DOCUMENT RESUME

ED 422 381 TM 028 943

AUTHOR	Ross, John A.; Rolheiser, Carol; Hoaboam-Gray, Anne
TITLE	Impact of Self-Evaluation Training on Mathematics Achievement in a Cooperative Learning Environment.
SPONS AGENCY	Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).; Ontario Dept. of Education, Toronto.
PUB DATE	1998-04-00
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998). Funding also provided by the Durham Region Roman Catholic Separate School Board.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Cooperative Learning; *Elementary School Students; Foreign Countries; Grade 5; Grade 6; Intermediate Grades; *Mathematics Achievement; Probability; *Self Evaluation (Individuals); Tables (Data); *Training
IDENTIFIERS	Authentic Assessment

ABSTRACT

This study examined the effects of self-evaluation on student performance in mathematics. Grade 5 and 6 students (N=300) were randomly assigned in intact classes to treatment and control groups. In both conditions a 2-week cooperative learning unit on probability was implemented. In the treatment condition, students received training in self-evaluation for 6 weeks prior to the probability unit. Treatment students became more accurate in their self-appraisals (ES=0.26 on the posttest and 0.35 on retention), an important finding since overestimates of performance reduce students' willingness to seek appropriate help. The treatment had a negligible impact on mathematics achievement, mainly because teachers preferred to teach self- evaluation skills in domains other than math (social skills and writing). The study demonstrated that self-evaluation training clarifies student understanding of curriculum expectations. The findings also weaken the argument for the consequential validity of authentic assessment practices, at least with respect to student achievement. (Contains 5 tables, 1 figure, and 60 references.) (Author/SLD)

Impact of Self-Evaluation Training on Mathematics Achievement In a Cooperative Learning Environment[1]

John A. Ross
Carol Rolheiser
Anne Hoaboam-Gray

Ontario Institute for Studies in Education
at the University of Toronto

The effects of student evaluation practices have rarely been investigated in cooperative learning classrooms. This study examined the effects of self-evaluation on student performance in mathematics. Grade 5-6 students (N=300) were randomly assigned in intact classes to treatment and control groups. In both conditions a two-week cooperative learning unit on probability was implemented. In the treatment condition only students received training in self-evaluation for 6 weeks prior to the probability unit. Treatment students became more accurate in their self-appraisals (ES=.26 on the posttest and .35 on retention), an important finding since overestimates of performance reduce students' willingness to seek appropriate help. The treatment had a negligible impact on mathematics achievement, mainly because teachers preferred to teach self-evaluation skills in domains other than math (social skills and writing). The study demonstrated that self-evaluation training clarifies student understanding of curriculum expectations. The findings also weakened the argument for the consequential validity of authentic assessment practices, at least with respect to student achievement.

Guidelines for giving students feedback on their group work abound in the cooperative learning literature (e.g., Bennett, Rolheiser-Bennett, & Stevahn, 1991), but the consequences of these strategies for students have rarely been investigated. A few studies (Johnson, Johnson, & Stanne, 1990; Johnson, Johnson, Stanne, & Garibaldi, 1990) have reported positive outcomes for student evaluation procedures but in these studies student evaluation procedures have been embedded in other instructional practices. It is difficult to disentangle the unique contribution to achievement.

The study reported here examined the effects of student self-evaluation on achievement. We focused on self-evaluation because we believed that its emphasis on student self-direction and sharing of control of instructional process were especially compatible with the philosophy of cooperative learning. In this study we focused on mathematics. Earlier we examined the effects of self-evaluation training on language skills (Ross, Rolheiser, & Hogaboam-Gray, 1998), finding that grade 4-6 students who had been taught how to evaluate their writing produced higher quality narratives than a comparable group of control students.

[1] Paper presented at the annual meeting of the American Educational Research Association in San Diego, April, 1998. This research was funded by the Social Sciences and Humanities Research Council of Council, the Ontario Ministry of Education, and the Durham Region Roman Catholic Separate School Board. The views expressed in the report do not necessarily reflect the views of the Council, the Ministry of the school district. Please send comments to John Ross, OISE/UT Trent Valley Centre, Box 719, 150 O'Carroll Ave., Peterborough, ON K9J 7A1

Theoretical Framework

Difficulty in Solving Data Management and Probability Problems

In the mid-1980s mathematics educators began to address probability concepts in the elementary and secondary curriculum (e.g., ASA-NCTM Joint Committee, 1985), linking curriculum expectations about probability to the broader domain of data management and statistical reasoning (e.g., Ontario Ministry of Education and Training, 1997). The motivation for addressing probability was based on the chronic inability of university students to master fundamental statistical concepts. It was also influenced by the movement to make mathematics relevant to everyday life. An understanding of probability is central to solving frequently occurring problems such as whether to buy an extended warranty on an appliance or take an umbrella to work. Yet researchers (e.g., Kahneman, Slovic, & Tversky, 1982) found inappropriate reasoning about probability to be widespread and resistant to instruction. Experts, including graduate mathematicians (Bramald, 1994), applied faulty heuristics when dealing with everyday situations involving probability.

Researchers attribute flawed performance to a variety of factors, including naïve conceptions. Mathematicians define probability as the relative frequency of event occurrences over a large number of trials, in contrast with the intuitive notion of probability as the strength of belief that a particular event will occur (Hacking, 1975). The lively sales of pamphlets recording numbers over-represented in winning lottery tickets testify to a prevailing misconception. For some readers these pamphlets reveal magical digits likely to reappear in lists of subsequent winners and unlucky numbers that should be shunned. Other readers find in these pamphlets numbers that have rarely come up and are therefore due to appear so that "the law of averages" prevails (the gambler's fallacy of Kahneman et al., 1982). Student misconceptions about probability get stronger with age (Garfield & Ahlgren, 1988). For example, Green (1983) suggested that students' ability to recognize randomness declines because science programs persuade students that everything can be explained.

Individuals may flip between naïve and sophisticated approaches to probability when moving from one problem to another. Konold, Pollatsek, Well, Lohmeier, and Lipson (1993) found that subjects held three different frameworks for probability estimates. Individuals used correct procedures to calculate probability, misleading heuristics identified by Kahneman et al. (1982), and a faulty "outcome approach" in which judgments are made on the basis of a single trial (Konold, 1989). Konold et al. (1993) found that some subjects simultaneously applied two or more framework s within the same probability situation without feeling any sense of contradiction.

Performance is depressed by a lack of prerequisite knowledge. The calculation of probability requires a grasp of proportional reasoning that develops late in most students (Wavering, 1984). In addition, nesting probability estimates in word problems evokes deficits in general problem solving skills. For example, students have difficulty categorizing types of probability problems when key words are absent or irrelevant information is provided (Hansen, McCann, & Myers, 1985). Translating problem content into a probability formula or moving from one representation to another is an unmet challenge for many students (Brenner et al., 1997).

