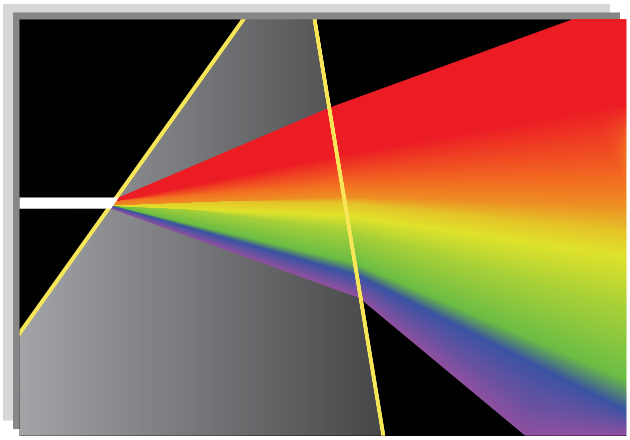

P rogramming
R emediation and
I ntervention for
S tudents in
M athematics

Kawartha Pine Ridge DSB
PRISM Project

Final Report[1]
May, 2006

John A. Ross – University of Toronto
John Ford - Kawartha Pine Ridge DSB
YunMei Xu – University of Toronto

[1] This project was funded by the Ontario Ministry of Education. The views represented in the report do not necessarily represent those of the Ministry. John Ross, Professor of Curriculum, Teaching, and Learning at OISE/UT took responsibility for all phases of the research, especially research design and reporting. John Ford, Intermediate Mathematics Consultant for the Kawartha Pine Ridge DSB, designed and implemented the in-service. YunMei Xu, Ph.D. candidate at OISE/UT contributed to the data analysis. Carolyn Brioux of the OISE/UT Trent Valley Centre formatted the surveys, scanned the data, and assisted with the data analysis. The research team expresses their thanks to the teachers and students who contributed data for the project.

Background to the Project

Introduction

Evidence has accumulated that implementation of standards-based mathematics teaching of the type embedded in the Ontario curriculum leads to higher achievement. For example, when teachers move from a traditional program emphasizing transmission of propositions and algorithms toward a constructivist approach, problem solving and conceptual understanding improve (Boaler, 1998; Brenner, et al., 1997; Cardelle-Elawar, 1995; Hamilton, McCaffrey, Stecher, Klein, Robyn, & Bugliari, 2003; Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000; Schoen, Fey, Hirsch, & Coxford, 1999) without loss of computational mastery (Mayer, 1999; Reys, Reys, & Koyama, 1996; Riordan & Noyce, 2001; Romberg, 1997; Schoen & Finn, 2001;Villasenor & Kepner, 1993; Wood & Sellers, 1997).

Despite consistent findings of positive student achievement effects, research reviewed in Ross, McDougall, and Hogaboam-Gray (2002) demonstrates that implementation of standards-based mathematics teaching is a formidable challenge for teachers. Sustained and purposeful professional development that provides teachers with the requisite skills and attitudes can have a beneficial impact on teacher ability to implement the central features of mathematics education reform.

Project Design

The research-base for the project has three distinct elements: an instructional design, a professional development design, and an evaluation design.

Instructional Design

Our first strategy for meeting the needs of at risk learners was to deliver instruction that is within the child's Zone of Proximal Development (Vygotsky, 1978). This Zone represents the difference between what a child can do with help from an adult or peer and what he or she can do without such guidance. It is knowledge that the child has the ability to learn but does not yet understand--knowledge that is just out of reach. We provided teachers with the diagnostic tests of PRIME (*Professional Resources and Instruction for Mathematics Educators)* to identify the Zone of Proximal Development for particular at risk learners. These tests place students on a developmental continuum for mathematics that has been empirically validated using samples of Canadian students, including students from the Kawartha Pine Ridge DSB (McDougall, Ben Jaafar, & Ross, in press). For each level of the continuum, PRIME provides developmentally appropriate teacher strategies. In addition to the diagnostic tests, PRIME provides a conception of the curriculum organized around a relatively small set of concepts and skills (e.g., for the Numbers and Operation strand there are five organizing concepts and three organizing skills). For each curriculum outcome there are five phases of performance. The practical benefit for the teacher is that the same activity can be addressed by learners at multiple levels, providing differentiated, developmentally appropriate instruction within the same curriculum context. PRIME links curriculum, assessment, and instruction, thereby meeting the four criteria for an effective feedback system: (a) data on the actual level of a measurable attribute, (b) the reference level of the attribute, (c) mechanisms for comparing the actual performance to a meaningful scale and generating information about the nature of the gap and (d) specific strategies for altering the gap (Black & William, 1998).

The second strategy for meeting the needs of at risk learners is recommended in the Expert Panel Report, *Learning Math Success: Mathematical Literacy Grades 7-12.* We will organize the PD sessions around the theme of effective use of manipulatives. Manipulatives can provide the additional scaffolding that at risk learners need because it enables students and teachers to be explicit about the mathematical thinking that characterizes deep understanding of mathematical concepts. Ben-Chaim, Fey, Fitzgerald, Benedetto, and Miller (1998) found that wise use of manipulatives contributed to higher mathematics achievement of Intermediate Division students. However, teacher use of manipulatives declines after the Primary years and simply providing grade 7 and 8 teachers with manipulatives is unlikely to lead to their use. The decline in implementation of manipulatives can be reversed through in-service on strategies for using manipulatives, although most of the research (e.g., Moyer & Jones, 2004) has been conducted with teachers of younger children.

Professional Development Design

The in-service in the project embodied features identified in Hill's (2004) review of effective mathematics PD: active learning by teachers (we engaged teachers in student tasks); examples from classroom practice (all tasks were drawn from the curriculum they were teaching); collaborative activities (teachers worked in groups of four at the sessions and in pairs between-sessions); modeling effective pedagogy (presenters demonstrated how to construct mathematical ideas using participant responses to student tasks and presenters gave particular attention to meeting the needs of at risk learners); opportunities for reflection, practice and feedback (teachers applied PD ideas in their own classrooms and brought student work to facilitate teacher reflection on implementation outcomes); focus on mathematics content (explicit attention was given to the mathematical concepts embodied in each task and to alternative strategies for eliciting these concepts).

In 2003-04 staff from Kawartha Pine Ridge DSB bundled these PD features into an in-service that was delivered to all grade 6 teachers in the district. A central feature of the in-service was the use of a teacher profile consisting of ten dimensions of mathematics that provide a continuum of teacher beliefs and practice. The teacher profile describes four levels ranging from direct instruction to constructivism. These levels are based on observations and interviews with teachers who ranged from traditional to high fidelity implementers of standards-based mathematics teaching (Ross, Hogaboam-Gray, McDougall, & Le Sage, 2003; Ross, Hogaboam-Gray, McDougall, & Bruce, 2001; Ross & McDougall, 2003). The dimensions guided the in-service design, the evaluation design, and provided teachers with self-assessment information.

The grade 6 in-service had a positive effect on teachers' beliefs about their ability to teach mathematics and it contributed to higher student achievement on EQAO assessments (Ross, 2004). In the current project we extended the grade 6 PD program to grade 7 and 8 teachers who had not participated in voluntary in-service in recent years. The key features extracted from Hill's (2004) review were retained along with a special emphasis on meeting the needs of at risk learners outlined in our instructional design. We continued using the ten dimensions as organizers, particularly since these are central to the implementation support materials of PRIME (McDougall, 2004).

The specific content of the in-service was developed by six teachers working as pairs, and supported by the district consultant for Intermediate Mathematics (John Ford). Each pair consisted of an experienced grade 7/8 teacher and the special education teacher from the same school. In June-August 2005 this development group participated in PRIME workshops, developing a firm understanding of the developmental continuum and how to apply it to the textbooks in use in Kawartha Pine Ridge.

The six teachers on the development team developed and simultaneously implemented the core instructional activities of the project, prior to providing in-service on these activities to teachers in the sample. The development teachers administered diagnostic tests to all students in their classroom. The focus of their instructional efforts was the five most at-risk students in their class. In pairs, the development team selected specific lessons that engaged at-risk learners in solving rich problems appropriate to the grade 7 and 8 curriculum. These lessons explicitly identified additional support for the at-risk students, using information from the continuum to provide sufficient scaffolding for them to be successful. These teachers used the mathematics resources that were currently available to them because the PRIME continuum and associated instructional strategies is not limited to the Nelson series.

