THE JOURNAL OF EXPERIMENTAL EDUCATION, *83*(3), 386–404, 2015

ISSN: 0022-0973 print/1940-0683 online
DOI: 10.1080/00220973.2014.907230

Effects of Strategy Training and Incentives on Students' Performance, Confidence, and Calibration

Antonio P. Gutierrez
Georgia Southern University

Gregory Schraw
University of Nevada, Las Vegas

This study examined the effect of strategy instruction and incentives on performance, confidence, and calibration accuracy. Individuals (N = 107) in randomly assigned treatment groups received a multicomponent strategy instruction intervention, financial incentives for high performance, or both. The authors predicted that incentives would improve performance, while strategy instruction would improve performance, confidence, and calibration accuracy as a result of better monitoring and self-regulation of learning. The authors compared pre- and posttest items and 20 new posttest-only items. They found significant effects for strategy training on performance, confidence, and calibration accuracy, as well as the interaction between strategy training and time on calibration accuracy. Incentives improved performance and calibration accuracy, either directly, or through an interaction with strategy training. Implications for future research are discussed.

Keywords *calibration accuracy, experiment, extrinsic incentives, metacognition, strategy training*

CALIBRATION ACCURACY refers to the discrepancy between a judgment of learning and learning itself. Research suggests that instructional interventions increase accuracy and improve learning (Huff & Nietfeld, 2009; K. W. Thiede, Griffin, Wiley, & Redford, 2009). These interventions are important because strategy instruction produces a durable set of self-regulation skills that can be used in a variety of settings by high school or college students who may otherwise experience a variety of learning difficulties (Greene & Azevedo, 2010; Winne & Nesbit, 2009). The present study investigated whether a multicomponent strategy intervention improved learning and calibration accuracy compared with an external monetary incentive. The strategy intervention was designed to be applicable in a wide variety of instructional settings with older learners. We believe this intervention is more comprehensive in scope than previous studies in the metacognition literature (e.g., Nietfeld & Schraw, 2002; Schraw, 1998; Volet, 1991) because

Address correspondence to Antonio P. Gutierrez, Department of Curriculum, Foundations, and Reading, Georgia Southern University, P.O. Box 8144, Statesboro, GA 30460-8144, USA. E-mail: agutierrez@georgiasouthern.edu

it was targeted to improve the monitoring and self-regulation functions of metacognition and was expected to significantly improve learning and calibration accuracy.

Metacognitive Monitoring

A number of theories address the role of self-monitoring in learning (Boekaerts & Rozendaal, 2010; de Bruin & van Gog, 2012; Efklides, 2008; Winne & Nesbit, 2009). However, Nelson and Narens (1990) proposed a historically important two-process model of metacognition in which monitoring processes enhance self-regulatory control that improve learning. This model has been used to develop a variety of active-processing intervention studies that focus on improving monitoring, which, in turn, improves self-regulation of learning (see K. W. Thiede et al., 2009, for a review).

Previous research indicates that calibration accuracy is related positively to prior knowledge (Tobias & Everson, 2009) and student achievement (Barnett & Hixon, 1997; Bol & Hacker, 2001; Grimes, 2002; Kruger & Dunning, 1999; Winne & Jamieson-Noel, 2002). In addition, accuracy improves when additional study time is provided to the learner (H. L. Thiede & Leboe, 2009), judgments of comprehension are delayed (Shiu & Chen, 2012), individuals receive feedback (Brannick, Miles, & Kisamore, 2005; Glengberg & Epstein, 1985; Glenberg et al., 1987; Walczyk & Hall, 1989), when learners are provided with an incentive (Hogarth, Gibbs, McKenzie, & Marquis, 1991; Tuckman, 1996), and learners receive practice and feedback (Bol, Hacker, O'Shea & Allen, 2005; Hacker, Bol, & Bahbahani, 2008; K. W. Thiede, Redford, Wiley, & Griffin, 2012). In sum, these studies suggest that calibration accuracy is a malleable skill that improves when a variety of scaffolding techniques are used to support learning and self-regulatory activities.

The Positive Role of Strategy Instruction on Metacognition

Strategy instruction is one of the most effective ways to increase student learning and metacognition about learning (McCormick, 2003; Pressley & Harris, 2006). Many studies suggest that the use of a specific strategy during learning improves performance and metacognitive awareness of learning. Some researchers (Dunlosky & Rawson, 2005; Dunlosky, Rawson, & Middleton, 2005; Griffin, Wiley, & Thiede, 2008) have shown positive effects for rereading. In addition, two recent studies reported a positive effect for self-explanation instruction on metacognition and monitoring during teacher-led (McNamara & Magliano, 2009) and computer-assisted (Roediger & Karpicke, 2006) instruction. Research also indicates that teaching metacognitive monitoring skills enhances learning outcomes and monitoring (Huff & Nietfeld, 2009). In contrast, Hacker and colleagues (2008) found that strategy training was not as effective as incentives given that low-achieving students benefitted from incentives but not from strategy training.

