC Scoring Rubric	1 (Incompetent)	2 (Developing Competency)	3 (Competent)	4 (Above Average Competency)	5 (Superior Competency)
Organization Appropriate use of essay structure (Introduction, Thesis statement, Body paragraphs, Conclusion, Transitional devices, etc.)	No clear or implied thesis statement; No clear introduction or conclusion; Paragraphing is missing, irregular or so frequent that it has no relationship to the essay; transitions are confusing or absent; Organizational problems make the essay near impossible to understand.	Thesis is vague or implied, not clear or specific, may simply breach prompt; Introduction and conclusion do not establish purpose or summarize arguments; Body paragraphs are poorly organized, ideas are strung together haphazardly; Ineffective transitions.	Generalized thesis addresses the prompt; Simple, but recognizable introduction and conclusion; Adequate incorporation of support for thesis in body paragraphs, though they may obtain some extraneous information; Transitions may be mechanical, but foster coherence.	Clearly-stated thesis addresses the prompt; Introduction begins to establish a foundation for the content and purpose; Conclusion summarizes arguments; Body paragraphs are sound and reinforce structure; Transitions connect ideas.	Clearly-stated, sophisticated thesis directly addresses the prompt; Introduction establishes the content and purpose; Conclusion effectively recounts and summarizes arguments; Body paragraphs include main points discussed separately and in detail; Effective use of thoughtful transitions that connect ideas.
Development & Analysis - Appropriate use of central ideas and concrete details that support the thesis and prompt	Supporting information is limited, unclear or not present at all; Thoughts are disconnected and have no discernible point; Essay length is not adequate for development.	Details may be too broad, narrow or inappropriate; Arguments are unclear or supporting evidence is insufficient, often unnecessarily repetitious.	Development is basic, ideas are reasonably clear, though they do not help flesh out some of the main arguments presented; Arguments on topic, but may not demonstrate in-depth understanding.	Details are present and support arguments; Arguments are clear and illustrate some awareness of the complexities of the issue being discussed.	Arguments effectively address all aspects of the prompt; Relevant, quality details enrich the central theme; Shows clear insight on the part of the writer.
Sentence Structure - Appropriate use of the construction of complete, complex sentences	Sequencing is random, most phrases are not sentences at all; Endless conjunctions or a complete lack thereof, which causes mass confusion.	Very little sentence variety, most are structured the same way; Some are awkward, others are fragments, run-ons, etc.	Sequencing shows logic, some sentence variety; Sentences are routine, but effective; A few fragments, run-ons, etc., but not to the point of distraction.	Sequencing is logical and effective, some sentence variety and use of complex sentence forms; Very few fragments, run-ons etc.	Complete sentences are well-built with complex and varied structure; Little to no sentence structure errors such as fragments, run-ons etc.
Grammar, Diction & Mechanics- Appropriate use of grammar such as tense, POV, language usage, punctuation (internal and external), spelling, capitalization, etc.	Frequent grammatical errors distort meaning and hinder communication; Little to no variation in word choice, language is used incorrectly and seriously impairs understanding; Gross errors in punctuation, spelling and capitalization that hinder meaning as well as understanding. (**15+ errors**).	Numerous grammatical errors that distort meaning in some instances; Language often used in odd ways; Jargon or clichés distract or mislead, redundancy is distracting; Many external and internal punctuation errors as well as numerous errors in spelling and capitalization (**10-14 errors**).	Problems with grammar are not serious enough to distort meaning, but may not be correctly applied in each instance; Attempts at colorful language apparent, but diction sometimes reaches beyond the scope of the argument; Punctuation sometimes missing or wrong; Some spelling and capitalization errors (**6-9 errors**).	A few grammatical errors, but grammar is correctly applied; Attempt at use of varied and advanced language that enhances arguments; Very few external punctuation and a few internal (i.e. comma, semi-colon, etc.) errors; Very few spelling and capitalization errors (**3-5 errors**).	Little to no grammatical errors (i.e. subject/verb agreement, tense, POV) used effectively and coherently throughout the essay; Language choices enhance meaning and clarify understanding in a precise, interesting way; Near perfect execution of internal and external punctuation, spelling and capitalization (**1-2 errors**).

Appendix B

Descriptive Statistics of Participants' Majors

Major	*N*	**%**
Accounting	**27**	**4.5**
Biology	27	4.5
Chemistry	5	.8
Computer Science	11	1.8
Education	23	3.9
Engineering	7	1.2
English	7	1.2
Finance	4	.7
Fine Arts	7	1.2
Exercise Science	24	4.0
History	9	1.5
Interdisciplinary Studies	32	5.4
Management	31	5.2
Marketing	17	2.9
Mass Communications	33	5.5
Mathematics	5	.8
Music	13	2.2
Nursing and Allied Health	61	10.2
Physics	1	.2
Political Science	11	1.8
Psychology	44	7.4
Secondary Education	1	.2
Social Work	41	6.9
Sociology	42	7.0
Technology	25	4.2
Tourism and Hospitality	13	2.2
Not listed	75	12.6

