A microanalytic study of self-regulated learning processes of expert, non-expert, and at-risk science students

by

Maria K. DiBenedetto

A dissertation submitted to the Graduate Faculty in Educational Psychology

in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The Graduate Center, City University of New York

2009

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the dissertation
requirement for the degree of Doctor of Philosophy.

Barry J. Zimmerman, Ph.D.

Date | Chair of Examining Committee

Mary Kopala, Ph.D.

Date | Executive Officer

Professor Barry J. Zimmerman

Professor David Rindskopf

Professor Peggy Chen

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

A microanalytic study of self-regulated learning processes

of expert, non-expert, and at-risk science students

by

Maria K. DiBenedetto

Mentor: Professor Barry Zimmerman

The present investigation sought to examine differences in the self-regulated learning processes and beliefs of students who vary in their level of expertise in science and to investigate if there are gender differences. Participants were 51 ethnically diverse 11th grade students from three parochial high schools consisting of 34 females and 17 males. Students were grouped as either *expert, non-expert*, or *at-risk* based on the school's classification.

Students were provided with a short passage on tornados to read and study. The two achievement measures obtained were the *Tornado Knowledge Test*: ten short-answer questions and the *Conceptual Model Test:* a question which required the students to draw and describe the three sequential images of tornado development from the textual description of the three phases. A *microanalytic methodology* was used which consists of asking a series of questions aimed at assessing students' psychological behaviors, feelings, and thoughts in each of Zimmerman's three phases of self-regulation: forethought, performance, and reflection. These questions were asked of the students while they were engaged in learning. Two additional measures were obtained: the *Rating*

Student Self-Regulated Learning Outcomes: A Teacher Scale (RSSRL) and the *Self-Efficacy for Self-Regulated Learning (SELF).*

Analysis of variance, chi square analysis, and post hoc test results showed significant expertise differences, large effect sizes, and positive linear trends on most measures. Regarding gender, there were significant differences on only two measures. Correlational analyses also revealed significant relations among the self-regulatory subprocesses across the three phases. The microanalytic measures were combined across the three phases and entered into a regression formula to predict the students' scores on the Tornado Knowledge Test. These self-regulatory processes explained 77% of the variance in the Tornado Knowledge Test, which was a significant and substantial effect.

Prior to this investigation, there have been no studies which have tested Zimmerman's three phase model on an academic task, such as science, within an expertise framework. Implications from the present study suggest that students varying in expertise level in science achievement also vary in self-regulatory behavior, and that gender is not a significant factor.

Acknowledgements

I am deeply grateful and lucky to have been mentored by Professor Barry Zimmerman, a truly brilliant scholar in the field of Educational Psychology. He has provided me with the inspiration, encouragement, and desire to contribute to the body of knowledge for which he has founded. He modeled self-regulation to me which helped me understand first-hand the significance of the theory. I am thankful for his never-ending patience, his kind demeanor, and his ability to help calm the often anxious and impatient feelings I have had over the years.

I would like to thank Professor David Rindskopf for serving on my dissertation committee and for providing me with the statistical foundations needed for research. I would also like to thank Professor Peggy Chen for serving on my dissertation committee and for being a ready resource and role model to me. Professor Helen Johnson has helped me throughout my academic career and I am extremely grateful to her for her guidance and kindness. I would also like to thank Professor Linnea Ehri for being a role model of classroom teaching and for her encouragement.

I would like to thank my friends and fellow doctoral students Irene R. Delgado and Heather B. Kadish for their emotional and intellectual support and encouragement over the past several years. I am extremely grateful to my network of colleagues at Baruch who have encouraged me and fostered my growth, especially Dean John Elliott who demonstrated his support by providing me with the time needed to conduct my research.

I would like to thank my Aunt Katie, a special woman who would often tell me not to give up, and my sisters who helped me time and again with child care, particularly while I was studying for comprehensive exams. I would like to thank my parents who believed education was the most important accomplishment one could achieve, especially my mother who taught me as a child that I could reach any goal I wanted - I am sure she is proudly smiling from heaven.

Most importantly of all, I would like to thank Michael who, without his help, I would never have been able to reach this dream; and my beautiful daughters Katie and Claire who know me as a student in addition to being a mom, and whose love has kept me on my path.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The purpose of this chapter is to lay the groundwork for a study of individual differences in American students' achievement in science, and whether these differences in achievement are related to their motivational beliefs and self-regulatory subprocesses during learning.

American Students' Achievement in Science

Current trends. There is considerable national concern about the quality of science learning in the United States. The U.S. participates in two international assessments that measure science skills: the TIMSS (Trends in International Mathematics and Science Study) and the PISA (Program for International Student Assessment). The TIMSS focuses on student performances on various academic measures including science in grades four and eight whereas the PISA measures the ability of 15 year-olds to apply science to real-life situations. Results of both of these assessments have disturbing implications regarding the American educational system and its competitiveness within an increasingly more educated world. For example, in the 2007 TIMSS study, American eighth graders performed more poorly in science than their peers in nine of the 47 participating countries, and those nine countries were countries located in Asia and Europe; and the U.S. performance was not measurably different from three other countries' averages (Gonzalez, Williams, Joceyln, Roey, Kastberg, & Brenwald, 2008).

In addition, only 10 percent of eighth grade U.S. students performed at or above the advanced international benchmark in science on the TIMSS.

The PISA assessment was conducted in 2006, and the results were also alarming. American 15 year-old students scored significantly below the science literacy average of the 29 participating countries, and U.S. students scored below 16 of the 29 participating countries (Baldi, Jin, Green, & Herget, 2007). This assessment required students to read science related material and to apply their understanding of what they have read to a real-life situation. Student performance on the combined science literacy on PISA is also placed on a proficiency scale ranging from 1 to 6, with 6 being the highest. U.S. students had the greatest percentage of students below and at level 1, than the other participating countries' percentages on the combined science literacy scale. Like the TIMSS results, those of the PISA indicate that American students are falling behind their international peers.

One additional source of concern involved significant gender effects in science where among eighth grade students, the international average was higher for boys than for girls (Baldi, Jin, Green, & Herget, 2007). These test results appear to have affected the subsequent lives of girls because fewer of them went on to careers in science (Blickenstaff, 2005). There are several other disturbing statistics regarding women in the sciences. Very few women choose careers in what has been coined the STEM professions: sciences, technology, engineering, and mathematics. According to the National Science Board (1998), only 22% of women in the workforce who entered the STEM professions are in the physical sciences, 33% are in technology or math, and only 9% are in engineering. Although it seems that females tend to prefer the biological sciences over the physical sciences, even in this category, only 40% of people within those careers are women. Additional information regarding educational attainment is equally alarming. According to Blickenstaff (2005), in 1999-2000, only 40% of the bachelor degrees awarded in the U.S. with majors in the physical sciences were awarded to women; and only 20% of the Bachelor degrees with majors in engineering were awarded to women. In addition, only 25% of the doctoral degrees in the physical sciences were awarded to women, and only 16% of the doctoral degrees in engineering were awarded to women. According to VanLeuvan (2004), when women are choosing careers in science, they tend to select careers that involve "helping others" such as in the biological sciences or in medicine; or they will choose careers in the social sciences – careers which address issues related to social concerns. And yet, even though women are more likely to study biological sciences than those in the STEM areas, in 2000, only 44% of the doctoral degrees in biology were awarded to women (Blickenstaff, 2005).

Understanding the reasons why girls, in particular, are not choosing the sciences. The lack of women in STEM careers has resulted in research directed at understanding why such gender differences in the sciences exist. One such study conducted by VanLeuvan (2004) examined girls' educational and career goals in the seventh grade and then again, in the 12th grade and she found that the girls' interest in science dropped as they aged. VanLeuvan and Blickenstaff (2005) offer several explanations for the decline in girls' interest as they mature: stereotypes and norms, role models and career information, participation in science related extracurricular activities, and self-efficacy beliefs. Self-efficacy is defined as personal judgments of one's capability to organize and execute courses of action to attain a designated goal, such as to obtain an A on a science test (Bandura, 1997).

Stereotypes and norms refer to adverse social influences by parents, peers, teachers, school counselors and others in society. Developmental research shows that parents often portray science as a male domain, and this would discourage females' efforts to study science (Meece & Courtney, 1992). For example, VanLeuvan (2004) found that messages are frequently given to girls that they may not be as capable as boys to study science. These messages may be direct, by telling them that certain careers are gender specific, or subtle, by the types of games that are given to them. According to VanLeuvan, scientific types of games are more typically given to boys than to girls. Parents tend to expect girls to engage in safe, stationary games whereas boys are expected to explore, take risks, and be independent. These gender biases may happen in the classroom as well as at home. Teachers also influence children by holding various stereotypes. In one study conducted by Spear (1987) that provided teachers with identical samples of students' work in science, teachers rated the papers they thought were prepared by boys with higher marks than those they thought were prepared by girls. School counselors may also play a role by directing girls to take classes other than science, and peers may rate girls on their looks and popularity and look down upon girls who take nontraditional courses (e.g. physics, engineering, etc.) (Blickenstaff, 2005).

Girls may be less likely to take subjects traditionally taken by boys because of a lack of role models for girls. VanLeuvan (2004) suggests that role models provide students with the opportunity to learn about various careers and may counter typical stereotypes. According to Bandura (1997), role models are essential to provide girls with examples of competency in fields traditionally held by men. He suggests that role models in the sciences can provide girls with the inspiration and encouragement needed for them to enter scientific career paths. Models provide children with many different benefits such as a belief in their competency or self-efficacy, skills, rules of behavior, and the proof that they too, can be scientists if they so desire. He suggests that it is particularly important for girls to have role models in scientific and technical fields traditionally dominated by men because these careers have "shunned" women. Perceived similarity to role models is particularly important for conveying information about self-efficacy (Schunk & Zimmerman, 1997). Role models convey a sense of ability, and for girls who may want to take science courses and study science, female role models may encourage young girls to aspire to scientific careers.

Experiences in scientific related activities have also been suggested as a means of providing girls with exposure to concepts, careers, role models, and positive attitudes towards science. Hanson (1996) for example, suggests that by middle school, compared to girls, boys will have been more exposed to science related experiences, such as attending museums, using microscopes, participating in science fairs, or even personally meeting scientists. These activities can affect a student's sense of competency (i.e., self-efficacy beliefs) in science related activities. This in turn may influence girls' decision to take advanced science courses in high school or college, which ultimately impacts their choice of science related careers.

The role of self-efficacy in career choice. Self-efficacy is a key construct in social cognitive theory. This theory examines what motivates individuals to work towards particular goals. Perceived self-efficacy affects one's choice of activities, persistence, and level of effort. Self-efficacy has also been found to affect one's motivation to learn (Zimmerman, 1997) and has been consistently linked to academic achievement (Pajaras, 1996, Schunk, 2001, Zimmerman, 1997). Self-efficacy has also been linked to course selection and career choice which has significant implications for men and women. For example, according to Bandura (1997), women generally judge themselves to be less scientific than men and tend to select nonscientific careers over men. He indicates that perceived self-efficacy in science is necessary in order to feel that one is capable of mastering scientific knowledge. If a student feels capable of learning scientific knowledge, he or she is likely to be successful in a scientific class, which ultimately predicts perseverance in a scientific field of study. This suggests a possible problem for girls in that if females tend to feel less self-efficacious in science related courses, they are likely to steer away from these courses, which will ultimately impact their career paths. In fact, Hackett (1997) has reported that there is a significant relationship between gender differences in self-efficacy and the percentages of males and females in various careers, including science. But what can be done to rectify these low student self-efficacy beliefs in science, especially among girls?

A Cyclical Theory of Academic Self-Regulation

Roots in social cognitive theory. There is a growing body of research indicating that higher levels of self-efficacy are associated with use of self-regulated learning processes, such as goal setting, self-monitoring, self-evaluation, and strategy use (Zimmerman, 1997, 2000). Recent research has shown that self-efficacy to regulate one's learning, mediates the effects of homework and studying on girls' grade point average in school (Zimmerman & Kitsantas, 2005). This finding suggests that students who self-regulate their science learning more effectively will achieve better than those who do not.

Social cognitive theory, when examined in the context of learning, has important implications for students of all ages and across all disciplines. This theory examines learners triadically in terms of interactions between their academic environment, behavior, and personal and cognitive factors (Bandura, 1986, Zimmerman, 2000).

Zimmerman (2000) further delineates these triadic components in terms of three self-oriented feedback loops. *Behavior self-regulation* takes place when the student strategically engages in a behavior, *environmental self-regulation* occurs when the student observes and alters his or her environment, and *covert self-regulation* happens when the learner reflects, monitors and adjusts his cognitive and affective states. In each of these forms of self-regulation, the learner self-regulates his academic functioning strategically to reach the goal he has set for himself and monitors the resulting feedback. Self-efficacy is particularly important here because it sustains the learners' motivation to do well (Zimmerman, 1997).

Developing into a self-regulated learner. Much research has been conducted to date to support the theory that academic self-regulation can lead to academic success. Academic self-regulatory subprocesses involve behaviors such as planning, managing time, goal setting, organizing, and rehearsing (Schunk & Zimmerman, 1997). Self-regulation also takes into account the motivational influences, such as self-efficacy beliefs and setting performance goals and outcomes. Studies have demonstrated links between elements of academic self-regulation and academic performance. In fact, in a large scale study conducted by PISA 2000, over 170,000 students around the globe were given a self-regulated learning questionnaire to complete (Artlet, Baumert, Julius-McElvany, & Peschar, 2003). The purpose of the questionnaire was to assess students' motivation, self-beliefs, and learning strategies alongside the assessment of learning in mathematics and literacy. Motivational questions were focused on the students' interest in learning, effort and persistence. Self-related belief questions were also obtained in this large scale study and included questions about students' self-efficacy. When examining effect size differences across gender, males tended to be more self-efficacious and interested in learning math whereas females tended to be more self-efficacious and interested in reading than their male counterparts. Students were also asked questions which focused on the use of specific learning strategies such as elaboration, memorization and control strategies. Correlational analysis showed that students who believed in their own efficacy as learners are likely to adapt "control strategies" or learning strategies to ensure that one's learning goals are obtained ($r = .54$). Control strategies involve checking what one has learned and what one still needs to learn, and adapting to the task at hand (p.13), subprocesses which Zimmerman has described as characteristic of self-regulated learning during the performance phase (Zimmerman, 2002).

