Student Self-Evaluation in Grade 5-6 Mathematics

Effects on Problem Solving Achievement

John A. Ross*

Anne Hogaboam-Gray

Carol Rolheiser

Ontario Institute for Studies in Education

University of Toronto

Cite as: Ross, J.A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics: Effects on problem-solving achievement. Educational Assessment, 8(1), 43-59.

*Corresponding Author
OISE/UT Trent Valley Centre,
Box 719, 633 Monaghan Rd. S.
Peterborough, ON K9J 7A1
CANADA

Student Self-Evaluation in Grade 5-6 Mathematics

Effects on Problem Solving Achievement

Abstract

We examined the effects of self-evaluation training on mathematics achievement. When Grade 5-6 students self-evaluated for 12 weeks ($N=259$ treatment, 257 control) treatment students outperformed controls ($ES=.40$). The findings contribute to the consequential validity argument for self-evaluation. Considered in the context of previous research, these results indicate that subject moderates the effects of self-evaluation on achievement.

Student Self-Evaluation in Grade 5-6 Mathematics

Effects on Problem Solving Achievement¹

Recent research on student assessment has focused on the consequences of test use. Some researchers, most notably Moss (1998), have framed consequential validity as the ultimate criterion for determining the adequacy of assessment programs. Even those who reject consequences as a dimension of validity (e.g., Popham, 1997) regard the impact of tests on student learning as an important consideration worthy of test developer attention. This article continues this line of research by examining the impact of self-evaluation on grade 5-6 student achievement of mathematics. It complements our previous research on the effects of student self-evaluation on teacher practice (Ross, Rolheiser, & Hogaboam-Gray, 1998a; 1998b; 1999b) and studies of the student and teacher effects of other methods of student assessment on student achievement such as portfolio (Underwood, 1998) and performance assessment (Fuchs, Fuchs, Karns, Hamlett, & Katzaroff, 1999).

Theoretical Framework

In this section we outline the theory that guided our inquiry into the student achievement consequences of teaching students how to evaluate their work. We then review previous studies of self-evaluation training and identify the research question that guided the inquiry.

Self-Evaluation Training

Figure 1 describes a model linking self-evaluation to student achievement. Models are simplifications. Achievement is influenced by myriad other factors, e.g., prior knowledge/ability, opportunity to learn, home background--to name but three salient variables not identified in Figure 1. We excluded these variables in order to focus attention on factors that are most directly

related to the influence of self-evaluation on student achievement. Although the constructs in the model are conceptually distinct, they overlap empirically (Wigfield, Eccles, & Rodriguez, 1998).

Figure 1 About Here

We adopted Klenowski's (1995) definition of self-evaluation as "the evaluation or judgment of 'the worth' of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes" (p. 146). Self-evaluation embodies three processes that self-regulating students use to observe and interpret their behavior (Schunk, 1996; Zimmerman, 1989). First, students produce self-observations, deliberately focusing on specific aspects of their performance relevant to their subjective standards of success. Second, students make self-judgments in which they determine how well their general and specific goals were met. Third, are self-reactions, interpretations of the degree of goal achievement that express how satisfied students are with the result of their actions. Not shown in this version of our model is the impact of attempts by teachers, peers and parents to influence each of these processes (e.g., Ross, Rolheiser, & Hogaboam-Gray, in press).

Self-evaluation contributes to perceived self-efficacy, i.e., "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 2). Students who perceive themselves to have been successful on the current task are more likely to believe that they will continue to be efficacious in the future (Bandura, 1997). Self-efficacy influences achievement in three ways.