Appendix C

Test Calibration Form - Global

Student ID _____

BEFORE THE TEST

Before you complete the Examination of Writing Competency, please estimate the total score from **1 (incompetency) to 20 (superior competency)** that you think you will receive. **Please circle the score below that you think you will achieve.**

Incompetency Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

AFTER THE TEST

Now that you have completed the Examination of Writing Competency, please estimate the total score from **1 (incompetency) to 20 (superior competency)** that you think you will receive. **Please circle the score below that you think you will achieve.**

Incompetency Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Appendix D

Test Calibration Form - Global and Criteria

Student ID _____

BEFORE THE TEST

Before you complete the Examination of Writing Competency, please estimate the total score from **1 (incompetency) to 20 (superior competency)** that you think you will receive. **Please circle the score you think you will achieve.**

Incompetence Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Also estimate the score you think you will receive from 1 (incompetency) to 5 (superior competency) in each category: (a) organization, (b) development and analysis, (c) sentence structure, and (d) grammar, diction, and mechanics. **Please circle the score you think you will achieve in each category.**

	Incompetency	Developing Competency	Competency	Above Average Competency	Superior Competency
Organization	1	2	3	4	5
Development & Analysis	1	2	3	4	5
Sentence Structure	1	2	3	4	5
Grammar, Diction, & Mechanics	1	2	3	4	5

AFTER THE TEST

Now that you have completed the Examination of Writing Competency, please estimate the total score from **1 (incompetence) to 20 (superior competency)** that you think you will receive. **Please circle the score you think you will achieve.**

Incompetency Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Also estimate the score you think you will receive from 1 (incompetency) to 5 (superior competency) in each category: (a) organization, (b) development and analysis, (c) sentence structure, and (d) grammar, diction, and mechanics. **Please circle the score you think you will achieve in each category.**

	Incompetency	Developing Competency	Competency	Above Average Competency	Superior Competency
Organization	1	2	3	4	5
Development & Analysis	1	2	3	4	5
Sentence Structure	1	2	3	4	5
Grammar, Diction, & Mechanics	1	2	3	4	5

Appendix E

Test Calibration Form – Global Scores and Detailed Criteria

Student ID _____

BEFORE THE TEST

Before you complete the Examination of Writing Competency, please estimate the total score from **1 (incompetency) to 20 (superior competency)** that you think you will receive. **Please circle the score you think you will achieve.**

Incompetency Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Also using the detailed EWC Scoring Rubric, estimate the score you think you will receive from 1 (incompetency) to 5 (superior competency) in each category: (a) organization, (b) development and analysis, (c) sentence structure, and (d) grammar, diction, and mechanics. **Please circle the score you think you will achieve in each category.**

EWC Scoring Rubric	1 (Incompetent)	2 (Developing Competency)	3 (Competent)	4 (Above Average Competency)	5 (Superior Competency)
Organization Appropriate use of essay structure (Introduction, Thesis statement, Body paragraphs, Conclusion, Transitional devices, etc.)	No clear or implied thesis statement; No clear introduction or conclusion; Paragraphing is missing, irregular or so frequent that it has no relationship to the essay; transitions are confusing or absent; Organizational problems make the essay near impossible to understand.	Thesis is vague or implied, not clear or specific, may simply breach prompt; Introduction and conclusion do not establish purpose or summarize arguments; Body paragraphs are poorly organized, ideas are strung together haphazardly; Ineffective transitions.	Generalized thesis addresses the prompt; Simple, but recognizable introduction and conclusion; Adequate incorporation of support for thesis in body paragraphs, though they may obtain some extraneous information; Transitions may be mechanical, but foster coherence.	Clearly-stated thesis addresses the prompt; Introduction begins to establish a foundation for the content and purpose; Conclusion summarizes arguments; Body paragraphs are sound and reinforce structure; Transitions connect ideas.	Clearly-stated, sophisticated thesis directly addresses the prompt; Introduction establishes the content and purpose; Conclusion effectively recounts and summarizes arguments; Body paragraphs include main points discussed separately and in detail; Effective use of thoughtful transitions that connect ideas.
Development & Analysis - Appropriate use of central ideas and concrete details that support the thesis and prompt	Supporting information is limited, unclear or not present at all; Thoughts are disconnected and have no discernible point; Essay length is not adequate for development.	Details may be too broad, narrow or inappropriate; Arguments are unclear or supporting evidence is insufficient, often unnecessarily repetitious.	Development is basic, ideas are reasonably clear, though they do not help flesh out some of the main arguments presented; Arguments on topic, but may not demonstrate in-depth understanding.	Details are present and support arguments; Arguments are clear and illustrate some awareness of the complexities of the issue being discussed.	Arguments effectively address all aspects of the prompt; Relevant, quality details enrich the central theme; Shows clear insight on the part of the writer.
Sentence Structure - Appropriate use of the construction of complete, complex sentences	Sequencing is random, most phrases are not sentences at all; Endless conjunctions or a complete lack thereof, which causes mass confusion.	Very little sentence variety, most are structured the same way; Some are awkward, others are fragments, run-ons, etc.	Sequencing shows logic, some sentence variety; Sentences are routine, but effective; A few fragments, run-ons, etc., but not to the point of distraction.	Sequencing is logical and effective, some sentence variety and use of complex sentence forms; Very few fragments, run-ons etc.	Complete sentences are well-built with complex and varied structure; Little to no sentence structure errors such as fragments, run-ons etc.
Grammar, Diction & Mechanics- Appropriate use of grammar such as tense, POV, language usage, punctuation (internal and external), spelling, capitalization, etc.	Frequent grammatical errors distort meaning and hinder communication; Little to no variation in word choice, language is used incorrectly and seriously impairs understanding; Gross errors in punctuation, spelling and capitalization that hinder meaning as well as understanding. (15+ errors).	Numerous grammatical errors that distort meaning in some instances; Language often used in odd ways; Jargon or clichés distract or mislead, redundancy is distracting; Many external and internal punctuation errors as well as numerous errors in spelling and capitalization (10-15 errors).	Problems with grammar are not serious enough to distort meaning, but may not be correctly applied in each instance; Attempts at colorful language apparent, but diction sometimes reaches beyond the scope of the argument; Punctuation sometimes missing or wrong; Some spelling and capitalization errors (5-10 errors).	A few grammatical errors, but grammar is correctly applied; Attempt at use of varied and advanced language that enhances arguments; Very few external punctuation and a few internal (i.e. comma, semi-colon, etc.) errors; Very few spelling and capitalization errors (3-5 errors).	Little to no grammatical errors (i.e. subject/verb agreement, tense, POV) used effectively and coherently throughout the essay; Language choices enhance meaning and clarify understanding in a precise, interesting way; Near perfect execution of internal and external punctuation, spelling and capitalization (1-3 errors).

