Enhancing the Calibration Accuracy of Adult Learners: A Multifaceted Intervention

Antonio P. Gutierrez
University of Nevada, Las Vegas, gutierr2@unlv.nevada.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations
Part of the Educational Psychology Commons

Repository Citation
https://digitalscholarship.unlv.edu/thesesdissertations/1568

This Dissertation is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
ENHANCING THE CALIBRATION ACCURACY OF ADULT LEARNERS: A
MULTIFACETED INTERVENTION

by

Antonio Partida Gutierrez

Bachelor of Arts
Political Science and Sociology
University of Nevada, Las Vegas
2000

Master of Science
Educational Psychology
University of Nevada, Las Vegas
2008

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy in Educational Psychology

Department of Educational Research, Cognition, & Development
College of Education
The Graduate College

University of Nevada, Las Vegas
May 2012
We recommend the dissertation prepared under our supervision by

Antonio Partida Gutierrez

entitled

Enhancing the Calibration Accuracy of Adult Learners: A Multifaceted Intervention

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Educational Psychology
Department of Educational Research, Cognition, and Development

Gregory Schraw, Ph.D., Committee Chair
Eunsook Hong, Ph.D., Committee Member
Gwen C. Marchand, Ph.D., Committee Member
Janelle M. Bailey, Ph.D., Graduate College Representative
Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2012
ABSTRACT

Enhancing the Calibration Accuracy of Adult Learners: A Multifaceted Intervention

by

Antonio P. Gutierrez

Dr. Gregory Schraw, Examination Committee Chair
Professor of Educational Research, Cognition, & Development
University of Nevada, Las Vegas

The present study employs the Nelson and Narens Model of Metacognition (NNMM) to examine the influence of metacognitive strategy training and extrinsic incentives on performance, level of confidence, and the calibration accuracy of undergraduate students’ metacognitive judgments within a pretest/posttest experimental design. Calibration of performance is crucial because it allows learners to engage in appropriate comprehension monitoring during a learning episode. As metacognition implies, those individuals who are better calibrators can more adequately adapt to the demands of the situation (monitoring), and thereby better prepare for learning episodes that are similar in format or content (control). Consequently, this aids in the improvement of performance, confidence, and calibration accuracy.

Findings suggest that strategy training and incentives enhanced performance and level confidence in performance. However, only strategy training increased the calibration accuracy of feeling-of-knowing judgments. Theoretically, both strategy training and incentives influence learners’ metacognitive monitoring and control; therefore, the results support the NNMM. Educational implications and directions for future research are discussed.
ACKNOWLEDGEMENTS

This path has most definitely been winding and full of thorns for me. But not all of it was bad. There were moments of joy and delight at professional and academic accomplishments; moments of emotional and motivational fatigue, when simply thinking about this process sapped me of energy; and moments of self-reflection and self-awareness. To this day, I still cannot believe that I will soon join the ranks of scholars and researchers in the field of educational psychology, advancing research in the area of higher-order thinking. I gladly confess that I could not have arrived to this point in my life, preparing to embark on a fantastic journey, without the aid of individuals.

Unfortunately, there is insufficient space here to lay down the names of all the individuals who have come to my rescue in my time of greatest need. Thus, I will name only those who have made the greatest contribution. I wish to thank my mother and my sister, Patty, as well as my dear friends Annette and Nick Marquart, Lori Candela, and Nancy Jo Hamilton from the bottom of my heart for their unconditional support. Also, thank you Michael Blume for helping me progress through some parts of this process. Finally, but not least, I wish to thank Gwen Marchand and Gregg Schraw for helping me develop myself as an outstanding scholar and researcher, and Janelle Bailey and Doug Lombardi for giving me very sage advice. May good fortune and happiness be with us all and follow us wherever we go!
TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... iii

ACKNOWLEDGEMENTS ................................................................................................. iv

LIST OF TABLES ............................................................................................................... viii

LIST OF FIGURES ............................................................................................................ ix

CHAPTER 1 INTRODUCTION .......................................................................................... 1
  Organization .................................................................................................................. 2
  Operational Definitions .............................................................................................. 2
  An Introduction to the Nelson and Narens Model of Metacognition ....................... 6
  Overview of Relevant Literature on Calibration ...................................................... 8
  Purpose and Overview of the Proposed Study ......................................................... 15
  Methods ...................................................................................................................... 15
    Experimental Interventions .................................................................................... 15
    Study Design .......................................................................................................... 17
    Hypotheses and Predictions .................................................................................... 18
    Data Analysis Plan ................................................................................................. 18
  Summary and Conclusion .......................................................................................... 19

CHAPTER 2 REVIEW OF THE LITERATURE .................................................................. 20
  Organization .............................................................................................................. 21
  A Taxonomy of Relevant Concepts .......................................................................... 22
  Comparison and Contrast Between SR, SRL, and Metacognition ......................... 23
    Models of SRL ......................................................................................................... 25
    Need for the Nelson and Narens Model of Metacognition .................................... 27
  Theoretical Framework ............................................................................................. 28
    The Nelson and Narens Model of Metacognition ............................................... 28
  Link Between Metacognition and Calibration .......................................................... 33
    Metacognition ......................................................................................................... 33
    Calibration .............................................................................................................. 37
  Antecedents → Calibration → Outputs ..................................................................... 38
    Antecedents/Inputs to Calibration ......................................................................... 38
    Calibration (Metacognitive) Judgments ................................................................. 44
    Outputs/Outcomes of Appropriate Calibration ..................................................... 46
  Review of Research on Improving Metacognition and Calibration ......................... 48
    Learners Are Imperfect Evaluators of their Own Internal Processes ..................... 51
    Training to Improve Metacognitive Monitoring ................................................... 54
    Review of Research on Interventions to Enhance Calibration .............................. 56
  Summary and Conclusion .......................................................................................... 67
  The Proposed Study ................................................................................................. 68
    Description of Strategy Training and Incentives .................................................... 70
    Proposed Research Question .................................................................................. 71
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>METHODOLOGY</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Participants and Setting</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Design and Materials</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Demographics</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>General Metacognitive Awareness</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Strategy Training Manipulation Fidelity Check</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Procedures</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Session 1</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Data Analysis</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>RESULTS</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Participant Attrition at Posttest</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Little’s MCAR χ² Test</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Equality of Outcome Means at Baseline</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Primary Analyses of the Present Study</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Equality Among Groups</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Descriptive Statistics and Correlations</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Strategy Training Intervention Fidelity Check</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Comparison Between Pretest and Posttest Performance</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Evaluating the Effectiveness of the Interventions</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Performance</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>104</td>
</tr>
<tr>
<td>5</td>
<td>DISCUSSION</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Research Hypotheses</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Performance Outcomes</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>Confidence Outcomes</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Calibration Outcomes</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>Theory and the Advancement of Knowledge</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>Contributions to Theory</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>Advancement of Knowledge on Calibration</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Limitations of the Present Study</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>Educational Implications</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>Directions for Future Research</td>
<td>118</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Key Terms, Definitions, and Sources</td>
</tr>
<tr>
<td>Table 2</td>
<td>Summary of Previous Relevant Research on the Enhancement of Calibration</td>
</tr>
<tr>
<td>Table 3</td>
<td>Summary of the Effects of Strategy Training and Incentives on the Control and Monitoring Processes within the Nelson and Narens Model of Metacognition</td>
</tr>
<tr>
<td>Table 4</td>
<td>Summary of Metacognitive Strategies and their Relation to Calibration and Theory</td>
</tr>
<tr>
<td>Table 5</td>
<td>Summary of Activities by Experimental Group</td>
</tr>
<tr>
<td>Table 6</td>
<td>Descriptive Statistics of Outcome Measures by Group</td>
</tr>
<tr>
<td>Table 7</td>
<td>Zero-Order Correlations for the Strategy Training and No Training Groups</td>
</tr>
<tr>
<td>Table 8</td>
<td>Zero-Order Correlations for the Incentives and No Incentives Groups</td>
</tr>
<tr>
<td>Table 9</td>
<td>Adjusted and Unadjusted Results of Posttest Performance</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. A visual representation of the Nelson and Narens Model of Metacognition...................................................................................................7

Figure 2. A taxonomy displaying the interrelations among self-regulation, self-regulated learning, metacognition, the Nelson and Narens Model of Metacognition, calibration, and calibration accuracy and bias. ...................... 23

Figure 3. The hypothesized explanatory trajectory of antecedents, calibration, and outcomes based on the two manipulations of strategy training and incentives used in the study. ................................................................. 75

Figure 4. A schematic of the proposed intervention, including the four-group breakdown with respect to the two components of the intervention, strategy training and incentives................................................................. 81
CHAPTER 1

INTRODUCTION

Students use cognitive skills and strategies as well as metacognitive knowledge and regulation to successfully prepare for future assessments of performance in the service of learning. Students are effective self-regulators of their learning when they can accurately determine what they know and do not know about domain content. This permits learners to focus attention and other cognitive resources on material they have not yet mastered and spend less time rehearsing material they already know, thus effectively demonstrating self-regulated learning (SRL) behavior. When asked how well they will perform in some future objective assessment of their knowledge, these more metacognitive learners will be able to accurately convey their actual performance because they can better calibrate confidence judgments of their knowledge of performance.

Conversely, students who are less metacognitive are not as readily able to accurately pinpoint what they know or do not know about a domain or topic, and thus, they are less capable of regulating their learning (i.e., they may be lacking in planning, evaluation, information management, or comprehension monitoring skills). This situation leads to inaccurate calibration because students are prone to too much confidence or insufficient confidence when making judgments of their performance on a criterion task; in other words, there is a larger disparity between metacognitive judgments in one’s knowledge and actual performance on a reference task, such as an exam. Consequently, these students are inefficiently allocating and utilizing cognitive resources when learning.

Research suggests that students can benefit greatly from tailored interventions targeted at increasing the accuracy of the metacognitive judgments and decreasing the
level of bias (i.e., over- or under-confidence) of their performance. The present study investigates one main aspect of calibration, namely accuracy, in an experimental framework. The main purpose was to ascertain whether a multifaceted educational intervention incorporating strategy instruction, incentives, and type of test (pretest/posttest) is successful at increasing the accuracy of undergraduate students’ metacognitive judgments regarding their knowledge. Students are predicted to improve their accuracy as a result of the proposed interventions.

Organization

In this Chapter, I first provide definitions of constructs under investigation. These operational definitions will be followed by an introduction to the Nelson and Narens Model of Metacognition (Nelson & Narens, 1990, 1994), which serves as the theoretical foundation for the present study. Next, a brief summary of the most relevant research regarding SRL, metacognition, calibration, and interventions aimed at enhancing calibration accuracy will be discussed. Finally, the chapter will end with a description of the present study, including purposes, research questions and hypotheses, methodology, and a summary and conclusion.

Operational Definitions

Table 1 includes a summary of key terms and their respective sources. Self-regulation (SR) generally refers to individuals’ ability to control, monitor, and regulate their cognition and behavior (Corno & Mandinach, 1983; Corno & Rohrkemper, 1985). In learning contexts, SR incorporates more specific functions, including: students’ metacognitive strategies for planning, monitoring, and modifying their cognition (e.g., Brown, Bransford, Campione, & Ferrara, 1983; Corno, 1986; Zimmerman & Pons, 1986,
management and control of their effort on learning tasks (Corno, 1986; Corno & Rohrkemper, 1985); and the actual cognitive strategies that students invoke in the service of learning (Corno & Mandinach, 1983; Zimmerman & Pons, 1986, 1988). While there is much overlap between SR and SRL, SRL is conceptualized as thoughts, feelings, and actions that are systematically invoked toward the fulfillment of learners’ own goals (Barak, 2010; Zimmerman & Schunk, 1989). Metacognition, in a very general sense, refers to knowledge about cognition and cognitive processes (Lin & Zabrucky 1998; McCormick 2003), which is a critical component of SR and SRL. On a finer grain, it is conceptualized as higher-order mental processes that individuals invoke during learning tasks (Dunslosky & Thiede, 1998; Flavell, 1979), as opposed to the mostly automated mental processes indicative of cognition that occur outside of awareness. The use of sophisticated problem solving strategies, the planning and allocation of resources, as well as monitoring learning and performance are all examples of metacognition.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Recent Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Regulation</td>
<td>Generally refers to individuals’ ability to control, monitor, and regulate their cognition and behavior.</td>
<td>Corno and Mandinach (1983); Corno and Rohr kemper (1985)</td>
</tr>
<tr>
<td>Self-Regulated Learning</td>
<td>The ability to control and influence one’s learning processes (e.g., planning, goal setting, strategy implementation, summarizing, and monitoring one’s progress). It is a multidimensional process which involves personal (cognitive and emotional), behavioral, and environmental components.</td>
<td>Barak (2010); Zimmerman and Schunk (1989)</td>
</tr>
<tr>
<td>Metacognition</td>
<td>Knowledge, including: declarative knowledge (i.e., knowledge about the self and strategies), procedural knowledge (i.e., how to appropriately apply strategies), and conditional knowledge (i.e., awareness of how and when to apply strategies), Regulation, including: planning (e.g., goal setting and allocating resources before a learning episode), information management strategies (i.e., skills and strategy sequences used online to process information more efficiently such as organizing, elaborating, summarizing, and selective focusing), debugging strategies (i.e., strategies used to correct comprehension and performance errors), and evaluation (i.e., analysis of performance and strategy effectiveness after a learning episode).</td>
<td>Artzt and Armour-Thomas (1992); Baker (1989); Brown (1987); Schraw and Dennison (1994); Schraw and Graham (1997); Schraw and Moshman (1995)</td>
</tr>
<tr>
<td>Calibration</td>
<td>The process of eliminating the discrepancy between perceived knowledge and actual knowledge (i.e., what one does and does not know). It has been variously described as feeling-of-knowing (FOK) judgments, ease-of-learning (EOL) judgments, and judgments of learning (JOL).</td>
<td>Nietfeld, Cao, and Osborne (2006); Nietfeld and Schraw (2002); Schraw (1995); Schraw (2009a, 2009b)</td>
</tr>
<tr>
<td>Calibration Accuracy</td>
<td>The level of correctness between metacognitive judgments (i.e., perceived knowledge) and actual knowledge, as assessed by some objective task (e.g., test/exam). In other words, the magnitude of said discrepancy.</td>
<td>Nietfeld and Schraw (2002); Schraw (2009b)</td>
</tr>
<tr>
<td>Calibration Bias</td>
<td>The direction of the discrepancy between metacognitive judgment and actual performance, which can be best understood as over- or underconfidence.</td>
<td>Nietfeld and Schraw (2002); Schraw (2009b)</td>
</tr>
</tbody>
</table>
Metacognitive judgments are central to calibration. They refer to a probabilistic judgment of one’s performance (e.g., tests/exams) before, during, or after performance (Schraw, 2009a). As such, calibration is defined as the process of eliminating the discrepancy between perceived performance and actual level of performance. The products of this calibration process are accuracy and bias with respect to the metacognitive judgments of confidence in one’s knowledge of a domain or topic and one’s performance. Accuracy is a measure of the magnitude of discrepancies between perceived versus actual knowledge whereas bias is the direction of discrepancies (Schraw, 1995). Highly accurate judgments of performance yield little or no bias; therefore, accuracy and bias are highly related constructs. When judgments are inaccurate, an individual may be biased in terms of overconfidence or underconfidence. Calibration involves selecting strategies and resources for the fulfillment of a learning task at the cognitive level and monitoring comprehension and evaluating the learning episode at the metacognitive level. These judgments can precede or follow the completion of a test item (local level) and/or the entire test (global level) (Nietfeld, Cao, & Osborne, 2006; Schraw, 2009a).

In sum, calibration of performance is both the process of making judgments about the information or knowledge one has learned about a domain, topic, or task in anticipation of performance on some task or objective assessment (e.g., an exam) and the product, or actual metacognitive judgment. Judgments may be accurate, or if they are inaccurate, they will be biased due to over- or under-confidence. Of special importance in the current study, these metacognitive judgments are heavily influenced by such
elements as strategy selection, ability to monitor progress when learning, and evaluating overall performance.

**An Introduction to the Nelson and Narens Model of Metacognition**

Within the context of calibration research, the Nelson and Narens Model of Metacognition (NNMM; Nelson & Narens, 1990, 1994) illustrates a theoretical framework that situates goal-directed action and volition (i.e., expression) as individuals use metacognitive processes to regulate their learning. Figure 1 presents the NNMM, which depicts the constant cyclical interaction between cognition and metacognition. These two interrelated components are linked together by two distinct processes governed by the actor, or more specifically the learner, namely monitoring and control. Control, or executive processes, refer to the manipulations learners impose on their environment in order to achieve some goal-directed end. In other words, the meta-level has the capacity to modify and make necessary adjustments to the object-level in order for the learner to come one step closer to a desired learning outcome. Monitoring, on the other hand, refers to the type and quality of the information received from the object-level so that the meta-level can make necessary changes. The more sophisticated the type of information and the more precise the information is from the object-level, the better the decisions and actions that the meta-level can make and take. In essence, monitoring can be conceptualized as a process by which the learner uses information at the object-level to evaluate progress toward a goal for learning at the meta-level.
Figure 1. A visual representation of the Nelson and Narens Model of Metacognition.

Note the dynamic and cyclical interaction between the object-level/environment and the meta-level (higher-order metacognitive processing), including explicit or implicit control and monitoring. These elements are central to effective self-regulated learning. The meta-level derives an imperfect model, or representation, of the object-level. Adapted from Nelson and Narens (1990).

This flow of information from the object-level to the meta-level is characterized by a representation of the meta-level of all the information available in the object-level (monitoring) in order for the meta-level to make accurate, precise decisions (control) given that this model will always, by nature, be imperfect. Presumably, accurate monitoring allows an individual to construct a mental model of his or her learning goals and plans at the meta-level, and it is a process by which learners evaluate the adequacy, precision, and accuracy of the model or their performance when striving to reach one of the goals and/or plans specified by the model.

The living brain, so far as it is to be successful and efficient as a regulator for survival must proceed, in learning, by the formation of a model (or models) of its environment. There can no longer be any question about
whether the brain models its environment: *it must* (Conant & Ashby, 1970, as quoted in Nelson & Narens, 1994, pp. 10-11)

Needless to say, the ability of our metacognitive mind to model our environment has profound implications for learning (Johnson-Laird, 1983; Rouse & Morris, 1986; Yates, 1985), and comprehension monitoring and calibration accuracy and bias in particular. This is especially significant considering that learners are perceived to be imperfect with respect to their ability to model their environments and to use models to regulate their performance (i.e., accuracy and bias). In other words, “people are construed as imperfect measuring devices of their own internal processes” (Nelson & Narens, 1994, p. 18), and thus the NNMM is an ideal framework in which to situate the present study on calibration accuracy. In general, the NNMM predicts that interventions that enhance either the control or monitoring aspects of metacognition will improve level of confidence in metacognitive judgments, calibration accuracy, and performance.

**Overview of Relevant Literature on Calibration**

Research suggests that monitoring accuracy is poor and that monitoring judgments affect strategy use. Many studies reveal poor calibration accuracy (see Brannick, Miles, & Kisamore, 2005 and Glenberg, Sanocki, Epstein, & Morris, 1987 for a review). In a series of experiments, Glenberg and associates (1987) found that poor calibration accuracy is not related to a particular type of performance test but rather it is found with inference tests, verbatim recognition tests, and idea recognition tests. Moreover, poor calibration accuracy was found when the test was provided immediately after reading as well as when the test was scheduled after a delay. Findings by Schraw et al. (1993) suggest that individuals harbor a strong response bias as they calibrate their performance; in other words, individuals were prone to report their accuracy consistently
irrespective of item difficulty and accuracy of response, indicating that negative feedback was not integrated to improve all aspects of calibration (i.e., accuracy and bias).

A second general finding is that feedback and incentives are related to students’ use of strategies, and performance on cognitive tasks (Pintrich, 2002; Schraw et al., 1993; Thomas & McDaniel, 2007; Tobias & Everson, 2002-2003; Yates, 1990). This point is relevant for the present study because incentives are an integral part of the proposed intervention. Based on the findings of previous research, it is reasonable to expect that both feedback and incentives will influence students’ calibration accuracy.

Glenberg et al. (1987) argued that the likely reason for poor calibration accuracy was that students assess familiarity with the general domain of a text instead of assessing knowledge obtained from a particular text. Research on this topic has concluded that calibration accuracy can be improved if students are provided with a pretest that furnishes self-generated feedback (Brannick, et al., 2005; Glenberg et al., 1987). However, calibration accuracy was improved only when the processes and knowledge invoked by the pretest are closely related to the processes and knowledge required on the criterion task (Glenberg et al., 1987). These findings are in stark contrast to research conducted by Schraw and associates (1993) in which results suggested that feedback had no effect on calibration accuracy of feeling of knowing judgments. On the other hand, incentives have been found to affect calibration accuracy (Hogarth, Gibbs, McKenzie, & Marquis, 1991; Schraw et al., 1993; Yates, 1990). Incentives to improve performance were found to negatively influence performance when contrasted with incentives to improve calibration accuracy, indicating that individuals are apt to rely on subjective feelings when calibrating their performance rather than more objective information (e.g., difficulty of
the test items; Hacker, Bol, & Bahbahani, 2008; Schraw et al., 1993). Therefore, interventions should be developed that seek to enhance the calibration accuracy of learners.

A review of relevant literature suggests that interventions targeted at increasing the accuracy of learners have been successful in a variety of domains. In problem solving and reasoning tasks, Kruger and Dunning (1999) found that students who received a brief instructional intervention were better equipped to judge the number of Wason selection problems they had solved correctly than those who were untrained. Similarly, Prowse-Turner and Thompson (2009) designed an experiment aimed at improving the calibration accuracy of undergraduate students’ syllogistic reasoning skills. The instructional intervention was successful in enhancing the accuracy of metacognitive judgments and was also effective in resolving students’ misunderstandings about the task. More specifically, those in the training condition were more proficient at estimating their overall performance (global level) than those in the no-training condition.

Along a similar vein, Mitchum and Kelley (2010) explored whether differences in problem solving strategies affect the ability of students to monitor their problem solving effectiveness as measured by confidence judgments. Their results demonstrated that spontaneous constructive matching in nonverbal spatial reasoning problems was related to improved confidence calibration and resolution when compared to response elimination. In the second experiment, they found that constructive matching yielded improved monitoring. The implication of this research is that instructed and spontaneous strategies are often not equivalent. In some instances, not all students are able to benefit
from an instructed strategy or benefit to the same extent as those who spontaneously construct the strategy.

Regarding reading comprehension, Glenberg and his colleagues (Glenberg & Epstein, 1985; Glenberg et al., 1987; Walczyk & Hall, 1989) conducted several experiments on enhancing the calibration accuracy of students’ text comprehension. In two separate experiments, Glenberg & Epstein (1985) found that providing training in calibration enhanced the accuracy of metacognitive judgments regarding text comprehension. In addition, they argued that improvement from initial calibration to recalibration of comprehension could be due to students’ access to their own knowledge as well as normative knowledge available to all students (e.g., students may be skilled at deducing which inference verification items are particularly challenging, and subsequently lower confidence). In a follow-up study, Glenberg et al. (1987) found that calibration accuracy may be improved if students are furnished a pretest that offers self-generated feedback. In such treatments, students are able to utilize feedback from the pretest to predict subsequent test performance with a fair degree of accuracy. Walczyk and Hall (1989) extended the previous research (Glenberg et al., 1987) by reporting that the accuracy of students’ confidence judgments can be enhanced if they read expository text that contains illustrative examples and embedded questions. They argued that examples and embedded questions furnish students with an opportunity to evaluate their own level of understanding of texts, which results in more accurate confidence judgments of comprehension than would be obtained from plain text.

Maki et al. (1990) investigated whether increased processing enhances calibration accuracy. In two experiments, they manipulated the amount of processing of text by
having students read intact text or text with deleted letters. They found that text with deleted letters yielded improved calibration accuracy. In essence, students can predict performance on text material with a relatively high degree of accuracy due to more active processing during reading afforded by text with missing letters. Henrion, Fischer, and Mullin (1993) examined the calibration accuracy of students in subjective probability distributions using three conditions: direct assessment (students directly estimated probabilities), experimenters’ decomposition (variables were determined by the experimenter and evaluated by the students), and students’ decomposition (variables and solutions were determined by the students). They found that, contrary to the divide and conquer principle, decomposition did not improve the accuracy of confidence judgments, as there were no significant differences between the experimenters’ and students’ decomposition conditions. Interestingly, students in the direct condition tended to underestimate their performance and students in the decomposition condition tended to overestimate their performance.

In an experimental study using prediction and postdiction judgments, Bol, Hacker, O’Shea, and Allen (2005) found that students in the overt calibration practice condition did not significantly improve the accuracy of their confidence judgments when compared to students assigned to the no-practice condition. However, they reported that higher achieving students were significantly more accurate, yet underconfident in their predictions, whereas lower achieving students were less accurate and overconfident. Therefore, this instructional intervention manipulating overt calibration practice had no effect on the accuracy of metacognitive judgments. However, Bol et al. did not include a pretest measure to establish group equivalence, which may have behaved as a confound
in the study, thus undermining any potential intervention effects. On a follow-up study investigating the effects of extrinsic rewards and reflection on calibration accuracy, Hacker, Bol, and Bahbahani (2008) found that high-achieving students were highly accurate in their confidence judgments and did not exhibit significant increases in their calibration accuracy. Low-achieving students, on the other hand, were less accurate in their confidence judgments and gained modest increases in their calibration accuracy due to reflection. Interestingly, students in the incentives condition demonstrated significant gains in their calibration accuracy compared to those in the no-incentives condition, although reflection on explanations for their cognitive judgments had no effect on calibration accuracy.

In two experiments, Nietfeld and Schraw (2002) tested the influence of prior knowledge and strategy training on monitoring accuracy. They found that knowledge of the domain was related positively to domain-specific performance and monitoring accuracy, and that strategy training enhances performance, confidence, and monitoring accuracy irrespective of aptitude. More specifically, strategy training enhanced performance and monitoring accuracy immediately following the test, but not following a one-week delay. Furthermore, the study lent credence to the trainable hypothesis (i.e., short-term strategy training enhances monitoring accuracy, even when such strategy training is not explicit). Along a related line, Schraw (1998) demonstrated that instructional strategies, such as promoting general metacognitive awareness and improving self-knowledge and regulatory skills, lead to enhanced metacognitive awareness, which subsequently positively informed students’ metacognitive judgments.
In an experiment manipulating cue familiarity and accessibility of FOK judgments, Koriat and Levy-Sadot (2001) found that both familiarity and accessibility improved students’ FOK judgments, but the effects of accessibility were found mostly when familiarity was high. This moderating pattern was found when FOK judgments were delayed, yet not when they were immediate. Finally, Lodewyk, Winne, and Jamieson-Noel (2009) examined the influence of task structure on calibration accuracy. They found that students tended to calibrate their knowledge better in ill-structured versus well-structured problems, presumably because students reported greater metacognitive strategy use in the ill-structured problems compared to their performance on well-structured problems. In addition, low achieving students showed poorer calibration accuracy on both tasks compared to high achieving students who demonstrated more proficient calibration accuracy.

Overall, these studies suggest that experimental interventions to improve calibration accuracy improve both performance and the monitoring accuracy of performance judgments. Nevertheless, interventions targeted at improving calibration accuracy have yielded modest/weak effect sizes as indices of the explained variance in the outcome measures that can be attributed to the experimental manipulations. For instance, Hacker et al. (2008) reported effect sizes ($\eta^2$) of 0.08 to 0.12 and Bol et al. (2005) found effect sizes ranging from 0.04 to 0.15. Moreover, Nietfeld et al. (2006) reported effect sizes ranging from 0.10 to 0.17. Hence, modest effect sizes in the present study would be in line with the body of literature on improving FOK judgments, albeit larger effect sizes may result. Moreover, although most studies did not link their findings to the NNMM, it appears that some studies improved calibration accuracy by improving
control processes (i.e., knowledge and feedback), while others improved monitoring processes (i.e., judgment strategies). Therefore, it is conceivable that interventions incorporating metacognitive strategy training and extrinsic incentives will positively influence learners’ control and monitoring processes. This should presumably improve the calibration accuracy of learners’ metacognitive judgments.