It is surprising that few studies have investigated the effect of integrated strategy instruction on calibration accuracy among college undergraduates. One exception is Nietfeld and Schraw (2002), who found that students who received strategy instruction showed higher learning and more accurate calibration. This study investigated the effects of strategy training when solving probability problems. Participants assigned randomly to the training condition received a 2-hr instructional sequence with four components: (a) understanding and solving problems involving percentages, (b) the addition rule (adding together separate probabilities of mutually exclusive outcomes to find the probability that any one outcome will occur), (c) the multiplication rule (multiplying separate probabilities of independent outcomes to find the probability that the outcomes will occur together), and (d) conditional probability. Nietfeld and Schraw (2002) directed participants to review five specific strategies discussed during the training (p. 137):

1. Draw a picture.
2. Look for key words. The word *or* signifies addition; the word *and* signifies multiplication.
3. Ask yourself whether events are independent or dependent.
4. Ask yourself whether there is replacement or no replacement.
5. Compute a probability by constructing a ratio comparing the sample space and total outcome space. To do so, identify the total number of possible events and use this as the denominator. Then, identify the number of observed events and use this as the numerator.

Nietfeld and Schraw (2002) found that domain knowledge was related positively to performance and calibration accuracy in Experiment 1 ($n = 93$). In Experiment 2 ($n = 58$), the strategy training described above enhanced performance, confidence, and accuracy irrespective of aptitude immediately following the intervention but not after a 1-week delay postintervention. More recently, Nietfeld, Cao, and Osborne (2006) found that distributed monitoring training with feedback produced higher performance, confidence, and accuracy among college students, while Huff and Nietfeld (2009) found that monitoring instruction with fifth-grade students improved performance and accuracy as well. These findings uniformly supported the hypothesis that calibration accuracy was trainable and could be improved during a single training session.

Incentives

Different types of incentives affect performance and calibration accuracy in different ways (Hogarth et al., 1991; Schraw, Potenza, & Nebelsick-Gullet, 1993; Yates, 1990). 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); (b) pride, enjoyment, or both; (c) a desire to impress or outperform others (see Deci & Ryan, 1985 for a review); or (d) any combination of the aforementioned sources. In contrast, 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 affect individuals' performance on tasks, the effect may not necessarily be positive (Kleinsorge & Rinkenauer, 2012), especially when a task is easy or unappealing (Bailey & Fessler, 2011). 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 (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 examined. 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 such as judging the difficulty of the test items (Hacker et al., 2008; Schraw et al., 1993). Incentives have a tendency to improve performance for tasks that are easily understood, such as simple or routine responses that can be executed quickly and frequently (Bailey & Fessler, 2011; McCullers, 1978). However, the effect of incentives is less obvious with respect to tasks that require flexibility and creative thinking (McCullers, 1978; McGraw, 1978). Studies also found that incentives have no influence on either calibration accuracy or performance (for a review, see Hogarth et al., 1991). An exception was Tuckman (1996), who found that learners exposed to incentives outperformed those who received strategy training. Hence, the literature suggests a complex relationship between incentives, calibration accuracy, and performance, at times resulting in mixed findings (e.g., Hacker et al., 2008; Schraw et al., 1993; Tuckman, 1996).

The Present Study

We used previous research (Nietfeld & Schraw, 2002; Schraw, 1998; Volet, 1991) to develop a general strategy intervention designed to improve both calibration and learning. On this view, strategy instruction that improves calibration accuracy should improve subsequent self-regulation of learning because better monitoring is assumed to increase self-regulation and control of learning processes (Greene & Azevedo, 2010; Nelson & Narens, 1990; Winne & Nesbit, 2009). We made three predictions consistent with the assumption that process-oriented strategy instruction improves learning and calibration accuracy (Huff & Nietfeld, 2009). One prediction is that strategy instruction will increase performance, confidence, and calibration accuracy because it helps learners to monitor and self-regulate their learning. A second prediction is that performance-based incentives will increase performance because the incentive (i.e., 80% or better on the posttest) is received only if the performance-based criterion is met. A third prediction is the interaction between strategy instruction and incentives will improve performance but have no effect on confidence or accuracy because they are relevant only to performance. These predictions reflect the assumption that strategy instruction enhances self-regulation processes because it uses information processing strategies such as identifying important information, synthesizing, and reflecting on to-be-learned information (McCormick, 2003; Pressley & Harris, 2006).

To test these predictions, we developed an integrated 1-hr strategy instruction intervention using seven general strategies that are well known to improve learning and self-regulation (Greene & Azevedo, 2010). The strategies required individuals to read, review, relate, and monitor information during learning; hence, we refer to the full strategy module as R^3M. The module is based on general strategy instruction principles (Pressley & Harris, 2006) and on specific strategies used in previous calibration research (Brannick et al., 2005; Bol et al., 2005; Dunlosky et al., 2005; Hacker et al., 2008; McNamara & Magliano, 2009; Nietfeld & Schraw, 2002). The present intervention sequenced seven individually effective strategies shown in Table 1 into an integrated intervention designed to improve self-regulation processes (i.e., strategic study, monitoring and control processes) during learning. Once introduced and explained in detail, the strategies were scaffolded, demonstrated, and practiced to maximize their effectiveness. This intervention constituted a broader training regimen than previous studies such as Nietfeld and Schraw (2002), which focused on probabilistic reasoning, or Huff and Nietfeld (2009), which focused on reading skills of children. We believe the current generalized intervention constitutes an important extension to the metacognitive strategies training literature.

