Student Achievement Effects of Teacher PD

Ross, J. A., Bruce, C. D., & Hogaboam-Gray, A. (2006). The impact of a professional development program on student achievement in grade 6 mathematics. *Journal of Mathematics Teacher Education*, 9, 551-577.

The Impact of a Professional Development Program on Student Achievement in Grade 6 Mathematics

Abstract

Grade 6 teachers (N=106) in one school district were randomly assigned to early or late professional development (PD) groups. The program focused on reform communication and incorporated principles of effective PD recommended by researchers, although the duration of the treatment was modest (one full day and four after school sessions over a ten-week period). At the posttest, there were no statistically significant differences in student achievement. Although it could be argued that the result demonstrates that PD resources should be redirected to more intensive PD delivered over longer periods, we claimed that the PD was assessed prematurely. After the completion of the study, the external assessments administered by the province showed a significant increase in student achievement from one year to the next involving both the early and late treatment groups, an increase that was not found for the same students in other subjects. The study had high ecological validity: it was delivered by district curriculum staff to all grade 6 teachers, volunteers and conscripts alike. The cost to the district, less than CAN\$14 [9 euros] per student, was comparable to the modest expenditures typically available for professional development in Canadian school districts.

Keywords: mathematics, student achievement, professional development, grade 6

Student Achievement Effects of Teacher PD

The Impact of a Professional Development Program on Student Achievement in Grade 6 Mathematics

More than 90% of 450 National Staff Development Council projects reviewed by Killion (1998) contained no student achievement measure. Research on professional development (PD) for mathematics teachers is no exception to this pattern. Positive teacher effects have been reported for intensive PD delivered over extended time periods to volunteers but such studies rarely include student outcome data. In addition, there is little research on the effects of the shorter and less intensive PD that is available to typical teachers. This study attempts to redress these deficiencies by examining the student achievement impact of PD delivered to all grade 6 teachers in a school district, using a randomized field trial with a delayed treatment design¹.

Rationale for Focusing on PD

In the 1990s, mathematics education reformers focused on materials development, giving lesser attention to PD (Boissé, 1995). For example, Riordan and Noyce (2001) compared student achievement in schools using mathematics texts written to reform standards against traditional texts using control schools, matched on prior achievement and percentage of students receiving free lunch. The effect sizes, favoring the reform texts, were $ES=.34$ for early implementers and $.15$ for late implementers. The student achievement outcomes were consistent across student subpopulations (ability quartile, race, socio-economic status), similar for each of four mathematics strands, and consistent for traditional as well as reform learning objectives.

In Riordan and Noyce and related studies, it is difficult to disentangle the effects of PD from the effects of introducing novel texts. Carpenter, Blanton, Cobb, Franke, Kaput and McClain (2004) argued that the teacher knowledge required to implement mathematics reform cannot be embedded in materials. This claim is supported by evidence that teachers ignore or transform

Student Achievement Effects of Teacher PD

textbook elements that conflict with their views of mathematics teaching (Remillard, 2000; Ross, Hogaboam-Gray, McDougall, & Le Sage, 2003). Boissé (1995) drew parallels between recent and previous mathematics reform movements. He attributed the failure of the New Math movement of the 1950s and 1960s primarily to its inability to provide teachers with the training they needed to master the challenging expectations of the curriculum. Boissé's call for a focus on teacher education reverberated with reformers who sought to develop PD that is generative, that provides teachers with the capacity to reconstruct their practice around core ideals.

The Effects of Professional Development on Teacher Attitudes, Beliefs and Actions

PD effects on teachers (as opposed to student effects) are well-documented in individual case studies. PD that simultaneously focuses on teachers' practice, their cognitions about mathematics teaching, and their knowledge of mathematics increases implementation of key elements of standards-based teaching. Borko, Davinroy, Bliem and Cumbo (2000) provide a good example. This study traced two teachers participating in a PD program in which 14 teachers met with mathematics education researchers weekly for a full year and monthly for a second year. The researchers presented expert views of mathematics teaching; teachers applied these ideas in their own classes and discussed the resulting student products with the experts and their peers. The PD themes were sharing control with students, emphasizing conceptual learning (by assigning high level tasks and listening to student talk), and increasing student expectations. Both teachers changed in the expected directions, with one making large strides; the other was still in transition from traditional to reform practices at the end of the two year intervention. Borko et al.'s results are replicated by other studies in which intensive interaction with experts, classroom practice, and collaborative peer discussions provide credible evidence of increased implementation of standards-based mathematics teaching (Farmer, Gerretson, & Lassak, 2003; Moreira, 1997; Ross & Bruce, in press).

Student Achievement Effects of Teacher PD

In reviewing the case study literature, Wilson and Berne (1999) noted that, in many studies, it is difficult to determine what teachers learned about mathematics teaching, other than how to engage in professional discourse, and that with the exception of the Cognitively Guided Instruction (CGI) studies, there is little attention to student outcomes. Hill (2004) argued that these well-documented studies are untypical of PD available to most teachers. She studied 13 PD programs identified as exemplary, finding that all were deficient in some way. ~~An, for example is,~~ in their failure to connect activities to core mathematical ideas, focusing on the mechanics of a classroom activity rather than when to use it, and/or emphasizing proceduralization rather than understanding.