**Purpose and Overview of the Proposed Study**

Calibration of confidence judgments of performance is crucial because it allows learners to engage in appropriate comprehension monitoring during a learning episode. As the use of the term *metacognition* implies, those individuals who are better calibrators can more adequately adapt to the demands of the situation, and thereby better prepare for future learning contexts that are similar in format or content (i.e., domain). Therefore, accurate and less biased metacognitive judgments are necessary ingredients in the recipe for a successful and metacognitively aware learner. In sum, individuals cannot be fully self-regulated if they cannot accurately calibrate their performance because they would not otherwise be able to monitor their comprehension as they engage in a learning task.

With this in mind the present study addressed whether the proposed interventions (i.e., strategy training and incentives) or their interaction enhance students’ metacognitive judgments by increasing calibration accuracy. I anticipate that each intervention will have a positive effect.

**Methods**

**Experimental Interventions**

The experimental interventions consisted of the following two components. The first component involved metacognitive strategy training specifically developed to
enhance the accuracy of metacognitive judgments in undergraduate students (e.g., Glenberg et al., 1987; Hogarth et al., 1991; King, Zechmeister, & Shaughnessy, 1980; Koriat, 1997; Nietfeld & Schraw, 2002). These metacognitive strategies were drawn from a previous mixed methods study in which students were asked a prediction (prospective judgment) and postdiction (retrospective) question on their expected performance on an exam. Individuals were asked a set of interview questions and then given a think-aloud protocol regarding the exam they took. The interview and think-aloud were intended to examine which metacognitive strategies were particularly effective at increasing accuracy, and they will be included as part of the metacognitive strategy training. The list of potentially beneficial strategies was reduced to seven strategies that were hypothesized to optimize calibration accuracy with respect to performance, including: (a) read and summarize in your own words; (b) use contextual cues in the items and responses, (e.g., bolded, italicized, underlined, or capitalized words); and (c) using diagrams, graphs, tables, etc. These strategies were derived from a content analysis of the interviews and think aloud. The strategy training independent variable is assumed to enhance the monitoring and control processes in the NNMM, as elaborated within Table 3 in the next chapter.

The second component involved providing students with a reward as an incentive to rehearse the metacognitive strategies, thereby improving the accuracy of their metacognitive judgments (e.g., Hacker et al., 2008; Hogarth et al., 1991; Nietfeld & Schraw, 2002; Swanson, 1990; Yates, 1990). Incentives are presumed to increase students’ motivation to more meaningfully learn strategies provided during training. Therefore, incentives are predicted to positively affect students’ information-gathering
capabilities by permitting them to learn strategies to improve calibration accuracy in a more in-depth manner; hence, the monitoring function of the NNMM will be enhanced. Furthermore, incentives could potentially influence not only the monitoring capacity of students but their control as well. Once learned, incentives should motivate students to apply the strategies to actually improve their calibration accuracy. In this fashion, incentives are expected to influence the control function of the NNMM as well. Although the literature regarding the role of incentives on calibration accuracy has spanned both external and internal reward structures (see Hogarth et al., 1991 for a review), the present study focuses exclusively on external incentives. The incentive involved a monetary reward of $10.00 for performance that meets or exceeds 80% of the items answered correctly at posttest. As suggested by the literature and supported by the NNMM, this improved performance is due to enhanced accuracy.

**Study Design**

The study design is a four-group pretest-posttest experimental design, in which two levels of the strategy training intervention are crossed with two levels of the incentives intervention. Participants were randomly assigned to one of four groups: (1) strategy training and incentives; (2) strategy training only; (3) incentives only; and (4) control (see Figure 4) in which they receive one of the four possible combinations of the experimental strategy training and incentive manipulations. Additionally, a pretest and posttest were administered to ascertain whether the intervention was effective at improving the accuracy of students’ metacognitive judgments.
Hypotheses and Predictions

The present study tested the general hypothesis that external interventions improve calibration accuracy by facilitating control and monitoring processes. I expect the strategy intervention to improve monitoring and the incentives to improve control through added attention to the use of these processes to self-regulate. With respect to the interaction between the strategy training and incentives manipulations and the type of test, I expect that those randomized into those conditions would exhibit more accurate calibration and increased performance and confidence at posttest when compared to those who were not. Moreover, I predict that the proposed strategy training and incentives enhance the accuracy of students’ metacognitive judgments, presumably due to improved performance from baseline to posttest. Finally, students’ performance on learning tasks, such as tests/exams, is predicted to increase.

Data Analysis Plan

In order to address the objective of this study, a one-way analysis of covariance (ANCOVA) was conducted, with posttest performance serving as the dependent variable while statistically controlling for baseline performance. Moreover, two 2 (strategy training, no training) x 2 (incentives, no incentives) factorial mixed-model analyses of variance (ANOVAs) were conducted, with type of test (pretest, posttest) serving as the within-subjects factor. Confidence and calibration accuracy each served as a dependent variable in a separate analysis in keeping with the data-analysis strategy in previous experiments. Furthermore, a series of repeated measures analyses of variance (RM ANOVAs) were conducted to compare performance on the pretest (i.e., 20) and posttest (i.e., all 40) items. The first RM ANOVA compared performance on the 20 items given at
pretest to those same 20 items given at posttest. The second RM ANOVA compared the performance on the 40 items given at posttest—that is, the previous 20 items given at pretest with the 20 items added at posttest. Correlations among all outcome variables were computed and reported as well. All appropriate data screening and assumption testing procedures were conducted prior to data analysis.

**Summary and Conclusion**

Metacognition—and more specifically metacognitive judgments like calibration accuracy—is a central component of SRL, and the NNMM provides a meaningful theoretical model by which to examine calibration accuracy and connect it to the larger framework of SRL. Students cannot be self-regulated learners if they are poor calibrators (i.e., rely primarily on inaccurate and biased metacognitive judgments). In order to successfully navigate learning episodes and succeed, students need to rely on effective cues and strategies to calibrate their performance. Therefore, finding interventions that succeed in increasing the accuracy of students’ metacognitive judgments has important implications not only for SRL and metacognition theory, but for educational practice and broader societal goals (e.g., producing competent, independent thinkers capable of accurately discerning what they know and do not know) as well. However, with research demonstrating that the effects of interventions aimed at improving calibration accuracy are short lived (e.g., Glenberg et al, 1987; Hacker et al., 2008; Nietfeld & Schraw, 2002), it is equally as important for theory, research, and practice to find ways to sustain meaningful long-lasting behavior change in students’ metacognitive judgments, such as through incentives.
CHAPTER 2
LITERATURE REVIEW

Finding ways to enhance learners’ metacognitive judgments is crucial. Students can become more competent self-regulated learners if instruction is tailored to improve their ability to evaluate their learning in terms of process and outcomes. However, without systematically vetting instructional practices via the inclusion of appropriate experimental and statistical means to control for extraneous factors, the development and implementation of strategies aimed at the enhancement of metacognitive judgments may very well be a wasted effort fraught with elusive and misleading results. Therefore, research-based interventions specifically developed for the improvement of metacognitive judgments should be sought, modified if necessary to fit specific contextual needs, and implemented within a rigorous experimental design. Preferably, experiments evaluating the effectiveness of interventions should be designed with adequate internal validity and external validity so that results can be generalized to other samples of the intended population.

Because undergraduate students engage in complex, higher order thinking in the course of their learning, they are ripe for interventions targeted at improving metacognitive judgments. For instance, interventions that maximize the accuracy of metacognitive judgments would be an excellent approach towards students’ attainment of educational success and the accomplishment of their learning goals—all of which exhibit different aspects of SRL. Moreover, students’ representation of their environment (i.e., the model represented by the meta-level) can be understood in terms of how well they calibrate their performance. Because individuals at-large are imperfect gauges of their
internal processes, how well their meta-level represents the object-level influences their calibration accuracy. Thus, the aim of this literature review is to critically examine the various interventions that have been designed and promulgated for the explicit purpose of enhancing the accuracy of metacognitive judgments in college undergraduates.

**Organization**

An effective literature review appropriately situates the research, research questions, hypotheses, and design in a broader theoretical context. Therefore, the following literature review places the present study in a theoretical context of SRL and metacognition via the Nelson and Narens Model of Metacognition (NNMM), and provides support for the necessity of the research. The Chapter begins with a taxonomy of relevant concepts to this research study that depicts how these concepts are interrelated. The next section includes a comparison and contrasts between the concepts of SR, SRL, and metacognition, and it sets the stage for the need to situate the research within the framework of the NNMM. The third section more fully describes the NNMM and how it helps explain the phenomenon of calibration as well as calibration accuracy. Next, the relation between metacognition and calibration is more fully developed. The fourth section includes a review of the antecedents/inputs to calibration, calibration itself, and the outcomes, or benefits, of more accurate calibration (i.e., inputs $\rightarrow$ calibration $\rightarrow$ outputs). The following section addresses the need for interventions aimed at enhancing the accuracy of metacognitive judgments. The Chapter ends with a summary/conclusion of the literature reviewed, the contributions of the present study to the broader body of literature, and a more detailed discussion of the present study.
A Taxonomy of Relevant Concepts

Figure 2 presents a taxonomy of the interrelations among concepts pertinent to the present study. The diagram begins with the broadest concept, SR, at the top followed by progressively finer-grained concepts toward the lower levels. The umbrella concept of SRL, which encompasses cognition, metacognition, and motivation, is subsumed within SR. SRL as a broad theoretical framework includes several different theoretical models (see below for a detailed explanation). However, although SR and SRL are an integral part of the present study, the focus will be on the concepts highlighted in black in the diagram. More specifically, metacognition is illustrated in the NNMM, which situates the study of calibration as well as calibration accuracy within the theory of metacognition.
Figure 2. A taxonomy displaying the interrelations among self-regulation, self-regulated learning, metacognition, the Nelson and Narens Model of Metacognition, calibration, and calibration accuracy and bias.

Comparison and Contrast Between SR, SRL, and Metacognition

Dinsmore, Alexander, and Loughlin (2008) posited that the heightened emphasis on self-regulation in academic settings led to the emergence of a new term, SRL, which sought to amalgamate a variety of different strands that addressed aspects of cognition, metacognition, and motivation. SRL emerged in the 1980s and gained momentum in the
1990s. And yet the developmental path of SRL is quite different from the trajectories of metacognition and SR. More specifically, while metacognition and SR developed in parallel, with little opportunity for the research paths to converge, most models of SRL incorporate aspects of both metacognition and SR to shape its lens on learner monitoring (Dinsmore et al., 2008). Theorists and researchers initially conceptualized SRL as an integrated theory of learning (Corno & Mandinach, 1983), intentionally seeking to address the interaction of cognitive, metacognitive, motivational, and contextual factors as opposed to their individual contributions.

Cognitive processes can be divided into two general categories, lower-order mental processes and operations, which reside mainly beyond one’s awareness, and the higher-order thought processes which involve knowledge, awareness, and control of the aforementioned lower-order cognition as well as reflection of one’s actions and performance. Since the time of Socrates and Aristotle, philosophers have been interested in the notion of consciousness and what it involves, such as higher-order thinking processes. Various philosophers such as Hume, Descartes, and Kant pondered which aspects of consciousness allowed humans to reason and convert sensory stimuli into thoughts and mental operations. Without a doubt these philosophers were toiling over the distinction between the lower-order mind and the higher-order mind. However, it was not until the advent of psychology in the mid 19th century that this distinction was more systematically considered. Nevertheless, researchers and philosophers acknowledged that both cognition and metacognition are essential components of SRL.
Models of SRL

SRL encompasses cognition, metacognition and motivation. Several models of SRL have been proposed (Barak, 2010). For instance, Zimmerman (2000) described SRL as a cyclical process involving three parts: (1) forethought (e.g., goal setting, strategic planning, self-efficacy beliefs, and intrinsic motivation); (2) performance and volitional control (e.g., attention focusing, self-instruction, and self-monitoring); and (3) self-reflection (e.g., self-evaluation, attributions, and self-reactions). Boekaerts (1999) proposed a three-layer model of SRL, including: (1) regulation of the self-choice of goals and resources; (2) monitoring of processing methods (i.e., the use of metacognitive knowledge and skills to direct one’s learning); and (3) regulation of processing modes (i.e., the choice of cognitive strategies). Schraw, Crippen, and Hartley (2006) presented a SRL model specifically tailored to science education, including: (1) knowledge (e.g., how to solve domain-specific problems); (2) metacognition (e.g., knowledge about oneself as a learner, goal setting, and implementing strategies); and (3) motivation (e.g., self-efficacy beliefs that affect one’s engagement and persistence in a task). Self-efficacy is vital to self-regulated learning because it influences the level at which learners take on and persevere during difficult tasks (Bandura, 1993; Pintrich, 1999). In relation to SRL, Bandura (2006) presented several aspects of human agency via social cognitive theory. His four-part model included: (1) intentionality, (2) forethought, (3) self-reactiveness, and (4) self-reflectiveness. It is this last component, self-reflectiveness, which most closely aligns with metacognition and calibration accuracy.

All of these models agree that learning is regulated by a variety of dynamic interacting and cyclical cognitive, metacognitive, and motivational factors (Butler &
Winne, 1995). With respect to the present study, all of these models of SRL include some metacognitive component, although not all of them necessarily refer to it as such. Schraw et al. (2006), for example, explicitly use metacognition whereas Zimmerman (2000) refers to it as self-reflection and Boekaerts (1999) as monitoring and regulation of cognitive processing. Conversely, while they all incorporate the same essential components, they differ in their complexity and specificity. For instance, Zimmerman’s model is specific and includes performance as an explicit element whereas Schraw et al.’s model is specific but broader, explicitly using the umbrella terms of cognition, metacognition, and motivation and not expressly including performance. Regarding metacognition, Zimmerman’s model is the one most deeply imbued with a metacognitive influence, as every component involves some aspect of reflection or regulation of cognition and/or behavior. Metacognition is also an integral part of Boekaert’s model, which includes terms such as monitoring and regulation. Bandura’s and Schraw et al.’s models are better organized insofar as metacognition/self-reflectiveness is an actual, unitary component as opposed to interspersed throughout, albeit Schraw et al.’s is the most easily recognizable due to its terminology (i.e., cognition, metacognition, and motivation). In sum, these models view metacognition as the regulator of strategy choice and processing (cognition) as well as behavior/action/volition (motivation).

Metacognition, and more specifically comprehension monitoring, is an essential component of SRL. All of the models of SRL proposed in the literature (e.g., Boekarts, 1999; Schraw et al., 2006; Zimmerman, 2000) include a metacognitive component in which students reflect and regulate (e.g., planning, goal setting, strategy implementation, summarizing, and monitoring one’s progress) their behavior in the service of learning.
Presumably, calibration involves forethought and self-reflection; however, equally as important is students’ awareness of their previous and future performance so that they can develop appropriate volitional control of their learning. In spite of the benefits these theoretical frameworks have for learning outcomes, they do not specifically address the process students undergo to calibrate their performance because they are essentially all-inclusive.

**Need for the Nelson and Narens Model of Metacognition**

Focusing on one aspect of SRL, more specifically calibration accuracy as a subcomponent of metacognition, necessitates a finer-grained model. Hence, a more relevant model specifically grounded in the theory of metacognition (Flavell, 1979; Hacker & Bol, 2004; Nelson & Narens, 1990, 1994) is necessary to better capture the process of calibration and explicate its outcomes, namely accuracy and bias.

The NNMM addresses two fundamental questions regarding metacognitive monitoring. The first involves studying the factors that influence individuals’ metacognitive judgments (Nelson & Narens, 1994). However, while this question attempts to elucidate the factors that increase the degree of confidence in performance, it does not address the accuracy or bias of individuals’ confidence judgments of performance. Therefore, the second question fills this gap by ascertaining the factors that affect the accuracy and bias of confidence judgments of performance. Nelson and Narens (1994) explained that individuals’ metacognitive monitoring is erroneous inasmuch as it omits important information from the context and it may also include information that is not actually germane to the context. Hence, even though monitoring may vary across situations and individuals (i.e., more or less refined), the individual never has an absolute
fix on the environment, thereby influencing accuracy and introducing bias. Consequently, the NNMM (Nelson & Narens, 1990, 1994) is an appropriate model in which to research calibration accuracy. According to this theoretical framework, more proficient calibrators will perform better and learn more efficiently than poor calibrators because their metacognitive judgments are more accurate and less biased vis-à-vis objective performance assessments. The next section includes a more detailed explanation of the NNMM.

**Theoretical Framework**

**The Nelson and Narens Model of Metacognition**

Fundamentally, the NNMM (see Figure 1) conceptualizes the process of thinking as cognitive (i.e., lower-order cognitive processes such as strategy selection during a learning episode) and metacognitive (i.e., the higher-order cognitive processes intended to govern and regulate the lower-order processes). Therefore, any metacognitive activity involves a combination of lower-order cognition and higher-order thinking. Presumably, calibration accuracy and bias influence and are influenced by lower-order and higher-order thinking. More specifically, calibration accuracy and bias affect the selection of strategies and skills at the cognitive level and the knowledge (e.g., procedural and conditional) and regulatory (e.g., planning, monitoring, and evaluation) elements that are necessary to successfully succeed in a learning task at the metacognitive level, contingent upon contextual needs and demands. By the same token, the cognitive and metacognitive elements continuously influence individuals’ calibration accuracy and bias in an on-going cyclical and reciprocal manner. In connection with SRL and metacognitive theories, individuals’ calibration accuracy and bias are affected by cognitive and metacognitive
levels of processing. In order for individuals to achieve optimum calibration accuracy, they need to be effective self-regulated learners as well as metacognitive thinkers (e.g., to adequately monitor comprehension and to evaluate overall performance in the service of learning). Therefore, the NNMM is a scientifically sound and appropriate theoretical framework in which to situate the present study on the enhancement of calibration accuracy in adult learners.

The strength of the NNMM is that metacognition serves as the connecting link in the chain between cognition, cognitive development, and motivational processes (Nelson & Narens, 1994). This, again, joins the broad theory of SRL with metacognition theory and, by extension, to the scientific study of calibration as a process (i.e., how are metacognitive judgments formed and what contributes to their development?) and a product (i.e., aspects of the metacognitive judgments themselves, such as accuracy and bias). Nelson and Narens (1994) remarked,

A college student studying for an examination is a conscious, self-directed [learner] who is continually making memory-relevant decisions about how difficult it will be to memorize a given item or set of items, about what kind of processing to employ during that memorization, [or] about how much longer to study this or that item … (p. 7)

Evidently, the model is quite appropriate for theoretically supporting the study of calibration accuracy among college students by specifying not only the cognitive and metacognitive aspects, but the self-regulatory one as well. As alluded to in Chapter 1, the NNMM has two basic components, or levels, the meta-level and the object-level. The meta-level is the executive which strives to model the object-level, albeit imperfectly. The levels are analogous to the mind (meta-level) and the environment (object-level).
Although somewhat simplistic, the model can be modified to varying levels of complexity. Nelson and Narens (1994) described this dynamic as such,

During monitoring, the metal-level uses information about the object-level—and perhaps about the relationship between the object-level and still other levels for which that level is in turn the meta-level. This information is used to update the meta-level’s model of what is occurring at the object-level. (p. 12)

Thus, the “model” at the meta-level is simply one’s representation and exchange of information (i.e., monitoring) between the object-level, or one’s environment, and the meta-level, or one’s mind. More specifically to this study, a learner’s meta-level model of the object-level includes all of the input related to a learning episode. For instance, if Tommy knows that he will have an exam next week, his meta-level model could conceivably include a representation of the exam study guide, his prior knowledge about the topics the exam will cover, knowledge of his repertoire of learning strategies, assessments about the time and effort it will take to prepare for the exam, and any study materials that he will use to prepare. All of these model components would then inform his confidence judgments of performance on the exam.

These two interrelated components, metal-level and object-level, are linked together by two similar yet distinct processes governed by the actor, or more specifically the learner, namely monitoring and control. These regulatory functions represent and fill the “void” between the meta-level and object-level in a constant, continuous flow of information between levels. It is important to note that the meta-level can influence or modify the object-level, but not vice-versa; in other words, the object-level cannot make adjustments to the meta-level.
In essence, control refers to the ability of the meta-level component of the system to make adjustments or modifications to the object-level. However, this ability is not mutual, as the object-level cannot modify the meta-level. “The information flowing from the meta-level either changes the state of the object-level process or changes the object-level process itself” (Nelson & Narens, 1994, p. 12). The intention and action this yields on the part of the actor could involve: (1) initiating an action; (2) proceeding with an action, which is not necessarily the same as what the actor was already doing because time has elapsed and the goal state may be closer than before; or (3) to cease an action. Therefore, “control processes are not conceptualized as being limited to the starting and stopping of object-level processes … control processes can also modify the object-level processes” (p. 15). “However, because control per se does not yield any information from the object-level, a monitoring component is needed that is logically (even if not psychologically) independent of the control component” (Nelson & Narens, 1990, as quoted in Nelson & Narens, 1994, p. 12).

Monitoring, on the other hand, involves information-gathering and represents the constant information flow between meta-level, object-level, and vice-versa. In essence, this monitoring process allows one to construct goals and plans as well as to evaluate the adequacy, precision, and accuracy of the meta-level model or one’s performance when striving to reach one of the goals and/or plans specified by the model. Nelson and Narens (1990) characterized this monitoring function as follows,

The meta-level is informed by the object-level. This changes the meta-level’s model of the situation, including “no change in state” (except perhaps for a notation of the time of entry, because the rate of progress may be expected to change as time passes). However, the opposite does not occur, i.e., the object-level has no model of the meta-level (as quoted in Nelson & Narens, 1994, p. 12).
Essentially, reflection and contemplation of one’s thoughts are transformed and applied to some goal-directed action/volition or purpose. From this perspective, metacognitive awareness is conceived as both process (reflection on one’s thinking) and product (application or expression of action for some meaningful purpose or end). Additionally, cognitive and metacognitive activities, such as calibration accuracy and bias, are seen as both individual and social activities inasmuch as “[meaning] must activate the object and the object must activate the [meaning]” (Moore, 1972, as cited in Hacker, Keener, & Kircher, 2010, p. 156). Necessarily, an individual must reflect and become aware of his or her own intentions, goals, and thoughts, but others must likewise be able to internalize and approximate the original meaning of the individual who is the initiator of said thoughts, goals, and intentions. In other words, this individual-social interaction is a metacognitive endeavor, and, by extension, indicative of SRL.

Initially conceptualized as a general metacognitive framework, the NNMM has been applied to domains such as writing (Hacker et al., 2010). Hacker and his colleagues (2010) posited that the NNMM is “a versatile theoretical framework for the conceptualization of metacognition as a heuristic for further theorizing and empirical research” (p. 161) on metacognition. In grounding written composition in this metacognitive framework, they declared, “Writing is the production of thought for oneself or others under the direction of one’s goal-directed metacognitive monitoring and control, and the translation of that thought into an external symbolic representation” (p. 154). This not only anchored the domain of written composition to metacognitive theory, but to the larger theory of SRL as well by incorporating motivational flavors (i.e., volition, goals, and intention). Likewise, the NNMM will be used as the central station
that will link this study on the enhancement of calibration accuracy of undergraduate students to metacognition, SRL, and SR.

Whereas the research on calibration of performance has spanned over 25 years, few empirical studies exist that utilize the NNMM as a guiding theoretical framework. To date the NNMM has been invoked by several researchers doing work in other domains, such as written composition (Hacker et al., 2010). However, these studies did not concern calibration accuracy. Moreover, the NNMM has never been adapted to an experimental study on the enhancement of calibration accuracy, and hence, the present study contributes to the literature by testing the utility of this model in an experimental context as it pertains to calibration accuracy. The model’s strength lies in its utility as a means to explain the calibration process and how individuals’ accuracy is influenced by their monitoring and control capacity. Of special relevance to this study, the NNMM assists in illustrating how metacognitive strategy training can benefit learners’ monitoring and control capabilities, and subsequently, enhance the accuracy of their metacognitive judgments. The model is also parsimonious, which makes it simple to understand for a wide range of readers. However, this parsimony serves as a liability as well because there are other metacognitive components which are not explicitly accounted for by the model, such as planning, evaluation, debugging, and metacognitive knowledge. Hence, its explanatory power is necessarily limited.

**Link Between Metacognition and Calibration**

**Metacognition**

Although terms like metacognition, metacomprehension, and metamemory—indicative of higher-order thinking—were not coined until 1979 by Flavell, well after the
birth of cognitivism, the main attributes of metacognition were being debated by philosophers well before then. In 1841, a Scottish physician by the name of John Abercrombie philosophized about consciousness and the notion of reflective thought. Abercrombie (1841) argued that consciousness simply meant any thought that is currently passing through the mind which one is attending to, necessitating awareness. In his chapter, Abercrombie noted,

That more extensive operation to which we ought to give the name of Reflection, as distinguished from simple consciousness, seems to be connected with a power of remembering past perceptions, and past mental processes—of comparing them with present feelings, so as to trace between them a relation, as belonging to the same sentient being—and, farther, of tracing the laws by which the mental processes themselves are regulated [italics added for emphasis]. It is employed also in tracing the relations and sequences of external things, and thus proves the source of certain notions expressive of these relations. It is therefore a compound operation of mind, including various mental processes, especially consciousness, memory, and the act of comparison or judgment [italics added for emphasis]. The knowledge which we derive from this source, whether we call it consciousness or reflection, is referable to three heads. (p. 69)

Abercrombie (1841) further suggested that knowledge (i.e., mental processes) that originates in either consciousness or reflection can be understood vis-à-vis: (1) knowledge of mental processes; (2) what he termed “compound notions” (e.g., time, cause-effect inferences, and motion); and (3) primary truths or intuitive beliefs. In essence, Abercrombie was describing the core ingredients of metacognition and SRL far in advance of both cognitivism and Flavell. Unfortunately, because he was not a researcher, Abercrombie was not able to systematically test his theory. However, it is doubtful whether his theory would have been given much attention, even if he had been systematic, considering that the founders of psychology, Wundt and James, fared no better. It was not until Flavell (1979), under the new zeitgeist of cognitivism, that
metacognition became an anchor in the map of cognitive latent constructs relevant to human learning and cognition.

Flavell’s (1979) model of metacognition involves four main categories: (1) metacognitive knowledge (world knowledge), (2) metacognitive experiences (insight into what information is necessary to fully understand a task/problem), (3) goals/tasks, and (4) actions/strategies. It is important to note that Flavell never truly intended for metacognition to be simply “cognitive,” as the fourth category suggests. In essence, metacognition has a behavioral/action component interwoven in its fabric. Arguably, metacognition plays an important role in such varied tasks as: oral communication of information, oral persuasion, oral comprehension, reading comprehension, writing, language acquisition, attention, memory, problem solving, social cognition, and, several forms of self-control and self-instruction (i.e., aspects of self-regulation). In a practical sense, metacognition is the ability to deliberately and intentionally reflect and think about tasks, situations, and problems that we face on a daily basis (e.g., attempting to consciously block distracting environmental stimuli, reflecting abstractly on better strategies to solve novel problems, etc.; Hacker, 1998).

Although Flavell (1979, 1987) sought to place metacognition at the reins of cognitive processes, he was unclear and rather vague about this distinction. He asserted that metacognitive knowledge and metacognitive experiences can be distinguished from other kinds—namely lower-order cognitive processes—only in their content and function, not in their form or quality. In this sense the imaginary line that divides cognitive from metacognitive activity becomes muddy, especially if one considers that the connection from cognitive to metacognitive activities is not linear. For instance,
research has demonstrated that metacognitive strategies can, and do with some frequency, become automatized, and hence, traverse to the cognitive arena (Livingston, 2003), as is the case as one gains expertise in a domain. Livingston (2003) posited that the distinction can be clarified by considering how the information is utilized by an actor.

The primary distinguishing factor between the two is in the goal or intent of the activity. For instance, cognitive activities seek to obtain, retain, and transfer knowledge for the implementation of activities while metacognitive activities are aimed at regulating and governing task implementation. Cognitive strategies are intended to assist an individual to achieve a particular goal (e.g., text comprehension) while metacognitive strategies are invoked to ensure that the goal has been achieved (e.g., reviewing and summarizing to evaluate one’s comprehension). Conversely, the NNMM views this relationship differently. Rather than focus on the distinction between cognitive and metacognitive activities, the NNMM underscores one’s metacognitive capabilities, namely monitoring and control, and how they function to represent at the meta-level the information flowing from the object-level. As such, the NNMM elucidates how the information-gathering function (monitoring) affects learners’ ability to adjust or modify the environment (control). Presumably, these metacognitive capabilities exert some influence over cognitive activities.

