cogent
education

EDUCATIONAL PSYCHOLOGY & COUNSELLING | RESEARCH ARTICLE

The effects of strategy training and an extrinsic incentive on fourth- and fifth-grade students' performance, confidence, and calibration accuracy

Antonio P. Gutierrez de Blume[1]*

*Corresponding author: Antonio P. Gutierrez de Blume, Department of Curriculum, Foundations, and Reading, Georgia Southern University, P.O. Box 8144, Statesboro, GA 30460-8144, USA
E-mails: agutierrez@georgiasouthern.edu, agutierrez_lv@yahoo.com

Abstract: This study investigated the influence of strategy training instruction and an extrinsic incentive on American fourth- and fifth-grade students' ($N = 35$) performance, confidence in performance, and calibration accuracy. Using an experimental design, children were randomized to either an experimental group (strategy training and an extrinsic incentive) or a comparison group in an after-school Brain Train Club. Data on performance and confidence ratings were collected pretest and again at posttest. Accuracy was calculated by comparing confidence in performance to actual performance. Results revealed that students in the experimental group demonstrated improved performance, confidence, and calibration accuracy compared to those in the comparison group. Implications for learning and instruction and avenues for future research are discussed.

Subjects: Social Sciences; Behavioral Sciences; Education

Keywords: calibration; children; performance; confidence; strategy training; metacognition

1. Introduction

Metacognitive awareness has been defined as having two main dimensions, knowledge and regulation (Brown & Palincsar, 1989; Palincsar & Brown, 1984; Schraw & Dennison, 1994). Knowledge

ABOUT THE AUTHOR

Antonio P. Gutierrez de Blume, PhD, is currently an Assistant Professor of Research at Georgia Southern University. He is interested in researching metacognition under the theory of self-regulated learning. More specifically, he is interested in how learners monitor their comprehension during learning episodes. His program of research includes examining the effects of dispositional characteristics (e.g. various aspects of motivation) and learning strategy training on learners' calibration (accuracy and bias), confidence, and performance as well as investigating the latent dimensions of calibration to improve its measurement.

Antonio P. Gutierrez de Blume

PUBLIC INTEREST STATEMENT

Understanding how children learn is an important endeavor. In this research article, I explore the mechanisms by which American elementary school students (fourth and fifth graders) convey what they know and do not know about a topic of interest, a key element of metacognitive monitoring known as calibration. Research suggests that children begin to develop comprehension monitoring skills and control of their cognitive processes by the age of 9–10 years. In addition, research indicates that learners are not very accurate in their calibration. Therefore, in this study, I outline an educational intervention that teaches crucial strategies to help students more effectively learn and monitor their comprehension. I also provide them with an extrinsic incentive to motivate them to apply strategies effectively given task demands. Results show that the intervention was successful at aiding these young learners to better monitor and control their learning, and hence calibrate their performance more effectively.

encapsulates declarative (knowledge of cognitive skills and strategies), procedural (knowledge of how to apply strategies), and conditional (knowledge of when where and why to apply strategies given task demands). Regulation, on the other hand, subsumes five elements: planning (anticipating cognitive resources and how to effectively allocate them to accomplish goals), information management (skills of how to manage and relate large volumes of information which may be discrepant), debugging (how to correct errors in judgment while learning), comprehension monitoring (monitoring one's understanding of information), and evaluation of learning (the act of reflecting after a learning episode and making appropriate adjustments for more effective future learning) (Schraw & Dennison, 1994; Sperling, Howard, Staley, & DuBois, 2004). All of these components are interrelated yet distinct. Previously, general metacognitive awareness has been researched in tandem with metacognitive monitoring ability and was found to be related (e.g. Schraw, 1998, 2001; Schraw & Graham, 1997).

Metacognitive monitoring is the process of judging what one knows and does not know about a topic (Huff & Nietfeld, 2009; Thiede, Griffin, Wiley, & Redford, 2009). In the metacognitive monitoring literature, judgments regarding one's metacognitive monitoring ability have been variously referred to as feelings of knowing (FOKs), judgments of learning (JOLs), or ease of learning judgments (EOLs; Schraw, 2009). A FOK is defined as a feeling that information that is necessary during a learning episode is accurately known and can be successfully recalled in the future when necessary, say during a criterion task such as a test; thus, it involves both a cognitive assessment of one's memory and a prediction (before the task; Hart, 1965; Metcalfe, 1986; Nelson, 1992; Schraw, 2009) or postdiction (after the task). Research suggests that learners are more accurate in their postdiction judgments when compared to predictions (e.g. Bol & Hacker, 2001; Bol, Hacker, O'Shea, & Allen, 2005; Gutierrez & Price, in press). On the other hand, JOLs are commonly referred to as judgments of learners' ability to convey how effectively something has been learned (Nelson & Dunlosky, 1991; Schraw, 2009). Finally, EOLs are referred to as judgments of the relative ease or difficulty of to-be-learned information (i.e. they are made in advance of learning novel information; Nelson & Narens, 1990).

Calibration is an outcome from FOK judgments and it refers to the relation between criterion task performance and a judgment about that performance (Boekaerts & Rozendaal, 2010; Efklides, 2008; Winne & Nesbit, 2009). The literature on calibration distinguishes between *relative* versus *absolute* judgments (Schraw, Kuch, Gutierrez, & Richmond, 2014). *Absolute* accuracy judgments refer to comparisons between confidence in performance and actual performance (e.g. G Index), whereas *relative* judgments refer to the extent to which said judgments discriminate performance (e.g. sensitivity, specificity, gamma, *d'*; Schraw et al., 2014; Serra & Metcalfe, 2009). Because these various judgments are related yet distinct and presumably tap into different aspects of memory (Leonesio & Nelson, 1990), choice of which types of judgments to measure has important implications for the inferences and conclusions researchers can draw from their data. Generally, these metacognitive judgments are collected as confidence ratings on a referent task such as tests and exams. Thus, calibration is a comparison of confidence in performance and actual performance, outcomes of which are calibration accuracy and bias (Boekaerts & Rozendaal, 2010; Efklides, 2008; Schraw, 2009; Winne & Nesbit, 2009). Calibration accuracy refers to how accurate one's confidence in performance is when compared to actual performance, whereas bias refers to the direction of the miscalibration (i.e. signed difference underconfidence or overconfidence) (Schraw, 2009). In the present investigation, learners were asked to make *absolute* accuracy judgments, as this was most appropriate for the aspect of metacognitive monitoring examined here. Thus, the purpose of the present study was to evaluate the effect of a compact, yet comprehensive, strategy training regimen that focused on learners' test-taking and study skills—all essential for effective self-regulated learning—combined with an extrinsic incentive on fourth- and fifth-grade students' performance, confidence in performance, and *absolute* calibration accuracy in an experimental pretest/posttest design.

1.1. Contributions to the extant literature on calibration

The present investigation contributes to the extant literature on metacognitive monitoring in that it is the first investigation of this type incorporating this population. Hitherto, much of the metacognitive monitoring literature investigating the effects of strategy training regimen on calibration

accuracy has focused mainly on adult learners, mainly college undergraduates. This poses a problem to the extent that generalizations from these studies apply mainly to other samples of adult learners and not children or adolescents. Hence, this research endeavor extends previous research by providing much needed understanding of how these phenomena operate in still cognitively and metacognitively developing learners, namely fourth- and fifth-grade students. Findings have shown that children of this age (9 to 10 years) are experiencing tremendous metacognitive development (e.g. Fatzer & Roebers, 2012; Roebers, 2002), and thus are honing important self-regulated learning skills such as their ability to monitor their comprehension during learning episodes, and thus this study will allow scholars and educators to better understand the relation among cognitive (performance), metacognitive (calibration accuracy), and motivation (use of extrinsic incentives) variables at this critical time in self-regulation of learning development.