AFTER THE TEST

Now that you have completed the Examination of Writing Competency, please estimate the total score from **1 (incompetency) to 20 (superior competency)** that you think you will receive. **Please circle the score you think you will achieve.**

Incompetency Superior Competency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Also using the detailed EWC Scoring Rubric, estimate the score you think you will receive from 1 (incompetency) to 5 (superior competency) in each category: (a) organization, (b) development and analysis, (c) sentence structure, and (d) grammar, diction, and mechanics. **Please circle the score you think you will achieve in each category**

EWC Scoring Rubric	1 (Incompetent)	2 (Developing Competency)	3 (Competent)	4 (Above Average Competency)	5 (Superior Competency)
Organization Appropriate use of essay structure (Introduction, Thesis statement, Body paragraphs, Conclusion, Transitional devices, etc.)	No clear or implied thesis statement. No clear introduction or conclusion; Paragraphing is missing, irregular or so frequent that it has no relationship to the essay; transitions are confusing or absent; Organizational problems make the essay near impossible to understand.	Thesis is vague or implied, not clear or specific, may simply breach prompt; Introduction and conclusion do not establish purpose or summarize arguments; Body paragraphs are poorly organized, ideas are strung together haphazardly; Ineffective transitions.	Generalized thesis addresses the prompt; Simple, but recognizable introduction and conclusion; Adequate incorporation of support for thesis in body paragraphs, though they may obtain some extraneous information; Transitions may be mechanical, but foster coherence.	Clearly-stated thesis addresses the prompt; Introduction begins to establish a foundation for the content and purpose; Conclusion summarizes arguments; Body paragraphs are sound and reinforce structure; Transitions connect ideas.	Clearly-stated, sophisticated thesis directly addresses the prompt; Introduction establishes the content and purpose. Conclusion effectively recounts and summarizes arguments; Body paragraphs include main points discussed separately and in detail; Effective use of thoughtful transitions that connect ideas.
Development & Analysis - Appropriate use of central ideas and concrete details that support the thesis and prompt	Supporting information is limited, unclear or not present at all; Thoughts are disconnected and have no discernible point; Essay length is not adequate for development.	Details may be too broad, narrow or inappropriate; Arguments are unclear or supporting evidence is insufficient, often unnecessarily repetitious.	Development is basic, ideas are reasonably clear, though they do not help flesh out some of the main arguments presented; Arguments on topic, but may not demonstrate in-depth understanding.	Details are present and support arguments; Arguments are clear and illustrate some awareness of the complexities of the issue being discussed.	Arguments effectively address all aspects of the prompt; Relevant, quality details enrich the central theme; Shows clear insight on the part of the writer.
Sentence Structure - Appropriate use of the construction of complete, complex sentences	Sequencing is random, most phrases are not sentences at all; Endless conjunctions or a complete lack thereof, which causes mass confusion.	Very little sentence variety, most are structured the same way; Some are awkward, others are fragments, run-ons, etc.	Sequencing shows logic, some sentence variety; Sentences are routine, but effective; A few fragments, run-ons, etc., but not to the point of distraction.	Sequencing is logical and effective, some sentence variety and use of complex sentence forms; Very few fragments, run-ons etc.	Complete sentences are well-built with complex and varied structure; Little to no sentence structure errors such as fragments, run-ons etc.
Grammar, Diction & Mechanics- Appropriate use of grammar such as tense, POV, language usage, punctuation (internal and external), spelling, capitalization, etc.	Frequent grammatical errors distort meaning and hinder communication; Little to no variation in word choice, language is used incorrectly and seriously impairs understanding; Gross errors in punctuation, spelling and capitalization that hinder meaning as well as understanding. (15+ errors).	Numerous grammatical errors that distort meaning in some instances; Language often used in odd ways; Jargon or clichés distract or mislead, redundancy is distracting; Many external and internal punctuation errors as well as numerous errors in spelling and capitalization (10-15 errors).	Problems with grammar are not serious enough to distort meaning, but may not be correctly applied in each instance; Attempts at colorful language apparent, but diction sometimes reaches beyond the scope of the argument; Punctuation sometimes missing or wrong; Some spelling and capitalization errors (5-10 errors).	A few grammatical errors, but grammar is correctly applied; Attempt at use of varied and advanced language that enhances arguments; Very few external punctuation and a few internal (i.e comma, semi-colon, etc.) errors; Very few spelling and capitalization errors (3-5 errors).	Little to no grammatical errors (i.e. subject/verb agreement, tense, POV) used effectively and coherently throughout the essay; Language choices enhance meaning and clarify understanding in a precise, interesting way; Near perfect execution of internal and external punctuation, spelling and capitalization (1-3 errors).