TABLE 1
Summary of Metacognitive Strategies and Their Relation to Calibration and Theory

Strategy	*Learning processes*	*Hypothesized influence on calibration*
Review main objectives of the text and focus on main ideas and overall meaning	Review and monitor	Enhance 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	Read and relate	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	Review, relate, and monitor	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	Relate	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	Review, relate, and monitor	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	Relate and monitor	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	Review and relate	These strategies help simplify complex topics by breaking them down to their constituent parts; this increases accuracy by decreasing errors

Note. These strategies were loosely adapted from the work of Nietfeld and Schraw (2002) and from results of a qualitative pilot study of participants.

The strategy training and incentive treatment variables were expected to affect outcome variables in different ways. We expected strategy instruction to instantiate and strengthen a set of internalized skills that were more effective than external incentives. Strategies were expected to improve self-regulatory control processes that should lead to better learning and calibration accuracy. In contrast, incentives were expected to enhance learning, but have no effect on accuracy.

METHOD

Participants

We recruited 160 undergraduates from psychology and educational psychology to participate in the experiment. Participants were 107 individuals (73 women, 34 men) who completed all three sessions. Individuals ranged from 18 to 65 years (M = 22.76, SD = 7.15). Nearly half of them (48.1%) reported enrollment in education-related majors ranging from early childhood education to secondary education. The remaining participants (51.9%) reported majors ranging from engineering to hospitality to art therapy. Participants varied with respect to undergraduate standing, with 15.6% freshmen, 31.3% sophomores, 35.0% juniors, and 18.1% seniors. Last, slightly fewer than half of participants (44.4%) reported their ethnicity as White/Caucasian, 20% reported Hispanic/Latino, 3.1% reported African American/Black, 21.9% reported Asian/Pacific Islander, 0.6% reported Native American/Alaskan Native, and 10.0% reported other/mixed. We used Little's missing completely at random chi-square statistics to verify that the missing data pattern for the cases lost to attrition was missing completely at random. A significant chi-square (i.e., $p < .05$) would suggest that the pattern of missing data was nonrandom, and perhaps biased because of systematic attrition (Little, 1988; Little & Rubin, 1989; Schaeffer & Graham, 2002). Results were not statistically significant, all p values $\geq .23$, suggesting that the loss of data was random.

Design and Materials

The study design was a 2 (strategy training: yes, no) × 2 (incentives: yes, no) × 2 (test time: pre, post) randomized true experiment. The strategy and incentive variables were manipulated between subjects, whereas the test time variable was repeated within subjects. At pretest, individuals read a passage, answered 20 multiple-choice items, and rated their confidence of performance for each test item using a 100-mm scale. At posttest, individuals reread the passage, answered 40 multiple choice items (20 new, 20 old), and rated their confidence of performance on the items on the same scale. Approximately 65% to 70% of the 20 old and new items were answered correctly based on pilot data. This was done to appropriately calibrate old and new items, and to assure that items were not too easy or too difficult overall. The Kuder Richardson 20 internal consistency reliability for the pretest was 0.60. The Kuder Richardson 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.

The materials included a 2,400-word expository chapter on behaviorism and related theories taken intact from Slavin (2009). The authors created a 40-item multiple-choice test in which each item contained four options. Items assessed factual, conceptual, and inferential knowledge based on the passage (see Appendix A). Each test item was followed by a 100-mm scale with 0% confidence on the lower end and 100% confidence at the upper end. Individuals read the story on a computer at a pace that was comfortable for them (i.e., at their own pace), and, they were instructed to read for comprehension. Students were allowed to reread and mark the passage on screen (e.g., highlighting, underlining, changing font color), but could not take notes on hardcopy paper. The performance test and confidence ratings took place after the reading phase was completed. Ratings were made in a paper booklet with one test item and rating scale per page. Students did not have access to the reading passage while completing the performance assessment and confidence judgments.

Performance was assessed by taking each participant's raw scores at pretest (i.e., 20 old items) and posttest (i.e., 20 old items, 20 new items). Confidence scores were averaged across all items to obtain a confidence score composite. To compute calibration accuracy, we converted raw scores to a proportion and subtracted this value from the composite confidence score. However, to avoid confusion and simplify interpretation, 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 calibration accuracy. Therefore, accuracy was evaluated by calculating the continuous difference score between the confidence judgment and actual performance on a scale from 0 to 100. A zero corresponded to perfect accuracy, whereas a higher nonzero score corresponded to lower calibration accuracy because the difference between confidence and performance was greater (e.g., 75 – 75 = 0 would indicate perfect calibration whereas 75 – 60 = 15 indicates miscalibration, with higher values indicating poorer calibration accuracy).