Instructional Responses

The *prototypical strategy for teaching probability* reported in professional journals is to have students conduct a large number of probability experiments involving coins, spinners, dice, telephone numbers, and other sources of random variation. Especially popular are activities in which students make predictions about the distribution of items (such as colored balls) to be found in a random sample drawn from a known population, test their predictions by drawing successive samples with replacement, tally and graph the results of their tests, and translate their findings into other mathematical representations. These activities emphasize features of mathematics education reform such as the use of manipulatives, embedding problems in real life situations, integrating mathematics with science and technology, using visual displays rather than abstract procedures, encouraging student talk about solutions, and enlisting the computer as an instructional tool.

Garfield and Ahlgren (1988) reviewed a few studies that reported success with a direct instruction, practice with feedback model. The reviewers were skeptical that real changes in children's probability skills had occurred. Student gains have been reported for programs in which probability concepts were embedded within a more complex task, such as correlational reasoning (Ross & Cousins,1993a; 1993b), but in these studies the program's contribution to probabilistic thinking could not be isolated.

The practice of using the computer as a tool for generating and/or displaying data in teaching probability (e.g., Akers, Finzer, Guiterrez, & Resek, 1987; Brutlag, 1994; Jiang & Potter, 1994) has had mixed results. Johnson (1985, in an unpublished dissertation cited by Garfield and Ahlgren, 1988) found that software that displayed the approximation of population parameters by adding successive random samples helped students understand the relationship between sample and population. It may also have had the unintended effect of increasing support for the gambler's fallacy. "Students watching the dynamic presentation showed particular interest in whether successive sample means would fill gaps or balance asymmetries that remained in the accumulating distribution" (p. 49). Reliance on computers to display data may impede students' understanding of the principles that drive the production of graphs (Leinhardt, Zaslavsky, & Stein, 1990). Using computers to generate graphs improves interpretation while depressing construction skills (Adams & Schrum, 1990), conceals student misconceptions (Bohren, 1989) and complex software may impede the development of interpretative skills (Ross & Cousins, 1993a).

Previous studies have found that explaining solutions promotes understanding of mathematical concepts (Webb, 1991). But the only study to examine learning of statistical concepts found that explainers were overloaded by the dual task of learning and teaching (Renkl, 1997). In addition, Konold et al. (1993) argued that the benefits of constructive cognitive conflict may not accrue when discussing solutions to probability problems because subjects are not disturbed by contradiction.

In summary, previous research demonstrates that even sophisticated adults have difficulty solving real-life problems involving probability due to naïve conceptions that interfere with rational strategies. Students are subject to similar misconceptions and also suffer from deficits in prerequisite knowledge. Probability is an important topic in the mathematics curriculum and there is broad support for a prototypical instructional procedure. Few studies of the effects of this procedure on student performance have been reported.

Self-Evaluation Training

In this study we examined whether adding a self-evaluation component to the prototypical strategy for teaching probability would enhance student performance. Figure 1 describes a model linking self-evaluation to student achievement. The model posits that achievement is the outcome of personal goals and student effort. Student goals can be categorized at a general level as students' orientation to the task. The highest performance is obtained when a task (or mastery) orientation predominates (Meece, Blumenfeld, & Hoyle, 1988). Although goal orientations overlap with effort, they are theoretically distinct. For example, students can be observed exerting enormous effort to avoid doing any work. Student effort influences how well students achieve their goals, since persistence increases accomplishment. Effort is also influenced by students' specific goals. For example, students are more likely to persist if they adopt goals that have unambiguous outcomes, that are achievable in the near future, and that are moderately difficult to achieve (Schunk, 1981).

Figure 1 About Here

Self-evaluation is the process of comparing actual task performance (achievement) to personal goals. First, the student makes a self-judgment, determining how well goals were met. Second is a self-reaction, an interpretation of the degree of goal achievement that expresses how satisfied the student is with the result. Self-reactions are influenced by students' explanations of the causes of success and failure (Weiner et al., 1971). The key issue is whether the student attributes outcomes to internal factors (such as the amount of effort expended) or to external forces beyond student control (such as luck). Although the evidence is not entirely consistent, most studies (reviewed by Fernandes, & Fontana, 1996) find internality to be associated with higher achievement.

The effect of students' self-evaluations on subsequent goal setting and effort is moderated by their confidence in their ability to perform the actions believed to produce the desired result (self-efficacy). If the performance is satisfactory and is attributed to the student's own efforts, confidence increases (Bandura, 1986).

Students with high confidence in their ability to accomplish the target task set higher goals for themselves and are more likely to visualize success than failure (Bandura, 1993). Student expectations about future performance also influence effort. Confident students persist. They are not depressed by failure but respond to setbacks with renewed effort. Confident students also set higher goals.

Self-evaluation plays a key role in fostering an upward cycle of learning when two conditions are met. The first condition is that the child's self-evaluations be positive. Positive self-evaluations encourage students to set higher goals and commit more personal resources to learning tasks (Bandura, 1986; Schunk, 1995). Negative self-evaluations lead students to embrace goal orientations that conflict with learning, select personal goals that are unrealistic, adopt learning strategies which are ineffective, exert low effort and make excuses for performance (Stipek, Recchia, & McClintic, 1992).

The second condition is that the child's self-evaluations be accurate. Accuracy is a matter of degree. The self-evaluations of even young children correlate reasonably well with their teachers' appraisals when students are asked to make a global assessments, comparing their ability to that of their classmates (Crocker & Cheeseman, 1988). Overestimates of specific performance are likely to lead to complacency and reduced effort. For example, the child who does not recognize the need for help will not seek it (Markman, 1979). Elementary students tend to over-estimate their success on school tasks, in part because they expect that teachers will give them tasks that they can complete (Schunk, 1996), but also because young students lack the cognitive skills required to integrate information about their abilities and they are more vulnerable to wishful thinking.

Studies of the Effects of Self-Evaluation Training

Fontana and Fernandes (1994) implemented a program to increase primary student control of learning. In the early phase of the 20 week program, students selected from a range of tasks identified by the teacher, negotiated learning contracts and determined whether they had fulfilled their commitments using assessment materials provided by the teacher. By the end of the program students were setting their own learning objectives, developing appropriate mathematical tasks, selecting suitable mathematical apparatus and developing their own self-assessment procedures. The program had a significant impact on student achievement for more able students but the effects were negligible for the less able. In this study, self-evaluation was embedded in a broader instructional treatment. The distinctive contribution of self-evaluation to the effects could not be disentangled from other program components.