Appendix F

Examination of Writing Competency - Retired

Directions:

- Write a well-developed academic essay that responds to ONE (1) of the three questions below. You are to write ONE essay that answers ONE of the questions.
- A competent academic essay will include: an introduction, thesis statement, body paragraphs, a conclusion, and transitional devices.
- Be sure to directly address the prompt/question throughout your essay and use details to support each of your arguments.
- A quote may be provided to help guide you as you think about how to answer the questions; it isn't necessary to include anything from the quote in your essay although you can use the quote if you wish.
- Develop your essay with specific details and examples drawn from history, literature, current events or personal experience. Your essay should be about 500-600 words long.

Question Code 01

"Tony Christopher, 26, says growing up in the Internet age has allowed him to quickly learn to use new technology. 'A lot of new technology makes my life easier,' says Christopher, office manager of a San Francisco law firm."

from *Poll: Many like tech gizmos but are frustrated, USA Today,* October 2005

Question: Which three (3) technological devices have made your life easier and why have they done so? Choose three (3) devices and discuss them thoroughly.

Question Code 02

"The Virginia Wesleyan students joined forces with the [Portsmouth Volunteers for the Homeless], which runs a winter shelter program that welcomes homeless people to church and synagogue auditoriums for a week at a time for a hot meal and a place to sleep."

from *For Wesleyan students, a life lesson, The Virginian-Pilot*, January 2007

Question: Virginia Wesleyan students are starting the New Year by sheltering some of the homeless in the area. What are three (3) activities that Norfolk State students could engage in to help those in the surrounding communities? Choose three (3) activities and discuss them thoroughly.

Question Code 03

"Meeting the needs of all students on our college campuses and helping them succeed is important to our higher education institutions," Chancellor Paul G. Risser said.

from *Oklahoma Higher Education Website*, December 2006

Question: Which three (3) college courses that you have had do you consider to be the most beneficial? Choose three (3) courses and discuss why they have been helpful to you.

Appendix G

The Writing Self-Regulatory Efficacy Scale

Student ID: _____

Directions: Think about your level of confidence in your ability to perform each of the following tasks. Indicate your level of confidence according to the 7-point confidence scale below.

	Not well						Very well
	1	2	3	4	5	6	7
1. When given a specific writing assignment, I can come up with a suitable topic in a short time.							
2. I can start writing with no difficulty.							
3. I can construct a good opening sentence quickly.							
4. I can come up with an unusual opening paragraph to capture the readers' interest.							
5. I can write a brief but informative overview that will prepare readers well for the main thesis of my paper.							
6. I can use my first attempts at writing to refine my ideas on a topic.							
7. I can adjust my style of writing to suit the needs of any audience.							
8. I can find a way to concentrate on my writing even when there are many distractions around me.							
9. When I have a pressing deadline on a paper, I can manage my time efficiently.							
10. I can meet the writing standards of an evaluator who is very demanding.							
11. I can come up with memorable examples quickly to illustrate an important point.							
12. I can rewrite my wordy or confusing sentences clearly.							
13. When I need to make a subtle or an abstract idea more imaginable, I can use words to create a vivid picture.							
14. I can locate and use appropriate reference sources when I need to document an important point.							
15. I can write very effective transitional sentences from one idea to another.							
16. I can refocus my concentration on writing when I find myself thinking about other things.							
17. When I write on a lengthy topic, I can create a variety of good outlines for the main sections of my paper.							
18. When I want to persuade a skeptical reader about a point, I can come up with a convincing quote from an authority.							
19. When I get stuck writing a paper, I can find ways to overcome the problem.							
20. I can find ways to motivate myself to write a paper even when the topic holds little interest for me.							
21. When I have written a long or complex paper, I can find and correct all my grammatical errors.							
22. I can revise a first draft of any paper so that it is shorter and better organized.							
23. When I edit a complex paper, I can find and correct all my grammatical errors.							
24. I can find other people who will give critical feedback on early drafts of my paper.							
25. When my paper is written on a complicated topic, I can come up with a short informative title.							