First, students with greater confidence in their ability to accomplish the target task are more likely to visualize success than failure. They set higher standards of performance for themselves. Student goals can be categorized at general and specific levels. General goals identify student motives for engaging in a classroom activity. Goal theorists distinguish two types of

general goal orientations, variously labeled mastery versus ability focused (Ames, 1992), learning versus performance (Dweck & Leggett, 1988), and task versus ego (Nicholls, 1984). Students who set mastery goals approach school tasks by focusing on what they might learn from their participation. They define success as mastering a skill or developing understanding and seek tasks that are more likely to generate these outcomes. Students who set ability goals focus on opportunities to demonstrate their abilities. Success is measured by higher grades or greater status in comparison to peers. Students who set ability goals seek familiar task situations. The highest academic performance is obtained when a mastery orientation predominates (Meece, Blumenfeld, & Hoyle, 1988). Although most research focuses on these two orientations, a third goal orientation, social affiliation, has been examined. In most studies, students who approach a school task by focusing on the opportunities it provides for social interactions with peers have lower achievement than students with high learning goal orientations (Atkinson & O'Connor, 1966; Schneider & Coutts, 1985; Schneider & Green, 1977). However, achievement and the need for affiliation can be positively associated, particularly in classrooms using cooperative learning techniques (Urdu & Maehr, 1995). At the specific goal level, achievement is likely to be higher when students focus on the lesson objectives embedded within the task.

Second, student expectations about future performance also influence effort. Confident students persist. They are not depressed by failure but respond to setbacks with renewed effort. For example, students with high self-efficacy interpret a gap between aspiration and outcome as a stimulus while low self-efficacy students perceive such a gap as debilitating evidence that they are incapable of completing the task (Bandura, 1997). Student effort, influences how well students achieve their goals, since persistence increases accomplishment. Effort is also influenced by students' 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).

Third, self-efficacy beliefs influence the impact of anxiety on achievement. Anxiety is a strong negative predictor of mathematics achievement (the meta-analysis of Schwarzer, Seip, & Schwarzer, 1989 found a correlation of $r = -.30$). Bandura (1997) argued that people do not generate anxiety in response to threats for which they feel a high coping efficacy. Confirming the moderating impact of self-efficacy on the anxiety-achievement relationship is evidence that self-efficacy is positively related to math achievement and negatively related to anxiety (Malpass, O'Neil, & Hocevar, 1999) and the finding that the predictive power of math anxiety on achievement diminishes when math self-efficacy is controlled (Pajares & Urdan, 1996).

Self-evaluation plays a key role in fostering an upward cycle of learning when the child's self-evaluations are positive. Positive self-evaluations encourage students to set higher goals and commit more personal resources to learning tasks (Bandura, 1997; 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).

Training in self-evaluation strategies could enhance achievement in several ways. Training could heighten self-observation. Since teachers try to assign tasks that students can complete successfully, student judgments of their performance would likely be positive and the upward cycle in which self-evaluation stimulates achievement could accelerate. Self-evaluation training could also modify the specific goals that students set, bringing them more closely in line with teacher expectations. Training might also spur students to greater efforts if they became more conscious of the specific gaps between their performance and their goals.

Studies of the Effects of Self-Evaluation Training on Mathematics Achievement

In Fontana and Fernandes (1994), self-evaluation was embedded in a broader instructional treatment--a mathematics program to increase students' control of their learning. In the early phase of this 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 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. The distinctive contribution of self-evaluation to the effects could not be disentangled from other program components.

Schunk (1996) implemented self-evaluation by asking grade 4 students to judge how certain they were they could solve computational problems. Self-evaluation had no effect on mathematics achievement in a learning goal condition (i.e., when students were told that the purpose of the activity was to learn how to solve fraction problems). 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). Despite the impressive effects attained in a very brief treatment, Schunk's study does not generalize well to classrooms because students were given no information about their performance and the instructional tasks (decontextualized short answer problems) did not match the rich problems embedded in meaningful situations (e.g., as in Fuchs et al., 1997) characteristic of current directions in mathematics education.

