Psychological Test and Assessment Modeling, Volume 53, 2011 (1), 141-160

Enhancing self-reflection and mathematics achievement of at-risk urban technical college students

Barry J. Zimmerman[1], Adam Moylan[2], John Hudesman[3], Niesha White[3], & Bert Flugman[3]

Abstract

A classroom-based intervention study sought to help struggling learners respond to their academic grades in math as sources of self-regulated learning (SRL) rather than as indices of personal limitation. Technical college students (N = 496) in developmental (remedial) math or introductory college-level math courses were randomly assigned to receive SRL instruction or conventional instruction (control) in their respective courses. SRL instruction was hypothesized to improve students' math achievement by showing them how to self-reflect (i.e., self-assess and adapt to academic quiz outcomes) more effectively. The results indicated that students receiving self-reflection training outperformed students in the control group on instructor-developed examinations and were better calibrated in their task-specific self-efficacy beliefs before solving problems and in their self-evaluative judgments after solving problems. Self-reflection training also increased students' pass-rate on a national gateway examination in mathematics by 25% in comparison to that of control students.

Key words: self-regulation; self-reflection; math instruction

---

[1] *Correspondence concerning this article should be addressed to:* Barry Zimmerman, PhD, Graduate Center of the City University of New York and Center for Advanced Study in Education, 365 Fifth Avenue, New York, NY 10016, USA; email: bzimmerman@gc.cuny.edu

[2] Now affiliated with the University of California, San Francisco, School of Medicine

[3] Graduate Center of the City University of New York and Center for Advanced Study in Education

Across America, faculty and policy makers at two-year and technical colleges have been deeply troubled by the low academic achievement and high attrition rate of at-risk students. According to the *Chronicle of Higher Education* (Evelyn, 2004), only 25% of the students entering a two-year college can be expected to graduate after six years. Ting (1998) concluded that a major reason for this high rate of attrition is the low academic proficiency of the following at-risk populations: minority students, special admission program students, students with poor high school preparation, students in lower socio-economic brackets, and students who are commuters. Low entry levels in math proficiency represent a major obstacle to students' academic success in technical programs in particular and two-year programs in general. Most interventions designed to address the needs of at-risk students in developmental subjects at the college level have focused on teaching academic content (e.g., math or science), along with a variety of academic/study skills (e.g., note-taking and test-taking). However, reviews of intervention studies have revealed that such programs do not help many students to attain their academic goals (Hattie, Biggs, & Purdie, 1996; Simpson, Hyned, Nist, & Burrell, 1997). Often, at-risk students continue to use maladaptive learning methods because their effects are not understood or are hard to discern.

To address this problem, a number of educational researchers have advocated performance-specific, process-related feedback during learning (Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Traditional academic feedback (i.e. grades) provides little adaptive help to at-risk students and often leads instead to counterproductive, defensive reactions. Frequently, students cannot identify why they made errors or how to correct their methods of learning. A key task-specific, process-related form of learning involves the use of self-regulatory processes, such as goal setting, strategy use, and self-monitoring (Schunk, 1990). In addition to instilling these and other self-regulatory processes, researchers have emphasized administering feedback designed to enhance strategic adjustments in learning activities (Butler & Winne, 1995). The present research evaluated the effectiveness of a SRL instructional program with at-risk math students attending an urban technical college.

Our research was guided by a social cognitive model of academic self-regulation comprised of three cyclical phases (Zimmerman, 2000). The forethought phase precedes learning or performance, and involves task analysis processes and sources of self-motivation. The performance control phase involves self-control processes and metacognitive monitoring. Self-reflection phase processes involve students' responses to performance outcomes, and include self-evaluative judgments and adaptive self-reactions. The model is cyclical in that self-reflective processes feed forward to the forethought phase of subsequent efforts to reach one's learning goals.

Self-efficacy beliefs are a key source of motivation during the forethought phase of the cyclical SRL model. These beliefs refer to one's capabilities to organize and implement actions necessary to attain designated performance of skill (Bandura, 1997). Research shows that self-efficacy beliefs directly predict academic performance (Pajares, 1996; Zimmerman, 2002a). Low self-efficacy beliefs can prompt students to avoid opportunities to learn (Pintrich, 2000; Zimmerman, 2002b). *Self-evaluative* judgments are a key self-reflective phase process in the cyclical SRL model. These self-judgments refer to comparing one's performance to a standard (Schunk, 1996; Zimmerman, 2002a). Students who engage in frequent self-evaluation tend to attain higher academic outcomes than those who do not self-evaluate (Kitsantas, Reiser, & Doster, 2004; Schunk, 1996; Schunk & Ertmer, 1999; 2000).

Recent research has revealed that overly confident self-beliefs can hinder the adaptive use of feedback (Schunk & Pajares, 2004). The accuracy or calibration of self-efficacy beliefs and self-evaluative judgments, in addition to their motivational strength, is a particular problem among learners who struggle academically. Struggling students often report more inflated self-appraisals than successful students, although the cause of this finding remains in doubt (Bol & Hacker, 2001; Campillo, Zimmerman, & Hudesman, 1999; Chen & Zimmerman, 2007; Klassen, 2002). A reliable methodology for assessing the calibration of self-efficacy and self-evaluation was developed by Pajares and Kranzler (1995), and there is evidence (Ramdass & Zimmerman, 2008) that calibration of students is enhanced by more effective instruction.