The pilot experiences gave the development team insights that enabled them to focus in-service discussions around particular student responses, especially those of at risk students, to the instructional strategies. Implementation in the developers' classroom also enabled them to identify students' mathematical misconceptions.

After teachers were randomly assigned to treatment and control conditions, we provided nine in-service sessions consisting of three full days and six half-day workshops from September 2005 to February 2006. The treatment group was divided into three subgroups for each of these events in order to reduce teacher travel time and to facilitate teacher learning. Each workshop was designed and delivered by the development team under the guidance of the Mathematics consultant and PRISM Project Steering Committee. The workshops integrated learning about the PRIME continuum, moderated marking of the PRIME diagnostic tests, pedagogical strategies for teaching mathematics, and techniques for supporting at-risk learners (especially the use of manipulatives). The specific lessons that were the focus of the development team's piloting of PRISM strategies were introduced to the treatment sample as exemplars of high quality mathematics teaching. Between in-service sessions teachers were encouraged to try these particular lessons in their classrooms and use them as the foundation for developing their own strategies which they will implement in their own classrooms between the PD sessions.

An example of a workshop is summarized below:

- Discourse about the theories behind manipulative use and cooperative learning strategies. Strategies for constructing cooperative groups that encourage equal participation by at risk learners in mixed ability groups will also be discussed.
- Teachers experienced a lesson set that allows students to investigate combinations of fractions using manipulatives, add fractions by connecting concrete with symbolic representations, and recognize the need for equivalent fractions with denominators when adding fractions with different denominators. The lesson incorporated the use of pattern blocks.
- Teachers participated in small group discussions about the strengths and challenges associated with the lessons' use of manipulative and cooperative learning strategies. Teachers also examined common misunderstanding and misconceptions related to the topic as well as suggestions for scaffolding the lesson for struggling students. The cooperative learning strategy think/pair/share was profiled.
- Teachers experienced a second lesson where the focus was multiplying mixed fractions. Teachers solved a problem in small groups using the placemat cooperative learning strategy. A variety of manipulatives were available, including fraction circles.
- Teachers participated in small group discussions about the strengths and challenges associated with the lessons use of manipulatives. Teachers examined common misunderstandings and misconceptions related to the topic and reviewed suggestions for additional scaffolding required by at risk learners.

- Teachers were asked to try the two lessons profiled in the workshop with their students over the two weeks between sessions. They were asked to keep a reflective journal on their learning and the learning of their students when using the manipulatives. Teachers were asked to track the learning of one at risk student in their classroom throughout the project. Teachers shared experiences at the next workshop.
- Teachers were given class sets of pattern blocks and fraction circles for use in their classrooms.

Subsequent sessions maintained the key themes: a focus on student thinking, rich talk about mathematical ideas, use of rich tasks that challenge students, and the role of scaffolding in learning, especially the learning of at-risk students.

Research Question

Did participation in PRISM in-service contribute to improved student attitudes, higher student achievement, and/or improved teacher beliefs and practices?

Evaluation Design

The evaluation design for the proposed activities followed best practices in program evaluation. We used a delayed treatment field test. Grade 7 and 8 teachers who had not participated in voluntary mathematics in-service in the past two years were randomly assigned to two conditions: (i) a treatment group who participated in the in-service, September 2005-February 2006 and (ii) a control group who did not participate in the in-service until February-June 2006. Randomized field trials provide a powerful defense against almost all of the threats to internal validity of quantitative evaluations (Posavac & Carey, 2003). In addition, the use of a delayed treatment design ensures that all teachers and students in the district receive the in-service, thereby avoiding the “denial of treatment to those who might benefit from it” problem of experimental designs.

Sample

The district mathematics consultant identified teachers who had not attended the TIPS program or any other district level mathematics in-service in 2003-04 or 2004-05. We compiled the mean school scores on the grade 6 EQAO mathematics assessment for a three year period, 2001-2004 for the 45 elementary schools with grade 6 and grade 7/8 students. We rank ordered these schools from lowest to highest, dividing them into matching pairs, and randomly assigned each school within the pair to treatment or control group. The prior achievement of the schools in the two conditions was virtually identical: 2.22 for the treatment and 2.28 for the control schools (on the provincial assessment scale that runs from less than level one to level four). We then randomly assigned seven schools that did not have grade 6 students to treatment and control conditions, produced a sample of 26 treatment (23 K-8 + 3 senior) and 26 control (22 K-8 + 4 senior) schools. The intended sample consisted of 99 grade 7/8 teachers. The achieved sample on the pretest was 42 treatment and 44 control group teachers. There was further attrition for the posttest, with four drop outs from each condition. The achieved sample on the posttest consisted of 38 treatment and 41 control teachers. We also collected data on the students in their classrooms. Achievement data was obtained on the six lowest achieving students in each class, as identified by the teachers **(**N=528; 251 treatment and 277 control). The other student measures were completed by all students in each class. After removing students who provided only a pretest or only a posttest, N=1770 (802 treatment and 968 control**).** We collected no data on the special education teachers, although 14 participated in the in-service.

Teacher Instruments

To measure PRISM's effects on teachers we administered teacher surveys at the beginning and end of the project. Teacher efficacy, the extent to which teachers believe they will be able to bring about student learning, consists of 12 items adapted for mathematics teaching from Tschannen-Moran and Wolfolk Hoy (2001): 4 items for efficacy for engagement, 4 items for efficacy for teaching strategies, and 4 items for efficacy for student management. Teacher efficacy is an important outcome of in-service because teachers' with positive beliefs about their instructional capacity are more likely to try out new teaching ideas, particularly techniques that are difficult, involve risks, and require control to be shared with students (Dutton, 1990; Czerniak & Schriver-Waldon, 1991; Moore, 1990; Riggs & Enochs, 1990; Ross, 1992; Shachar & Shmuelevitz, 1997), characteristics of instruction that match the Ontario mathematics curriculum.

Mathematics teaching practices was measured with 20 items aligned to the ten dimensions of standards-based mathematics teaching that characterize the Ontario curriculum. Evidence of the validity and reliability of this scale are provided in Ross et al. (2003). In previous research this scale has been internally consistent, a significant predictor of EQAO achievement in mathematics, and it correlates with observations of teaching. Items 6,11, 15, 16, 18, 19, and 20 were reverse coded prior to analysis.

Implementation Measures

Treatment teachers completed student learning logs on 20 occasions from November to February. Rowan, Camburn and Correnti (2004) found that 20-30 logs per teacher were required because of considerable within-teacher variation. These logs tracked the curriculum experience for a single student, two days per week for ten weeks (i.e., the same student was tracked over this time). We drew items from Rowan, Harrison, and Hayes (2004) who demonstrate that these logs correlate well with observations of mathematics classrooms. The logs asked teachers to identify

- The phase, concept, and skills of Numbers or Operations the student was working on that day.
- The representations used by the student such as numbers or symbols, concrete materials, real-life situations or word problems, tables or charts and whether the teacher made explicit links between these representations.
- Whether the target student worked on mathematics tasks likely to generate deep understanding. Response options included: listen to me present the definition for a term or steps in a procedure (negative); perform tasks requiring ideas or methods already introduced to the student (negative); assess a problem and choose a method to use from those already introduced to the student; explain an answer or a solution method for a particular problem; analyze similarities and differences among representations, solutions, or methods; prove that a solution is valid or that a method works for all similar cases.
- Additional probes of the balance between recall and constructivist tasks included options such as: orally answering recall questions (negative); working on textbook, worksheet, or board work exercises for practice or review (negative); working on problem(s) that have multiple answers or solution methods, or involve multiple steps; discussing ideas, problems, solutions, or methods in pairs or small groups; using flashcards, games, or computer activities to improve recall or skills (negative); writing extended explanations of mathematical ideas, solutions, or methods; working on an investigation, problem, or project over an extended period of time.

Teacher Satisfaction Measures

In addition to formal measures of program effects on teachers' beliefs and practices, we also collected less formal data on teacher satisfaction with the in-service. In the fall and in February treatment teachers responded to open-ended probes such as "how has PRISM impacted teaching and learning in your classroom". They also completed a13 Likert items described as a professional development reflection, e.g., "I will be able to apply learning from these sessions in my school."