Because calibration accuracy is a component of the comprehension monitoring aspect of metacognition, a general measure of metacognitive awareness was included to ascertain whether general metacognitive awareness influenced the results (i.e., as a potential confound). We measured general metacognitive knowledge using the Metacognitive Awareness Inventory (Schraw & Dennison, 1994), which included 52 questions that assessed separate *knowledge of cognition* (declarative, procedural, and conditional knowledge) and *regulation of cognition* (planning, monitoring, debugging, information management, and evaluation) factors. Sample items included "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), answered using a 5-point Likert scale. Schraw and Dennison (1994) reported—in two separate experiments—the Metacognitive Awareness Inventory to have a stable and consistent two-factor structure. For the present study, the two Metacognitive Awareness Inventory scales demonstrated high internal consistency reliability: for the knowledge scale, Cronbach's alpha = .87; for the regulation scale, Cronbach's alpha = .92.

Procedures

The study was conducted in three separate 1-hr sessions separated by 1 week between sessions 1 and 2 and 1 week between sessions 2 and 3. All students completed the demographics form and Metacognitive Awareness Inventory first. They next read the passage and completed the 20-item multiple-choice performance assessment simultaneously in session 1. 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. All students completed the same 20 of the 40 items of the performance assessment at pretest. All data were collected prior to assigning individuals to groups.

At the end of this session, individuals were randomized into groups and given written instructions about session two. Individuals in the strategy training plus incentives group received a brief overview of the strategies that would be covered in the upcoming 1-hr training session as well as the instructions regarding incentives for posttest performance (i.e., ≥80% of the total 40 items correct). They were explicitly instructed that their payment would depend on how well they performed (i.e., better performance) 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 US$10 contingent upon meeting or exceeding the test performance criterion 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. As such, students needed to correctly respond to at least 80% of the 40 items on the performance assessment at posttest to receive the incentive. Students in the strategy training only group were given the same brief overview of the 1-hr training, whereas those in the incentives only group were furnished instructions regarding the incentives for better posttest performance. Students in the control group received no special instructions other than information about when and where to report for session 2.

Session 2 included strategy training instruction in a 1-hr 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 1 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.

Students first 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 researchers 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 a different technical text on seasonal affective disorder as well as an 18-item practice test. During this apply-practice part of each session, one of the researchers 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. Students not receiving the strategy instruction participated in an activity unrelated to the experiment—namely, watching a 1-hr film on the history of psychology.

In session 3, students again read the stimulus text and completed the performance assessment. Students in the strategy training plus incentives and the strategy training only groups received a brief 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.

Strategy Training Manipulation Perceptions of Effectiveness

A seven-item strategy training perceptions of effectiveness scale developed by the researchers was used to ascertain the effectiveness 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 B). Participants responded to the items on a 4-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). A higher mean score indicated effective training. Internal consistency reliability for this scale was high, Cronbach's alpha = .92.

RESULTS

We conducted four sets of analyses. The first evaluated the fidelity of training. The second compared knowledge and regulation of cognition scores on the Metacognitive Awareness Inventory. The third compared pretest (i.e., 20 old items) and posttest (i.e., same 20 old items) scores. This enabled us to evaluate the effects of training, incentives, and time on pre- and posttest scores. The fourth examined the 20 new items available only at posttest to examine how training and incentives affected previously unseen items. This analysis was of special interest in order to gauge the effect of the interventions on new items that had not been seen previously. We did not compare old pretest and new posttest items directly since old items were answered at pre- and posttest, whereas new items were answered only at posttest and were unaffected by improvement attributable to time (i.e., a second viewing). All p values were adjusted accordingly by analysis using the Bonferroni adjustment to obviate the inflation of familywise type I error rate. All reported effect sizes are eta square (η^2). Last, only statistically significant results are reported for the main analyses; all other results did not reach statistical significance (i.e., $p > .05$), and hence, were not reported.

Strategy Training Intervention Self-Perceptions of Training Effectiveness

The strategy training self-perceptions of training effectiveness assessed whether the training was effective. Results demonstrated that both groups (combined $M = 3.51$, $SD = 0.47$) on a 1 to 4 scale perceived the training to be effective. Although the strategy training only group reported higher ratings ($M = 3.63$, $SD = 0.37$) than the strategy training and incentives group ($M = 3.38$, $SD = 0.56$), this difference was not significant based on an independent samples t-test, $p = 0.06$.

Metacognitive Awareness Inventory Scores

We conducted a 2 (strategy training: yes, no) × 2 (incentives: yes, no) × 2 (Type of Metacognitive Awareness Inventory score: knowledge of cognition, regulation of cognition) mixed-model repeated measures analysis, using the strategy and incentive variables as between-subject factors and Metacognitive Awareness Inventory score as a repeated factor. There were no differences between any of the groups, $p > 30$. However, there was a significant repeated main effect in which knowledge scores ($M = 73.55$, $SD = 10.52$) exceed regulation scores ($M = 67.06$, $SD = 11.25$). This suggested that all participants reported slightly higher knowledge of cognition compared to regulation of cognition.