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 group interactions and were trained how to interpret them. They used an instrument 1-2 times per week for 12 weeks to record the frequency of positive interactions. Self-assessment increased the frequency of productive help giving, help seeking, and attitudes about asking for help. However, in this study no measure of mathematics achievement was administered. Other studies have also reported that self-evaluation has a positive effect on factors associated with achievement without measuring achievement directly. For example, self-evaluation increases persistence (Henry, 1994; Hughes, Sullivan, & Mosley, 1985; Schunk, 1996).

Research Question

Our approach to teaching students how to evaluate their work began with a study of the student assessment practices of cooperative learning teachers identified by their peers and supervisors as exemplary (Ross, Rolheiser, Hogaboam-Gray, 1998a). 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. Our four stages were similar to the three essential elements of self-evaluation training independently identified by Klenowski (1995): use of criteria negotiated by teacher and students, teacher-student dialogue that focuses on evidence for judgments, and ascription of a grade by students (alone or in collaboration with teachers). Strategies for each stage and classroom usable tools (Rolheiser, 1996) were elaborated by a school-university partnership. Use of these strategies had a positive effect on student attitudes to evaluation (Ross et al., 1998b; 1999b) and on language achievement (Ross, Rolheiser, & Hogaboam-Gray, 1999a). Our goal was to extend these findings to mathematics. The specific

research question that guided our research was: Will grade 5-6 students who are trained how to evaluate their work have higher mathematics achievement?

Method

Participants

Twelve grade 5-6 teachers, seven of whom participated in the treatment group in an 8-week pilot study conducted a year earlier, volunteered to be in the 12-week treatment described below. They were matched (exactly on grade and gender and within 5 years of teaching experience) with 12 teachers from an adjacent district who were not using systematic self-evaluation procedures but were interested in innovative approaches to student assessment. Control group teachers were promised that at the end of the study they would have the opportunity to attend a half-day workshop (held during school hours) on self-evaluation techniques and would receive a collection of student assessment materials

Measures

Students completed a performance task at the beginning and end of the project. Both were adapted from Kulm (1994). On the pretest students were given \$10 to purchase items for a puppy. There were prices from three retail outlets for each of the three items. The task was to “Choose a possible selection of collar, dish and toy that you could buy. What is the cost? How much change would you receive? How many different ways could you buy the three items and still spend \$10 or less. Show each combination.” On the posttest the task was “You have been hired by Hanson [a musical group popular with grade 5-6 students at the time of the study] to design their stage for their upcoming concert in Toronto. The stage should be rectangular and have an area of 200 square meters. It will have a security rope on both sides and the front. Make some

drawings of different rectangles that have 200 square meters of area. How many meters of security rope are needed for each one? Which shape would you use? Why?"

The tests were independently coded by two experienced teachers using a rubric of three dimensions: strategy for generating a solution, accuracy of concepts and computations, and communication of solution. For each dimension the rubric described four levels: unsatisfactory, minimal competence, mastery, showing extensions beyond expectations for the grade. Markers first holistically assigned a level to each response and then determined whether it was high or low within the assigned level, creating an 8-point scale (1-, 1+, 2-, 2+, 3-, 3+, 4-, and 4+). Inter-rater reliability was acceptable: Cohen's $\kappa=.73$ for exact agreement and $.90$ for within one point (i.e., half level) on the scale. The pretest predicted posttest performance ($r=.37$, $p<.001$, $n=476$).

The student tests of sample equivalence consisted of 8 instruments. Alpha reliabilities were calculated using pretest scores ($N=514$). Self-evaluation was measured with 6 items ($\alpha=.91$). After completing the pretest achievement task students rated their overall performance with a 1-10 scale (anchored by 1=not well and 10= very well). They used the same scale to rate five dimensions of their problem solving ("How well you...understood the problem, made a plan, solved the problem, checked the solution, and explained the solution."). Student self-efficacy ($\alpha=.91$) consisted of 6 items identical to the self-evaluation measure except that each asked about expectations about future performance "how sure are you that you could..." rather than focusing on past performance ($\alpha=.89$). Attitudes to self-evaluation consisted of 10 Likert items adapted from Paris, Turner, and Lawton (1990) and Wiggins (1993). There were 5 positive items (e.g., "I like self-evaluation because...it tells me the areas I need to improve on") and 5 negative items (e.g., "I dislike self-evaluation because...we really don't know how to mark ourselves"). Ross et al. (1998b) presented evidence of the validity and reliability of the instrument