Student Instruments

The second desirable feature of our evaluation design is that we examined student as well as teacher outcomes. More than 90% of 450 National Staff Development Council projects reviewed by Killion (1998) had no student achievement measure. We began with affective variables that consistently predict mathematics achievement. These were administered to all students in both experimental conditions. If the project is effective we would expect to see greater student gains on these affective measures in the treatment group than in the control group. The focus on affective outcomes mirrors the attention given to affect in the Expert Panel Report (Ministry of Education, 2004).

- *Math self-efficacy* consisted of six Likert items measuring expectations about future mathematics performance (from Ross, Hogaboam-Gray, & Rolheiser, 2002). Pajares (1996) demonstrated that self-efficacy is a better predictor of mathematics achievement than closely associated variables such as math anxiety and mathematical self concept.
- *Negative affect for failure* consisted of six items from Turner et al. (2003). This scale measures students' fear of failing, a powerful inhibitor of mathematical motivation.
- *Effort* was measured with eight items developed for our assessment of the grade 6 PD program in 2003-04.
- *Student beliefs about mathematics and mathematics learning* have been identified as predictors of student achievement (Muis, 2004). We administered 19 items selected from Schoenfeld's (1989) student beliefs survey, which has been used extensively in previous research. However, to date no attempt to convert Schoenfeld's items to scales has been reported.
- *Rank* in the class was determined by asking teachers to identify the ten lowest achieving students in their class using a scale in which 1 meant the lowest, 2 meant next lowest, etc.
- *Achievement of at risk students.* We administered the diagnostic tests of the PRIME program. (The manual for the diagnostic tests indicates that the same instrument can be used for pre-post comparisons, as long as the administrations are three months apart.) There were two tests: Number and Operations, each consisting of 14 items. Student responses were scored as incorrect (0), partially correct (1) or correct (2), following the scoring key in the manual. Teachers reported the ten lowest achieving students in their class in September; they also indicated whether these students had been formally identified. We selected the six lowest students for whom we had complete pretest data. Four markers were trained to 90% accuracy. The mathematics consultant independently marked a random sample of 14 papers from each marker (i.e., 56 students, each completing 20 items N=1120 decisions). There was 91% perfect agreement across markers (88-98% for each marker). The chance adjusted Kappa score was.87. Bakeman and Gottman (1997) described Kappa scores of .40-.60 as fair; .60-.75 as good and .75+ as excellent reliability.

Data Analysis

After establishing the reliability of the measures used in the study, we applied univariate and multivariate procedures of General Linear Modeling. The dependent measures for the teacher data were the posttest survey scale scores listed above; the covariates were the pretest scale scores on the same measures; and the independent variable was experimental condition (early or late treatment—hereafter treatment or control group). The same procedures were used for the student outcomes.

Results

Student Achievement Data

Psychometric Characteristics of the Achievement Measures

There were two student achievement instruments, administered at the beginning and end of the in-service to students in the treatment and control groups. The measures used in the study were each internally consistent; i.e., Cronbach's alpha was .70+. There were 14 items in each scale. The reliabilities were: pretest Operations (.76, N=513), pretest Numbers (.80, N=505), posttest Operations (.79, N=466), posttest Numbers (.81, N=457)[2]. In addition, each achievement variable was normally distributed (skewness ranged from -.502 to .345; kurtosis ranged from -.284 to -.891). We defined an outlier as more than three standard deviations beyond the mean. Only one score reached this level: we left it alone.

Sample Equivalence. Table 1 shows the means and standard deviations for each group on each test and on the rank order assigned to the student by the teacher. There were no significant differences on the pretest between students in the treatment and control conditions on either the Operations [t(526)=-.721 , p=.471] or Numbers [t(522)=1.642, p=.101] test.

Table 1
Means and Standard Deviations of Achievement Variables by Study Condition

Test	Study Condition	N	Mean	SD
OP pre	Control	277	.68	.40
	Treatment	251	.71	.42
NU pre	Control	276	.95	.48
	Treatment	248	.88	.43
OP post	Control	248	.90	.46
	Treatment	219	.93	.46
NU post	Control	248	1.02	.44
	Treatment	212	1.07	.46

OP=Operations test; NU=Numbers test

We used GLM to examine the student achievement impact of PRISM in-service. In the first analysis, the dependent variable was the posttest Operations score, the covariate was pretest Operations score, and the independent variable was experimental group (treatment or control). Table 2 shows there was a statistically significant main effect for the pretest but not for experimental condition. Students who had done well on the pretest tended to do well on the posttest, regardless of whether or not their teachers participated in the in-service. After adjusting for the effects of pretest performance, the means were slightly higher for students in the treatment group (.917) than for students in the control (.896). In fact the lower bound for the treatment group mean (set at the 95% confidence interval) exceeded the mean of the control group. However, Table 2 shows that the differences between the treatment and control groups were not so large that we could rule out chance as an explanation. PRISM had no statistically significant effect on achievement in Operations.

[2] The number of cases in the tables in this and subsequent sections of the Results varies because we used pairwise deletion prior to the imputation of missing values when determining scale reliability; there was attrition from pre- to posttest in both experimental conditions; and comparisons between conditions used merged databases that consisted of complete data sets only.

Table 2
Posttest Operations Score by Experimental Condition, Adjusting for Pretest Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	22.521(a)	2	11.261	72.738	.000	.245
Intercept	33.113	1	33.113	213.893	.000	.323
Operations pre	22.418	1	22.418	144.805	.000	.244
Group	.052	1	.052	.335	.563	.001
Error	69.356	448	.155			
Total	461.990	451				
Corrected Total	91.877	450				

a R Squared = .245 (Adjusted R Squared = .242)

The Operations test consisted of items that were more algorithmic than conceptual. In the second analysis we look at the results for the Numbers test. These items were more conceptual than algorithmic. Since the in-service was more focused on the conceptual aspects of mathematics than on performance of routine algorithms, we anticipated that we would be more likely to find a student achievement effect for the in-service.

In the second analysis, the dependent variable was the posttest Numbers score, the covariate was pretest Numbers score, and the independent variable was experimental group (treatment or control) and rank in class. Table 3 shows that there was a statistically significant effect for both the pretest and experimental condition. As in the previous analysis, students who had done well on the pretest tended to do well on the posttest and conversely, those who had done poorly on the pretest continued to do poorly on the posttest. There was also a statistically significant effect for PRISM. Students who were in the treatment group outperformed students in the control group: the means, adjusted by pretest scores, were 1.081 for the treatment group and .985 for the control. As in the Operations test, the mean of students in the treatment group exceeded the lower bound for the mean of the control group but the differences were statistically significant for the Numbers test.

Table 3
Posttest Numbers Score by Experimental Condition, Adjusting for Pretest Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	20.875(a)	2	10.438	68.725	.000	.238
Intercept	28.535	1	28.535	187.880	.000	.300
Numbers pre	20.586	1	20.586	135.544	.000	.236
Group	1.009	1	1.009	6.641	.010	.015
Error	66.674	439	.152			
Total	555.524	442				
Corrected Total	87.549	441				

a R Squared = .238 (Adjusted R Squared = .235)

In the third analysis, we examined the effect of teacher assigned rank on student performance. Teachers gave the lowest achieving student a rank of 1 (i.e., most needy), the second lowest a rank of 2, and so on up to 10. We selected the first six students in each class (i.e, ranks 1-6), although in some instances (where data were missing) we selected the next ranked student. We anticipated that since PRISM was designed to address the needs of the least able student, that the lower ranked students would benefit more from their teachers being in the in-service program than students who were ranked by their teachers as being less needy.

To test this possibility, we repeated the GLM analysis, adding student assigned rank as a second independent variable. We began with: the dependent variable was the posttest Operations score, the covariate was pretest Operations score, and the independent variables were experimental group (treatment or control) and the rank (from 1-10 assigned to the student by the teacher). We then repeated the analysis, replacing the Operations scores with the Numbers score. After some exploration of the data we found that student rank was most visible when we collapsed the scale into lowest student and all other students nominated by teachers as low achievers. The results are shown in Tables 4 (Operations) and 5 (Numbers).