Studies involving larger samples of teachers experiencing more typical PD suggest that PD influences teachers' practice. Wenglinsky (2002) analyzed the 1996 grade 8 NAEP database using multilevel structural equation modeling. He found that PD (focusing on higher order thinking skills) strongly influenced classroom practice. Despite the methodological rigor of the analysis, Wenglinsky's claims are weakened by their correlational nature—there is no way to tell whether commitment to reform practice was a consequence of PD experience or a motivator for seeking it. A similar problem weakens Cohen and Hill's (2000) finding that teachers who participated in a more extensive PD (longer than one day) ~~that was~~ focused on student curriculum topics were significantly more likely to engage in reform teaching practices. Reys, Reys, Barnes, Beem and Papick (1997) found that, after the first year of a three year PD program, participating teachers had adopted many mathematics education reform principles, even though they fell considerably short of reform ideals. Reys et al. provided little information on teacher practice prior to entering the program and their original sample appeared to be elite: 80% had master's degrees and 40% were members of NCTM. It is impossible to tell whether the practices reported by Reys et al. were the result of the PD as opposed to prior teacher characteristics.

The Effects of Teacher PD on Student Achievement

A number of studies have reported positive effects of standards-based reform in which PD is one of several bundled initiatives. For example, Hamilton, McCaffrey, Stecher, Klein, Robyn and Bugliari (2003) conducted a meta-analysis of the student achievement outcomes in eleven sites receiving National Science Foundation funds for mathematics (and science) reform. The outcome measures were tests currently in use in the sites supplemented with standardized multiple choice and open ended items. The independent variables were self-reported teacher practices (two independent scales representing traditional and reform practices) and student demographics. The covariate was prior achievement (state test scores). Hamilton et al. found that teachers who implemented mathematics reform (defined as emphasis on conceptual understanding, real world applications of mathematical ideas, active engagement of students in constructivist tasks, and new forms of assessment) produced significantly higher student achievement, after controlling for other salient variables. The effects of reform teaching varied within- and between-sites and the effects were small. Hamilton et al. provides some evidence about the student achievement effects of PD in that one third of the NSF funds were allocated to PD. However, their design was not able to extract the unique contribution to student outcomes—it is possible that other factors, such as the provision of innovative curriculum materials, accounted for the student achievement effects. In addition, the external validity of the findings is weakened by the decision of Hamilton et al. to select the best cases in each site for their study. At best, their evaluation is a study of the student achievement efficacy of PD (i.e., in somewhat ideal conditions), not an effectiveness study (i.e., conducted in typical settings).

The few studies which isolated student achievement effects found that PD had mixed results. The strongest methodologically, Wenglinsky (2000), found that teacher PD had a small positive effect ($ES=.33$) on students' mathematics achievement. The results varied with the other

Student Achievement Effects of Teacher PD

variables in the model but in all cases the independent effect of PD was statistically significant and stronger than student socio-economic status, although weaker than classroom practice variables. However, Wenglinsky's study was cross-sectional rather than longitudinal, the variable of interest (teacher PD) could not be manipulated, and as Wenglinsky noted, the constructs in the model were derived from survey items created by NAEP for other purposes.

Cohen and Hill (2000) found that grade 2-5 teachers who reported participating more extensively in PD based on student curriculum topics had higher student achievement than teachers who did not but when classroom practice variables were included in the model, PD effects dwindled to insignificance. Although it could be argued that these findings suggest an indirect effect of PD on achievement (i.e., through changes in classroom practice), the student achievement results were based on a much reduced and unrepresentative sample. Teachers could opt out of the student testing and most did. The 27% who were included in this phase of the study were more reform oriented than those excluded. In addition, the type of PD program chosen by teachers was self-selected (i.e, uncontrolled).

Huffman and Thomas (2003) examined the effects of five types of PD on student achievement as measured by state assessments. For mathematics teachers, PD involving curriculum development was the only significant predictor, accounting for 16% of the variance in student achievement, a large effect. There were several methodological problems in this study including the use of step-wise regression (which produces results that are highly sample dependent); no other variables that might contribute to achievement were included in the equation (which inflates PD effects); and PD experiences were based on teacher self-reports and were not experimentally manipulated. The largest threat to the validity of the finding is the alternate explanation that teachers may have had access to curriculum development PD because they were recognized as leaders in implementing reform initiatives.

Student Achievement Effects of Teacher PD

Saxe, Gearhart and Nasir (2001) compared the student achievement effects of two approaches to PD. The Integrating Mathematics Assessment approach focused on teacher understanding--of the mathematics they taught, children's mathematics, and student motivation--and provided opportunity for teachers to reflect collaboratively on their teaching. The Collegial Support approach included only the last component. These two PD experiences were compared to each other and to a no-PD control condition consisting of teachers who were committed to using traditional texts. Saxe et al. found that the multi-dimensional PD approach produced higher upper elementary student understanding of key mathematics concepts. However, the internal validity of the comparison of the two PD approaches was threatened by the fact that teachers in the multi-dimensional PD condition had more PD time: a five day summer institute, followed by 12 evening sessions every two weeks and a full Saturday. Teachers in the collegial support PD received only two full days and 7 evening sessions. The external validity of the comparison of the two PD approaches to the control was weakened by the fact that teachers in the reform condition had demonstrated commitment to reform by using reform texts in their teaching at least once prior to the study. We have no way of knowing whether teachers with a lower commitment to reform experiencing similar PD would enjoy comparable student achievement benefits. In addition, the multi-dimensional PD was delivered by researchers while the one-dimensional approach was delivered by school district staff, which raises issues of the feasibility of scaling up the more successful treatment.