Schunk (1996) found that asking grade 4 students to judge how certain they were they could solve computational problems (of the type they had just been taught) influenced achievement in a performance goal condition. Students given a performance goal (i.e., the directions made no reference to *learning how to* solve fraction problems) had higher achievement if they self-evaluated on six occasions (once after each lesson). Self-evaluation had no effect on the mathematics achievement of students in a learning goal condition (i.e., they were told that the purpose of the activity was to learn how to solve fraction problems, the typical classroom condition). In this study students were given no feedback on their performance, contrary to usual classroom practice.

Ross (1995) found that self-evaluation training increased cooperative student interactions associated with achievement. Grade 7 mathematics students working in cooperative groups were given edited transcripts of their interactions and trained in how to interpret them. They used an instrument 1-2 times per week for 12 weeks to record the frequency of positive interactions. The self-assessment procedures increased the frequency of productive help giving, help seeking and attitudes about asking for help.

There is also evidence from other domains of achievement effects. Self-evaluation training improved students' writing skills (Arter, Spandel, Culham, & Pollard, 1984; Ross et al., 1998), musical performance (in Sparks, 1991 but not in Aitchison, 1995), and persistence (Henry, 1994; Hughes, Sullivan, & Mosley, 1985; Schunk, 1996).

Research on university students indicates that the accuracy of self-appraisal improves when professors and students agree on assessment criteria (e.g., Falchikov & Boud, 1989) and when students are required to justify their assessments (Boud, Churches, & Smith, 1986). There is also evidence from short duration lab studies that self-evaluation accuracy of elementary students can be improved by influencing goal conditions (Butler, 1990) and drawing attention to previous performance (Stipek, Roberts, & Sanborn, 1984). Ross et al. (1998) found that self-evaluation training had a positive impact on the ability of 8-13 year olds to assess the quality of their writing. Students who were initially accurate in judging their work were more likely than controls to continue to be accurate and students who initially over-estimated their performance were more likely than controls to become accurate.

In summary, although no previous studies have focused on probability skills, there is sufficient evidence from previous studies to suggest that teaching students how to evaluate their work might have a positive impact on students' mathematics performance and their ability to assess it accurately.

Research Questions and Predictions

Our approach to teaching students how to evaluate their work began in a study of the student assessment practices of exemplary cooperative learning teachers (Ross, Rolheiser, & Hogaboam-Gray, in press-a). We organized their strategies as a four-stage process: (i) involve students in defining evaluation criteria, (ii) teach students how to apply the criteria, (iii) give students feedback on their self-evaluations, and (iv) help students use evaluation data to develop action plans. Strategies for each stage were elaborated by a team of teachers and reported as a series of action research case studies and classroom usable tools (Rolheiser, 1996). Use of these strategies had a positive effect on student attitudes to evaluation in some but not all of the pilot test classrooms (Ross, Rolheiser, & Hogaboam-Gray, in press-b; in press-c). Our goal in this study was to determine whether teaching students how to evaluate their work would improve achievement in solving probability problems in grades 5-6. Our research questions and hypotheses were:

1. Will self-evaluation training increase the accuracy of students' self-assessments? We anticipated that students in the treatment group would evaluate their work more accurately

because all four stages in our model reduce uncertainty about the criteria for judging academic work.

2. Will self-evaluation training contribute to mathematics achievement? We anticipated that focusing student and teacher attention on performance criteria (Stages 1 and 2) would enhance achievement.

Method

Sample

Fourteen grade 5-6 classes (mean age 11 years, 9 months), in a large school district in Ontario (Canada), were randomly assigned within schools to two conditions. The treatment group received training in self-evaluation for a six-week period and then completed a two-week unit on probability. The control group did not receive the self-evaluation training but completed the same two-week probability unit. The treatment group (N=176 students) was slightly older (median age 11 years) than controls (N=174, median age 10 years). There were equal numbers of males and females in both groups (51% male).[1]

Instruments

Outcome Measures. Students completed an achievement test on three occasions: pre, post (after 8 weeks), and retention (4 weeks after the posttest). The post and retention items were probability problems measuring the Data Management and Probability expectations of the 1997 Ontario Curriculum for grade 5-6 (pp. 66-67). The pretest was a general measure of problem solving (selected from Kulm, 1994) because students had little instruction on probability prior to the study. A marker with an Ed.D in mathematics education coded the achievement items. The marker was blind to the experimental conditions of the students and to study goals.

In the pretest students designed rectangular dog pens using 24 meters of fencing. They were asked to make drawings of possible pens, select one that they would build, and explain why. The coding scheme for interpreting student responses had three levels: does not meet expectations (level 1), minimally satisfactory (level 2), and satisfactory answer demonstrating understanding of mathematics concepts (level 3). The levels were divided into high and low responses, creating a scale with six values: 1-, 1+, 2-, 2+, 3-, 3+. Each response was holistically placed in one of these levels using three criteria: reasoning or strategy for solving the problem, accuracy of concepts and computations, and communication of argument. The Appendix contains the rubric.

The posttest of achievement consisted of two items. Students were given a list of 17 movies, their starting times (9 values), and their categorization according to audience suitability (3 values). The first task (data management) was to create a bar chart showing how many movies were offered at each time of day. The second task (probability) was to find the probability of seeing a movie with a particular rating at a particular time. Student responses were coded into the same 6-point scale used in the pretest.

The retention test consisted of 16 probability items (e.g., what is the probability of getting heads when a coin is tossed?) that were scored correct or incorrect. The number correct was converted to the six point scale used in the pre test: 0-1 correct=level 1-, 2-3 correct=level 1+, 4-8 correct=level 2-, 9-12 correct=level 2+, 13-14 correct=3- and 15-16 correct (including question 2bi or 2bii, the most difficult items)=level 3+.

The inter-rater reliability based on a random sample of 60 responses for each test was acceptable: for the pretest Cohen's κ=.62 for exact agreement and .92 for within one point on the 6-point scale; for the two post-test items Cohen's κ=.57 and .93 for exact agreement and .92 and .97 for within one point on the scale; for the retention test: Cohen's κ=.96 for exact agreement and 1.0 for within one point.

Accuracy of self-evaluation was calculated from the achievement data and from student responses to survey items administered after the achievement task. In the survey students used a 1-10 scale (anchored by 1=not well and 10= very well) to rate their overall performance on the math test. They then used the same 1-10 scale to rate five dimensions of their performance. "How well you...understood the problem, made a plan, solved the problem, checked the solution, and explained the solution". These six items were averaged to create a 1-10 mean score for each student (Cronabach's alpha for the scale =.93 on each of the three administrations). The scores were transformed to a 1-6 scale (corresponding to the achievement scale).[2] Accuracy variables were created so that if the student's self-evaluation score matched that student's achievement score, the accuracy value was 6. If the student's self-evaluation was within one point on the six-point scale of that student's achievement, the accuracy value was 5, etc. Self-evaluation accuracy scores ranging from 1 to 6 were calculated for each student for the pre-, post-, and retention tests.