Table 4
Posttest Operations by Experimental Condition and Student Rank, Adjusting for Pretest Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	25.264(a)	4	6.316	42.288	.000	.275
Intercept	22.409	1	22.409	150.034	.000	.252
Operations pre	19.112	1	19.112	127.965	.000	.223
Group	.274	1	.274	1.834	.176	.004
newrank	2.282	1	2.282	15.277	.000	.033
Group * newrank	.265	1	.265	1.775	.183	.004
Error	66.613	446	.149			
Total	461.990	451				
Corrected Total	91.877	450				

a R Squared = .275 (Adjusted R Squared = .268)

Table 5
Posttest Numbers by Experimental Condition and Student Rank, Adjusting for Pretest Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	22.917(a)	4	5.729	38.738	.000	.262
Intercept	21.957	1	21.957	148.457	.000	.254
Numbers pre	17.724	1	17.724	119.840	.000	.215
Group	.885	1	.885	5.986	.015	.014
newrank	1.735	1	1.735	11.729	.001	.026
Group * newrank	.157	1	.157	1.062	.303	.002

Error	64.632	437	.148
Total	555.524	442	
Corrected Total	87.549	441	

a R Squared = .262 (Adjusted R Squared = .255)

Tables 4 and 5 show that when student rank was included in the equation, scores on the posttests were significantly influenced by pretest scores and by teacher assigned rank. Students who were nominated as the least able student in the class scored significantly lower on the posttest than other students identified by teachers as needy. The effect was statistically significant even when student pretest scores were taken into account; i.e., teachers included in their judgments student ability to learn characteristics that were not visible in the pretest results.

The addition of teacher assigned rank in the class did not change the treatment effects: PRISM continued to have a significant effect on student achievement in Numbers but the differences were not significant on Operations. In addition the treatment * student rank interaction was not statistically significant. Table 6 summarizes the interaction. The table shows that the adjusted posttest means (i.e., after adjustment by pretest scores) are higher in the treatment group than in the control for students given the lowest rank in the class by their teachers. The advantage of the treatment group over the control is much smaller for the other students nominated by teachers as low achievers. Although the effects are consistent across achievement tests they are too small to be statistically significant. However, for the Numbers posttest, the lower bound (representing the 95% confidence interval) of the treatment group was higher than the control mean for the lowest ranked students (.811 versus .772); this was also the case for the other ranks (i.e., lower bound of treatment=1.040 versus the control mean of 1.019).

Summary of Student Achievement Effects

Students in classrooms of teachers who participated in the in-service learned more than students in the control group. There was a small but statistically significant program effect on the Numbers placement test. This test measured conceptually oriented items. The gains were achieved with no loss of performance of routine algorithms, the traditional focus of instruction for at-risk students, represented in this study by the Operations test. The effects of the program were robust; i.e., even students identified as least able in the classroom benefited from the treatment.

Table 6
Adjusted Posttest Scores for Lowest Ranked Students by Experimental Condition

	Operations Posttest		**Numbers Posttest**	
Group	**Lowest Rank**	**Other Ranks**	**Lowest Rank**	**Other Ranks**
Treatment	.963	1.097	.934	.935
Control	.772	1.019	.644	.934

Student Survey Data

Psychometric Properties of the Student Surveys.

There were four student instruments, administered at the beginning and end of the in-service to students in the treatment and control groups. An exploratory factor analysis of the student responses produced similar results in both the pre- and posttest data. The items for Math self-efficacy, negative affect for failure, and math effort loaded on separate scales. The 19 items selected from Schoenfeld (1989) did not produce interpretable scales on either the pretest or the posttest. We tried several iterations of principal axis with varimax and oblim rotation. We then conducted a reliability analysis of the 19 items. When we removed five items (#4, 7, 12, 14, 16) we produced an acceptable scale. We labeled the 14 item scale, *dysfunctional beliefs about mathematics learning*. The items that were removed consisted of items that expressed functional beliefs (e.g., “when the teacher asks a question in class there are lots of possible right answers you could give”). However, when the five deleted items were examined they did not comprise a coherent scale so we decided to treat this set as five separate measures.

Table 7 shows that each instrument was internally consistent on both administrations: all reached the .70 + criterion for Cronbach's alpha. We identified a few outliers (16 scores across all variables in 1770 cases). Each outlier was corrected by reducing the score to +/- three standard deviations beyond the mean. Table 8 displays the descriptive statistics for the student instruments. The table shows that each of the scales were normally distributed on both pre- and posttest (i.e., skewness was less than 3.0 and kurtosis was less than 10.0). There were relatively few missing values--these were replaced using the expectation maximization procedure of Missing Values Analysis.

Table 7
Means, Standard Deviations, and Cronbach’s Alpha for Student Attitude Measures, by Test Occasion

Scales	N of Students	N of Items	Mean	SD	Alpha	Comment
Math Self-efficacy (pre)	1770	6	4.31	0.94	.89	No adjustments
Math Self-efficacy (post)	1770	6	4.29	0.96	.89	No adjustments
Dysfunctional Math Beliefs (pre)	1770	14	3.48	0.65	.73	Item Q4, Q7, Q12, Q14, Q16, removed
Dysfunctional Math Beliefs (post)	1770	14	3.38	0.68	.77	Item Q4, Q7, Q12, Q14, Q16, removed
Negative affect for failure (pre)	1770	6	3.17	1.10	.81	No adjustments
Negative affect for failure (post)	1770	6	3.14	1.14	.83	No adjustments
Math Effort (pre)	1770	4	4.34	0.91	.88	No adjustments
Math Effort (post)	1770	4	4.08	0.94	.89	No adjustments

Table 8
Means, Standard Deviations, Skewness and Kurtosis, by Test Occasion

	N	Mean	SD	Skewness	Kurtosis
Math Self-efficacy (pre)	1770	4.31	0.94	-.51	-.10
Math Self-efficacy (post)	1770	4.29	0.96	-.58	-.04
Dysfunctional Math Beliefs (pre)	1770	3.48	0.65	.09	-.09
Dysfunctional Math Beliefs (post)	1770	3.38	0.68	.20	.078
Negative affect for failure (pre)	1770	3.17	1.10	.27	-.30
Negative affect for failure (post)	1770	3.14	1.14	.27	-.42
Math Effort (pre)	1770	4.34	0.91	-.56	.42
Math Effort (post)	1770	4.08	0.94	-.44	.06

Sample Equivalence.

Table 9 shows the means for the students in each experimental condition. There were no significant treatment-control differences on three of the four measures: self-efficacy, dysfunctional math beliefs, and negative affect for failure. On the fourth measure (Math Effort), students in the treatment group reported significantly higher effort on the pretest than students in the control group. This result was unexpected: with random assignment of 52 schools to experimental conditions we expected to find no pretest differences on any student characteristic. Consequently, it was essential that we control for pretest scores when calculating the effects of the in-service on students' self-reported effort.

Table 9
Means and Standard Deviations by Experimental Condition, by Test Occasion

	Group	N	Mean	SD	t-test results
Math Self-efficacy (pre)	treatment	802	4.26	0.94	T(1768)= -1.649, *p*=.099
	control	968	4.34	0.95	
Math Self-efficacy (post)	treatment	802	4.29	0.95	
	control	968	4.29	0.97	
Dysfunctional Math Beliefs (pre)	treatment	802	3.49	0.63	T(1768)= 0.235, *p*=.814
	control	968	3.48	0.66	
Dysfunctional Math Beliefs (post)	treatment	802	3.36	0.68	
	control	968	3.40	0.68	
Negative affect for failure (pre)	treatment	802	3.15	1.09	T(1768)= -0.620, *p*=.536
	control	968	3.18	1.11	
Negative affect for failure (post)	treatment	802	3.08	1.12	
	control	968	3.19	1.15	

Math Effort (pre)	treatment	802	4.40	0.90	T(1768)= 2.489, p=.013
	control	968	4.29	0.91	
Math Effort (post)	treatment	802	4.11	0.94	
	control	968	4.05	0.95	

Effects of the In-service on Students.

The correlations between the student measures were not substantial, except that student self-efficacy was moderately correlated with math effort. Consequently, we examined the effects of the in-service separately for each measure.