Although there is a growing body of research investigating the impact of self-regulatory interventions (Dignath & Buettner, 2008; Dignath, Buettner, & Langfeldt, 2008), few studies have been true experiments (involving random assignment of students to experimental and control groups) in authentic contexts. In the present study, these authentic features included: instruction by regular classroom instructors, learning by at-risk populations of students, using regular curricular tasks in math, and receiving official academic grades as outcomes. These features of authentic contexts represent significant challenges to the effective implementation of an experimental intervention. For example, random assignment required departmental permission to assign students to experimental and control conditions. The training of actual educators involved not only initial descriptions and discussions of the program, but also periodic monitoring and weekly follow-up to ensure fidelity to intervention standards. In addition, experimental and control instructors not only needed to agree on a common textbook, but they also had to work out the details about the test content and methods of grading.

The present study investigated the effectiveness of a semester-long classroom intervention designed to enhance the self-regulatory processes of at-risk undergraduate students in mathematics. More specifically, we sought to enhance students' self-reflection responses to academic feedback through: (a) instructor modeling of error correction, (b) guided self-reflection opportunities as part of a formative assessment process, and (c) an incentive system that rewards subsequent attempts at learning. These curricular components were designed to help students to self-reflect more effectively on their errors in mathematical learning and to change their perceptions of academic feedback. Instead of viewing academic feedback as an end-point of learning, the students were taught to view it as opportunities for further learning. In conclusion, we hypothesized that students trained to use SRL processes would outperform control students on instructor-developed mathematics tests as well as would display greater calibration in their self-efficacy beliefs and self-evaluative judgments than control students.

Method

Participants and design

The study was conducted at an urban public technological college that offers both associate and baccalaureate degree programs. This college has an enrollment of over 15,000 students, of which almost 90% of these students are from diverse minority groups. More than half of the students were born outside of the United States and speak a language other than English at home. Fifty-eight percent of the students work at least 20 hours a week. The income of these students averaged at the 29^{th} percentile of a national sample of two-year college students. The first year attrition rate of first time, full time, associate degree freshmen averaged 40% over the last eight years, and the graduation rate for Associate degree students averaged only 21% after six years.

This randomized controlled study involved six developmental mathematics courses and twelve introductory college-level mathematics courses. Six-hundred forty-seven students were randomly assigned to either an experimental classroom (receiving the self-regulatory intervention) or a control classroom (receiving conventional instruction) for a 15-week semester. During the semester, 151 students withdrew course enrollment and were not included in statistical analyses of math course outcomes, such as periodic and final math exams, self-regulation measures, and self-reflection measures. The latter measures were analyzed for 496 students who completed the course (260 males and 236 females). Students' class enrollment is presented in Table 1 based on their mathematics course, experimental group, class instructor, and gender. Four instructors (1 male and 3 females) taught one or more of the nine SRL classes, and nine other instructors (5 males and 4 females) each taught a control class. All of the instructors participating in the study had more than two years' experience teaching their respective mathematics courses.

Intervention procedures

SRL teacher training. The training of SRL instructors involved initial meetings before the beginning of the semester and weekly follow-up meetings to review implementation by instructors. The pre-semester training led by the investigators included information about academic self-regulation, strategic modeling, and a formative assessment process that coupled frequent feedback on problem solving quizzes with grade-point incentives for in-depth correction of errors on *self-reflection forms*, which are described below.

Errors as learning opportunities. The classroom intervention included the instructors' use of coping modeling techniques to teach students about error detection and strategy adaptation in solving mathematical problems. The experimental instructors emphasized the importance of using feedback regarding errors to make changes in approaches to particular math topics or tasks as well as to make adjustments in their general math learning methods. In addition, classroom activities included individual and group practice in detecting and correcting errors in math problems that were solved incorrectly. Students practiced describing the implementation of math strategies or procedures.

Table 1:
Class Enrollment based on Mathematics Course, Experimental Group, Classes, and Student Gender

Math Course	Experimental Group	Class	Student Gender	Student frequency
Developmental	Control	A	Female	17
		A	Male	8
		B	Female	8
		B	Male	9
		C	Female	12
		C	Male	12
	SRL	D	Female	12
		D	Male	10
		E	Female	15
		E	Male	13
		F	Female	13
		F	Male	11
Introductory	Control	G	Female	12
		G	Male	17
		H	Female	9
		H	Male	22
		I	Female	8
		I	Male	21
		J	Female	11
		J	Male	14
		K	Female	14
		K	Male	16
		L	Female	17
		L	Male	15
	SRL	M	Female	14
		M	Male	24
		N	Female	12
		N	Male	18
		O	Female	9
		O	Male	15
		P	Female	20
		P	Male	11
		Q	Female	22
		Q	Male	10
		R	Female	11
		R	Male	14

Every two to three class sessions, students in the SRL classrooms were administered a 15-20 minute quiz involving four or five math problems as a vehicle for frequent feedback to students and instructors. Quizzes required students to make task-specific self-efficacy judgments before solving individual problems and self-evaluative judgments after attempting to solve each math problem. These self-efficacy and self-evaluation measures are described in the measures section. Quizzes were scored and returned to students either within the same class period or at the start of the following class meeting.