Tests of Sample Equivalence Three instruments (derived from the model in Figure 1) were administered on the pretest to determine sample equivalence. The goals orientation survey consisted of 16 items from Meece et al., (1988) distinguishing three orientations toward learning: mastery (e.g., "The work made me want to find out more about the topic."), ego (e.g., "I wanted others to think I was smart.") and affiliation ("I wanted to help others with the work").. The internal consistency of the scales were adequate for the 9 item mastery scale (alpha=.83) but less so for the 3 item ego (alpha=.63) and 3 item affiliation (alpha=.65) scales. Attributions for success and failure consisted of 14 items selected from Vispoel and Austin (1995). It produced four scores: internal and external attributions for success and failure. The internal consistencies were poor for each of these 3-4 item scales (alphas=.42-.57). Student self-efficacy consisted of 6 items identical to the self-evaluation measure except that each asked "how sure are you that you could..." rather than "how sure are you that you [did]" (alpha=.89).

Experimental Conditions

In the first six weeks treatment students were given direct instruction on how to evaluate their work. There were 6-30 minute lessons in which the teacher demonstrated a particular self-evaluation technique or engaged students in a discussion of their self-evaluations. For example, in one activity students cooperatively developed a rubric. The activity began with students individually solving a mathematics problem. In whole class setting, students suggested criteria for judging the quality of their performance. The teacher recorded the suggestions on the board and asked groups of four to vote on which criteria were most important. After determining the top four criteria (by summarizing the vote and combining categories), the teacher had each group describe high, medium, and low performance on one criterion. Outside of class, the teacher reworked student suggestions to construct a rubric that used student ideas and language, while reflecting expectations of the curriculum. Students then used the rubric to evaluate their work. Students worked through other activities based on the four-stage model, including 11 short practice sessions in which students completed a 3-5 minute self-evaluation using a form provided by the teacher. The activities implemented by teachers were based on suggestions in a teacher handbook (Rolheiser, 1996) and ideas developed in working sessions developed by teachers during in-service sessions (described below). Few of the self-evaluation examples provided to teachers focused specifically on mathematics. Most focused on assessing social skills and language development. Teachers had complete control of how they adapted these materials to their mathematics programs.[3] Teachers also received a handbook suggesting ways of linking assessment to mathematics instruction (Ontario Association for Mathematics Education, 1996). It provided examples of performance assessments and rubrics for them, although the handbook gave little attention to self-evaluation.

In the last two weeks teachers implemented a unit that enacted the prototypical strategy for teaching probability. Students conducted a number of experiments in which they made predictions about the probability of various results produced by random generation devices such as number cubes. Students drew samples and represented their results in a variety of ways (pictographs, bar and line graphs, tally sheets). The activities were designed to demonstrate equal likelihood of outcomes and test probability estimates by counting. Other features of the unit were the use of manipulatives, real-life examples, and student dialogue about reasons for solutions. Teachers in the treatment condition continued to assign self-evaluation tasks during the probability unit.

Teachers in the treatment condition attended 3 three-hour, after-school in-service sessions distributed over the eight weeks. The three sessions modeled classroom activities (e.g., a tangram task was used to model the development of a rubric), provided structured opportunities for teachers to share successful self-evaluation activities and identify problems, and enabled teachers to collaboratively plan self-evaluation activities for their own classrooms. During these sessions the three authors recorded teacher plans, successes and problems. In addition artifacts (primarily lesson plans) were collected. Treatment teachers also attended four brief team meetings in their schools to review progress and solve problems that arose during their enactment of the treatment.

During the eight weeks of the project control group teachers continued teaching mathematics as they usually without overt self-evaluation training. In the last two weeks control group classes worked through the probability unit without providing self-evaluation activities. Control group teachers received the probability unit in a three-hour after school workshop. They (unlike treatment teachers) were also given a half-day of release time in their schools to plan how to use the probability unit. Control teachers also received a handbook of performance tasks (Ontario Association for Mathematics Education, 1996). Immediately prior to the retention test, control group teachers received a three-hour, after-school in-service on self-evaluation, a copy of the project handbook (Rolheiser, 1996), and suggestions for teaching self-evaluation following the final data collection.

Analysis

Descriptive statistics (means, standard deviations, reliabilities) for all student and teacher variables were compiled. Prior to inferential statistics all variables were normalized using log transformations. The effects of the treatment were determined through a series of repeated measures analyses of variance (General Linear Modeling in SPSS).

Results

Table 1 summarizes the means and standard deviations of the student variables for each experimental condition on three occasions. Table 2 displays the results of t-tests comparing treatment and control groups on pretest variables. The table shows four differences and the extent to which each variable correlated with the study's outcome measures. Control group students scored significantly higher on self-evaluation. That is, they gave themselves higher ratings on the pretest achievement task, a judgment that was warranted because they scored higher on the task (although the latter differences were not statistically significant). Control group students were also more accurate than treatment students in their self-assessments of pretest performance. Control group students tended to be younger: they were more likely to 9 or 10 years old than 11; the reverse was the case for the treatment group. Control students were more likely than treatment students to attribute their success to internal causes. All of these differences between the samples were positively correlated with achievement on the post- and retention-tests. These data suggest that control group students began the project with a significant advantage over control group students. Table 2 also shows that the variables that distinguished the two student samples did not correlate with self-evaluation accuracy on either post- or retention-test.

Tables 1 & 2 About Here

General Linear Modeling (GLM) was used to determine the effect of the treatment on the outcome measures. GLM is a form of trend analysis that estimates effects over three test occasions. It first conducts multivariate tests, using less stringent assumptions, to determine the total variance in the outcome measure over time and the interaction of time with treatment. It then separates within-subject factors (variance from one test occasion to another) from between-subject factors (the effects of the treatment and covariates after the removal of within-subject effects). For the within-subjects analysis, GLM shows linear and quadratic trends. In the between-subjects tests, GLM reports for each test occasion the intercept (average) and the effect of the treatment when the within-subject variance is removed.