Previous researchers have employed a variety of strategy training interventions. For instance, Bol and Hacker (2001) used practice tests when compared to traditional review and Bol et al. (2005) incorporated overt practice in their study. Similarly, Hacker, Bol, and Bahbahani (2008) used reflection and incentives, whereas Bol, Hacker, Walck, and Nunnery (2012) incorporated a comparison between individual and group guidelines. Other researchers have employed some form of strategy instruction to influence metacognitive monitoring accuracy (e.g. Gutierrez & Schraw, 2015; Dunlosky, Rawson, & Middleton, 2005; Huff & Nietfeld, 2009; Nietfeld, Cao, & Osborne, 2005; Nietfeld & Schraw, 2002). Other varied educational interventions used by researchers have examined the effects of general self-regulated learning training (Azevedo & Cromley, 2004), self-explanations of reading comprehension (McNamara, 2004), main idea comprehension (Jitendra, Kay Hoppes, & Xin, 2000), and a comparison of the infusion method and the instrumental enrichment program (Lizarraga, Baquedano, Mangado, & Cardelle-Elewar, 2009). As is evident, in the literature on educational interventions, researchers have employed a variety of methods; however, it is important to note that these researchers did not necessarily all target metacognitive monitoring per se. The present investigation employs a strategy intervention similar to those used previously (e.g. Huff & Nietfeld, 2009; Nietfeld & Schraw, 2002) but also incorporates an extrinsic incentive similar to the one used by Hacker et al. (2008) and Gutierrez and Schraw (2015).

With respect to population, several researchers have conducted related research on various aspects of self-regulated learning with a variety of populations ranging from children to adolescents. Even though these studies did not specifically address strategy training or instruction, they warrant brief mention here. Barnett and Hixon (1997), for instance, examined the effect of grade and domain on second-, fourth-, and sixth-grade learners' tests score predictions. Chen and her colleagues (Chen, Cleary, & Lui, 2015; Digiacomo & Chen, 2016) in a series of studies investigated self-regulated learning skills in mathematics among a cohort of middle school students. Along a similar vein, Bol, Riggs, Hacker, and Nunnery (2010) explored calibration accuracy in mathematics among middle school learners. Hence, while the present study is informed by these studies, it represents an important extension of this work by employing strategy instruction with an incentive among children in an experimental, pretest/posttest design.

2. Theoretical Framework

Self-regulated learning (SRL) theory encompasses cognition, metacognition, and motivation. Several models of SRL have been proposed in the literature. 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). The theory of SRL posits that metacognitive monitoring is a crucial component of self-regulated learning. Previous findings indicate that proficient self-regulated learners are more able to leverage cognitive resources, select appropriate strategies to match task

demands, effectively monitor comprehension during learning episodes, and exhibit greater meta-cognitive awareness when compared to less proficient learners (Boekaerts & Rozendaal, 2010; de Bruin & van Gog, 2012; Efklides, 2008; Winne & Nesbit, 2009). Greene and Azevedo (2007) describe Winne and Hadwins' model of SRL as more complex because it incorporates monitoring and control processes at each phase of learning as well as distinguishing between task understanding and goal setting. They also describe similarities among various models of SRL. Therefore, by enhancing indi-viduals' comprehension monitoring through explicit instruction, it is plausible for them to improve higher order thinking skills such as calibration accuracy and metacognitive monitoring skills. Thus, the present investigation incorporates all three components of SRL theory—cognition, metacogni-tion, and motivation (i.e. use of an extrinsic incentive).

3. Developmental trajectory of metacognition

Given the emphasis of children rather than adults in the present study, it is important to understand the developmental trend of metacognition across the lifespan. Although these research endeavors are cross-sectional and largely descriptive in nature, they provide an understanding of how relevant metacognitive skills function at different points along the developmental continuum. Studies on the development of metacognition reveal that metacognitive skills are present as early as 5–7 years of age (Ghetti & Angelini, 2008; Krebs & Roebers, 2010; Lyons & Ghetti, 2011; Roderer & Roebers, 2010). However, these studies indicate that self-awareness and monitoring processes begin to emerge at this early stage of metacognitive development. More advanced metacognitive skills such as control and inhibitory skills begin to surface in children approximately between 8 and10 years of age (Roebers, Schmid, & Roderer, 2009; Schneider, Knopf, & Sodian, 2009). Research results have shown that the metacognitive skills requisite for effective comprehension monitoring arise at about 9 or 10 years of age, in which children begin to display adequate control and monitoring of their own cognitive resources and abilities (Schneider, Visé, Lockl, & Nelson, 2000; Schneider et al., 2009). Interestingly, children begin to develop the ability to calibrate their performance at about age 8 (Roebers, von der Linden, Schneider, & Howie, 2007; Son, 2005; Souchay & Isingrini, 2004), although their calibration accuracy tends to be rather poor at this age. Additional research in later years found that individuals hone their metacognitive skills such as metamemory, reflection, comprehension monitoring, and self-regulation throughout adolescence (Kitchner, 1983; Veenman, Kok, & Blöte, 2005). Nevertheless, it is during adulthood that individuals' metacognitive skills are most proficient, suggesting that as individuals age, their metacognitive monitoring improves (Koriat, Ackerman, Adiv, Lockl, & Schneider, 2013; Vukman, 2005). It is perhaps for this reason that the majority of re-search on metacognitive monitoring has involved adult learners rather than children. Therefore, this study elucidates how metacognitive monitoring functions in children between 9 and 10 years old, in which metacognitive monitoring and control processes are present.

4. Metacognitive monitoring training and performance, confidence, and calibration accuracy

McCormick (2003) and Pressley and Harris (2006) argued that metacognitive monitoring training is an optimal way to improve comprehension monitoring. A series of studies suggested that individu-als' performance and calibration accuracy benefit from explicit metacognitive monitoring training. For example, research on reading comprehension indicated a positive effect of specific metacogni-tive monitoring training on rereading (Dunlosky & Rawson, 2005; Dunlosky et al., 2005; Griffin, Wiley, & Thiede, 2008). It is important to note that research on the influence of metacognitive strategy training on performance, confidence in performance, and calibration accuracy has been inconclu-sive. Some studies have demonstrated a positive effect of metacognitive strategy training on perfor-mance, confidence judgments, and calibration accuracy (Gutierrez & Schraw, 2015; Huff & Nietfeld, 2009; Magliano, Little, & Graesser, 1993; Nietfeld et al., 2005; Nietfeld, Cao, & Osborne, 2006; Nietfeld & Schraw, 2002). In contrast, research also points to the lack of any effects of training on these out-comes. For instance, Bol and Hacker (2001), Bol et al. (2005), and Hacker et al. (2008) found that metacognitive monitoring training had no effect on calibration accuracy and confidence, although it marginally improved performance. Differences in findings among these studies may be due to meth-odological choices (cf. Gutierrez & Schraw, 2015; Bol et al., 2005).