The Effect of Training and Incentives on Old Items

We conducted a 2 (strategy training: yes, no) × 2 (incentives: yes, no) × 2 (time: pretest, posttest) mixed-model repeated measures analysis, using the strategy and incentive variables as between-subject factors and time as a repeated factor. Means and standard deviations for all outcome measures are reported by group in Table 2.

TABLE 2
Descriptive Statistics of Outcome Measures, by Group for Pretest–Posttest Data

	Strategic training + incentives (n = 33)		*Strategic training (n = 27)*		*Incentives (n = 23)*		*No strategic training or incentives (n = 24)*	
Variable	M	SD	M	SD	M	SD	M	SD
MAI knowledge scale	74.61	10.08	72.09	11.10	73.34	10.92	73.67	10.26
MAI regulation scale	67.16	10.81	66.94	11.91	66.11	11.06	68.11	11.49
Pretest performance	13.96	2.54	12.96	2.40	12.95	2.32	12.25	2.06
Pretest confidence	65.54	13.98	67.55	13.93	63.71	12.58	66.85	7.56
Pretest calibration accuracy	12.09	9.47	15.84	9.12	8.34	5.98	8.65	8.41
Posttest performance	16.03	1.87	13.77	2.40	14.56	2.17	13.04	2.59
Posttest confidence	79.49	11.71	74.32	11.42	71.23	15.52	75.35	10.48
Posttest calibration accuracy	7.70	5.65	11.38	8.83	8.05	6.59	10.90	6.78

Note. MAI = Metacognitive Awareness Inventory.

Performance

The Incentives × Time interaction was statistically significant, $F(1, 103) = 7.38$, $p = .01$, $\eta^2 = 0.067$. A comparison of the four marginal means using the Newman-Kuels test showed that those who received incentives at posttest ($M = 15.42$, $SD = 2.11$) performed significantly better than did those who received the incentives at pretest ($M = 13.55$, $SD = 2.48$) and those who did not receive incentives at posttest ($M = 13.43$, $SD = 2.50$) and pretest ($M = 12.62$, $SD = 2.26$). The no incentives group at posttest and incentives group at pretest also differed significantly from the no incentives group at pretest.

In addition, the main effect for strategy training was statistically significant, $F(1, 103) = 5.77$, $p < .05$, $\eta^2 = 0.053$, with the strategy training group ($M = 14.19$, $SD = 2.28$) outperforming the no training group ($M = 13.20$, $SD = 2.11$). The incentive main effect was significant as well, $F(1, 103) = 11.27$, $p = .001$, $\eta^2 = 0.10$, with those in the extrinsic incentives condition ($M = 71.97$, $SD = 10.66$) performing better than did those in the no incentive condition ($M = 68.05$, $SD = 10.66$). The main effect for time was significant, $F(1, 103) = 48.28$, $p < .001$, $\eta^2 = 0.319$, with higher posttest performance ($M = 14.47$, $SD = 2.50$) when compared to pretest performance ($M = 13.11$, $SD = 2.41$).

Confidence

Results of the factorial mixed-model analysis of variance with confidence as the dependent variable revealed a significant main effect for time, $F(1, 103) = 54.82$, $p < .001$, $\eta^2 = 0.347$, with students at posttest ($M = 75.48$, $SD = 12.51$) showing greater confidence than at pretest ($M = 66.34$, $SD = 12.75$).

TABLE 3
Descriptive Statistics on Outcome Measures, by Group for New Posttest Data

	Strategic training + incentives (n = *33)*		*Strategic training* (n = *27)*		*Incentives* (n = *23)*		*No strategic training or incentives* (n = *24)*	
Variable	M	SD	M	SD	M	SD	M	SD
Performance	15.24	2.12	13.18	1.21	12.56	2.46	13.09	2.99
Confidence	73.93	11.75	66.22	8.20	65.73	14.73	73.93	11.02
Calibration	9.54	5.55	5.95	5.33	9.67	5.43	12.96	11.24

Calibration Accuracy

The Strategy Training × Time interaction was statistically significant, $F(1, 103) = 8.99$, $p = .01$, $\eta^2 = 0.08$. A comparison of the four marginal means using the Newman-Kuels test showed that those who received strategy training at posttest ($M = 9.34$, $SD = 7.41$) were significantly better calibrated (i.e., smaller scores in Table 3) than those in the strategy training at pretest ($M = 13.78$, $SD = 8.89$) and those who did not receive strategy training at posttest ($M = 9.50$, $SD = 6.77$) and pretest ($M = 8.50$, $SD = 7.24$).

There was a significant main effect for strategy training, $F(1, 103) = 5.08$, $p < .05$, $\eta^2 = 0.047$, with the strategy training group ($M = 11.74$, $SD = 8.37$) outperforming the no training group ($M = 8.99$, $SD = 6.97$). The main effect for the incentive was also significant, $F(1, 103) = 4.65$, $p = .05$, $\eta^2 = 0.043$, with those in the extrinsic incentives condition ($M = 9.19$, $SD = 7.21$) showing better calibration accuracy than those in the no incentive condition ($M = 11.81$, $SD = 10.39$).