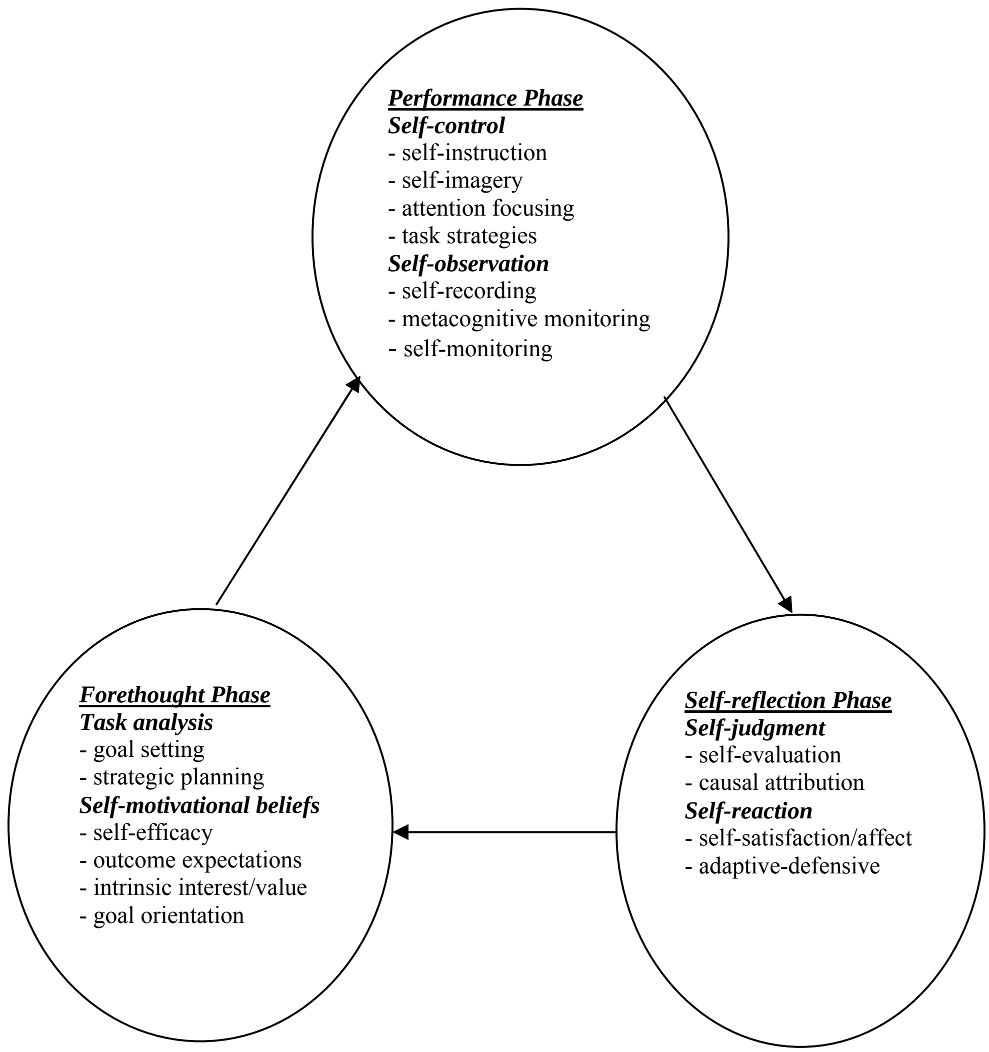
There is reason to expect that self-regulatory processes would be predictive of success in school. In Zimmerman's (2002) seminal article, he outlines how children develop self-regulatory skills as they progress through the different grades in education. Zimmerman describes how the learning of children, in the early grades, is regulated by their teachers and their parents. During the early school years, the young student is not expected to engage in self-directed behavior with regard to learning. For example, teachers will often provide students with outlines to study for tests, and parents will often be asked to sign their children's homework in order to ensure it was completed. When students enter the middle school, they will be exposed to many different teachers and are expected to monitor assignments and exhibit greater independence with homework assignments. In the high school years, students experience a further decline in the structure of their academic environment. Zimmerman indicates that during these years students are expected to be self-directed in their studying habits, both inside and outside of the classroom. What are these aspects of academic self-regulation that a student must engage in to perform well?

Phases and subprocesses of self-regulation. According to Zimmerman (2002) there are three phases of self-regulation which students engage in when performing an academic task (see Figure 1). The forethought phase is the first phase in which there are two major categories: task analysis and self-motivational beliefs. Task analysis involves goal setting and strategic planning. In this phase, the student sets a goal for himself or herself and a plan for how to accomplish this goal. In the second major category, the student's motivation is affected by his or her self-efficacy beliefs, outcome expectations, intrinsic interest and goal orientation. The student's self-motivation beliefs are important because they will determine the choice of the student's goal, how hard he or she is willing to work, and how long the student will persist to reach his or her goal. The forethought phase involves self-initiation or self-direction on the part of the learner, and Zimmerman emphasizes how this is part of what a student experiences as the student develops into a self-regulated learner. This phase is pivotal in the self-regulatory cycle because it is here that a student will plan his or her next course of action, be it a study schedule, what course he or she will take next term, or the path he or she will take in pursuit of a career.

Performance phase consists of two main categories: self-control and self-observation. Self-control subprocesses involve the use of various strategies such as self-instruction, imagery, attention focusing, and other task-related strategies. The self-observation subprocesses involve self-recording and metacognitive monitoring. In this phase the student is performing a task towards reaching his or her goal. Here the student is employing the various strategies he or she outlined in the forethought phase and is monitoring his or her progress towards the goal. This phase is extremely active for the highly self-efficacious student in that while the student is performing the behavior, the student is self-regulating as he or she monitors and adjusts his or her strategy while working to achieve goals.

In self-reflection, which is the third phase, the student uses self-judgment and self-reacts. Within self-judgment the student self-evaluates his or her performance and attributes causal significance to the outcomes. In addition, the student self-reacts. The self-reaction subprocesses involve determining whether or not the student is satisfied with the results and whether the student responds in an adaptive or defensive way. The results of the third phase are assumed to feed into the forethought phase for future behavior resulting in the cyclical feedback loop as seen in Figure 1. For example, a student who has low self-efficacy for a particular academic task in science is likely to not have been motivated to implement the strategies needed to learn better. The outcome is then likely to result in low achievement in the science task sustaining the student's original belief in his or her inability to perform well, thus the cycle continues.

Figure 1: Zimmerman's three phases and subprocesses of academic self-regulation



Obtaining a Complete Picture of the Self-Regulated Learner

To test this three phase model while students are engaged in learning, an online measure referred to as a *microanalytic methodology* is needed (Bandura, 1986). A microanalytic methodology is similar to Ericsson's think-aloud methodology in that the participant is talking while engaging in a task (2006). A main difference, however, is that the microanalytic approach applies specific questions targeted to address specific psychological processes at key times during the act of learning where the think-aloud asks the participants to verbalize their thoughts without any outside direction (Kitsantas & Zimmerman, 2002). In a microanalysis, the participant is performing a task and is asked simple open or closed-ended questions which are context-specific and provide online qualitative and quantitative data. The questions are short and asked at very specific times during the performance in order to minimize any threats to validity. Although a microanalytic methodology has been used in various athletic studies, to date, this methodology has not been used within an academic context and in the present study, has been used to test Zimmerman's three phase model of academic self-regulation on a science task.

Two additional measures of self-regulation provide information about the processes students may engage in when learning. The *Self-Efficacy for Learning Form-A* (SELF) was developed by Zimmerman and Kitsantas (2007) to measure students' self-efficacy judgments for learning. In this study, the SELF was shown to predict students' university course grades even when the effects of prior SAT scores were controlled statistically. In addition to predicting course grades, the SELF was also predictive of the quality and quantity of students' homework and their acceptance of their responsibility for course outcomes, including negative ones. Use of this measure can provide one with a scale to measure self-efficacy which is highly correlated with student achievement (Bandura, 1997).

The *Rating Student Self-Regulated Learning Outcomes: A Teacher Scale* (RSSRL) was developed by Zimmerman and Martinez-Pons (1989) to measure teacher's perceptions of the self-regulated learner in class. The RSSRL measure can be informative because it provides information about the behavior students engage in while learning. In the study conducted by Zimmerman and Kitsantas (2007), the SELF and the RSSRL were highly correlated, .49 at the .01 level. This suggests that the RSSRL scale, along with the SELF and a microanalytic analysis can provide a complete picture of the self-regulatory behavior of a learner and can help distinguish among students who are expert regulators from those who are poor academic self-regulators, which ultimately affects academic achievement. It takes years of practice as a student to be able to engage in expert academic self-regulatory behavior that produces good grades. But what does practice as a student entail?

Theory of Expert Performance

An important method for studying individual differences in skill involves determining participants' level of expertise. In the current research, students were classified according to their level of expertise in science. The expert performance methodology has been used widely in other areas of skill and has been shown to have important benefits when studying self-regulation of learning (Zimmerman, 2006).

Expert performance and its link to academic self-regulation. Extensive research has been conducted on expert performance in areas such as math, writing, history, and music (Ericsson, Charness, Feltovich, & Hoffman, 2006). Expert performance has been defined as performance that is consistently superior on a specific set of tasks and is achieved through years of deliberate practice (Ericsson & Charness, 1994). Expert performance is closely linked to self-regulation in that "developing experts focus *proactively* on learning processes during the forethought and performance control phases, rather than only *reactively* on personal outcomes during self-reflection" (Zimmerman, 2006, p.708).

Expert students are expected to excel during efforts to learn because of years of deliberate practice. Deliberate practice involves a high level of concentration and the structuring of specific tasks to facilitate setting appropriate personal goals, monitoring information feedback, and providing opportunities for repetition and error correction (Ericsson, Krampe, & Tesch-Romer, 1993). According to Ericsson (2006), it takes at least ten years of experience in a particular domain for someone to reach expert status. He refers to this as the ten-year rule. This idea of ten years has interesting implications for children in school. Drawing upon Ericsson's work and Zimmerman's theory, one hypothesis may be that it takes a student who is a high achiever, or an "expert", at least ten years of academic training and practice to reach that level, suggesting that by the 11^{th} grade, students who are academic experts, are successful self-regulators in specific academic domains when compared to students who are struggling or at-risk.

According to Ericsson (2006) experts are able to rapidly perceive the situation at hand and to generate potential moves or a plan of action. This implies that students, who are experts, can identify the academic task before them and make a decision as to the correct strategies to use to perform or master the task. Ericsson also indicates that expertise is domain specific and that "the core assumption of deliberate practice is that expert performance is acquired gradually and that effective improvement of performance requires the opportunity to find suitable training tasks that the performer can master sequentially…" (p. 692). What distinguishes the student who excels from the one who is struggling and what are the differences in their self-regulatory behaviors that may contribute to distinguishing between the two?

Expert versus non-expert learners. The comparison of expert and non-expert learners has advantages when studying individual differences in academic learning. As mentioned earlier, an expert in a particular domain is someone who is not only knowledgeable about the task at hand but also has the performance skill to accomplish that task at a high level consistently. By contrast, a novice lacks detailed task knowledge and performs at a consistently low level. Novice students would be at risk for academic failure. Cleary and Zimmerman (2001) recommended including a second "non-expert" comparison group to control for a variety of background variables, such as a general valuing of the task and engagement in practice. The non-experts are distinguished from experts by their lower level of performance. The non-experts were hypothesized to lack important self-regulatory skills that experts use to bring consistency to their high level of performance. More specifically, expert learners should display higher quality self-regulatory processes than non-experts during each of the three phases of learning that were outlined above.

Study time. Pant, Ericsson, Hill, and Asberg (2005) have examined the use of study time and its role in deliberate practice on achievement. Their findings suggest that what is important for achievement is not the amount of time spent, but the *quality* of the activities students engaged in while studying. In addition, Zimmerman (2000) has argued that self-regulated learning requires students in the forethought phase to come up with strategic plans and during the performance phase, in addition to using these strategies, to be metacognitively aware of whether their plans are working. Greene, Moos, Azevedo, and Winters (2008) examined the self-regulatory learning processes with hypermedia and compared gifted students from grade-level students. Their findings suggest that students who were gifted used more sophisticated strategies in learning as compared to the grade-level students.

One of the purposes of the current investigation was to test whether or not differences in students who are expert learners as a group spend more time studying than students who are at-risk and if this is related to the number of strategies planned and used. In the present study male and female students selected to participate were at three levels of expertise in science (experts, non-experts, and novices or at-risk) in order to assess differences in self-regulation at each level of expertise. Prior research indicating gender differences in self-regulation (Zimmerman & Martinez-Pons, 1990) did not include measures of expertise or a time study measure.

Research Questions

The present study attempted to address several research questions using the microanalytic methodology to test Zimmerman's three phase model of academic self-regulation on an academic task. The first research question asks whether or not there are expertise or gender differences on achievement measures: the Tornado Knowledge Test and the Conceptual Model Test. The second research question asks whether or not there are expertise differences in the established person measures of self-regulation, namely the teacher ratings (RSSRL) and the student ratings (SELF). The third research question asks whether or not there will be expertise differences in the three phases of academic self-regulation. The fourth research question asks whether or not there are gender differences in self-regulated learning. Next to test the relationship among the subprocesses within the three phases, the fifth research question asks whether or not there are correlational relationships among these variables. A sixth research question is regarding the overall predictive power of the microanalytic measures on achievement; and lastly, the final research question asks whether or not there are differences in the amount of study time used by students in the three expertise groups.

Chapter 2

Literature Review

Issues related to science education have begun to draw attention in the United States despite an intense focus on reading and math. The low science scores when compared to those of other developed countries and the lack of women in the STEM occupations are of concern to many educators. The United States National Assessment of Educational Progress (NAEP) has reported that on average, students' science scores in the 12th grade assessments in 2005, had declined since 1996 (Grigg, Lauko, & Brockway, 2006). It can be noted that while fourth grade students performed higher in recent years, in eighth grade, there was no significant improvement when compared to previous years. In addition, many states have begun to examine how their own assessments measure up to the standards in science set by NAEP (Timms, Schneider, Lee, & Rolfhus, 2007). However, as indicated in chapter one, students may perform poorly in science for reasons related to their self-beliefs and the processes they engage in, or do not engage in, when learning.

Measuring Self-Efficacy

The Self-Efficacy for Learning Form. Students' beliefs about their capabilities to perform well have been significantly associated with their performance on various academic tasks (Bandura, 1997, Bandura 1997, Pajaras, 1996, Schunk, 2001, Zimmerman, 1997). According to Bandura (1997) social cognitive theory suggests that students' self-efficacy beliefs will influence their motivation and learning activities.

Perceived academic self-efficacy promotes intellectual achievement directly through the use of self-regulatory processes and covertly by raising their academic aspirations (Zimmerman & Martinez-Pons, 1992).

Zimmerman and Kitsantas (2005) developed self-efficacy scale for high school students' self-efficacy beliefs and tested this measure on a selective group of high school girls. The scale was designed to measure perceived self-efficacy beliefs as they relate to various academic tasks such as reading, note taking, test taking, writing and studying. The scale consisted of 57 items and their findings suggest that students who were highly self-efficacious about regulating their learning, were more likely to accept responsibility for the learning outcomes.

In a second study conducted by the same researchers, the scale was reduced to 19 items which focused on studying, test preparation, and note-taking and tested the predictive validity using a sample of college students (Zimmerman & Kitsantas, 2007). The 57 item scale was administered to college students along with scales on personal data, perceived responsibility for learning, homework, and teacher rating of self-regulated learning. Two forms of the SELF measures were correlated with each other along with students' grades in an educational psychology course and their SAT scores. The results from this study demonstrated superior predictive validity of students' scores on the revised SELF-A compared to the longer form on all of the measures with the exception of SAT scores. There were no significant differences in predicting SAT scores between the two forms of the SELF. In addition, the researchers also examined whether or not there would be differences across gender and found there were not significant differences. The questions in the SELF-A were designed to measure students' beliefs about their capability to overcome obstacles to learning. As indicated earlier, these beliefs are important in providing students with self-directed agency needed to engage in self-regulatory behavior. Although this measure has demonstrated predictive validity, it does not provide an online measure of students' self-regulatory behavior while they are performing an academic task.

Measuring Self-Regulation

The structured interview. Zimmerman and Martinez-Pons (1990) conducted a study that examined the differences in self-regulated learning and self-efficacy based on students' grade, sex and giftedness. In their sample, 45 boys and 45 girls from regular and gifted elementary, middle, and high schools were selected to participate in the study. The students completed self-efficacy scales on their verbal comprehension and their mathematical problem solving. In addition, the students were presented with the following eight different learning contexts: different classroom situations, when preparing for a test, when taking a test, when studying at home, when not motivated to do homework, when checking science or English homework, and when completing math and writing assignments. For each different learning context, students were asked to describe what methods they would use. This format is considered a structured interview.

There were several interesting findings regarding the students' self-efficacy beliefs and reports of self-regulated strategies. Self-efficacy increased with age. In addition, gender differences were found in the use of self-regulatory learning strategies such as record keeping, monitoring, and goal setting. Girls reported significantly *more* use of these strategies than did boys. Both findings are extremely interesting when linked to current trends in science studying and achievement. There is evidence that fewer girls are selecting to study science and that girls tend to be less self-efficacious than boys in science (Bandura, 1997). However, girls used more self-regulatory strategies in this study. Again, perhaps this suggests that self-efficacy and self-regulatory processes are more predictive of academic performance than gender, particularly when a third finding of this investigation revealed that higher perceptions of math and verbal efficacy resulted in higher efforts to regulate their learning strategically.