Effect of Treatment on Self-evaluation Accuracy

The GLM analysis for self-evaluation accuracy showed there was a multivariate effect for test occasion [$F(2, 307)=11.250$, $p<.001$] but the treatment-test occasion interaction did not reach statistical significance [$F(2, 307)=2.929$, $p<.055$]. The within-subject contrasts found significant linear [F(1,308)=16.206, p<.001] and quadratic [F(1, 308)=8.857, p<.003] trends. Student performance increased from pre to post and from post to retention. The treatment-test occasion interaction was significant for the linear trend [F(1, 308)=5.568, p<.019] but not for the quadratic [F(1, 308)=.686, p<.408]. Table 3 (to be read in conjunction with the means of Table 1) explains these results. On the pretest, treatment students were less accurate than control students. Treatment students were more likely to overestimate their performance. On the posttest, treatment students caught up to controls. There were no significant differences between the two groups. Table 3 also shows that the accuracy gains of treatment students continued in the retention test. The initial advantage of the controls did not reappear four weeks after the end of the treatment. The effect size of the treatment, using the formula of Glass, McGaw, and Smith (1981), was .26 on the post-test and .35 on the retention test.

Table 3 About Here

No further analyses of the effect of the treatment on self-evaluation accuracy were conducted. Although pretest differences were found on three of the measures used to determine sample equivalence, none of these pretest measures predicted accuracy of self-appraisal on either the post or retention test (as shown in Table 2).

Effect of Treatment on Achievement

The GLM analysis for achievement showed there was a multivariate effect for test occasion [F(2, 315)=36.904, p<.001] but the treatment-test occasion interaction was not significant [F(2, 315)=.315, p<.730]. The within-subject contrasts found significant linear [F(1,316)=72.876, p<.001] and quadratic [F(1, 316)=8.752, p<.003] trends. Student performance increased from pre to post and from post to retention. The treatment-test occasion interaction was not significant for either the linear trend [F(1, 316)=.155, p<.694] or for the quadratic trend [F(1, 316)=.579, p<.447]. The between-subject tests, displayed in Table 4, found no significant effects for experimental condition. Although the differences between the treatment and control groups narrowed, the effect size of the treatment was negligible: .08 on the post-test and .02 on the retention test.

Table 7 About Here

The GLM analysis was repeated using as covariates three variables that predicted achievement and for which there were pretest differences between the treatment and control groups. There were also pretest differences on accuracy of self-evaluation and this variable correlated with achievement. Self-evaluation accuracy was not included as a covariate of achievement because achievement was one of the factors used to calculate accuracy.

In the first run, the dependent variable was achievement, the independent variable was experimental condition, and the covariate was student age. There was a significant multivariate effect for test occasion [F(2,309)=12.240, p<.001] and for the test occasion X age interaction [[F(2,309)=8.775, p<.001]. The test occasion X treatment interaction [[F(2,309)=.838, p=.433] and the two-way interaction of test occasion X age X treatment [[F(2,309)=.635, p=.531] were not significant. The univariate effects, displayed in Table 5, indicated that on all three test occasions the controls scored higher than the treatment group. Self-evaluation training did not contribute to mathematics achievement. Age was a significant predictor of achievement only on the pretest and there was significant age X treatment effect on all occasions. Inspection of the means indicated that 11-13 year olds had higher achievement than 9-10 year olds on all test occasions. The two-way interaction indicated that there was a monotonic relationship between age and achievement in all three test periods in the treatment group: 12-13 year olds scored higher than 11year olds who outperformed 9-10 year olds. For the control group the relationship was more complex: on the pre- and post-tests 11years outperformed 12 year olds who outscored 9-10 year olds. On the retention test the pattern was monotonic: 9-10 < 11 < 12-13 year olds. The interactions of age with treatment do not change the overall picture: the treatment had no achievement impact.

The pattern continued in subsequent analyses (not reported). In the second run, pretest self-evaluation scores replaced age as the covariate. In the third run, pretest internal attributions for success was the covariate. In all of these runs the covariate was a significant predictor of achievement and experimental condition had no independent effect. A final run in which age, self-evaluation scores and internal attributions for success were entered as covariates also produced no treatment effects. The treatment did not contribute to higher achievement in mathematics.

Variation in Treatment Implementation

Teacher self-reports and the artifacts collected during the in-service sessions indicated that some of the treatment teachers taught self-evaluation mainly in subjects other than mathematics. The reasons teachers gave for avoiding math were: students were reluctant because they lacked key terms for describing their work, they were uncomfortable due to math anxiety, and some students had difficulty seeing gradations in performance, believing answers were correct or not. These features made it more difficult to involve students in talking about and applying evaluation criteria. Four of the seven teachers reported parallel innovations. These teachers taught students how to self-evaluate in language and social skills and they simultaneously changed their mathematics instruction (by increasing attention to problem solving in general and to data management and probability skills in particular). What they did not do was combine the innovations.

In contrast, two treatment teachers and to a lesser degree a third, taught students how to evaluate their performance in mathematics. Although they also used self-evaluation in other subjects, their primary attention was to mathematics. These teachers involved students in the design of rubrics for mathematical tasks and helped them see common features across problem types. The teachers demonstrated how to apply evaluation criteria and had students practice the criteria by looking at alternate solutions generated by their peers. These sharing procedures often led to a revision of the criteria, usually making the rubrics more precise. One teacher had students maintain a running record linking their self-evaluations to an accumulating list of strategies for solving problems and to a personal goal-setting procedure.[4]

Discussion

The first finding of the study is that students became more accurate when taught how to evaluate their work. The tendency of students to inflate their self-evaluations declined in the treatment group, while remaining unchanged among controls. Treatment students began the study significantly behind controls but erased the disadvantage. They caught up to controls on the posttest and continued at the higher level on the retention test. Although the effect sizes were small, the treatment had a statistically significant effect. This finding confirms the results of Ross et al. (1998) who found that a similar treatment increased students' ability to assess their writing accurately.

This is an important finding because elementary students consistently over-estimate their performance. These exaggerations have a negative effect on students' willingness to seek help or to change their learning strategies. In addition, many teachers believe that if students are allowed to evaluate their own work they will give themselves inflated scores. A substantial number of students share this view, believing that self-evaluation is unfair because it rewards cheaters (Ross, Rolheiser, & Hogaboam-Gray, in press-c). This belief reduces teachers' willingness to share classroom control with students, a key indicator of reform in mathematics education.

The second finding of the study is that training in self-evaluation had no impact on student achievement. The treatment group began the study with significantly lower achievement scores. These differences were never overcome. Even when a variety of covariates were introduced into the analysis there were no treatment effects. There are several reasons why the treatment had no impact:

First, despite random assignment of classes, the groups were not equivalent. The control group began the study with higher problem solving skills, more accurate self-appraisals and took greater responsibility for their school success (i.e., they were more willing to attribute success to factors within their control). Although we made statistical adjustments to control for these differences, we could not adjust for other impacts ability differences might have had. Previous research has found that the ability level of the class predicts student opportunities to learn through teacher pacing decisions and observation of peer performance (Gamoran, Nystrand, Berends, & LePore, 1995).