Shepard et al. (1996) provided weekly after school workshops for a year to grade 3 teachers. Treatment teachers developed rubrics and devised performance assessments focused on mathematics and language reform agendas. Students in treatment classes were matched against control classes (on socio-economic status and prior achievement - CTBS scores). Outcome measures were a battery of standardized (CTBS) and alternate assessments. There were small

Student Achievement Effects of Teacher PD

gains in mathematics achievement ($ES=.13$) on the state assessment but not on the alternate assessments. However, the state assessment was voluntary (raising external validity issues), the controls had higher prior achievement (raising internal validity concerns), and there was no explicit attempt to link changes in assessment practice to other dimensions of teaching.

Simon and Schifter (1993) examined the student achievement effects of a summer PD program that emphasized learning mathematics concepts through constructivist methods--teachers solved problems in groups and wrote journals. There was evidence of increased student understanding of key concepts (based on teacher reports of what students learned) but there were no changes in standardized test scores. The results were limited by methodological flaws: the study was a pre-post cohort design without control groups; the researchers used grade-equivalent scores rather than raw scores; the performance measures varied (teachers were from different states); and researchers treated individual survey items as independent variables (inflating Type I error).

Research Question

In summary, research on PD for mathematics teachers demonstrates mixed student achievement results, perhaps due to methodological problems that reduce credibility within and across studies. Even those studies employing rigorous analytic methods suffer from a lack of experimental controls at the design phase. To address these deficiencies in the literature, we conducted a study of the student achievement effects of a PD program offered to all grade 6 teachers in a single school district. Our research question was: Does teacher professional development enhance student achievement in mathematics?

Method

Sample

Student Achievement Effects of Teacher PD

The study was a randomized field trial involving all elementary schools in a single Canadian district. Over 95% of the students in the district were Canadian born, only 2% spoke a language other than English at home, 15% were identified as special needs, and average family income in the district was near the mean for the province of Ontario. The population consisted of 120 grade 6 teachers and we drew a random sample of six students per class. The teacher sample reduced to 106 teachers when teachers with incomplete student assessments were removed (i.e., there were 14 classes for which there were fewer than six student responses due to absences and a few cases of teacher misinterpretation of our directions.) The achieved sample represented 85% of the grade 6 teacher population for the district. The student sample represented 24% of the grade 6 student population. All grade 6 teachers in each school were randomly assigned to the early (September-December) PD group (i.e., the treatment) or to the late (January-May) PD group (i.e., the control).

Sources of Data

Student achievement was measured with a performance assessment comparable to the mandated assessments conducted by the Education Quality and Accountability Office (hereafter EQAO). Our test was shorter (60-90 minutes on each of three days rather than 150 minutes on each of five days), it covered only two mathematical strands (Number Sense & Numeration and Patterning & Algebra), and used different content (i.e., the September assessment used end of grade 5 content; the December assessment used mid-grade 6 content). The assessment was ~~made produced~~ by the teacher team that produced the 2002 grade 6 mathematics EQAO test. The performance assessments were field tested with 140 students in two adjacent districts. Students in both conditions completed the pre and post achievement tests.

On each administration, students read a short information booklet about water or wheels². On each of the three days they completed a mathematics investigation based on the same theme.

Student Achievement Effects of Teacher PD

The pretest focused on the theme of water and the posttest focused on the theme of wheels. Each investigation contained tasks that required Number Sense & Numeration and Patterning & Algebra. For example, in a Number Sense & Numeration task, students were shown a figure displaying a scooter wheel and a bicycle wheel in the ratio of 1:15, and were told that the bicycle wheel rotates 5 times every 10 metres. Three tasks were posed: (i) Students had to calculate how many times the scooter wheel will rotate every 10 metres. Students were required to show their work and explain how they solved the problem. (ii) Students were given the information that bikes have 64 spokes and that spokes are packaged in boxes of various sizes that combine to create multiples of 64. Students were required to show all possible combinations of spokes and boxes. Again, students were required to show their work and explain their answer. (iii) Students were told that there are 24 teeth on the front gear and 18 teeth on back gear of a bike. Students were given three options for representing the relationship between the gears: $1\frac{1}{3}$; 40%, and 4:3. They had to select the best representation and justify their choice. We considered this item to be a grade-appropriate problem solving task because it provided for a (i) variety of solution strategies, (ii) involved the identification and use of curriculum-relevant mathematical concepts, (iii) drew upon knowledge from children's world, (iv) provided for different ways of representing the problem, and (v) required solution justification.

Each booklet generated eight scores: four aspects of mathematics achievement (problem solving, concept understanding, application of mathematical procedures, and communication of mathematical ideas) · two strands of mathematics (Number Sense & Numeration and Patterning & Algebra). The most complex dimension was communication. The rubric provided four sets of indicators: (i) justification of solution as reasonable (i.e., how well the student provided evidence to support his/her arguments), (ii) use of mathematical language (i.e., how well the student incorporated mathematical words and symbols into the argument), (iii) use of sketches, diagrams

and charts to communicate mathematical ideas, and (iv) purposes for using multiple representations (i.e., how well the student combined representations to communicate solutions and solution strategies). The full rubric is available from the authors.