Two additional findings from this study involved the examination of students in the gifted versus the regular groups. Gifted students tended to be more self-efficacious and to use more self-regulatory strategies than regular students. These findings imply that students who are consistently successful (gifted or experts) at learning, such as these students who were attending a highly selective school, are also more confident in their ability to learn math and verbal material and are also more likely to engage in self-regulatory behavior, such as greater organizing and transforming, seeking peer assistance, and reviewing notes. Although findings from this study are provocative, the methodology does not measure self-regulated learning as it takes place.

The microanalytic method. Although the structured interview was originally developed by Zimmerman and Martinez-Pons in 1986 when they attempted to assess self-regulatory processes in naturalistic-like settings, it does not provide a dynamic, online measure of learning. According to Zimmerman (2000) self-regulated learning involves three cyclical phases. It is hypothesized that the forethought phase influences the performance phase which in turn, influences the self-reflection phase. According to Zimmerman (2006), the student who is *proactive* in learning engages in high quality forethought and performance phase subprocesses, which then increases his or her ability to perform well. In contrast, a student who is *reactive* examines his or her performance outcomes afterward, which can lead to poor self-regulatory control during subsequent efforts to self-regulate. Social cognitive researchers sought a way to measure these specific behavioral and psychological processes while they were happening and changing. As described in chapter one, the microanalytic methodology, provides an online measure by assessing self-regulation while it is (or is not) taking place.

Studies using the microanalytic methodology. Cleary and Zimmerman (2001) used the microanalytic methodology to test the three phases of self-regulation on an athletic task within the expert-novice framework. In their study, basketball players were grouped as Expert, Non-Expert, or Novices and were examined during their free-throw shooting practices. Male high school students participated in the study and were placed in one of three groups based on their free-throw accuracy. Cleary and Zimmerman found significant differences across the three groups in their use of the self-regulatory subprocesses. Students in the Expert groups for example, set more process oriented goals and were more strategic in the techniques they used than free throwers in the other two groups. Experts were also more self-efficacious and had higher intrinsic interest than Novices. During the self-reflection phase, the Experts attributed their mistakes to ineffective strategy use and made strategic adjustments while the Non-Expert groups blamed their mistakes on more general-like issues such as lack of concentration.

Kitsantas and Zimmerman (2002) also used microanalysis of an athletic task to test the cyclical nature of Zimmerman's three phases of self-regulation. This study closely parallels the proposed study in that it attempted to use a microanalysis to examine

Expert, Non-Experts, and Novices except for the nature of the task. In their investigation, Kitsantas and Zimmerman asked microanalytic questions of female university students who had been grouped into one of the three categories based on their level of skill. Students in the Expert group were on the university's volleyball team; the Non-Expert players were members of the university's volleyball club for at least three years; and the Novices were women who did not play in an organized sport, but rather played volleyball informally.

The primary goal for the Kitsantas and Zimmerman (2002) was to examine the differences in the three groups' practice methods of the overhand serve. To assess prior knowledge, the volleyball players were asked to identify three most important parts of a well-executed overhand serve. This was followed by a video presentation of the overhand serve motion, then questions addressing the subprocesses of the forethought phase such as: self-efficacy beliefs, perceived instrumentality (outcome expectations), intrinsic interest, goal setting and planning. After the participants responded to these questions, they were asked to practice the overhead serve for ten minutes which was then followed by a post-test for serving skill, strategy use, self-monitoring, which are subprocesses of the performance phase. This was followed by scales measuring the subprocesses of the self-reflection phase: self-evaluation, causal attribution, self-satisfaction, and adaptation. The final step for the players was for them to serve in one of the "highest priority areas" which was followed by questions on attribution, adaptation, and self-efficacy perceptions. Although causal attribution and adaptation questions are subprocesses within the third phase of self-reflection, the self-efficacy question connects the cycle by bringing the player back to the forethought phase. This initiation of another feedback loop is important with regard to Zimmerman's (1998) theory because it provides another opportunity to improve their performance.

The results from this study showed significant differences across groups in the use of each of the self-regulatory phases and their subprocesses. Separate statistical analyses were conducted on each of the measures assessed in the microanalytic approach. Findings revealed significant differences among the expertise groups in the predicted direction for each measure. Experts, for example, indicated they thought about their mistakes and would likely do things differently next time; while Non-Experts focused on general attributes, and Novices indicated their mistakes were due to weaknesses in "power". Although this study provides compelling support for the three phase model, there has been no attempt to apply this methodology to an academic task to date.

The Rating Student Self-Regulated Learning Outcomes: A Teacher Scale. Although the measures described above assess students' beliefs and their self-regulatory behavior, they do not provide information on how a student is viewed from a teacher's perspective. In a study conducted by Zimmerman and Martinez-Pons (1989) they examined the relationship between students' reports of self-regulatory learning and teacher observations. In their investigation, a rating scale was developed by the researchers on readily observable strategies in school, and immediate outcomes of strategy use such as completing assignments on time and coming to class prepared. In addition, intrinsic questions were developed to measure indirect as well as direct use of strategies. The purpose of the study was to assess teachers' ratings of students' self-regulated learning, and to correlate these teacher ratings with the structured interview and standardized measures of achievement.

The Rating Student Self-Regulated Learning Outcomes: A Teacher Scale (RSSRL) was developed by the Zimmerman and Martinez-Pons (1989) from an original pool of 25 items that indicated students' self-regulatory learning strategies. The pool of items was then submitted to a panel of four experienced teachers which resulted in reducing the pool to a total of 12 items. Items involved students' help-seeking activities, self-evaluation activities, goal-setting and planning, seeking of information, organization and transforming strategies, and intrinsic motivation to learn. Eighty 10th grade students participated in the study. Their teachers were given the RSSRL and the students were interviewed individually. The results from this study indicate that the two different methodological approaches to measuring students' use of self-regulated learning strategies converged to reveal an underlying construct. In addition, student scores on standardized mathematics and English sections of the Metropolitan Achievement Test correlated with the teacher ratings and the structured interview reports. These researchers conclude their findings with the recommendation that further research be conducted in which self-regulated learning strategies might be explored within the contexts of which students are tested. This is one of the purposes of the current study.

Measuring Expertise

In Ericsson, Charness, Feltovich, and Hoffman's edited book: *The Cambridge Handbook of Expertise and Expert Performance* (2006) there are many articles addressing various domains in which experts have been identified. For example, there are chapters on Expertise in Chess (Gobet, & Charness), Music (Lehmann, & Gruber), Artistic Performance: Acting, Ballet, and Contemporary Dance (Noice, & Noice, 2006)

Mathematical Expertise (Butterworth), Expertise in History (Voss, & Wiley) and others. Although many of these articles describe many aspects of expertise in these fields, they seldom describe the activities of experts as they learn new material.

Ericsson (2006) concludes that expertise is developed over time and that personal improvement requires the opportunity to find tasks that are challenging, and this is often done in cooperation with a teacher or a coach. He indicates that "concentration" is required while engaging in the practice, and that without high levels of concentration, increased skill is unlikely. Concentration is needed in order to distinguish effective practice from a routine activity which is automatic, and unlikely to enhance one's expertise further. In addition, Ericsson indicates that tasks must be carefully structured and that the individual must plan, analyze, execute, and monitor his or her performance. Zimmerman (2006) suggests that a skilled teacher will provide the opportunities for students to practice these skills while learning. In addition, children must have the opportunity to be exposed to models, teachers, and peers who will provide examples through modeling and instruction. Zimmerman states that as children gain higher levels of performance, the external supports previously received from models, teachers and peers, will gradually be eliminated as the learner becomes more independent and self-directed

The think-aloud methodology. Ericsson (2006) has recommended a think-aloud measure to assess what experts may be "thinking" as they perform a task. According to this methodology the participant will verbalize his or her thought processes as he or she engages in a task. Greene and Azevedo (2007) have used the think-aloud methodology to assess self-regulated learning of a science task using hypermedia. In their study, middle and high school students were presented with material on the circulatory system using a Hypermedia Learning Environment called Microsoft Encarta DVD (2003). After having completed a pretest, the participants were instructed to "learn all you can" using the hypermedia and to "think aloud continuously" as they learned about the circulatory system. They were specifically also told that "it is important to say everything that you are thinking while you are working on this task" and were reminded to verbalize their thinking when they were silent for more than three seconds throughout their learning session. Students were then presented with a posttest measure. Student responses were audio and video-recorded and transcribed and coded. Findings from this study using the think-aloud protocol suggest that students who demonstrated a qualitative shift in learning more frequently used six self-regulatory processes such as feeling of knowing (metacognitive monitoring), expecting the adequacy of information (monitoring their learning), and learning strategies such as control of context, coordinating informational sources, inferences, and knowledge elaboration. Although this study is promising in that it offers information about the self-regulatory processes a learner may engage in while learning is taking place, it does not examine motivational beliefs or the differences among expert and novice learners in the use of the subprocesses of Zimmerman's (2002) three phase model of self-regulated learning.

Studies using the think-aloud and verbalizations. In another study, Winters and Azevedo (2005) examined self-regulated learning of science tasks with hypermedia, and by pairing high and low prior knowledge students using verbalizations to determine self-regulatory behavior. In their study, ninth and tenth grade students were grouped in heterogeneous pairs based on their pretest scores on prior knowledge of genetics material and science inquiry skills. The methodology employed was not a think-aloud, verbalizations of the students as they worked collaboratively were audio-recorded. The tapes were then transcribed and coded for self-regulatory behavior. The study took place over four of the students' regular science periods, during a two week period, and in collaboration with their science teachers. The hypermedia program of GenScope, a genetics and evolution-themed simulation program, was used to allow the students to learn about genetics and, inheritance and natural selection. The program is interactive in that it allows students to pose questions, manipulate the data, and analyze the information produced. In addition to the audio-recordings, students were administered a posttest on knowledge learned from the GenScope.

Findings from this study suggested that students who had low prior knowledge regulated their learning qualitatively differently than students in with high prior knowledge. For example, low prior knowledge students tended to seek procedural information from their partners whereas the high prior knowledge students tended to verbalize conceptual information which was instrumental in helping them answer the inquiry questions. With regard to goal setting, the low prior knowledge students tended to verbalize the teacher-set goals where the high prior knowledge students spent more time verbalizing their own goals. Surprisingly, posttest findings of students with low prior knowledge experienced significant gains, while the students with the high prior knowledge did not. One possible explanation for this is that the students with the high prior knowledge may have experienced ceiling effects. Interestingly, although this study did not use the think-aloud protocol, it attempted to measure self-regulation from the conversations of paired students.

Conclusions

The current study also attempted to examine self-regulatory behavior differences among students on an academic science task. However, the current investigation differs from those described above in a number of ways. Even though technology has made its way into the classroom, the typical learning tasks for students in school involve reading texts in subjects such as science (biology, chemistry, physics), social studies (e.g. American history, world history, geography), art history, English (e.g. literature, fiction, current events), and foreign language among others. Students are regularly required to read information texts, to study the material, and to demonstrate proficiency through a test. The current investigation attempted to simulate a classroom setting in which students were presented with science material, informed that they would be tested on the material, given the opportunity to prepare for the exam, tested, and presented with the results of the test score. Their individual self-regulatory processes were examined throughout the session.

Unfortunately, the previous studies did not address self-motivational beliefs such as self-efficacy, outcome expectations, intrinsic interest/value and goal orientation. These subprocesses are essential to the cyclical nature of self-regulation in that they have been found to significantly influence self-regulatory performance and are significantly influenced by the self-reflection subprocesses which occur after the performance has taken place (Zimmerman, 2002). The current investigation attempted to measure motivational beliefs that students may have before they begin to study and prepare for a test, and what they think about their ability to perform well on a test once they have seen the test, experiences and beliefs which many students have a daily basis.

The present investigation used the microanalytic methodology rather than the think-aloud in an attempt to assess the specific self-regulatory subprocesses which were (or were not) taking place during the act of learning. The microanalytic approach was used in an attempt to target specific psychological processes. Bandura (1986) raises a few concerns about measuring thoughts by means of verbal self-reports. Bandura suggests that thoughts are not always easy to put into words and that people may intentionally, or unintentionally, misrepresent what they are thinking. In addition, the assumption that verbal descriptions may mean the same things to different people can be questioned. Each of these concerns may result in measurement error, which is an important reason for using microanalytic method.

The microanalytic approach provides information about differences which exist across expertise level and gender. Psychological behaviors, feelings, and thoughts were examined as students within each of the groups engaged in self-regulatory behavior while they performed a science task. Students in the 11th grade were selected to participate because it was assumed that after ten years of experience in the school, they would have reached a level of consistency in expertise regarding their study habits and their motivational beliefs in science. Students in the 11th grade have had been exposed to science for at least ten years because science is a required academic subject across the entire school curriculum. In addition, the students were selected based on their prior year's final grades and teacher reports (RSSRL) on the students' self-regulatory behavior for the entire 10th grade in science.

The students' microanalytic responses, their test score on the science task, the teacher scale and the scale on self-efficacy for learning completed by the student, were all examined in an attempt to obtain a complete picture of students' self-regulatory processes across expertise level and gender. Findings from this study have significant educational implications for students of all ages in that Zimmerman's three phase model could be used in classroom instruction to assist students who are non-experts or at-risk to reach higher academic achievements, particularly girls who have often been pegged as not having the aptitude for science related subjects. Findings from this study could be used to develop students' self-regulatory behavior and to raise their self-beliefs so that students may have more confidence in their ability to master difficult tasks, both inside and outside of the classroom. This could have a significant impact on students' lives as they take their academic learning experiences and apply them to real world settings and lifelong learning.

Chapter Three

Method

Participants

The participants in this study were fifty-one 11th grade students from three middle class, suburban parochial high schools. The ethnicities of the students were 9.8% Asian or Pacific Islander, 19.6% African American, 19.6% Hispanic, and 51% White. Individual and parental consents were obtained from the children and their parents or legal guardians.

The students were grouped according to three categories: experts, non-experts, and at-risk based on the recommendation of the assistant principals at each school. *Experts* were students in advanced placement and honors biology courses in 10th grade who earned a final grade average of 90 or higher. To qualify for these advanced courses, the students had to display consistently high levels of achievement in science. *Non-experts* were students who earned what the assistant principals deemed to be the school's "average" final term grades of 84-87" in 10th grade biology. The equivalent of the novice level of expertise in biology were *at-risk* students who either failed 10th grade biology or scored the minimum passing grade for that school with the range being from 65-75. The sample consisted of 17 students in each expertise level with 17 boys and 34 girls as indicated in Table 1 below, and each expertise level had students represented from each school as indicated in Table 2 below.

Table 1. Sample: Expertise level by gender

		Expertise level			
		At-risk	Non-expert	Expert	Total
Gender	Male	6	6	5	17
	Female	11	11	12	34
	Total	17	17	17	51

Table 2. Sample: Expertise level by school

		Expertise level			
		At-risk	Non-expert	Expert	Total
School	1	4	11	5	20
	2	10	2	7	19
	3	3	4	5	12
	Total	17	17	17	51

Task and Materials

Students were given a science reading passage on the topic of tornadoes. The passage is three pages, double spaced with a formula embedded in the text and a table to refer to at the end. The passage was taken primarily from an undergraduate college-level textbook on natural disasters with additional information from a book on meteorology and two websites as cited in the list of references. The passage briefly describes how a tornado is formed, the types of damage it can cause, and how to protect oneself. In the body of the text a formula of how to calculate the speed of a tornado is described and a table indicating the intensity of a tornado based on the calculated wind speed. (See Appendix A for a copy of the tornado passage). In addition to the science reading passage, an audio-tape recorder was used to record each microanalytic session.