We conducted a series of univariate GLM analyses in which the dependent variables were, alternately, students' self-efficacy, dysfunctional math beliefs, negative affect for failure, and student effort. In each analysis, the pretest score was the covariate and the independent variable was experimental condition (treatment or control group).

In every analysis, students' posttest scores were significantly influenced by their pretest scores. The pretest accounted for from 22.3% to 42.2% of the variance in the posttest scores. Table 10 shows the adjusted posttest scores for each measure (i.e., the posttest score after regressing over pretest scores) and the effect of experimental condition on the student attitude measures. Treatment group students scored slightly better than control group student scores on all of the measures: The treatment group was higher than control group students on self-efficacy and self-reported student effort. In addition, the treatment group was lower than the control on support for dysfunctional math beliefs and negative affect for failure. However, only one of the comparisons was statistically significant. Students in the treatment group report lower negative affect for failure than students in the control group. The size of the effect was very small (eta squared < .01).

Table 10
Effects of In-service on Student Survey: Adjusted Means and GLM Results

	Adjusted Posttest Means		
Posttest variable	**Treatment**	**Control**	**Univariate GLM Results**
Math Self-efficacy	4.313	4.260	F(1,1767)=1.946, p=.163
Dysfunctional Math Beliefs	3.356	3.405	F(1,1767)=2.933, p=.087
Negative affect for failure	3.092	3.182	F(1,1767)=3.939, p=.047
Math Effort	4.081	4.080	F(1,1767)=0.000, p=.987

We conducted an exploratory investigation of the student beliefs measures. In this series of univariate analyses, the dependent variable was the posttest belief, the covariate was the pretest score on the same belief, and the independent variable was experimental condition. We found significant differences on five of the 19 comparisons. Table 10a shows that four were in the expected direction. Q10, Q11, and Q15 are dysfunctional beliefs that impede mathematics learning. On these measures, students in classrooms taught by teachers who attended the in-service were *less likely* to believe that when they don’t know the answer to a question posed by the teacher that another student will be asked (Q10) or that the teacher will answer the question (Q11). These results suggest that treatment students believe that not knowing the immediate answer does not absolve one from continuing to think about the problem. Treatment students were also *less likely* to believe that math problems can be solved in only one way, suggesting that there had been progress toward richer problems in the classrooms of treatment teachers. However, contrary to our expectation, treatment students were *more likely* to indicate that when asked a question a student must remember the right answer, a result that suggests that even after the teacher in-service, students expected recall items to dominate. The final significant item indicated improvement in treatment students' beliefs about mathematics: they were *more likely* to view the subject as one in which you can be creative and discover things by yourself. These individual item differences provide further support for the argument that students' beliefs about mathematics and mathematics learning improved during the teacher in-service. However, we need to treat these results with caution because making multiple comparisons within a set of items inflates Type I error.

Table 10a
Effects of In-service on Selected Student Belief Items: Adjusted Means and GLM Results

	Adjusted Means		
Posttest variable	**Treatment**	**Control**	**Univariate GLM Results**
Q6: When the teacher asks a question in math class . . . You have to remember the right answer to answer it correctly	3.466	3.319	F(1,1767)=4.553, *p*=.033
Q10: When the math teacher asks me a question that I can't answer right away . . . Someone else will be asked	3.314	3.495	F(1,1767)=7.435, *p*=.006
Q11: When the math teacher asks me a question that I can't answer right away . . .The teacher will answer the question	2.561	2.681	F(1,1767)=4.450, *p*=.035
Q14: In mathematics you can be creative and discover things by yourself	4.364	4.247	F(1,1767)=4.042, *p*=.045
Q15: Math problems can be done correctly in only one way	2.199	2.325	F(1,1767)=4.589, *p*=.032

Summary of Student Attitude Effects
Students in the treatment group scored better than students in the control group on all measures. However, the differences were statistically significant only for negative affect for failure and for a few of the individual belief items. Students in the treatment group were less fearful of failing in mathematics than students in the control group, after adjustment for pretest scores on the same measure. In addition treatment students were more likely to espouse positive beliefs about mathematics and less likely to support dysfunctional beliefs that impede mathematics learning.

Teacher Data

Psychometric Properties of the Teacher Surveys.
There were four teacher instruments, administered at the beginning and end of the in-service. Table 11 shows that each instrument was internally consistent on both administrations: all reached the .70 + criterion for Cronbach's alpha. One item was removed from the efficacy for student engagement scale on both administrations.

Table 11.
Means, Standard Deviations, and Cronbach's Alpha of Teacher Instruments

Scales	N	Items	Mean	SD	Alpha	Comment
Math Practices (pre)	80	20	4.44	.43	.76	No adjustments
Math Practices (post)	77	20	4.50	.51	.81	No adjustments
Efficacy in Student Engagement (pre)	87	3	3.71	.60	.82	Item Q31 removed (alpha=.74)
Efficacy in Student Engagement (post)	79	4	3.78	.67	.85	Item Q31 removed (alpha=.82)
Efficacy in Instructional Strategies (pre)	86	4	3.70	.55	.74	No adjustments
Efficacy in Instructional Strategies (post)	79	4	3.80	.66	.83	No adjustments
Efficacy in Classroom Management (pre)	86	4	4.15	.55	.82	No adjustments
Efficacy in Classroom Management (post)	78	4	4.16	.60	.84	No adjustments

A few outliers (11 in 688 cases) were identified and corrected (by reducing the score to +/- three standard deviations beyond the mean). Table 12 displays the descriptive statistics for the teacher instruments. The table shows that each of the scales were normally distributed on both pre- and posttests (i.e., skewness was less than 3.0 and kurtosis was less than 10.0).

Table 12
Means, Standard Deviations, Skewness and Kurtosis of Teacher Variables

	N	Mean	SD	Skewness	Kurtosis
Math teaching pre	87	4.4428	.42950	.257	-.328
Math teaching post	79	4.5015	.50931	-1.683	9.201
TE engagement pre	87	3.7050	.60452	.232	-.406
TE engagement post	79	3.7764	.67034	-.204	.078
TE instructional strategies pre	86	3.6948	.54960	-.188	-.292
TE instructional strategies post	79	3.8038	.66104	-.879	1.313
TE student management pre	87	4.1494	.55308	-.503	.519
TE student management post	79	4.1582	.59804	-.832	1.397

Sample Equivalence. Table 13 also shows the means for the teachers in each experimental condition. The table shows that on every pretest measure, teachers in the treatment group scored higher than teachers in the control group. On two of these measures (teacher efficacy for instructional strategies and teacher efficacy for student management) the differences were large enough to be statistically significant. This result was unexpected: with random assignment of 52 schools to conditions we expected to find no pretest differences on any teacher characteristic.

Consequently, it was essential that we control for pretest scores when calculating the effects of the in-service on teachers.

Table 13
Means and Standard Deviations by Experimental Condition

	group	N	Mean	SD	t-test results
Math teaching pre	treatment	43	4.48	.44	T(85)=.801, p=.426
	control	44	4.41	.42	
Math teaching post	treatment	36	4.59	.47	
	control	43	4.46	.41	
TE engagement pre	treatment	43	3.80	.59	T(85)=1.434, p=.155
	control	44	3.61	.61	
TE engagement post	treatment	36	3.87	.67	
	control	43	3.70	.67	
TE Instruction pre	treatment	43	3.82	.50	T(84)=2.154, p=.034
	control	43	3.57	.58	
TE Instruction post	treatment	36	3.97	.42	
	control	43	3.68	.77	
TE management pre	treatment	43	4.28	.49	T(85)=2.315, p=.011
	control	44	4.02	.58	
TE management post	treatment	36	4.34	.56	
	control	43	4.01	.58	

Effects of the In-service on Teachers

We conducted a univariate GLM analysis in which the dependent variable was teachers' self-reported beliefs and practices in teaching mathematics. A high score on this measure shows greater fidelity to the ideals of standards-based mathematics teaching, while a low score shows greater commitment to traditional approaches to teaching mathematics. Teachers' posttest scores were significantly influenced by their pretest scores [F(1, 76)=44.082, p<.001]. The pretest accounted for virtually all of the variance explained (i.e., .367 of .381). Treatment group teacher scores were slightly higher than control group teacher scores after adjusting for pretest effects [treatment=4.527 versus control=4.480] but the differences were not statistically significant [F(1,76)=.911, p=.343].