EQAO reports scores consisting of levels 1-4 and several categories below level 1. ~~E~~each level is defined in provincial policy (Ontario Ministry of Education and Training, 1997). To increase discrimination, we used a six-point scale that corresponds to the distinctions made by teachers; i.e., Level 1 or below; Level 2 low (close but not fully at level 2); Level 2; Level 3 low; Level 3; and Level 4. All assessments were marked by a central team of teachers that were trained in a full day session in February. All assessments were previously coded so that no information about the school, the student or the experimental condition was available to the markers³. The marking session began with a review of the marking rubric and anchor papers illustrating each level. After marking in pairs to establish consistency, each marker scored sets of six papers, coded to conceal teacher, school and treatment group. There were two levels of reliability checks. At the first level, all papers and their assigned grades were reviewed by the team leader. If there were discrepancies between the team leader and the marker's assessment, the team leader and marker negotiated the differences. If the discussion did not lead to agreement, a master scorer arbitrated the decision. A second reliability check was conducted at the beginning, midpoint, and end of the marking session by having a random sample of items independently scored by a second marker. The reliability sample over the three sessions comprised 20% of the total items. Agreement within one level of the scale was Kappa=.73, .97, and .97 respectively. (Kappa adjusts the proportion of units on which judges agree by the proportion of units for which agreement is expected by chance. Stemler (2004) suggests that Kappa scores over .60 indicate substantial agreement.)

In May students completed the mandated EQAO assessment. This assessment, held over five days, was in the same format as the September and December assessments except that five

Student Achievement Effects of Teacher PD

strands were assessed (i.e., Probability & Data Management, Measurement and Geometry were added to Number Sense & Numeration and Patterning & Algebra) and there was an added multiple choice component. The May assessments were marked by EQAO and grades were adjusted (using the protected multiple choice component) to ensure equivalence from one year to the next. *EQAO achievement* consisted of a 0-4 score for each student; i.e., the 1-4 scale reported by EQAO and level 0 for all categories below level 1.

Students in both experimental conditions completed surveys in September to test the equivalence of treatment and control groups on eight motivational measures associated with student achievement in previous research. All the items specifically addressed mathematics class experiences, beliefs, and attitudes.

(i) *Mathematics self-efficacy* consisted of six Likert items measuring expectations about future performance (from Ross, Hogaboam-Gray, & Rolheiser, 2002). For example, “As you work through a math problem how sure are you that you can:... a) understand the math problem. The response options were a six-point scale anchored by “not sure” and “really sure”. Pajares (1996) reviewed evidence that self-efficacy predicts achievement directly and indirectly through goal setting: students with high self-efficacy are more likely to be successful.

Formatted: Font: (Default) Times New Roman, 12 pt

The goal orientations survey consisted of six items from Midgely et al. (1998) for each of three scales (ii) *task goal orientation* (e.g., “The work made me want to find out more about the topic.”), (iii) *ability-approach goal orientation* (e.g., “I want to do better than other students in my math class.” and (iv) *ability-avoid goal orientation* (e.g., “It's very important to me that I don't look stupid in math class.”. Response options were a six-point scale anchored by “not at all true” and “very true”. Goal orientations represent student aims for engaging in a classroom activity. Students with a task orientation focus on the intrinsic value of learning; students with an ability-approach orientation focus on demonstrating their ability; students with an ability-avoid

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Student Achievement Effects of Teacher PD

orientation focus on concealing their ability. High task orientation is consistently associated with high achievement; high ability-approach orientation is inconsistently but usually positively associated with high achievement; and high ability-avoid orientation is negatively associated with achievement (Wigfield, Eccles, & Rodriguez, 1998).

Closely associated with goal orientations is (v) *negative affect for failure* (fear of failure).

We adapted six items from Turner, Meyer, Midgley and Patrick (2003); for example, “If I were to do poorly in math, I would try not to let anyone know.”. Response options were a six-point scale anchored by “not at all true” and “very true”. Scores on this scale are negatively correlated with achievement.

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Turner et al. provided a theoretical argument for attending to student perceptions of classroom goal structures as well as individual goal orientations. They argued that individual orientations were influenced by student interpretations of the motivational climate of the classroom. We administered from Turner et al. (2003) six items for (vi) *classroom mastery goal*

structure (e.g., “My teacher wants us to understand our math work, not just memorize it.”) and five items for (vii) *classroom performance goals structure*; (e.g., “My teacher lets us know which students get the highest scores on a math test.”) Response options were a six-point scale anchored by “not at all true” and “very true”. Turner et al. argued that these perception of classroom motivation scales should be combined, with the expectation that student achievement would be highest when students perceived the goal structure to be high on both scales.

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Our final measure for testing the equivalence of the groups was (viii) *effort*. It was measured with eight items developed for this study measuring how hard students work in mathematics class. For example, “how hard do you study for math tests?” The response scale was a six-point scale anchored by “not hard at all” and “as hard as I can”. It is a reasonable inference that effort will correlate with student achievement.

A *Chronology of the Study* is attached as a table in the Appendix.