In fact, Flavell (1979, 1987) was the first to make connections between comprehension monitoring and FOK. For instance, the sudden feeling that one does not comprehend something another person just said, such as a teacher during a learning episode, is indicative of comprehension monitoring and FOK. While he was far from discussing the role and function of calibration as a means to determine bias and accuracy
in comprehension monitoring, Flavell paved the way for future researchers to bridge the gap between regulation of cognition ➔ comprehension monitoring ➔ and calibration of performance. The arrows portray the expected causal flow between metacognition at a broad level and calibration accuracy and bias at a more specific level. However, while Flavell provided the foundation for the conceptualization of metacognition, Baker and Brown (1984) provided the dichotomous definition of metacognition that is now widely accepted. These two distinct elements are the knowledge about cognition and the self-regulatory mechanisms that situate monitoring as a primary focus. The self-regulatory processes involve checking the outcome, planning, monitoring effectiveness, testing, revising, and evaluating strategies (Baker & Brown, 1984).

**Calibration**

Calibration of performance is crucial to effective self-regulation (Winne & Jamieson-Noel, 2002) and it is an important cognitive and metacognitive process (Nietfeld et al., 2006; Nietfeld, Enders, & Schraw, 2006; Schraw, 1995; Schraw, Dunkle, Bendixen, & Roedel, 1995). Unlike comprehension monitoring, which focuses on the assessment of one’s learning or strategy use toward the fulfillment of some desired end, calibration more specifically involves the confidence judgments of the state of one’s knowledge vis-à-vis performance on a criterion task that requires one to draw upon such knowledge. The products of this calibration process are the actual metacognitive judgments themselves as well as the level of accuracy and bias of those judgments.

In general, students tend to calibrate tasks better when they use their metacognitive knowledge and beliefs to inform their cognition and behavior regarding the aspects and demands of a task and make ongoing and productive adjustments based
on task information (Butler & Winne, 1995; Winne & Jamieson-Noel, 2002). Thus, as students engage in tasks, they self-regulate their learning in a variety of ways that include metacognitively representing the task and invoking strategies by acquiring information from their background knowledge and from monitoring aspects associated with the task and themselves (Butler & Cartier, 2004). This feedback loop permits students to make metacognitive judgments regarding progress and mastery and, with this information, make informed decisions about whether to continue with current strategies or modify them to address gaps in cognitively representing the task and performance (Lodewyk, Winne, & Jamieson-Noel, 2009).

**Antecedents → Calibration → Outputs**

**Antecedents/Inputs to Calibration.**

Metacognitive experiences typically precede or follow a cognitive activity. They are elicited when cognitive processes fail, such as the recognition that one did not comprehend a passage one just read (Flavell, 1979, 1987; Livingston, 2003). This breakdown in the cognitive flow is thought to trigger metacognitive processes as the learner attempts to remedy the situation (Roberts & Erdos, 1993). This disruption in comprehension presumably significantly influences the accuracy and bias of learners’ metacognitive judgments.

As learners engage in learning activities (e.g., in preparation for an upcoming exam), they invoke cognitive (e.g., rehearsing, summarizing, elaborating, and transforming learned information into meaningful individual knowledge) and metacognitive learning strategies (comprehension monitoring, planning and allocating resources prior to the learning episode, and understanding when information has been
sufficiently rehearsed) with respect to the knowledge they expect the criterion task to involve. In order to successfully navigate and perform the task (e.g., the exam itself), individuals must be able to appropriately calibrate what they do and do not know about the topic(s) the task will cover.

The dilemma lies in the fact that learners are incapable of perfectly assessing their internal cognitive processing (Nelson & Narens, 1990, 1994) and, by extension, unable to precisely and accurately diagnose gaps in their knowledge. Consequently, more highly metacognitive learners are better able to allocate cognitive resources and monitoring to achieve greater accuracy and less bias in their metacognitive judgments because they utilize more sophisticated cognitive strategies and metacognitive knowledge (e.g., declarative and conditional knowledge) as well as regulatory skills (e.g., planning, monitoring, and evaluation). Less metacognitive learners, conversely, are ill-equipped to perform these functions effectively because they lack a repertoire of more sophisticated cognitive strategies and the metacognitive awareness to regulate their learning (Artzt & Armour-Thomas, 1992; Baker 1989; Brown, 1987; Nelson & Narens, 1990, 1994; Schraw & Dennison, 1994; Schraw & Graham, 1997; Schraw & Moshman, 1995). Therefore, these less metacognitive learners will exhibit greater bias and decreased accuracy in their metacognitive judgments because they are less confident of what they do and do not know.

Research on calibration has involved a variety of factors that influence accuracy and bias in metacognitive judgments, including personal characteristics, task parameters, and text and test parameters. For instance, individuals bring certain characteristics to learning settings (e.g., dispositions, attitudes, beliefs, aptitude, etc.) which are impacted
by influences embedded in the tasks (e.g., strategy training and delaying judgments) as well as the text and tests (e.g., test format, item difficulty, and complexity of text). These then presumably affect individuals’ calibration accuracy and bias.

**Personal characteristics.** Personal characteristics include any characteristic that an individual brings to the setting (Schraw, 2009b). Several personal characteristics found in the literature are: the role of working memory in the accuracy of judgments; verbal ability; and sociocultural differences. In terms of working memory, working memory capacity is associated with probability judgments (i.e., comprehension monitoring), although it is not necessarily related to frequency judgments (Dougherty & Hunt, 2003; Sprenger & Dougherty, 2006). Along a similar vein, the relative accuracy of FOK judgments is related to executive cognitive functions, although JOLs exhibited no such relationship (Souchay, Isingrini, Clarys, & Eustache, 2004). In regards to verbal ability, more proficient readers were found to be better equipped to predict and judge performance on a reading task than poor readers (Gillstrom & Ronnberg, 1995). Moreover, reading ability is related to absolute accuracy (a measure of the precision of a confidence judgment vis-à-vis performance on the same criterion task in which the confidence judgments were made), but not to relative accuracy (a measure of the relationship between confidence judgments and performance scores on a criterion task) (Maki, Shields, Wheeler, & Zacchilli, 2005). In essence, poor readers were more overconfident than proficient readers, warranting interventions to eliminate this bias. Finally, research on cultural differences suggests that culture influences monitoring accuracy, presumably because of differences in overconfidence (Lundeberg, Fox, Brown & Elbedour, 2000; Wallsten & Gu, 2003). For instance, individuals reared in cultures
that emphasize community and deference to authority (e.g., Asian and Hispanic cultures) may be prone to underconfidence in their judgments whereas individuals reared in more individualistic cultures (e.g., American and Western European cultures) that accentuate the role of the individual may be more prone to overconfidence.

**Task parameters.** Task parameters are aspects of the task that influence accuracy. A variety of aspects are found in the literature, such as differences between immediate versus delayed metacognitive judgments, different types of judgment tasks, and strategy training. Metacognitive judgments have been found to be more accurate when they are made after a delay (Kimball & Metcalfe, 2003; Nelson, Narens, & Dunlosky, 2004; Thiede, Anderson & Therriault, 2003; Thiede, Dunlosky, Griffin & Wiley, 2005) due to more time to engage in metacognitive monitoring (Schraw, 2009b). Apparently, different types of judgments influence performance on recall and recognition tests because they are more sensitive to verbal ability and information (Metcalf & Kornell, 2005). Presumably, different metacognitive judgment tasks necessitate different underlying metacognitive processes (Keleman, Frost, & Weaver, 2000). For example, some aspects of calibration accuracy and bias may tap into different elements of metacognitive awareness, such as knowledge (i.e., procedural, declarative, and conditional) and regulation (e.g., information management or debugging strategies or evaluation) of cognition. Strategy use and strategy training have also been found to be related to accuracy and bias. For instance, studies have found that summarizing texts increased the relation between confidence and performance (Thiede & Anderson, 2003), and that distributed study enhanced relative accuracy (Son, 2004). Along a similar line, research has demonstrated that monitoring accuracy remained stable throughout the
semester, in spite of continual monitoring practice (Bol et al., 2005; Neitfeld et al., 2006). However, individuals may be more prone to overconfidence with practice (Koriat, Sheffer, & Ma’ayan, 2002), as would be the case if self-efficacy transforms into overconfidence due to over-practice in one domain.

**Text and test parameters.** Text and test parameters are elements of the text, such as length and complexity, and test items, such as difficulty, that influence accuracy and bias (Schraw, 2009b). Text difficulty was evidently related to relative accuracy such that accuracy was optimum when text is at the reader’s current reading level as opposed to overly difficult or simple (Lin, Moore, & Zabrucky, 2001). This finding is in line with Vygotsky’s (1978) zone of proximal development (ZPD), in which individuals need the right amount of challenge (i.e., not too difficult or easy) in order to optimize learning. Other text and test parameters that influence calibration accuracy and bias include, test item difficulty (which significantly constrained performance accuracy judgments that influenced both poor and proficient task performers; Burson, Larrick & Klayman, 2006; Schraw & Roedel, 1994), test item format (especially degree of overconfidence; Juslin, Wennerholm, & Olsson, 1999), deception (decreases accuracy; Brewer, Sampiano, & Barlow, 2005), and information that is perceived to be important to a judgment task (by increasing accuracy; irrelevant information, on the other hand, undermined accuracy; Dougherty & Franco-Watkins, 2003; Dougherty & Sprenger, 2006).

Other studies have unearthed similar factors which contribute to one’s monitoring accuracy and bias, including: (1) characteristics of the testing environment (e.g., difficulty, format, length, and time the test was given vis-à-vis when the learner prepared for the exam; Baker, 1989; Pressley & Ghatala, 1989, 1990); (2) externally-imposed
processing constraints (e.g., feedback, incentives, and explicit instruction; Nietfeld & Schraw, 2002); and (3) individual traits the learner brings to the setting (e.g., expertise and aptitude; Nietfeld & Schraw, 2002).

In sum, the literature indicates that individual differences, task constraints, test and text parameters, expertise, and aptitude, among others, significantly influence monitoring accuracy and bias. However, Schraw (2009b) warned that it is not well established whether differences examined in research involving these elements is due to shifts in metacognitive ability, or to the influence of external scaffolding such as additional time to make judgments, longer texts, or additional test items.

In spite of this lack of clarity, research has demonstrated that several characteristics influence confidence judgment accuracy and bias, and that these fluctuate within individuals based on the type of judgment they make, among other factors, such as those surveyed above. For instance, personal characteristics of the learner, such as motivation and disposition, may either benefit or hinder the calibration process by either increasing or decreasing accuracy and bias. Additionally, environmental characteristics, such as complexity of the domain (e.g., assessing learning in English, physics, or engineering), type and method of evaluation of learning (e.g., selecting performance assessments using holistic- or rubric-based evaluation), and format and difficulty of the test/exam used to assess learning (e.g., using multiple-choice- or essay-format tests) exert an influence on calibration processes and inevitably impact learners’ ability to calibrate accurately. On a more macro level, these various elements affect the monitoring (information-gathering) and control (ability to adjust and modify the environment) mechanisms of the learner (see Figure 1), which in turn, influence the calibration
accuracy and bias of their metacognitive judgments and subsequently outcomes (e.g., achievement) at the micro level. Therefore, these different factors have a bearing on learning outcomes through the following trajectory: antecedent → metacognition → monitoring and control → calibration → accuracy and bias → outcomes.

**Calibration (Metacognitive) Judgments.**

Calibration can be described as the process of eliminating the discrepancy between perceived performance and actual performance on an objective task, such as a test/exam assessing declarative knowledge of a domain or topic. As learners calibrate their performance in preparation for a test/exam, their calibration will be either more or less accurate and more or less biased. Calibration accuracy refers to the level of correctness between metacognitive judgments (i.e., perceived knowledge) and actual performance, as assessed by some objective task (e.g., test/exam); in other words, the magnitude of said discrepancy. Calibration bias, on the other hand, refers to the direction of the discrepancy between metacognitive judgment and actual performance, which can be best understood as over- or under-confidence.

Metacognitive judgments are categorized into three types of judgments regarding monitoring one’s performance, known as prospective, concurrent, and retrospective judgments. They differ in terms of when the judgment is made vis-à-vis the performance task of interest (Schraw, 2009b). Prospective judgments (i.e., predictions) require the individual to make a judgment about learning or performance prior to performing the criterion task. Concurrent judgments ask individuals to make confidence or performance judgments during task performance. Typically, examinees complete a task (e.g., an exam) and make a confidence or performance accuracy judgment after each item (i.e.,
local level). With retrospective judgments (i.e., postdictions), on the other hand, individuals evaluate the ease of learning or performance after completing a task. Unlike concurrent judgments, in which individuals assess performance on an item-by-item basis, prospective and retrospective judgments are more often than not holistic (i.e., individuals judge their performance on all test items at once) (Schraw, 2009b). The present study will focus exclusively on concurrent judgments at the local level.

**Prospective.** Three different types of prospective judgments are common in the literature on metacognitive monitoring; these include judgments of learning (JOLs), feeling of knowing (FOK), and ease of learning (EOL) judgments. JOLs gauge an individual’s ability to study to-be-learned information (e.g., word lists) and subsequently make predictions of recollection for each item (e.g., words or word pairs). JOLs are theorized to invoke metacognitive judgments about one’s capacity to encode and store information in memory. FOK judgments refer to an individual’s ability to predict whether he will recognize information that could not be recalled from long term memory or from a previous learning episode. FOK judgments gauge one’s ability to monitor the contents of memory and to query memory for information. EOL judgments are defined as evaluations regarding the time it takes or the necessary effort to learn information for some future use (e.g., performance assessments). EOL judgments are theorized to tap one’s ability to monitor the relative difficulty of the comprehension process (Nelson & Narens, 1994; Schraw, 2009b).

**Concurrent.** Concurrent judgments refer to continual evaluations of one’s learning or performance (Schraw, 2009b). Examples include confidence judgments regarding learning or performance, ease of solution, and performance accuracy judgments.
during task performance. Research on these three judgments as outcomes is similar inasmuch as an individual responds to a test item or performs a task and immediately thereafter makes a judgment regarding confidence, ease of problem solution, or performance accuracy. Concurrent confidence judgments gauge individuals’ ability to monitor their performance continually in real time. Ease of solution judgments assess individuals’ ability to monitor task difficulty vis-à-vis their available cognitive resources. Judgments of performance accuracy furnish information about individuals’ ability to monitor their performance on the task (Nelson & Narens, 1994).

**Retrospective.** Retrospective judgments occur after the task has been performed. Examples include local as well as global judgments in which a single ease of learning or performance evaluation is made for the overall task. Retrospective ease of learning or ease of solution judgments are quite similar to concurrent and prospective, with the notable exception that they are made post hoc as opposed to a priori (prospective) or during (concurrent). For instance, after completion of a test, individuals may make global retrospective ease of learning or solution judgments even if they have already made local judgments. Retrospective performance accuracy judgments are made according to a similar process (Nelson & Narens, 1994; Schraw, 2009b). Such a situation is more comprehensive insofar as individuals provide information at both the local and global levels. It is plausible for individuals to more accurately calibrate performance of one or the other but not necessarily both.

**Outputs/Outcomes of Appropriate Calibration**

As the NNMM (Nelson & Narens, 1990, 1994) suggests, the calibration process is dynamic and cyclical in nature. Presumably, highly metacognitive learners will continue
to adequately monitor and regulate their learning effectively because the antecedents to
good calibration accuracy will inform the calibration process (the monitoring function in
the NNMM), and thus, yield successful performance on learning tasks (the control
function in the NNMM). Clearly, these individuals are better equipped and able to more
accurately represent the meta-level model of the object-level, which in turn produces an
enhanced capacity to subsequently adjust or modify the object-level (i.e., control).
Unfortunately, such cannot be said regarding the less metacognitive learners. Because
their assessment of their own internal cognitive processes is more flawed (i.e.,
inaccurate), their calibration processing is likely to be more shallow and superficial,
thereby perpetuating the cycle of inaccurate and biased metacognitive judgments. These
individuals do not perform as successfully on tasks because the feedback (information-
gathering or monitoring function) they receive on their poor performance is not
incorporated in, or even communicated to, the meta-level, and thus, their metacognitive
judgments remain inaccurate and biased. Stated differently, less metacognitive learners
have more limited monitoring and control capabilities than their more metacognitive
counterparts; therefore, the meta-level model of the object-level is more
imperfect/inaccurate, which leads to poor performance on learning tasks due to more
inaccurate and biased assessments of their actual knowledge. It is these less
metacognitive individuals that can benefit the most from interventions targeted at
increasing the accuracy of metacognitive judgments.

The benefits to appropriate calibration are varied and broad in scope. For instance,
with respect to the NNMM, learners who engage in appropriate calibration processing
have the benefit of enhanced monitoring (i.e., information-gathering) and control (i.e.,
ability to adjust and modify the environment) capabilities. On a finer grain, the cognitive and metacognitive skills and strategies invoked prior to and during calibration processing can be honed and refined as antecedents to future calibration processing, thus continuously augmenting its efficiency and effectiveness. More specifically, cognitive skills and strategies include: summarizing; elaborating; chunking; mnemonics; putting knowledge into one’s own words for added significance and meaning to increase retention; exerting extra effort on more complex information and less effort on simpler information; and knowing when information has been learned so as not to expend unnecessary effort on already-learned information (i.e., overlearning). Metacognitive skills and strategies at the regulatory end involve: planning for appropriate allocation, investment, and expenditure of resources and effort; improved monitoring for better comprehension of learned information; more sophisticated information management and debugging strategies to handle incoming information and correct errors in judgment, respectively; and more effective evaluation of learning. At the knowledge end, metacognitive skills and strategies include more in-depth knowledge of one’s cognitive capabilities, increased ability to appropriately apply strategies, and a better grasp of when and why to apply strategies. Beyond benefitting learners’ cognitive and metacognitive abilities, good calibration accuracy greatly contributes to one’s actual performance on learning tasks, and thus, to improved achievement outcomes (Nelson & Narens, 1990, 1994; Nietfeld & Schraw, 2002; Schraw & Dennison, 1994).

Review of Research on Improving Metacognition and Calibration

The following section includes a review of relevant literature addressing the enhancement of monitoring accuracy. Table 2 includes a summary of the most relevant
previous research on interventions developed specifically for increasing the accuracy and decreasing the bias of metacognitive judgments. However, before proceeding with the review of interventions, it is necessary to first establish the need for an intervention.

Table 2

*Summary of Previous Relevant Research on the Enhancement of Calibration*

<table>
<thead>
<tr>
<th>Study</th>
<th>Independent Variable</th>
<th>Outcomes</th>
<th>Synthesis of Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kruger and Dunning (1999); Mitchum and Kelley (2010); Prowse-Turner and Thompson (1999)</td>
<td>Instructional training on problem solving and reasoning</td>
<td>Accuracy</td>
<td>Enhanced the accuracy of metacognitive judgments and resolved students’ misunderstandings of problem solving and reasoning tasks.</td>
</tr>
<tr>
<td>Glengberg and Epstein (1985); Glenberg et al. (1987); Walczyk and Hall (1989)</td>
<td>Instructional training on expository text with embedded questions; providing students with feedback on calibration accuracy</td>
<td>Accuracy and Bias</td>
<td>Enhanced the accuracy of metacognitive judgments regarding text comprehension. Students benefit from benefit from individual and normative knowledge in improvements in accuracy and bias from pretest to posttest.</td>
</tr>
<tr>
<td>Glenberg et al. (1987)</td>
<td>Provision of pretest</td>
<td>Accuracy and Bias</td>
<td>Calibration accuracy and bias were improved for students who were furnished a pretest that offers self-generated feedback. However, this effect is limited to pretest that are highly related to the actual objective performance assessment used for the purposes of calibration.</td>
</tr>
<tr>
<td>Chu, Jamieson-Noel, and Winne (2000)</td>
<td>Different testing conditions</td>
<td>Confidence, Bias, and Discrimination</td>
<td>Feedback that had no effect on calibration due to its domain generality. Confidence, bias, and discrimination are not per se influenced by testing conditions, suggesting a general monitoring skill.</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Accuracy/Bias</td>
<td>Findings</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Maki et al. (1990)</td>
<td>Different texts</td>
<td>Accuracy</td>
<td>Text with deleted letters yielded enhanced calibration accuracy due to increased processing time.</td>
</tr>
<tr>
<td>Thiede and Leboe (2009)</td>
<td>Different study times</td>
<td>Accuracy and Bias</td>
<td>Self-paced study times suggested that encoding fluency was not closely linked to the magnitude of overconfidence. Students may have been immersing themselves in strategic responding to maximize correct recall.</td>
</tr>
<tr>
<td>Henrion et al. (1993)</td>
<td>Different training conditions on probability recall</td>
<td>Accuracy and Bias</td>
<td>Although the training failed to improve accuracy, the directional biases were helpful in informing students’ probability judgments.</td>
</tr>
<tr>
<td>Bol et al. (2005); Hacker et al. (2008)</td>
<td>Calibration practice and no practice</td>
<td>Accuracy and Bias</td>
<td>Bol et al. found that the overt practice condition did not significantly improve accuracy; however, higher achiever tended to be less biased than lower achievers. Hacker et al. found that calibration practice somewhat improved accuracy, but only for low achieving students. Incentives were also successful at enhancing accuracy.</td>
</tr>
<tr>
<td>Nietfeld and Schraw (2002)</td>
<td>Monitoring accuracy training and no training</td>
<td>Accuracy and Bias</td>
<td>Strategy training enhanced performance, confidence, and monitoring accuracy irrespective of aptitude, albeit training had no impact on bias. Incentives were not effective at sustaining the intervention effects after a one-week delay.</td>
</tr>
<tr>
<td>Koriat and Levy-Sadot (2001)</td>
<td>Cue familiarity and accessibility training</td>
<td>Accuracy</td>
<td>Cue accessibility and familiarity improved calibration accuracy.</td>
</tr>
<tr>
<td>Lodewyk et al. (2009)</td>
<td>Task structure and task complexity</td>
<td>Accuracy and Bias</td>
<td>Students’ calibration accuracy and bias was better in ill-structured as opposed to well-structured tasks. Task structure and complexity influence calibration.</td>
</tr>
</tbody>
</table>
Learners Are Imperfect Evaluators of their Own Internal Processes

Researchers have argued that there are crucial connections between students’ initial assessments, their use of strategies, and their learning and performance on cognitive tasks (Pintrich, 2002; Thomas & McDaniel, 2007; Tobias & Everson, 2002-2003). In fact, some researchers (e.g., Pintrich, 2002; Tobias & Everson, 2002-2003) have argued that metacognitive judgments are the most foundational of the metacognitive processes because there can be little corrective strategy use with erroneous assessments. However, students’ metacognitive assessments are often inaccurate and biased (Grimes, 2002; Kennedy, Lawton, & Plumlee, 2002; Nelson & Narens, 1990, 1994), although results reported by Lin et al. (2001) suggested that students were capable of monitoring comprehension and performance in spite of relatively poor calibration accuracy, especially for prediction judgments. In addition, they found that more skilled prediction calibrators tended to be more skilled postdiction calibrators as well, and that students’ calibration accuracy remained stable across measures and assessment conditions.

Contrary to popular belief, the relation between calibration accuracy and actual performance is typically low. In other words, poor calibration accuracy is the rule, rather than the exception (Brannick et al., 2005; Glenberg et al., 1987). The high levels of calibration accuracy that have been reported in studies on the calibration of probabilities and FOK research may be the product of utilizing feedback from taking the test to assess the probability of correct performance on the test (Glenberg et al., 1987). For instance, in a series of experiments, Glenberg and associates (1987) found that poor calibration accuracy is not related to a particular type of performance test but rather it is found with inference tests, verbatim recognition tests, and idea recognition tests. Moreover, poor
calibration accuracy was found when the test was provided immediately after reading as well as when the test was scheduled after a delay. These general findings are supported by additional research (e.g., Epstein, Glenberg, & Bradley, 1984; Glenberg & Epstein, 1985; Glenberg, Wilkinson, & Epstein, 1982; Maki et al., 1990; Schraw, 1994; Schraw et al., 1993). On a similar vein, findings by Schraw et al. (1993) suggest that individuals harbor a strong response bias as they calibrate their performance; in other words, individuals are prone to report their accuracy consistently irrespective of item difficulty and accuracy of response, indicating that negative feedback was not integrated to improve both aspects of calibration per se.

This rather small association between confidence about performance and actual performance reported in the literature is critical, partly due to its consequences for student strategies while studying (Brannick, et al., 2005). Optimal calibration accuracy is arguably necessary for sustained effort. For instance, students are likely to persevere while tackling a difficult problem if previous experience has demonstrated that they will ultimately succeed in solving it. Nevertheless, if students believe that efforts to master a subject or solve a problem are fruitless, the likelihood that they will persist in such efforts decreases. However, if students feel that they have mastered a topic, they are less likely to expend additional time studying it. Therefore, poor calibration accuracy can be expected to result in students misallocating effort in wasteful endeavors. Conversely, enhanced calibration accuracy should permit students to become more aware of their own strengths and weaknesses, thereby improving the ability to determine where to best expend effort (Brannick, et al., 2005; Glenberg et al., 1987; Schraw et al., 1993).
Glenberg et al. (1987) argued that the likely reason for poor calibration accuracy was that students assess familiarity with the general domain of a text instead of assessing knowledge obtained from a particular text. Nevertheless, calibration accuracy can be improved if students are provided with a pretest that furnishes self-generated feedback (Brannick, et al., 2005; Glenberg et al., 1987). These findings are in stark contrast to research conducted by Schraw and associates (1993) in which results suggested that feedback had no effect on calibration accuracy of FOK.

On the other hand, incentives have been found to affect calibration of performance (Hogarth et al., 1991; Schraw et al., 1993; Yates, 1990). The literature examining the effects of incentives on calibration accuracy has typically distinguished between internal versus external incentives. Internal incentives draw upon individuals’ intrinsic motivation to perform well on a task (Hogarth et al., 1991), which is driven by inherent enjoyment of the task itself. Sources of this intrinsic motivation include: (a) a need to achieve true mastery of the material (Hogarth et al., 1991; White, 1959); (b) pride and/or enjoyment; and/or (c) a desire to impress or outperform others (see Deci & Ryan, 1985 for a review). Conversely, external incentives are driven by tangible rewards, such as money or extra credit, which are heavily reliant on individuals’ performance on a criterion task. Whereas incentives have the potential to impact individuals’ performance on tasks, the effect may not necessarily always be positive. For example, Lepper, Greene, and Nisbett (1973) and Levine and Fasnacht (1974) found that when individuals ceased to receive external incentives to motivate task performance on an intrinsically enjoyable task their performance and interest on the task waned. Furthermore, the provision of incentives to influence task performance has been found to have deleterious effects on the
amount of incidental learning that individuals attain during learning episodes, presumably because attention is focused on the salient aspect of the task that is rewarded (Bahrick, 1954; Bahrick, Fitts, & Rankin, 1952; Hogarth et al., 1991) rather than the task at hand.

The body of research on the effects of incentives has yielded inconclusive results, especially when performance and calibration accuracy are paired together. For instance, incentives to improve performance were found to negatively influence performance when contrasted with incentives to improve calibration accuracy, indicating that individuals are apt to rely on subjective feelings when calibrating their performance rather than more objective information (e.g., difficulty of the test items; Hacker et al., 2008; Schraw et al., 1993). Yet other studies have demonstrated that incentives have no influence on either calibration accuracy or performance (see, Hogarth et al., 1991, for a review). Hence, the literature suggests a complex dynamic between incentives, calibration accuracy, and performance. Incentives, for example, have a tendency to improve performance for tasks that are easily understood, such as simple, routine, and consistent responses that can be executed quickly and frequently (McCullers, 1978). However, the effect of incentives is less obvious with respect to tasks that require flexibility and creative thinking and creativity (McCullers, 1978; McGraw, 1978; McGraw & McCullers, 1979).

**Training to Improve Metacognitive Monitoring**

Although most individuals of normal intelligence engage in some degree of metacognitive regulation when faced with an effortful cognitive endeavor, some are more metacognitive than others. The more proficient metacognitive students tend to be more successful in their learning (Livingston, 1997; Schraw, 1994) when compared to their counterparts with poor command of their regulatory mechanisms. However, in spite of
the research that demonstrates that by and large individuals, including adult learners, are not metacognitive (e.g., Livingston, 2003; Schraw & Moshman, 1995; Schraw et al., 1995), fortunately, individuals can be taught skills and strategies on how to enhance their regulation of cognitive activities while learning.