Appendix H

Notification Letter to Students

Dear Student,

My name is Katrice Hawthorne. I am a doctoral student at Old Dominion University in the Darden College of Education and Director of Assessment at ---. I am conducting research on students' confidence judgments and performance on writing exams. I would appreciate your help with this project, as it will assist us in better evaluating student learning and achievement. If you agree to participate, then you will join a study consisting of nearly 500 other students. You will be asked to self-assess your performance on the Examination of Writing Competency (EWC) using the EWC Scoring Rubric and to complete the Writing Self-Regulatory Efficacy Scale, a 25-item survey. I expect your time commitment to be 15 – 20 minutes.

The potential benefit of your participation is that it will help us to better understand students' use of writing rubrics and the relationship between student self-evaluations, self-regulation, and performance. Risks are minimal. The researchers will maintain strict confidentiality. We will remove any information that might identify you. The results of this study may be used in reports, presentations, and publications, but the researcher will not disclose your identity.

Your participation is voluntary. You can decline or withdraw at any time. Your participation will not affect your score on the EWC or your standing at the university. We hope you will allow your responses to be used for this project.

You are encouraged to ask questions about anything you do not understand before completing the rubric. Should you have additional questions later or if you want to know more about this research, please contact Katrice A. Hawthorne at 757-823-8375 or khawt002@odu.edu or Linda Bol at 757-683-4584 or lbol@odu.edu.

Thank you very much for your consideration.

Sincerely,

Katrice A. Hawthorne

Appendix I

Human Subject Review Approval

February 24, 2014

Approved Application Number 201401074

Dr. Linda Bol
Department of Educational Foundations and Leadership

Dear Dr. Bol:

Your Application for Exempt Research with Katrice A. Hawthorne entitled "Global and Criteria Based Judgments of an Undergraduate Exist Writing Examination," has been found to be EXEMPT under Categories 6.1 and 6.2 from IRB review by the Human Subjects Review Committee of the Darden College of Education with the condition that provide me with a copy of your updated CITI certificates. Faculty members must update their training each calendar year and the certificate you submitted is out of date.

The determination that this study is EXEMPT from IRB review is for an indefinite period of time provided no significant changes are made to your study. If any significant changes occur, notify me or the chair of this committee at that time and provide complete information regarding such changes.

In the future, if this research project is funded externally, you must submit an application to the University IRB for approval to continue the study.

Best wishes in completing your study.

Sincerely,

Theodore P. Remley, Jr., J.D., Ph.D.
Professor and Batten Endowed Chair in Counseling
Department of Counseling and Human Services
ED 110
Norfolk, VA 23529

Chair
Darden College of Education Human Subjects Review Committee
Old Dominion University

tremley@odu.edu

Appendix J

Descriptive Statistics of Examination of Writing Competency (EWC) Scores

	M	*SD*
Global EWC Score	11.30	2.25
Organization Subscore	2.88	.73
Development & Analysis Subscore	2.91	.67
Sentence Structure Subscore	2.81	.64
Grammar, Diction, and Mechanics Subscore	2.72	.67

Appendix K

Descriptive Statistics of Examination of Writing Competency (EWC) Scores by Gender

	Global EWC Score		Organization Subscore		Development & Analysis Subscore		Sentence Structure Subscore		Grammar, Diction, & Mechanics Subscore	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Male	11.26	2.35	2.84	.74	2.91	.67	2.81	.65	2.67	.69
Female	11.38	2.22	2.91	.73	2.91	.68	2.82	.64	2.75	.65

Appendix L

Descriptive Statistics of Examination of Writing Competency (EWC) Scores by Class Standing

	Global EWC Score		Organization Subscore		Development & Analysis Subscore		Sentence Structure Subscore		Grammar, Diction, & Mechanics Subscore	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Freshman	10.95	1.93	2.79	.71	2.79	.63	2.68	.49	2.68	.67
Sophomore	11.28	2.38	2.93	.68	2.87	.64	2.78	.68	2.70	.74
Junior	11.45	2.09	2.95	.69	2.97	.66	2.79	.58	2.76	.62
Senior	11.26	2.30	2.83	.76	2.89	.68	2.83	.66	2.72	.68

Appendix M

Descriptive Statistics of Examination of Writing Competency (EWC) Scores by School/College

	Global EWC Score		Organization Subscore		Development & Analysis Subscore		Sentence Structure Subscore		Grammar, Diction, & Mechanics Subscore	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
College of Liberal Arts	11.17	2.20	2.78	.77	2.89	.65	2.83	.62	2.70	.64
College of Science, Engineering, & Technology	11.64	2.29	2.98	.73	2.96	.69	2.88	.65	2.81	.66
School of Business	11.10	2.17	2.82	.70	2.86	.65	2.76	.62	2.66	.69
School of Education	11.61	2.11	3.10	.73	3.00	.63	2.78	.58	2.73	.57
School of Social Work	10.76	2.11	2.78	.57	2.76	.69	2.61	.59	2.61	.67