Pens, pencils, magic-markers, highlighters, blank index cards, and lined paper were on the desk and available for the student to use. Each room contained a large clock which was directly in front of the student's desk.

Measures

Tornado Knowledge Test. This scale involves a 10 item short-answer completion test that requires both recall and explanation of scientific knowledge about tornadoes. Questions are primarily short answer. For example: "Define tornados."…*and*… "List four ways in which tornados cause damage."

Tornado Conceptual Model Test. This test question required students to form a conceptual model of the three stages of tornado development. Students were asked to describe and draw a diagram of these three stages. There was no diagram in the passage, therefore students must have been able to form an abstract image of each stage as they read and studied the passage. This question was placed within the Tornado Knowledge Test so that it would appear to be one examination consisting of 11 test questions. (See Appendix B for the ten question Tornado Knowledge Test and the Conceptual Model Test: question two.)

Rating Student Self-regulated Learning Outcomes: A Teacher Scale (RSSRL). This questionnaire involves 12 items that are answered using a 5 point rating scale ranging from *never* (1) to *always* (5). The RSSRL was found to have high levels of reliability and predictive validity (Zimmerman & Martinez-Pons, 1989). Questions in this scale assess a teacher's observations of a student's academic self-regulation in class, such as "Does this student solicit additional information about the exact nature of forthcoming tests?" and "Is this student prepared to participate in class on a daily basis?" Teachers score the student on a scale of one to five with one being "never", three being "often" and five being "always". This questionnaire was given to the 10th grade biology teacher who taught each of the students during the previous academic year. In five cases where the biology teacher was no longer at the school, an alternative teacher was selected by the assistant principal to complete the form. (See Appendix C for the RSSRL).

Microanalytic processes measure of self-regulated learning. These scales involved on-line measures of processes presented earlier in Figure 1. Students' scores on these scales have shown high levels of reliability and predictive validity in prior research using an athletic task (Kitsantas & Zimmerman, 2002). In the present study, students were asked questions as they read and studied the science textual materials aimed at measuring the various subprocesses in Zimmerman's three phase model of self-regulation. This online measure is outlined below in the study session protocol.

Self-Efficacy for Self-Regulated Learning Form-A (SELF). This scale is a 19 item questionnaire involving a 5 point rating scale. It has demonstrated high levels of alpha reliability and predictive validity (Zimmerman & Kitsantas, 2007). Questions in this questionnaire involve assessing a student's self-efficacy judgments of his/her capability to self-regulate learning and "science" was added to the questions to make them more specific to science content. Examples of questions were "When you miss a science class, can you find another student who can explain the lecture notes as clearly as your teacher did?" and "When problems with your friends and peers conflict with school work, can you keep up with your science assignments?" Student responses were on a scale ranging from 0% to 100% with 10% being *definitely cannot do it* and 50% being *maybe*, and 100% being definitely *can do it* (See Appendix D for the SELF-A).

Procedure

Students were tested individually outside of the classroom during the winter at a time when the school assistant principals determined it to be least disruptive to their lessons. Each student was escorted from the class to a study room outside of the classroom. The entire session was audio-taped in order to accurately record student responses. In addition, the investigator kept written records of the students' behavior observed during the sessions. The duration of the session varied based on the student, but lasted no longer than 60 minutes. Table 3 below outlines when each of the measures described above were administered.

Table 3. Measures given during research phase

Research Phase	**Measures**
Pre-Research Phase: Scale completed by previous year's biology teachers	Rating Student Self-Regulated Learning Outcomes: A Teacher Scale
Phase 1 of Academic Self-Regulation: Forethought	Microanalytic Forethought Phase Questions: Self-regulation is assessed before learning
Phase 2 of Academic Self-Regulation: Performance	Microanalytic Performance Phase Questions: Self regulation is assessed during learning Tornado Knowledge Test
Phase 3 of Academic Self-Regulation: Self-Reflection	Microanalytic Self-Reflection Phase Questions: Self-regulation is assessed after learning
Post-Research Phase: Upon completing the protocol and just prior to returning to class	Student Scale: Self-Efficacy for Self-Regulated Learning Form-A

Study session protocol. Each student was individually tested by the investigator. Students were first shown their signed copy of the consent form and asked to review the form and asked if they had any questions. As outlined in the table above, once the session began, the microanalysis took place where students were asked questions as they read and studied for the Tornado Knowledge Test and the Conceptual Model Test, and then immediately after completion of the test.

In conducting the microanalysis, students were presented response choices on cue cards (unless otherwise indicated) and asked to point to their answer. Each answer was carefully recorded by the investigator and audio-recorded to ensure accuracy.

After reviewing the consent form and prior to beginning the microanalysis, students' prior knowledge about tornados was assessed to determine if there were any differences that existed between schools, by expertise level, or by gender, by asking them the following question: "What do you know about tornados?" The number of correct facts was recorded. The following introduction was then given to students:

> "I am going to give you a science passage on tornados to read, and then I will test you on it. It is really important that you learn as much as possible about tornados so that you can do as well as you can on a test. You have up to 40 minutes to study for the test, and when you are ready to take the test, let me know. As you read and study the passage for the test, I am going to ask you some questions and will tape record your answer so I will not forget anything important that you say, ok?"

Microanalysis of Phase 1: Forethought

Self-efficacy for learning measure. Self-efficacy questions were developed to assess the students' self-confidence about his/her ability to learn and remember the material for the tornado assessment and on new science topics. For this construct, and two other constructs, students were asked two questions, one referring to the topic at-hand of tornados, and a second item referring to the field of science in general.

The questions for self-efficacy were asked of the students after they had scanned the passage. The investigator gave the students time to look over the passage. The self-efficacy questions asked are listed below:

"OK, before we begin I'm going to ask you a few quick questions. Please respond by pointing to the number on the scale that applies to you.

Q.1 How self-confident do you feel in your capability to completely learn and remember all of the material on tornados in this passage?"

|..........|.........|..........|..........|..........|..........|..........|..........|..........|
10 20 30 40 50 60 70 80 90 100
Not sure at all — somewhat unsure — pretty sure — very sure

Q.2. "How self-confident do you feel in your capability to completely learn a new topic in science?"

|..........|.........|..........|..........|..........|..........|..........|..........|..........|
10 20 30 40 50 60 70 80 90 100
Not sure at all — somewhat unsure — pretty sure — very sure

Intrinsic interest/value measures. Students were asked two questions to assess their intrinsic value/interest:

Q.3. "How interested are you personally in learning about tornados?"

|…...……….|………...….|……..……….|…...……….|
1 2 3 4 5
really not interested — not interested — neutral — interested — really interested

Q.4. "How interested are you personally in learning about new topics in science?"

1	2	3	4	5
really not interested	not interested	neutral	interested	really interested

Outcome expectations measures. Students were asked two questions to assess their outcome expectations:

Q.5. "How important is information about tornados to your future?"

1	2	3	4	5	6	7
really unimportant	unimportant	somewhat unimportant	neutral	somewhat important	important	really important

Q.6. "How important is information about science to your future?"

1	2	3	4	5	6	7
really unimportant	unimportant	somewhat unimportant	neutral	somewhat important	important	really important

Goal setting measure. Students were asked the following question to assess their goal setting.

Q.7. "What percentage grade will you set as your goal on this test of knowledge?"

10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

Task procedures. Students were reminded of the purpose of the task and were directed to prepare for the assessment as follows:

> "OK, you have the material you are going to be tested on. As I said, it is really important that you do as well as you can on the tornado test. You can use any of these materials (pointing to markers, etc) if you want, but you don't have to. After you read the passage, tell me when you are ready to take the test, and I will give it to you. When I give you the test, I will

have to take away the tornado passage and any study material you have. If you have any questions, now is a good time to ask me. I am going to be working on something while you read and prepare for the test. Do you have any questions?"

Strategic planning measure. After the students had responded as to whether or not they had any questions, the students were asked a question to determine their strategic planning. A list of possible responses was available for the investigator to check off with space available for additional answers to be noted.

Q.8. "Oh, I meant to ask you before you get started; do you have any particular plans for how to read this passage about tornados and take the test?"

If yes: "Can you tell me about them?"

If no: "Do you have any particular methods of studying if you find the text material difficult to understand or remember? Can you tell me about them?"

Possible responses to this question include:

- Reading the passage carefully ☐
- Rereading ☐
- Reading aloud ☐
- Underlining ☐
- Highlighting ☐
- Taking notes ☐
- Writing down facts ☐
- Studying facts/information ☐
- Other___________________________

____

No methods used ☐

"Ok, great, you can go ahead and begin."

Microanalysis of Phase 2: Performance

This phase consisted of three parts: 1. while reading, 2. while studying, and 3. post science assessment. In addition, a time measure was recorded at the point in which the student began to read the passage and at the point in which the student indicated he or she was ready to take the test.

Part I: While reading (before studying)

Task strategy measure.

Q. 9. "I noticed while you are reading that you are highlighting (or rereading, underlining, or note taking in the margins, or talking aloud as you studied, etc.) could you explain to me what you are doing and why?"

Part II: After the students have completed the initial reading of the passage but prior to taking the test (while studying).

Task strategy and metacognitive monitoring measures.

"You may begin to study."

The investigator recorded what the students were doing while studying and then after 10 minutes asked:

Q. 10. "Can you explain to me how you are preparing for the test? What exactly are you doing?"

Possible responses to this question included:

"I am….

- Reading the passage carefully ☐
- Rereading ☐
- Reading aloud ☐
- Underlining ☐
- Highlighting ☐
- Taking notes ☐
- Writing down facts ☐
- Studying facts/information ☐

Other__________________________

____

Not using any particular method □

"OK, please continue. Let me know when you are ready to take the test."

Self-efficacy of performance measure (a Phase 1 measure for test-taking). At the time in which the student indicated he or she was ready to take the test, the investigator recorded the time and removed the reading passage and any study material. The investigator gave the student the test and asked him/her to read the test questions without responding. A self-efficacy question was asked to assess the students' self-confidence about his/her ability to earn a perfect score on the test.

Q.11. "How confident do you feel in your capability to earn 100% on this test?"

|……….|……….|……….|……….|……….|……….|……….|……….|……….|

10	20	30	40	50	60	70	80	90	100
Not sure at all			somewhat unsure			pretty sure			very sure

Once the student had responded to the above question, the student was instructed to take the test.

Metacognitive monitoring for each of the questions on the Tornado Knowledge Test and on the Conceptual Model Test. Upon completion of the test, the students were asked the following for each of the eleven questions:

Q. 12-22. "How confident do you feel about your answer to question 1 (…2, 3…)?"

|……….|……….|……….|……….|……….|……….|……….|……….|……….|

10	20	30	40	50	60	70	80	90	100
Not sure at all			somewhat unsure			pretty sure			very sure

Part III: Upon completion of the test, the investigator asked students a question to assess whether their performance matches what they thought their scores would be.

Metacognitive monitoring measure.

Q. 23. “What score do you think you got on the test?”

|……….|………|……….|……….|……….|……….|……….|……….|……….|……….|
10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Q. 24. “What score do you think you got on question #2 on drawing and labeling the three phases of tornado development? Each drawing and each description was worth one point out of six.”

|……….|………|……….|……….|……….|……….|……….|
0 1/6 2/6 3/6 4/6 5/6 6/6

Microanalysis of Phase 3: Self-Reflection

Self-evaluation measure. Upon grading the Conceptual Model Test (diagram and labeling of the three phases of a tornado), the investigator showed the Conceptual Model Test grade to the students.

Q.25. “How well did you learn about the three phases of tornado development?”

|……….|………|……….|……….|……….|……….|……….|……….|……….|……….|
10 20 30 40 50 60 70 80 90 100
poorly pretty poorly pretty well very well

Causal attribution measure. Students were asked a question to assess their causal attributions to their performance. (This is an open-ended question to which all responses were recorded.)

Q.26. “Why do you think you didn’t do better on this particular test question on tornado development? Please explain.”

For students who answered this question 100% correctly:

"Why do you think you did so well on this particular test question on tornado development? Please explain."

Self-satisfaction measure.

Q.27. "How satisfied are you with your score on this particular test item?"

|…...……….|……….…..|…….……….|…...……….|…….……….|…….……….|

1	2	3	4	5	6	7
very dissatisfied	dissatisfied	somewhat dissatisfied	neutral	somewhat satisfied	satisfied	very satisfied

Adaptive/defensive measure. Students were asked a question to assess their adaptive or defensive inferences about future attempts to learn about the topic of tornados.

Q.28. "Is there anything you would do differently on this particular test question on tornado development if you were given another chance to study the material on tornados? Please explain."

Table 4 presents a summary of the questions listed above as they addressed Zimmerman's three phases of self-regulation and their subprocesses.

Distribution of the SELF-A

Finally, the students were given SELF-A questionnaire to complete. Upon completion, they were thanked for their participation and escorted back to the classroom.

Table 4: Summary of microanalytic questions addressing Zimmerman's three phases of academic self-regulation and their subprocesses

Phase	***Microanalytic Measures Obtained***
Phase 1: Forethought (Pre-reading)	- self-efficacy (learning Q.1, Q.2 and performance Q.11) - outcome expectations Q.5, Q.6 - intrinsic values/interest Q.3, Q.4 - goal setting Q.7 - strategic planning Q.8
Phase 2: Performance: Part I (While reading)	- task strategy Q.9
Phase 2: Performance: Part II (While studying)	- task strategy Q.10 - metacognitive monitoring Q.12-22
Phase 2: Performance: Part III (Upon completion of the test)	- metacognitive monitoring Q.23, 24
Phase 3: Self-Reflection (Post assessments)	- self-evaluation Q.25 - self-satisfaction Q. 27 - causal attribution Q. 26 - adaptive/defensive responses Q.28

Chapter 4

Results

Person Measures

Tornado Knowledge Test. Because the questions on the Tornado Knowledge Test were open-ended, it was necessary to establish their reliability using an index of inter-rater agreement. The scores of the experimenter and an independent scorer were correlated using Pearson's Product correlation. The correlation between tornado knowledge ratings was $r = .75$, $p < .01$. These results indicate substantial inter-rater agreement regarding scoring of the Tornado Knowledge Test.

A principal component factor analysis was conducted to examine the factor structure of the test. The results indicated that there was one main factor which accounted for 37% of the variance and was labeled *tornado knowledge*. Two test items also loaded negatively on a second factor that involved problem solving using the Fujita Scale. This second factor explained 11.30% of the variance and was labeled *tornado problem solving*. The first item involved using a mathematical formula and loaded -.66 on the second factor. The second item involved applying the solution obtained in the prior test item to a scale. This item also loaded negatively -.44 on the second factor. Although these two items also loaded highly on the first factor (.54 and .66 respectively), the two factors are independent and were therefore, included as part of the 10 question test format. The negative tornado problem solving loadings imply that a lack of math skill could adversely affect scores on these two items.

Tornado Conceptual Model Test. This measure involved illustrating and labeling the three phases of tornado development. Although the tornado knowledge passage provided students with a written description of the three phases, there were no illustrations. The conceptual model measure required students to be able to form three sequential images of tornado development from the textual description of the three phases. Because these questions on the Tornado Conceptual Model Test were also open-ended, it was necessary to establish their reliability using an index of inter-rater agreement. The scores of the experimenter and an independent scorer were correlated using Pearson's Product correlation, and a coefficient of $r = .72$, $p < .01$ inter-rater agreement was obtained. These results indicate substantial inter-rater agreement regarding the scoring of the Tornado Conceptual Model Test.

This measure was included in the principal component factor analysis, and it loaded the highest on the main factor (.82) indicating it was a factorially valid measure of tornado knowledge. The Pearson Product Correlation between the Tornado Knowledge Test score and the Conceptual Model Test was $r = .70$, $p < .01$.