Formatted: Font: Times New Roman,
12 pt

Treatment

The PD consisted of one full day, followed by three 2-hour after-school sessions delivered over a ten-week period. Sessions were held in three sites to reduce group size and teacher travel time. Communication of mathematical ideas was the organizing theme, chosen because it impacts multiple aspects of mathematics teaching. The PD goals included moving teacher practice toward (i) the use of rich tasks (i.e., complex, open-ended problems embedded in real life contexts that provide multiple solutions and/or multiple solution strategies), (ii) sharing and appraising mathematical ideas in student groups and whole class discussions, and (iii) teachers and students collaboratively constructing mathematical knowledge. Both groups received the same PD except that the September group received it before the December posttest and the January group participated after the posttest.

The full day distinguished constructivist conceptions of mathematics teaching from transmission approaches. Ten dimensions of teaching (from Ross et al., 2003) were displayed and teachers in small groups arranged descriptions of teaching for selected dimensions into partial rubrics representing a continuum of teaching from traditional to reform practice. We presented research evidence (from the review in Ross, McDougall, & Hogaboam-Gray, 2002) to argue that moving toward standards-based teaching was likely to increase student achievement. Teachers worked in groups on a rich task drawn from the grade 6 curriculum (determining what constitutes a triangle) and developed three ways of communicating their findings. Teachers identified examples of “good talk” (e.g., justifications of solutions) that occurred in their groups. Participants worked through a Mobius strip investigation in which a workshop leader modeled teaching. During the presentation, participants recorded what students would see and identified what the teacher might be thinking. Teachers developed a rubric for communicating mathematical

Student Achievement Effects of Teacher PD

ideas which they used to assess their group's talk about processes and solutions. Additional rich tasks were provided and teachers planned between-session activities with peers from their schools. We asked teachers to have students work on a rich task (from the resource booklet distributed or their textbooks) that required communication of mathematical ideas, use teaching ideas discussed at the previous PD session, and bring student responses to the next session.

The second session began with teachers reflecting on their experiences with the between-session investigations, focusing on what distinguishes strong from weak student performances. Each group member displayed student responses to one of the rich tasks, and described his/her strategies for enhancing sharing of solutions and processes. Presenters displayed a rubric for mathematical communication, synthesized from the products of each of the three sites in session one. The rubric emphasized explanations of solution strategies using mathematical concepts, justifying solutions, using mathematical language, and using multiple representations. Teachers applied the rubric to the examples of student communication brought to the session and generated prompts (from the rubric) that would elicit higher quality mathematical communication. A presenter demonstrated how to use such prompts to stimulate communication about a rich task. Additional examples of rich tasks (toothpick patterns, number patterns, the amazing number 1089) were distributed for classroom use between-sessions.

The third session began with small group reflection on between-session activities. As each teacher shared his/her strategies and student responses, another member of the group recorded the prompts the teachers used to stimulate communication. The reporting teacher indicated which prompts worked well/poorly and the group suggested why. Teachers generated prompts to improve communication among students, using a transcript of students working in a group as raw material (from Ross, 1995). Examples of prompts used in research to stimulate explanations (e.g.,

King, 1994) were distributed. Examples of between-session tasks (blockhouse pattern, growing dots-triangular numbers, building a pool and deck) were distributed.

The fourth session began by debriefing of between-session activities. Teachers worked in groups on a rich task in which they had to find the pattern between the number of tables and number of seats around it. Teachers constructed a group log in which they recorded their solution strategies. Teachers generated prompts that they could teach to students and use themselves to improve communication. They gave particular attention to strategies for getting students to write mathematical explanations. Resources (e.g., classroom posters) and a self-checking rubric for students were also distributed.

What made the specific activities undertaken in the PD illustrative of reform was their explicit attention to the three PD goals. For example, the triangle task modeled by the presenters consisted of three images of a triangle, each showing three more or less straight lines enclosing a two-dimensional space. The images varied in how straight the lines were. PD participants were asked to decide which of these images was a triangle and why.

- PD Goal (i) The use of rich tasks: The task was nontraditional because it did not provide for a single solution or single strategy for reaching it. One could make a plausible case for 0, 1, 2 or 3 of the images as triangles.
- PD Goal (ii) Sharing and appraising mathematical ideas in student groups and whole class discussions: The presenters elicited from individuals and groups their reasons for arguing a particular image was a triangle. In doing so, presenters stayed with a single response for longer than would be the case in a traditional classroom; they emphasized the mathematics concepts embedded in the arguments made by participants; they linked the arguments to those likely to be encountered in grade 6 classrooms; and they linked mathematical ideas embedded in the triangle lesson to other lessons.

- PD Goal (iii) Teachers and students collaboratively constructing mathematical ideas: The presenters made explicit how their actions contributed to shared knowledge about triangles, emphasizing actions that develop a mathematical community comparable to the learning communities teachers should construct in their own classrooms. In constructing knowledge within a learning community, presenters emphasized the role of the teacher as a co-learner.

After the post tests, the January teachers began the treatment which continued until early May. In February, after the assessments were scored, teachers in both groups participated in a marking session. We gave each teacher the pre and post assessments for their students, including the randomly chosen six students included in the data analysis. We distributed a training booklet consisting of anchor papers illustrating each of the levels of the rubric, the scoring guide, and the test blueprint showing which sections of the tests measured each of the rubric dimensions.

Teachers constructed a personal set of “look fors” for each level based on the anchor papers.

Teachers marked one example together and compared their results. In pairs, teachers marked one of the students from their own class and compared their codes to those of the researchers.

Teachers continued to work in pairs marking whichever student papers they chose.