but in contrast with previous results in this study the internal consistency was poor ($\alpha = .53$). Math confidence (i.e., less threatening expressions of math anxiety, $\alpha = .78$) and math anxiety (i.e., more threatening expressions of anxiety, $\alpha = .87$) consisted of 10 Likert items, originally developed by Betz (1978), that have been extensively used in previous research. Pajares and Urdan (1996) presented evidence of their reliability and validity. The goal orientations survey consisted of 16 items from Meece et al., (1988) distinguishing three orientations toward learning: mastery ($\alpha = .85$), ego ($\alpha = .60$), and affiliation ($\alpha = .56$).

We also administered a battery of teacher instruments finding no significant pretest differences between treatment and control teachers on teacher efficacy (Gibson & Dembo, 1984), self-reported use of student assessment procedures (Haydel, Oescher, Kirby, & Brooks, 1997) and beliefs about teaching (Rowan, 1995). These data are not reported because the sample size was too small for meaningful comparisons.

Study Conditions

The treatment was based on the theoretical framework in Figure 1. Treatment activities targeted each of the self-assessment elements (self-observation, self-judgment, and self-reaction) through four instructional stages: (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. Self-observation was influenced by involving students in the identification of the evaluation criteria used to assess their work (stage 1). The purposes were to make student learning goals explicit so that they would attend to the central rather than the peripheral elements of the lesson; to align student goals more closely with those of the teacher; and to ensure that students set attainable goals that would lead to mastery experiences. Self-judgment was influenced by teaching students how to recognize goal attainment

by modeling the application of criteria to student work (stage ii) and giving students feedback on the accuracy of their self-judgments (stage iii). The purpose of these activities was to help students recognize when they were successful. Self-reaction was influenced by helping students translate their self-judgments into action plans, primarily through teacher-student conferences (stage iv). The purposes of these activities were to set attainable goals for the future and to increase the likelihood that students perceived themselves to have mathematics ability.

There were six-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 reach consensus and then vote as a class on which criteria were most important. After determining the top four criteria, the teacher had each small 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 co-developed rubric to evaluate their work and for comparative purposes received a teacher evaluation using the same rubric. Students also set learning goals based on their self-evaluations and received comments from their teacher on the utility and feasibility of their plans.

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. For example, students might be asked to assess how well they performed the social skill of the day (such as “praising good ideas and disagreeing agreeably”) using a 4-

point “poorly” to “very well” scale. This form also had a place for the teacher to evaluate the student’s performance on the same scale. In a practice activity focused on mathematics, students used the same 4-point scale to rate how well they performed four recurrent problem solving skills (understanding the problem, devising a plan, carrying out the plan, and checking the solution). The form provided space for self-assessments on four separate occasions to provide students with progress over time. 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).

Training for treatment teachers included 15 after school hours with an additional 3 hours of self-directed in-school release time. Five half-day sessions modeled classroom activities (e.g., a tangram task² 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. Treatment teachers received a handbook on teaching self-evaluation (Rolheiser, 1996), a handbook containing performance tasks (Ontario Association for Mathematics Education, 1996), and a document providing examples of how to teach each of the four stages of self-evaluation in mathematics. Control teachers received no in-service on self-evaluation but they did receive 3 hours of self-directed in-school release time to develop activities for teaching problem solving and the performance tasks handbook.

Results

Table 1 displays the means and standard deviations for the treatment and control groups. Table 2 displays the correlations of each of the measures used to determine sample equivalence with pre- and posttest achievement.