Prior knowledge. A measure of prior knowledge of tornados was analyzed to determine if there were differences between students in different schools and by expertise level and gender. Students were asked what they knew about tornados, and all correct facts were tallied for each student, whereas incorrect facts were ignored. Inter-rater reliability was $r = .77$, $p < .01$. These results indicate substantial inter-rater agreement regarding the scoring of prior knowledge. The mean for school one was 2.85, school two was 2.74 and school three was 2.84. Analysis of variance tests revealed no significant differences based on school affiliation and prior knowledge of students. A two-way analysis of variance did not reveal a significant main effect for expertise. There was also no main effect for gender or interaction between expertise level and gender on prior knowledge.

Tornado knowledge, expertise level, and gender analysis. A two-way analysis of variance was used to assess differences in achievement scores based on students' expertise level and gender. When significant differences were found, a linear trend analysis was conducted to determine the overall relation between expertise and tornado learning outcomes. The trend analysis was followed by post hoc Tukey tests to determine pairwise differences between expertise groups.

Tornado Knowledge Test. The two-way analysis revealed a statistically significant main effect for expertise on the Tornado Knowledge Test, F (2, 45) = 15.99, $p < .01$. The effect size is large (partial eta squared .42) suggesting a large portion of the variance is explained by the students' expertise level in science (Cohen, 1988). A trend analysis revealed a significant positive linear increase in tornado knowledge, F (1, 48) = 37.27, $p < .01$, based on the students' expertise level. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between all three expertise groups. The mean and standard deviation scores for the expert group were M = 75.74, SD = 9.97, while the mean and standard deviation scores for the non-expert group were M = 61.01, SD = 12.70 and the mean and standard deviation scores for the at-risk group were M = 49.27, SD = 14.77. There was no main effect for gender or interaction between expertise level and gender on the Tornado Knowledge Test.

Tornado Conceptual Model Test. The two-way analysis of variance revealed a statistically significant main effect for expertise on the Conceptual Model Test, F (2, 45)

= 16.68, $p < .01$. The effect size for expertise level (partial eta squared) was .43, which is considered large. A trend analysis revealed a significant positive linear trend, F (1, 48) = 39.20,

$p < .01$, for expertise regarding achievement on the Conceptual Model Test. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were M=5.18, SD=.95, while the mean and standard deviation scores for the non-expert group were M = 3.29, SD = 1.83 and the mean and standard deviation scores for the at-risk group were M = 1.94, SD = 1.60. There was also a significant main effect for gender, F (1, 45) = 5.26, $p < .03$. The effect size for gender (partial eta squared) was .10, which is considered moderate. There was no interaction between expertise level and gender.

Rating Student Self-Regulated Learning Outcome scale. Cronbach's alpha was found to be .91. A two-way between group analysis of variance revealed a statistically significant main effect for expertise on the RSSRL, F (2, 45) = 14.77, $p < .01$. The effect size for expertise level on these teacher ratings of self-regulation (partial eta squared) .40, is considered large. A trend analysis revealed a significant positive linear trend for expertise on the RSSRL, F (1, 48) = 33.30, $p < .01$. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were M = 3.33, SD = 1.80, while the mean and standard deviation scores for the non-expert group were M = 2.74, SD = 1.70 and the mean and standard deviation scores for the at-risk group were M = 1.99, SD = 1.70. There was no main effect for gender or interaction between expertise level and gender on the RSSRL.

Self-Efficacy for Learning scale. Cronbach's alpha was found to be .92. A two-way analysis of variance revealed a marginally significant effect of expertise, $F(2, 45) = 2.14$, $p < .07$, one-tailed. A trend analysis revealed a significant negative trend for expertise on the SELF, $F(1, 48) = 9.71$. Post hoc comparisons using Tukey tests indicated significant pairwise differences between expertise and at-risk groups, $F(2, 45) = 3.77$, $p < .01$. The mean and standard deviation scores for the expert group were $M = 69.44$, $SD = 2.79$, while the mean and standard deviation scores for the non-expert group were $M = 72.75$, $SD = 2.79$, and the mean and standard deviation scores for the at-risk group were $M = 77.75$, $SD = 2.92$. There was also no main effect for gender or interaction between expertise level and gender.

Microanalytic Measures

As previously outlined in Figure 1, there are three phases of self-regulation in Zimmerman's theory. A two-way analysis of variance was used to assess differences in microanalytic measures for all three phases based on students' expertise level and gender. For the strategic planning and the task strategies measures, inter-rater reliability was assessed first and followed by two-way analyses, trend analyses, and Tukey HSD. In the self-reflection phase, two measures were categorical, and after determining inter-rater reliability, these measures were analyzed using chi-square tests. Following the findings on each of Zimmerman's three phases of academic self-regulation is a summary table (Table 7) which presents the means and standard deviations for each of the subprocesses assessed across expertise level, and a summary table outlining when and where there were significant differences across expertise groups (Table 8).

Forethought phase measures. This phase of self-regulation consists of two major categories: task analysis and self-motivational beliefs. The task analysis category involves two subprocesses: goal setting and task strategic planning. Regarding *goal-setting,* a two-way analysis of variance did not reveal a significant main effect of expertise. There was also no main effect for gender or interaction between expertise level and gender for this measure.

The *strategic planning* subprocess involved counting and analyzing the number of strategies listed by the students. Inter-rater reliability was $r = .83$, $p <.01$. The two-way analysis of variance revealed a statistically significant main effect for expertise on strategic planning, F (2, 45) = 8.76, $p < .01$. The effect size for expertise level (partial eta squared) was .28, which is considered large. A trend analysis revealed a significant positive linear trend for expertise, F (1, 48) = 10.24, $p < .01$, on strategic planning. Post hoc comparisons using Tukey tests indicated significant pairwise differences with expert and non-expert differing from the at-risk group. The mean and standard deviation scores for the expert group were $M = 2.29$, $SD = .92$, while the mean and standard deviation scores for the non-expert group were $M = 2.35$, $SD = .92$ and the mean and standard deviation scores for the at-risk group were $M = 1.35$, $SD = .70$. There was no main effect for gender or interaction between expertise level and gender on the strategic planning measure.

The next category in the forethought phase involved self-motivational beliefs. Three subprocesses that were measured in this category: *self-efficacy, outcome expectations*, and *intrinsic interest*. With regard to *self-efficacy,* three measures of self-efficacy were obtained. Two measures assessed students' self-efficacy for learning, one measure assessed students' self-efficacy for performance.

The *self-efficacy for learning and remembering the material in the tornado passage* was analyzed using the two-way analysis of variance. It did not reveal a significant main effect for expertise, gender, or the interaction between expertise level and gender.

With regard to *self-efficacy for learning new topics in science,* the two-way analysis of variance did not reveal a significant main effect for expertise gender, or the interaction between expertise and gender.

The *self-efficacy for performance* was analyzed using the two-way analysis of variance. It did not reveal a significant main effect for expertise on self-efficacy for performance. There was also no main effect for gender or interaction between expertise and gender.

The next set of self-motivational beliefs involved *outcome expectations.* Two outcome expectations measures were obtained. The first outcome measure assessed students' *outcome expectations about the importance of information about tornados to his/her future.* The second measure assessed students' outcome expectations about the importance of information about science to his/her future. The two-way analysis of variance did not reveal a significant main effect for expertise, gender, or interaction between expertise level and gender.

The two-way analysis of variance did not reveal a significant main effect of expertise on *outcome expectations regarding the importance of information about science* *to one's future*. There was also no main effect for gender or interaction between expertise level and gender on this outcome expectation measure.

The last set of self-motivational beliefs involved students' *intrinsic interest.* Two intrinsic interest measures were obtained. The first intrinsic interest measure assessed students' *interest in learning about tornados*. The second intrinsic interest measure assessed students' *interest in learning new topics in science*. The two-way analysis of variance revealed a marginally significant main effect for expertise on intrinsic interest in learning about tornados, $F(2, 45) = 2.92$, $p = .06$. The effect size for expertise level (partial eta squared) was .12, which is considered moderate; however, the trend analysis did not reveal a significant positive linear trend. Post hoc comparisons using Tukey tests indicated significant pairwise differences between non-expert and the other two groups. The expert group did not differ significantly from the at-risk group. The mean and standard deviation scores for the expert group were $M = 3.49$, $SD = .21$, while the mean and standard deviation scores for the non-expert group were $M = 3.92$, $SD = .20$ and the mean and standard deviation scores for the at-risk group were $M = 3.26$, $SD = .20$. There was no main effect for gender or interaction between expertise level and gender on the intrinsic interest in learning about tornados measure.

Regarding the second intrinsic interest measure, *intrinsic interest in learning about new topics in science*, the two-way analysis of variance did not reveal a significant main effect for expertise and intrinsic interest in learning about new topics in science. There was also no significant main effect for gender or interaction between expertise level and gender.

Performance phase measures. As previously outlined in Figure 1, this phase consists of two major categories: self-control and self-observation. Within the self-control category the sub-process of *task-strategy* was assessed. Two observational measures of task strategies were obtained. The first measure, was assessed while students were reading the tornado passage. The second assessment was made while students were studying the tornado passage.

Regarding the *task-strategy* of *reading*, the number of strategies listed by the students were counted and analyzed. Inter-rater reliability was $r = .76$, $p < .01$. These results indicate substantial inter-rater agreement regarding scoring of the task-strategy of reading measure. The two-way analysis of variance revealed a statistically significant main effect for expertise in reading, F (2, 45) = 4.80, $p < .01$. The effect size for expertise level (partial eta squared) .18, is considered large. A trend analysis revealed a significant positive linear trend for expertise, F (1, 48) = 7.57 $p < .01$, on students' task-strategy for reading. Post hoc comparisons using Tukey tests indicated significant pairwise differences between the expert and the at-risk groups while the non-expert group did not differ significantly from either expert or at-risk groups. The mean and standard deviation scores for the expert group were $M = 2.59$, $SD = 1.00$, while the mean and standard deviation scores for the non-expert group were $M = 1.88$, $SD = 1.17$ and the mean and standard deviation scores for the at-risk group were $M = 1.65$, $SD = .79$. There was no main effect for gender or interaction between expertise level and expertise on students' task-strategy for reading measure.

The second *task-strategy* was *studying.* Again, the number of strategies listed by the students were counted and analyzed. Inter-rater reliability was $r = .92$, $p < .01$. These results indicate substantial inter-rater agreement regarding scoring the task-strategy for studying measure. The two-way analysis of variance revealed a statistically significant main effect for expertise on the task-strategy for studying, F (2, 45) = 5.44, p = .01. The effect size for expertise level (partial eta squared) .20, is considered large. A trend analysis revealed a significant positive linear trend for expertise, F (1, 48) = 15.56 p = .01. Post hoc comparisons using Tukey tests indicated significant pairwise differences between the expert and the at-risk groups while the non-expert group did not differ significantly from either expert or at-risk groups. The mean and standard deviation scores for the expert group were M = 3.35, SD = 1.32, while the mean and standard deviation scores for the non-expert group were M = 2.59, SD = 1.58, and the mean and standard deviation scores for the at-risk group were M = 2.0, SD = .79. There was no main effect for gender or interaction between expertise level and expertise on students' task-strategy for studying measure.

Within the second category of the performance phase: self-observation, the subprocess of *metacognitive monitoring* was assessed. *Metacognitive measures* were obtained on the Tornado Knowledge Test and on the Tornado Conceptual Model Test as follows. On the Tornado Knowledge Test, students were asked *a metacognitive question about each test item:* "How confident do you feel about your answer to question 1…11?" A metacognitive mean score was then calculated. The two-way analysis of variance revealed a statistically significant main effect for expertise on metacognitive monitoring, F (2, 45) = 15.80, p < .01. The effect size for expertise level (partial eta squared) .41, is considered large. A trend analysis revealed a significant positive linear trend for expertise, F (1, 48) = 34.90, p < .01, on students' metacognitive monitoring of test questions. Post hoc comparisons using Tukey tests indicated significant pairwise differences between the at-risk groups and both the expert and non-expert groups; however the difference between expert and non-expert groups was only marginally significant ($p < .20$). The mean and standard deviation scores for the expert group was $M = 91.82$, $SD = 2.21$, while the mean and standard deviation scores for the non-expert group were $M = 88.12$, $SD = 2.12$ and the mean and standard deviation scores for the at-risk group were $M = 75.62$, $SD = 2.12$. There was no main effect for gender or interaction between expertise level and expertise on students' metacognitive monitoring of test questions.

Regarding the Tornado Knowledge Test as a whole, students were asked the *metacognitive question of what score they think they earned on the Tornado Knowledge Test*. A two-way analysis of variance revealed a statistically significant main effect for expertise, $F(2, 45) = 15.53$, $p < .01$. The effect size for expertise level (partial eta squared) was .41, which is considered large. A trend analysis revealed a significant positive linear trend, $F(1, 48) = 36.09$, $p < .01$, for expertise on metacognitive monitoring of test score**.** Post hoc comparisons using the Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were $M = 90.60$, $SD = 2.17$, while the mean and standard deviation scores for the non-expert group $M = 83.48$, $SD = 2.07$ and the mean and standard deviation scores for the at-risk group was $M = 74.00$, $SD = 2.07$. There was no main effect for gender or interaction between expertise level and metacognitive monitoring of Tornado Knowledge Test score.

Regarding the Tornado Conceptual Model Test, students were asked the *metacognitive question on how confident they felt about their answer to the Tornado Conceptual Model Test.* A two-way analysis of variance revealed a statistically significant main effect for expertise on metacognitive monitoring of the Conceptual Model Test, F (2, 45) = 9.24, $p < .01$. The effect size for expertise level (partial eta squared) was .29, which is considered large. A trend analysis revealed a significant positive linear trend, F (1, 48) = 19.94, $p < .01$, for expertise on metacognitive monitoring. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were M = 91.38, SD = 4.70, while the mean and standard deviation scores for the non-expert group M = 85.68, SD = 4.49 and the mean and standard deviation scores for the at-risk group was M = 65.10, SD = 4.49. There was no main effect for gender or interaction between expertise level and gender on metacognitive monitoring of Tornado Conceptual Model Test.

Regarding the Tornado Conceptual Model Test score, students were asked a *metacognitive question of what score they think they earned on the Conceptual Model Test question*. A two-way analysis of variance revealed a statistically significant main effect for expertise on metacognitive monitoring of Conceptual Model Test score, F(2, 45) = 5.23, $p < .01$. The effect size for expertise level (partial eta squared) was.19, which is considered large. A trend analysis revealed a significant positive linear trend, F (1, 48) = 9.91, $p < .01$, for expertise on metacognitive monitoring of the Tornado Conceptual Model Test score. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between expert and non-expert and at-risk groups. The non-expert group, however, did not significantly differ from the at-risk group. The mean and standard deviation scores for the expert group were *M* = 4.98, *SD* = .32, while the mean and standard deviation scores for the non-expert group were *M* = 4.03, *SD* = .30 and the mean and standard deviation scores for the at-risk group were *M* = 3.59, *SD* = .30. There was no main effect for gender or interaction between expertise level and gender regarding metacognitive monitoring of the Tornado Conceptual Model Test score**.**

Self-reflection phase measures. As previously outlined in Figure 1, this phase consists of two major categories: self-judgment and self-reaction. Each measure in the two categories was based on the Conceptual Model Test. There were two measures obtained that are part of the self-judgment category: self-evaluation and causal attribution. With regard to *self-evaluation,* a two-way analysis of variance revealed a statistically significant main effect for expertise and self-evaluation, F (2, 45) = 12.77, $p < .01$. The effect size for expertise level (partial eta squared) .36 is considered large. A trend analysis revealed a significant positive linear trend for expertise, F (1, 48) = 31.51, $p < .01$. Post hoc comparisons using Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were *M* = 88.38, *SD* = 4.70, while the mean and standard deviation scores for the non-expert group were *M* = 72.89, *SD* = 4.48 and the mean and standard deviation scores for the at-risk group were *M* = 55.61, *SD* = 4.48. There was no main effect for gender or interaction between expertise level and self-evaluation.