Data Analysis

After examining the characteristics of study variables, we tested the equivalence of the treatment and control groups, using separate sample t-tests for the student motivational variables. The teacher was the unit of analysis in subsequent procedures. We conducted a multivariate analysis of covariance (using the General Linear Model program in SPSS) in which the dependent variables were posttest achievement; the pretest score was the covariate and the independent variable was experimental condition. We also conducted a separate samples t-test comparing 2004 EQAO scores to the EQAO average for the preceding three years.

Results

Descriptive Analysis

We examined the distributional properties of all variables. Outliers were defined as 3.0 standard deviations above or below the mean and were reduced to the mean $\pm$ 3.0 SD. Variables were defined as normally distributed if the skewness index was below 3.0 and kurtosis was below 10.0. All variables met these criteria.

Table 1 displays the number of cases, means, standard deviations, reliabilities (Cronbach's alpha) for the student survey variables in the study and shows separate sample t-tests comparing the two groups on pretest variables. There were more cases in the September than in the January group; i.e., a few teachers drifted from the late to the early PD group. We contacted each of these teachers about why they switched groups: their reasons for violating random assignment were idiosyncratic rather than systemic (e.g., "I thought I was in the fall group because the fall PD sessions were held at my school"). There were no significant differences between the groups on any of the pretest surveys. Table 1 also indicates that all instruments were of adequate reliability, except that *classroom mastery goal structure* fell slightly below the $\alpha=.70$ criterion. In summary, Table 1 demonstrates that the student groups were equivalent on motivational measures that affect achievement. The Table also demonstrates that these motivational measures were internally consistent—i.e., the failure to find differences could not be attributed to lack of reliability. Together the Table 1 information reduces a threat to the validity of the study by eliminating a possible alternate explanation (motivational differences between the groups) for any achievement differences we might find.

Table 1 About Here

The student achievement assessment produced eight scores based on four aspects of mathematics achievement (problem solving, concept understanding, application of mathematical

procedures, and communication of mathematical ideas) · two strands of mathematics (Number Sense & Numeration and Patterning & Algebra). The eight scores were entered into an exploratory factor analysis (principal axis with promax rotation). There was a single factor solution, *mathematics achievement*: only one factor had an eigenvalue above 1.0; the second factor had an eigenvalue one-sixth as large (first factor=5.14; the second=.84); and the single factor accounted for 64.3% of the variance, using the pretest data. The results were virtually identical for the posttest data (factor 1 had an eigenvalue of 5.08; factor 2=.753; factor 1 accounted for 63.5% of the variance). In other words we could represent the eight scores as a single variable because each score represented the same attribute: grade 6 mathematics performance.

Research Question: Does teacher professional development enhance student achievement in mathematics?

We conducted an analysis of covariance in which the dependent variable was the composite achievement score on the posttest, the covariate was pretest composite achievement, and the independent variable was experimental condition. There was a statistically significant pretest effect [$F(1, 103)=.111.256, p<.001$]: 52% of the variance in posttest scores was explained by the pretest. There were no significant differences between the treatment and control groups [$F(1, 103)=.193, p=.662$].

We explored further by conducting a multivariate analysis of variance for four of the achievement scores that contributed to the composite achievement variable. Two are associated with mathematics education reform and were given more attention in the PD: posttest scores on Communication of Mathematical Ideas (averaged across the two mathematics strands) and posttest scores on Patterning & Algebra (averaged across four aspects of mathematics achievement). The other two variables were outcomes given more attention in traditional

programs: posttest scores on Number Sense & Numeration (averaged across four aspects of mathematics achievement) and posttest scores on problem solving dimensions other than communication of mathematical ideas (i.e., problem solving, concept understanding, and application of mathematical procedures; these were averaged across the two mathematics strands). The covariates in the analysis consisted of the pretest scores on the same variables. For example, pretest scores on Communication of Mathematical Ideas consisted of pretest scores in Communication averaged across the two mathematics strands. The independent variable was group assignment.

The multivariate results indicated there was a significant pretest effect [$F(1, 103)=37.352$, $p<.001$] accounting for 53% of the variance in posttest performance. To interpret the size of this pretest effect, consider that a variable accounting for 15% of the variance is considered to be large (Cohen, 1988). That more than half of the posttest variance was accounted for by pretest performance indicates that mathematics achievement was very stable over the duration of the in-service. There were no significant differences between the treatment and control groups [$F(1, 103)=2.529$, $p=.062$]. The univariate results, displayed in Table 2 showed the same results for each of the subtests. Pretest scores accounted for 47-50% of the posttest variance. There were slight differences in the adjusted posttest student achievement means favoring the treatment group but none was large enough to reach statistical significance. In summary, in Table 2 we explored the possibility that the PD program had an effect that was limited to one dimension of achievement and the effect was diluted when we averaged all dimensions together in a single composite measure of achievement. Table 2 demonstrated that was not the case. Teacher participation in the PD program had no statistically significant effect on any of the four sub-measures of student achievement.

Table 2 About Here

Student Achievement Effects of Teacher PD

Finally, we examined changes in the annual provincial assessments from 2003 to 2004. In this analysis we collapsed the treatment groups into a single category since all grade 6 teachers had participated in the PD by the time of the provincial assessment in May 2004. In this comparison we used the EQAO definitions of levels in a 0-4 scale rather than the 0-6 scale developed for the main analysis in this study. We found that grade 6 mathematics achievement increased significantly from 2003 to 2004 [$t(6105)=3.73, p<.001$], although the effect was small ($ES=.10$). In contrast there were no significant district differences in grade 6 Reading [$t(6105)=0.839, p=.401$] or Writing scores [$t(6105)=1.749, p=.080$].