The three teacher efficacy measures were highly inter-correlated. We conducted a multivariate GLM in which the dependent variables were the three teacher efficacy posttest scores, the covariates were the three teacher efficacy pretest scores, and the independent variable was treatment condition. However, Box’s M test [19.340, p=.005] indicated that the covariances of the posttests were not the same for the two teacher groups, suggesting that the multivariate test was invalid. Consequently, we examined the effects of the in-service separately for each measure.

Table 14 shows that the posttest scores were adjusted by pretest scores, all three of the teacher efficacy variables were higher in the treatment than the control group teachers. However, these differences were so small that we could not rule out chance as an explanation for them. Table 14 shows that the in-service program had no significant effect on teacher beliefs about their ability to implement standards-based mathematics teaching.

Table 14
Effects of In-service on Teacher Efficacy: Adjusted Means and GLM Result

Posttest variable	Adjusted Means: Treatment	Adjusted Means: Control	Univariate GLM Results
TE for engagement	3.802	3.754	F(1,76)=.131, p=.719
TE for instruction	3.847	3.774	F(1,76)=.374, p=.543
TE for student management	4.264	4.070	F(1,76)=.2.734, p=.102

In addition to these formal measures completed by both treatment and control group teachers, there were also less formal teacher satisfaction instruments completed by treatment teachers only. Eighty percent of the teachers gave high ratings to the in-service (i.e., they agreed or strongly agreed to such statements as “The sessions were well-paced within the allotted time.” and “Student achievement will improve as a result of my implementation of what I have learned.” When asked if the in-service affected their classroom teaching, most teachers indicated that it had positive effects for teachers and for students. For example:

- “It has given me confidence, enthusiasm and the necessary teaching methods/tools to provide an engaging program for my students.”
- “I am incorporating manipulatives into my lessons on a regular basis. The students are enjoying Math quite a bit more now and are better able to visualize concepts.”
- “I am able to determine where a student is with respect to their knowledge of concepts and skills. I am able to help that student move on by using different strategies/manipulatives. Students will have the opportunity to explore and discover concepts rather than being shown.”
- “It’s made math less frightening for me.”

Summary of Effects on Teachers Beliefs and Practices

Teachers who participated in the in-service expressed satisfaction with the professional development it provided and reported multiple ways in which it contributed to enhanced teaching, greater teacher confidence, and other benefits for teachers and students. The formal data collection instruments indicated that after pretest performance was taken into account, treatment teachers had higher posttest scores on the teacher beliefs and self-reported practices measure and on the three teacher efficacy scales. However, none of these treatment-control differences was large enough to be statistically significant.

Discussion

Student Achievement

At the beginning of this report we noted that the vast majority of studies of teacher professional development make no attempt to measure student achievement (Killion, 1998; Knapp, 2003). The most important finding of our study is that participation in PRISM in-service made a small but statistically significant contribution to student achievement. Treatment students had higher scores on the Numbers test, which measures conceptual understanding, than control group students. This higher conceptual achievement was attained without loss of computational accuracy, as measured by the Operations test. The key questions with regard to this finding are (1) can we believe it? (i.e., what is the internal validity of the study?), and (2) can teachers in other settings expect to get similar results? (i.e., what is the external validity of the study?)

Internal Validity

Refers to the credibility of the claim that the program had an impact on the specific students and teachers who participated in the study. The gold standard for establishing causality is the true experiment. The logic of experimental designs is that if two groups are equivalent at the beginning of the study and only one group receives the intervention, any subsequent changes between the groups can be attributed to that intervention. The status of the true experiment in educational research is shown in the standards of the What Works Clearinghouse in the United States. This Clearinghouse stores studies that claim effects for educational programs. Their goal is to identify programs that have produced positive effects across multiple settings. The Study Review Standards of the Clearinghouse (n.d.) identifies the randomized field trial, the design we used in our PRISM study, as the optimal method. It provides strong protection against the principal threats to internal validity (Posavac & Carey, 2003). For example, selection is unlikely to be a threat in a randomized field trial because any factors likely to influence the outcome will be found in both experimental conditions, provided the number of cases is reasonably large. In our study we implemented a true experimental design by randomly assigned 52 schools to treatment and control groups.

We also guarded against the unintended consequences of random assignment that diminish the value of true experiments. We chose schools (rather than students or teachers) as our unit of assignment to minimize the disruption of the study and reduce leakage of PRISM activities from treatment to control group teachers. Since we had fewer schools than teachers, we used a stratified random assignment strategy in which we rank ordered K-8 schools on the basis of their scores on EQAO mathematics assessments over the three years prior to the study. We grouped these schools into pairs (beginning at the high and low ends so that the last school assigned, an unpaired orphan, would be at the mid-point of the rank order). We randomly assigned each school within the pair to treatment or control, resulting in two groups of schools that had virtually identical records of prior achievement. We assigned seven grade 7 and 8 schools, for which there was no prior achievement data, randomly to the two conditions. The result was that on the pretest, there were no significant differences on student achievement between the groups.

A major concern about true experiments in education is ethics: how can we deny programs that we believe to be worthwhile to children who might benefit from them? Our solution was the delayed treatment design (also called a crossover design); that is, at the end of the study teachers in the control group received exactly the same PRISM training as the treatment teachers received during the study. Providing both groups with the program protects the study from such threats to validity as compensatory rivalry (i.e., control group members try harder to show they are as good as the treatment group) and resentful demoralization (i.e., control group members feel so deprived they reduce their effort) (Woortman, 1994).

The credibility of the study was also strengthened by the use of standardized measures (in the case of the student achievement tests) or measures that have been field tested in other research studies. Meta-analyses have consistently shown that effect sizes are higher for investigator-designed measures than for standard tools (Glass, McGaw, & Smith, 1981). In addition, our PRISM study included tests of instrument reliability: We found that markers of the achievement tests were highly consistent (based on Kappa, a chance-adjusted measure of inter-rater reliability). All other measures used in the study met the standard (Cronbach's alpha=.70 or more) for internal consistency.

Other methodological features used in our PRISM study were the use of implementation measures (although we have not reported the teacher log data), tests of the assumptions of statistical tests (e.g., we tested for homogeneity of variance when using GLM and established that our measures were normally distributed), identification and adjustment of outliers (which can distort statistical tests if uncorrected), blind marking (i.e., the markers did not know which students were from treatment and which from control schools), and adequate statistical power.

External Validity

Refers to the likelihood that students in other jurisdictions will attain the same results as the students who participated in the study. It is useful to distinguish between efficacy and effectiveness studies. Efficacy studies (sometimes called existence proofs) examine a program's effects in optimal conditions, for example, with eager, volunteer teachers with high interest and skill in mathematics teaching. The logic is that if the program does not work in the best conditions it will not work anywhere. In contrast, effectiveness studies (sometimes called ecologically valid experiments) test a program's effects in typical situations, for example, with teachers ranging from the eager to the conscripted, from the expert to the novice, from those serving advantaged student populations to those serving at-risk students, and so on.

Our study was very much of the second type: the teachers chosen to participate consisted of those who had not volunteered for previous mathematics initiatives. These teachers were not volunteers. The invitations they received were gracious and encouraging but opting out was not an option. Release time was provided for all sessions to over-ride the contractual right of teachers to opt out of after-school in-service. There was some attrition of data (as much in the control as in the treatment sample) but attendance at the in-service sessions was consistently high. Virtually all the teachers were generalists, not mathematics specialists. Given the characteristics of the teachers in the study, it is reasonable to anticipate that teachers in other jurisdictions would do as well.

The second important finding is that the effects were robust across teacher-assigned student ranks. Even the students ranked lowest in the class benefited from the treatment (although the scores of the lowest ranked student were below the scores of higher ranked students in the same experimental condition). Previous research has produced mixed results for the effects of mathematics reform on low ability students. Mayer (1998) found that low ability students did not benefit from the Standards-based instruction but high ability students did. Fuchs et al. (1997) found benefits for mathematics program that included some elements of reform for low achieving students but not for those identified as Learning Disability. In a review of 15 studies on teaching mathematics to low achieving students, Baker, Gersten, and Lee (2002) found that one study produced ES=.48 on tests of transfer of conceptual understanding but in this successful study students were taught basic skills before engaging in complex math (contrary to reform ideals). In Baker et al., one study showed no effects and the other two studies showed negative effects for teaching Standards-based mathematics to low achieving students. Our finding of positive results for low achieving students, including the lowest achievers in the class, is an important addition to this research.