Summary

These results suggest that incentives increased performance over time and that strategy training, incentives, and time increased performance overall. Confidence increased over time but was unaffected by strategy training and incentives. Calibration increased over time as a result of strategy training and the separate interaction between strategy training and incentives.

The Effect of Training and Incentives on New Items

We conducted a 2 (strategy: yes, no) × 2 (incentives: yes, no) analysis of variance using scores on the 20 new items to test the effect of the interventions on previously unseen test information. Means and standard deviations for all outcome measures are reported by group in Table 3.

Performance

The Strategy Training × Incentives interaction was statistically significant, $F(1, 103) = 8.20$, $p = .005$, $\eta^2 = 0.074$. A comparison of the four marginal means using the Newman-Kuels test showed that those who received strategy training with incentives ($M = 15.24$, $SD = 2.12$) performed better than did those who received the strategy training without incentives ($M = 13.18$, $SD = 1.30$) and better than did those in the groups with no strategy training without incentives ($M = 13.04$, $SD = 2.99$) and no strategy training with incentives ($M = 12.56$, $SD = 2.46$).

There was a significant main effect for strategy training as well, $F(1, 103) = 10.16$, $p < .005$, $\eta^2 = 0.09$, with the strategy training group ($M = 14.31$, $SD = 2.06$) outperforming the no training group ($M = 12.80$, $SD = 2.73$).

Confidence

The Strategy Training × Incentives interaction was statistically significant, $F(1, 103) = 12.34$, $p = .001$, $\eta^2 = 0.107$. A comparison of the four marginal means using the Newman-Kuels test showed that those who received strategy training with incentives ($M = 73.98$, $SD = 11.87$) had higher confidence than did those in the no strategy training with incentives group ($M = 65.13$, $SD = 14.73$) but did not differ from the strategy training without incentives ($M = 66.99$, $SD = 8.20$) and the no strategy training without incentives ($M = 73.13$, $SD = 11.02$) groups.

Calibration Accuracy

The Strategy Training × Incentives interaction was statistically significant, $F(1, 103) = 6.04$, $p = .01$, $\eta^2 = 0.055$. A comparison of the four marginal means using the Newman-Kuels test showed that those who received strategy training without incentives ($M = 5.95$, $SD = 5.33$) had significantly lower (i.e., better calibrated) scores than those in the no strategy training without incentives group ($M = 12.96$, $SD = 11.24$), but did not differ from the strategy training with incentives ($M = 9.54$, $SD = 5.55$) and the no strategy training with incentives ($M = 9.67$, $SD = 5.53$) groups.

There was a significant main effect for strategy training as well, $F(1, 103) = 6.59$, $p < 0.01$, $\eta^2 = 0.060$, with the strategy training group ($M = 7.92$, $SD = 5.70$) demonstrating better calibration accuracy than the no training group ($M = 11.35$, $SD = 8.97$).

Summary

Strategy training and incentives interacted to increase performance at posttest and the strategy training increased performance compared to no strategy training. Similar to performance, the Strategy Training × Incentives interaction increased confidence at posttest. Strategy training also improved calibration significantly. A Strategy Training × Incentives interaction also occurred in which strategy training without incentives improved calibration compared to strategy training with incentives.

DISCUSSION

This study examined the effect of a 1-hr strategy instruction intervention and a monetary incentive on performance, confidence, and calibration accuracy. Our first prediction was that strategy instruction would increase performance, confidence, and calibration accuracy as a result of better monitoring and self-regulation of learning. We developed an integrated 1-hr strategy instruction intervention using seven general strategies designed to improve self-regulation based on recent research (Brannick et al., 2005; Bol et al., 2005; Dunlosky et al., 2005; Greene & Azevedo, 2010; Hacker et al., 2008; Pressley & Harris, 2006). This intervention constituted a broader, better sequenced training regimen than previous metacognitive training studies such as Nietfeld and Schraw (2002) or Huff and Nietfeld (2009).

Strategy training had several important positive effects on outcome variables. In the comparison of pretest-posttest results, the strategy training improved performance and calibration, and the Strategy × Time interaction showed that strategies improved calibration at posttest compared to pretest. In the posttest-only analysis of new items, strategy training improved performance and calibration, and the Strategies × Incentives interaction improved performance, confidence, and calibration. These findings supported our prediction that strategy training, either alone or in conjunction with incentives, improves performance, confidence and calibration accuracy. We attribute these gains to better self-regulation (Pressley & Harris, 2006) and better monitoring through increased awareness and a better match of confidence to performance (Greene & Azevedo, 2010; Winne & Nesbitt, 2009).

These findings indicate that a 1-hr intervention positively affects subsequent performance and calibration accuracy. Our findings replicated previous strategy training and incentives research with respect to performance (e.g., Bol et al., 2005; Hacker et al., 2008; Huff & Nietfeld, 2009; Nietfeld & Schraw, 2002; Nietfeld et al., 2006; Swanson, 1990; Yates, 1990). The strategy training and incentives manipulations also were effective at improving confidence from pretest to posttest, which is consistent with previous research on confidence judgments (e.g., Koriat & Levy-Sadot, 2001; Kruger & Dunning, 1999; Maki, Foley, Kajer, Thompson, & Willert, 1990; Mitchum & Kelley, 2010; Shynkaruk & Thompson, 2006). The effect sizes (η^2) observed in our performance analyses, which ranged from 0.07 to 0.15, were in the medium to large range, and were as large or larger than those in other strategy training interventions with college students (Nietfeld & Schraw, 2002).