Discussion

The PD investigated in this study constituted a valid instance of the PD available to typical teachers. It incorporated features of effective mathematics PD identified in Hill's (2004) review: active learning by teachers (teachers worked through student tasks); examples from classroom practice (all tasks were from the curriculum they taught); 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); 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 content (explicit attention was given to the mathematical concepts embedded in each task and to alternative strategies for eliciting these concepts). The sessions gave less attention to two other dimensions identified by Hill: attention to student learning theory was implicit and teacher involvement in planning was restricted to a narrow range of choices.

The evaluation of the PD program had high internal and external validity. Internal validity was established by randomly assigning teachers to early and late treatment conditions. Although there were minor departures from random assignment, the reasons given by the few teachers who

Student Achievement Effects of Teacher PD

disregarded their assignments indicated that violations were unsystematic and unrelated to the focus of the PD. In addition, a battery of measures associated with student achievement in mathematics found no significant prior differences between the student groups. Reported elsewhere (Ross, 2004) is evidence that there were no significant prior differences between the teachers in the two conditions with respect to teacher characteristics that influence implementation of reform, self-reported teaching practices and beliefs about instructional capacity. There were no differences in the textbooks available in the two conditions and both teacher groups worked within the same district policies. Although some studies make the student the unit of analysis, thereby inflating statistical power, this study used the appropriate unit of analysis, the teacher. External validity was established by including all grade 6 teachers in the district in the study. Teachers were not selected on the basis of their predisposition to standards-based mathematics teaching or on the basis of their willingness to participate in the study. In addition, the PD was mainly delivered by district staff and the costs were within the district's existing PD budget without being supplemented by external grants.

The study provided an unambiguous answer to the research question: the ten-week teacher PD program did not contribute to student achievement. But this result can be interpreted in two ways.

First, it may mean that PD of this type is a waste of resources. Although research on the achievement effects of PD offered to substantial groups of teachers is limited, the results reported to date are discouraging. Other researchers found that PD, to the extent that it could be disentangled from other reform initiatives, had no student achievement effect (Cohen & Hill, 2000), had a positive effect on some measures but not others (Saxe et al., 2001; Simon & Schifter, 1993), or that some forms of PD were effective but others were not (Huffman & Thomas, 2003).

Student Achievement Effects of Teacher PD

These weak and inconsistent results from quantitative studies contrast with qualitative case studies reporting substantial changes in teacher practice and student performance. For example, Steinberg, Empson and Carpenter (2004) found that a teacher who had been trained in Cognitively Guided Instruction (CGI) principles made large gains in her responsiveness to individual differences in students' mathematical thinking, once she had adopted an inquiry stance toward her teaching. The teacher changed from level 3 to level 4b on the CGI scale in one year.

At the outset of the study, the teacher was implementing some elements of reform teaching but she was not challenging students about their mathematical understanding--she tended to accept what they said and was not able to predict their problem solving strategies. The changes this teacher made were still visible when the authors observed the teacher **years** later, confirming the evidence presented by Franke, Carpenter, Levi, and Fennema (2001) about the enduring effects of CGI. Although Steinberg et al. did not collect student achievement data, previous research involving randomized field trials found that students in classroom of teachers implementing CGI had significantly higher achievement than students in control groups (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989). Steinberg et al. attributed the impact of this teacher's PD to 1) the immersion of the teacher in a discourse community of CGI teachers; i.e., an experienced teacher in the same school and the support provided by the researchers; 2) to the teacher acquiring processes for reflectively generating, debating and evaluating new knowledge and practices when she shifted from being a passive to active observer of children's thinking; and 3) to her ownership of the change process. In the Steinberg et al. study, the teacher interacted with an expert observer, a researcher who observed her children's thinking in 34 lessons over a five month period, interviewed each of her students, and then discussed what the researcher had observed and provided specific advice when meeting with the teacher for 30-40 minutes every week for 13 weeks. The intensity of Steinberg et al.'s intervention contrasts starkly with the support provided

Comment [d1]: How many?

Formatted: Highlight

to teachers in our study. From this perspective, the PD evaluated in our study was insufficiently intensive to make a meaningful difference. Perhaps it would have been better to redirect the resources expended across the district to a few teachers who were most amenable to professional growth.

A second interpretation is that we evaluated the PD program prematurely. In Chen's (2005) taxonomy of evaluations, we conducted an outcome evaluation that measured program effectiveness in actual, as opposed to ideal, conditions. Chen persuasively argued that, until a program has reached mature implementation and has been demonstrated to be viable in ideal conditions, an outcome evaluation is premature. The duration of PD programs reported in case studies far exceeds our ten-week delivery. For example, Zaslavsky and Leikin (2004) reported a case study of interactions among teachers, teacher educators, and a PD program director in a site where teachers met on professional issues for six hours each week; the PD program extended for five years. Multi-year interventions are not unusual (cf. Franke et al., 2001; Borko et al., 2000). Supovitz and Turner (2000) found that the duration of PD was a strong predictor of change in science teaching--at least 60 hours were required for substantive change and the ability to construct an investigative culture required at least 160 hours. We suspect that similar time allocations are required in mathematics education. From this perspective, the ten-week PD program, a total of 12 hours, was evaluated too soon. However, this particular PD effort was part of a long term agenda of the district to improve mathematics achievement through a variety of initiatives. It was preceded and followed by further opportunities for teachers to deepen their understanding and application of standards-based teaching.