Despite the fact that developmental research on metacognition previously surveyed has demonstrated that children begin to realize metacognitive monitoring and control skills by 9 to 10 years of age, Beal (1996) argues that children in the elementary school years are highly miscalibrated in their reading comprehension skills such that they are overconfident in their performance, underscoring the need to implement metacognitive monitoring training among this population of learners. Although there is a great wealth of research on the effects of metacognitive monitoring training on performance, confidence in performance, and calibration accuracy among adult learners, the research on these constructs among children and adolescents is scarce. Among a cohort of middle school students, Cataldo and Cornoldi (1998) showed that explicitly instructing students to monitor their reading comprehension did not significantly influence comprehension monitoring. In contrast, de Bruin, Thiede, Camp, and Redford (2011) demonstrated that instructing children on generating keywords improved metacomprehension monitoring accuracy in reading among fourth- and sixth-grade students. In sum, the scant research on the role metacognitive monitoring training plays in performance, confidence, and calibration accuracy improvement among children and adolescence varies widely. Yet, the general consensus among these studies is that monitoring skills are malleable and trainable and that, for those studies revealing an effect, training has positive effects for these outcomes, albeit small.

5. The role of extrinsic incentives

The extant literature on the role of incentives converges on the finding that incentives differentially affect performance and calibration accuracy as a function of the type of incentive (Hogarth, Gibbs, McKenzie, & Marquis, 1991; Schraw, Potenza, & Nebelsick-Gullet, 1993; Yates, 1990). Extrinsic incentives are driven by tangible rewards, such as money and extra credit, which instead draw on individuals' performance on a criterion task (e.g. to outperform others or to avoid underperforming). Although incentives have been shown to influence task performance, the effect is not always positive (Kleinsorge & Rinkenauer, 2012), especially when a task is easy or unappealing (Bailey & Fessler, 2011). For example, Sankaran and Bui (2001) found that when individuals no longer received incentives to motivate performance on an intrinsically enjoyable task of their choosing, their performance and interest on the task dampened. Sinkavich (1994) provided an extrinsic incentive—in this case, extra credit points—to enhance monitoring accuracy. In addition, when individuals are provided an incentive to sustain high achievement on a referent task, this significantly decreased their incidental learning (i.e. learning that occurs unintentionally or that which is not initially planned; Hogarth et al., 1991), presumably because attention is focused on performance rather than actual learning.

The literature on the effects of incentives on performance and calibration accuracy has reported mixed findings. For instance, Hacker et al. (2008) found that students in the incentives condition significantly improved calibration accuracy but this was the case only for lower performing students. On the other hand, research by Schraw et al. (1993) found that incentives provided explicitly to improve calibration accuracy improved individuals' monitoring accuracy, whereas incentivizing them for results on specific tasks had no effect on performance, suggesting that individuals invoke subjective feelings of knowing when calibrating their performance rather than more objective information such as judging item difficulty on performance assessments. On another vein, findings have indicated nil effects of incentives on either calibration accuracy or performance (see, Hogarth et al., 1991; for a review), albeit one study revealed that individuals who received incentives performed significantly better than those who were exposed to strategy training (Tuckman, 1996). Conversely, Gutierrez and Schraw (2015) found that extrinsic incentives improve performance, confidence, and calibration accuracy among adult learners. In sum, the extant literature on these topics points to a dynamic interplay between incentives, calibration accuracy, and performance (e.g. Hacker et al., 2008; Schraw et al., 1993; Tuckman, 1996), thus underscoring the need for additional research to clarify this complex dynamic.

6. The present study

The purpose of the present investigation was to examine the influence of a compact strategy training regimen targeting learners' self-regulated learning skills (e.g. test-taking and study strategies)

combined with an extrinsic incentive on fourth- and fifth-grade students' performance on a declarative knowledge assessment, confidence in performance, and calibration accuracy. This was evaluated using an experimental design with a repeated measures component. The compact strategy training was developed using previous research (Gutierrez & Schraw, 2015; Nietfeld & Schraw, 2002; Schraw, 1998; Volet, 1991). This line of inquiry posits that strategy instruction targeted to increase calibration accuracy should in tandem improve subsequent self-regulation of learning because more proficient monitoring presumably increases self-regulation and control of learning processes (Greene & Azevedo, 2010; Nelson & Narens, 1990; Winne & Nesbit, 2009). Gutierrez and Schraw (2015) used the same strategy instruction implemented here among adults using a four-group experimental design (strategy training and incentive, incentive only, strategy training only, and control).

More specifically, the integrated one-hour strategy instruction intervention used seven domain-general strategies that have been demonstrated to improve learning and self-regulation (Greene & Azevedo, 2010). Individuals were expected to read, review, relate, and monitor information during learning as a result of the strategy instruction. The full strategy module is known as R^3M and was developed and tested previously with adult learners (Gutierrez & Schraw, 2015). The module is based on domain-general strategy instruction principles (Pressley & Harris, 2006) as well as specific strategies used in previous calibration research (Gutierrez and Schraw, 2015; Nietfeld & Schraw, 2002; Schraw, 1998; Volet, 1991). The present study presented the seven strategies shown in Table 1 in an integrated one-hour educational training session designed to increase proficiency with respect to self-regulation processes (i.e. strategic study, monitoring, and control processes) during learning. This strategy training represents a more comprehensive training regimen than previous research on this topic (e.g. Bol et al., 2005; Brannick, Miles, & Kisamore, 2005; Dunlosky et al., 2005;

Table 1. Summary of metacognitive strategies and their relation to calibration and theory		
Strategy	**Metacognitive processes**	**Hypothesized influence on calibration**
Review main objectives of the text and focus on main ideas and overall meaning	Monitoring	Enhances calibration through clarifying misunderstandings and tying details to main ideas
Read and summarize material in your own words to make it meaningful; use elaboration and create your own examples	Monitoring and self-regulation	Enhances calibration by transforming knowledge into something personally meaningful
Reread questions and responses and reflect on what the question is asking; go through and take apart the question paying attention to relevant concepts	Monitoring and self-regulation	Purposefully slowing information processing allows for a more accurate representation of the problem, thus decreasing errors in judgment
Use contextual cues in the items and responses, e.g. bolded, italicized, underlined, or capitalized words	Monitoring	Using contextual cues allows the mind to focus on salient aspects of the problem rather than seductive details, thereby increasing accuracy
Highlight text; underline keywords within the question to remind yourself to pay attention to them; use different colors to represent different meanings	Self-regulation	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
Relate similar test questions together and read them all before responding to any	Self-regulation	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
Use diagrams, tables, pictures, graphs, etc. to help you organize information	Monitoring and self-regulation	These strategies help simplify complex topics by breaking them down to their constituent parts; this increases accuracy by decreasing errors

Hacker et al., 2008; McNamara & Magliano, 2009; Nietfeld & Schraw, 2002). With the majority of research on the effect of strategy training on calibration focusing on adult learners rather than children who are at a critical juncture in their metacognitive development, the present investigation represents an important extension to the calibration literature.

6.1. Research question and hypothesis
The present study sought to answer the following research question.

1. What is the effect of strategy training instruction combined with an extrinsic incentive on fourth- and fifth-grade students' performance on a 20-item declarative knowledge test, confidence in performance, and calibration accuracy?