The third important finding concerned the contribution of the in-service to student beliefs and attitudes. Students in classrooms taught by teachers who participated in the in-service were less likely than control group students to fear failure, were less likely to support dysfunctional beliefs that impede mathematics learning, and were more likely to support positive beliefs about mathematics.

The fourth important finding was that although teachers expressed high levels of satisfaction with the in-service, there were no significant differences between teachers in the treatment and control groups on any of the standard teacher measures. One interpretation might be that the instruments were too weak to detect in-service effects. However, in previous research we found a treatment effect for grade 6 teacher mathematics in-service for teacher efficacy (Ross & Bruce, 2005) and for teacher self-reported beliefs and practices (Ross et al., 2003). A second interpretation might be that changes in teachers follow changes in students, i.e., that teachers try out new instructional ideas but do not incorporate them into their teaching make-up until they have evidence of positive student effects (this view of teacher change is elaborated in Ross & Regan, 1993). From this perspective we anticipate that changes in teacher beliefs and attitudes will develop as teachers assimilate the student achievement results.

Conclusion

The results of the PRISM study provide evidence that systematic in-service can contribute to improved student achievement, even when the focus is on students' of low ability and the teachers involved in the study are generalists with little prior experience in mathematics in-service. The rigor of the research design provides the foundation for our expectation that other jurisdictions that adopt PRISM in-service will attain similar benefits.

References

Baker, S., Gersten, R., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal*, *103*(1), 51-74.

Ben-Chaim, D., Fey, J., Fitzgerald, W., Benedetto, C., & Miller, J. (1998). Proportional reasoning among 7th grade students with different curricular experiences. *Educational Studies in Mathematics, 36*(3), 247-273.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, Principles, Policies & Practices, 5*(1), 7-74.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29*(1), 41-62.

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.

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11*(1), 81-95.

Cohen, E., & Lotan, R. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal, 32*(1), 99-120.

Czerniak, C., & Schriver-Waldon, M. (1991, April). *A study of science teaching efficacy using qualitative and qauntitative research methods.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Lake Geneva, WI.

Dutton, M. (1990). *An investigation of the relationship between training in cooperative learning and teacher job satisfaction*. Unpublished doctoral dissertation, Portland State University, Portland, OR.

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.

Fuchs, L., Fuchs, D., Hamlett, C., Phillips, N., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *Elementary School Journal, 97*(3), 223-250.

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

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1-29.

Hill, H. C. (2004). Professional development standards and practices in elementary school mathematics. *Elementary School Journal, 104*(3), 215-231.

Hoek, D., van den Eeden, P., & Terwel, J. (1999). The effects of integrated social and cognitive strategy instruction on the mathematics achievement in secondary school. *Learning and Instruction, 9*, 427-448.

Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of Standards-based mathematics education: A study of the Core-Plus Mathematics Project algebra and functions strand. *Journal of Research in Mathematics Education, 31*(3), 328-361.

Killion, J. (1998). Scaling the elusive summit. *Journal of Staff Development, 19*(4), 12-16.

King, L. (1993). High and low achievers' perceptions and cooperative learning in two small groups. *Elementary School Journal, 93*(4), 399-416.

Knapp, M. S. (2003). Professional development as a policy pathway. In R. Floden (Ed.), *Review of Research in Education* Vol. 27 (pp. 109-158). Washington: American Educational Research Association.

Kohler, F. W., Ezell, H. K., & Paluselli, M. (1999). Promoting changes in teachers' conduct of student pair activities: An examination of reciprocal peer coaching. *Journal of Special Education, 33*(3), 154-165, 88.

Licklider, B. L. (1995). The effects of peer coaching cycles on teacher use of a complex teaching skill and teachers' sense of efficacy. *Journal of Personnel Evaluation in Education, 9*(1), 55-68.

Mark, M. M., Feller, I., & Button, S. B. (1997). Integrating qualitative methods in a predominantly quantitative evaluation: A case study and some reflections. *New Directions for Evaluation, 74*, 47-59.

Mayer, D. P. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*, *20*(2), 53-73.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data. *Educational Evaluation and Policy Analysis, 21*(1), 29-45.

McDougall, D. (2004). *PRIME school mathematics improvement: Leadership handbook*. Toronto: Nelson.

McDougall, D., Ben Jaafar, S., & Ross, J. A. (in press). *PRIME research study*. Toronto: Nelson.

Moore, P. (1990). The effect of science inservice programs on the self-efficacy belief of elementary school teachers. *Dissertation Abstracts International, 51*(3), 823-A.

Moyer, P. S., & Jones, M. G. (2004). Controlling choice: Teachers, students, and manipulatives in mathematics classrooms. *School Science and Mathematics, 104*(1), 16-31.

Mulryan, C. (1995). Fifth and sixth graders' involvement and participation in cooperative small groups in mathematics. *Elementary School Journal, 95*(4), 297-310
.
Nattiv, A. (1994). Helping behaviors and math achievement gain of students using cooperative learning. *Elementary School Journal, 94*(3), 285-297.

Ontario Ministry of Education. (2004). Leading math success. Mathematical literacy grades 7-12. The report of the expert panel on student success in Ontario. Toronto: author.

Posavac, E., & Carey, R. (2003). *Program evaluation: Methods and case studies. 6/E.* Englewood Cliffs, NJ: Prentice-Hall.

Reys, B., Reys, R., & Koyama, M. (1996). The development of computation in three Japanese primary-grad textbooks. *Elementary School Journal, 96*(4), 423-438.

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*(6), 625-637.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two Standards-Based curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*(4), 368-398.

Romberg, T. A. (1997). The influence of programs from other countries on the school mathematics reform curricula in the United States. *American Journal of Education, 106*, 127-147.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51-65.

Ross, J.A. & Bruce, C. (2005, October). Teachers' beliefs in their instructional capacity: The effects of in-service. *Proceedings of the 27th annual meeting of the North American Chapter of the International Psychology of Mathematics Group*. Roanoke, VA: Virginia Tech.

Ross, J. A., & Cousins, J. B. (1995a). Giving and receiving explanations in cooperative learning groups. *Alberta Journal of Educational Research, 41*(1), 104-122.

Ross, J. A., & Cousins, J. B. (1995b). Impact of explanation seeking on student achievement and attitudes. *Journal of Educational Research, 89*(2), 109-117.

Ross, J. A., Hogaboam-Gray, A., & Bruce, C. (2004). *Student achievement effects of professional development for grade 6 mathematics teachers In the Kawartha Pine Ridge District School Board*. Peterborough, ON: OISE/UT Trent Valley Centre.

Ross, J. A., Hogaboam-Gray, A., McDougall, D., & Bruce, C. (2001). The contribution of technology to the implementation of mathematics education reform: Case studies of grade 1-3 teaching . *Journal of Educational Computing Research, 26*(1), 123-140.

Ross, J. A., Hogaboam-Gray, A., McDougall, D., & Le Sage, A. (2003). A survey measuring implementation of mathematics education reform by elementary teachers. *Journal of Research in Mathematics Education, 34*(4), 344-363.

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002b). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment, 8*(1), 43-58.

Ross, J. A., & McDougall, D. (2003). The development of education quality performance standards in grade 9-10 Mathematics teaching. Peterborough, ON: OISE/UT Trent Valley Centre.

Ross, J. A., McDougall, D., & Hogaboam-Gray, A. (2002). Research on Reform in Mathematics Education, 1993-2000. *Alberta Journal of Educational Research, 48*(2), 122-138.

Ross, J. A., & Regan, E. M. (1993). Sharing professional experience: Its impact on professional development. *Teaching and Teacher Education*, *9*(1), 91-106.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal* , *105*(1), 49-74

Rowan, B., Harrison, D. M., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal , 105*(1), 103-127.