Second, control group teachers may have overcompensated. Cook and Campbell (1979) observed that subjects assigned to what they perceive to be a less desirable condition (the situation here) may engage in compensatory rivalry with treatment units. In addition control group teachers had a less onerous task. The treatment teachers had the dual responsibility of implementing the probability unit and experimenting with self-evaluation. The control group teachers were able to focus all their energy on the probability unit, a topic that had previously been given relatively little attention.

Third, we under-estimated how difficult it was for teachers to apply the self-evaluation activities in our handbook to the mathematics curriculum. Many of the treatment group teachers had students practice their self-evaluation skills in language or in social skills rather than in mathematics. Although we demonstrated how to translate self-assessment instruments developed for social skills into a mathematics instrument, the process was time consuming and teacher preparation time was limited.

Fourth, the treatment may have been too short. Students' understanding of their role in evaluation develops over their entire school career. An 8-week intervention may be insufficient to overturn the belief that evaluation is something done to students rather than a process in which they have a personal responsibility. In our previous research we found that student cognitions about evaluation changed when they were taught self-evaluation techniques but many of the misconceptions they had about the process, particularly its contribution to improved performance, continued unabated (Ross et al., in press-c).

Finally, it is possible that our quantitative measures did not capture changes in students' thinking that might have been revealed by observing students discussing their solutions in small groups or interviews focused on their grasp of key probability concepts.

This study contributes to our knowledge in three ways. First, it demonstrates that self-evaluation training clarifies student understanding of math curriculum expectations, thereby increasing their ability to accurately assess their performance. Second, it is one of the few studies to assess the effectiveness of the prototypical strategy for teaching probability. Although the study did not include a no-treatment control group, it provided data on the effects of the prototypical strategy on samples that differed in their use of self-evaluation. Third, the study provided evidence of the null effects of self-evaluation training on mathematics achievement, in contrast with Shepard et al. (1996) who found a small but significant effect (ES=.13) for performance assessment on mathematics learning. This finding weakens the case for the consequential validity of authentic assessment instruments (e.g., Wiggins, 1993), at least with respect to the argument that it will enhance student achievement.

In conclusion, the study suggests that training in self-evaluation has promise but it has yet to deliver in the field of mathematics.

References

Aitchison, R. E. (1995). The effects of self-evaluation techniques on the musical performance, self-evaluation accuracy, motivation and self-esteem of middle school instrumental music students. Unpublished doctoral dissertation, University of Iowa.

Adams, D., & Shrum, J. (1990). The effects of microcomputer-based exercises on the acquisition of line graph construction and interpretion skills of high school biology students. Journal of Research in Science Teaching, 27(8), 777-788.

Akers, J., Finzer, W., Gutierrez, J., & Resek, D. (1987). Sampling. Journal of Mathematical Behavior, 6, 149-156.

American Statistical Association--National Council of Teachers of Mathematics Joint Committee. (1985). Teaching statistics within the K-12 curriculum. Washington: American Statistical Association.

Arter, J., Spandel, V., Culham, R., & Pollard, J. (1994, April). The impact of training students to be self-assessors of writing. Paper presented at the American Educational Research Association, New Orleans.

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

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. Educational Psychologist, 28(2), 117-148.

Bennett, B., Rolheiser-Bennett, C., & Stevahn, L. (1991). Cooperative learning: Where heart meets mind. Toronto: Educational Connections.

Bohren, J. (1989). A nine-month study of graph construction skills and reasoning strategies used by ninth grade students to construct graphs of science data by hand and with computer graphing software. Doctoral dissertation, Ohio State University. Dissertation Abstracts International, 49, 2984-A.

Boud, D. J., Churches, A. E., & Smith, E. M. (1986). Student self-assessment in an engineering design course: An evaluation. International Journal of Applied Engineering Education, 2(2), 83-90.

Bramald, R. (1994). Teaching probability. Teaching Statistics, 16(3), 85-89.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B. S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. American Educational Research Journal, 34(4), 663-689.

Brutlag, D. (1994). Choice and chance in life: The game of "Skunk". Mathematics Teaching in the Middle School, 1(1), 28-33.

Butler, R. (1990). The effects of mastery and competitive conditions on self-assessment at different ages. Child Development, 61, 201-210.

Cook, T., & Campbell, D. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago, IL: Rand McNally.

Crocker, A. C., & Cheeseman, R. G. (1988). The ability of young children to rank themselves for academic ability. Educational Studies, 14(1), 105-110.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. Review of Educational Research, 59(4), 395-430.

Fernandes, M., & Fontana, D. (1996). Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. British Journal of Educational Psychology, 66, 301-313.

Fontana, D., & Fernandes, M. (1994). Improvements in math performance as a consequence of self-assessment in Portuguese primary school pupils. British Journal of Educational Psychology, 64, 407-417.

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. American Educational Research Journal, 32(4), 687-715.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. Journal for Research in Mathematics Education, 19(1), 44-63.

Glass, G., McGaw, B., & Smith, M. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Green, D. R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & C. Constable (Eds.), Proceedings of the first international conference on teaching statistics (pp. 766-783). Sheffield, UK: Teaching Statistics Trust.

Hacking, I. (1975). The emergence of probability. A philosophical study of early ideas about probability, induction, and statistical inference. London: Cambridge University Press.

Hansen, R., McCann, J., & Myers, J. (1985). Rote versus conceptual emphases in teaching elementary probability. Journal for Research in Mathematics Education, 16(5), 364-374.

Henry, D. (1994). Whole language students with low self- direction: A self-assessment tool. Virginia: University of Virginia. ED 372359.

Hughes, B., Sullivan, H., & Mosley, M. (1985). External evaluation, task difficulty, and continuing motivation. Journal of Educational Research, 78(4), 210-215.

Jiang, Z., & Potter, W. D. (1994). A computer microworld to introduce students to probability. Journal of Computers in Mathematics and Science Teaching, 13(2), 197-222.

Johnson, D., Johnson, R., & Stanne, M. (1990). Impact of goal and resource interdependence on problem-solving success. Journal of Social Psychology, 129(5), 621-629.

Johnson, D., Johnson, R., Stanne, M., & Garibaldi, A. (1990). Impact of group processing on achievement in cooperative groups. Journal of Social Psychology, 130(4), 507-516.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. London: Cambridge University Press.

Konold, C. (1989). Informal conceptions of probability. Cognition and Instruction, 6(1), 59-98.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies about students' reasoning about probability. Journal for Research in Mathematics Education, 24(5), 392-414.

Kulm, G. (1994). Mathematics assessment: What works in the classroom. San Francisco: Jossey-Bass.

Leinhardt, G., Zaslavsky, O., & Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. Review of Educational Research, 60(1), 1-64.