    H_1: The experimental condition (strategy training and incentive) was expected to positively affect students' performance, confidence judgments, and calibration accuracy. More specifically, students in the experimental group were predicted to improve their performance, confidence, and calibration accuracy at posttest when compared to students in the comparison group. Thus, an experimental condition x occasion (pretest, posttest) interaction was hypothesized, along with a significant main effect for experimental group. Given that the comparison group was not exposed to the treatment, a main effect for occasion was not predicted.

7. Method

7.1. Participants and sample
Participants were 35 fourth- ($n = 23$) and 5th ($n = 12$) grade students from a Title I school in the southeast United States. This sample of students consisted of 19 females (16 males), with a mean age of 9 years and 6 months ($SD = 0.57$). The majority of the students were white ($n = 26$), with five identifying as black, two as Hispanic, and two as Mixed Race. School administration reported that the racial distribution of the sample approximates that of the school at large, and hence it is representative of this particular school. Furthermore, the educator who assisted in coordinating the study reported that, with respect to ability and achievement, the sample was typical of the school population.

7.2. Design and materials
This study employed an experimental pretest/posttest design. Students were randomized into one of the two groups, strategy training and an extrinsic incentive ($n = 18$) and a comparison group ($n = 15$) who completed tasks unrelated to the experiment (i.e. played educational games provided by the school in a tablet). Therefore, the monitoring and incentive condition was manipulated between-subjects, whereas occasion (pretest, posttest) was repeated within-subjects. Table 2 presents descriptive statistics and internal consistency reliability coefficients (Cronbach's α) for the performance variable at pretest and posttest for the sample and Table 3 includes the zero-order correlations for pretest and posttest measures.

Table 2. Descriptive statistics and internal consistency reliability coefficients of performance, confidence, absolute accuracy, and bias						
Variable	**Pretest**			**Posttest**		
	M	*SD*	*α*	*M*	*SD*	*α*
Performance	65.57	16.92	0.69	67.97	17.73	0.74
Confidence	76.23	18.74	–	78.68	20.09	–
Absolute accuracy	9.11	17.96	–	9.75	12.68	–
	Lower bound, upper bound			Lower bound, upper bound		
Absolute signed bias	−22.50, 60.00			−21.05, 37.45		

Notes: Reliability coefficients were not computed for calibration (absolute accuracy and bias) and confidence judgments because these data are not survey-based; $N = 33$.

cogent ·· education

Table 3. Zero-order correlation coefficients for pretest and posttest performance, confidence, and absolute calibration accuracy			
Variable	**1**	**2**	**3**
1. Performance	–	.38*	−.60**
2. Confidence	.58**	–	.56**
3. Calibration accuracy	.55**	.36*	–

Notes: Correlations above the diagonal are for pretest variables and those below the diagonal are for posttest measures; N = 33.

*p < .05.

**p < .01.

Performance and confidence judgments: Performance was assessed using a researcher-developed 20-item declarative knowledge test. Items included a variety of domains that incorporated information on what students were learning at school at the time of the experiment. Sample items included, "What is the answer to 203–159?"; "The sixth planet from the sun is:"; and "What is the answer? 17 2 × 5?" All items had only one correct response and four available choices. Correct responses were dummy coded as 1 and incorrect responses as 0. This permitted for a continuous raw score for performance, which was subsequently transformed to a proportion of correct responses (i.e. percent correct) to facilitate interpretation and to more logically compare confidence judgments to actual performance. Confidence judgments were collected locally (i.e. item-by-item) by asking participants the following question after each item, "How confident are you in your response to this item?" Participants responded to confidence ratings on a 0 to 100 continuous scale. They were instructed that any value from 0 to 100 is valid and that the closer the value is to 0, the less confidence they have whereas choosing a value closer to 100 indicated greater confidence. Confidence ratings were then averaged across all items to produce a mean confidence rating.

Absolute calibration accuracy: Absolute accuracy scores were calculated by comparing participants' confidence in performance against their actual assessment percent correct score—that is, the *residual score approach*. Raw scores were converted to a proportion and subtracted from the composite confidence in performance ratings to calculate absolute accuracy. Absolute values—to avoid negative values—of the discrepancy between students' self-reported level of confidence and actual performance were used as the measure of accuracy to facilitate interpretation. Comparing confidence in performance against actual performance yielded continuous, *absolute* calibration accuracy scores, as described by Schraw (2009). A score of "0" indicates perfect calibration; on the other hand, the higher the value, and thus the farther away from "0", the greater the inaccuracy. In essence, the higher the accuracy scores, the greater the *miscalibration* exhibited by the participant. In addition, the upper bound and lower bound signed scores were reported to gather information regarding participants' bias (i.e. underconfidence or overconfidence).

7.3. Procedure

University IRB approval was secured before the commencement of data collection activities. The research was conducted across four one-hour sessions that occurred every Tuesday after school between 2:15 and 3:15 pm. Two separate data collection sessions occurred from September to December 2015 to recruit enough participants for an adequate sample size. Participants first completed child assent forms after receiving parent permission forms for voluntary participation at the beginning of Session 1. Next, students completed pretest measures, including the performance assessment with item-by-item confidence in performance ratings (i.e. confidence ratings were collected immediately after participants responded to each item). All data collection activities occurred online in a computer lab at the participating school via Qualtrics. Participants randomized to the experimental R³ M condition were taken to a separate room and received an advance organizer activity in which they were provided with the strategies in Table 1 and the instructions for receiving

the incentive for improved performance at posttest. These instructions were read aloud to the experimental group to obviate confusion, and students were permitted to ask questions if they did not understand any part of the instructions. The choice of candy as the incentive was based on the professional advice of three educators in the participating school. Finally, students received a summary of the strategy training to take place in Session 2 and were instructed to reflect on the seven strategies and how they apply to their learning. Students in the comparison condition completed educational games on tables (e.g. Math Genius, Chess, and/or History) unrelated to the present study in a separate room for the remainder of Session 1.

During Session 2, participants in the experimental condition received the strategy training. The training session involved direct instruction and individual practice in using strategies with scaffolded feedback in a face-to-face lecture format. First, participants 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 author 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 using an expository text on emotions as well as an 18-item practice test. The researcher walked around during this apply practice portion 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. Students not receiving the monitoring instruction participated in an activity unrelated to the experiment in a separate room—namely, playing educational games on a tablet provided by the school.

Session 3 involved providing participants in the experimental condition with a brief 10-min summary of the activities of Session 2 while the comparison condition participants played educational games on a tablet in a separate room. Next, all participants returned to the computer lab to complete the posttest measure—the performance assessment in which confidence ratings were again completed immediately after responding to each item. After completion of the posttest, students were permitted to play educational games on a tablet. Finally, Session 4 involved providing the comparison group with the strategy training. Participants in the experimental condition who met the criteria for the incentive were given a candy of their choice by the educator who assisted in coordinating the study. Additional candy was given to school administration to distribute to all remaining participants after the conclusion of Session 4.