The second subprocess in self-judgment is *causal attribution*. This involves an open-ended question that asks students to what cause do they attribute their score on the Conceptual Model Test. The responses were coded according to the following categories:

don't know/not sure, ability (for example: I can't draw), effort (for example: I didn't study hard enough), and strategy (I did not draw a diagram when I was studying). Inter-rater reliability and chi-square analyses were conducted. The inter-rater reliability was r = .81, $p < .01$. These results indicate substantial inter-rater agreement regarding the attribution measure. Table 5 below indicates the frequency of responses based on expertise level and attribution response.

Table 5. Attribution subprocess frequency count in the self-reflection phase

Attribution: Why do you think you did not do better on the Conceptual Model Test?	Expertise Level Count				
	Attribution	Expert	Non-expert	At-risk	Total
	Do not know	0	2	1	3
	Ability	3	6	9	18
	Effort	5	6	6	17
	Strategy/perfect score	9	3	1	13
	Total	17	17	17	51

A chi-square analysis revealed significant differences across expertise level, χ^2 (6) = 13.12, $p < .04$, $\varphi = .51$ on the attribution measure. A secondary chi-square was conducted to determine if there were differences between the expertise level of students who indicated either ability or strategy on their attribution measure. This chi-square analysis revealed a stronger significant difference between expertise level and attribution of ability or strategy, χ^2 (2) = 9.9, $p < .01$, $\varphi = .56$. No significant differences across gender were found.

With regard to the subprocess: *self-satisfaction* in the self-reaction category, a two-way analysis of variance revealed a statistically significant main effect for expertise and self-evaluation, F (2, 45) = 15.37, $p < .01$. The effect size for expertise level (partial eta squared**)** was .41, which is considered large. A trend analysis revealed a positive linear trend, F (1, 48) = 31.37, $p = .01$, across expertise level and self-evaluation. Post hoc comparisons using Tukey tests indicated significant pairwise differences between all three groups. The mean and standard deviation scores for the expert group were M = 5.97, SD = .35, while the mean and standard deviation scores for the non-expert group were M = 4.57, SD = .34 and the mean and standard deviation scores for the at-risk group were M = 3.27, SD = .34. While there was no interaction between expertise level and self-satisfaction, there was an effect for gender, F (2, 45) = 8.36, $p < .01$; the mean for females was 5.17 and males 4.03 and respective standard deviations were .23 and .32.

The second subprocess in self-reaction category is the subprocess of *adaptive* or *defensive inferences*. This is an open-ended question that asks students whether they would do anything differently if given another opportunity to prepare for the Conceptual Model Test. The responses were coded according to the following categories: no, I would not do anything differently, ability (e.g., I can't do any better), effort: (e.g., I would try harder next time), and strategy (e.g., I would have visualized the three phases). Students who earned a perfect score on the Conceptual Model Test and responded by indicating they would not do anything differently, were coded as strategic. Both inter-rater reliability and chi-square analyses were conducted. The inter-rater reliability was r = .93, $p < .01$. These results indicate substantial inter-rater agreement regarding the scoring of adaptive/defensive measure. Table 6 below indicates the frequency of responses based on expertise level and adaptation response.

Table 6. Adaptive/Defensive subprocess frequency count in the self-reflection phase

Adaptive/defensive: Is there anything you would do differently if given another chance to study the material on tornado development?	Expertise Level Count				
	Adaptive/defensive	Expert	Non-expert	At-risk	Total
	No, I would not do anything differently	0	1	2	3
	Ability	1	0	1	2
	Effort	1	10	7	18
	Strategy/perfect score	15	6	7	28
	Total	17	17	17	51

A chi-square analysis revealed significant differences across expertise level, χ^2 (6) = 15.21, $p < .02$, $\varphi = .49$ on the adaptive/defensive measure. Based on the frequency of responses, a secondary chi-square analysis was conducted to determine if there were differences between the expertise level of students who indicated either effort or strategy on their adaptive/defensive measure. This analysis revealed a stronger significant difference between expertise level and adaptive/defensive response based on effort or strategy use, χ^2 (2) = 11.63, $p < .01$, $\varphi = .44$. Chi-square tests revealed no significant differences across gender. Table 8 presents a summary of the differences found or not found in the subprocesses across expertise level**.**

Table 7. Means and standard deviations for metric microanalytic measures by expertise level

	Expert		**Non-Expert**		**At-Risk**	
	Means	**SD**	**Means**	**SD**	**Means**	**SD**
Forethought: Goal setting	88.23	2.90	88.98	2.76	90.19	2.76
Forethought: Strategic planning	2.44	.22	2.31	.21	1.27	.21
Forethought: Self-efficacy for learning about tornados	71.83	2.92	70.11	2.79	72.27	2.79
Forethought: Self-efficacy for learning new topics in science	68.58	3.97	73.98	3.78	79.09	3.78
Forethought: Self-efficacy for performance	72.98	4.00	77.77	3.81	75.61	3.81
Forethought: Outcome expectations: tornados	3.95	.36	4.3	.34	4.55	.34
Forethought: Outcome expectations: science	5.63	.35	5.41	.33	5.79	.33
Forethought: Intrinsic interest: tornados	3.40	.21	3.92	.20	3.26	.20
Forethought: Intrinsic interest: science	3.53	.22	3.99	.21	4.17	.21
Performance: Task-strategy: reading	2.71	.27	1.83	.26	1.61	.26
Performance: Task-strategy: studying	3.48	.34	2.46	.32	1.96	.32
Performance: Metacognitive monitoring: Tornado Knowledge Test	91.82	2.20	88.11	2.11	75.62	2.11
Performance: Metacognitive monitoring: Conceptual Model Test	91.38	4.70	85.68	4.49	65.10	4.49
Performance: Metacognitive Monitoring: Tornado Knowledge Test score	90.60	2.17	83.48	2.07	74.00	2.07
Performance: Metacognitive Monitoring: Conceptual Model Test score	4.98	.32	4.08	.30	3.59	.30
Self-Reflection: Self-evaluative standards	88.38	4.70	72.89	4.48	55.61	4.48
Self-Reflection: Attribution	-	-	-	-	-	-
Self-Reflection: Self-satisfaction	5.97	.35	4.57	.34	3.27	.34
Self-Reflection: Adaptive/defensive	-	-	-	-	-	-

Table 8: Summary of differences found across expertise levels in Zimmerman's three phase model of academic self-regulation

	Expert	Non-expert	At-risk
Forethought: Goal setting	-	-	-
Forethought: Strategic planning	Significantly differed from at-risk	Significantly differed from at-risk	Significantly differed from expert and non-expert
Forethought: Self-efficacy for learning about tornados	-	-	-
Forethought: Self-efficacy for learning new topics in science	-	-	-
Forethought: Self-efficacy for performance	-	-	-
Forethought: Outcome expectations: tornados	-	-	-
Forethought: Outcome expectations: science	-	-	-
Forethought: Intrinsic interest: tornados	Marginally differed from non-expert	Marginally differed from expert and at-risk groups	Marginally differed from non-expert
Forethought: Intrinsic interest: science	-	-	-
Performance: Task-strategy: reading	Significantly differed from at-risk	-	Significantly differed from expert
Performance: Task-strategy: studying	Significantly differed from at-risk	-	Significantly differed from expert
Performance: Metacognitive monitoring: Tornado Knowledge Test	Significantly differed from at-risk; marginally differed from non-expert	Significantly differed from at-risk; marginally differed from expert	Significantly differed from expert and non-expert
Performance: Metacognitive monitoring: Conceptual Model Test	Significantly differed from non-expert and at-risk	Significantly differed from expert and at-risk	Significantly differed from expert and non-expert
Performance: Metacognitive Monitoring: Tornado Knowledge Test score	Significantly differed from non-expert and at-risk	Significantly differed from expert and at-risk	Significantly differed from expert and non-expert
Performance: Metacognitive Monitoring: Conceptual Model Test score	Significantly differed from non-expert and at-risk	Significantly differed from non-expert	Significantly differed from expert
Self-Reflection: Self-evaluative standards	Significantly differed from non-expert and at-risk	Significantly differed from expert and at-risk	Significantly differed from expert and non-expert
Self-Reflection: Attribution	Chi-square: Significant differences across groups	Chi-square: Significant differences across groups	Chi-square: Significant differences across groups
Self-Reflection: Self-satisfaction	Significantly differed from non-expert and at-risk	Significantly differed from expert and at-risk	Significantly differed from expert and non-expert
Self-Reflection: Adaptive/defensive	Chi-square: Significant differences across groups	Chi-square: Significant differences across groups	Chi-square: Significant differences across groups

Correlation and regression analysis. Zimmerman's model of academic self-regulation suggests there relationships exist among the three phases: forethought, performance, and self-reflective. To test this, correlations were calculated among the subprocesses, and significant correlations were found among some variables as indicated in Table 9 below. As can be noted, significant correlations were found in strategic planning and four of the performance phase measures. In addition, as noted earlier the strategic planning measure is the one measure in the forethought phase where expertise level differences were also significant. Students in the expert and non-expert groups were both significantly different from the at-risk groups (See Table 8). The strategic planning measure's correlation with the performance measures suggests that the relation between the number of strategies that students will report that they will use to study for an exam and the number of strategies they use while they are studying are related. In addition, the number of strategic plans is also related to students' ability to be metacognitively aware of the responses to both achievement measures and to their metacognition on their overall test score.

Significant correlations were also found among each of the performance phase measures and two of the subprocesses in the self-reflection phase: self-evaluative standards and students' self-satisfaction as indicated in Table 9 below. Particularly interesting are correlations between students' metacognitive monitoring of their score on the Tornado Knowledge Test and their self-evaluative standards and their self-satisfaction reactions. These data suggest a strong relation between students' use of these two self-reflection processes and their science learning.

Table 9. Correlations among the three phases of academic self-regulation.

Forethought: Goal setting	Performance: Task-strategy: reading	Performance: Task-strategy: studying	Performance: Metacognitive monitoring: TKT Responses	Performance: Metacognitive monitoring: CMT Response	Performance: Metacognitive monitoring: TKT Score	Performance: Metacognitive monitoring: CMT Score
Forethought: Goal setting	-.04	-.21	-.08	.01	.01	-.16
Forethought: Strategic planning	.079	.31*	.35*	.28*	.30*	.21
Forethought: Self-efficacy for learning about tornados	-.01	-.16	-.06	-.11	.10	-.06
Forethought: Self-efficacy for learning new topics in science	-.25	-.24	-.13	-.05	-.03	-.03
Forethought: Self-efficacy for performance	-.08	.05	.13	.21	.11	-.03
Forethought: Outcome expectations: tornados	-.27	-.12	-.08	-.06	-.12	-.14
Forethought: Outcome expectations: science	-.09	.00	-.06	.03	-.08	.03
Forethought: Intrinsic interest: tornados	.09	-.08	.08	-.06	.03	-.30*
Forethought: Intrinsic interest: science	-.09	-.14	-.01	.02	-.01	-.21
Self-Reflection: Self-evaluative standards	.36**	.28*	.66**	.77**	.62**	.58**
Self-Reflection: Attribution	.24	.34*	.26	.16	.24	.13
Self-Reflection: Self-satisfaction	.30*	.29*	.52**	.61**	.49**	.59**
Self-Reflection: Adaptive/ defensive	.17	.19	.08	.01	.25	.08

* significant at the .05 level
** significant at the .01 level

Next, to test the overall effectiveness of the model, a regression analysis was conducted. As listed in Table 4 above, there were nineteen microanalytic measures used to assess academic self-regulation in this investigation. Two of these nineteen measures (attribution and adaptive/defensive) were categorical and needed to be transformed into metric measures. The attribution measure was originally coded based on four possible categories as described above. To transform this measure into a metric one, responses were coded based on the presence or absence of a particular attribution. For example, students who indicated they *do not know why* they did not do better on the conceptual model question were coded as 1 on the *do not know why* attribution scale whereas students who indicated ability, effort, or strategy were coded as 0 on the *do not know why* attribution scale. Students who indicated *ability* were coded as 1 on the *ability* attribution scale and students who indicated do not know why, effort, or strategy were coded as 0 on the *ability* attribution scale. Students who indicated an *effort* attribution were coded as 1 on the *effort* scale whereas students who indicated do not know why, ability, or strategy were coded as a 0 on the *effort* attribution scale. Students who indicated a *strategy* attribution were coded as 1 on the *strategy* attribution scale whereas students who indicated do not know why, effort, or ability were coded as 0 on the *strategy* attribution scale. These changes resulted in one microanalytic measure becoming four measures in the regression, thus bringing the total number of measures from nineteen to twenty-one.

The adaptive/defensive measure was also transformed into a metric scale based on the presence or absence of a response**.** For example, students who indicated they *no, I would not do anything differently* when studying for the conceptual model question were coded as 1 on the *no, I would not do anything differently* adaptation scale whereas students who indicated ability, effort, or strategy were coded as 0 on the *no, I would not do anything differently* adaptation scale. Students who indicated *ability* were coded as 1 on the *ability* adaptation scale and students who indicated do not know why, effort, or strategy were coded as 0 on the *ability* adaptation scale. Students who indicated an *effort* adaptation were coded as 1 on the *effort* adaptation scale whereas students who indicated do not know why, ability, or strategy were coded as 0 on the *effort* adaptation scale. Students who indicated a *strategy* adaptation were coded as 1 on the *strategy* adaptation scale whereas students who indicated do not know why, effort, or ability were coded as 0 on the *strategy* adaptation scale. These transformations resulted in 4 metric scale measures of attributions and 4 metric scale measures of adaptation, bringing the total to twenty-five self-regulatory microanalytic measures. The tornado conceptual model measures were included as self-regulatory process measures because the experimenter provided corrective feedback regarding these problem solving responses. No feedback regarding the results of the Tornado Knowledge Test was provided during learning, and thus, this measure can be used as an outcome measure.

These microanalytic measures were entered into a regression formula to predict the students' scores on the Tornado Knowledge Test. The following twenty-five measures were included in the analysis: self-efficacy for learning the tornado passage; self-efficacy for learning new topics in science; self-efficacy for performance, intrinsic interest for learning about tornados; intrinsic interest for learning new topics in science; outcome expectations about tornados; outcome expectations about new topics in science; goal setting for the Tornado Knowledge Test; strategic planning; task strategy for reading; task-strategy for studying; metacognitive measure for the answers on the Tornado Knowledge Test; metacognitive measure for the score on the Tornado Knowledge Test; metacognitive measure for the answer to the Conceptual Model Test; metacognitive measure for the score on the Conceptual Model Test; self-satisfaction; the four measures of attribution; self-satisfaction; and the four measures of adaptation. Self-regulatory processes explained a significant proportion of the variance in the Tornado Knowledge Test score, F (24) = 3.47, $p < .01$, $R^2 = .77$. This result reveals that 77% of the variance in students' on this test was predicted by their microanalytic processes.

Supplementary calibration bias analyses. The Forethought phase motivational measures were not predictive of tornado learning outcomes, which failed to support expectations. These motivational measures were given before the students had an opportunity to learn from personal feedback about tornados, and it was possible that some students overestimated their motivation, especially if they lacked expertise in science. To test this emergent hypothesis, I conducted a calibration analysis designed to assess students' bias in reporting their self-efficacy scores. This analysis was only possible with self-efficacy measures of motivation in the present study because they involved test item-specific judgments. Bias scores measure over- or underconfidence regarding knowing the answer to a test item.