Appendix N

Prediction and Postdiction Accuracy by Criteria, Condition, and Achievement (SAT CR)

	High Achievers (Score > 420)			Low Achievers (Score < 420)		
	N	*M*	*SD*	*N*	*M*	*SD*
Organization – Prediction						
Global & Criteria	59	.78	.87	70	1.24	.91
Global & Detailed Criteria	39	.95	.72	41	.90	.77
Organization – Postdiction						
Global & Criteria	59	.95	.92	70	1.27	.87
Global & Detailed Criteria	39	.87	.83	41	.88	.90
Development & Analysis – Prediction						
Global & Criteria	59	.78	.96	70	1.14	.97
Global & Detailed Criteria	39	.92	1.01	41	.76	.77
Development & Analysis – Postdiction						
Global & Criteria	59	.92	1.04	70	1.11	.86
Global & Detailed Criteria	39	1.00	1.05	41	.90	.92
Sentence Structure – Prediction						
Global & Criteria	59	1.09	.99	70	1.23	.84
Global & Detailed Criteria	39	.85	.78	41	.95	.74
Sentence Structure – Postdiction						
Global & Criteria	59	.95	1.02	70	1.07	.92
Global & Detailed Criteria	39	.82	.94	41	1.02	.82
Grammar, Diction, & Mechanics – Prediction						
Global & Criteria	59	.98	.94	70	1.26	.88
Global & Detailed Criteria	39	.97	.90	41	1.07	.82
Grammar, Diction, & Mechanics – Postdiction						
Global & Criteria	59	1.00	.91	70	1.24	.91
Global & Detailed Criteria	39	.87	.98	41	1.17	.92

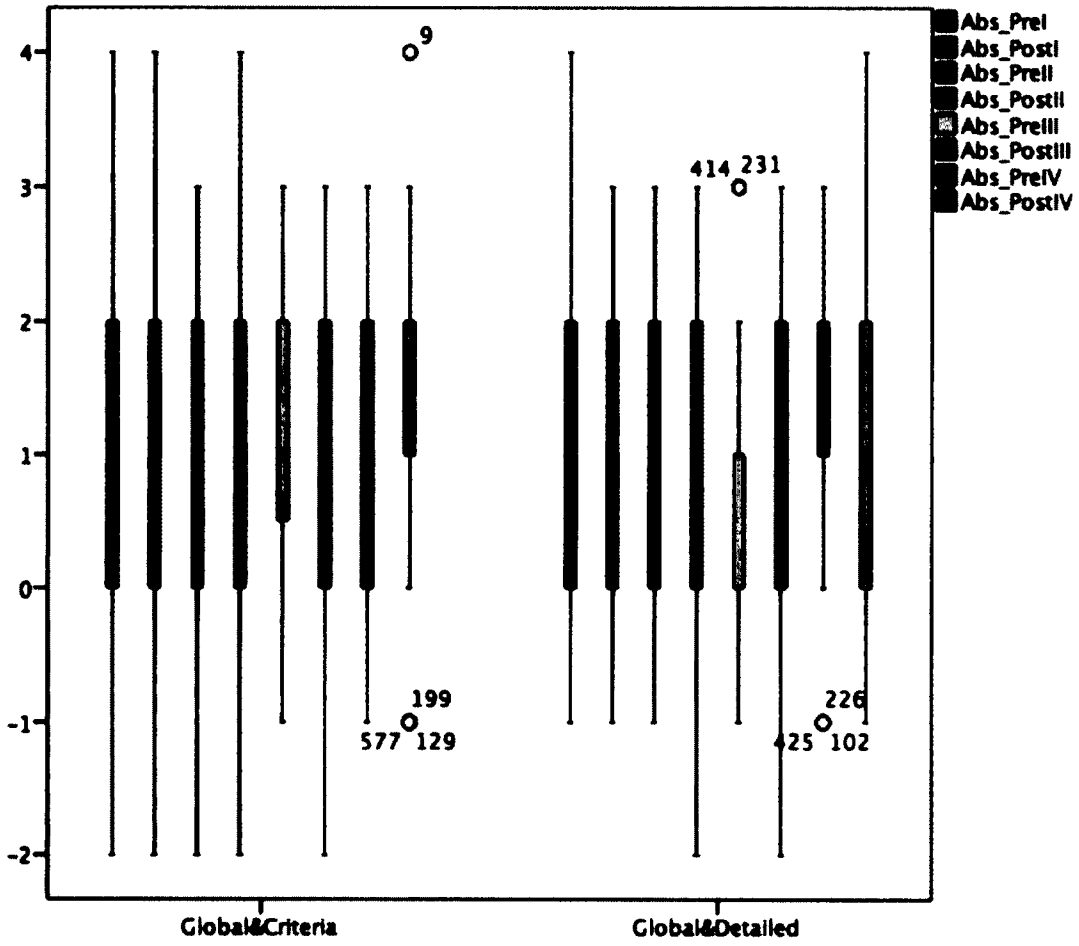
Appendix O

Global Prediction and Postdiction Accuracy by Criteria and Achievement (EWC)