7.4. Strategy instruction intervention

The one-hour strategy training intervention was implemented by introducing each of the seven strategies in Table 1 sequentially. During the modeling and scaffolding portions of the training, specific examples were used to increase the utility and application of the strategy in authentic learning situations. For instance, when explaining the strategy on diagramming, participants were shown how this strategy could be applied when attempting to learn and understand the notions of division and multiplication by drawing squares or circles to understand the concept of breaking larger numbers into equal groups and remainders or compounding in multiplication. Another example used was how students could use highlighting and underlining effectively to facilitate reading comprehension in math word problems and when attempting to understand main ideas in texts. Students were permitted to ask questions for clarification and to practice the strategy with other concepts of their choice. This aided students in understanding the "when, where, and why" (i.e. conditional metacognitive knowledge) of strategy use in their everyday learning environment. At the end of the training, participants were instructed to reflect how they could apply any of the strategies while learning at school, albeit these process data were not captured for subsequent data analysis.

7.5. Data analysis plan

Prior to data analysis, data were first screened for univariate outliers and evaluated against requisite statistical assumptions according to the procedures outlined by Tabachnick and Fidell (2013) via IBM

SPSS version 22. No extreme outliers that would otherwise undermine the trustworthiness of the data were detected. A missing values analysis demonstrated that two cases (one from each group; 0.03%) had missing data. Systematic bias in the pattern of missing data could pose a problem to the trustworthiness and accuracy of the data, and hence the validity of the inferences and conclusions drawn from such data. Thus, Little's MCAR χ^2 was requested from the missing values analysis to ascertain if the pattern of missing data was missing at random (Little & Rubin, 1989; Schafer & Graham, 2002). A significant χ^2 (i.e. $p < .05$) would suggest that the pattern of missing data is 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, the result of this test was non-significant, $p = .53$, suggesting that the missingness pattern in the data was missing at random. Thus, all final analyses were conducted with 33 complete cases.

Data were also tested for univariate normality using histograms with the normal curve overlay and skewness and kurtosis statistics. Data approximated a normal distribution. Furthermore, data were evaluated for assumptions including multicollinearity (all correlations were $< r = .85$), homogeneity of variance (all Levene's test p-values $> .05$), and sphericity. All of the aforementioned assumptions were met, and thus data analysis proceeded without making any adjustments to the data.

A series of 2 (experimental condition: monitoring training and incentive, comparison) × 2 (occasion: pretest, posttest) factorial mixed-model (between-subjects and within-subjects) analyses of variance (ANOVAs) were conducted to answer the research questions. In each of these analyses, performance, confidence judgments, and calibration accuracy served as the dependent variables separately. The Bonferroni adjustment to statistical significance was used to control for the family-wise Type I error rate inflation. All effect sizes for the factorial ANOVA results were reported as *partial* η^2 (η_p^2). Cohen (1988) specified the following interpretive guidelines for η_p^2: .010–.059 as small; .060–.139 as medium; and $\geq .140$ as large.

8. Results

Pearson's product moment correlation coefficients in Table 3 were all within range, in the theoretically expected direction, and all of the correlations were weak to strong and statistically significant. Of special significance, absolute accuracy was significantly and inversely associated with performance at pretest, and descriptive statistics in Tables 2 and 4 show that participants' signed absolute bias scores tended to be positive, suggesting that students were generally overconfident in their confidence judgments. Moreover, this correlation was attenuated at posttest, indicating that students generally adjusted confidence ratings from pretest to posttest performance. Descriptive statistics by group are presented for all relevant variables in Table 4.

8.1. Establishing group equivalence at baseline

Prior to proceeding with data analyses, group equivalence was evaluated for all variables at baseline. Significant differences among groups on pretest measures would point to the need to control for those variables. However, a series of independent samples t-tests revealed no statistically significant differences between the experimental and comparison groups on any of the pretest variables (all p-values $\geq .31$). Moreover, analyses between grades demonstrated that grade level had no significant effect on any of the outcome measures (all p-values $\geq .26$). Therefore, all analyses proceeded as planned.

8.2. Main analyses

Performance: Performance results revealed a significant experimental condition x occasion interaction, $F_{(1,31)} = 3.98$, $p = .03$, $\eta_p^2 = .104$ (see Figure 1). Neither main effect reached statistical significance (occasion, $p = .10$; experimental condition, $p = .29$). Simple contrasts with the Bonferroni adjustment for multiple comparisons revealed no significant differences for either group within occasion, all p-values $\geq .90$. Conversely, simple main effects with the Bonferroni adjustments for multiple comparisons indicated that the experimental group exhibited improved performance at posttest when compared to pretest performance, $\eta_p^2 = .090$, $p = .04$. No significant difference was found for the comparison group.

⁂ cogent ·•· education

Table 4. Descriptive statistics for pretest and posttest variables for the strategy training and incentive and comparison groups

Variable	Pretest				Posttest			
	ST + I		Comparison		ST + I		Comparison	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Performance	67.05	16.58	67.72	13.85	72.08	9.85	71.36	19.25
Confidence judgments	78.15	9.22	82.05	15.72	79.42	7.89	91.33	8.85
Absolute accuracy	12.64	11.81	15.84	13.72	8.54	5.66	15.83	10.47
Absolute bias (lower bound, upper bound)	(−11.40, 40.00)		(−22.50, 60.00)		(−6.30, 37.45)		(−21.05, 32.10)	

Notes: *N* = 33 (ST + I, *n* = 18; Comparison, *n* = 15).

Key: ST + I represents the experimental metacognitive strategy training and incentive group.

Figure 1. Performance results for the strategy training and incentive and control groups across testing occasion.


Performance Results

	ST+I	Control
Pretest	67.05	67.72
Posttest	72.08	71.36
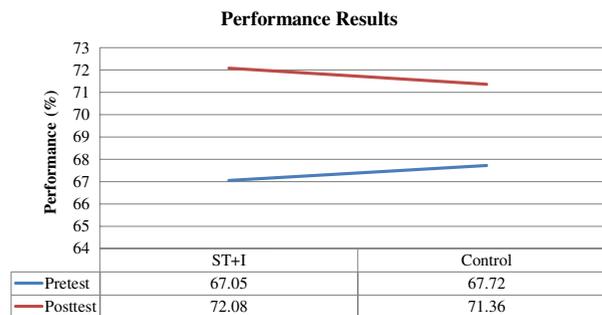
Figure 2. Confidence ratings for the strategy training and incentive and control groups across testing occasion.

Confidence in Performance Results

	ST+I	Control
Pretest	78.15	82.05
Posttest	79.42	91.33
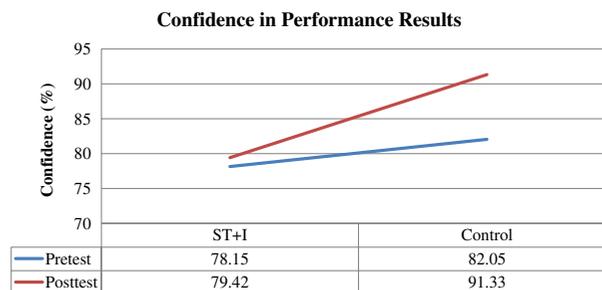
Confidence: Confidence in performance results indicated a statistically significant experimental condition x occasion interaction, $F_{(1,31)} = 4.97$, $p = .03$, $\eta_p^2 = .115$ (see Figure 2), as well as significant main effects for experimental condition, $F_{(1,31)} = 4.71$, $p = .03$, $\eta_p^2 = .176$, and occasion, $F_{(1,31)} = 5.67$, $p = .03$, $\eta_p^2 = .184$. Simple contrasts with the Bonferroni adjustment for multiple comparisons revealed that, at posttest, the comparison group reported significantly higher confidence when compared to the experimental group, $\eta_p^2 = .353$, $p = .001$. Simple main effects with the Bonferroni adjustment showed that the difference in confidence ratings for the comparison group was significantly higher at posttest than pretest, $\eta_p^2 = .23$, $p = .01$. The main effect for occasion indicated that students' confidence increased from pretest to posttest and the main effect for experimental condition found that the confidence ratings were higher for the comparison group.