Our second prediction was that the monetary incentive would increase performance because it is received only if the performance criterion is met. The incentive yielded a positive main effect on performance and calibration accuracy. In addition, the incentive interacted positively with time to increase performance. These findings suggest that the incentive encouraged individuals to perform better and monitor their performance with greater accuracy after the intervention.

Our third prediction was the interaction between strategy instruction and the incentive will improve performance, but have no effect on confidence or accuracy because they are relevant only to performance. As described earlier, incentives had a positive impact on performance and calibration, while the interaction between strategies and incentives also had a positive effect on performance, confidence, and calibration accuracy of new items. One explanation of the positive effects of the incentive on confidence and accuracy is that performance incentives may incentivize individuals to use the strategies acquired in the strategy training intervention sequence (Hogarth et al., 1991), which improves performance and increases confidence and accuracy in one's performance.

Our results are consistent with the processing-oriented hypothesis supported by Huff and Nietfeld (2009), which states that strategy instruction improves both performance, monitoring and self-regulation by providing explicit techniques to process and monitor new information. The fact that performance, confidence, and calibration accuracy improved as a result of strategy instruction supported this assumption. Our findings also suggested that incentives may improve the value of strategy instruction through a Strategy Training × Incentives interaction in which incentives incentivize individuals to use the strategies that are acquired during the intervention. This conjecture was supported further by the finding that the incentive alone does not have a significant main effect on either confidence or calibration accuracy.

Of special importance, strategy training played a strong role in calibration accuracy in the pretest-posttest and posttest-only analyses. For example, in the pretest–posttest analysis, the Strategy × Time interaction had an η^2 of 0.08, which corresponds to a moderate effect size. There also were moderate η^2 values of 0.09 and 0.07 for the Strategy Training × Incentives interaction on performance, as well as moderate effects of 0.06 for strategies and the Strategy × Incentives interaction on calibration accuracy. This is consistent with previous research that has shown that strategy training has a significant positive effect on measures of calibration (Huff & Nietfeld, 2009; Nietfeld & Schraw, 2002; H. L. Thiede et al., 2009).

It was our working assumption when we undertook this experiment that strategy instruction is more beneficial than incentives for improving calibration accuracy because strategies, once learned, are durable and can be adapted to a variety of learning situations. We hypothesized that incentives would be of limited value and might have negative side effects on calibration accuracy (Schraw et al., 1993). Our findings suggested otherwise in that incentives enhanced the strategy training, perhaps by motivating individuals to use strategies. However, our experiment did not determine the exact reason that incentives interacted with strategy training to increase performance and confidence. For this reason, the role of incentives in strategy learning should be studied in more detail. Using online verbal reports may help researchers to better understand these processes. We also did not examine the effects of strategy training or incentives after an extended delay, which may affect the long term success of these interventions.

It is especially important to emphasize that only strategy training—as a stand-alone component—had a positive effect on calibration accuracy, with η^2 values in the current study as large as or larger than previous studies that reported effect sizes for strategy training (Hattie, Biggs, & Purdie, 1996). This suggested that even a short strategy intervention may have a significant impact on self-regulation processes. Although we did not examine the generality of strategy training, it is likely that strategy training with practice will be sustained over a long-term delay of 2 to 6 weeks and also transfer to other domains and activities (Pressley & Harris, 2006; Serra & Metcalfe, 2009).

Our findings raised several questions for future research. Foremost among these is the extent to which an integrated strategy training intervention is sustainable over time and transferable to new domains? A related question is whether a strategy intervention program would have comparable results on younger students in the 6–12 grade levels? Consistent with the strategy instruction literature, we believe a case can be made that strategy instruction may benefit younger students more than older students (McNamara & Magliano, 2009; Pressley & Harris, 2006). One question of special theoretical interest is how strategy instruction affects control and monitoring processes separately? A final question is whether intrinsic motivation would have a different effect on performance and calibration compared to extrinsic incentives?

Limitations

The present investigation has several limitations. The strategy training, although effective, was compact and brief. Furthermore, additional research is needed to investigate the sustainability of the strategy training examined in the present study, and whether such training could be generalized to other samples. Such studies could inform and help clarify the results presented here.

Conclusions

Our findings suggest two important conclusions. One is that both an integrated strategy intervention program and a monetary incentive for normative high performance was successful at improving students' performance. Strategy instruction has been found consistently to improve performance and calibration, whereas 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) while others have not (e.g., Hogarth et al., 1991). The present research supported the former group of studies, as incentives were effective at improving performance by demonstrating that both interventions can effectively help undergraduate students to exhibit better performance. These discrepancies could be the result of procedural differences across studies. For example, dissimilarities in choice of outcomes, type of incentive, number of trials, and type of training may have contributed to differences in findings.