	High Achievers (Score > 12)		Low Achievers (Score < 12)	
	M	*SD*	*M*	*SD*
Organization				
Prediction	.37	.87	1.42	.93
Postdiction	.39	.80	1.37	.91
Development & Analysis				
Prediction	.43	.94	1.36	.99
Postdiction	.33	.98	1.34	.96
Sentence Structure				
Prediction	.66	.86	1.36	.88
Postdiction	.48	.92	1.31	.90
Grammar, Diction, & Mechanics				
Prediction	.62	.84	1.36	.93
Postdiction	.68	.86	1.43	.92

Appendix P

Boxplot of Univariate Outliers by Criteria Condition



Appendix Q

Pearson Correlation Coefficients for Multicollinearity Test by Calibration Criteria

	Pre 1	Post 1	Pre 2	Post 2	Pre 3	Post 3	Pre 4	Post 4
Pre 1	1	.741	.617	.551	.503	.395	.378	.349
Post 1	.741	1	.524	.618	.456	.495	.290	.437
Pre 2	.617	.521	1	.754	.588	.557	.465	.442
Post 2	.551	.618	.754	1	.571	.665	.391	.512
Pre 3	.503	.456	.588	.571	1	.701	.612	.556
Post 3	.395	.495	.557	.665	.701	1	.489	.699
Pre 4	.378	.290	.465	.391	.612	.489	1	.712
Post 4	.349	.437	.442	.512	.556	.669	.712	1

Appendix R

Prediction and Postdiction Accuracy by Criteria, Condition, and Achievement (EWC)

	High Achievers (Score > 12)			Low Achievers (Score < 12)		
	N	*M*	*SD*	*N*	*M*	*SD*
Organization – Prediction						
Global & Criteria	45	.27	.99	99	1.48	.94
Global & Detailed Criteria	28	.54	.64	69	1.38	.94
Organization – Postdiction						
Global & Criteria	45	.31	.87	99	1.46	.93
Global & Detailed Criteria	28	.54	.64	69	1.26	.89
Development & Analysis – Prediction						
Global & Criteria	45	.33	.93	99	1.48	1.00
Global & Detailed Criteria	28	.61	.96	69	1.25	.93
Development & Analysis – Postdiction						
Global & Criteria	45	.24	.96	99	1.43	1.01
Global & Detailed Criteria	28	.50	1.00	69	1.22	.87
Sentence Structure – Prediction						
Global & Criteria	45	.53	.92	99	1.52	.86
Global & Detailed Criteria	28	.89	.74	69	1.13	.84
Sentence Structure – Postdiction						
Global & Criteria	45	.36	.98	99	1.47	.88
Global & Detailed Criteria	28	.71	.81	69	1.09	.90
Grammar, Diction, & Mechanics – Prediction						
Global & Criteria	45	.53	.81	99	1.51	.94
Global & Detailed Criteria	28	.79	.88	69	1.22	.91
Grammar, Diction, & Mechanics – Postdiction						
Global & Criteria	45	.64	.86	99	1.54	.91
Global & Detailed Criteria	28	.75	.89	69	1.28	.94

Appendix S

Descriptive Statistics for Global Prediction, Postdiction, and Actual Scores

	Global			Global & Criteria			Global & Detailed		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Prediction	175	15.61	2.74	227	15.78	2.72	177	15.49	2.49
Postdiction	175	15.27	3.12	217	15.73	2.87	172	15.39	2.71
Actual	178	11.44	2.43	232	11.26	2.39	185	11.22	1.88

Appendix T

Global Prediction and Postdiction Accuracy by Achievement (SAT CR)

	High Achievers (Score > 420)		Low Achievers (Score < 420)	
	M	*SD*	*M*	*SD*
Actual Score	12.03	2.02	11.00	1.90
Global Prediction Accuracy	3.66	3.27	4.51	2.92
Global Postdiction Accuracy	3.48	3.47	4.46	3.01

Appendix U

Global Prediction and Postdiction Accuracy by Self-Efficacy Level

	High Self-Efficacy (Score > 5)			Low Self-Efficacy (Score < 5)		
	N	*M*	*SD*	*N*	*M*	*SD*
Global Prediction	268	5.37	2.87	280	3.41	3.16
Global Postdiction	268	5.10	3.13	280	3.39	3.27