Absolute calibration accuracy: There was a statistically significant experimental condition x occasion interaction, $F_{(1,31)} = 3.96$, $p = .03$, $\eta_p^2 = .090$ (see Figure 3). There was also a significant main effect

cogent ·· education

Figure 3. Absolute accuracy results for strategy training and incentive and control groups across testing occasion.

Absolute Accuracy Results

	Pretest	Posttest
ST+I	12.64	8.54
Control	15.84	15.83

Y-axis: Accuracy (Confidence - Performance)
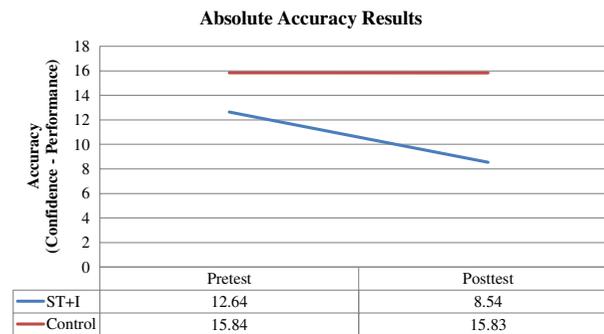
for experimental condition, $F_{(1,31)} = 4.01$, $p = .03$, $\eta_p^2 = .089$. The main effect for occasion did not reach statistical significance ($p = .42$). Simple contrasts with the Bonferroni adjustment to the p-value revealed that within occasion, the experimental group exhibited significantly greater absolute accuracy at posttest when compared to the comparison group, $\eta_p^2 = .181$, $p = .03$. Within experimental condition, simple main effects with the Bonferroni adjustment for multiple comparisons indicated that the experimental group improved absolute accuracy from pretest to posttest, $\eta_p^2 = .067$, $p = .04$, and that the comparison group showed no statistically significant change in accuracy ($p = .99$). The main effect of experimental condition showed that the experimental group exhibited more proficient absolute accuracy when compared to the comparison group.

Summary: Taken together, these results indicate that strategy training, combined with an extrinsic incentive, yields positive effects for a variety of critical learning outcomes. The compact one-hour strategy training regimen (R^3 M) presented here had practically and meaningfully significant effects on fourth- and fifth-grade students' performance, confidence, and absolute calibration accuracy. Of special significance, students randomized to the comparison group reported significantly higher confidence at posttest when compared to the experimental group, suggesting that they did not adjust confidence ratings to more accurately match performance (i.e. they remain *miscalibrated*). More importantly, the experimental group reported greater absolute accuracy at posttest than the comparison group due to improved performance and lower confidence at posttest when compared to the control group. This is in spite of the test–retest effect, which apparently had deleterious effects for the control group participants, who reported greater confidence at posttest. Although tentatively, due to a small sample and a two-group design, these findings highlight the utility of the experimental condition at improving metacognitive monitoring skills.

9. Discussion

The purpose of the present investigation was to evaluate the effects of compact strategy training and an extrinsic incentive on fourth- and fifth-grade students' performance, confidence judgments, and absolute calibration accuracy. It was predicted that students in the experimental condition would improve their performance, confidence judgments, and calibration accuracy as a function of the strategy training and incentive. The significant interaction for performance indicated that only the experimental group showed significant gains in performance while the comparison group did not. This is in line with research conducted by Gutierrez and Schraw (2015) and Nietfeld and Schraw (2002), who found that strategy training improved students' performance (cf. Bol et al., 2005 regarding overt calibration practice). The present investigation found that calibration accuracy improved significantly for the experimental group but not the comparison group. This is congruent with previous research, which found a positive effect for strategy training on students' calibration (e.g. Gutierrez & Schraw, 2015; Azevedo & Cromley, 2004; Bol et al., 2012; Greene & Azevedo, 2010; Nietfeld & Schraw, 2002). Conversely, research by Bol et al. (2005) and Hacker et al. (2008) uncovered that while training improved performance, it had no significant effect of calibration accuracy.

Findings of performance suggest that students benefited from the strategy training which improved their performance, and by extension enhanced their self-regulation of learning skills. Besides improving regulatory capacity, the incentive motivated students in the experimental group to improve performance from pretest to posttest. This positive effect of incentives on performance has been found in previous studies (e.g. Gutierrez & Schraw, 2015; Schraw et al., 1993; Tuckman, 1996; Yates, 1990). The fact that the experimental group adjusted confidence to more closely align with actual performance from pretest to posttest supports the hypothesis and is sensible, given what we know about SRL theory and metacognitive monitoring. The theory of SRL specifies that a proficient learner will be better able to monitor his comprehension (Boekaerts & Rozendaal, 2010; de Bruin & van Gog, 2012; Efklides, 2008). Likewise, those who are more proficient at monitoring their comprehension as they learn exhibit greater calibration accuracy. Therefore, those students in the experimental group exhibited appropriate self-regulation via increased comprehension monitoring, and hence were able to adjust their confidence ratings to actually more adequately match what they know and do not know to actual performance. Presumably, this enhanced monitoring led students in the experimental group to adjust confidence at posttest to more closely align to actual performance when compared to comparison group students. In essence, the increased confidence judgments of comparison group students may be, quite possibly, evidence of greater *miscalibration* when compared to those of the experimental group (see Gutierrez & Schraw, 2015).

Although strategy training has been found to enhance calibration accuracy, the influence of incentives on calibration has not been thoroughly explored. Gutierrez and Schraw (2015), however, found that an extrinsic incentive combined with strategy training targeting SLR skills improved calibration accuracy among adult learners. In the present sample of elementary school students, it may be that the study and test-taking strategies improved self-regulatory ability and comprehension monitoring skills while the extrinsic incentive provided students with the motivational edge necessary to apply the strategies appropriately. In essence, what this is demonstrating is that the strategy training provided students with the conditional knowledge needed to know when, where, and why to apply a strategy, whereas the incentive motivated students to actually apply the strategies, although this needs to be empirically examined in future research.

9.1. Implications for learning and instruction

The present investigation provided tentative support that essential learning outcomes such as self-regulated learning skills (e.g. performance, confidence judgments, and calibration accuracy) are malleable and trainable. Although with a small sample size, the present study showed the potential for educators of fourth- and fifth-grade students to adopt a strategy training regimen such as R³M that employs study and test-taking strategies to successfully enhance key monitoring skills, strategy use, and performance. R³M is appealing because it is a compact (one-hour training session) approach to improving students' performance, and by extension, metacognitive monitoring accuracy, making them more proficient at understanding what they know and do not know about a topic and constantly adjusting confidence judgments to more closely approximate these feelings of knowing and avoid *illusions of knowing* and *illusions of not knowing* (Serra & Metcalfe, 2009). For practitioners in particular, the R³M can be successfully and efficiently employed either together in a compact one-hour training such as the one utilized in this study, or they can segment the training by covering one strategy per day to provide a more in-depth modeling and scaffolding approach. Either approach can be easily incorporated in any lesson of the educator's choosing. Ultimately, educators should incorporate metacognitive monitoring instruction to their typical curricula to have longer, more sustained effects. This is important given that research such as that conducted by Nietfeld and Schraw (2002) showed that students' monitoring and performance reverted to baseline levels after a two-week delay. Use of an extrinsic incentive demonstrated that dispositional characteristics have a significant impact on strategy use, performance, confidence, and calibration accuracy. Therefore, educators should consider not only cognitive and metacognitive factors but motivational ones as well, and thus more closely align instructional practice to SRL theory. Finding ways to incentivize students to actually apply effective strategies and be flexible in strategy use as task demands fluctuate may be an effective way to improve learning outcomes.