Arguably students who fare well on tasks (e.g., tests) monitor with better accuracy than those who perform poorly (Alexander, Carr, & Schwanenflugel, 1995; Baker, 1989; Baker & Brown, 1984; Garner, 1987; Garner & Alexander, 1989; Pressley & Afflerbach, 1995; Schraw & Moshman, 1995). In addition, students who can count on a large store of strategies solve problems and monitor with greater accuracy than learners with less knowledge of strategies. Research by Pressley and Wharton-McDonald (1997) demonstrated a strong positive relationship between strategy training and improved performance. Moreover, research has shown that strategy use (Pressley & Wharton-McDonald, 1997) and strategy training (Jacobs & Paris, 1987) enhance performance on learning tasks. For instance, Nietfeld and Schraw (2002) found that students who have undergone metacognitive skills training enhanced their learning and academic achievement. Moreover, underachieving students have benefitted greatly from metacognitive training (Swanson, 1990) irrespective of intellectual ability (Pressley & Ghatala, 1990; Yan, 1994). The implication for educational practice is that students with well-developed metacognitive awareness exhibit higher achievement, even those who may not necessarily have high aptitude.

Research conducted by Brown and Pressley (1994) and Pressley, Van Etten, Yokoi, Freebern, and Van Meter (1998) found that teaching metacognitive monitoring skills enhances learning outcomes. Paris, Cross, and Lipson (1984), for example, found
that students who underwent metacognitive awareness training were more capable of learning new information effectively and they had more knowledge of reading strategies when compared to students who did not receive such training. Delclos and Harrington (1991) studied the influence of metacognitive training on problem solving. They found that students in the metacognitive training with problem solving skills group outperformed students in the problem solving only and the comparison group. Moreover, Nietfeld and Schraw (2002) concluded that research on metacognitive monitoring training explains significant incremental variance of performance when controlling for the effects of other forms of instruction, such as problem solving skills.

Research on general metacognitive training provides insight into the cognitive processes involved in learning and what characteristics distinguish successful from unsuccessful students. Furthermore, it has implications for instructional interventions, such as teaching students how to be more reflective in their learning processes and outcomes as well as how to regulate those processes for more productive learning (Livingston, 2003).

**Review of Research on Interventions to Enhance Calibration**

It is immensely valuable for students to understand when they have learned the material because studying insufficiently may be costly whereas studying too much can be a wasteful expenditure of resources. In order to achieve this complex task, students must learn to effectively and efficiently calibrate the comprehension of what they have learned (Glenberg et al., 1987; Zabrucky & Moore, 1994; Zabrucky, Agler, & Moore, 2009). The NNMM suggests that students who are highly metacognitive will be successful in achieving this task whereas less metacognitive learners will find it challenging, perhaps
never succeeding at the task. Interestingly, the research on interventions targeting the enhancement of the accuracy in metacognitive judgments is somewhat conflicting. Glenberg and Epstein (1987), on the one hand, found a negative relationship between performance and monitoring accuracy. On the other hand, Schraw and his associates (1995) reported a positive relationship between knowledge and monitoring accuracy. Presumably, learners who possess relevant strategies solve problems and monitor with greater accuracy than learners who are less strategic (Nelson & Narens, 1990, 1994; Nietfeld & Schraw, 2002).

Pertinent to the present study, monitoring accuracy is malleable and is honed as students obtain additional information from a test or when they gain a deep store of metacognitive knowledge. For instance, making tests longer or increasing exposure of test information enhances monitoring accuracy. Furthermore, manipulating external processing effects (e.g., feedback and information about test preparedness) influences monitoring accuracy in positive ways. Conversely, general aptitude does not seem to affect monitoring accuracy (Nietfeld & Schraw, 2002).

Metacognition, and more specifically calibration accuracy, is enhanced in a variety of instructional environments (Alexander et al., 1995; Livingston, 2003; Schraw & Moshman, 1995). Schraw and Graham (1997), for example, proposed that one possible mechanism for the enhancement of metacognitive judgments is the connection between the instructional environment and domain specific knowledge, illustrated as follows: instruction $\rightarrow$ increased knowledge base $\rightarrow$ metacognition. The instructional environment is defined broadly to include explicit monitoring strategy instruction, which is subsequently combined with either explicit or implicit scaffolding. The mechanism
presumes that each of these instructional techniques enhances metacognitive monitoring by initially increasing students’ knowledge store, and thus, guiding their knowledge acquisition; in other words, shifts in students’ knowledge store mediate the relationship between instructional interventions and metacognition. This model for the enhancement of monitoring skills has received support from several research studies. For instance, Palincsar and Brown (1984) found that scaffolded instruction yields significant increases in strategy use and comprehension monitoring during an explicit strategy training intervention and following a six month delay. Scruggs, Mastropien, Jorgensen, and Monson (1986) found that explicit metacognitive strategy instruction yielded a similar increase in monitoring strategy use and subsequent performance.

Accurate confidence judgments are essential in order for students to become successful self-regulated learners. A critical element in this calibration process is students’ ability to rate how well they will perform prior to tests (prediction/prospective) and subsequently rate how well they feel they performed after completing tests (postdiction/retrospective) (Bol et al., 2005). In fact, predicting and postdicting test performance has been defined as a core element of the process of calibration (Lin & Zabrucky, 1998). Calibration accuracy has been significantly correlated with both metacognitive skills and achievement (Butler & Winne, 1995; Kruger & Dunning, 1999; Nietfeld & Schraw, 2002; Schraw, et al., 1993).

Previous attempts to enhance students’ calibration accuracy have been inconclusive in terms of their success. On the one hand, some research studies have found modest gains in students’ ability to predict and postdict performance (Hacker, Bol, Horgan, & Rakow, 2000; Horgan, 1990; Koriat & Goldsmith, 1994, 1996; Nietfeld &
Schraw, 2002; Pressley, Synder, Levin, Murray, & Ghatala, 1987; Walczyk & Hall, 1989). Other studies have resulted in no significant change in calibration accuracy after practice or other types of interventions (Bol & Hacker, 2001; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Koriat, 1997; Koriat, Lichtenstein, & Fischhoff, 1980). A plausible explanation for the mixed results of the success of calibration accuracy enhancement interventions lies in methodology. Schraw (2009b) argues that the method of measurement chosen to gauge calibration accuracy (i.e., absolute versus relative accuracy versus discrimination) can have profound repercussions on results and interpretation. Schraw suggests, among other recommendations, to carefully align measurement to the goals of the research study [i.e., research question(s)] and to use multiple forms of measurement (e.g., using absolute, relative, and discrimination measures) whenever possible so as to obviate these methodological pitfalls. Therefore, the inconsistent results of some of these studies may be methodological artifacts rather than failure of or gaps in the theoretical frameworks invoked to explain results, especially considering that many of these studies used only a single measure of calibration accuracy.

Students’ achievement has been found to be strongly associated with accuracy (Barnett & Hixon, 1997; Bol & Hacker, 2001; Grimes, 2002; Hacker et al., 2000; Kruger & Dunning, 1999; Winne & Jamieson-Noel, 2002). Higher achieving students have consistently shown better accuracy, but greater underconfidence, in their predictions when compared to lower achieving students, who are less accurate yet more overconfident in their predictions. Kruger and Dunning (1999) posited that the most underachieving tended to be the most miscalibrated, overestimating their performance even when faced with negative feedback, thus demonstrating poor comprehension.
monitoring and evaluation skills. Therefore, it would appear that underachieving, miscalibrated students are the most ripe for effective interventions aimed at improving calibration accuracy.

**Considerations for interventions.** An important consideration for research and practice is determining the reasons why metacognitive judgments are impervious to enhancement, despite some well-designed interventions in terms of internal validity. A step in this direction was attempted by Hacker et al. (2000) who found that students based their calibrations on previous calibrations, not previous performance scores, which are objective measures. Along a similar vein, Schraw (1997) reported that confidence judgments on a particular test were related to confidence judgments on unrelated tests, indicating that students base domain-specific judgments on information unrelated to the domain being assessed. In addition, Schraw and his associates (1993) demonstrated that reliability of performance judgments was considerably higher than the reliability of actual performance scores. Therefore, rather than anchoring their judgments of performance on objective learning activities, students appear to be anchoring their judgments on enduring subjective feelings of their personal characteristics (e.g., aptitude/ability).

Some researchers have argued that predictions of performance are subjective experiences of memory. For example, Hertzog, Dixon, and Hultsch (1990) described performance predictions as self-efficacy judgments based on: global and local memory self-efficacy; an evaluation of the memory task; and a set of general processes that transform students’ memory self-efficacy into confidence judgments. Kelley and Jacoby (1996) stated that subjective experiences of memory can take the form of analytic judgments, in which students can list the factors on which a judgment is made, and
nonanalytic judgments, in which assistance from several sources work in tandem to yield a memory experience. However, these sources have yet to be clearly and specifically identified, although research has already uncovered several of them, including motivational factors, volitional behavior, locus of control, self-esteem, self-confidence, and interest (Lin & Zabrucky, 1998; Osman & Hannafin, 1992; Pressley & Ghatala, 1990; Zimmerman, 1995).

**Interventions to enhance calibration.** Calibration of students’ confidence judgments is essential for problem solving, reasoning, and sound decision making insofar as poor decision making should yield an appropriate lack of confidence (Prowse-Turner & Thompson, 2009). Enhanced calibration accuracy is critical because low confidence in a conclusion should be a sign to the learner to derive another (Shynkaruk & Thompson, 2006), just as low confidence in a memory retrieval is a sign that additional resources may be necessary to make certain that the target memory is retrieved (e.g., Kruglanski, Peri, & Zakay, 1991; Zakay, 1998). Reasoners express high levels of confidence in faulty conclusions (Johnson-Laird & Savary, 1999; Quayle & Ball, 2000; Shynkaruk & Thompson, 2006), and variables that influence accuracy may have no or different effects on confidence (Shynkaruk & Thompson, 2006). Shynkaruk and Thompson (2006), for instance, reported that reasoners expressed more confidence in conclusions that they were given additional time to evaluate, even if this additional time did not enhance accuracy. Conclusion believability yielded different effects on confidence and accuracy insofar as, in relation to neutral conclusions, reasoners reported higher levels of confidence when evaluating conclusions that could be accepted or rejected on the basis of belief, even though only unbelievable conclusions yielded more accurate reasoning. Therefore,
training students on the skills necessary to reason effectively should enhance the accuracy of metacognitive judgments (Kruger & Dunning, 1999; Shynkaruk & Thompson, 2006).

Kruger and Dunning (1999), for example, found that students who received a brief instructional intervention were better equipped to judge the number of Wason selection problems they had solved correctly than those who were untrained. Other instructional interventions also have been successful in enhancing the accuracy of metacognitive judgments and resolving students’ misunderstandings about problem solving and reasoning tasks (e.g., Mitchum & Kelley, 2010; Prowse-Turner & Thompson, 2009). The implication of these studies is that instructed and spontaneous strategies are often not equivalent. In some instances, not all students are able to benefit from an instructed strategy or benefit to the same extent as those who spontaneously construct the strategy. This is a critical consideration for the present study, as strategies selected for the metacognitive strategy training were spontaneously developed by college undergraduates.

Considering the immense importance of reading comprehension, Glenberg and his colleagues (Glenberg & Epstein, 1985; Glenberg et al., 1987; Walczyk & Hall, 1989) conducted several experiments on enhancing the calibration accuracy of students’ text comprehension. These studies found that students were poor calibrators of text comprehension, unable to distinguish what they did and did not comprehend. Moreover, providing training in calibration accuracy (e.g., by providing expository text that contains illustrative examples and embedded questions, furnishing students with feedback on their calibration accuracy) enhanced the accuracy of metacognitive judgments regarding text comprehension. Interestingly, the improvement from initial calibration to recalibration of comprehension accuracy could be due to students’ access to their own knowledge as well
as normative knowledge available to all students (e.g., students may be skilled at
deducing which inference verification items are particularly challenging, and
subsequently lower confidence) (Glenberg & Epstein, 1985; Glenberg et al., 1987).
Glenberg et al. (1987) found that calibration accuracy may be improved if students are
furnished a pretest that offers self-generated feedback. In such treatments, students are
able to utilize feedback from the pretest to predict subsequent test performance with a fair
degree of accuracy. The effect of self-generated feedback is limited, however, as
calibration accuracy improves only when the processes and knowledge drawn by the
pretest are highly correlated with the processes and knowledge needed on the objective
performance test.

Nevertheless, these results were later replicated by Chu, Jamieson-Noel, and
Winne (2000), who reported that process feedback had no effect on calibration accuracy
due to its domain-generality. Moreover, they found that confidence, bias, and
discrimination are not per se influenced by testing conditions, indicating a general
monitoring skill. Of particular importance to the present study is that the calibration
training used in these research studies on text comprehension only yielded small
improvements in calibration accuracy primarily because the researchers involved were
unable to determine the causes of calibration accuracy failures. Therefore, the proposed
study takes into account locating the purported causes of calibration accuracy failures
among students (i.e., miscalibration) and including more specific training procedures and
curricula so as to increase intervention effect sizes.

Maki and associates (1990) investigated whether increased processing enhances
calibration accuracy. They found that text with deleted letters yielded improved
calibration accuracy. In essence, students can predict performance on text material with a relatively high degree of accuracy due to more active processing during reading afforded by text with missing letters. This research is in line with findings from Thiede and Leboe (2009). They asserted that miscalibration of competence is believed to occur when JOLs made in the presence of intact cue-target pairs during study create a “foresight bias,” which inflate JOLs due to the apparent relationship between a cue and a target. In two experiments, Thiede and Leboe found that self-paced study times indicated that encoding fluency was not closely linked to the magnitude of overconfidence. Error data showed that students may have been immersing themselves in strategic responding to maximize correct recall. Their results highlight the relevance of considering factors that influence both JOLs and recall performance when examining sources of miscalibration in absolute accuracy. Similarly, Castel, McCabe, and Roediger (2007) found miscalibration of competence for identical word pairs. These studies suggest the necessity for interventions, such as the one proposed in the present research, to consider processing time and aspects such as foresight bias.

Henrion, Fischer, and Mullin (1993) examined the calibration accuracy of students in subjective probability distributions using three conditions: direct assessment (students directly estimated probabilities), experimenters’ decomposition (variables were determined by the experimenter and evaluated by the students), and students’ decomposition (variables and solutions were determined by the students). They found that, contrary to the divide and conquer principle, decomposition did not improve the accuracy of confidence judgments, as there were no significant differences between the experimenters’ and students’ decomposition conditions. Interestingly, students in the
direct condition tended to underestimate their performance and students in the decomposition condition tended to overestimate their performance, demonstrating a relatively high degree of bias. While the training failed to improve calibration accuracy, the directional biases reported in the study have practical implications in terms of students’ probability judgments.

In an experimental study using prediction and postdiction judgments, Bol and her colleagues (2005) found that students in the overt calibration practice condition did not significantly improve their calibration accuracy when compared to students assigned to the no-practice condition. However, they reported that higher achieving students were significantly more accurate, yet underconfident, in their predictions whereas lower achieving students were less accurate and overconfident. Thus, this instructional intervention manipulating overt calibration practice had no effect on the accuracy of metacognitive judgments. This finding is particularly relevant to the proposed study, as it provides evidence that interventions to enhance calibration accuracy may not always be successful. However, Bol et al. did not include a pretest measure to establish group equivalence, which may have behaved as a confound in the study, thus undermining any potential intervention effects. On a follow-up study investigating the effects of extrinsic rewards and reflection on calibration accuracy, Hacker et al. (2008) found that high-achieving students were highly accurate in their confidence judgments and did not exhibit significant increases in their calibration accuracy. Low-achieving students, on the other hand, were less accurate in their metacognitive judgments and gained modest increases in their calibration accuracy due to reflection. Interestingly, students in the incentives condition demonstrated significant gains in their calibration accuracy compared to those
in the no-incentives condition, although reflection on explanations for their cognitive judgments had no effect on calibration accuracy. These two studies are noteworthy because they included calibration accuracy training and incentives; however the mixed findings, especially as they relate to calibration accuracy justify the need to develop stronger, more appropriate interventions to enhance calibration accuracy.

In two experiments, Nietfeld and Schraw (2002) tested the influence of prior knowledge and strategy training on monitoring accuracy. Results suggest that knowledge of the domain was related positively to domain-specific performance and monitoring accuracy. Moreover, strategy training enhances performance, confidence, and monitoring accuracy irrespective of aptitude immediately following the intervention but not after a one-week delay post-intervention. In sum, this study demonstrated that performance, monitoring accuracy, and confidence scores were all enhanced by the intervention, which in turn supports the inference that metacognitive judgment skill enhancements are followed by concomitant increases in the accuracy of monitoring judgments. Of major significance to the present study, this research lent credence to the trainable hypothesis, which stipulates that short-term strategy training enhances monitoring accuracy, even when such strategy training is not explicit per se.

In an experiment manipulating cue familiarity and accessibility of FOK judgments, Koriat and Levy-Sadot (2001) posited that both cue familiarity and accessibility contribute to FOK; however whereas the effects of familiarity happen early, those of accessibility happen subsequently and only when cue familiarity is sufficiently high to drive the interrogation of memory for possible responses. They found that both familiarity and accessibility improved students’ FOK judgments, but the effects of
accessibility were found mostly when familiarity was high. This moderating pattern was found when FOK judgments were delayed, yet not when they were immediate. When considering interventions to enhance the accuracy of metacognitive judgments it is important to understand the influence of cue familiarity and accessibility on calibration accuracy, as the aforementioned research indicates.

Finally, Lodewyk, Winne, and Jamieson-Noel (2009) examined the influence of task structure on calibration accuracy. They found that students tended to calibrate their knowledge better in ill-structured versus well-structured problems, presumably because students reported greater metacognitive strategy use in the ill-structured problems compared to their performance on well-structured problems. In addition, low achieving students showed poorer calibration accuracy on both tasks compared to high achieving students who demonstrated more proficient calibration accuracy. Hence, it is valuable to the success of the intervention to acknowledge that tasks structure and the complexity of the task significantly influence not only the calibration process itself, but calibration accuracy as well.

**Summary and Conclusion**

In conclusion, research demonstrates that students benefit from instruction targeted at improving metacognitive monitoring via more sophisticated strategy use and that several of these instructional interventions have been developed and successfully implemented. Furthermore, a preponderance of evidence exists suggesting that several instructional interventions specifically tailored to enhance the calibration accuracy of students’ metacognitive judgments in various domains (e.g., text comprehension, problem solving and reasoning, math) are successful, if only modestly. Nevertheless,
interventions targeted at improving calibration accuracy have yielded modest/weak effect sizes as indices of the explained variance in the outcome measures that can be attributed to the experimental manipulations. For instance, Hacker et al. (2008) reported effect sizes ($\eta^2$) of 0.08 to 0.12 and Bol et al. (2005) found effect sizes ranging from 0.04 to 0.15. Moreover, Nietfeld et al. (2006) reported effect sizes ranging from 0.10 to 0.17. Hence, modest effect sizes in the present study would be in line with the body of literature on improving metacognitive judgments, albeit larger effect sizes may result.

However, in order to develop and effectively implement such instructional interventions to enhance students’ accuracy of their metacognitive judgments, researchers need to consider other aspects beyond the intervention, including the influences on (e.g., properties of the testing environment, externally imposed processing constraints, and personal attributes of the learner) and mechanism for (e.g., relationship between the academic environment and domain knowledge) metacognitive strategy use, as these will influence not only the structure and type, but the effectiveness of selected interventions. Given that many of the interventions aimed at the enhancement of calibration accuracy reviewed here have resulted in inconclusive findings and/or small practical significance, it is necessary to continue to refine and modify them. By developing more sound and rigorous interventions that include various components such as training, incentives, and feedback, and which are subsequently implemented in an experimental context, many of the flaws in previous studies can be corrected.

**The Proposed Study**

Well-developed metacognitive skills are necessary to navigate the intricacies and heavy cognitive demands of modern life. Presumably, higher-order thinking skills such as
metacognition, critical thinking, and creative thinking and creativity contribute to the development and promotion of independent thinkers who are capable of critically consuming information in a fast-paced digital age, such as the present. With the environment rapidly in flux, highly metacognitive thinkers will become ever more necessary. For instance, metacognition plays a critical role in reading comprehension, writing, memory and metamemory, and problem-solving as well as other learning domains (Nietfeld et al., 2006; Schraw & Graham, 1997). Thus, metacognitively aware individuals can select appropriate problem-solving strategies in the service of learning and allocate resources effectively by invoking skills such as planning, monitoring, and evaluation. Moreover, metacognition affords individuals the ability to monitor their current knowledge and skills, plan and distribute cognitive resources efficiently, and evaluate the learning process (Schraw et al., 2006). Finally, metacognition places the individual at the very center of his cognitive processes as a reflective, volitional, and knowledgeable individual in control of his own thoughts and actions predicated on the strategies to problem solve derived from those thoughts (self-monitoring). This self-monitoring component allows learners to invoke self-regulatory schemes to intentionally and deliberately choose strategies that maximize the likelihood of success in a task or finding a solution to a problem (Reynolds & Wade, 1986).

This literature review has proposed that the NNMM (Nelson & Narens, 1990, 1994) is an appropriate theoretical framework in which to situate the present study on the enhancement of calibration accuracy among adult learners, more specifically, college undergraduates. According to this perspective, more highly metacognitive thinkers have more refined monitoring (information-gathering) of the environment (the object-level)
which subsequently continuously informs the metal-level’s model (representation of the environment) of the object-level. Although still imperfect in terms of their representation of the environment and knowledge of their own internal processes, these highly metacognitive thinkers have a *more* perfect representation and increased awareness of their cognitive processes vis-à-vis their less metacognitive cohorts. Therefore, highly metacognitive students will be better able to engage in the calibration process and produce metacognitive judgments with more appropriate levels of confidence and greater accuracy than less metacognitive students. It is these less metacognitive students that are the main target audience for the proposed study.

**Description of Strategy Training and Incentives**

**Strategy training.** The strategy training component of the intervention involved providing students with instruction regarding more sophisticated and adaptive strategies that are more conducive to enhancing calibration accuracy. Examples of strategies that are part of the training include: (a) read and summarize in your own words; (b) use contextual cues in the items and responses, e.g., bolded, italicized, underlined, or capitalized words; and (c) using diagrams, graphs, tables, etc. These strategies were derived from a content analysis of the interviews and think aloud and they have been demonstrated to maximize improvements in calibration accuracy with respect to performance (e.g., Bol et al., 2005; Hacker et al., 2008; Nietfeld & Schraw, 2002), and hence, they are more spontaneous in nature. Once introduced and explained in detail, the strategies were scaffolded and demonstrated so that students perceived their value with respect to improved calibration accuracy. Finally, students were provided with an
opportunity to apply and to practice the skills/strategies to bolster their confidence in the application of said strategies/skills.

**Incentives.** Although the literature regarding the role of incentives on calibration accuracy has spanned both external and internal reward structures, the present study focused exclusively on external incentives for more accurate performance. Participants in the strategy training and incentives and the incentives only groups were explicitly instructed that their pay would depend on how well they perform at posttest: “Your pay for participation in the experiment **WILL** depend on how **WELL** you perform on the assessment at posttest. **WELL** is defined as getting ≥ 80% of the items correctly.” The incentive involved a monetary reward of $10.00. From pilot study data of 76 undergraduates, a median percentage of 79 was calculated for performance on the test. The median score was selected because it is not susceptible to undue influence by outliers, as would be the case with the mean. As such, students need to answer at least 80% of the items correctly on the performance assessment at posttest to receive the incentive. This procedure permits the isolation of the true influence of incentives on calibration accuracy with respect to performance.

**Proposed Research Question**

The following research question guided the conduct of the proposed study:

Are the proposed interventions, including metacognitive strategy training and incentives, effective at increasing the accuracy of students’ metacognitive judgments?

**Hypotheses**

Predicated on the aforementioned question, the following hypotheses are proposed:
H_1: With respect to the combination of the strategy training and incentives manipulations, I hypothesize that the interactive effect will yield more accurate calibration, higher performance, and higher levels of confidence at posttest for the groups that are exposed to the combined interventions. Interestingly, Hacker et al. (2008) did not find an interaction effect between their training and incentives conditions. However, there is reason to believe that a combination of the two conditions in the present study will lead to improved performance, calibration, and confidence. When combined, strategy training and incentives positively influence both the information-gathering (monitoring) and control processes of the NNMM. These more effective monitoring and control processes in turn lead to better performance, presumably due to increased accuracy (i.e., more accurate confidence judgments).

H_2: The proposed strategy training is predicted to enhance the calibration accuracy of students’ metacognitive judgments at posttest (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; Nelson & Narens, 1990, 1994). It is noteworthy to mention that Bol and her colleagues (2005) and Henrion et al. (1993) found that their training was not successful at enhancing calibration accuracy.

H_3: Incentives are expected to positively influence calibration in such a way as to enhance accuracy (e.g., Hacker et al., 2008; Hogarth et al., 1991; Schraw et al., 1993; Yates, 1990). However, the findings regarding incentives also have been inconclusive, as some studies suggest it has no effect on monitoring accuracy (Hogarth et al., 1991). These mixed findings with respect to the effects incentives on the enhancement of calibration accuracy warrant further investigation, which this study proposes to achieve.
Moreover, students’ performance on learning tasks, such as tests/exams, is expected to increase for those exposed to the strategy training and incentives interventions. As students’ monitoring and control is enhanced due to the effects of the strategy training and incentives, as proposed by the NNMM, their subsequent performance on learning tasks is likely to improve as well (e.g., Hacker et al., 2008; Henrion et al., 1993; Nietfeld & Schraw, 2002). Enhanced information-gathering capabilities increase metacognitive awareness because the meta-level’s model of the object-level is more precise. As a result of this more accurate model, students’ control capabilities (e.g., applying appropriate and effective strategies while learning) are concomitantly enhanced, which consequently leads to improved performance.

**Situating the Hypotheses within the Nelson and Narens Model of Metacognition**

The four hypotheses described above can be explained in terms of metacognitive processes modeled within the NNMM. When learners are explicitly taught more sophisticated and specific strategies intended to produce good calibration accuracy their ability to gain a better understanding of their own internal cognitive processes will be enhanced. As they become more aware of their inner functioning, the NNMM suggests that learners’ information-gathering (monitoring) capability and ability to modify the environment (control) will be augmented. The delayed judgment effect, for instance, describes situations in which individuals’ accuracy improves, which allows for rehearsal and the opportunity for meaningful learning of strategies. Presumably this increased processing time has a positive effect on performance due to better accuracy caused by greater awareness of one’s internal processes (i.e., monitoring and control).
Likewise, extrinsic and intrinsic incentives have been shown to have positive effects on performance (e.g., Hogarth et al., 1991) and calibration accuracy (e.g., Hacker et al., 2008), arguably because of intensified effort and motivation (Deci & Ryan, 1985) on the part of the learner. Extrinsic rewards, as the exclusive form of incentives in the present study, should positively affect calibration accuracy as well. Conceivably this is achieved by motivating students to exert additional effort to hone their monitoring and control capabilities, thereby enhancing calibration accuracy, which should subsequently improve performance.

Therefore, whereas strategy training directly influences monitoring and control, and subsequently calibration accuracy and performance through greater self-regulation, incentives do so indirectly through motivational factors. In summary, the argument proffered is that both strategy training and incentives enhance individuals’ control and monitoring processes through greater, clearer insight to their own internal cognitive functioning, thereby improving calibration accuracy, resulting in improved performance. However, it is important to note that extrinsic incentives are intended to lead to higher performance as a vehicle to better accuracy. This can be succinctly conceptualized as follows:
Figure 3. The hypothesized explanatory trajectory of antecedents, calibration, and outcomes based on the two manipulations of strategy training and incentives used in the study. These are predicated on theoretical claims of the Nelson and Narens Model of Metacognition.