Appendix V

Calibration Accuracy by Self-Efficacy Level and Condition

	High Self-Efficacy (Score > 5)			Low Self-Efficacy (Score < 5)		
	N	*M*	*SD*	*N*	*M*	*SD*
Organization						
Global & Criteria Prediction	99	1.40	.92	121	.75	.96
Global & Detailed Criteria Prediction	78	1.32	.89	75	.81	.77
Global & Criteria Postdiction	99	1.37	.90	121	.88	1.00
Global & Detailed Criteria Postdiction	78	1.13	.90	75	.83	.86
Development & Analysis						
Global & Criteria Prediction	99	1.37	.96	121	.72	.95
Global & Detailed Criteria Prediction	78	1.18	.79	75	.64	.97
Global & Criteria Postdiction	99	1.33	.98	121	.73	1.00
Global & Detailed Criteria Postdiction	78	1.14	.89	75	.71	.99
Sentence Structure						
Global & Criteria Prediction	99	1.39	.95	121	.88	.81
Global & Detailed Criteria Prediction	78	1.26	.71	75	.64	.82
Global & Criteria Postdiction	99	1.21	.99	121	.86	.91
Global & Detailed Criteria Postdiction	78	1.17	.78	75	.64	.92
Grammar, Diction, & Mechanics						
Global & Criteria Prediction	99	1.33	.93	121	.88	.89
Global & Detailed Criteria Prediction	78	1.30	.77	75	.81	.85
Global & Criteria Postdiction	99	1.33	.90	121	.97	.89
Global & Detailed Criteria Postdiction	78	1.12	.84	75	.83	1.04

Appendix W

Means and Standard Deviations for the Writing Self-Regulatory Efficacy Scale

	n	*M*	*SD*
1. When given a specific writing assignment, I can come up with a suitable topic in a short time.	587	5.03	1.46
2. I can start writing with no difficulty.	587	4.71	1.54
3. I can construct a good opening sentence quickly.	580	4.76	1.45
4. I can come up with an unusual opening paragraph to capture the readers' interest.	582	4.75	1.38
5. I can write a brief but informative overview that will prepare readers well for the main thesis of my paper.	580	4.91	1.30
6. I can use my first attempts at writing to refine my ideas on a topic.	578	4.96	1.29
7. I can adjust my style of writing to suit the needs of any audience.	583	4.96	1.33
8. I can find a way to concentrate on my writing even when there are many distractions around me.	577	4.63	1.61
9. When I have a pressing deadline on a paper, I can manage my time efficiently.	581	5.21	1.47
10. I can meet the writing standards of an evaluator who is very demanding.	581	4.87	1.35
11. I can come up with memorable examples quickly to illustrate an important point.	580	5.16	1.35
12. I can rewrite my wordy or confusing sentences clearly.	588	5.16	1.29
13. When I need to make a subtle or an abstract idea more imaginable, I can use words to create a vivid picture.	582	5.07	1.33
14. I can locate and use appropriate reference sources when I need to document an important point.	580	5.18	1.30
15. I can write very effective transitional sentences from one idea to another.	586	5.07	1.33
16. I can refocus my concentration on writing when I find myself thinking about other things.	582	5.01	1.38
17. When I write on a lengthy topic, I can create a variety of good outlines for the main sections of my paper.	579	4.90	1.33
18. When I want to persuade a skeptical reader about a point, I can come up with a convincing quote from an authority.	580	4.81	1.44

	n	*M*	*SD*
19. When I get stuck writing a paper, I can find ways to overcome the problem.	583	5.09	1.39
20. I can find ways to motivate myself to write a paper even when the topic holds little interest for me.	580	4.97	1.46
21. When I have written a long or complex paper, I can find and correct all my grammatical errors.	583	4.72	1.44
22. I can revise a first draft of any paper so that it is shorter and better organized.	585	4.99	1.34
23. When I edit a complex paper, I can find and correct all my grammatical errors.	585	4.70	1.44
24. I can find other people who will give critical feedback on early drafts of my paper.	584	5.47	1.41
25. When my paper is written on a complicated topic, I can come up with a short informative title.	588	5.14	1.40

VITA

Katrice Alexandria Hawthorne

Old Dominion University
5115 Hampton Boulevard
Norfolk, VA 23529
Email: khawt002@odu.edu

EDUCATION

December 13, 2014	Expected Date of Completion Ph.D. in Education, Curriculum and Instruction Old Dominion University, Norfolk, VA Dissertation: Global and Criteria Based Judgments of an Undergraduate Exit Writing Examination
2003	Master of Arts in English Virginia Commonwealth University, Richmond, VA
2001	Bachelor of Arts in English Virginia Union University, Richmond, VA

PUBLICATONS

Hawthorne, K. A., Bol, L., Pribesh, S., & Suh, Y. (In Press). Effects of motivational prompts on motivation, effort, and performance on a low-stakes standardized test. *Research & Practice in Assessment.*

Butler, B. M., Burns, E., Hawthorne, K., Innes, A., Frierman, C., & Parrott, J. A. (2014). Becoming teacher educators: The impact of a pedagogy of teacher education seminar on emerging teacher educator identities. *Studying Teacher Education.* doi: 10.1080/17425964.2014.956716

Butler, B. M., Burns, E., Hawthorne, K., Innes, A., Frierman, C., & Parrott, J. A. (2014). Reflections on educator and teacher educator development: Implications of a pedagogy of teacher education seminar on practice. *Critical Issues in Teacher Education, 21*, 4-20.

Michaels, H. R., Hawthorne, K., Cuevas, N., & Matveev, A. G. (2011). Creating seamless K-16 pathways: Role of assessment. *Research & Practice in Assessment, 5*, 15-21.