Self-reflection processes. After students in the experimental instruction group received graded quizzes from the instructor, they had an opportunity to correct quiz errors by completing self-reflection forms (see Appendix A). Typically, these forms were submitted to the instructor during the next class meeting, but instructors periodically extended the time allotted for students to submit completed self-reflection forms. The self-reflection form was designed to guide students' self-reflection processes regarding erroneous answers to items on a mathematics quiz, and their subsequent cyclical self-regulatory efforts to solve transfer problems. For example, the self-reflection form required students to compare their self-efficacy and self-evaluative judgments with their outcome on the quiz item, explain their ineffectual strategies, establish new effective strategies, and indicate their confidence for solving another problem. On the form, students also solved a similar problem and specified the strategies and procedures involved in their work. If the student failed to solve the problem, they were encouraged to seek assistance from the instructor or a peer. If the student solved the problem, he or she would receive as an incentive quiz points that had been lost during the quiz. To help students understand and complete the self-reflection forms, instructors initially modeled the completion of the forms and supported them with in-class group and individual practice with completing the forms.

Fidelity. During the semester, the investigators observed all classrooms using a checklist that focused on strategy instruction and error analysis. The observation results from SRL classrooms were used primarily in weekly support meetings with the SRL instructors as a formative measure. Three measures were ultimately selected: (a) the instructor makes deliberate errors during presentation, (b) the instructor encourages students to go to the board to demonstrate problem solving and error detection strategies, and (c) the instructor encourages students to verbalize error detection and problem solving strategies while working through practice problems. These same three instructional practices were observed in control classes to provide a base rate for comparison with SRL classes. The observers were trained to a criterion of agreement of at least 80%. After this training was completed, observers rated the frequency of these three instructional practices during a one-hour session on a 5-point scale, from (1) never to (5) very often. The SRL instruction group instructors displayed a significantly higher frequency of the targeted behaviors ($M = 7.71$) than the control instruction group ($M = 5.01$), $F(1, 16) = 9.36$, $p < .01$, $d = 1.18$.

The research design. The math exams and self-belief measures were given at four points during the semester. Periodic math test 1 was administered after approximately a month of class attendance, and periodic math test 2 was given after roughly the second month of class attendance. Periodic test 3 was scheduled after approximately the third month of class attendance, and final exam in math was given after roughly the fourth month of class attendance. The only measure that preceded the first periodic test was a standardized entrance exam in math (i.e., COMPASS), the results of which were retrieved from their school records. Self-efficacy and self-evaluation measures were not administered at the outset of the semester, but instead were administered as part of the periodic math testing. Self-reflection form measures that were completed by students in the SRL group were turned in throughout the semester. These measures will be described in detail in the next section.

Performance measures

Math placement test. Students' need for mathematics remediation in the college is based on the sum of the arithmetic and algebra subtest scores from the Computer Adaptive Placement Assessment and Support System (COMPASS) Mathematics Test (ACT). Scores from this multiple-choice, adaptive test were used as measures of prior mathematics performance. Students in the developmental course who received a passing course grade are allowed to retake the COMPASS examination, which is a gateway for taking higher-level mathematics as well as science, technology, and engineering courses.

Math periodic examinations. Each pair of treatment and control instructors developed three cumulative mathematics problem-solving examinations and administered them during the semester. An outside grader scored each treatment-control pair periodic test.

Math final examinations. Each mathematics course has a comprehensive departmental final examination. The developmental mathematics examination consists of approximately 25 questions. Students must show all of their computation work. The passing score is 70%, and the students can earn partial credit if their grade is at least 65%. Students must pass this final exam in order to retake the COMPASS test. In the introductory level mathematics course, the final examination consists of approximately 20 problems, and all computation work must be shown. The passing score is 60%, and partial credit is given. The test accounts for 30% of the students' course grade. No self-regulation measures were administered during math final examinations.

Self-regulation measures

Math exam self-efficacy scale. Self-efficacy was assessed by asking students to rate their confidence before solving each mathematics problems on the periodic examination (Bandura & Schunk, 1981). This scale represents a task-specific assessment of how students perceive their mathematical capability. The scale items varied in math content between the developmental and introductory mathematics courses. For both courses, the math self-efficacy items were worded as follows: "How confident are you about solving this math problem correctly?" Students indicated their response on a 5-point scale: (1) definitely not confident, (2) not confident, (3) undecided, (4) confident, and (5) very confident.

This self-efficacy scale was listed at the top of the periodic exam form, and each math problem was presented below in a format that required students to render a numerical rating *before* attempting to solve each problem. Although the periodic exams varied from 9 to 20 problems in developmental math classes and from 4 and 16 problems in introductory math classes, the exams were identical in form and number of math problems for paired SRL and control classes. To correct for differences in the number of math problems on periodic exams, we averaged individual self-efficacy scores for each student to produce an overall self-efficacy measure. To assess the reliability of this self-efficacy scale, Cronbach's alpha reliability was computed separately for each pair of classrooms (SRL and control). The obtained alphas averaged .95 for the developmental course and .80 for the introductory math course .80. In prior research (Chen, 2006), the alpha reliability of a math task-specific self-efficacy scale was .83.