Tables 1 and 2 About Here

There was a significant difference between the samples on affiliative orientations to learning (the treatment group was higher) and on age (the treatment group was younger). Neither affiliative orientation nor age was associated with posttest achievement but pretest achievement was ($r=.366$, $n=476$, $p=.001$). An analysis of covariance was conducted in which the outcome measure was mathematics achievement, the covariate was pretest achievement, and the independent variable was study condition (treatment or control). Pretest scores predicted posttest achievement [$F(1,475)=82.20$, $p=.000$] as suggested by the correlation of pre-and posttest scores. The treatment had a significant effect [$F(1,475)=14.58$, $p=.000$]. Treatment students outperformed controls. The effect was small ($ES=.40$).

Discussion

The results indicated that self-evaluation training in mathematics had a positive effect on mathematics achievement. The effect size was .40 SD, meaning that a student at the 50th percentile in the control group would have performed at the 66th percentile if he or she had been in the treatment group. If the 50th percentile were viewed as the cut-point defining a pass, the proportion of successful students increased by 32% in the treatment. Lipsey (1990)'s review of 186 meta-analyses of instructional improvement efforts found that the median effect size for all studies was .40, suggesting that the impact of self-evaluation training was comparable to that of other initiatives. However, the effect was higher than that reported for some other assessment based improvement efforts. For example, Shepard, Flexer, Hiebert, Marion, Mayfield, & Weston

(1996) reported that a performance assessment program had an $ES=.13$ on mathematics achievement, although higher impacts have been reported for other performance assessment training programs (Fuchs et al., 1999).

The results conflicted with the findings from a pilot study conducted a year earlier with teachers from the same district (Ross, Hogaboam-Gray, & Rolheiser, 2001). The pilot found no differences between treatment and control students. However, this study was flawed in several ways: treatment students were more confident than control students about their mathematics performance on entry to the program and the achievement measures used in the study matched each other and the curriculum poorly. Most importantly the treatment was one-third shorter (8 weeks rather than 12 weeks) and there was evidence that the program was too short for teachers and students. For example, teachers reported that students were reluctant to self-evaluate in mathematics because they lacked key terms for describing their work, they were anxious about the subject, and some students had difficulty seeing gradations in performance, believing answers were correct or not. Teachers in the pilot reported difficulty applying the principles of self-evaluation to mathematics and reported that many of the opportunities they provided for students to self-evaluate occurred in subjects other than mathematics or focused on social skills. We reduced both problems in the current study by extending the duration of the study, by increasing the amount of in-service, and by providing materials containing more math-specific examples.

Training in self-evaluation had an impact on mathematics achievement only when students experienced self-evaluation for a longer period and additional in-service and curricular support were provided to teachers. The treatment in the pilot was of the same type and duration as a study that found that the writing performance of grade 4-6 students improved when they were trained how to evaluate their narratives. Trained students increased integration of story elements around a

central theme and more consistently adopted a narrative voice than students in the control group (Ross et al., 1999a). The latter finding was congruent with previous research on the effects of self-evaluation training on language achievement. Hillocks (1986) reviewed seven studies (all but one unpublished) that found that student writing improved when students were given scales for judging writing samples and used them to appraise the work of their peers and themselves. Arter, Spandel, Culham, and Pollard (1994) reported significant treatment effects when grade 5 students were taught how to apply a trait analysis scheme to their writing. These results suggest that subject has a moderating effect on the impact of self-evaluation on achievement.