Table 3 summarizes how strategy training and incentives influence the control and monitoring functions of the NNMM, which subsequently affect calibration accuracy.
Table 3

**Summary of the Effects of Strategy Training and Incentives on the Control and Monitoring Processes within the Nelson and Narens Model of Metacognition**

<table>
<thead>
<tr>
<th>Predicted Implicit Processes in each Hypothesis</th>
<th>Strategy Training Condition</th>
<th>Incentives Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H₁:</strong> Monitoring and Control</td>
<td>Through the combined effects of strategy training and extrinsic incentives combined, learners are able to not only meaningfully learn and retain strategies to yield better accuracy through improved performance (information-gathering or monitoring) but also have the disposition/desire to apply those calibration strategies (modify or control their environment)</td>
<td>Extrinsic incentives are expected to motivate learners to want to enhance their control capabilities. As such, an augmented control function should result in greater accuracy</td>
</tr>
<tr>
<td><strong>H₂:</strong> Monitoring</td>
<td>A greater repertoire of strategies that are meaningful and sophisticated increases learners’ information-gathering capabilities which should make their metacognitive judgments more accurate. It is important to keep in mind that as highly related elements, what influences bias will affect accuracy and vice-versa</td>
<td></td>
</tr>
<tr>
<td><strong>H₃:</strong> Control</td>
<td>Deeper insight to their internal cognitive processes afforded by strategy training leads to enhanced control and monitoring functions, and greater awareness of their capabilities is expected to increase accuracy, thereby resulting in better performance</td>
<td>As learners’ desire to enhance their monitoring and control capabilities increases—leading to better accuracy—their performance on tests/exams is predicted to improve</td>
</tr>
<tr>
<td><strong>H₄:</strong> Control and Monitoring</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Contributions of the Present Study to the Literature**

While previous studies on the enhancement of calibration accuracy have involved a plethora of different interventions, the proposed study will involve a distinct training
program. The training that is included as part of the intervention was developed from a pilot study I conducted during the Spring 2011 semester with college undergraduates. The mixed methods study involved a structured interview and a think-aloud protocol in which five proficient and four poor calibrators were selected to participate based on their quantitative prediction and postdiction scores. From these qualitative data, successful (versus unsuccessful) strategies, cues, and skills were extracted, which form an integral and innovative aspect of the proposed intervention. These strategies and skills are likely to be successful across a diverse range of undergraduates because they were spontaneously developed by the students themselves. Moreover, every attempt was made to include undergraduates with different cognitive styles, dispositions, and profiles as part of the pilot study. Incentives also form part of the proposed intervention. These are used to ascertain their influence on confidence, accuracy, and performance. As previously highlighted, the research has been mixed in terms of not only the effects of various training programs, but also the influence of incentives. Hence, this study will contribute to the literature by providing additional evidence as to the influence of incentives and training on the calibration accuracy of college undergraduates.

Likewise, the proposed study contributes to theory by affirming the utility and value of the NNMM (Nelson & Narens, 1990, 1994), which recently has fallen under heavy criticism by calibration scholars and researchers (Boekaerts & Rozendaal, 2010; Efklides, 2011). The process of calibration as well as calibration accuracy are influenced and explained by the processes described in the model, as enumerated throughout this discussion. For instance, because of our inability to perfectly and accurately represent the environment (object-level) in the meta-level’s model, even the most highly metacognitive
learners are prone to some level of inaccuracy and bias in their metacognitive judgments. This imprecise monitoring (information-gathering) and limited control (ability to modify and adjust the environment) functions inevitably have profound consequences for learners’ ability to engage in the calibration process and on the accuracy of their metacognitive judgments. It is for these very reasons that the proposed intervention is necessary. Highly metacognitive learners who exhibit high accuracy are without a doubt better able to monitor their environment, inform the meta-level’s model, and make appropriate adjustments and modifications to their environment, thus subsequently yielding better learning and achievement outcomes. This is the ultimate goal of educational research, and this study contributes to the fulfillment of this goal.
CHAPTER 3

METHODS

Participants and Setting

An a priori power analysis with power equaling .80, a \( p \)-value of .05, and a medium effect size yielded a total of 128 participants as adequate power for this design. However, every attempt was made to recruit sufficient participants, taking into account a 20% attrition rate, in order to find statistical and practical significance. Because such large samples would be difficult to secure utilizing only the Educational Psychology Experiment Management System, participants were also recruited from the Department of Psychology’s subject pool. Appropriate permissions were secured from the Department of Psychology Subject Pool Coordinator. Therefore, participants were undergraduate students enrolled in either general psychology or educational psychology courses.

One hundred-sixty participants were recruited to participate in the study. Participants’ age ranged from 18-65 years (\( M = 22.76, SD = 7.15 \)). There were 49 male (30.6%) and 111 female (69.4%) participants, nearly half of them [77 (48.1%)] reporting enrollment in education-related majors ranging from early-childhood education to secondary education. The remaining 83 participants (51.9%) reported majors ranging from engineering to hospitality to art therapy. Participants varied with respect to undergraduate standing, with 25 (15.6%) freshmen, 50 (31.3%) sophomores, 56 (35.0%) juniors, and 29 (18.1%) seniors. Finally, slightly fewer than half of participants [71 (44.4%)] reported their ethnicity as White/Caucasian, 32 (20%) reported Hispanic/Latino, 5 (3.1%) reported African American/Black, 35 (21.9%) reported Asian/Pacific Islander, 1
(0.6%) reported Native American/Alaskan Native, and 16 (10.0%) reported Other/Mixed. Nevertheless, approximately 30% of participants were lost to attrition, yielding 107 cases with complete data. Little and Rubin (1987) argue that this level of attrition in studies using repeated measures designs or some variation is typical.

**Design and Materials**

The study design is a four-group pretest-posttest experimental design, in which two levels of the strategy training intervention are crossed with two levels of the incentives intervention. Additionally, a pretest and posttest were administered to ascertain whether the intervention was effective at improving the accuracy of students’ metacognitive judgments. At pretest, individuals read a passage, answered 20 multiple choice items, and rated their confidence of performance for each test item. At posttest, individuals reread the passage, answered 40 multiple choice items, and rated their confidence of performance on the items. Twenty of the posttest questions constituted “old” items that followed the pretest, while 20 items constituted “new” items. Figure 4 presents a schematic of this design.
Figure 4. A schematic of the proposed intervention, including the four-group breakdown with respect to the two components of the intervention, strategy training and incentives.

Table 4 presents a summary of strategies that are included in the strategy training component of the intervention as well as the level of processing, the affected calibration component, the NNMM function in operation, and the hypothesized influence of each strategy on calibration.
Table 4

*Summary of Metacognitive Strategies and their Relation to Calibration and Theory*

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Affected Calibration Component</th>
<th>NNMM Function in Operation</th>
<th>Hypothesized Influence on Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review main objectives of the text and focus on main ideas and overall meaning</td>
<td>Both</td>
<td>Monitoring</td>
<td>Enhance calibration through clarifying misunderstandings and tying details to main ideas</td>
</tr>
<tr>
<td>Read and summarize material in your own words to make it meaningful; use elaboration and create your own examples</td>
<td>Both</td>
<td>Both</td>
<td>Enhances calibration by transforming knowledge into something personally meaningful</td>
</tr>
<tr>
<td>Reread questions and responses and reflect on what the question is asking; go through and take apart the question paying attention to relevant concepts</td>
<td>Both</td>
<td>Both</td>
<td>Purposefully slowing information processing allows for a more accurate representation of the problem, thus decreasing errors in judgment</td>
</tr>
<tr>
<td>Use contextual cues in the items and responses, e.g., bolded, italicized, underlined, or capitalized words</td>
<td>Accuracy</td>
<td>Monitoring</td>
<td>Using contextual cues allows the mind to focus on salient aspects of the problem rather than seductive details, thereby increasing accuracy</td>
</tr>
<tr>
<td>Highlight text; underline keywords within the question to remind yourself to pay attention to them; use different colors to represent different meanings</td>
<td>Accuracy</td>
<td>Control</td>
<td>Highlighting and underlining can assist one to focus on main ideas and what is truly important, increasing accuracy; however, relying too much on this can be counterproductive and may potentially increase errors</td>
</tr>
<tr>
<td>Relate similar test questions together and read them all before responding to any</td>
<td>Accuracy</td>
<td>Control</td>
<td>Relating information together provides a clearer understanding of the material and may highlight inconsistencies that need to be resolved; it may point to information the learner may have missed, increasing accuracy</td>
</tr>
<tr>
<td>Use diagrams, tables, pictures, graphs, etc. to help you organize information</td>
<td>Accuracy</td>
<td>Both</td>
<td>These strategies help simplify complex topics by breaking them down to their constituent parts; this increases accuracy by decreasing errors</td>
</tr>
</tbody>
</table>
The proposed intervention incorporates several aspects from interventions previously demonstrated to be successful at improving students’ metacognitive judgments, as thoroughly discussed in the previous chapter. Nietfeld and Schraw (2002) and Schraw (1998) argued that direct strategy instruction is effective at enhancing metacognitive judgments. For instance, Schraw (1998) asserted that instructional strategies, such as promoting general metacognitive awareness and improving self-knowledge and regulatory skills, lead to enhanced metacognition; this improved metacognitive awareness should subsequently positively influences students’ metacognitive judgments. Likewise, Nietfeld and Schraw (2002) found that a short strategy training session was effective at increasing both domain-specific content knowledge and accuracy of metacognitive judgments. Strategy instruction included training on improving self-knowledge and regulatory skills (i.e., monitoring accuracy). These studies have demonstrated that strategies similar to the seven strategies illustrated in Table 4 are the most effective at increasing calibration accuracy with respect to performance.

Offering students incentives has also been shown to increase the accuracy of calibration (e.g., Hacker et al., 2008; Hogarth et al., 1991; Schraw et al., 1993; Yates, 1990). Hacker and his associates (2008) and Schraw and his colleagues (1993) reported that students in the incentives condition demonstrated significant gains in their calibration accuracy when compared to those in the no-incentives condition. Therefore, incentives also play a role in the proposed intervention. More specifically, students randomly assigned to the incentives conditions will receive rewards to ascertain the effect of incentives on metacognitive judgments.
Demographics

A brief, researcher-developed demographic form (see Appendix A) was utilized to obtain demographic information from participants. This form included questions soliciting participants’ gender, age (on a continuum), major, ethnic identity, and undergraduate standing. These demographic variables were used to describe the population.

General Metacognitive Awareness

The eight components of metacognition (i.e., knowledge of cognition: declarative, procedural, and conditional; regulation of cognition: planning, monitoring, debugging, information management, and evaluation) were measured using the Metacognitive Awareness Inventory (MAI; see Appendix B) developed by Schraw and Dennison (1994). The MAI is a 52-item instrument which measures domain-general metacognition through its constituent components. Sample items include: “I ask myself periodically if I am meeting my goals” (monitoring); “I try to use strategies that have worked in the past” (procedural knowledge); “I reevaluate my assumptions when I get confused” (debugging); and “I ask myself if I have considered all options after I solve a problem” (information management).

Ratings on each item were marked by a vertical slash on a continuous 0-100 bipolar (i.e., “not at all true of me” representing 0 and “very true of me” representing 100) scale line that is 10 cm (i.e., 4 inches) in length. This rating scheme is superior to an ordinal Likert scale because it enhances the reliability of the instrument by increasing the variability of responses (Schraw & Dennison, 1994; Weaver, 1990). Each participant’s scores on the individual scales was derived by summing all the items from that scale and
taking the average. Hence, each participant had eight composite scores, one for each of
the components of metacognition. The MAI has been used extensively as a measure of
domain-general metacognitive awareness; studies using this instrument have consistently
reported internal consistency reliability coefficients above 0.75 for all of the scales.
Moreover, the MAI has been validated via exploratory factor analyses with common
factor extraction methods with oblique rotations, and it has demonstrated good construct
validity (see Schraw & Dennison, 1994).

In spite of the argument that metacognition can be conceptualized as eight
dimensions—three under knowledge of cognition and five under the regulation of
cognition—Schraw and Dennison (1994) concluded that their two validation experiments
for the MAI demonstrated little support for an eight-scale instrument. Instead, their
results supported a two-factor solution separating metacognition into two components,
knowledge of cognition and regulation of cognition. Therefore the three knowledge
scales and the five regulation scales were collapsed to form one knowledge score and one
regulation score. For the present study, the two MAI scales demonstrated high internal
consistency reliability—Knowledge Scale, Cronbach’s alpha = 0.87; Regulation Scale,
Cronbach’s alpha = 0.92. This demonstrates that respondents provided consistent
responses across the two dimensions of metacognition as measured by the MAI,
suggesting low measurement error in the hypothesized constructs.

**Calibration**

Calibration accuracy was assessed using a continuous scale. Schraw (2009b)
posits that metacognitive judgments can be understood in terms of absolute and relative
accuracy. Absolute accuracy measures the precision of a confidence judgment vis-à-vis
performance on a criterion task. Absolute accuracy is the discrepancy between a confidence judgment and performance and it is obtained by computing the squared deviation between confidence and performance on the same scale. Smaller deviations correspond to better accuracy. This is a measure of “absolute” accuracy in the sense that a student’s confidence judgment is compared in an absolute fashion to his or her performance on the same task.

Participants were asked to make continuous confidence judgments on a 0-100 point scale on an item-by-item basis (see Appendix C) on a 100 millimeter bipolar scale (0% Confidence to 100% Confidence). This maintained a ratio scale rather than rely on the Gamma coefficient which is a 2x2 matrix of correct versus incorrect responses which is subsequently compared to either low or high confidence. Confidence scores were averaged across all items to obtain a confidence score composite. This score was subsequently compared against the proportion (i.e., percentage) of correct responses. Therefore, accuracy was evaluated by calculating the continuous difference score between the confidence judgment and actual performance (i.e., squared difference).

**Performance**

The performance outcome was measured using a declarative knowledge test adapted by the researcher with the permission of the instructors who developed the tests (see Appendix C). The test included 40 multiple choice items with four responses per item (a correct response and three distractors), which all participants completed. Having all participants complete the same test ensures equity and obviates potential confounds. The test covers topics pertaining to general psychology, such as classical conditioning, cognitivism, information processing, memory, and metacognition.
For analytic purposes, the correct response to each item was coded as “1” whereas the incorrect response was coded as “0”, thereby yielding a dichotomous coding scheme. Each participant’s raw scores at pretest and posttest were computed by summing the total across all items such that performance ranged from 0-20 at pretest and 0-40 at posttest. In an effort to facilitate pretest and posttest performance comparisons, raw scores were transformed to proportions of correct responses by dividing the number of correct responses by the total number of items and multiplying the total by 100 to obtain a percentage of items with a correct response. Cronbach’s alpha becomes the Kuder-Richardson (KR) 20 formula when assessing the internal consistency reliability of items with dichotomous response sets (SPSS, 2010). The KR 20 internal consistency reliability for the pretest was 0.60. The KR 20 reliability for the posttest items was as follows: first 20 items (i.e., same items as pretest) = 0.68; last 20 items = 0.62; and for all 40 items combined = 0.78. Reliability coefficients of ≥ 0.70 are considered good (Tabachnick & Fidell, 2007).

**Strategy Training Manipulation Fidelity Check**

A seven-item Strategy Training Fidelity Check scale developed by the researcher was used to ascertain the utility and relevance of the strategy training manipulation with respect to improving performance and calibration accuracy from the perspective of the participants randomized into the strategy training condition (see Appendix H). Sample items included, “The strategy training was clear and understandable.” and “Overall, I feel that the strategy training has adequately prepared me to increase the accuracy and confidence of my calibration of performance judgments.” Participants responded to the items on a 4-point Likert scale as follows: 1, “Strongly Disagree”; 2, “Disagree”; 3,
“Agree”; and 4, “Strongly Agree”. The higher the mean on the fidelity check, the higher the utility and relevance of the strategy training condition to the improvement of performance and calibration accuracy whereas a lower mean suggests lower utility and relevance. Internal consistency reliability for this scale was high, Cronbach’s alpha = 0.92.

**Procedures**

The appropriate permissions were obtained from the Psychology Department’s subject pool coordinator and the Experiment Management System administrator. Subsequently, an institutional review board (IRB) protocol was prepared and submitted; the university’s IRB approved the dissertation research study as “Exempt” (see Appendix D). Students recruited to the study were randomly assigned to either one of four groups—incentives, no incentives; strategy training or no training—making every effort to have equal sample sizes in each group. Table 5 contains a list of step-by-step activities that each of the four experimental groups experienced.
Table 5

Summary of Activities by Experimental Group

<table>
<thead>
<tr>
<th>Experimental Session</th>
<th>Strategy Training and Incentives</th>
<th>Strategy Training Only</th>
<th>Incentives Only</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 2</strong></td>
<td>1. PowerPoint presentation providing an overview of strategies 2. Each strategy will be separately introduced, discussed/ scaffolded, and demonstrated 3. After instruction, students will be provided psychology texts and a practice test unrelated to the study to apply and practice learned strategies 4. Students will be allowed to ask questions throughout 5. A brief summary of instruction will be provided at the end of the session</td>
<td>1. PowerPoint presentation providing an overview of strategies 2. Each strategy will be separately introduced, discussed/ scaffolded, and demonstrated 3. After instruction, students will be provided psychology texts and a practice test unrelated to the study to apply and practice learned strategies 4. Students will be allowed to ask questions throughout 5. A brief summary of instruction will be provided at the end of the session</td>
<td>Watch a psychology-related film</td>
<td>Watch a psychology-related film</td>
</tr>
<tr>
<td><strong>Session 3</strong></td>
<td>1. Receive a summary of strategy training completed in Session 2 2. Read stimulus 3. Complete posttest measures 4. Receive incentive if the ≥ 80% of items answered correctly criterion is met</td>
<td>1. Receive a summary of strategy training completed in Session 2 2. Read stimulus 3. Complete posttest measures</td>
<td>1. Read stimulus 2. Complete posttest measures 3. Receive incentive if the ≥ 80% of items answered correctly criterion is met</td>
<td>1. Read stimulus 2. Complete pretest measures</td>
</tr>
</tbody>
</table>
Session 1

All students completed the demographics form, MAI, and the performance assessment simultaneously in a one-hour session. This session served as a baseline pretest to establish group equivalence on general metacognitive awareness and performance as well as to gather calibration accuracy scores of performance prior to the experimental manipulations. Students reported to different classrooms based upon the random group to which they were assigned for experimental purposes. All students completed the same 20 of the 40 items of the performance assessment at pretest. Students were exposed to the same stimulus—a text containing information related to the test (Slavin, 2009)—prior to completing the test (see Appendix E). At the end of this session, students randomized to the strategy training and incentives group received an overview of the strategies that were covered in the one-hour training session as well as the instructions regarding incentives for posttest performance (i.e., ≥ 80% of the items correct). Namely, they were explicitly instructed that their payment would depend on how well they performed (i.e., better performance, as manifested by greater accuracy) at posttest: “Your pay for participation in the experiment WILL depend on how WELL you perform on the assessment at posttest. WELL is defined as getting ≥ 80% of the items correctly.”

The incentive involved a monetary reward of $10.00 contingent upon higher levels of calibration accuracy with respect to test performance at posttest. The 80% criterion was obtained from pilot study data of 76 undergraduates. A median percentage of 79 was calculated for performance on the test from the pilot study. The median score was selected because it is not susceptible to undue influence by outliers, as would be the case with the mean. As such, students needed to correctly respond to at least 80% of the
items on the performance assessment at posttest to receive the incentive. Students in the strategy training only group were given an overview of the one-hour training whereas those in the incentives only group were furnished instructions regarding the incentives for better posttest performance. Finally, students in the control group were excused after completion of all measures.

**Session 2**

Instruction for those individuals in the strategy training and incentives as well as the strategy training only groups occurred in a one-hour session. The strategy training component of the intervention involved providing students with instruction regarding more sophisticated and adaptive strategies that are more conducive to enhancing calibration accuracy with respect to performance. Table 4 includes a summary list of strategies that are part of the strategy training component of the intervention. The training session involved direct instruction and individual practice in using strategies with scaffolded feedback in a face-to-face lecture format.

First, students were provided with a brief introduction to the goal of the session and an overview of the types of strategies that would be covered. Next, the researcher covered each of the strategies separately. For each strategy, students were provided direct instruction that included explaining the strategy, identifying when it is applicable, and modeling as well as scaffolding the strategy so that students perceived its value with respect to improved calibration accuracy. Subsequently, students were provided opportunities to apply and practice each strategy covered during the session (see Appendix F) as well as a practice test (see Appendix G) to bolster their confidence in the application of the strategies. During this apply-practice part of each session, the
researcher walked around to provide additional guidance individually, where necessary. Students were afforded opportunities to ask questions and discuss strategies after they were introduced and modeled to clarify any misunderstandings.

In sum, the session focused on strategy training, practice, and reflection via informational feedback. Following the strategy training, students in both the strategy training and incentives and strategy training only groups completed a brief strategy training fidelity check survey (see Appendix H) intended to gauge participants’ overall evaluation of the strategy training intervention. Students in the incentives only and control groups participated in an activity unrelated to the present study.

This method of strategy training is warranted for two main reasons. First, the direct instruction affords students functional knowledge about strategies to enhance calibration accuracy regarding performance as well as when and why to apply them appropriately (conditional knowledge). Moreover, this approach furnishes students with an opportunity to actually practice and apply the newly internalized strategies as well as to receive scaffolding in the form of feedback and modeling.

**Session 3**

All students again read the stimulus text and completed the performance assessment simultaneously. Students were again placed in separate rooms. Those students in the strategy training and incentives and the strategy training only groups received an overview/summary of strategies covered during the training session (Session 2) before completing the assessment. All students first read the same stimulus as they were exposed to at pretest and completed the same 20 items on the performance assessment they completed at pretest as well as an additional 20 items to counter any potential testing
effects. All students, regardless of group, had the same time frame between the two data collection points to further control for any potential confounds.

**Data Analysis**

In order to address the objective of this study, a one-way analysis of covariance (ANCOVA) was conducted, with posttest performance serving as the dependent variable while statistically controlling for baseline performance. Moreover, two 2 (strategy training, no training) x 2 (incentives, no incentives) factorial mixed-model analyses of variance (ANOVAs) were conducted, with type of test (pretest, posttest) serving as the within-subjects factor. Confidence and calibration accuracy each served as a dependent variable in a separate analysis in keeping with the data-analysis strategy in previous experiments.

Furthermore, a series of repeated measures analyses of variance (RM ANOVAs) were conducted to compare performance on the pretest (i.e., 20) and posttest (i.e., all 40) items. The first RM ANOVA compared performance on the 20 items given at pretest to those same 20 items given at posttest. The second RM ANOVA compared the performance on the 40 items given at posttest—that is, the previous 20 items given at pretest with the 20 items added at posttest. However, whereas the first analysis compared performance between pretest and posttest on the same 20 items, the second analysis was for posttest performance only—20 previous items compared to the 20 additional items. Correlations among all outcome variables were computed and reported as well.

Data screening and assumption testing procedures proceeded by splitting the file by group and conducting these procedures for each group separately for each of the variables under consideration. This method is more accurate inasmuch as data screening
and assumption testing for the entire sample is meaningless when conducting between-subjects analyses, as in the present study. Data screening was done by requesting box plots by group. Deletion of outliers from the dataset is preferred over transformation because transforming the variables in an attempt to normalize data complicates interpretation because the data is no longer in its original scale (Tabachnick & Fidell, 2007); however, deletion of outliers may not be possible in situations in which deleting the outliers would lead to a severe loss of power— that is, datasets with smaller numbers of cases.

Tabachnick and Fidell (2007) warn that leaving outliers untreated is inappropriate because they unduly influence group means with respect to the outcome variables, and thus, they lead to results that are misleading and inaccurate. Reporting such misleading results is an unethical practice. For the present study, data screening detected no outliers that would undermine the trustworthiness of the data. Furthermore, the data met all requisite assumptions, including normality (all skewness and kurtosis values were < |2|) and homogeneity of error variance (all $p$-values were > .05 for Levene’s Test) for each of the outcome variables by group as well as homogeneity of regression coefficients and sphericity. Therefore, data analysis proceeded as planned with all 107 cases.
CHAPTER 4
RESULTS

Participant Attrition at Posttest

In order to strengthen the trustworthiness of the data, analyses were conducted to verify the absence of non-random sampling bias as well as to establish equality of outcome means at pretest—in which data were available for all 160 participants—between participants who remained in the study and participants who were lost to attrition at posttest. These analyses were conducted prior to any data analysis with respect to the research question of the present study. Results of these preliminary analyses follow.

Little’s MCAR $\chi^2$ Test

In an effort to verify that the missing data pattern was missing completely at random (MCAR), Little's MCAR $\chi^2$ statistics (Little & Rubin, 1989; Schaeffer & Graham, 2002) were requested from the missing values analysis for each group separately. A significant $\chi^2$ (i.e., $p < 0.05$) would suggest that the pattern of missing data is not MCAR (i.e., missing not at random (MNAR)], which poses a problem for interpretation of results because they may be biased due to systematic differences in non-responses. However, for the present study, all results were not statistically significant, all $p$-values $\geq 0.23$, suggesting that the missingness pattern in the data was MCAR.

Equality of Outcome Means at Baseline

A series of independent samples $t$-tests were conducted to ascertain whether there were any significant differences between those who completed the study and those who were lost to attrition. All $p$-values were adjusted accordingly using the Bonferroni
adjustment to obviate the inflation of familywise Type I error rate. There were no statistically significant differences between the two aforementioned groups with respect to any of the outcome measures at pretest, all $p$-values $\geq 0.19$, suggesting the absence of non-random sampling bias in the data. Given these results, data analysis proceeded as planned with the 107 complete cases.

**Primary Analyses of the Present Study**

**Equality Among Groups**

A series of one-way ANOVAs were conducted to establish equivalence among the groups on the various baseline measures. For all analyses, strategy training (training, no training) and incentives (incentives, no incentives) served as the independent variables. All $p$-values were adjusted accordingly by analysis using the Bonferroni adjustment to obviate the inflation of familywise Type I error rate.

**Metacognitive awareness.** Results demonstrated that there were no statistically significant differences among the groups in either condition with respect to the knowledge and regulation components of metacognitive awareness, all $p$-values $\geq 0.22$. Although not significantly different, the strategy training group reported the highest knowledge mean score ($M = 74.61, SD = 10.08$) whereas the no incentives group reported the highest regulation mean score ($M = 68.10, SD = 11.49$) when compared to the other groups. Due to the lack of significance between the groups on these measures, neither metacognitive awareness component was included as a covariate in the remaining analyses.

**Performance.** There were statistically significant differences with respect to performance between the groups for the strategy training condition, $F_{(1,105)} = 3.96, p =$
.02, $\eta^2 = 0.05$, with the strategy training group ($M = 67.58, SD = 12.57$) outperforming the no training group ($M = 62.98, SD = 11.02$), and the incentives condition, $F_{(1,105)} = 4.06, p = .02, \eta^2 = 0.05$, in which the incentives group ($M = 67.77, SD = 12.42$) outperformed the no incentives group ($M = 63.13, SD = 11.31$). Therefore, pretest performance was statistically controlled in the analysis of posttest performance.

Confidence. Results showed that there were no statistically significant differences on confidence judgments at baseline between the groups in either condition, all $p$-values $\geq 0.44$. Therefore, pretest confidence judgment score was not statistically controlled in subsequent analyses.

Descriptive Statistics and Correlations

Descriptive statistics (e.g., means and standard deviations) for all outcome measures are reported by group in Table 6. Zero-order correlations are reported for the strategy training group and the no training group in Table 7; Table 8 presents the correlations of the incentives group and the no incentives group.
Table 6

*Descriptive Statistics of Outcome Measures by Group*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Group 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Group 3&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Group 4&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>MAI K</td>
<td>74.61</td>
<td>10.08</td>
<td>72.09</td>
<td>11.10</td>
</tr>
<tr>
<td>MAI R</td>
<td>67.16</td>
<td>10.81</td>
<td>66.94</td>
<td>11.91</td>
</tr>
<tr>
<td>PreP</td>
<td>67.58</td>
<td>12.57</td>
<td>62.93</td>
<td>11.01</td>
</tr>
<tr>
<td>PreC</td>
<td>67.15</td>
<td>14.40</td>
<td>65.32</td>
<td>10.33</td>
</tr>
<tr>
<td>PostP1</td>
<td>75.00</td>
<td>12.89</td>
<td>69.21</td>
<td>12.84</td>
</tr>
<tr>
<td>PostP2</td>
<td>71.26</td>
<td>10.84</td>
<td>64.54</td>
<td>13.56</td>
</tr>
<tr>
<td>PostP3</td>
<td>73.13</td>
<td>9.88</td>
<td>66.87</td>
<td>11.85</td>
</tr>
<tr>
<td>PostC</td>
<td>73.58</td>
<td>10.60</td>
<td>71.66</td>
<td>12.54</td>
</tr>
<tr>
<td>Calibration2</td>
<td>7.84</td>
<td>5.20</td>
<td>9.65</td>
<td>6.75</td>
</tr>
</tbody>
</table>

*Note.* MAI K=MAI Knowledge Scale; MAI R=MAI Regulation Scale; PreP=Pretest Performance; PreC=Pretest Confidence; PostP1=Posttest Performance First 20 Items Percentage; PostP2=Posttest Performance Latter 20 Items Percentage; PostP3=Posttest Performance All 40 Items Percentage; PostC=Posttest Confidence; Calibration1=Pretest Calibration Accuracy; Calibration2=Posttest Calibration Accuracy.