Markman, E. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

Meece, J., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. Journal of Educational Psychology, 80(4), 514-523.

Ontario Association for Mathematics Education. (1996). Linking assessment and instruction in mathematics. Rosseau, ON: author.

Ontario Ministry of Education and Training. (1997). Grades 1-8 Mathematics. Toronto: author.

Renkl, A. (1997, March). Learning by explaining - or better by listening? Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Rolheiser, C. (Ed.). (1996). Self-evaluation: Helping students get better at it. Toronto: Visutronx.

Ross, J. A. (1995). Effects of feedback on student behavior in cooperative learning groups in a grade 7 math class. Elementary School Journal, 96(2), 125-143.

Ross, J.A. & Cousins, J. B. (1993a). Enhancing secondary school students' acquisition of correlational reasoning skills. Research in Science & Technological Education, 11(3), 191-206.

Ross, J.A., & Cousins, J. B. (1993b). Patterns of student growth in reasoning about correlational problems. Journal of Educational Psychology, 85(1), 1-17.

Ross, J.A., Rolheiser, C., & Hogaboam-Gray. (1998, May). Effects of self-evaluation training on narrative writing. Paper presented at the annual meeting of the Canadian Association for Curriculum Studies, Ottawa.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (in press-a). Student evaluation in cooperative learning: Teacher cognitions. Teachers and Teaching.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (in press-b). Teaching students how to self-evaluate when learning cooperatively: The effects of collaborative action research on teacher practice. Elementary School Journal.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (in press-c). Skills training versus action research in-service: Impact on student attitudes to self-evaluation. Teaching and Teacher Education.

Schunk, D. (1981). Modelling and attributional effects on children's achievement: A self-efficacy analysis. Journal of Educational Psychology, 73(1), 93-105.

Schunk, D. (1995, April). Goal and self-evaluative influences during children's mathematical skill acquisition. Paper presented at the annual conference of the American Educational Research Association, San Francisco.

Schunk, D. (1996). Goal and self-evaluative influences during children's cognitive skill learning. American Educational Research Journal, 33(2), 359-382.

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1996). Effects of introducing classroom performance assessments on student learning. Educational Measurement: Issues and Practice, 15(3), 7-18.

Sparks, G. E. (1991). The effect of self-evaluation on musical achievement, attentiveness and attitudes of elementary school instrument students. Unpublished doctoral dissertation, Louisiana State University and Agricultural and Mechanical College.

Stipek, D., Roberts, T., & Sanborn, M. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. Child Development, 59, 1983-1989.

Stipek, D., Recchia, S., & McClintic, S. (1992). Self- evaluation in young children. Monographs of the Society for Research in Child Development, 57(1), 1-84.

Vispoel, W., & Austin, J. (1995). Success and failure in junior high school: A critical incident approach to understanding students' attributional beliefs. American Educational Research Journal, 32(2), 377-412.

Wavering, M. (1984). Interrelationships among Piaget's formal operations schemata: Proportions, probability, and correlation. Journal of Psychology, 118(1), 57-64.

Webb, N. (1991). Task-related verbal interaction and mathematics learning in small groups. Journal for Research in Mathematics Education, 22(5), 362-365.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. (1971). Perceiving the causes of success and failure. New York, NY: General Learning Press.

Wiggins, G. (1993). Assessing student performance: Explore the purpose and limits of testing. San Francisco, CA: Jossey-Bass

Endnotes

[1] No systematic data on students' social class or ethnicity were collected. Teachers in both conditions reported that students came from a range of economic circumstances with few visible minorities (i.e., orientals, blacks). A few (<2%) mentally handicapped students were excluded from both samples. We also administered a battery of teacher instruments finding no significant pretest differences between treatment and control teachers on teacher efficacy, self-reported use of student assessment procedures, gender, teaching experience, and beliefs about teaching. These data are not reported because the sample size is too small for meaningful comparisons.

[2] The formula was the original score X 3.5 divided by 10.

[3] Although teachers varied in the specific topics they addressed during the first 6 weeks of the project, most focused on number sense and numeration with some attention to measurement. Very few addressed grade 5-6 topics in geometry and number sense or patterning and algebra.

[4] Pre to post and pre to retention scores were modestly higher in the classes of teachers who integrated self-evaluation training with mathematics instruction than in the classes of those who maintained separate innovations. These data are not displayed because the sample size is so small.

Table 1 Means and Standard Deviations of Student Variables, by Experimental Condition on Each Test Occasion

	Pre				Post				Retention			
	Treatment (n=164)		Control (n=142)		Treatment (n=164)		Control (n=142)		Treatment (n=164)		Control (n=142)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Achievement	1.86	.84	2.02	.74	2.12	.62	2.22	.55	2.21	.70	2.36	.61
Self-evaluation	7.43	1.85	7.00	2.32	7.22	2.05	7.80	1.80	7.47	1.79	7.57	1.92
Accuracy of Self-evaluation	4.15	1.27	4.56	1.20	4.64	1.02	4.74	.97	4.69	1.12	4.75	1.09
Self-efficacy	7.47	1.61	7.74	1.73								
Locus of control												
Internal success	4.05	.61	4.18	.57								
Internal failure	2.36	.89	2.30	.69								
External failure	2.35	.77	2.37	.76								
External success	3.42	.66	3.33	.71								
Goal orientations												
Mastery	3.71	.69	3.84	.66								
Ego	2.83	.91	2.89	.91								
Affiliation	3.71	.79	3.76	.71								

	Treatment (n=176)		Control (n=174)	
	% Male	% Female	% Male	% Female
Gender	49	51	54	46

	Treatment (n=176)			Control (n=174)		
	% 9-10	% 11	% 12-13	% 9-10	% 11	% 12-13
Age	25	56	19	56	29	16

Table 2 Pretest Equivalence of Student Groups (N=323)

				Correlations with			
				Accuracy		Achievement	
Variables	t	df	p	Post	Retention	Post	Retention
Self-efficacy	-1.48	321	.140				
Self-evaluation	2.09	294.55	.037*	-.10	.04	.18**	.27**
Evaluation attitudes	.66	321	.508				
Locus of control							
Internal success	-2.103	321	.036*	-.09	.00	.18**	.22**
External success	1.26	321	.209				
Internal failure	.71	321	.480				
External failure	-.12	321	.908				
Goal orientations							
Mastery	-1.87	321	.062				
Ego	-.57	321	.571				
Affiliation	-.17	321	.866				
Achievement	-1.89	320.24	.060				
Accuracy	-3.20	321	.002*	--	--	.16**	.25**
Age	11.76	287.72	.000**	.04	.09	.18**	.19**

*$p < .05$. **$p < .001$.