9.2. Methodological reflections and limitations

No research endeavor involving human participants is ever without limitations. Although this study found significant and robust practical effects (i.e. moderate-to-large effect sizes given the sample size) of strategy training and an extrinsic incentive on the outcomes of interest, especially calibration accuracy, the sample size was small by comparison. Moreover, because of time and other logistical constraints, it was not possible to implement a four-group design, as described in the Avenues for Future Research section, and hence it was not possible to isolate the individual effects of monitoring training and the extrinsic incentive separately. This approach would have provided a clearer, more nuanced explanation of the relations among these variables. In spite of these limitations, however, this research study was a randomized experiment in a repeated measures context, and thus increases the validity of the inferences and conclusions drawn from the findings. In addition, this study involved children, namely fourth- and fifth-grade students, and hence extends the extant calibration literature in this understudied population.

9.3. Avenues for future research

The present investigation indicated that a compact strategy training regimen that underscores enhancement of test-taking and study skills combined with an extrinsic incentive enhances self-regulatory skills such as those examined here (e.g. increased calibration accuracy). While the R^3 M module contains domain-general strategies focusing on learners' test-taking and study strategies, it is not presently clear whether these strategies perform equally effectively across academic domains or among other spans of development. Thus, future research should explore the invariance of these strategies across multiple domains and among adolescents. Also, additional research with children is warranted to ascertain the stability of the results of the present investigation across multiple samples of this population, especially studies using more robust sample sizes. For instance, process data regarding actual application of strategies in everyday learning environments would help researchers better tailor interventions and help researchers and educators understand under what situations students actually apply (or do not apply) strategies for more effective learning. Moreover, future research should replicate these findings in more authentic, ecologically valid settings such as actual classrooms. Further, the role incentives play in influencing these learning outcomes should be further examined. For instance, future research should investigate whether intrinsically incentivizing students is more, equally, or less effective than using extrinsic incentives. Such a study would help researchers and educators compare and contrast intrinsic and extrinsic motivators. Due to the nature of conducting research in the schools, it was not logistically feasible to implement an experiment with four groups—monitoring training and incentive; monitoring training only; incentive only; and comparison—to disentangle the effects of the combined condition and the individual manipulations. Thus, future research should implement this four-group design to better understand how these manipulations work in tandem and alone in this population of children.

10. Conclusion

Our main task as educators of learners with mixed ability, achievement, and dispositions is to find the most effective means to teach critical content while maintaining fidelity to the individual differences inherent in our learners. The utility of R^3 M is that it uses domain-general strategies targeting essential self-regulated learning skills such as study and test-taking strategies, which may be beneficial for diverse learners. The present inquiry showed that this regimen of strategies is effective in improving learners' performance, confidence in performance, and calibration accuracy. Hill, Bloom, Black, and Lipsey (2008) discuss the importance of using empirical benchmarks for a more realistic approach to interpreting effect sizes, which allows for cross-study comparisons. The effect sizes (η_p^2) observed in this study, which ranged from 0.090 to 0.115, were in the medium range, and were comparable to those in other strategy training interventions for the improvement of metacognitive monitoring (e.g. Bol et al., 2005; Gutierrez & Schraw, 2015; Hacker et al., 2008; Nietfeld & Schraw, 2002), albeit these studies were with college students. Moreover, findings pointed to the need to consider cognitive, metacognitive, *and motivational* characteristics, as use of an extrinsic incentive was found to enhance the aforementioned outcomes as well. Because it is a relatively short strategy instructional intervention, it may be appealing to educators in classroom settings to assist students in becoming more self-regulated learners.

cogent •• education

Author details
Antonio P. Gutierrez de Blume[1]
E-mails: agutierrez@georgiasouthern.edu, agutierrez_lv@yahoo.com
ORCID ID: http://orcid.org/0000-0001-6809-1728
[1] Department of Curriculum, Foundations, and Reading, Georgia Southern University, P.O. Box 8144, Statesboro, GA 30460-8144, USA.

References
Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523–535. https://doi.org/10.1037/0022-0663.96.3.523

Bailey, C. D., & Fessler, N. J. (2011). The Moderating effects of task complexity and task attractiveness on the impact of monetary incentives in repeated tasks. *Journal of Management Accounting Research, 23*, 189–210. https://doi.org/10.2308/jmar-10104

Barnett, J. E., & Hixon, J. E. (1997). Effects of grade level and subject on student test score predictions. *The Journal of Educational Research, 90*, 170–174. https://doi.org/10.1080/00220671.1997.10543773

Beal, C. (1996). The role of comprehension monitoring in Children's revision. *Educational Psychology Review, 8*, 219–238. https://doi.org/10.1007/BF01464074

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research, 31*, 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*, 372–382. doi:10.1016/j.learninstruc.2009.03.002.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*, 133–151. https://doi.org/10.1080/00220970109600653

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*, 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*, 280–287. https://doi.org/10.1016/j.cedpsych.2012.02.004

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

Brannick, M. T., Miles, D. E., & Kisamore, J. L. (2005). Calibration between student mastery and self-efficacy. *Studies in Higher Education, 30*, 473–483. https://doi.org/10.1080/03075070500160244

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393–451). Mahwah, NJ: Erlbaum.

Cataldo, M. G., & Cornoldi, C. (1998). Self-monitoring in poor and good reading comprehenders and their use of strategy. *British Journal of Developmental Psychology, 16*, 155–165. https://doi.org/10.1111/bjdp.1998.16.issue-2

Chen, P. P., Cleary, T. J., & Lui, A. M. (2015). Examining parents' ratings of middle-school students' academic self-regulation using principal axis factoring analysis. *School Psychology Quarterly, 30*, 385–397. https://doi.org/10.1037/spq0000098

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

de Bruin, A. B., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology, 109*(3), 294–310. doi:10.1016/j.jecp.2011.02.005

de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*, 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Digiacomo, G., & Chen, P. P. (2016). Enhancing self-regulatory skills through an intervention embedded in a middle school mathematics curriculum. *Psychology in the Schools, 53*, 601–616. https://doi.org/10.1002/pits.2016.53.issue-6

Dunlosky, J., & Rawson, K. A. (2005). why does rereading improve metacomprehension accuracy? evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes, 40*, 37–55. doi:10.1207/s15326950dp4001_2.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypothesis. *Journal of Memory and Language, 52*, 551–565. https://doi.org/10.1016/j.jml.2005.01.011

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*, 277–287. https://doi.org/10.1027/1016-9040.13.4.277

Fatzer, S. T., & Roebers, C. M. (2012). Language and executive functions: The effect of articulatory suppression on executive functioning in children. *Journal of Cognition and Development, 13*, 454–472. doi:10.1080/15248372.2011.608322.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development, 79*, 339–358. https://doi.org/10.1111/j.1467-8624.2007.01129.x

Greene, J. A., & Azevedo, R. (2007). A theoretical review of winne and hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*, 334–372. https://doi.org/10.3102/003465430303953

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*, 203–209. https://doi.org/10.1080/00461520.2010.515935

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processes and cue validity as constraints on metacomprehension accuracy. *Memory and Cognition, 36*, 93–103.
https://doi.org/10.3758/MC.36.1.93

Gutierrez, A. P., & Price, A. F. (in press). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *The Journal of Experimental Education.* doi:10.1080/00220973.2016.1180278.