<sup>a</sup> Strategy training, *n*=60; <sup>b</sup> No training, *n*=47.
<sup>c</sup> Incentives, *n*=56; <sup>d</sup> No incentives, *n*=51.
Table 7

Zero-Order Correlations for the Strategy Training and No Training Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MAI K</td>
<td>-</td>
<td>.64**</td>
<td>-.10</td>
<td>.13</td>
<td>-.15</td>
<td>-.04</td>
<td>-.12</td>
<td>.07</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>2. MAI R</td>
<td>.78**</td>
<td>-</td>
<td>-.13</td>
<td>.15</td>
<td>-.26*</td>
<td>-.16</td>
<td>-.26*</td>
<td>-.11</td>
<td>.16</td>
<td>.00</td>
</tr>
<tr>
<td>3. PreP</td>
<td>-.03</td>
<td>-.13</td>
<td>-</td>
<td>.23</td>
<td>.71**</td>
<td>.47**</td>
<td>.72**</td>
<td>.36</td>
<td>-32</td>
<td>-26*</td>
</tr>
<tr>
<td>4. PreC</td>
<td>.14</td>
<td>.20</td>
<td>.50**</td>
<td>-</td>
<td>.16</td>
<td>.08</td>
<td>.15</td>
<td>.51**</td>
<td>-.20</td>
<td>-.21</td>
</tr>
<tr>
<td>5. PostP1</td>
<td>-.06</td>
<td>-.07</td>
<td>.63**</td>
<td>.40**</td>
<td>-</td>
<td>.38**</td>
<td>.86**</td>
<td>.47**</td>
<td>-.31</td>
<td>-.39**</td>
</tr>
<tr>
<td>6. PostP2</td>
<td>.04</td>
<td>-.10</td>
<td>.51**</td>
<td>.30*</td>
<td>.61**</td>
<td>-</td>
<td>.80**</td>
<td>.57**</td>
<td>-.23</td>
<td>-.19</td>
</tr>
<tr>
<td>7. PostP3</td>
<td>-.01</td>
<td>-.09</td>
<td>.64**</td>
<td>.39**</td>
<td>.89**</td>
<td>.90**</td>
<td>-</td>
<td>.62**</td>
<td>-.33</td>
<td>-.36**</td>
</tr>
<tr>
<td>8. PostC</td>
<td>.12</td>
<td>.16</td>
<td>.36**</td>
<td>.64**</td>
<td>.59**</td>
<td>.61**</td>
<td>.67**</td>
<td>-</td>
<td>-.44**</td>
<td>-.31*</td>
</tr>
<tr>
<td>9. Calibration1</td>
<td>-.04</td>
<td>.09</td>
<td>-.32*</td>
<td>.17</td>
<td>-.14</td>
<td>-.30*</td>
<td>-.25</td>
<td>-.02</td>
<td>-</td>
<td>.46*</td>
</tr>
<tr>
<td>10. Calibration2</td>
<td>.02</td>
<td>.16</td>
<td>-.41**</td>
<td>.13</td>
<td>-.40**</td>
<td>-.52**</td>
<td>-.15</td>
<td>-.01</td>
<td>.68**</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note.* Correlations above the diagonal are for the strategy training group² and those below the diagonal are for the no training group². MAI K=MAI Knowledge Scale; MAI R=MAI Regulation Scale; PreP=Pretest Performance; PreC=Pretest Confidence; PostP1=Posttest Performance First 20 Items Percentage; PostP2=Posttest Performance Latter 20 Items Percentage; PostP3=Posttest Performance All 40 Items Percentage; PostC=Posttest Confidence; Calibration1=Pretest Calibration Accuracy; Calibration2=Posttest Calibration Accuracy.

²n=60; ²n=47

*Ap<.05 **Ap<.01 (two-tailed)
### Table 8

Zero-Order Correlations for the Incentives and No Incentives Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MAI K</td>
<td>-.71**</td>
<td>-.04</td>
<td>.09</td>
<td>-.21</td>
<td>.09</td>
<td>-.05</td>
<td>.10</td>
<td>-.03</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>2. MAI R</td>
<td>.73**</td>
<td>-.00</td>
<td>.02</td>
<td>-.20</td>
<td>-.04</td>
<td>-.13</td>
<td>-.02</td>
<td>-.02</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>3. PreP</td>
<td>-.04</td>
<td>-.24</td>
<td>-.49**</td>
<td>.73**</td>
<td>.51**</td>
<td>.71**</td>
<td>.42**</td>
<td>-.05</td>
<td>-.20</td>
<td></td>
</tr>
<tr>
<td>4. PreC</td>
<td>.22</td>
<td>.35*</td>
<td>.13</td>
<td>-</td>
<td>.51**</td>
<td>.33*</td>
<td>.48**</td>
<td>.68**</td>
<td>-.24</td>
<td>-.20</td>
</tr>
<tr>
<td>5. PostP1</td>
<td>.06</td>
<td>-.09</td>
<td>.63**</td>
<td>.06</td>
<td>-</td>
<td>.49**</td>
<td>.84**</td>
<td>.64**</td>
<td>-.08</td>
<td>-.40**</td>
</tr>
<tr>
<td>6. PostP2</td>
<td>-.04</td>
<td>-.20</td>
<td>.45**</td>
<td>-.03</td>
<td>.51**</td>
<td>-</td>
<td>.88**</td>
<td>.65**</td>
<td>-.08</td>
<td>-.21</td>
</tr>
<tr>
<td>7. PostP3</td>
<td>.01</td>
<td>-.16</td>
<td>.63**</td>
<td>.02</td>
<td>.89**</td>
<td>.85**</td>
<td>-</td>
<td>.75**</td>
<td>.00</td>
<td>-.35*</td>
</tr>
<tr>
<td>8. Calibration1</td>
<td>.14</td>
<td>.24</td>
<td>-.42**</td>
<td>.13</td>
<td>-.18</td>
<td>-.39**</td>
<td>-.32*</td>
<td>-.23</td>
<td>-</td>
<td>.39**</td>
</tr>
<tr>
<td>9. Calibration2</td>
<td>.14</td>
<td>.24</td>
<td>-.42**</td>
<td>.13</td>
<td>-.18</td>
<td>-.39**</td>
<td>-.32*</td>
<td>-.23</td>
<td>-</td>
<td>.39**</td>
</tr>
<tr>
<td>10. Calibration2</td>
<td>.00</td>
<td>.17</td>
<td>-.46**</td>
<td>.06</td>
<td>-.39**</td>
<td>-.58**</td>
<td>-.54**</td>
<td>.04</td>
<td>.51**</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note.* Correlations above the diagonal are for incentives group\(^a\) and those below the diagonal are for no incentives group\(^b\). MAI K=MAI Knowledge Scale; MAI R=MAI Regulation Scale; PreP=Pretest Performance; PreC=Pretest Confidence; PostP1=Posttest Performance First 20 Items Percentage; PostP2=Posttest Performance Latter 20 Items Percentage; PostP3=Posttest Performance All 40 Items Percentage; PostC=Posttest Confidence; Calibration1=Pretest Calibration Accuracy; Calibration2=Posttest Calibration Accuracy.

\(^a\)\(n=56;\) \(^b\)\(n=51\)

*\(p<.05\) **\(p<.01\) (two-tailed)

### Strategy Training Intervention Fidelity Check

The strategy training fidelity check was administered to ascertain whether participants who received the strategy training perceived the strategy training to be useful and utilitarian with respect to improved performance, confidence, and calibration accuracy. Results demonstrated that both groups \((M = 3.51, SD = 0.47)\) perceived the training to be highly useful, relevant, and utilitarian in helping them to improve
performance, confidence, and calibration accuracy, with the strategy training only group rating the training more favorably ($M = 3.63, SD = 0.37$) than the strategy training and incentives group ($M = 3.38, SD = 0.56$). An independent samples $t$-test demonstrated that this difference in rating between the groups was not statistically significant, $p = 0.06$.

### Comparison Between Pretest and Posttest Performance

Results of the first RM ANOVA between pretest performance and posttest performance on the same 20 items demonstrated that there were statistically significant differences across time, $F_{(1,106)} = 51.02, p < 0.0005, \eta^2 = 0.33$, with participants performing better at posttest ($M = 72.45, SD = 13.13$) than at pretest ($M = 65.56, SD = 12.08$), as expected due to the role of feedback. The subsequent RM ANOVA results between the first 20 items at posttest—the same that were given at pretest—and the 20 items that were unique to the posttest showed statistically significant differences, $F_{(1,106)} = 11.71, p < 0.01, \eta^2 = 0.10$, with participants exhibiting increased performance in the first 20 items ($M = 72.45, SD = 13.13$) when compared to the new items ($M = 68.31, SD = 12.51$) (i.e., new items participants had never been exposed to).

### Evaluating the Effectiveness of the Interventions

For the remaining analyses regarding the strategy training and incentives interventions, performance was the only pretest measure to exhibit significant differences between the four groups, whereas confidence and calibration accuracy did not. Hence, pretest performance was the only variable used as a covariate in the analysis of posttest performance.
Performance

The strategy training x incentives interaction with respect to posttest performance while controlling for pretest performance was statistically significant, $F_{(1,102)} = 6.61, p = 0.03, \eta^2_p = 0.12$. Simple main effects follow up analyses were requested with the Bonferroni adjustment for the inflation of familywise Type I error rate. The strategy training x incentives interaction within strategy training demonstrated that those who received strategy training and extrinsic incentives ($M = 74.83, SD = 9.03$) performed significantly better than those who received the incentives only with no training ($M = 68.47, SD = 8.51$), $F_{(1,102)} = 7.26, p = 0.01, \eta^2 = 0.15$. The difference between those who received strategy training and no training with no incentive was not statistically significant, $p = 0.69$. The incentives simple main effects demonstrated that there were significant differences between those who received incentives and no incentives and also received strategy training, $F_{(1,102)} = 9.79, p = 0.03, \eta^2 = 0.09$, with those who received incentives performing better than those who did not. The difference between those who received incentives and no incentives with no training was not statistically significant, $p = 0.52$.

In addition, main effects were interpreted in the presence of an ordinal interaction. The strategy training main effect was statistically significant, $F_{(1,102)} = 4.59, p < 0.05, \eta^2 = 0.07$, with the strategy training group ($M = 72.65, SD = 9.88$) outperforming the no training group ($M = 68.37, SD = 11.85$) at posttest. The incentive main effect was also statistically significant, $F_{(1,102)} = 6.52, p = 0.01, \eta^2 = 0.08$, with those in the extrinsic incentives condition ($M = 71.97, SD = 10.66$) performing better at posttest than those in
the no incentive condition \((M = 68.05, SD = 10.66)\) (see Table 9 for adjusted means, unadjusted means, and standard deviations).

Interestingly, approximately 75\% of the participants who received the extrinsic incentive for improved posttest performance (i.e., \(\geq 80\%\) of the items correct at posttest) were from the strategy training and incentives group whereas only about 25\% were from the incentives only group.

Table 9

<table>
<thead>
<tr>
<th>Group 1(^a)</th>
<th>Group 2(^b)</th>
<th>Group 3(^c)</th>
<th>Group 4(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M^1)</td>
<td>(M^2)</td>
<td>(SD)</td>
<td>(M^1)</td>
</tr>
<tr>
<td>72.65</td>
<td>73.13</td>
<td>9.88</td>
<td>68.37</td>
</tr>
</tbody>
</table>

\(^a\) Strategy training, \(n=60\); \(^b\) No training, \(n=47\); \(^c\) Incentives, \(n=56\); \(^d\) No incentives, \(n=51\)

1 Adjusted means; 2 Unadjusted means

Confidence

Results of the factorial mixed-model ANOVA with confidence as the dependent variable demonstrated that the strategy training x incentive x time interaction was statistically significant, \(F_{(1,103)} = 6.26, p = 0.01, \eta^2_p = 0.10\). Simple effects were requested following the significant interaction. An inspection of the simple effects of the strategy training x incentives x time interaction within strategy training demonstrated that the difference between the strategy training group \((M = 76.88, SD = 11.31)\) and the no training group \((M = 69.56, SD = 8.18)\) which received incentives, was statistically significant, \(F_{(1,103)} = 7.65, p = 0.01, \eta^2 = 0.07\), with the strategy training group exhibiting greater confidence than the no training group at posttest. All other group comparisons were not statistically significant, all \(p\)-values \(\geq 0.11\). The simple effects analysis of
incentives was statistically significant, $F_{(1,103)} = 6.49, p = 0.01, \eta^2 = 0.09$, with those in the extrinsic incentives condition who received strategy training exhibiting higher confidence at posttest than those who did not receive the incentive. Those who received incentives and no strategy training ($M = 74.64, SD = 9.48$) also showed greater confidence at posttest than those who did not receive the incentive ($M = 68.56, SD = 14.67$), $F_{(1,103)} = 4.54, p = 0.04, \eta^2 = 0.06$. All other mean comparisons were not significant, all $p$-values $\geq 0.41$.

The simple effects results of time demonstrated that the change in confidence across time was statistically significant for: the strategy training group that received incentives (Pretest: $M = 66.82, SD = 14.93$; Posttest: $M = 76.88, SD = 11.31$), $F_{(1,103)} = 26.35, p < 0.0005, \eta^2 = 0.21$; the no training group which received incentives (Pretest: $M = 63.72, SD = 12.58$; Posttest: $M = 68.56, SD = 14.67$), $F_{(1,103)} = 4.26, p = 0.04, \eta^2 = 0.06$; and the no training group that received no incentives (Pretest: $M = 66.85, SD = 7.56$; Posttest: $M = 74.64, SD = 9.48$), $F_{(1,103)} = 11.50, p = 0.001, \eta^2 = 0.12$. The greatest gain in confidence across time was achieved by the strategy training group which also received incentives. The change in confidence across time for the strategy training group that received no incentives was not significant, $p = 0.36$.

**Calibration**

The factorial mixed-model ANOVA results with calibration accuracy as the dependent variable demonstrated a statistically significant ordinal strategy training x time interaction, $F_{(1,103)} = 25.37, p < 0.0005, \eta^2 = 0.20$. Neither the strategy training x incentive x time ($p = 0.18$) interaction nor the incentives x time interaction ($p = 0.70$) was significant. A review of the simple effects of strategy training demonstrated that the
difference between the strategy training group ($M = 13.78$, $SD = 9.43$) and the no training group ($M = 8.50$, $SD = 7.25$) was statistically significant at pretest, $F_{(1,103)} = 10.86$, $p = 0.001$, $\eta^2 = 0.11$, with the strategy training group exhibiting the lowest overall calibration accuracy at pretest. Moreover, the difference between the strategy training group ($M = 7.84$, $SD = 5.20$) and the no training group ($M = 9.65$, $SD = 6.75$) at posttest was significant, $F_{(1,103)} = 5.24$, $p = 0.03$, $\eta^2 = 0.09$, with the strategy training group exhibiting greater calibration accuracy than the no training group; the strategy training group also showed the highest overall calibration accuracy at posttest.

The simple effects results of time demonstrated that the change in calibration accuracy across time was statistically significant for the strategy training group (Pretest: $M = 13.97$, $SD = 9.42$; Posttest: $M = 7.90$, $SD = 5.49$), $F_{(1,103)} = 40.79$, $p < 0.0005$, $\eta^2 = 0.28$, with this group showing significantly improved calibration accuracy at posttest as well as the greatest improvement in calibration accuracy across time. The change in calibration accuracy across time for the no training group was not significant, $p = 0.29$.

Significant main effects for time were also interpreted in the presence of an ordinal interaction. The time main effect was statistically significant, $F_{(1,103)} = 11.92$, $p < 0.01$, $\eta^2 = 0.13$, with the strategy training group demonstrating significantly improved calibration accuracy at posttest (Pretest: $M = 13.78$, $SD = 9.43$; Posttest: $M = 7.83$, $SD = 5.20$); neither the no training group nor the incentives and no incentives groups exhibited a significant change in calibration accuracy across time, albeit those in the incentives group exhibited higher accuracy at posttest than those in the no incentives group.
CHAPTER 5
DISCUSSION

The main research objective of the present investigation was to examine the separate and interactive effects of two interventions with respect to performance, metacognitive confidence judgments, and calibration accuracy within the theoretical framework of the NNMM. Individuals were assigned randomly to conditions. The first intervention involved metacognitive strategy training in which undergraduate students received instruction on the seven strategies outlined in Table 4 for one hour. The second intervention consisted of an incentive manipulation in which individuals were told, if they performed at or better than 80% (i.e., at least 80% of the items answered correctly on the posttest performance assessment), they would receive a cash payment for their posttest performance. Evaluation of the influence of these four conditions—strategy training, no training; incentives, no incentives—on confidence, performance, and calibration accuracy was the primary goal of the present study.

I predicted that each intervention would have a positive effect on the aforementioned constructs with respect to the monitoring and control processes described in the NNMM. I expected the strategy training intervention to improve monitoring and the incentives to improve control through added attention to the use of these processes to self-regulate learning. Providing students with metacognitive training increases their comprehension monitoring, as they reflect and become more aware of their own internal cognitive processes. Extrinsic rewards, on the other hand, increase students’ motivation to more closely attend to the task. These interventions, therefore, aided students in more
precisely feeling what they know and what they do not know, thereby improving performance, confidence, and calibration accuracy.

With respect to the interaction between the strategy training and incentives manipulations and the type of test, I expected that those randomized into those conditions would exhibit more accurate calibration and increased performance and confidence at posttest when compared to those who were not. Moreover, I predicted that the proposed strategy training and incentives enhance the accuracy of students’ metacognitive judgments, presumably due to improved performance from baseline to posttest. Finally, students’ performance on learning tasks, such as tests/exams, was predicted to increase.

**Research Hypotheses**

**H₁:** With respect to the combination of the strategy training and incentives manipulations, I hypothesized that the interactive effect would yield more accurate calibration, higher performance, and higher levels of confidence at posttest for the groups that were exposed to the combined interventions. When combined, strategy training and incentives are expected to positively influence both the information-gathering (monitoring) and control processes of the NNMM. These more effective monitoring and control processes in turn lead to better performance, presumably due to increased accuracy (i.e., more accurate confidence judgments).

**H₂:** The proposed strategy training was predicted to enhance the calibration accuracy of students’ metacognitive judgments at posttest.

**H₃:** Incentives were expected to positively influence calibration in such a way as to enhance accuracy.
Moreover, students’ performance on learning tasks, such as tests/exams, was expected to increase for those exposed to the strategy training and incentives interventions. As students’ monitoring and control is enhanced due to the effects of the strategy training and incentives, as proposed by the NNMM, their subsequent performance on learning tasks is likely to improve as well. Enhanced information-gathering capabilities increase metacognitive awareness because the metal-level’s model of the object-level is more precise. As a result of this more accurate model, students’ control capabilities (e.g., applying appropriate and effective strategies while learning) are concomitantly enhanced, which consequently leads to improved performance.

**Performance Outcomes**

The interaction between the metacognitive strategy training and incentives was statistically significant with respect to posttest performance after adjusting for the effects of pretest performance. Those in the strategy training condition who also received incentives performed significantly better at posttest when compared to those who received training and no incentives. This supports the first prediction (H1) that strategy training and extrinsic incentives combined to have an interactive effect regarding posttest performance, as neither manipulation alone—either strategy training or incentives—appear to be as effective as the two combined, with the combination of the manipulations yielding nearly double the effect size than either alone. The fourth hypothesis (H4) also received additional support from the main effects results, which indicated that those who received strategy training outperformed those who received no training; the same pattern was found in the incentives manipulation, in which those who received incentives performed significantly better than those who did not. Moreover, a larger majority of
those who received the incentive also received strategy training, further supporting H1. Therefore, the data support the notion that metacognitive strategy training and incentives positively influence student performance. These findings are in line with previous research on strategy training and incentives with respect to performance (e.g., Bol et al., 2005; Hacker et al., 2008; Nietfeld & Schraw, 2002; Swanson, 1990; Yates, 1990).

Confidence Outcomes

The hypothesis with respect to confidence judgments (H1) also received overwhelming support from the data. The groups which received strategy training and incentives only demonstrated an increase in confidence judgments from pretest to posttest. Interestingly, however, participants in the group which received neither manipulation (i.e., the control group) also exhibited a gain in confidence across time, although their confidence judgments were misaligned with their performance, thus showing high miscalibration.

Equally as important, the strategy training and incentives group demonstrated the highest improvement in confidence judgments from pretest to posttest, suggesting that the combined effect of the strategy training and incentives manipulations was more powerful than either one alone, especially given that the change in confidence across time was not significant for the group that received strategy training with no incentives. The fact that the strategy training only group did not have a significant increase in confidence levels at posttest suggests that the incentives manipulation was more successful at improving confidence from pretest to posttest by enhancing the value of the strategy training manipulation. In terms of the NNMM, the extrinsic incentives augmented students’ comprehension monitoring (i.e., information-gathering capabilities) with respect to test
items, which allowed them to make adjustments to their confidence (control). In sum, the strategy training and incentives manipulations were effective at improving confidence from pretest to posttest, which is supported by the body of literature on confidence judgments (e.g., Koriat & Levy-Sadot, 2001; Kruger & Dunning, 1999; Maki et al., 1990; Mitchum & Kelley, 2010; Shynkaruk & Thompson, 2006).

Calibration Outcomes

Hypothesis 1 (H1) pertaining to the interaction between strategy training and incentives with respect to calibration accuracy did not receive support from the data, as only the metacognitive strategy training manipulation significantly improved calibration accuracy at posttest. However, this finding fully supports hypothesis 2 (H2). It is noteworthy to mention that the strategy training group demonstrated the poorest calibration accuracy at baseline when compared to the other three groups, including the no incentives/no training group. However, the strategy training group not only demonstrated improved calibration accuracy at posttest when compared to the no training group, but it also showed the greatest improvement in calibration accuracy overall. In contrast, the change in calibration accuracy across time for the no training group was not significant. Therefore, the strategy training group displayed significantly improved calibration accuracy at posttest as well as the greatest improvement in calibration accuracy across time, thus strongly supporting H2.

Overall, extrinsic incentives did not significantly aid in the improvement of calibration accuracy, which supports the general findings of Hogarth et al. (1991) and Yates (1990); hence, hypothesis 3 (H3) found no support from the data. Nevertheless, the strategy training significantly increased calibration accuracy (H2). These findings are
congruent with research conducted by Bol et al. (2005) and Nietfeld and his colleagues (Nietfeld et al., 2006; Nietfeld & Schraw, 2002).

**Conclusions**

There are several important conclusions that can be drawn from the results of the present study. First, external manipulations, such as metacognitive strategy training and extrinsic incentives, are successful at improving students’ performance as well as confidence in their performance. Previous research studies utilizing similar interventions have found that some form of instruction or training can successfully help learners to perform better due to increased confidence in what they know and do not know about the topic, presumably because of increased calibration accuracy in their confidence with respect to actual performance (e.g., Dunlosky & Nelson, 1992; Hacker et al., 2008; Nelson & Dunlosky, 1991; Nelson & Narens, 1990, 1994; Nietfeld et al., 2006; Nietfeld & Schraw, 2002).

The research on the use of extrinsic incentives, however, has not been as definitive, as some studies have found that extrinsic rewards aid in improving performance and confidence (e.g., Hacker et al., 2008; Schraw et al., 1993; Yates, 1990) whereas others have found their effects to be inconsequential (e.g., Hogarth et al., 1991). The present research supported the former group of studies, as incentives were effective at improving performance and level of confidence in feeling-of-knowing judgments. Extrinsic incentives enhance students’ control capabilities by motivating them to slow down their cognitive processing, thus allowing them to more closely attend to what test items are actually asking. Consequently, they are able to better comprehend, yielding not only increased performance but increased levels of confidence through increased control.
Nevertheless, very few studies have combined the use of instruction or training and extrinsic incentives (e.g., Bol et al., 2005; Hacker et al., 2008) with the express purpose of improving performance and/or confidence judgments. Therefore, the present study demonstrated that combining these two interventions can effectively help undergraduate students to exhibit better performance and increased confidence.

Moreover, extrinsic incentives do not seem to affect calibration accuracy, which is supported by work conducted by Hacker et al. (2008) and Hogarth et al. (1991). The explanations as to why incentives assist in the improvement of performance and confidence yet not calibration accuracy remain unclear. One possible explanation is that incentives focus the learner’s attention on performance outcomes rather than confidence or calibration accuracy per se. More specifically, providing instructions that improved performance would lead to an extrinsic incentives may have focused students’ attention directly on performance, and perhaps only implicitly on confidence and calibration accuracy. Therefore, although it led to higher performance at posttest, incentives may not have necessarily led to improved performance because students’ confidence judgments relative to performance may still have been somewhat inaccurate. The inclusion of calibration bias may help clarify this matter.

The metacognitive strategy training component of the study, however, was effective at improving the calibration accuracy of undergraduate students from pretest to posttest, which supports work done by Nietfeld and associates (e.g., Nietfeld et al., 2006; Nietfeld & Schraw, 2002) but runs counter to the research of Hacker and his colleagues (Bol et al., 2005; Hacker et al., 2008), who found no significant differences between the training and no training group. However, it is important to note that, unlike the present
study, these two latter studies were conducted in a quasi-experimental framework, which may have created issues of internal validity and help explain the lack of significance between-groups with respect to calibration accuracy.

In sum, although previous research on calibration accuracy using similar interventions has yielded somewhat mixed findings, especially as they pertain to accuracy of performance, the present study supports the contention that metacognitive strategy training and extrinsic rewards can be invoked in college classrooms to improve not only the performance and confidence in performance judgments of undergraduates but the accuracy of those judgments vis-à-vis performance as well.

Theory and the Advancement of Knowledge

Contributions to Theory

The NNMM has been the prevailing theoretical framework in the literature on calibration since its inception in 1990. However, the NNMM has recently fallen under heavy criticism by calibration scholars and researchers as simplistic and lacking (e.g., Boekaerts & Razendaal, 2010; Efklides, 2011). For instance, the model treats metacognition as essentially a unidimensional construct, and it does not adequately distinguish whether and how students process correct and incorrect responses differently (Schraw, Kuch, & Gutierrez, 2012). This distinction is necessary for the calibration process, especially as students gauge levels of confidence in their performance for purposes of determining accuracy. Nevertheless, with little empirical evidence with respect to the superiority of alternative frameworks, the NNMM is appropriate to help explain the findings of the present study.
The results of this study suggest that the metacognitive strategy training intervention enhances the monitoring and control processes in the NNMM. The strategies learned during the training aided students in being more alert and attentive to the task by deliberately slowing down cognitive processing, which improves reflection and awareness of one’s internal cognitive processes. This, in turn, allowed students to more accurately and effectively gather information regarding the assessment, such as what the performance assessment items were truly asking. Consequently, they were able to more precisely feel what they knew and did not know about the concepts the items were tapping, and thus, were better able to respond to the items and adjust confidence accordingly to yield better accuracy, when their confidence judgments were compared to actual performance. This clearly shows how the strategy training influenced students’ monitoring capabilities, which is congruent with the NNMM.

Offering students a reward as an incentive to rehearse the metacognitive strategies led to significantly improved performance and the level of confidence of their metacognitive judgments (e.g., Hacker et al., 2008; Hogarth et al., 1991; Nietfeld & Schraw, 2002; Swanson, 1990; Yates, 1990). As predicted, incentives increased students’ motivation to more meaningfully learn strategies provided during training. Therefore, incentives positively affected students’ information-gathering capabilities by permitting them to learn strategies to improve calibration accuracy in a more in-depth manner, thereby enhancing the monitoring function of the NNMM. Furthermore, incentives were found to influence students’ performance. It appears that incentives helped students to apply the strategies to actually improve their performance and degree of confidence, thus positively influencing the control function of the NNMM.
In sum, enhanced information-gathering capabilities increased metacognitive awareness because the metal-level’s model of the object-level was more precise for the students exposed to the manipulations. As a result of this more accurate model, students’ control capabilities (e.g., applying appropriate and effective strategies while learning) were concomitantly enhanced, which consequently led to improved performance, presumably due to increased accuracy.

**Advancement of Knowledge on Calibration**

Few studies (Bol et al., 2005; Hacker et al., 2008) have combined multiple external interventions to evaluate their effects on calibration components (e.g., accuracy and level of confidence), which for the present study were metacognitive strategy training and extrinsic incentives. One of the contributions of this study is adding evidence in support of the NNMM as a vehicle to explain phenomena related to calibration accuracy. In addition, the strategy training intervention was innovative insofar as the strategies used for the training were drawn directly from a pilot study involving undergraduate students, rather than existing strategies pulled from previous research which may not have been intended for use with adult populations such as undergraduate students. The direct relevance of these strategies to the intended population in the present study may have contributed to the significant influence of the training intervention on outcomes of interest when compared to other studies that have failed to detect significant differences between students who were trained and those who were not (e.g., Bol et al., 2005; Hacker et al., 2008).