A second conclusion is that only the strategy training improved calibration accuracy from pretest to posttest. This is supported by Nietfeld and colleagues (e.g., Nietfeld et al., 2006; Nietfeld & Schraw, 2002), but was inconsistent with Hacker and colleagues (2008), who found no significant differences between the training and no training groups on calibration as well as Bol and colleagues (2005), who found a detrimental effect of practice and feedback on calibration. However, it is important to note that, unlike the present study, the Hacker 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 or detrimental effects of strategy training with respect to calibration accuracy.

AUTHOR NOTES

Antonio P. Gutierrez, Ph.D., is currently an Assistant Professor of Educational 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 metacognitive strategy training on learners' calibration, confidence, and performance as well as investigating the latent dimensions of calibration to improve its measurement. **Gregory Schraw**, Ph.D., is currently a Professor of Educational Psychology at the University of Nevada, Las Vegas. His lines of research include: text comprehension; the seductive detail effect; the design and refinement of visual displays in learning; self-regulated learning; the domain-specificity or generality of metacognition; and comprehension monitoring—that is, calibration. Dr. Gutierrez, Dr. Schraw, and their colleagues are presently involved in a series of experiments related to the effects of delayed judgments on calibration and the influence of various aspects of motivation on calibration.

REFERENCES

Azevedo, R., & Witherspoon, A. M. (2009). Self-regulated use of hypermedia. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 319–339). Mahwah, NJ: Erlbaum.

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.

Barnett, J. E., & Hixon, J. E. (1997). The effects of grade level and subject on student test score predictions. *The Journal of Educational Research*, *90*, 170–174.

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

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.

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.

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

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*, 245–252.

Deci, E. L., & Ryan, R. M. (1985). Operant and attributional theories. *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.

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.

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

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 702–718.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologists*, *45*, 203–209.

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.

Grimes, P. W. (2002). The overconfident principles of economics students: An examination of metacognitive skill. *Journal of Economic Education*, *1*, 15–30.

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

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*, 99–136. doi:10.3102/00346543066002099

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.

Kleinsorge, T., & Rinkenauer, G. (2012). Effects of monetary incentives on task switching. *Experimental Psychology*, *59*, 216–226.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of cue familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 34–53.

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

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*, 129–137.

Levine, F. M., & Fasnacht, G. (1974). Token rewards may lead to token learning. *American Psychologist*, *29*, 816–820.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Little, R. J. A., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research*, *18*, 292–326.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 609–616.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*, 723–731. doi:10.1037/0022-0663.97.4.723

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

McCullers, J. C. (1978). Issues in learning and motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward* (pp. 5–18). Hillsdale, NJ: Erlbaum.

McGraw, K. O. (1978). The detrimental effects of reward on performance: A literature review and a prediction model. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward* (pp. 33–60). Hillsdale, NJ: Erlbaum.

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.

Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 699–710.

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). New York, NY: Academic Press.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy. *Metacognition and Learning*, *2*, 159–179. doi: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.

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.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Schaeffer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

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

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

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*, 460–475.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, *18*, 455–463.

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.

Shiu, L., & Chen, Q. (2013). Self and external monitoring of reading comprehension. *Journal of Educational Psychology*, *105*, 78–88. doi:10.1037/a0029378

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory and Cognition*, *34*, 619–632.

Slavin, R. E. (2009). *Educational psychology: Theory and practice (*9th ed.*)*. Boston, MA: Pearson.

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, *82*, 306–314.

Thiede, H. L., & Leboe, J. P. (2009). Illusions of competence for phonetically, orthographically, and semantically similar word pairs. *Canadian Journal of Experimental Psychology*, *63*, 294–302.

Thiede, K. W., Dunlosky, J., Griffin, D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1267–1280.

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). New York, NY: Routledge.

Thiede, K. W., Redford, J. S., Wiley, J., & Griffin, T. D. (2012). Elementary school experience with comprehension testing may influence metacomprehension accuracy among seventh and eighth graders. *Journal of Educational Psychology*, *104*, 554–564.

Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 107–127). New York, NY: Routledge.

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

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

Walczyk, J. J., & Hall, V. C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology*, *81*, 435–437.

Winne, P. H., & Jamieson-Noel, D. L. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*, 551–572.

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.

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

Appendix A

Sample 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.
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.

0%____100%
CONFIDENCE CONFIDENCE

Appendix B

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.	1	2	3	4
2. The strategies covered in the training were appropriately and thoroughly explained.	1	2	3	4
3. The strategies were sufficiently scaffolded and modeled for me to understand how, when, and in which situations to apply them.	1	2	3	4
4. I had sufficient opportunity to practice and apply each strategy.	1	2	3	4
5. The trainer demonstrated the utility value of each strategy with respect to calibration of performance.	1	2	3	4
6. Overall, I feel that the strategy training has adequately prepared me to increase the accuracy and confidence of my calibration of performance judgments.	1	2	3	4
7. The strategy training intervention was useful in enhancing my performance confidence judgments.	1	2	3	4