Math exam self-evaluation scale. Students' self-evaluative judgments were assessed by asking them to rate their confidence about their solution of each problem on the periodic math examinations. This self-evaluation scale was embedded in the same periodic exam form that was used to assess students' self-efficacy beliefs except that the self-evaluation scale was positioned to be answered *after* completing each problem on the exam. The self-evaluation question was worded as follows: "How confident are you that you solved this math problem correctly?" As we noted with regard to self-efficacy beliefs, the periodic exams were identical in form and number of math problems for paired SRL and control classes, but the exams for other classes differed in the number of problems that were presented and self-evaluated. To correct for differences in the number of math problems on periodic exams, we averaged individual self-evaluation scores for each student. To compute the reliability of this self-evaluation scale, Cronbach's alpha reliability was computed separately for each pair of classrooms (SRL and control). The obtained alphas for the developmental course averaged .95 and for the introductory math course averaged .89.

Self-efficacy bias. The bias calibration or direction of error between students' self-efficacy beliefs and their actual performance was assessed by subtracting '5' from their self-efficacy rating if the problem was solved correctly, or by subtracting '1' if there was an error (Pajares & Graham, 1999; Schraw, Potenza, & Nebelsick-Gullet, 1993). For example, if students were "confident" about their answer (a rating of 4) to a math problem that they answered correctly, their bias score would be -1 or a slight underestimation of self-efficacy. The scale ranged from -4 to +4 for each math problem. Thus, self-efficacy bias was assessed for each math problem and then averaged across other math problems.

Self-evaluation bias. Bias calibration of post-performance self-evaluative judgments was assessed similarly to self-efficacy bias except that self-evaluation judgments were compared with math item accuracy.

Academic benchmarks

Passage of two key academic benchmarks was assessed: passage of the course and COMPASS posttest exams for developmental math students. For the introductory math students, the key benchmark was passage of the course. These data were coded dichotomously for a student's passage or not of each benchmark criterion.

Results

Preliminary analyses

The effectiveness of the random assignment procedure in equalizing the two instruction groups in terms of students' prior scores from the math COMPASS placement test was assessed using a single factor ANOVAs. For developmental math, there were no significant differences between SRL instructional and control groups on either the arithmetic or algebraic subtests. There also were no significant differences between SRL instruction and control groups in the introductory math course for either the arithmetic or algebra subtest. Clearly, the SRL instruction and control groups were comparable statistically in math skill before participating in the study.

Main math analyses

Because we sought to determine the effectiveness of the SRL intervention at two different levels of mathematical competence, we analyzed the results for the developmental math course and the introductory math course separately. The students in the two courses had different entrance requirements, different math content, and different math difficulty. Table 2 presents both descriptive and inferential statistics for the primary achievement and SRL measures. Zero-order correlation coefficients for these measures are presented in Table 3.

A two group (SRL instruction, control) multivariate analysis of covariance (MANCOVA) was conducted with the developmental math students, using the pretest COMPASS scores as a covariate and the three periodic math exams as the dependent measures. This analysis showed that the SRL instructional group outperformed the control group on the three periodic math examinations, $F(1,135) = 3.80$, $p > .02.$, $\eta^2 = .09$. Separate follow-up univariate *F-tests* for each dependent measure revealed no significant difference between the SRL and control groups on the first periodic test, but a significant difference was found on the second and third periodic tests favoring the SRL group. A univariate *F-test* of the standardized math final examination in developmental math course revealed that the SRL experimental group significantly surpassed the control group.

The same MANCOVA model was conducted with students enrolled in the introductory math course, and the results were significant, $F(3/210) = 7.66$, $p > .01$. Follow-up univariate *F-tests* for each dependent measure revealed significant differences between the SRL and control groups for each periodic math test. A univariate *F-test* of the standardized math final examination scores revealed that the SRL experimental group signifi-

Table 2:
Descriptive Statistics, Univariate F-tests, and Effect Sizes for Dependent Measures by Instruction Group and Mathematics Course

		Control			SRL Instruction					
Course	**Measure**	***n***	***M***	***SD***	***n***	***M***	***SD***	**F**	**p**	**d**
Developmental	Periodic math exam 1	65	72.93	15.31	65	71.14	21.06	.311	.58	-.10
	Periodic math exam 2	60	62.10	17.96	.69	70.64	24.50	4.96	.03	.39
	Periodic math exam 3	61	54.05	23.64	67	68.06	25.65	10.26	.03	.55
	Final math exam	65	58.03	27.69	72	73.18	28.39	9.96	.02	0.50
	Self-efficacy	65	3.53	.76	71	3.55	.89	.02	.89	0.02
	Self-evaluation	66	3.68	.76	71	3.59	.83	.45	.51	-0.11
	Self-efficacy bias	65	.62	.69	71	.18	.95	9.49	.03	-0.53
	Self-evaluation bias	66	.81	.78	71	.22	.92	16.08	.01	-0.69
Introductory	Periodic math exam 1	165	58.82	23.47	174	67.86	22.41	13.16	.01	0.39
	Periodic math exam 2	165	51.92	23.73	170	57.85	23.14	7.33	.01	.29
	Periodic math exam 3	153	43.32	28,10	169	55.88	27.70	16.29	.01	.44
	Final exam	158	47.46	24.05	167	59.04	25.15	17.40	.01	0.47
	Self-efficacy	164	3.33	.81	178	3.23	.91	1.25	.27	-0.12
	Self-evaluation	167	3.47	.80	179	3.31	.91	2.83	.09	-0.19
	Self-efficacy bias	164	.70	.92	178	.31	.98	13.71	.01	-0.41
	Self-evaluation bias	167	.77	.90	179	.37	.96	16.64	.01	-0.43