Finally, the study has added evidence in support of the positive influence of metacognitive strategy training and extrinsic rewards on performance and confidence,
and the positive effect of training on calibration accuracy. This is crucial to research on calibration of performance, as findings have been mixed with respect to strategy training and extrinsic rewards.

**Limitations of the Present Study**

As with any research involving human subjects, the present study is not without limitations. Thus, the results should be interpreted cautiously. The study involved a convenience sample of undergraduate students enrolled in psychology and educational psychology courses. Consequently, the sample of students, although randomly assigned to groups, was not randomly selected from the target population. Moreover, participant attrition may have influenced the results, as approximately 30% of participants did not complete the posttest. A plausible explanation for this is that the study spanned three weeks, which may have discouraged some from completing all three sessions. Bias may also have been introduced in measurement. Even though the performance assessment and confidence judgments were objective in nature, the MAI may have been prone to bias in self-reports because students may have overestimated their metacognitive awareness, thus potentially biasing results.

Next, this study did not include other potential variables that may affect performance, confidence, and/or calibration accuracy, such as achievement in psychology, interest in the topic (i.e., situational or personal interest), or calibration bias. As such, the effects of these confounding variables on the outcomes relevant to this investigation could be neither ascertained nor controlled. Furthermore, the use of artificial forms of extrinsic incentives—cash payments in the present investigation—limit
the generalizability of these results to classroom contexts with respect to the effects of extrinsic incentives on calibration accuracy.

Finally, the NNMM does not come without shortcomings in spite of the fact that it presently remains the prevailing framework in calibration research. For instance, the model is necessarily limited in its explanatory power regarding calibration due to its simplistic view of metacognition. Recent research has elucidated the weaknesses of the NNMM, and it has proposed a different conception of calibration—as consisting of sensitivity and specificity components that learners use to judge inaccurate versus accurate responses—with respect to metacognition (Schraw et al., 2012). The model can be improved by shifting the focus away from the individual to contextual factors, such as affect (i.e., academic emotions), relevance of instruction, extrinsic utility value, and situational interest (e.g., Boekaerts & Razendaal, 2010). Moreover, additional metacognitive components beyond monitoring and control—for example, planning and evaluation—can be included to further strengthen the model’s explanatory power.

Despite these limitations, I believe that the present study offers new insights to the scientific inquiry regarding the role of metacognitive strategy training and incentives with respect to performance, calibration accuracy, and confidence judgments, and thus, represents a unique contribution to the literature on calibration.

**Educational Implications**

Beyond contributions to research and theory, the findings of the present study also have direct application to educational practice. For instance, educators of undergraduate students should include metacognitive strategy training as part of their curriculum. However, the training, to be successful, should be explicit, clear, succinct, and directly
relevant and applicable to undergraduate students. This tailored strategy training should subsequently have a positive effect on student outcomes beyond performance, such as improved levels of confidence and calibration accuracy brought about by students’ ability to more accurately feel and attend to what they know and what they do not know about course content. Additionally, extrinsic rewards appear to positively affect performance and levels of confidence, albeit they had no significant effect on calibration accuracy. While the present study used cash payments as the reward structure, educators can adopt other extrinsic reward structures that are more suitable and sustainable in the classroom. For instance, educators can offer students the choice to opt out of a homework assignment or quiz as a form of negative reinforcement. These types of extrinsic incentives are far more feasible in classroom settings.

**Directions for Future Research**

There has been little research examining the influence of multiple interventions on calibration accuracy, particularly within the lens of the NNMM. The present study demonstrated that a combination of metacognitive strategy training and incentives improves not only performance and confidence in performance judgments but calibration accuracy as well. However, this study did not investigate the influence of these interventions on calibration bias, another major index of calibration. Future research should focus on examining the effects of strategy training and incentives on both calibration accuracy and bias and their presumed impact on metacognitive monitoring and control. Moreover, the incentives manipulation involved the use of extrinsic rewards only. This is in line with the majority of research on the effects of incentives (e.g., Hogarth et al., 1991; Swanson, 1990; Yates, 1990). Nevertheless, understanding how and
to what extent extrinsic and intrinsic rewards influence performance, confidence, and calibration accuracy is critical to advancing research, practice, and theory not only as it pertains to metacognition and the NNMM but motivation as well (e.g., goal orientation, self-determination, self-efficacy, etc.). Therefore, future research should endeavor to include training as well as extrinsic and intrinsic incentives. Finally, as more research is uncovered highlighting the weaknesses of the NNMM, researchers should extend the findings of this study utilizing other theories of metacognition or other models of self-regulated learning to ascertain whether these frameworks more clearly and completely capture the essence of calibration and explain the findings of this study.
APPENDICES

Appendix A: Demographics

Demographics Form

Directions: Please answer each of the following demographic questions.

1. What is your age? __________

2. What is your gender?  M___  F___

3. What is your ethnicity? Hispanic___ Caucasian___ African-American___
   Asian/Pacific Islander___  Native American/Alaskan Native___
   Other___ (please specify)____________________________

4. Please indicate your major:

   Major____________________________

5. What is your current undergraduate standing?

   a. Freshman  b. Sophomore
   c. Junior  d. Senior
Appendix B: General Metacognitive Awareness

Directions: Please place a vertical slash at the point on the continuous line under each statement that best corresponds to how true each statement is about you.

Not at all true of me              Very true of me
0_______________________________100

For instance, the closer the slash is to “Not at all true of me” the LESS true that statement is about you. Conversely, the closer the slash is to “Very true of me” the MORE true that statement is about you. Likewise, drawing a slash on either end of the line (0 or 100) indicates that the statement is either not at all true of you (0) or very true of you (100).

1. I ask myself periodically if I am meeting my goals.

   Not at all true of me              Very true of me
   0_______________________________100

2. I consider several alternatives to a problem before I answer.

   Not at all true of me              Very true of me
   0_______________________________100

3. I try to use strategies that have worked in the past.

   Not at all true of me              Very true of me
   0_______________________________100

4. I pace myself while learning in order to have enough time.

   Not at all true of me              Very true of me
   0_______________________________100

5. I understand my intellectual strengths and weaknesses.

   Not at all true of me              Very true of me
   0_______________________________100
6. I think about what I really need to learn before I begin a task.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

7. I know how well I did once I finish a test.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

8. I set specific goals before I begin a task.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

9. I slow down when I encounter important information.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

10. I know what kind of information is most important to learn.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

11. I ask myself if I have considered all options when solving a problem.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

12. I am good at organizing information.

Not at all true of me

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
13. I consciously focus my attention on important information.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

14. I have a specific purpose for each strategy I use.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

15. I learn best when I know something about the topic.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

16. I know what the teacher expects me to learn.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

17. I am good at remembering information.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

18. I use different learning strategies depending on the situation.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

19. I ask myself if there was an easier way to do things after I finish a task.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
20. I have control over how well I learn.
   Not at all true of me
   0
   Very true of me
   100

21. I periodically review to help me understand important relationships.
   Not at all true of me
   0
   Very true of me
   100

22. I ask myself questions about the material before I begin.
   Not at all true of me
   0
   Very true of me
   100

23. I think of several ways to solve a problem and choose the best one.
   Not at all true of me
   0
   Very true of me
   100

   Not at all true of me
   0
   Very true of me
   100

25. I ask others for help when I don’t understand something.
   Not at all true of me
   0
   Very true of me
   100
26. I can motivate myself to learn when I need to.

Not at all true of me
Not at all true of me

0-------100

27. I am aware of what strategies I use when I study.

Not at all true of me
Not at all true of me

0-------100

28. I find myself analyzing the usefulness of strategies while I study.

Not at all true of me
Not at all true of me

0-------100

29. I use my intellectual strengths to compensate for my weaknesses.

Not at all true of me
Not at all true of me

0-------100

30. I focus on the meaning and significance of new information.

Not at all true of me
Not at all true of me

0-------100

31. I create my own examples to make information more meaningful.

Not at all true of me
Not at all true of me

0-------100
32. I am a good judge of how well I understand something.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

33. I find myself using helpful learning strategies automatically.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

34. I find myself pausing regularly to check my comprehension.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

35. I know when each strategy I use will be most effective.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

36. I ask myself how well I accomplished my goals once I am finished.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

37. I draw pictures or diagrams to help me understand while learning.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

38. I ask myself if I have considered all options after I solve a problem.

<table>
<thead>
<tr>
<th>Not at all true of me</th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
39. I try to translate new information into my own words.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

40. I change strategies when I fail to understand.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

41. I use the organizational structure of the text to help me learn.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

42. I read instructions carefully before I begin a task.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

43. I ask myself if what I am reading is related to what I already know.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

44. I reevaluate my assumptions when I get confused.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>

45. I organize my time to best accomplish my goals.

<table>
<thead>
<tr>
<th></th>
<th>Very true of me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>100</td>
</tr>
</tbody>
</table>
46. I learn more when I am interested in the topic.

Not at all true of me

Not at all true of me

0

100

47. I try to break studying down into smaller steps.

Not at all true of me

Not at all true of me

0

100

48. I focus on overall meaning rather than specifics.

Not at all true of me

Not at all true of me

0

100

49. I ask myself questions about how well I am doing while I am learning something new.

Not at all true of me

Not at all true of me

0

100

50. I ask myself if I learned as much as I could have once I finish a task.

Not at all true of me

Not at all true of me

0

100

51. I stop and go back over new information that is not clear.

Not at all true of me

Not at all true of me

0

100

52. I stop and reread when I get confused.

Not at all true of me

Not at all true of me

0

100
Appendix C: Performance Assessment

Instructions: Circle the BEST response for each item. After you have responded to each item, please rate how much confidence you have in your response to that item by drawing a vertical slash along the line under each item. The closer the line is to “0% CONFIDENCE” the LESS confident you are in your response; the closer the line is to “100% CONFIDENCE” the more confident you are in your response. For example, if you draw a line at “0% CONFIDENCE” you have NO confidence in your response to that item whereas if you draw a line at “100% CONFIDENCE” you are indicating that you have TOTAL confidence in your response to that item. On the other hand, if you draw a slash through the middle of the line you are indicating that you have “50% confidence” in your response to that item.

1. Behavioral learning theories emphasize:
   a. thinking.
   b. development.
   c. observable actions.
   d. heredity over environment.

2. Which of the following best describes operant conditioning?
   a. Pairing of a neutral stimulus with an unlearned response (such as an eye blink).
   b. Establishing an association between a learned behavior and its consequences.
   c. Establishing an association in memory.
   d. Learning through observation.

3. Reinforcement causes behavior to ____________, punishment causes behavior to ____________.
   a. decrease, increase
   b. accommodate, assimilate
   c. increase, decrease
4. Alex turns in a perfect essay and receives an A+. He has apparently received:
   a. Punishment I.
   b. Positive reinforcement.
   c. Negative reinforcement.
   d. Punishment II.

5. Slot machines are programmed to payoff using ___________, the strongest schedule of reinforcement.
   a. fixed interval
   b. variable interval
   c. fixed ratio
   d. variable ratio

6. Working memory capacity is limited to an estimated ___________ items.
   a. 9 + or - 2
   b. 7 + or - 2
   c. 5 + or - 2
   d. 3 + or - 2

7. Applying knowledge, reasoning, problem solving, and strategic learning are all examples of:
   a. behavioral learning
   b. higher-order cognition
c. declarative memory

d. bottom-up processes

8. Schemas (such as the "restaurant schema") are BEST thought of as:
   a. specific memories for events,
   b. abstract memory representations, with slots that can be filled in.
   c. genetically endowed ways of organizing information in memory.
   d. networks of associations in memory.

9. Procedural knowledge is:
   a. knowledge of facts.
   b. network organized.
   c. knowledge of "how to."
   d. knowing when and why.

10. The best way to move information from working memory to long-term memory is:
  a. maintenance rehearsal.
  b. elaborative rehearsal.
  c. generalized rehearsal.
  d. rote memorization.

11. Functional fixedness is the tendency when solving problems to:
  a. respond in the most familiar way.
  b. stick to the same strategy, even if it is not working.
c. see objects for the intended purpose only.

d. apply algorithmic strategies when heuristics are called for.

12. Social learning theory as described by Bandura suggests that we learn by:
   a. doing
   b. observing
   c. engaging in activities with a more knowledgeable other
   d. being reinforced for our behaviors

13. Richard is studying both French and Spanish. In the same week, he learns that the French word for "mother" is *mere* and that the Spanish word for "mother" is *madre*. One day his French teacher asks Richard, "Who is married to your father?" and Richard erroneously answers, "Madre." Richard's memory error can best be explained in terms of:
   a. decay
   b. interference
   c. failure to store
   d. insufficient wait time

14. In which one of the following examples is *metacognition* most clearly illustrated?
   a. Mary knows all the letters of the alphabet before she begins kindergarten.
   b. Fran knows how much of a book she is likely to remember a month later.
   c. Billy can read fourth-grade-level books at the age of six.
   d. Alex has a photographic memory that enables him to remember almost everything he sees.
15. Which one of the following best illustrates concept mapping?

a. Alexandra lists the defining and correlational features of the concept canine.

b. Bob draws a chart listing the sequence of events leading up to World War II in chronological order.

c. Christina puts the words force, gravity, velocity, acceleration, and time on a piece of paper; she then draws lines between pairs of related words and describes the relationships.

d. Darnell makes a chart showing the hierarchy that biologists use to classify animals; his chart includes such concepts as vertebrates, invertebrates, mammals, fish, birds, mollusks, crustaceans, and so on.

16. Which one of the following is the best example of positive transfer?

a. Robert is trying to learn the spelling of the word shepherd; he remembers how he learned to spell lighthouse by putting two words together and so writes ”shepherd.”

b. Vince notices that rules of grammar are not always the same in English and Japanese.

c. Zelda uses the formula for calculating the area of a circle when she wants to figure out how much bigger a 10-inch pizza is than a 7-inch pizza.

d. David is trying to learn to program a computer. He reads his programming manual but is confused by some of its instructions.

17. Nathan has been playing golf with his parents for many years. When he goes out for the school baseball team, he has trouble hitting the ball because he keeps confusing the swing of the bat with how he swings a golf club. Nathan's difficulty reflects:

a. Negative transfer

b. General transfer

c. Rote learning

d. Mental set in problem solving
18. Which one of the following statements best describes a situated cognition perspective of transfer?
   a. Transfer is more likely to be transferred when it is consciously retrieved from long-term memory.
   b. Knowledge and skills acquired in one context are unlikely to be used in a very different context.
   c. Studying principles of deductive and inductive reasoning leads to more logical thought processes in a variety of contexts.
   d. Students are more likely to transfer new knowledge when their teacher describes the situations in which they might use it.

19. Which one of the following statements characterizes a well-defined problem?
   a. It has a clear goal.
   b. It can be solved only by a heuristic.
   c. It has several possible correct solutions.
   d. It is missing information essential for a solution.

20. Albert Bandura uses the term _____ to describe the fact that people observe their own behavior, judge it, and themselves reinforce or punish.
   a. Self-confidence.
   b. Self-regulation.
   c. Self-denial.
   d. Self-efficacy.
21. The teacher would like Kerri to learn the entire alphabet. At first she reinforces Kerri for recognizing two letters, then four, then six, and so on. The teacher is using the technique called:

a. Negative reinforcement.

b. Generalization.

c. Shaping.

d. Punishment.

0% —————————————————— 100%
CONFIDENCE	CONFIDENCE

22. For which component of the memory system is rehearsal most important?

a. Long-term episodic.

b. Short-term memory.

c. Long-term procedural.

d. Sensory register.

0% —————————————————— 100%
CONFIDENCE	CONFIDENCE

23. Before a person connects the dots in a connect-the-dots task, he guesses what the figure is. This is an example of the Gestalt principle of:

a. Figure-ground.

b. Proximity.

c. Closure.

d. Pointilism.

0% —————————————————— 100%
CONFIDENCE	CONFIDENCE

24. The process by which tasks require less and less attentional capacity as they become better learned is known as:

a. Pre-attentional behavior.

b. Fading.

c. Automaticity.
d. Rehearsal.

25. In a series of names on list, which are likely to be most difficult to remember?
   a. Those towards the beginning.
   b. Those towards the middle.
   c. Those towards the end.
   d. None of the above, as order is irrelevant to memory.

26. Which one of the following italicized concepts is an example of undergeneralization?
   a. Fred thinks that spiders are insects.
   b. Ivan thinks that birds are not animals.
   c. Lenny thinks that “you” is a noun.
   d. Oscar thinks “you” is not a noun.

27. The surest sign that students have mastered a concept is that they:
   a. Tend to overgeneralize rather than undergeneralize.
   b. Tend to undergeneralize rather than overgeneralize.
   c. Can accurately identify at least one example and one nonexample of the concept.
   d. Can consistently distinguish between examples and nonexamples.

28. Which of the following statements best describes psychologists’ belief that children’s knowledge sometimes takes the form of theories?
   a. The ways in which children categorize their experiences usually have little relevance to physical reality.
b. Children form hypotheses about the characteristics of members of a particular concept category and then test those hypotheses against specific examples of the concept they encounter.

c. In the early years, children develop concrete understandings of events; these understandings become increasingly more abstract as they reach adolescence.

d. Children develop general belief systems about how aspects of the world operate.

29. Which of the following concepts best describes conceptual understanding?
   a. Students learn all the facts that a teacher or textbook presents related to a topic.
   b. Students can describe two opposing perspectives about a controversial issue.
   c. Students come to the realization that a particular belief they have is incorrect.
   d. Students learn ideas related to a topic in a meaningful and integrated fashion.

30. Which of the following is the best example of problem-based learning?
   a. Learning the logic behind certain problem-solving procedures in math.
   b. Learning history by reading detective novels set in certain historical eras.
   c. Solving a series of mathematical word problems that gradually progress in difficulty.
   d. Devising a way to move a large, heavy object using principles of physics.

31. Students misconceptions about the world may come from a variety of sources. Which one of the following is not likely a source that theorists have identified?
   a. Students form general theories based on how the world appears to be.
   b. Teachers and textbooks sometimes provide misinformation.
   c. Students usually believe explanations that younger children give them.
d. Common expressions in language (e.g., the sun “sets” in the west) misrepresent reality.

32. Which of the following best describes the process of conceptual change?
   a. Developing new categories to classify objects and events.
   b. Revising one’s beliefs after receiving information that contradicts those beliefs.
   c. Achieving the instructional objectives that a teacher has established for a lesson.
   d. Acquiring more sophisticated vocabulary with which to describe the events in one’s life.

33. After getting in line quietly, Tony is told by the teacher, “Nice going! For setting such a good example, you do not have to help with the finger painting clean-up this afternoon.” Tony has apparently received:
   a. Punishment I.
   b. Positive reinforcement.
   c. Negative reinforcement.
   d. Punishment II.

34. Joe receives a piece of gum when he works three problems. Joe is being reinforced on what schedule?
   a. Fixed interval.
   b. Fixed ratio.
   c. Variable interval.
   d. Variable ratio.
35. A child yells at his brother but not his mother. This is an example of:

a. Generalization.

b. Negative reinforcement.

c. Antecedent stimulus.

d. Discrimination.

36. Mark and Tom each got a plastic model airplane as a gift. Mark watched Tom put his together and noted his mistakes. Mark then put his own plane together without making any mistakes. Mark learned new behavior through:

a. Cueing.

b. Positive reinforcement.

c. Modeling.

d. Shaping.

37. Alice is a student who frequently tries to talk to others during seatwork. On Thursday, when Alice is working quietly, Mrs. Barnette says, “Good, Alice, I like it when you do your work so well.” Which intervention strategy is being used?

a. Praise of other students.

b. Simple verbal reminder.

c. Praise for correct behavior.

d. Nonverbal cues.

38. Although Allen hasn’t typed in 10 years, he finds that while practicing on his new word processor the skills come back to him fairly quickly. Which component of memory is most directly involved?

a. Long-term episodic.

b. Short-term.
c. Long-term semantic.

d. Long-term procedural.

0%—--------------------------------------------------------100%
CONFIDENCE CONFIDENCE

39. One problem with students’ practice of underlining material to be used for studying is the tendency to:
   a. Not underline enough material.
   
   b. Underline too much material.
   
   c. Consider underlining to be sufficient and not study the material as well at a later date.
   
   d. Using different colors when underlining.

0%—--------------------------------------------------------100%
CONFIDENCE CONFIDENCE

40. The process of repeating an item of information to yourself several times without altering its form is called:
   a. Perception.
   
   b. Encoding.
   
   c. Retrieval.
   
   d. Rehearsal.

0%—--------------------------------------------------------100%
CONFIDENCE CONFIDENCE
Appendix D: Institutional Review Board Approval

Social/Behavioral IRB – Exempt Review
Deemed Exempt

DATE: September 23, 2011
TO: Dr. Gregory Schraw, ERCD
FROM: Office of Research Integrity – Human Subjects
RE: Notification of review by Josi dos Santos/ Ms. Josi dos Santos, CIP
Protocol Title: Enhancing the Calibration Accuracy of Adult Learners: A Multifaceted Intervention
Protocol # 1109-3916M

This memorandum is notification that the project referenced above has been reviewed as indicated in Federal regulatory statutes 45CFR46 and deemed exempt under 45 CFR 46.101(b)1.

PLEASE NOTE:
Upon Approval, the research team is responsible for conducting the research as stated in the exempt application reviewed by the ORI – HS and/or the IRB which shall include using the most recently submitted Informed Consent/Assent Forms (Information Sheet) and recruitment materials. The official versions of these forms are indicated by footer which contains the date exempted.

Any changes to the application may cause this project to require a different level of IRB review. Should any changes need to be made, please submit a Modification Form. When the above-referenced project has been completed, please submit a Continuing Review/Progress Completion report to notify ORI – HS of its closure.

If you have questions or require any assistance, please contact the Office of Research Integrity – Human Subjects at IRB@unlv.edu or call 895-2794.
Appendix E: Performance Assessment Stimulus

Instructions: Please read the following passage carefully. It covers topics related to general psychology, such as behaviorism and cognitivism, among others. After reading the passage, you will be given a performance assessment on topics related to the passage.

Behaviorism and Related Theories

Children are excellent learners. What they learn, however, may not always be what we intend to teach. Researchers believe that children develop general belief systems about how aspects of the world operate based on how they appear to be, known as naïve theories. Ms. Esteban is trying to teach students how to behave in class, but by paying attention to Rebecca's outburst, she is actually teaching them the opposite of what she intends. Rebecca craves her teacher's attention, so being called on (even in an exasperated tone of voice) rewards her for calling out her answer. Not only does Ms. Esteban's response increase the chances that Rebecca will call out answers again but also Rebecca now serves as a model for her classmates' own calling out. What Ms. Esteban says is less important than her actual response to her students' behaviors.

Behavioral learning theories focus on the ways in which pleasurable or unpleasant consequences of behavior change individuals' behavior over time and ways in which individuals model their behavior on that of others. Social learning theories focus on the effects of thought on action and action on thought. Later chapters present cognitive learning theories, which emphasize unobservable mental processes that people use to learn and remember new information or skills. Behavioral learning theorists try to discover principles of behavior that apply to all living beings. Cognitive and social learning theorists are concerned exclusively with human learning. Actually, however, the boundaries between behavioral and cognitive learning theories have become increasingly indistinct in recent years as each school of thought has incorporated the findings of the other.

Pavlov: Classical Conditioning

In the late 1800s and early 1900s, Russian scientist I van Pavlov and his colleagues studied the digestive process in dogs. During the research, the scientists noticed changes in the timing and rate of salivation of these animals. Pavlov observed that if meat powder was placed in or near the mouth of a hungry dog, the dog would salivate. Because the meat powder provoked this response automatically, without any prior training or conditioning, the meat powder is referred to as an unconditioned stimulus. Similarly, because salivation occurred automatically in the presence of meat, also without the need for any training or experience, this response of salivating is referred to as an unconditioned response. Whereas the meat will produce salivation without any previous experience or training, other stimuli, such as a bell, will not produce salivation. Because these stimuli have no effect on the response in question, they are referred to as neutral.

---

stimuli. Pavlov's experiments showed that if a previously neutral stimulus is paired with an unconditioned stimulus, the neutral stimulus becomes a conditioned stimulus and gains the power to prompt a response similar to that produced by the unconditioned stimulus. In other words after the bell and the meat are presented together, the ringing of the bell alone triggers the salivation in the dog. This process is referred to as classical conditioning.

Skinner: Operant Conditioning

Some human behaviors are clearly prompted by specific stimuli. B.F. Skinner proposed that the reflexive behavior accounts for a small proportion of all actions. Skinner proposed another class of behavior, which he labeled operant conditioning because they operate on the environment in the apparent absence of any unconditioned stimuli, such as food. Skinner’s work focused on the relation between behavior and its consequences. For example, if an individual’s behavior is immediately followed by pleasurable consequences, the individual will engage in that behavior more frequently. The use of pleasant and unpleasant consequences to change behavior is known as operant conditioning.

Perhaps the most important principle in behavioral learning theories is that behavior changes according to its immediate consequences. Pleasurable consequences strengthen behavior while unpleasant consequences weaken it. A reinforcer is defined as any consequence that strengthens or increases the frequency of behavior. However, the effectiveness of the reinforcer must be explicitly demonstrated to the person whose behavior one wishes to modify. Positive reinforcers include praise, grades, stars, money, etc. because they give something pleasurable. Nevertheless, another way to strengthen a behavior is to have the behavior’s consequences be an escape from an unpleasant situation or a way of preventing something unpleasant from occurring. These types of reinforcers are known as negative reinforcers. Negative reinforcement is often confused with punishment because individuals are negatively reinforced for doing something unfavorable. One way to avoid this confusion is to remember that reinforcement, whether positive or negative, strengthens behavior whereas punishment is intended to decrease behavior.

Punishment refers to consequences that are intended to weaken or decrease behavior. Presentation punishment, or Punishment I, is the use of unpleasant consequences or aversive stimuli, as when a student is scolded. Removal punishment, or Punishment II, is the removal of a pleasant consequence, such as loss of privileges or time out. Most students need some form for reinforcement along the way. When teachers guide students towards goals by reinforcing the many steps that lead to successful learning they are using a technique called shaping. The term shaping is used in behavioral learning theories to refer to the teaching of new skills or behaviors by reinforcing learners for approximating the desired final outcome.

By definition, reinforcers strengthen behavior. But what happens when reinforcers are withdrawn? Eventually the behavior will be weakened, and ultimately, it will disappear
or become extinct. The effects of reinforcement depend on many factors, one of the most important which is the schedule of reinforcement, or the frequency with which reinforcers are given. One common schedule of reinforcement is the fixed ratio, in which reinforcers are given after a fixed number of behaviors. A variable ratio schedule of reinforcement is one in which the number of behaviors required for reinforcement is unpredictable, although it is certain that the behaviors will eventually be reinforced. In fixed interval schedules, reinforcement is available only at certain periodic times. Finally, in a variable-interval schedule, reinforcement is available at some times but not at others, and the individual being reinforced has no idea when a behavior will be reinforced. However, maintenance reinforcement occurs with behaviors that do not need to be reinforced because they are intrinsically reinforcing, that is engaging in these behaviors is pleasurable in itself.

Antecedent stimuli, or events that precede a behavior, are also known as cues because they inform individuals about what behavior will be reinforced and/or what behavior will be punished. Discrimination is the use of cues, signals, or information to know when behavior is likely to be reinforced. Also, generalization, or transfer, refers to situations in which behaviors learned under one set of conditions is applied successfully to other situations different from the one in which they were learned. Overgeneralization involves including objects or events that are not true members of the category whereas undergeneralization involves having too narrow a view about which objects or events concepts include. This perspective differs from situated cognition, which stipulates that knowledge learned in one context is unlikely to transfer to a very different context. Positive transfer refers to situations in which learning from one situation assists learning in another whereas negative transfer is when learning from one situation interferes with learning in another situation. It is important to note that students sometimes develop misconceptions, or inaccurate information, about facts and information they encounter. Misconceptions are an important type of negative transfer. This is particularly problematic because misconceptions can lead to additional inaccurate learning and erroneous knowledge of the way the world operates (i.e., naïve theories). In order to dispel misconceptions, students are typically made aware of their misconception, introduced to the correct concept (e.g., ideas related to the topic in which the misconception exists is learned in a meaningful and integrated fashion so as to yield appropriate conceptual understanding), and then challenged to modify their understanding of the concept, a process known as conceptual change. Sometimes people are unable to generalize, such as when they see things for their intended primary purpose only, a phenomenon known as functional fixedness. Mastery of concepts involves learners consistently distinguishing between examples and non-examples.