Two measures of self-efficacy bias were obtained to determine whether students over- or underestimated their competence in science. The first measure was *self-efficacy for learning and remembering the material in the tornado passage bias.* This measure was calculated by subtracting students' score on the *Tornado Knowledge Test* from their self-efficacy for learning and remembering the material in the tornado passage score. The Means and Standard deviations for these measures of self-efficacy bias are presented in Table 10. A two-way analysis of variance revealed a statistically significant main effect for expertise on self-efficacy for learning bias, F (2, 45) = 9.25, $p < .01$. The effect size for expertise level (partial eta squared) was .30 and is considered large. A trend analysis revealed a significant negative linear trend for expertise, F (1, 48) = 26.28, $p < .01$, on self-efficacy for learning bias. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between expert and the other two groups. The non-expert group differed marginally from the at-risk group ($p < .06$). The mean and standard deviation scores for the expert group were M = -2.96, SD = 4.50, while the mean and standard deviation scores for the non-expert group were M = 10.31, SD = 4.29 and the mean and deviation scores for the at-risk group were M = 23.76, SD = 4.30. There was no main effect for gender or interaction between expertise level and gender on the self-efficacy for learning bias measure.

The second measure of self-efficacy bias obtained was *self-efficacy for performance bias.* This measure was obtained by subtracting students' score on the *Tornado Knowledge Test* from their self-efficacy for performance score**.** A two-way analysis of variance revealed a statistically significant main effect for expertise and self-efficacy for performance bias, F (2, 45) = 7.29, $p < .01$. The effect size for expertise level (partial eta squared) was .23 and is considered large. A trend analysis revealed a significant negative linear trend for expertise, F (1, 48) = 17.99, $p < .01$, on self-efficacy for performance bias. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between expert and the other two groups. The non-expert group did not differ from the at-risk group. The mean and standard deviation scores for the expert group were M = -1.81, SD = 5.58, while the mean and standard deviation scores for the non-expert group were M = 17.96, SD = 5.34 and the mean and standard deviation scores for the at-risk group were M = 27.12, SD = 5.34. There was no main effect for gender and interaction between expertise level and gender on the self-efficacy for performance bias measure.

Table 10. Means and standard deviations for self-efficacy calibration measures by expertise level

	Expert		Non-expert		At-risk	
	Means	SD	Means	SD	Means	SD
Self-efficacy for learning bias	-2.96	4.5	10.31	4.29	23.76	4.30
Self-efficacy for performance bias	-1.81	5.58	17.96	5.34	27.12	5.34

These results reveal that on both measures of self-efficacy (i.e., for learning and performance), at-risk students overestimated their competence due to their lack of expertise in science the most, the non-expert group showed lower levels of overestimation, and the expert group showed a slight underestimate of self-efficacy. In future research using designs that involve multiple feedback cycles, it will be interesting to see whether forethought measures will become less biased by students who are lower in expertise in science.

Study Time

Students' *study time* was assessed to determine if there were differences due to expertise or gender. The time gap between when students began to read and the time in which students indicated they were ready to take the test were recorded. The difference between the two was termed a study time measure. A two-way analysis of variance revealed a statistically significant main effect for expertise on the study time measure, $F(2, 45) = 4.35$, $p < .02$. The effect size for expertise level (partial eta squared) .16, is considered large. A trend analysis revealed a significant linear increase, $F(1, 48) = 6.57$, $p < .01$, for expertise level. Post hoc comparisons using the Tukey tests indicated significant pairwise differences between expert and at-risk groups, and marginal differences between non-expert and at-risk students. The mean and standard deviation scores for the expert group were $M = 34.87$, $SD = 3.73$, while the mean and standard deviation scores for the non-expert group were $M = 34.53$, $SD = 3.56$ and the mean and standard deviation scores for the at-risk group were $M = 21.74$, $SD = 3.56$. There was no main effect for gender or interaction between expertise level and gender for the study time measure**.** Pearson correlational analyses revealed that the length of students' study time correlated positively with their number of self-regulated learning processes, such as task strategies for reading ($r = .47$, $p < .01$) and for studying ($r = .32$, $p < .05$) and their level of self-evaluative standards ($r = .30$, $p < .05$).

Chapter 5

Discussion

Expertise Differences in Achievement

The first research question asks whether or not there are differences in achievement and in the self-regulatory behavior of students across expertise levels. With regard to achievement, significant differences were found among the three expertise levels with the experts performing better than the non-expert group, which performed better than the at-risk group on both the Tornado Knowledge Test and the Conceptual Model Test. The latter test is particularly important because it involved having to abstract from the reading a three-phase diagram of tornado development. Greene and Azevedo (2007) found that students who were more self-regulatory were better able to abstract a model of the blood circulatory system than those who were less self-regulatory. The present study reveals similar findings. It should be noted that the visual diagrams of the blood system were presented during instruction in Greene and Azevedo's study whereas only written descriptions of the conceptual models were presented in the present study. Thus, students in the present study were able to generalize from that linguistic knowledge to create their own mental images.

Expertise Differences in Person Measures of Self-Regulation

The second research question asks whether there are differences with regard to expertise level on the use of the two pre-established measures of self-regulation, the RSSRL and the SELF. The RSSRL findings revealed significant differences between all three expertise groups suggesting that teachers who have observed the students over an academic year, are able to determine which students engaged in self-regulation and which did not. This has important implications for teachers. If teachers have information about students who may not be engaging in academic self-regulation, they may be able to address these issues during the academic year to help the students achieve to the level of the expert performers.

With regard to the SELF, the ANOVA revealed marginally significant differences in self-efficacy for self-regulated learning across expertise level. Post hoc tests found significant differences among the two extreme expertise groups: experts and at-risk students. This suggests that the self-efficacy for students who were in the expert group were different from the self-efficacy for students who were in the at-risk group. However, the mean score for students in the at-risk group was *higher* than that of the expert students - suggesting that students who are at-risk may be overly confident in their ability to self-regulate in science. This finding is consistent with the self-efficacy bias findings described below.

Expertise Differences in the Three Phases of Academic Self-Regulation

Forethought phase. The third research question asks whether there are significant differences between expertise level and gender on the microanalytic measures of self-regulation. Beginning with the forethought phase, the microanalytic task analysis measure of goal setting revealed no significant differences across expertise level. However, with regard to the forethought phase task analysis measure of strategic planning, both the experts and non-experts differed significantly from the students in the at-risk group. This finding indicates that students who are at-risk have fewer strategies in mind as they prepare to study an academic task. One important implication from this finding is that teachers may want to instruct students who are at-risk about how to plan for academic tasks, and this may involve providing them with an inventory of strategies that can be used.

Two measures of students' self-motivational beliefs, namely self-efficacy and outcome expectations did not differ significantly across expertise groups. The self-motivational beliefs of intrinsic interest for learning about tornados produced a marginally significant difference between non-experts and the other two groups, although the trend analysis was not significant. This suggests that expertise differences in intrinsic motivation to study tornados were weak.

One possible reason for the lack of differences in pre-existing motives regarding tornados is their lack of relevance in a part of the country where few tornados occur. Even experts or non-experts in science may not be particularly motivated to learn a topic that is irrelevant to their lives. Furthermore, at-risk students may not be able to judge their self-efficacy or outcome expectations accurately. The calibration results regarding students' self-efficacy beliefs provide some support for this hypothesis. Unfortunately, time constraints prevented the tracking of these measures during additional cyclical efforts to learn. Once the students had an opportunity to try to learn information about tornados, their self-efficacy beliefs about their competence and their outcome expectations can be expected to improve their predictive power. They may also form a more reliable sense of intrinsic interest in the topic of tornados as well.

Performance phase. The results from measures of performance phase processes were much more supportive of Zimmerman's theory than forethought phase measures. With regard to self-control strategies, experts differed significantly from the at-risk students in their use of the strategies of reading and studying. These findings reveal that students who are expert self-regulators used more strategies while reading a science passage and while studying for a science test when compared to students who are at-risk. An important educational implication from these findings is that science teachers may work with students to help them identify and use strategies when reading and studying for exams.

Regarding all four microanalytic self-control measures of metacognitive monitoring, expert science students were significantly higher in the quality of their monitoring than non-expert or at-risk students. This suggests that students who are expert self-regulators may be more aware of their self-regulatory behavior and academic performance than students in the other two groups. Because metacognition involves being aware of one's thinking and learning, teachers may work with students who are non-expert or at-risk to help them become more aware of what they are reading and what they do and do not know before taking an exam, thus resulting in more strategic self-regulatory behavior.

Self-reflection phase. Significant science expertise differences were found regarding the processes of self-evaluative standards, attributions, self-satisfaction, and adaptive/defensive responses. Interestingly, females were significantly different in their satisfaction levels suggesting that there may be differences in what academic grades satisfy boys versus girls. Overall, students who are experts held higher standards and attributed their performance to self-regulatory strategies are more satisfied with their performance measures and are more adaptive than defensive when compared to students in the other two expertise level groups. Findings such as these imply that working with students to help them form more accurate self-judgments and to react by examining the self-regulatory strategies used rather than by attributing success or failures to ability or to effort may ultimately result in higher achievement.

Gender Differences in Self-Regulated Learning

Gender differences were found in achievement on the Conceptual Model Test suggesting that females tended to abstract better than males. Significant gender differences were also found on the microanalytic measure of self-satisfaction suggesting that females tended to be more satisfied with their academic grades than males. There were no gender differences on any of the other microanalytic measures, the Tornado Knowledge Test, RSSRL, and the SELF. In addition, the lack of gender differences across most of the self-reflection phase measures suggests that self-regulatory performance activities, rather than gender, are the target of attributions for academic achievement.

Correlational Relations among Processes Assessed during the Three Phases of Self-Regulation

The strategic planning in the forethought phase was significantly correlated with several of the measures in the performance phase suggesting that there is a relation between students who plan in advance of studying for an academic task and their actual studying and performance on the task. This implies that it is important for students to think in advance of studying. Students should develop a course of action before they actually begin studying because their performance may be affected by this very first step. An unexpected finding was that intrinsic interest correlated negatively with students' awareness of their score on the Conceptual Model Test. This indicates that the more interested students were in the topic of tornados, the less aware they were of what their score was on this achievement measure. It is unclear why this negative relationship was obtained.

Several measures in the performance phase also correlated significantly with measures from the self-reflection phase. The performance phase measure of task strategy for reading correlated significantly with self-evaluative standards and self-satisfaction standards. The performance phase measure of task strategy for studying also correlated significantly with these two self-reflective measures, in addition to the attribution measure. These two performance measures were based on a count of the number of strategies that they used. The correlations between these two task strategies and the self-reflective measures suggests that the number of strategies described is related to the ways in which students evaluate their performance, what they attribute their performance to be, and how satisfied they are with their performance. Implications of these findings are that it is important for educators to help students who are engaging in academic tasks to set appropriate standard levels and to examine the degree of satisfaction they may experience with various grades.

The metacognitive measures correlated significantly with the self-evaluation measure suggesting that students who are in the three expertise groups are very aware of their performance on achievement assessments. This has interesting implications for students in that it suggests that the standards students set for themselves are in synch with their thoughts about their performance on academic tasks. Self-satisfaction correlations with performance measures were also significant suggesting that students who set a level of satisfaction for themselves are also metacognitively aware of their performance on achievement measures. It is also provides support for Zimmerman's model indicating that the subprocesses in the performance phase are directly linked to subprocesses in the self-reflection phase.

Surprisingly, the attribution and adaptive/defensive measures did not reveal significant correlations with most of the measures in the performance phase. This may be attributed to the open-endedness of these two measures. Students may not have had experience in reflecting upon their attributions and how they would adapt or change based on performance on achievement tests, and when asked an open-ended question, they were required to think and explain (react) on-the-spot about their performance and about whether or not they would change their behavior if given another opportunity (adapt or not adapt). Perhaps if the cycle were repeated, the students would have been familiar with thinking about their attributions and whether they would adapt the next time given a similar academic task.

In sum, as described in Table 8 above, there were several significant differences across expertise levels on measures in each of the three phases of academic self-regulation with students in the expertise group differing significantly from students who are at-risk. This has important implications for educators who are working with students who are varying in levels of achievement. Perhaps if teachers were to work with students on each of the subprocesses in the three phases, students who are at-risk may be provided with the coaching and training needed to achieve at the level of experts.

Overall Predictive Power of Microanalytic Measures

The regression analysis was conducted to determine how well the microanalytic measures of self-regulation would collectively predict achievement. A significant and substantial amount of variance (77%) was found, supporting the hypothesis that the measures included in the cyclical model do in fact predict students' science achievement. As indicated earlier, the microanalytic measures involved very specific questions assessing psychological processes before, during, and after learning. The educational implications of this finding are that students who have a history of performing well engage in self-regulatory processes when preparing for a typical academic task, such as a content matter test. The differences found between the expertise group and the other two groups suggest that teachers may use Zimmerman's model as a tool for promoting the academic growth and success of students who have a history of not reaching their potential.

Calibration Bias Analyses

The calibration analyses indicate that students who are experts tend to underestimate their capability to perform a task while students who are at-risk tend to overestimate it. These expertise differences in calibration on performance phase measures indicate that the students were differentially monitoring their effectiveness. The adverse effect of overestimation of one's competence during science learning is that at-risk students may not study to a sufficient level to do well on the exam. By contrast, expert science students will over-study slightly because of their underestimates of their effectiveness in studying.

Study Time

One interesting finding was the significant differences found across expertise level with regard to study time. Students in the expert and non-expert groups spent significantly more time preparing for the exam (mean score of 35 minutes for expert and non-expert groups) when compared to students in the at-risk groups (mean score of 22 minutes). One possible implication from this finding is that the amount of time that students use to prepare for examinations, may in fact, be the result of their use of more self-regulatory strategies while studying, which ultimately impacts achievement. There is some support for this interpretation in the significant correlations between the length of students' study time and their level of self-evaluative standards and with their number of task strategies for reading and studying.

Recommendations for Further Research

Students in the current sample were drawn from three parochial high schools and may differ from students attending a public high school. Each of these parochial schools requires students to apply for admission based on an entrance examination and junior high school grades. In addition, parochial schools may promote conformity to authority by requiring students to wear uniforms, to conduct themselves as young adults at all times, and to behave in line with the religious beliefs of the school. These factors may have resulted in a sample of students who tended to perform higher academically than students in typical public schools. Conducting this study within regular public high schools could involve students at even greater levels of risk. Thus, it is possible that expertise differences in self-regulation in these schools may be more pronounced and may be more generalizable to 11^{th} grade high school students as a population group.

As suggested earlier, the current study did not provide the students with a second or third follow-up academic task, which would test whether students' self-reflective processes provided the necessary feedback to change or impact the forethought processes. Feedback is an important element in academic self-regulation, and the present investigation did not test for this beyond the third phase of the first cycle, thus information about full impact of the feedback loop was not examined.

Although these microanalytic measures provided a reliable and valid assessment of the processes used in self-regulation, the subject matter content may have had limited interest to students living in a geographic region where tornados do not regularly occur. One recommendation is to use an academic task that is more directly relevant to what students may experience in their current living environment.

A recommendation for further research is to test Zimmerman's three phase model using an academic task in an authentic classroom setting. If the researcher could collaborate with the teachers to use an academic task that is part of the students' records, the students may be more engaged and more challenged. This may result in differences across expertise levels in the forethought phase processes, which were expected but not found in the present investigation.

Appendix A: Tornado Passage

Name____________________________________ Date____________

Tornados

Thunderstorms generate a number of weather hazards, including lighting strikes, damaging winds, hailstorms, heavy rain, flash floods, and tornados. Tornados are bodies of air that have a funnel shape, usually rotate counterclockwise (viewed from above), have very low interior pressure, and have very high velocities in their "walls." Hail, severe thunderstorms and wind usually accompany tornados. People have reported that a tornado sounds like a roaring train approaching from a distance.

Tornados develop from thunderstorms in three phases. Phase one occurs just prior to a thunderstorm developing. During this phase, cold polar air collides with warm moist tropical air and this initiates horizontal rolling near the ground. During the second phase, the spinning roll of air "lifts" at one end, tilting it from horizontal to vertical. In the third phase, the rising column of air begins to open into a funnel shape as it rises towards the cloud. The tornado generated within a severe thunderstorm extends down from the parent cloud, sweeping along the surface as the cloud travels above.