Schoen, H., & Finn, K. F. (2001). *Teacher variables that relate to student achievement in a standards-oriented curriculum.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Schoen, H. L., Fey, J. T., Hirsch, C. R., & Coxford, A. F. (1999). Issues and options in the math wars. *Phi Delta Kappan, 80* (6), 444-453.

Shachar, H., & Shmuelevitz, H. (1997). Implementing cooperative learning, teacher collaboration and teachers' sense of efficacy in heterogeneous junior high schools. *Contemporary Educational Psychology, 23*, 53-72.

Slavin, R. E., & Madden, N. A. (1999, April). *Success for All/Roots and Wings: 1999 summary of research on achievement outcomes*. Paper presented at the annual meeting of the American Educational Research Associated, Montreal.

Tschannen-Moran, M., & Wolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

Turner, J. C., Meyer, D. K., Midgley, C., & Patrick, H. (2003). Teachers' discourse and sixth graders' reported affect and achievement behaviors in two high-mastery/high-performance mathematics classrooms. *Elementary School Journal, 103*(4), 357-382.

Villasenor, A., & Kepner, H. S. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24*(1), 62-69.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

What Works Clearinghouse. (nd) *WWC Study Review Standards* [Web Page]. URL http://www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf [2005, March 24].

Whicker, K. M., Bol, L., & Nunnery, J. A. (1997). Cooperative learning in the secondary mathematics class. *Journal of Educational Research, 91*(1), 42-48.

Wineburg, M. S. (1995, April). *The process of peer coaching in the implementation of cooperative learning structures*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Wood, T., & Sellers, P. (1997). Deepening the analysis: A longitudinal assessment of a problem-based mathematics program. *Journal for Research in Mathematics Education, 28*, 163-186.

Wortman, P. (1994). Judging research quality. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 98-109). New York: Russell Sage Foundation.

Appendix – All instruments

Student Survey

KAWARTHA PINE RIDGE DSB - OISE/UT FEBURARY 2006

Student Name:______________________________________ ○ male ○ female

For teacher use only ☐ IEP

IMPORTANT
Completely fill the appropriate bubble.

Section 1 This feeling scale shows how sure you are about something. It goes from 1 to 6. If you are feeling sure, pick a high number. If you are not sure, pick a low number. Fill in the bubble that best describes how sure you are about each question.

	not sure 1	2	3	4	5	really sure 6
1. How sure are you that you could solve a math problem?	○	○	○	○	○	○

2. As you work through a math problem how sure are you that you can:	not sure 1	2	3	4	5	really sure 6
a) understand the math problem.	○	○	○	○	○	○
b) make a plan.	○	○	○	○	○	○
c) solve the problem.	○	○	○	○	○	○
d) check the problem.	○	○	○	○	○	○
e) explain the solution.	○	○	○	○	○	○

Section 2 This section asks for your opinion about learning mathematics. Using the 1 to 6 scale where 1 = not at all true and 6 = very true, fill in the bubble that best describes how true each statement is for you.

The math that I learn in school is . . .	not at all true 1	2	3	4	5	very true 6
3. Mostly facts and procedures that have to be memorized	○	○	○	○	○	○
4. Thought provoking	○	○	○	○	○	○
5. Just a way of thinking about space, numbers, and problems	○	○	○	○	○	○

When the teacher asks a question in math class . . .	not at all true 1	2	3	4	5	very true 6
6. You have to remember the right answer to answer it correctly	○	○	○	○	○	○
7. There are lots of possible right answers you might give	○	○	○	○	○	○
8. You have to think really hard to answer it	○	○	○	○	○	○
9. The students who understand only need a few seconds to answer correctly	○	○	○	○	○	○

When the math teacher asks me a question that I can't answer right away . . .	not at all true 1	2	3	4	5	very true 6
10. Someone else will be asked	○	○	○	○	○	○
11. The teacher will answer the question	○	○	○	○	○	○
12. Everyone waits until I have thought about it and given my best try at answering	○	○	○	○	○	○

Section 3 This section asks for your opinion about the nature of mathematics. Fill in the bubble that best describes how true each statement is for you.

	not at all true 1	2	3	4	5	very true 6
13. Everything important about mathematics is already known by mathematicians	○	○	○	○	○	○
14. In mathematics you can be creative and discover things by yourself	○	○	○	○	○	○
15. Math problems can be done correctly in only one way	○	○	○	○	○	○
16. Real math problems can be solved by common sense instead of the math rules you learn in school	○	○	○	○	○	○
17. To solve math problems you have to be taught the right procedure, or you can't do anything	○	○	○	○	○	○
18. The best way to do well in math is to memorize all the formulas	○	○	○	○	○	○

When you get the wrong answer to a math problem . . .	not at all true 1	2	3	4	5	very true 6
19. It is absolutely wrong---there's no room for argument	○	○	○	○	○	○
20. You only find out when it's different from the book's answer or when the teacher tells you	○	○	○	○	○	○
21. You have to start all over in order to do it right	○	○	○	○	○	○

Draft

Section 4 Students have a lot of different thoughts and feelings when they are working in math class. We want to know how true these statements are for you. Fill in the bubble that best describes how true each statement is for you.

	not at all true 1	2	3	4	5	very true 6
22. If I gave the wrong answer to my teacher's math question, I would feel terrible.	○	○	○	○	○	○
23. If I were to do poorly in math, I would try not to let anyone know.	○	○	○	○	○	○
24. If I were to get a low grade in math, it would make me feel very sad.	○	○	○	○	○	○
25. I worry a lot about making errors on my math work.	○	○	○	○	○	○
26. I would get very discouraged if I made errors on a math assignment.	○	○	○	○	○	○
27. I really dislike math work on which I make mistakes.	○	○	○	○	○	○

Section 5 Some students put more effort forth in math class than others. We want to know how hard you work in math class. Using the 1 to 6 scale where 1 = not hard at all and 6 = as hard as I can, fill in the bubble that best describes how hard you are working in math.

	not hard at all 1	2	3	4	5	as hard as I can 6
28. How hard are you working to learn about math?	○	○	○	○	○	○
29. How hard do you study for math tests?	○	○	○	○	○	○
30. How hard are you working to solve math problems?	○	○	○	○	○	○
31. As you work through a math problem how hard are you working to understand the problem?	○	○	○	○	○	○
32. As you work through a math problem how hard are you working to make a plan? .	○	○	○	○	○	○
33. As you work through a math problem how hard are you working to solve the problem?	○	○	○	○	○	○
34. As you work through a math problem how hard are you working to check the problem?	○	○	○	○	○	○
35. As you work through a math problem how hard are you working to explain the solution?	○	○	○	○	○	○

In Section 6 and Section 7 we want to find out about gender issues.

Section 6 Using the 1 to 6 scale where 1 = very little and 6 = very much, fill in the bubble that best describes what you think about how much males and females like math.

	very little 1	2	3	4	5	very much 6
a) How much do most girls like math?.	○	○	○	○	○	○
b) How much do most boys like math?	○	○	○	○	○	○
c) How much do most women like math?	○	○	○	○	○	○
d) How much do most men like math?	○	○	○	○	○	○

Using the 1 to 6 scale where 1 = not good at all and 6 = very good, fill in the bubble that best describes what you think about how good males and females are at math.

	not good at all 1	2	3	4	5	very good 6
e) How good are most boys at math?.	○	○	○	○	○	○
f) How good are most girls at math?.	○	○	○	○	○	○
g) How good are most men at math?.	○	○	○	○	○	○
h) How good are most women at math?.	○	○	○	○	○	○

Section 7 Using the 1 to 6 scale where 1 = strongly disagree and 6 = strongly agree, fill in the bubble that best describes your feelings about participating in a new activity.

37. **I would be more willing to participate in a new class activity if I knew . . .**

	strongly disagree 1	2	3	4	5	strongly agree 6
a) boys are better at the activity than girls	○	○	○	○	○	○
b) boys like the activity more than girls	○	○	○	○	○	○
c) boys do the activity more often than girls	○	○	○	○	○	○
d) boys do the activity for a longer time than girls	○	○	○	○	○	○
e) girls are better at the activity than boys	○	○	○	○	○	○
f) girls like the activity more than boys	○	○	○	○	○	○
g) girls do the activity more often than boys	○	○	○	○	○	○
h) girls do the activity for a longer time than boys	○	○	○	○	○	○

Draft

Thank you!