cantly surpassed the control group in introductory math course. The zero-order correlation between COMPASS math pretest and first periodic math exam was .30 ($p < .01$) for the developmental math course and .26 ($p < .01$) for the introductory math course.

Self-efficacy and self-evaluation analyses

A two group (SRL instruction, control) multivariate analysis of variance was conducted on the self-efficacy beliefs of students in developmental and introductory math classes. The self-efficacy beliefs for the three periodic math exams served as dependent measures. This MANOVA revealed no significant differences between the SRL and control group students on the three periodic exam measures of self-efficacy in either developmental, $F(3,87) = 1.96$, $p = .13$, $\eta^2 = .06$ or introductory math courses $F(3,215) = .44$, $p = .73$, $\eta^2 = .01$. The same two group MANOVA was conducted on the self-evaluation beliefs of students in developmental and introductory math classes. The self-evaluation beliefs for the three periodic math exams served as dependent measures. This MANOVA revealed no significant differences between the SRL and control group students on the three measures of self-evaluation in either developmental, $F(3,89) = .49$, $p = .23$, $\eta^2 = .05$, or introductory math courses, $F(3,212) = .65$, $p = .59$, $\eta^2 = .01$.

Although the self-efficacy and self-evaluation measures were not affected by SRL training, these measures were significantly predictive of math examination outcomes (see

Table 3). For students in developmental math, their self-efficacy mean correlated with their periodic math exam mean, $r = .50$, $p < .001$, and with their final math exam score, $r = .41$, $p < .001$. Their self-evaluation judgments mean also correlated with their periodic math exam mean, $r = .49$, $p < .001$, and with their final exam score, $r = .38$, $p < .001$). For students in introductory college math, their self-efficacy mean correlated with their periodic exam mean, $r = .39$, $p < .001$ and with their final examination score, $r = .31$, $p < .001$. These introductory math students' self-evaluation mean correlated with their periodic exam means, $r = .39$, $p < .01$; and with their final examination score, $r = .34$, p < .01.

The failure of the self-efficacy and self-evaluation measures to detect SRL and control group differences on the periodic exams was unexpected, especially in view of the high reliability of and significant positive correlation between these measures of self-belief and math outcomes. One possible explanation for these results is that these low-functioning students may have over-estimated their self-efficacy and self-evaluative judgments.

Table 3:
Zero-order Correlations between Performance and SRL Measures by Course Level
(all ps < .001)

Scale	1	2	3	4	5	6
1. Periodic examinations		.79	.50	.49	-.48	-.51
2. Final examination	.58		.39	.35	-.41	-.50
3. Self-efficacy	.39	.29		.90	.40	.29
4. Self-evaluation	.39	.32	.90		.39	.41
5. Self-efficacy bias	-.47	-.26	.52	.42		.91
6. Self-evaluation bias	-.47	-.23	.44	.50	.92	

NOTE: Developmental math course correlation coefficients are above the diagonal, and introductory math course correlation coefficients are below the diagonal. The three periodic math exams and self-efficacy and self-evaluation means are averaged for these correlational analyses.

Calibration measures of self-efficacy and self-evaluation

To test for the presence of over-estimations among low functioning students, a calibration analysis was conducted. A well-established measure of directional errors in self-efficacy calibration, namely *bias,* was selected (Pajares & Graham, 1999). As we described in the method section, self-efficacy bias scores were derived from periodic math exam results and self-efficacy beliefs. As hypothesized, these self-efficacy bias scores revealed significantly less overconfidence in periodic exam self-efficacy beliefs among the SRL instruction group than in the control group in either the developmental math course, $F(1, 134) = 6.65$, p < .01, $d = -.52$ or the introductory math course, $F(1, 340) = 13.71$, p < .01, $d = -.39$.

To test for the presence of overconfidence in the students' post-performance self-evaluation judgments, a bias transformation was also performed with these scores. Self-evaluation bias scores were derived from periodic math exam results and self-evaluation beliefs. In the developmental math course, students in the SRL instruction group displayed significantly less overconfidence, $F(1, 135) = 16.08$, $p < .01$, $d = -.69$, than the control group. In introductory college math, the same pattern of results was found. That is, the SRL instruction group had significantly less overconfidence than the control group for self-evaluation bias, $F(1, 344) = 16.64$, $p < .01$, $d = -.43$.