The reason why self-evaluation requires more implementation effort to have an effect in mathematics than in language might be attributable to the nature of mathematics and teachers' representation of it. The deep structure of mathematics may be more difficult for teachers to recover than the structures underlying writing. Most teachers understand mathematics as a set of arbitrary rules and algorithms to be memorized (Prawat & Jennings, 1997), in contrast with the reform view of mathematics as a dynamic set of intellectual tools for solving meaningful problems (e.g., NCTM, 2000). Teacher images of mathematics affect how they teach it. For example, Fennema, Franke, and Carpenter (1993) observed an exemplary teacher who was able to engender a high degree of student metacognition because her knowledge of mathematics was extensive, accurate, and hierarchically organized. In contrast, teachers who are uncertain about their grasp of the subject might find it difficult to identify performance criteria that run through a variety of topics within a course. Stein, Baxter, and Leinhardt (1990) found that a generalist teacher who lacked subject content knowledge was limited in her ability to make connections within grade 5 mathematics. Teachers might also be reluctant to share responsibility for assessment if they are uneasy about their ability to defend their own assessment decisions. Grade 5-6 teachers are more

likely to have taken undergraduate courses in language than in mathematics and assign greater priority to the former than the latter when making in-service selections (Spillane, 2000).

Experienced teachers teaching topics that are less familiar to them are more likely to use low risk strategies that reduce opportunities for student involvement in classroom decision making than when teaching familiar topics (Carlsen, 1993; Lee, 1995).

Students' ability to self-evaluate might also be negatively affected by the nature of the discipline. For example, students might find it difficult to identify and apply assessment criteria and to generate precise goals to remedy their deficiencies if they represent mathematics learning as the mastery of discrete facts and procedures. Resnick (1988) found that reciprocal teaching did not translate easily from language to mathematics classrooms, in large part because of difficulty in creating generic math prompts and student roles comparable to those used in the language version. The challenge faced in teaching self-evaluation in math may be embedded in the larger issue of the constructivist classroom in which teacher-student talk is a central vehicle for learning. Researchers (e.g., Williams & Baxter, 1996) have frequently observed how difficult it is for teachers, even those committed to constructivist approaches and trained in their use, to sustain productive dialogue on mathematics concepts.

Conclusion

The study reported in this article contributes to knowledge in two ways. First, research on the effects of mathematics education reform (reviewed in Ross, McDougall, & Hogaboam-Gray, forthcoming) indicates that implementation of reform contributes to problem solving achievement, that changing mathematics instruction is very difficult, and that teacher change has been obtained in multi-dimensional interventions. The nonsignificant result of the pilot study suggests that an intervention based only on changing assessment strategies may be insufficient to impact teaching

and learning of mathematics. In the study reported here greater attention was given to specific teacher concerns about using self-evaluation in mathematics and a booklet of subject-specific examples was provided. The positive results suggest that student assessment can be usefully combined with other dimensions of reform, particularly in-service that focuses on teachers' cognitions about the nature of mathematics.

Second, the study provides evidence of the positive effects of self-evaluation, providing additional support for the consequential validity argument for alternatives to traditional short answer tests. It has been argued that shifting to assessment practices that are tightly integrated into daily instruction will have beneficial consequences for teachers and students (e.g., Wiggins, 1993; 1998). It is predicted that such assessments will focus teacher attention on the objectives to be measured and provide teachers with more useful information than is afforded by traditional tests. This focusing of teacher attention will contribute to improved student performance. Very little empirical data is available to test these predictions, especially the claims about student effects. One of the few studies to explore the effects of alternative assessments (Shepard et al., 1996) found a small but significant effect ($ES=.13$) for performance assessment on mathematics learning but not for reading. Our studies found positive effects for self-evaluation: in writing (Ross et al., 1999a) and in mathematics (but only with increased support beyond that provided for the writing experiment). The findings suggest that the effects of student assessment vary with subjects in which they are embedded. Proponents of innovative approaches to student assessment should seek to identify the conditions under which particular assessment practices are more effective than traditional assessment, rather than assume that one approach will be universally superior.