Table 3 Between-Subject Effects of Treatment on Accuracy of Self-Evaluation: GLM Univariate Effects

Dependent Variable	Parameter	B	Standard Error	t	p
Pretest	Intercept	.733	.010	76.711	.000
	Treatment	-3.6E-02	.013	-2.756	.006
Posttest	Intercept	.752	.007	110.685	.000
	Treatment	-8.4E-03	.009	-.903	.367
Retention	Intercept	.745	.008	89.859	.000
	Treatment	-2.3E-04	.011	-.020	.984

Table 4 Between-Subject Effects of Treatment on Achievement: GLM Univariate Effects

Dependent Variable	Parameter	B	Standard Error	t	p
Pretest	Intercept	.464	.010	48.829	.000
	Treatment	-2.4E-02	.013	-1.856	.064
Posttest	Intercept	.501	.007	76.490	.000
	Treatment	-1.4E-02	.009	-1.598	.111
Retention	Intercept	.516	.008	66.909	.000
	Treatment	-1.9E-02	.011	-1.819	.070

Table 5 Between-Subject Effects of Treatment and Age on Achievement: GLM Univariate Effects

Dependent Variable	Parameter	B	Standard Error	t	p
Pretest	Intercept	.239	.068	3.495	.001
	Treatment	-.489	.119	-4.129	.000
	Age	.350	.105	3.336	.001
	Treatment x Age	.653	.175	3.729	.000
Posttest	Intercept	.464	.049	9.419	.000
	Treatment	-.358	.085	-4.188	.000
	Age	5.6E-02	.076	.738	.461
	Treatment x Age	.496	.126	3.928	.000
Retention	Intercept	.419	.058	7.167	.000
	Treatment	-.335	.101	-3.310	.001
	Age	.150	.090	1.672	.096
	Treatment x Age	.450	.150	3.011	.003

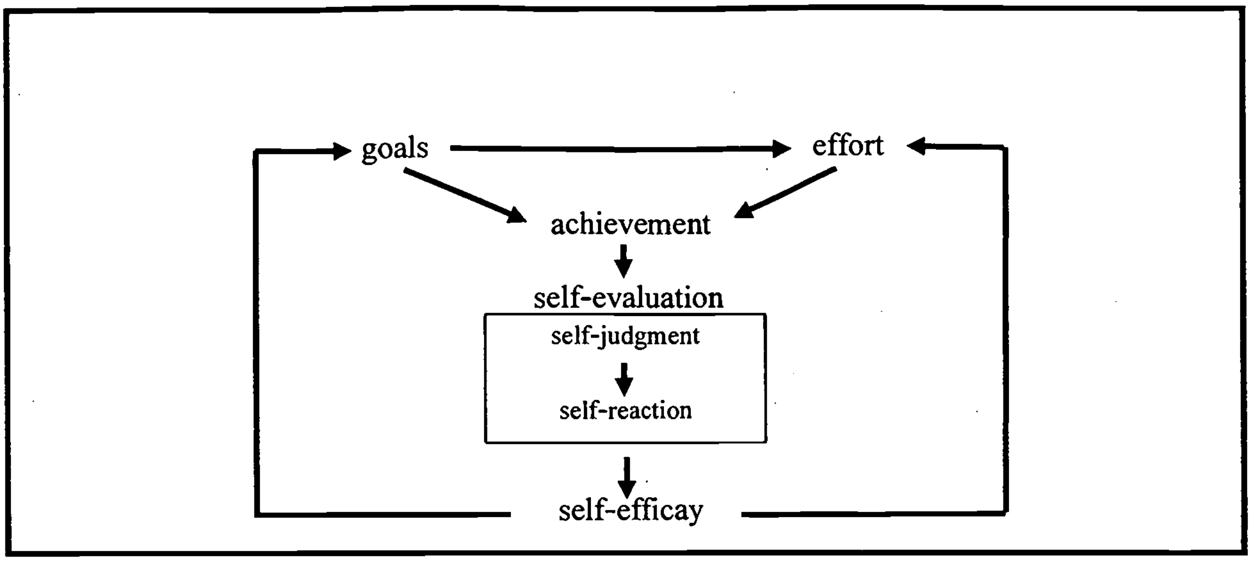

Figure 1 How Self-Evaluation Contributes to Learning

Appendix Rubric/Coding Scheme

Indicators	Level 1	Level 2	Level 3
General Characteristics	Does not meet the outcome; unsatisfactory	Meets the outcome, minimally satisfactory	Mastery, solid response, clear understanding of concepts, procedures & communication
Dog Pen Item			
Reasoning	No overt strategy	No apparent strategy for generating pens but more than one illustrated	Systematic strategy for generating pens; at least 3 illustrated
Accuracy of concepts & computations	Conceptual confusion (area, perimeter, rectangle), mixture of units, unlabeled diagrams	Pens not accurately labelled (e.g., units not given or error in dimensions)	All pens accurately labeled, showing relative size, correct measurements
Communication	No argument	Weak argument	Credible argument for selecting one pen that links real life to math
[Distinguishing high from low responses]	[1+ some attempt at labels, multiple pens, argument; 1- no labels, single diagram, not rectangles]	[2+ virtually a 3 but missing a major element such as units on diagram or error; 2- very poor reasoning]	[3+ as above; 3- if minor error made, e.g., missing labels]
Data Management Item	1+ graph does not correspond to table or graph missing	2+ more than one error or item missing	3+ accurate & complete table; graph has equally spaced columns, with scales, labels and titles on each axis
	1- little or no response	2- table incomplete or major errors but graph drawn; serious flaws in graph or incomplete	3- minor errors in table or 1 item missing from graph
Probability Item	1+ does not understand probability as a ratio but has one correct element	2+ correct reasoning but expresses only part of the correct ratio (e.g., 3)	3+ Correct format (fraction or ratio of 3/11); expressed accurately using math terms
	1- unintelligible or missing response	2- correct reasoning; expresses an incorrect ratio or incorrect format	3- Correct format but wrong ratio (e.g., ¾); expressed accurately in words or correct answer but unclear explanation

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

TM028943

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Impact of Self-Evaluation Training on Mathematics Achievement in a Cooperative Learning Environment	
Author(s): John A. Ross, Carol Rolheiser, Anne Hogaboam-Gray	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title: Dr. John Ross, Professor
Organization/Address: OISE - University of Toronto	Telephone: / FAX:
	E-Mail Address: jross@oise.utoronto.ca / Date: Apr 17/98

Box 719, 150 O'Carroll Ave
Peterborough, Ontario
K9J 7A3 CANADA

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:	
Address:	
Price:	

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:	
Address:	

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

The Catholic University of America
ERIC Clearinghouse on Assessment and Evaluation
210 O'Boyle Hall
Washington, DC 20064
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com