Self-reflection form analyses

During the study, the teachers reported that students in the SRL experimental group varied considerably in their individual use of the self-reflection forms. We hypothesized that high self-reflectors would achieve more on the periodic and final exams than low self-reflectors. It was also hypothesized that high self-reflectors would be better calibrated in their self-efficacy and self-evaluation judgments. Students' use of self-reflection form was assessed by dividing the number of self-reflection forms that they completed by their number of quiz errors. This formula adjusts for differences in students' opportunities to use the form because students who made fewer errors on the quizzes would have fewer chances to self-reflect. A median split of the self-reflection rate for each math course (developmental math: Mdn = .50; introductory math: *Mdn* = .47) was used to compare performance of high self-reflectors with low self-reflectors (see Table 4).

To compare self-reflection outcomes with the developmental math students, a 2 group (high versus low self-reflectors) MANCOVA was conducted with the students' pretest COMPASS scores serving as a covariate and the three periodic math exams as dependent measures. This statistical analysis revealed that high self-reflectors outperformed low self-reflectors on the three periodic math exams, $F(3, 57) = 3.79$, $p < .02$, $\eta^2 = .17$. Separate follow-up univariate *F-tests* for each dependent measure revealed no significant differences between the high and low self-reflector groups for the first periodic test, but significant higher math scores for the high self-reflectors for second and third periodic tests. Regarding the final math exam, a follow-up univariate F-test revealed that the high self-reflectors outscored the low self-reflectors.

With introductory math students, a 2 group (high versus low self-reflectors) MANCOVA was conducted. The students' pretest COMPASS scores served as a covariate and three periodic math exams served as dependent measures. This statistical analysis revealed that high self-reflectors outperformed low self-reflectors on the three periodic math exams, $F(3, 155) = 3.81$, $p = .02$, $\eta^2 = .07$. Follow-up univariate *F-tests* for each dependent measure revealed a marginally significant difference between the high and low self-reflector groups for the first periodic test, but significantly higher math scores for the high self-reflectors on the second and third periodic tests. Regarding the standardized final math exam, a univariate *F-test* reveals that the high self-reflectors outscored the low self-reflectors significantly.

Table 4:
Descriptive Statistics, F-tests, and Effect Sizes for Dependent Measures by Self-Reflection Rate and Mathematics Course

Math course	Measure	Low self-reflection *n*	*M*	*SD*	High self-reflection *n*	*M*	*SD*	F	p	d
Developmental	Periodic math exam 1	28	69.82	20.74	37	72.14	21.54	.19	.67	.11
	Periodic math exam 2	29	61.41	27.95	40	77.32	19.41	7.80	.01	.65
	Periodic math exam 3	26	58.92	28.08	41	73.85	22.43	13.66	.01	.48
	Final exam*	29	64.76	29.23	43	78.86	26.66	4.48	.04	.50
	Self-efficacy	29	3.53	.91	42	3.55	.88	.01	.93	.02
	Self-evaluation	29	3.54	.90	42	3.62	.79	.19	.66	.10
	Self-efficacy bias	29	.40	.88	42	.03	.98	2.69	.11	.39
	Self-evaluation bias	29	.42	.88	42	.08	.93	2.33	.13	.37
Introductory	Periodic exam 1	84	64.75	23.36	90	70.77	21.21	3.17	.08	.27
	Periodic exam 2	81	51.22	23.11	89	63.89	21.57	13.66	.01	.55
	Periodic exam 3	81	48.95	27.12	88	62.26	26.81	10.28	.01	.48
	Final exam*	80	52.98	25.94	84	64.62	23.17	9.39	.003	.46
	Self-efficacy	88	3.23	.90	87	3.24	.92	.00	.97	.01
	Self-evaluation	88	3.30	.92	88	3.33	.90	.03	.85	.03
	Self-efficacy bias	88	.53	.84	87	.08	1.07	9.99	.01	.46
	Self-evaluation bias	88	.58	.89	88	.14	.96	9.92	.01	.46

NOTE: The lower *n* sizes in these analyses are due to the focus on differences in self-reflection among students in experimental SRL classes. Control group students were excluded for these analyses.

Benchmark analyses

An important educational outcome is the number of students who successfully met academic benchmarks for the two math courses. A key gateway criterion for students enrolled in to math courses is student passage of the course. This passage criterion depends not only on students' exam grades but also on other academic criteria, such as homework completion and class participation. A significantly greater percentage of students in SRL instruction classrooms (68%) passed the developmental math course than students in control classrooms (49%,), χ^2 (1) = 5.24, $p < .05$, $\phi = .19$. In addition to their passage of the course, students in the developmental math course were required to pass college-wide entrance test, the COMPASS, in order to enroll in credit bearing courses and make progress toward their degree. A significantly greater percentage of the SRL group (64%) passed the posttest COMPASS exam than in the control group (39%), χ^2 (1) = 8.13, $p < .01$, $\phi = .15$. Regarding the introduction to college math course, a significantly greater percentage of students in SRL instruction classrooms (76%) passed the course than students in control classrooms (62%,), χ^2 (1) = 7.70, $p < .01$, $\phi = .24$.

Discussion

Although meta-analyses have revealed a positive effect of self-regulatory instructional interventions (Dignath & Buettner, 2008; Dignath, Buettner, & Langfeldt, 2008), few prior studies have involved true experiments in authentic contexts, such as (a) instruction by regular classroom teachers, (b) learning by at-risk populations of students, (c) instruction using regular curricular tasks in math, and (d) feedback involving official academic grades as outcomes. Despite these challenges, SRL teachers in the present study produced significantly higher mathematics exam performance than conventional instructors, and these training effects were large (d = .40 or greater) or near large in size (Cohen, 1988). These findings represent an important addition to the literature on self-regulatory training interventions, but they also raise important questions.