References

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84, 261-271.
- 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.
- Atkinson, J. W., & O'Connor, P. (1966). Neglected factors in studies of achievement-oriented performance: Social approval as an incentive and performance decrement. In J. W. Atkinson, & N. T. Feather (Eds.), A theory of achievement motivation (pp. 299-326). New York: Wiley.
- Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W. H. Freeman.
- Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology*, 25, 441-448.
- Carlsen, W. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classrooms. Journal of Research in Science Teaching, 30(5), 471-481.
- Dweck, C., & Leggett, E. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95, 256-273.
- Fennema, E., Franke, M., & Carpenter, T. (1993). Using children's mathematical knowledge in instruction. American Educational Research Journal, 30(3), 555-583.
- 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.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. American Educational Research Journal, 36(3), 609-646.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. American Educational Research Journal, 34(3), 513-543.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569-582.

Haydel, J. B., Oescher, J., Kirby, P. C., & Brooks, C. R. (1997, March). Measuring the evaluative culture of classrooms. Paper presented at the annual meeting of the American Psychological Association, Chicago.

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

Hillocks, G. (1986). Research on written composition: New directions for teaching. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

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

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. Assessment in Education, 2(2), 145-163.

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

Lee, O. (1995). Subject matter knowledge, classroom management, and interactional practices in middle school science classrooms. Journal of Research in Science Teaching, 32(4), 423-440.

- Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.
- Malpass, J. R., O'Neil, H. F., & Hocevar, D. (1999). Self-regulation, self-efficacy, worry, and high-stakes math achievement for mathematically gifted high school students. *Roeper Review*, 21(4), 281-288.
- Meece, J., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80(4), 514-523.
- Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 6-12.
- National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.
- Ontario Association for Mathematics Education. (1996). Linking assessment and instruction in mathematics. Rosseau, ON: author.
- Pajares, F., & Urdan, T. (1996). Exploratory factor analysis of the Mathematics Anxiety Scale. *Measurement and Evaluation in Counselling and Development*, 29(1), 35-47.
- Paris, S., Turner, J., & Lawton, T. (1990, April). Students' views of standardized achievement tests. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Popham, W. J. (1997). Consequential validity: Right concern—wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Prawat, R., & Jennings, N. (1997). Students as context in mathematics reform: The story

of two upper-elementary teachers. Elementary School Journal, 97(3), 251-270.

Resnick, L. (1988). Teaching mathematics as an ill-structured discipline. In R. Charles & E. Silver (Eds.), The teaching and assessing of mathematical problem solving (pp. 32-60).

Hillsdale, NJ: Erlbaum.

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

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

Ross, J. A., McDougall, D., & Hogaboam-Gray, A. (forthcoming). Research on reform in mathematics education, 1993-2000. Submitted to Alberta Journal of Educational Research.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998a). Student evaluation in cooperative learning: Teacher cognitions. Teachers and Teaching, 4(2), 299-316.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998b). Skills training versus action research in-service: Impact on student attitudes to self-evaluation. Teaching and Teacher Education, 14(5), 463-477.

Ross, J.A., Rolheiser, C., & Hogaboam-Gray. (1999a). Effects of self-evaluation training on narrative writing. Assessing Writing, 6(1), 107-132.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999b). Teaching students how to self-evaluate when learning cooperatively: The effects of collaborative action research on teacher practice. Elementary School Journal, 99(3). 255-274.

Ross, J.A., Hogaboam-Gray, A., & Rolheiser, C. (2001, April). Effects of self-evaluation training on mathematics achievement. Paper presented at the annual meeting of the American Educational Research Conference, Seattle.

Ross, J.A., Rolheiser, C., & Hogaboam-Gray, A. (in press). Influences on student cognitions about evaluation. Assessment in Education.

Rowan, B. (1995, April). Teachers' instructional work: Conceptual models and directions for future research. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Schneider, F. W., & Coutts, L. M. (1985). Person orientation of male and female high school students: To the educational disadvantage of males? Sex Roles, 13, 47-63.

Schneider, F. W., & Green, J. E. (1977). Need for affiliation and sex as moderators of the relationship between need for achievement and academic performance. Journal of School Psychology, 15, 269-277.