There are a number of important characteristics of tornados such as air pressure and wind-speed, diameter, path and predictability, and ground speed which make them potentially very dangerous. For example, with regard to air pressure, tornados are serious local hazards because of their low air pressure and high swirling winds. Air pressure can drop up to 3 inches of mercury inside a tornado's funnel, which is very low. The wind speed of a tornado is usually less than 140 miles per hour but can be as fast as 350 miles per hour. A tornado's low air pressure and fast rotary winds typically destroy whatever they touch. A tornado can range from 30 feet to almost one mile in diameter. The average destructive path is about 4 miles long. Tornados generally travel in a southwest-to-northeast direction, but their direction can be erratic and may change abruptly. Most tornados move at an average speed of about 30 miles per hour, although speeds varying from stationary to over 68 miles per hour have been reported. A fast moving tornado can be very dangerous because it gives little time for people to prepare or to seek safety. Tornados are much less predictable than hurricanes and northeasters, striking with little warning, which is why they are so terrifying.

Each year tornados kill about a hundred people and cost hundreds of million dollars of damage. When compared to other weather hazards like hurricanes and northeasters, tornados damage the smallest area; but, the damage done is usually very severe. Tornados cause damage in multiple ways. The rapid, rotating winds knock down weaker structures. Extremely low pressure inside a tornado creates such great differences in pressure between the inside and outside of structures that roofs are lifted and removed by the resulting high winds. The wind force and high pressure pick up smaller objects (cars, small structures, animals, people) and can transport them up to hundreds of meters. Another way in which tornados cause damage is through the tremendous amounts of debris that are mobilized in a tornado. This debris becomes shrapnel that destroys other structures. It is important to remember that in the high winds of a tornado, normally innocuous debris such as cardboard, asphalt roofing tiles, sticks, and fine gravel can become lethal objects.

Survival in a tornado involves sheltering oneself from the high winds and flying debris. One way you can do this is by staying away from windows, doors and outside walls. In homes and small buildings, you should go to a basement or reinforced area in the interior of the lowest floor. In addition, you should get under something sturdy or lie in the bathtub and cover yourself with a thick blanket. In schools, nursing homes, shopping centers and malls, go to pre-designated shelter areas which are usually in the hallways or the lowest levels. If you are caught in the open, the best advice is to lie below the ground surface in a drainage ditch or at least flat on the ground, and again try to cover your body and always remember to protect your head.

Tornado frequency, state by state, is quite variable; according to the National Weather Service, only Alaska does not experience tornados. The south-central United States/Gulf Coast region has many tornados. Northern Texas, Oklahoma, Kansas, and southern Nebraska have the greatest frequency.

How are the wind speed and intensity of tornados measured? The U.S. Weather Service issues a tornado watch when the weather conditions are favorable for the development of severe thunderstorms that are capable of producing tornadoes. A tornado warning is issued when a tornado has been spotted and may be approaching. Doppler radar is a tracking system that determines the speed of wind within a storm. It is used with to help identify distinctive "hook-shaped" areas within a particular part of a thunderstorm line that are likely to form a tornado.

The Fujita scale is used to classify a tornado based on the potential intensity of the damage, ranging from F-0 to F-5 as outlined in the table on page 4. An F-0 tornado refers to "light" with wind speeds of less than 72 miles per hour and damage is likely to be to chimneys, TV antennas, roof shingles, trees, signs, and windows. An F-5 tornado refers to "incredible" damage with wind speeds of more than 260 miles per hour and damage is likely to result in no structures left standing. When determining the intensity of a tornado, one must add the rotational wind speed and the forward motion of the tornado. For example, if the rotational wind speed of a tornado is 70 miles per hour, and the forward motion of the tornado is 30 miles per hour, the total wind speed would be 70 + 30 or 100 miles per hour. On the Fujita intensity scale on the following page, this places the tornado in an F-1 category.

Science has enabled people to predict the possibility, path of potential destructions of tornados, which can be the difference between life and death.

Table: Fujita intensity scale for tornados.

Fujita Category	Wind speed in miles per hour
F-0	Less than 72
F-1	72-112
F-2	113-157
F-3	158-206
F-4	207-260
F-5	More than 260

Passage taken from the following sources with modifications made:

Coch, N. K. (1995). *Geohazards, natural and human*. Englewood Cliffs, NJ: Princeton-Hall, Inc.

Gibilisco, S. (2006). *Meteorology demystified*. New York, NY: McGraw Hill.

http://en.wikipedia.org/wiki/Tornado_watch

http://www.windows.ucar.edu/tour/link=/earth/Atmosphere/tornado/formation.html

Appendix B: Tornado Knowledge Test and the Conceptual Model Test

Name_____________________________________ Date____________

Please answer the following questions based on the passage to the best of your ability:

1. Define *tornados*.

2. Tornados develop from thunderstorms in three phases. Please draw a diagram and briefly describe what is happening in each of the three phases.

3. List four characteristics of tornados and explain how they contribute to making tornados potentially very dangerous.

4. List four ways in which tornados cause damage.

5. Explain how debris can cause damage.

6. Describe at least five ways in which you can protect yourself if a tornado strikes.

7. Which four states have the greatest frequency of tornados?

8. What is the difference between a *tornado watch* and a *tornado warning*?

9. Describe the *Doppler radar* and how it is used.

10. If the rotational wind speed of a tornado is 100 miles per hour and the forward motion of the tornado is 50 miles per hour, what is the total wind speed of the tornado?

11. Using the table below, what category of intensity is this tornado likely to have on the Fujita scale?

Table: Fujita intensity scale for tornados.

Fujita Category	Wind speed in miles per hour
F-0	Less than 72
F-1	72-112
F-2	113-157
F-3	158-206
F-4	207-260
F-5	More than 260

Appendix C: Rating Student Self-Regulated Learning Outcomes: A Teacher Scale

<u>Rating Student Self-regulated Learning: A Teacher Scale</u>

Teacher:____________________________________ **Class: __________________**

Student: ____________________________________ **Date: __________________**

Choose the appropriate scale value and mark the number in the space before each item:

Never Seldom Often Usually Always
1……………...2…………….3………………4……………..5

____ 1. Did this student solicit additional information about the exact nature of forthcoming tests?

____ 2. Did this student solicit additional information about your expectations or preferences concerning homework assignments?

____ 3. Did this student display awareness concerning how well he/she had done on a test before you had graded it?

____ 4. Did this student complete assignments on or before the specified deadline?

____ 5. Did this student come prepared to participate in class on a daily basis?

____ 6. Did this student spontaneously express interest in course matter?

____ 7. Did this student offer relevant information that was <u>not</u> mentioned in the textbook or previous class discussions?

____ 8. Would this student seek assistance from you on his/her own when he/she was having difficulty understanding schoolwork?

____ 9. Did this student ask unusual or insightful questions in class?

____ 10. Did this student volunteer for special tasks, duties, or activities related to schoolwork?

____ 11. Did this student express and defend opinions that may have differed from yours or those of classmates?

____ 12. Did this student solicit further information regarding your grades or evaluation of his or her schoolwork?

Thank you!

Appendix D: Self-Efficacy for Learning Form-A

Name:___________________________ID#:_________________ Date:____________

Definitely ***cannot do it***			**Probably** ***cannot***		**Maybe**		**Probably** ***can***			**Definitely** ***can do it***
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

Choose a percentage to indicate your answer:

_____1. When you miss a science class, can you find another student who can explain the science lecture notes as clearly as your teacher did?

_____2. When your teacher's science lecture is very complex, can you write an effective summary of your original science notes before the next class?

_____3. When a science lecture is especially boring, can you motivate yourself to keep good science notes?

_____4. When you had trouble understanding your instructor's science lecture, can you clarify the confusion before the next science class by comparing notes with a classmate?

_____5. When you have trouble studying your science class notes because they are incomplete or confusing, can you revise and rewrite them clearly after every science lecture?

_____6. When you are taking a science course covering a huge amount of material, can you condense your science notes down to just the essential facts?

_____7. When you are trying to understand a new science topic, can you associate new science concepts with old ones sufficiently well to remember them?

_____8. When another student asks you to study together for a science course in which you are experiencing difficulty, can you be an effective study partner?

_____9. When problems with friends and peers conflict with schoolwork, can you keep up with your science assignments?

_____10.When you feel moody or restless during studying science, can you focus your attention well enough to finish your assigned science work?

_____11. When you find yourself getting increasingly behind in a science course, can you increase your study time sufficiently to catch up?

_____12. When you discover that your science homework assignments for the semester are much longer than expected, can you change your other priorities to have enough time for studying science?

_____13. When you have trouble recalling a science abstract concept, can you think of a good example that will help you remember it on the test?

_____14. When you have to take a science test in an area of science that you dislike, can you find a way to motivate yourself to earn a good grade?

_____15. When you are feeling depressed about a forthcoming science test, can you find a way to motivate yourself to do well?

_____16. When your last science test results were poor, can you figure out potential questions before the next test that will improve your score greatly?

_____17. When you are struggling to remember the technical details of a concept for a science test, can you find a way to associate them together that will ensure recall?

_____18. When you think you did poorly on a science test you just finished, can you go back to your science notes and locate all the information you had forgotten?

_____19. When you find that you had to “cram” at the last minute for a science test, can you begin your test preparation much earlier so you won’t need to cram next time?

Thank you!

References

Artlet, C., Baumert, J., Julius-McElvany N., & Peschar, J. (2003). Learners for life. Student approaches to learning, results from PISA 2000. Organisation for Economic Cooperation and Development, Paris (France). 133 pp.

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

Bandura, A. (1997). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp.1-45). New York, NY: Cambridge University Press.

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

Baldi, S., Jin, Y., Green, P., & Herget, D. (2007). Highlights from PISA 2006: Performance of U.S. 15 year old students in science and mathematics literacy in an international context. *National Center for Education Statistics.* 74 pp.

Blickenstaff, J.C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*, 369-386.

Butterworth, B. (2006). Mathematical expertise. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 553-568). New York, NY: Cambridge University Press.

Cleary, T.J., & Zimmerman, B.J. (2001). Self-regulation differences during athletic practice by experts, non-experts, and novices. *Journal of Applied Sport Psychology, 13,* 185-206.

Coch, N. K. (1995). *Geohazards, natural and human.* Englewood Cliffs, NJ: Princeton-Hall, Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^{nd} ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ericsson, A.K. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.223-242). New York, NY: Cambridge University Press.

Ericsson, A.K. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.683-704). New York, NY: Cambridge University Press.

Ericsson, A.K., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747.

Ericsson, A.K., Charness, N., Feltovich, P.J., & Hoffman, R.R. (2006). *The Cambridge handbook of expertise and expert performance.* New York, NY: Cambridge University Press.

Ericsson, A.K., Krampe, R.T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Gibilisco, S. (2006). *Meteorology demystified.* New York, NY: McGraw Hill.

Gobet, F., & Charness, N. (2006). Expertise in chess. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.523-538). New York, NY: Cambridge University Press.

Gonzalez, P., Williams, T., Jocelyn, Roey, S., Kastberg, D., & Brenwald, S. (2008). Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth and eighth grade students in an international context. *National Center for Education Statistics,* 112 pp.

Greene, J.A., & Azevedo, R. (2007). Adolescents' use of self-regulatory processes and their relation to qualitative mental model shifts while using hypermedia. *Journal of Educational Computing Research, 36*(2), 125-148.

Greene, J.A., Moos, D.C., Azevedo, R., & Winters, F.I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers & Education, 50*(3), 1069-1083.

Grigg, W.S., Lauko, M.A., & Brockway, D.M. (2006). The nation's report card: Science 2005. Retrieved October 1, 2007 from http://nces.ed.gov/nationsreportcard/pubs/main2005/2006466.asp

Hackett, G. (1997). Self-efficacy in career choice and development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp.232-258). New York, NY: Cambridge University Press.

Hanson, S.L. (1996). *Lost talent: Women in the sciences*. Philadelphia, PA: Temple University Press.

Kitsantas, A., & Zimmerman, B. J. (2002). Comparing self-regulatory processes among novice, non-expert, and expert volleyball players: A microanalytic study. *Journal of Applied Sport Psychology, 14,* 91-105.

Lehman, A.C., & Gruber, H. (2006). Music. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.457-470). New York, NY: Cambridge University Press.

Meece, J.L., & Courtney, D.P. (1992). Gender differences in students; perceptions: Consequences for achievement-related choices. In D. J. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 209-228). Hillsdale, NJ: Erlbaum.

National Science Board. (1998). *Science and engineering indicators-1998*. (NSB 98-1). Arlington, VA: National Science Foundation.

Noice, H., & Noice, T. (2006). Artistic performance: Acting, ballet, and contemporary dance. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.489-504). New York, NY: Cambridge University Press.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

Pant, E.A., Ericsson, K.A., Hill, L., & Asberg, K. Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology, 30*(1), 96-116.

Schunk, D.H. (2001). Social cognitive theory and self-regulated learning. In B.J. Zimmerman & D.H. Schunk (Eds.) *Self-regulated learning and academic* *achievement: Theoretical perspectives.* (pp.125-152). Mahwah, JN: Lawrence Erlbaum Associates, Inc.

Schunk, D.H., & Zimmerman, B.J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32*(4), 195-208.

Spear, M.G. (1987). The biasing influence of pupil sex in a science marking exercise, In: A. Kelly (Ed.) *Science for girls?* Milton Keynes, Open University Press, (pp.46-51).

Timms, M., Schneider, St., Lee, C., & Rolfhus, E. (2007). *Aligning science assessment standards: Texas and the 2009 National Assessment of Educational Progress (NAEP). Issues & Answers.* Regional Educational Laboratory Southwest (NJ1).

Tornado Watch. Retrieved on June 1, 2007 from http://en.wikipedia.org/wiki/Tornado_watch

University Corporation for Atmospheric Research (2000-2006). Windows to the universe: How a tornado forms. Retrieved on June 1, 2007 from http://www.windows.ucar.edu/tour/link=/earth/Atmosphere/tornado/formation.html

VanLeuvan, P. (2004). Young women's science/mathematics career goals from seventh grade to high school graduation. *Journal of Educational Research, 97*(5), 24-268.

Voss, J.F., & Wiley, J. (2006). Expertise in history. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.569-587). New York, NY: Cambridge University Press.

Winters, F., & Azevedo, R. (2006). High-school students' regulation of learning during

Computer-based science inquiry. *Journal of Educational Computing Research, 33*(2), 189-217.

Zimmerman, B.J. (1997). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp.202-231). New York, NY: Cambridge University Press.

Zimmerman, B.J. (1998). Academic studying and the development of personal skill: A self-regulatory perspective. *Educational Psychologist, 33,* 73-86.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Self-regulation: Theory, research, and applications* (pp.13-39). Orlando, FL: Academic Press.

Zimmerman, B.J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82-91.

Zimmerman, B. J. (2002). Achieving self-regulation: The trial and triumph of adolescence. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (vol.2, pp.1-27). Greenwich, CT: Information Age.

Zimmerman, B. J. (2006). Development and adaptation of expertise. In A. K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.705-722). New York, NY: Cambridge University Press.

Zimmerman, B. J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology, 30,* 397-417.

Zimmerman, B. J., & Kitsantas, A. (2007). Validity and reliability of Self-Efficacy for Learning Form (SELF) scores of college students. *Journal of Psychology, 215*(3). 157-163.

Zimmerman, B.J., & Martinez-Pons, M. (1986). Development of a structured interview of assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*. 614-628.

Zimmerman, B. J., & Martinez-Pons, M. (1989). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology, 80*. 284-290.

Zimmerman, B.J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51-59.

Zimmerman, B.J., & 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.