Were the results due merely to quiz-taking?

This SRL intervention involved a number of components that were designed specifically to help at-risk college students in math. For example, instructors introduced frequent math quizzes, and they trained students to self-reflect on and correct their errors on these quizzes. Evidence of the effectiveness of self-reflection training with math problem solving tasks has been limited to date, but evidence of a “testing effect” is quite extensive in the prose learning literature (Roediger & Karpicke, 2006). The present research design did not regulate quiz-taking by students in the control group, but the self-reflection form, a within-group analysis, did yield evidence that math learning was related significantly to students’ self-reflection activities. To assess the impact of this self-reflection training component, we compared SRL students who were above the median in amount of quiz error corrections with classmates who were below the median. The higher self-reflection group significantly outperformed the lower self-reflection group on the periodic and final exams in both developmental and introductory math courses. Clearly, students’ engagement in self-reflective learning experiences was associated with higher levels of math skill.

Were there trends in students’ acquisition of math skill during the semester?

The multivariate test of students’ periodic math exam scores showed a pattern of greater effectiveness during the semester by students in the SRL instructional group. Differences in learning between the SRL and control groups in developmental math did not attain significance on the first periodic exam, but did so during the second and third periodic exams. By contrast, SRL students in introductory math course showed significant differences on the first periodic exam as well as on the latter two periodic exams. Apparently, students who had greater background math knowledge were better prepared to profit from self-regulatory training during the first periodic math exam.

What might be a source of the students' math learning trend across the periodic exams? A similar learning trend to that of the developmental math students was evident in the SRL students' use of self-reflection forms. More specifically, the high self-reflectors did not significantly surpass the low self-reflectors on the first periodic exam but did so on the second and third exams. A parallel pattern was observed with students in the introductory math course. The difference between the high and low self-reflectors narrowly missed significance ($p < .08$), during the first periodic exam, but it did attain significance on the second and third periodic exams. Together these results indicate a key aspect of students' self-regulatory training, their use of the self-reflection form was not evident on the first periodic exam. However, by the second periodic exam, the effectiveness of self-reflection training had reached statistical significance. These efforts to track self-reflection processes and assess their impact on math performance merit further study.

Why weren't the effects of SRL instruction significant on students' self-efficacy and self-evaluation beliefs?

We hypothesized initially that students in the SRL classes would display significantly higher self-efficacy and self-evaluative judgments than students in control classes, but neither of these comparisons attained statistical significance. However, both measures correlated significantly with the students' grades on the periodic and final math exams in both the developmental and introductory math courses. This pattern of findings indicates that self-efficacy and self-evaluation measures were predictive of math outcomes but were insensitive to the effects of SRL training. One possible explanation that we tested was that the students in the control classes may have overestimated their math proficiency. The calibration bias analyses provided support for this hypothesis: Students in SRL classes displayed less overestimation bias in their pre-performance self-efficacy judgments and in their post-performance self-evaluative judgments than was displayed by students in control classes. These differences in bias were evident on periodic math tests in both developmental and introductory math courses. Thus, inaccuracy in students' self-regulative judgments was associated with poorer achievement in both the developmental and introductory math courses.

Overestimation of learning by poorer achieving students has been reported in a variety of areas of academic functioning (Bol & Hacker, 2001), but there has been little research support that these self-judgments could be improved, even with prolonged metacognitive training (Hacker & Bol, 2004). This lack of evidence has led researchers, such as Maki (1998), to question whether self-prediction is a teachable skill. The present study revealed that self-regulatory training did lead to significantly greater accuracy in students' self-efficacy and self-evaluation judgments, and these two self-regulatory judgments were each correlated negatively with periodic math exam means (both *p*s $< .01$; see Table 3). Students' use of the self-reflection form was also related to more accurate self-efficacy and self-evaluation scores. However, it is unclear whether the nature of the training or the types of assessment instruments was responsible for the positive results in the present study. The self-efficacy and self-evaluation measures in this study are different from classic metacognitive measures, such as feelings of knowing, that have been used in prior research.

Still, there are other differences in the present study that contrast with prior research. Distinctive features of the present study were the extensive length of the training (one semester) and the use of a targeted training device (self-reflection form). Receiving an academic grade is an important transactional event between an instructor and a student. From the students' standpoint, feedback from traditional forms of assessment is seldom adequate to guide self-regulation of learning and is often perceived as punishing. An instructional approach that emphasizes errors as sources of learning is more likely to empower students to respond adaptively to academic feedback by making strategic learning improvements. The self-reflection analyses revealed that students who more frequently corrected their quiz errors reported significantly less bias than students who did not utilize this SRL option.

What was the impact of self-regulatory instruction on students' career paths?