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education, 83*, 386–404. doi:10.1080/00220973.2014.907230.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*, 101–121. doi:10.1007/s11409-008-9021-5.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.
https://doi.org/10.1037/h0022263

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives, 2*, 172–177. doi:10.1111/j.1750-8606.2008.00061.x.

Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 734–752.

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning, 4*, 161–176.
https://doi.org/10.1007/s11409-009-9042-8

Jitendra, A. K., Kay Hoppes, M. K., & Xin, Y. P. (2000). Enhancing main idea comprehension for students with learning problems: The role of summarizing strategy and self-monitoring instruction. *The Journal of Special Education, 34*, 127–139.
https://doi.org/10.1177/002246690003400302

Kitchner, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three level model of cognitive processing. *Human Development, 4*, 222–232.
https://doi.org/10.1159/000272885

Kleinsorge, T., & Rinkenauer, G. (2012). Effects of monetary incentives on task switching. *Experimental Psychology, 59*, 216–226.
https://doi.org/10.1027/1618-3169/a000146

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2013). The effects of goal-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology, 143*, 386–386. doi:10.1037/a0031768.

Krebs, S., & Roebers, C. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology, 80*, 325–340.
https://doi.org/10.1348/000709910X485719

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 464–470.

Little, R., & Rubin, D. (1989). *Statistical analysis with missing data*. New York, NY: Wiley.

Lizarraga, M. L. S., Baquedano, M. T. S., Mangado, T. G., & Cardelle-Elewar, M. (2009). Enhancement of thinking skills: Effects of two intervention methods. *Thinking Skills and Creativity, 4*, 30–43.
https://doi.org/10.1016/j.tsc.2008.12.001

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*, 1778–1787.
https://doi.org/10.1111/cdev.2011.82.issue-6

Magliano, J. P., Little, L. D., & Graesser, A. C. (1993). The impact of comprehension instruction on the calibration of comprehension. *Reading Research and Instruction, 32*, 49–63.
https://doi.org/10.1080/19388079309558124

McCormick, C. B. (2003). Metacognition and learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 79–102). Hoboken, NY: Wiley.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30.
https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., & Magliano, J. P. (2009). Self-explanation and metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 60–81). Mahwah, NJ: Erlbaum.

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology, 12*, 623–634.

Nelson, T. O. (1992). *Metacognition: Core readings.* Needham Heights, MA: Allyn & Bacon.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed JOL effect". *Psychological Science, 2*, 267–270.
https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, 125–173). New York, NY: Academic Press.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*, 7–28.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition Learning, 1*, 159–179.
https://doi.org/10.1007/s10409-006-9595-6

Nietfeld, J. L., & Schraw, G. (2002). The role of knowledge and strategy training on monitoring accuracy. *Journal of Educational Research, 95*, 134–142.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175.

Pressley, M., & Harris, K. R. (2006). Cognitive strategy instruction: from basic research to classroom instruction. In P. Alexander, & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 265–286). San Diego, CA: Academic Press.

Roderer, T., & Roebers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. *Metacognition and Learning, 5*, 229–250.
https://doi.org/10.1007/s11409-010-9059-z

Roebers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology, 38*, 1051–1067.

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79*, 749–767. doi:10.1348/978185409X429842.

Roebers, C. M., von der Linden, N., Schneider, W., & Howie, P. (2007). Children's metamemorial judgments in an event recall task. *Journal of Experimental Child Psychology, 97*, 117–137. doi:10.1016/j.jecp.2006.12.006.

Sankaran, S. R., & Bui, T. (2001). Impact of learning strategies and motivation on performance: A study in web-based instruction. *Journal of Instructional Psychology, 28*, 191–198.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schneider, W., Knopf, M., & Sodian, B. (2009). Verbal memory development from early childhood to early adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from a 20 year longitudal study* (pp. 63–90). New York, NY: Psychology Press.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134. https://doi.org/10.1016/S0885-2014(00)00024-1

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125. https://doi.org/10.1023/A:1003044231033

Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (Vol. 19, pp. 3–16). Netherlands: Springer. https://doi.org/10.1007/978-94-017-2243-8

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). Mahwah, NJ: Erlbaum.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475. https://doi.org/10.1006/ceps.1994.1033

Schraw, G., & Graham, T. (1997). Helping gifted students develop metacognitive awareness. *Roeper Review, 20*, 4–8. https://doi.org/10.1080/02783199709553842

Schraw, G., Kuch, F., Gutierrez, A. P., & Richmond, A. S. (2014). Exploring a three-level model of calibration accuracy. *Journal of Educational Psychology, 106*, 1192–1202. doi:10.1037/a0036653.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology, 18*, 455–463. https://doi.org/10.1006/ceps.1993.1034

Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 278–298). Mahwah, NJ: Erlbaum.

Sinkavich, F. J. (1994). Metamemory, attributional style, and study strategies: Predicting classroom performance in graduate students. *Journal of Instructional Psychology, 21*, 172–182.

Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies. *The Journal of General Psychology, 132*, 347–364. https://doi.org/10.3200/GENP.132.4.347-364

Souchay, C., & Isingrini, M. (2004). Age-related differences in metacognitive control: Role of executive functioning. *Brain and Cognition, 56*, 89–99. https://doi.org/10.1016/j.bandc.2004.06.002

Sperling, R. A., Howard, R. S., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research & Evaluation, 10*, 117–139. https://doi.org/10.1076/edre.10.2.117.27905

Tabachnick, B. G., & Fidell, L. S. (2013). Cleaning up your act: Screening data prior to analysis. In B. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (6th ed., pp. 676–780). New York, NY: Pearson.

Thiede, K. W., Griffin, D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). Mahwah, NJ: Erlbaum.

Tuckman, B. W. (1996). The Relative Effectiveness of Incentive Motivation and Prescribed Learning Strategy in Improving College Students' Course Performance. *The Journal of Experimental Education, 64*, 197–210. doi:10.1080/00220973.1996.9943803.

Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills at the onset of metacognitive skill development. *Instructional Science, 33*, 193–211. https://doi.org/10.1007/s11251-004-2274-8

Volet, S. E. (1991). Modelling and coaching of relevant metacognitive strategies for enhancing university students' learning. *Learning and Instruction, 1*, 319–336. https://doi.org/10.1016/0959-4752(91)90012-W

Vukman, K. (2005). Developmental differences in metacognition and their connections with cognitive development in adulthood. *Journal of Adult Development, 12*, 211–221. https://doi.org/10.1007/s10804-005-7089-6

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York, NY: Routledge/Taylor & Francis Group.

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice Hall.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7