Bandura: Social Learning Theory

Social learning theory is a major outgrowth of the behavioral learning tradition. Developed by Albert Bandura, social learning theory accepts most of the principles of behavioral theories but focuses to a much greater degree on the effects of cues on behavior and on internal mental processes, emphasizing the effect of thought on action and action on thought. Bandura noted that the Skinnerian emphasis on the effects of the
consequences of behavior largely ignored the phenomenon of modeling, or the imitation of others’ behavior, and of vicarious experience—learning from others’ successes and failures. Bandura’s observational learning involves four phases: attentional, retention, reproduction, and motivational. Another important concept in social learning theory is self-regulation, or rewarding or punishing one’s own behavior. Students can be taught to monitor and regulate their own behavior through an approach known as cognitive behavior modification. Likewise, problem-based learning is intended to develop both problem solving strategies and disciplinary knowledge bases and skills by placing students in the active role of problem solvers confronted with an ill-structured problem that mirrors real-world problems. However, some students may not appropriate utilize some strategies, such as the tendency to underline too much, thus rendering the strategy useless.

Cognitivism

Information-Processing Approaches

Information-processing theory is a cognitive theory that describes the processing, storage, and retrieval of knowledge in the mind. Two widely known information-processing models include the levels-of-processing theory, which posits that people subject stimuli to different levels of mental processing and retain only the information that has been subjected to the most thorough processing, and the dual code theory of memory, which posits that information is stored in long-term memory as either visual or verbal/auditory. Sensory registers receive large amounts of information from each of the five senses and hold it for a very short time, no more than a few seconds. If nothing happens, information in the sensory registry is quickly lost. Perception of stimuli is not as straightforward as reception of stimuli. Instead, it involves mental representation and it is influenced by our mental state, past experience, knowledge, motivations, and other factors. Along a related line, attention is the active focus on certain stimuli while excluding others; it is a limited resource.

Memory

In the memory system, short-term memory is a storage system that can hold a limited amount of information, usually 5 to 9 pieces of information simultaneously, for several seconds. When we stop thinking about something, it disappears from our working memory, which is another term for short-term memory. In order to learn facts and other information individuals need to repeat the information multiple times, a process known as rehearsal. Long-term memory, on the other hand, is that part of the memory system where information is stored for long periods of time; it is believed to be a very large, very long-term memory repository. Elaboration or elaborative rehearsal is one of the most effective ways to transfer information from short-term to long-term memory. Theorists divide long-term memory into three main parts: episodic, semantic, and procedural memory. Episodic memory is our memory of personal experience. Conversely, semantic memory contains the facts and generalized information that we know while procedural memory holds the “how to,” or the steps necessary to fulfill a task. Episodic memory
contains images of experiences organized by when and where they occurred. Flashbulb memory is the phenomenon in which visual and auditory memories about an important event are fixed in long-term memory. Semantic or declarative memory is organized in networks of connected ideas or relationships called schemata (singular: schema). Finally, procedural memory is organized in a series of stimulus-response pairings.

One important reason people forget is interference, which happens when information gets mixed up with, or pushed aside by, other information. Retroactive inhibition, one form of interference, occurs when previously learned information is lost because it is mixed up with new and somewhat similar information. Proactive inhibition, on the other hand, occurs when learning one set of information interferes with learning later information. However, there are several mechanisms to help people remember. Proactive facilitation refers to the increased ability to learn new information based on the presence of previously learned information whereas retroactive facilitation refers to increased comprehension of previously learned information because of the acquisition of new information. Interestingly, when people learn list of words and are subsequently tested immediately they tend to recall the first few (primacy effect) and last few (recency effect) items better than those in the middle. Information or skills may exist in long-term memory, but may take so much time or effort to retrieve that they are of limited value when speed of access is required. This is where automaticity, or a level of speed and ease such that tasks can be performed or skills utilized with little mental effort, comes in handy. However, it is important to note that automaticity takes much time, often years, to achieve.

It is important to distinguish between rote and meaningful learning. Rote learning refers to the memorization of facts or association whereas meaningful learning is not arbitrary and it relates to information or concepts learners already have. Conversely, inert knowledge could and should be applicable to a wide range of activities and situations but is applied to a restricted set of circumstances. In the realm of meaningful learning, metacognition is important. Metacognition refers to knowledge about one’s own learning or about how to learn. Study skills, problem solving and reasoning, and thinking skills are examples of metacognitive (i.e., higher-order thinking) skills. In the area of problem solving and reasoning, researchers distinguish between well-defined problems, or those that tend to have a simple structure and a clear goal with one correct solution, and ill-defined problems, or those with complex structures that have no one correct solution or have many possible solutions. A related family of strategies is outlining and mapping. Outlining presents the main points of the material in a hierarchical format, with each detail organized under a higher level category. In networking and concept mapping, students identify main ideas and then diagram connections between them. Also, advance organizers orient students to material they will learn subsequently and help them recall related information that could help them incorporate the new information.

Finally, Gestalt psychology, which emphasizes the unified or meaningful whole, helps explain why individuals are able to see the whole object even when it is incomplete, a concept known as closure.
Appendix F: Strategy Training Practice Text

WINTER DEPRESSION: A CASE OF BEING SAD

A young woman lies asleep on an overcast winter morning. At 4 a.m., a faint glow emanates from a light bulb placed beside her bed. The glow gradually gains intensity and bathes the room in soft light until 6 a.m. when she wakes. The woman has just experienced a simulated dawn. After several more mornings of artificial sunrise, the clouds begin to lift from her "winter depression" that appears in most cases from November to April.

As the study of seasonal affective disorder, or SAD, enters its second decade, researchers have marshalled an abundance of new findings. Along with feelings of sadness, anxiety, and tiredness, people who experience SAD often display symptoms not seen in other cases of depression. These include extreme weight gain, trouble waking up, daytime drowsiness, cravings for sweets or starches, and a significant drop in work performance.

A surprisingly large percentage of people suffer from SAD, with numbers increasing as one travels north of the equator. About 1 percent of Floridians are affected; 6 percent of New Yorkers, and 10 percent of New Hampshirites. The majority of those hospitalized with SAD experience their first symptoms in their early 20s, although most fail to recognize their winter malaise as SAD and therefore do not seek treatment. The majority of those who seek treatment are women. Currently it is unknown whether women are more susceptible to the disease or are merely more likely to report it.

A number of theories have been proposed to explain the onset of SAD. One suggests that major changes occur in the body's biological clock, or circadian rhythms, due to decreased exposure to sunlight. Consistent with this theory, regular exposure to bright light, usually 20 times brighter than normal room light, decreases the severity of SAD in most severe cases. Studies with rats show that increased exposure to bright light prior to waking increases the uptake of melatonin, a hormone secreted to the brain in larger quantities during sleep. No comparable studies have been performed on humans.

Another possible cause of SAD is differences in the sensitivity of people's eyes. Researchers speculate that SAD patients may have retinas incapable of squeezing more light out of short winter days. Studies report that SAD individuals, compared to non-depressed controls, experience greater difficulty seeing dim light when placed in dark surroundings. In a related study, the retinas of 19 SAD patients generated weaker electrical activity across the retina in response to light than control subjects. Oddly, this phenomenon occurs in SAD patients only during the winter months. Researchers currently have few explanations for this puzzling change.

Most physicians rely on exposure to intense light to combat SAD. Light therapy is usually conducted using one of two schedules. In the first, individuals are seated in front of a large screen emitting light 5 times brighter than normal room light for 2 hours. In the second, patients receive light 10 to 20 times brighter than normal room light for 30 minutes. SAD treatments are usually conducted once each day. Treatments given in the early morning appear to be most effective followed by those administered at night. Treatments administered during midday are least effective. Approximately 60 percent of SAD cases report substantial improvement using these techniques. In comparison, only 10 percent of patients report the same degree of improvement when given a "placebo" treatment (i.e., a treatment intended to have no effect) of dim light (i.e., 3 to 4 times normal room light).
Another mode of therapy is to gradually expose individuals to light prior to waking. This so called "simulated dawn" treatment appears to be quite effective. In one study, 23 SAD patients used special incandescent bulbs that gradually produced light comparable to a natural dawn between the hours of 4 and 6 a.m. each morning for one week. Another 18 SAD patients received a week of placebo treatments which consisted of 30-minute dawns with a peak intensity equal to that of moonlight. Those receiving the full-scale dawn exposure reported substantial improvement; those receiving the placebo reported little change.

A third and more recent approach to treatment is to equip SAD patients with "light visors." Treatments usually last 30 minutes per day. Several preliminary studies indicate that visors have limited effectiveness; only about 40 percent of patients report substantial improvement under this treatment. Surprisingly, the intensity of the dosage of light had little relationship to subsequent improvement, causing some researchers to suggest that visors improve SAD for reasons other than exposure to light. Currently there are no studies that compare the effectiveness of visors with the other treatments described above. Much additional research is needed.

Finally, some researchers have observed that many SAD patients experience improvements similar to the treatments described above after taking daily 1-hour walks in full sunlight. Recent studies indicate that SAD patients spend less times outdoors during the winter months due to increased depression. It is especially ironic that depression caused by reduced exposure to high intensity light should cause individuals to forego an activity that may actually reduce their depression.

Fortunately, most of those who suffer from severe cases of SAD can expect substantial improvement using one of the four treatment modes described above (i.e., light screens, simulated dawns, visors, and walking). This represents a dramatic change from only two decades ago when light-induced depression was thought to be "psychological." Perhaps the next decade will see further breakthroughs in understanding the origin and treatment of SAD. Indeed, it is fair to say the future looks brighter every day for individuals suffering from winter depression.2

---

2 Adapted with permission from Dr. Gregory Schraw.
Appendix G: Strategy Training Practice Test

SAD Test³

Instructions: Circle the BEST response for each item. After you have responded to each item, please rate how much confidence you have in your response to that item by drawing a vertical slash along the line under each item. The closer the line is to “0% CONFIDENCE” the LESS confident you are in your response; the closer the line is to “100% CONFIDENCE” the more confident you are in your response. For example, if you draw a line at “0% CONFIDENCE” you have NO confidence in your response to that item whereas if you draw a line at “100% CONFIDENCE” you are indicating that you have TOTAL confidence in your response to that item. On the other hand, if you draw a slash through the middle of the line you are indicating that you have “50% confidence” in your response to that item.

1. Research on SAD has been on-going for _____ years?
   a) 0-5
   b) 10-20
   c) 20-30
   d) 30-40

   0% ———————————————————— 100%
   CONFIDENCE                   CONFIDENCE

2. The so called "simulated dawn" treatment consists of:
   a) exposing patients to a bright light screen
   b) equipping patients with a visor
   c) watching a videotape of the sun coming up
   d) gradually exposing patients to light prior to waking

   0% ———————————————————— 100%
   CONFIDENCE                   CONFIDENCE

3. SAD is related to:
   a) physiological factors
   b) environmental factors
   c) physiological and environmental factors
   d) none of the above

   0% ———————————————————— 100%
   CONFIDENCE                   CONFIDENCE

³ Taken with permission from Dr. Gregory Schraw
4. SAD usually begins in _____ and ends in _____?
   a) October-May
   b) November-April
   c) December-March
   d) April-November

5. Based on the article, one can conclude that light-related depression occurs with certainty in:
   a) humans
   b) rats and humans
   c) all mammals
   d) all animals

6. Light screen therapy is most effective when administered during the:
   a) morning
   b) afternoon
   c) evening
   d) none of these times differ

7. Studies indicate that intense light affects melatonin levels in:
   a) rats
   b) humans
   c) rats and humans
   d) all animals

8. Light therapy is usually conducted ______ each day?
   a) once
   b) twice
   c) three times
d) up to six times

0%-----------------------------100%
CONFIDENCE                   CONFIDENCE

9. One cause of SAD is:
   a) production of melatonin during physical activity
   b) the way the retina perceives light
   c) average winter temperature
   d) general level of fitness

0%-----------------------------100%
CONFIDENCE                   CONFIDENCE

10. The acronym SAD stands for:
   a) seasonally activated depression
   b) seasonally acute disfunction
   c) seasonal affective depression
   d) seasonal affective disorder

0%-----------------------------100%
CONFIDENCE                   CONFIDENCE

11. The current status of light visors as a treatment suggests that:
   a) it improves SAD in 60% of cases
   b) it does not improve SAD at all
   c) it is as effective as other treatments
   d) it needs much more research

0%-----------------------------100%
CONFIDENCE                   CONFIDENCE

12. One of the symptoms of SAD that is not associated with other forms of depression is:
   a) sadness
   b) anxiety
   c) tiredness
   d) extreme weight gain

0%-----------------------------100%
CONFIDENCE                   CONFIDENCE

13. Which of the following statements is true?
a) brighter light is related to better treatment outcomes  
b) longer light exposure is related to better treatment outcomes  
c) length of exposure and brightness compensate for each other  
d) all of the above

14. In the "simulated dawn" study, the placebo group received a treatment which:

a) was a less intensive dawn  
b) was a more intensive dawn  
c) was a simulated dusk  
d) was a simulated sunset

15. Which of the following IS NOT a current explanation of SAD?

a) changes in one's biological clock due to decreased exposure to sunlight  
b) greater difficulty seeing dim light in dark surroundings  
c) changes in how the body metabolizes blood sugar  
d) retinas incapable of squeezing more light out of winter days

16. Most people hospitalized for SAD experience their first symptoms in their:

a) childhood  
b) teens  
c) early 20's  
d) mid 30's

17. Light therapies in general result in:

a) swift improvements or none at all  
b) gradual improvements for a minority of patients  
c) gradual improvements for many patients  
d) gradual improvement for all patients
18. The 30-minute simulated dawns described in the article are:

   a) very effective  
   b) moderately effective  
   c) produced little change  
   d) increased the severity of SAD
Appendix H: Strategy Training Intervention Fidelity Check Survey

Instructions: Please complete this brief survey regarding your overall impression and evaluation of the strategy training intervention. Please be honest in your ratings! Thank you!

Rate each item on the following scale:

1 = Strongly Disagree; 2 = Disagree; 3 = Agree; 4 = Strongly Agree

1. The strategy training was clear and understandable.  

2. The strategies covered in the training were appropriately and thoroughly explained.  

3. The strategies were sufficiently scaffolded and modeled for me to understand how, when, and in which situations to apply them.  

4. I had sufficient opportunity to practice and apply each strategy.  

5. The trainer demonstrated the utility value of each strategy with respect to calibration of performance.  

6. Overall, I feel that the strategy training has adequately prepared me to increase the accuracy and confidence of my calibration of performance judgments.  

7. The strategy training intervention was useful in enhancing my performance confidence judgments.
REFERENCES


doi:10.1016/j.learninstruc.2009.03.002


definition, measurement, and instruction. Educational Psychologist, 22, 255-278.

threshold priming on metamemory and recall. Acta Psychologica, 73, 55-68.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language,


probability calibration. Journal of Experimental Psychology: Learning, Memory,
and Cognition, 25, 1038-1052. doi: 0278-7393/99/S3.00

metacognition: Evidence against a general metacognitive ability. Memory &
Cognition, 28, 92-107.

and feeling of knowing. In L. M. Reder (Ed.), Implicit memory and metacognition

unrecognized incompetence and academic performance. Journal of Marketing
Education, 24, 243-252.

Kimball, D. R., & Metcalfe, J. (2003). Delayed judgments of learning affects memory,


CURRICULUM VITAE

Name and Academic Standing

Antonio Partida Gutierrez, M.Sc., Ph.D. Candidate

Home Address

8411 Lower Trailhead Ave.
Las Vegas, NV 89113
(702) 617-2720 (Home)
(702) 324-2695 (Mobile)
agutierrez_lv@yahoo.com

Academic Address

University of Nevada Las Vegas
Department of Educational Research, Cognition, & Development
4505 Maryland Parkway, Box 453003
Las Vegas, NV 89154-3003
(702) 895-3556
antonio.gutierrez@unlv.edu

Specialty Areas

Research Methods, Statistics, and Learning and Cognition

Research Interests

Metacognition
Metarepresentation
Self-Regulated Learning
Critical Thinking
Problem Solving
Reasoning
Epistemic Beliefs
Executive Function
Creativity and Creative Thinking

I. Education

2008  
*M.Sc. in Educational Psychology.* University of Nevada, Las Vegas, Department of Educational Psychology.

2004  
*B.A. in Political Science and Sociology.* University of Nevada, Las Vegas, Departments of Political Science and Sociology; Senior Thesis Title: *Revolutions and Social Movements of the 20th Century.* University of Nevada, Las Vegas Honors College.

II. Professional Experience

March 2007 – Present  
*Faculty Researcher,* University of Nevada, Las Vegas School of Nursing (UNLV SON)
- Engaged in research design planning and the conduct of research
- Prepared and reviewed grant documents (e.g., budgets, budget justifications, research design, timeline, etc.)
- Formulated and disseminated institutional policies and standard operating procedures
- Enforced research compliance and related policies and procedures
- Supervised/managed graduate assistants and student workers
- Taught research methods and statistics courses at the graduate-level
- Assisted in strategic planning efforts

September 2005 – March 2007  
*Assistant Manager,* New York, New York Hotel & Casino
- Formulated and disseminated budgets
- Placed merchandise/inventory orders and requisitions
- Supervised staff
- Created policies and procedures regarding staff conduct
- Enforced unit policies and procedures
- Provided mentoring and skills counseling to staff
- Evaluated staff performance
- Resolved conflicts between staff and guests/clients

October 2004 – September 2005  
*Research Supervisor,* Strategic Solutions
- Conducted background research
- Prepared interview and questionnaire/survey items
- Supervised/managed staff on research projects
- Analyzed data using Statistical Package for the Social Sciences (SPSS)
- Interpreted research findings
- Prepared executive summary of research results
- Presented research results before clients
- Conducted staff performance evaluations
January 2003 – July 2004
Policy Analyst, Legislative Counsel Bureau, Nevada State Legislature
- Provided constituent services
- Conducted research for state legislators
- Attended legislative committee meetings to provide expert testimony on pertinent legislative bills
- Prepared legislative reports, memoranda, and executive summaries for legislative committees and legislators
- Coordinated information with directors and heads of departments of state agencies.

III. Teaching Experience

2012 Spring
Instructor, NUR 707, Biostatistics for Evidence-Based Practice, 3 credits (Graduate in a distance education/on-line format). Touro University Nevada School of Nursing.

Instructor, NURS 775, Statistical Methods for Nursing Research I-Univariate Statistics, 3 credits (Graduate in a distance education/on-line format). UNLV School of Nursing.

2011 Fall
Co-Instructor, NURS 780, Research Methods in Nursing, 3 credits (Graduate in a distance education/on-line format). UNLV School of Nursing.

Co-Instructor, OCCT 523, Evidence-Based Research and Qualitative Research Methodology, 3 credits, (Undergraduate in a traditional/face-to-face format). Touro University Nevada School of Occupational Therapy.

2011 Summer
Instructor, NURS 776, Statistical Methods for Nursing Research II-Multivariate Statistics, 3 credits (Graduate in a distance education/on-line format). UNLV School of Nursing.

2011 Spring
Teaching Assistant, EPY 702, Research Methods, 3 credits (Graduate in a traditional/face-to-face format). UNLV Department of Educational Psychology.

Teaching Assistant, EPY 303, Educational Psychology, 3 credits (Undergraduate in a traditional/face-to-face format). UNLV Department of Educational Psychology.
Teaching Assistant, *EPY 451, Educational Assessment*, 3 credits (Undergraduate in a traditional/face-to-face format). UNLV Department of Educational Psychology.

2010 Fall
Teaching Assistant, *EPY 702, Research Methods*, 3 credits (Graduate in a traditional/face-to-face format). UNLV Department of Educational Psychology.

2010 Summer
Instructor, *NURS 776, Statistical Methods for Nursing Research II-Multivariate Statistics*, 3 credits (Graduate in a distance education/on-line format). UNLV School of Nursing.

*Courses Prepared to Teach*

Research Methods (Quantitative and Qualitative)
Univariate Inferential Statistics (including descriptive statistics)
Multivariate Statistics
Multiple Regression and Path Analysis with Observed Variables
Latent Variable Models: Factor Analysis and Structural Equation Modeling
Cognitive Development
Human Learning and Cognition
Cognition and Instruction
Motivation
Problem Solving and Reasoning

*Developed Courses*

*NUR 707, Biostatistics for Evidence-Based Practice*
*NURS 775, Statistical Methods for Nursing Research I - Univariate Statistics*
*NURS 776, Statistical Methods for Nursing Research II - Multivariate Statistics*
*NURS 780, Research Methods in Nursing*

**IV. Publications**

*Refereed Publications*


**Book Chapters**

Submitted Manuscripts


V. Research Presentations

Refereed Posters: National


Candela, L. L., & **Gutierrez, A. P.** (2010, May). *Nurturing a research collaboration to prevent and reduce obesity in teenagers*. Poster session presented at the Obesity Treatment and Prevention National Conference 2010, Washington, DC.


Refereed Posters: State


Refereed Papers & Symposia: National


Gutierrez, A. P., & Dougherty, E. (2009, August). Mixed-method approach in evaluation research. In L. G. Putney (Chair), The role of the evaluation process in action research and teacher education. Symposium conducted at the national meeting of The Association of Teacher Educators, Reno, NV.

Refereed Papers & Symposia: Regional


VI. Research in Progress


Candela, L. L., Keating, S. B., & Gutierrez, A. P. (in progress). *Nurse faculty members’ intention to become, stay, or leave the faculty role*. Manuscript in preparation.


VII. Contracts and Grants

(Date Submitted, Role; Title; Agency; Dollar Amount; Funding Outcome and Amount [if funded])

- **December 2010** Co-Investigator; "Research Methods Class is Boring"... *Think Again! Effects of Instructional Support, Emotions, and Utility Value on Situational Interest, and Consequently, Student Engagement*; UNLV Graduate & Professional Student Association (GPSA); $1,152; **Funded for $1,000**.

- **September 2010** Fellow; Marjorie Barrick Fellowship; UNLV Graduate College; $14,000; **Funded for $14,000**.

- **August 2010** Statistics Consultant/Project Coordinator; *Collaborative Approach to Expanding RN to BSN Education*; Human Resources and Services Administration (HRSA)—Nurse Education, Practice, and Retention (NEPR); $723,862; **Funded for $723,862**.

- **November 2009** Co-Investigator; *Influences on Cognitive Engagement and Learning in Graduate Level Research Methods Courses*; UNLV GPSA; $750; **Funded for $500**.

- **August 2009** Co-Principal Investigator; *Translating the Diabetes Prevention Program (DPP) to a School Setting*; UNLV School of Nursing (SON); $50,000; **Funded for $46,000**.

- **November 2008** Co-Investigator; *Assessing the Biobehavioral Correlates of Low Back Pain Progression to Disability*; National Institutes of Health (NIH); $388,953; Not funded.
September 2008  Co-Investigator; Reducing Disability Associated with Work-Related Musculoskeletal Disorders in Hispanic Construction Workers; The Center for Construction Research and Training (CPWR); $596,662; Not funded.

July 2008  Community Liaison; English and Health Literacy for Hispanic Workers and Families; National Center of Minority Health and Health Disparities (NCMHD); $375,000; Not funded.

March 2007  Researcher; Establishing Biomarkers for Low Back Pain Disability; UNLV Research Development Award; $9,500; Funded for $9,500.

VIII. University Service

2007 – 2008  Member, Professional Staff Committee, University of Nevada, Las Vegas

2008 – 2010  Member, Academic Freedom and Ethics Committee of the Faculty Senate, University of Nevada, Las Vegas.

2010 – 2011  Member, Committee on Graduate Student Funding, University of Nevada, Las Vegas

2010 – 2011  Member, Fiscal Affairs Committee, University of Nevada, Las Vegas

2010 – 2011  Member, Student Coordinating Committee, Department of Educational Psychology, University of Nevada, Las Vegas

IX. Honors, Awards, and Academic Memberships

 Honors and Awards

2010  Recipient, Marjorie Barrick Fellowship, University of Nevada, Las Vegas. Awarded in recognition of excellence in research and scholarship.


2010  Recipient, 2nd Place in Graduate Research Forum Poster Presentation. Graduate and Professional Student Association (GPSA), University of Nevada, Las Vegas.
2009  Honorable Mention List, Ford Foundation Predoctoral Fellowship Competition.


2006  Recipient, Nevada Centennial Medallion. Awarded for having earned the highest GPA in my graduating undergraduate class.

2004  Alumnus, Honors College, University of Nevada, Las Vegas.

2001-2004  Recipient, Dean’s Honor Award Scholarship. Awarded in recognition of excellence in academic achievement.

2004  Recipient, Michael Morales Hispanic in College Award. Awarded in recognition of excellence in academic achievement.

2003  Recipient, National Society of Collegiate Scholars Diligence Award. Awarded in recognition of leadership and community service.


2002  Recipient, Outstanding Academic Achievement Award. Awarded in recognition of excellence in academic achievement.


Academic Memberships
2002 – Present  Member, Phi Kappa Phi National Honor Society.

2002 – Present  Member, National Society of Collegiate Scholars.

2001 – Present  Member, Phi Eta Sigma National Honor Society.

2001 – Present  Member, McNair Scholars Institute.
X. Professional Memberships

2008 – Present  American Psychological Association (APA). Division 15, Educational Psychology.


XI. Other Service and Leadership

2011 – Present  Peer-Reviewer, The Internet and Higher Education

2010 – Present  Peer-Reviewer, American Journal of Industrial Medicine

2010 – 2011  Vice-President and Chair of Grants Committee, Graduate & Professional Student Association (GPSA), University of Nevada, Las Vegas.

2009 – Present  Peer-Reviewer, American Educational Research Association (AERA), Special Interest Group (SIG)-Mixed Method Research; SIG-Research Use; Division H-Research, Evaluation, and Assessment in Schools, Section 1-Applied Research in the Schools and Section 2-Program Evaluation in School Settings; Division D-Measurement and Research Methodology, Section 3-Qualitative Research Methods

2006  Volunteer Coordinator, Clark County Homeless Youth Count Project 2006.

2001 – 2002  Volunteer Tutor/Mentor, University of Nevada Las Vegas Tutoring & Mentoring Program.

XII. Language Fluency (Read, Write, and Speak)

Spanish, English, and French

XIII. Professional and Academic References

Dr. Gregory Schraw, Professor, UNLV Department of Educational Research, Cognition, & Development
E-mail: gschraw@unlv.nevada.edu  Phone: (702) 895-2606
Address:
4505 Maryland Parkway, Box 453003
Las Vegas, Nevada 89154-3003
Dr. Gwen Marchand, Assistant Professor, UNLV Department of Educational Research, Cognition, & Development  
E-mail: gwen.marchand@unlv.edu  Phone: (702) 895-4303  
Address: 
4505 Maryland Parkway, Box 453003 
Las Vegas, Nevada 89154-3003

Dr. LeAnn G. Putney, Professor and Chair, UNLV Department of Educational Research, Cognition, & Development  
E-mail: putneyl@unlv.nevada.edu  Phone: (702) 895-4879  
Address: 
4505 Maryland Parkway, Box 453003 
Las Vegas, Nevada 89154-3003

Dr. Eunsook Hong, Professor, UNLV Department of Educational Research, Cognition, & Development  
E-mail: eunsook.hong@unlv.edu  Phone: (702) 895-3246  
Address: 
4505 Maryland Parkway, Box 453003 
Las Vegas, Nevada 89154-3003

Dr. Lori Candela, Associate Professor and Chair, Department of Psychosocial Nursing  
E-mail: lori.candela@unlv.edu  Phone: (702) 895-2443  
Address: 
4505 Maryland Parkway, Box 453018 
Las Vegas, Nevada 89154-3018

Dr. Carolyn Yucha, Dean, UNLV School of Nursing and School of Allied Health Sciences  
E-mail: carolyn.yucha@unlv.edu  Phone: (702) 895-3906  
Address: 
4505 Maryland Parkway, Box 453018 
Las Vegas, Nevada 89154-3018

Dr. Michele C. Clark, Associate Professor, UNLV School of Nursing  
E-mail: michele.clark@unlv.edu  Phone: (702) 895-5978  
Address: 
4505 Maryland Parkway, Box 453018 
Las Vegas, Nevada 89154-3018