The benchmark analyses revealed that the SRL intervention positively influenced the career paths of at-risk students in both developmental and introductory math courses. One benchmark involved passing the developmental or introductory math courses, and the second benchmark involved passing the COMPASS exam by developmental math students. The differences in pass-rates on these benchmark measures by self-regulated students were not only statistically significant, they were substantial in size. For example, the pass-rate for the SRL group on the COMPASS exam was 25% higher than that of the control group on this standardized national test that is widely used as a gateway to college level math courses. Thus, students who received SRL training fared significantly better than control students not only on curriculum-based tests but also on a high-stakes standardized test. However, a word of caution is warranted since 36% of the SRL group still did not meet the benchmark. Future research needs to focus on this important group of students.

What are the limitations of the present research and the need for future of research?

Because the present investigation involved a limited number of instructors in the SRL and control classes in developmental and introductory math, we were unable to assess the effects of variations in instructors with sufficient power. Furthermore, some instructors taught the course more than once, and all instructors were nested within a single school. To overcome these limitations in future research, we recommend the inclusion of a larger sample of teachers from multiple schools. A second limitation is the nature of the control group. We did not attempt to alter the instructional practices of teachers in the control group on matters such as having them give the same number of quizzes or provide academic grade incentives for revising quiz errors. Future research is needed to separately assess the effectiveness of these components.

Conclusions

The present research demonstrated the effectiveness of a SRL intervention designed to improve at-risk students' success in collegiate math courses. Deficiencies in key self-regulatory processes, such as bias in self-efficacy and self-evaluation measures, were found to undermine students' math exam performance, but a self-reflection intervention, which focused on correcting errors, was effective in reducing over-estimates of math self-efficacy and self-evaluation. Instead of viewing the reception of an academic grade as an end-point of learning, these students learned to view it as an opportunity for further learning.

Author note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H060018 to the City University of New York. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

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

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7-71.

Bol, L., & Hacker, D. (2001). The effect of practice tests on students' calibration and performance. *Journal of Experimental Education, 69*, 133-151.

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

Campillo, M., Zimmerman, B. J., & Hudesman, J. (1999). *Enhancing Academic Study Skill, Self-efficacy, and Achievement through Self-Regulatory Training*. A paper presented at the annual meeting of the American Psychology Association. Boston, MA, August.

Chen, P. (2006). Relationship between students self-assessment of their capabilities and their teachers' judgments of students' capabilities in mathematics problem-solving. *Psychological Reports, 98*, 765-778.

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *Journal of Experimental Education, 75*, 221-244.

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

Dignath, C., & Buettner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231-264.

Dignath, C., Buettner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively: A meta-analysis of self-regulation training programmes. *Educational Research Review, 3*, 101-129.

Evelyn, J. (2004). Community colleges struggle to foster 'engagement,' survey finds. *Chronicle of Higher Education, 51* (15), A37.

Hacker, D. J., & Bols, L. (2004). Metacognitive theory: Considering the social-cognitive influences. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (vol. 4, pp. 275-297). Greenwich CT: Information Age Publishing.

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.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112.

Kitsantas, A., Reiser, B., & Doster, J. (2004). Goal setting, cues, and evaluation during acquisition of procedural skills: Empowering students' learning during independent practice. *Journal of Experimental Education, 72*, 269-287.

Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly, 25*, 88-102.

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.

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

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*, 124-139.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 29*, 426-443.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner, (Eds.), *Handbook of Self-Regulation* (pp. 452-502). San Diego: Academic Press.

Ramdass, D. H., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics, 20,* 18-41.

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

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

Schunk, D. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist, 25*, 71-86

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33*, 359-382.

Schunk, D. H., & Ertmer, P. A. (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Educational Psychology, 91*, 251-260.

Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 631-649). San Diego: Academic Press.

Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (pp. 115-138). Greenwich, CT: Information Age.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189.

Simpson, M., Hyned, C., Nist, S., & Burrell, K. (1997). College academic assistance programs and practices. *Educational Psychology Review, 9*, 39-87.

Ting, S. (1998). Predicting first-year grades and academic progress of college students of first-generation and low-income families. *Journal of College Admission, 158*, 14-23.

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

Zimmerman, B. J. (2002a). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*, 64-70.

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

Appendice

Appendix A. SRL Math Revision sheet (i.e., Self-Reflection Form)

SRL Math Revision Sheet, Quiz #____ Item # ____ **Student: ______________________ Date: ________**
Instructor: ________________________

Now that you have received your corrected quiz, you have the opportunity to improve your score. Complete all sections thoroughly and thoughtfully. Use a separate revision sheet for each new problem.

PLAN IT [ ] 8pts

1 a. How much time did you spend studying for this topic area? _______
b. How many practice problems did you do in this topic area ______________________________in preparation for this quiz? (circle one) 0 – 5 / 5 – 10 / 10+
c. What did you do to prepare for this quiz? (use study strategy list to answer this question)

2. After you solved this problem, was your confidence rating too high (i.e. 4 or 5)? yes no

3. Explain what strategies or processes went wrong on the quiz problem.

PRACTICE IT [ ] 8pts

4. Now re-do the original quiz problem and write the strategy you are using on the right.

	Definitely not confident	Not confident	Undecided	Confident	Very confident
5. How confident are you now that you can correctly solve this similar item?	1	2	3	4	5

6. Now use the strategy to solve the alternative problem. [ ] 4pts

7. How confident are you now that you can correctly solve a similar problem on a quiz or test in the future? 1 2 3 4 5