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.

Schwarzer, R., Seip, B., & Schwarzer, C. (1989). Mathematics performance and anxiety: A meta-analysis. In R. Schwarzer, H. M. van Der Ploeg, & C. D. Spielberger (Eds.), Advances in Test Anxiety Research (pp. Vol. 6, pp.105-119). Berwyn, PA: Swets North America.

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.

Spillane, J. P. (2000). A fifth-grade teacher's reconstruction of mathematics and literacy teaching: Exploring interactions among identity, learning, and subject matter. Elementary School Journal, 100(4), 307-330.

Stein, M., Baxter, J., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. American Educational Research Journal, 27(4), 639-663.

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

Underwood, T. (1998). The consequences of portfolio assessment: A case study. Educational Assessment, 5(3), 147-194.

Urdan, T., & Maehr, M. (1995). Beyond a two-goal theory of motivation and achievement: A case for social goals. Review of Educational Research, 65(3), 213-243.

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. In P. D. Pearson, & A. Iran-Nejad (Eds.), Review of Research in Education, Vol. 23 (pp. 73-118). Washington, DC: American Educational Research Association.

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

Wiggins, G. (1998). Educative Assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.

Williams, S., & Baxter, J. (1996). Dilemmas of discourse-oriented teaching in one middle school mathematics classroom. Elementary School Journal, 97(1), 21-38.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. Journal of Educational Psychology, 81 (3), 329-339.

Table 1

Means & Standard Deviations, by Study Conditions and Test Occasion

	Pre							Post			
	Treatment (<u>n</u> =259)		Control (<u>n</u> =255)		t-test			Treatment (<u>n</u> =248)		Control (<u>n</u> =244)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>t</u>	<u>df</u>	<u>p</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Achievement	1.86	.84	1.98	.85	-1.57	496	.117	1.42	.90	1.20	.94
Self-evaluation	7.56	1.91	7.61	2.16	-.29	512	.770				
Self-efficacy	7.59	1.79	7.66	1.93	-.47	512	.638				
Evaluation attitudes	7.56	1.91	7.61	2.16	-.20	495	.842				
Math confidence	3.92	.75	3.93	.86	-.10	509	.923				
Math anxiety	3.76	.98	3.88	1.11	-1.29	509	.200				
Goal orientations											
Mastery	3.80	.70	3.73	.84	1.42	492	.157				
Ego	3.10	.96	3.07	.98	.36	492	.717				
Affiliative	3.52	.95	3.21	1.06	3.42	492	.001				
	Treatment(<u>n</u> =257)				Control (<u>n</u> =250)						
Percent Male	56				51						
Age (in percent)											
9	15				0						
10	41				40						
11	37				45						
12-13	8				16						

Table 2

Correlations of Sample Equivalence Measures with Pre- and Post-test Achievement

	Pre-test Achievement (<u>N</u> =514)	Post-test Achievement (<u>N</u> =494)
Self-evaluation	.28*	.17*
Self-efficacy	.21*	.18*
Evaluation attitudes	.01	.07
Math confidence	.18*	.12*
Math anxiety	.29*	.15*
Goal orientation		
Mastery	.15*	.02
Ego	-.22*	-.03
Affiliation	-.14*	-.07

* $p < .001$.

Figure 1: How self-evaluation contributes to student learning.

Endnotes

¹ This research was funded by the Social Sciences and Humanities Research Council of Canada and the Ontario Ministry of Education. The views expressed in the report do not necessarily reflect the views of the Council or the Ministry. An earlier version of the article study was presented at the annual meeting of the American Educational Research Association in Seattle, April 2001. We thank the teachers and students of Durham Catholic District School Board for their assistance.

² Tangram is an ancient Oriental toy which has 7 pieces, 5 triangles in different size, 1 square and 1 diamond. Thousands of shapes can be made by moving these simple shaped pieces around.