The best argument for the claim that the evaluation was premature was the change in student achievement on the provincial assessments written in May 2004: scores increased significantly from 2003 to 2004, although the effect size was small. Given the finding of Linn and

Student Achievement Effects of Teacher PD

Haug (2002) that external assessment results for individual schools are highly unstable from one year to the next, it could be argued that the differences between 2003 and 2004 achievement in our study represent changes in the ability of the two cohorts and/or changes in the difficulty of the tests. However, there were over 3000 students involved in the assessments in both years, diluting the effects of between-cohort fluctuations. The validity of the comparison is strengthened by EQAO procedures to ensure the validity of year over year comparisons. EQAO uses expert panels to review test content, conducts field tests to calibrate the difficulty of its tests, and corrects student scores using performance on a protected multiple choice item battery. In addition, the grade 6 Reading and Writing assessments completed by the same students showed no significant EQAO gains from 2003 to 2004.

The evaluation ended after ten weeks but the PD continued. Since there were no differences between the groups at the posttest, the PD effect is likely attributable to the supervised marking session that occurred after the posttests. EQAO does not release student test papers to teachers and few participate in the summer marking sessions. Teachers who do participate in summer sessions report that assessing student responses and discussing them with teacher colleagues gives them greater understanding of provincial curriculum requirements and insight into how students think about key mathematics concepts. The only study we located that examined the effects of supervised marking, found that teachers who participated in scoring produced higher student achievement in algebra than teachers who were in the control condition (Schafer, Swanson, Bené, & Newberry, 2001). A key question is whether the effects of such preparation represent true improvements in learning or merely increased scores. We argue, with Haladyna, Nolan, and Haas (1991), that test preparation that focuses on the concepts to be learned and the criteria for their appraisal, rather than on superficial features of test format and content, is appropriate.

Conclusion

The threats to the internal validity of post only comparisons of two cohorts are formidable. We cannot say with certainty that the PD program had a delayed impact on the annual assessments. Our study found that PD of modest duration (one full day and four after-school sessions) was associated with improvement in student achievement on external assessments: the proportion of students achieving the provincial standard increased from 50% to 54% (an 8% increase on the 2003 base). The cost to the district, less than CAN\$14 (9 euros) per student, was comparable to the modest expenditures typically available for PD in Canadian school districts.

The key practical question is whether these funds would have been better allocated to engage a smaller group of teachers at a deeper level. The argument is that concentrating resources would lead to more authentic application of reform principles, albeit in fewer sites. This argument would be more persuasive if we could be assured that recipients of intensive PD would respond as the exemplary teacher in Steinberg et al. (2004) did, by making substantial improvements in her practice. But case studies reveal that teacher response is highly variable. For example, Borko et al. (2000) presented evidence that two teachers in the same school who experienced intensive PD changed in expected directions. But one made considerably larger strides than the other. Borko et al. attributed the changes to the PD and to the organization in which it occurred. They also noted that teacher characteristics beyond the control of PD staff or the district moderated PD effects. The teacher who did not progress as far was constrained by a family situation that limited the time she could devote to changing her practice and by beliefs about the teacher's role that inhibited her from sharing responsibility for learning with students.

Even if we could accurately select the teachers most likely to benefit from intensive PD, how could we deny the learning opportunities of standards-based mathematics teaching to the

Student Achievement Effects of Teacher PD

students of teachers not chosen? School districts have a moral imperative to equalize opportunity across schools and classrooms.

We recommend that districts divide their resources between two levels of PD:

- Intensive PD experiences with selected teachers in ideal circumstances could constitute existence proofs for teacher change. Qualitative case studies could identify processes in these sites most plausibly linked to improved teaching.
- Large scale PD activities would involve short duration, less intensive learning opportunities based on findings from existence proof studies. Annual achievement monitoring would measure progress over time. Central to this strategy is the integrated design of intensive and distributed PD offerings so that each informs the other and to ensure that the small effects of short duration PD accumulate over time.

The generalizability of the findings of this study is limited by the fact that it was conducted in a single school district. Knapp (1997) found that mathematics education reform is affected by characteristics of the district, with superficial implementation in some settings and deeper engagement in others. The district in which this research was conducted had been energized by a new director, improvement of student achievement was the core goal, and district activities were aligned toward its realization. District capacity was enhanced by capable curriculum staff. We do not regard these features as unique to this district. We conclude from this study that committed staff can harness modest resources to deliver PD that makes an incremental contribution to student achievement.

References

- Borko, H., Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers' experience in a mathematics and literacy staff development project. *Elementary School Journal*, 100(4), 273-306.
- Bossé, M. J. (1995). The NCTM Standards in light of the New Math movement: A warning! *Journal of Mathematical Behavior*, 14, 171-201.
- Carpenter, T. P., Blanton, M. L., Cobb, P., Franke, M. L., Kaput, J., & McClain, K. (2004). *Scaling up innovative practices in mathematics and science*. Madison: University of Wisconsin-Madison, National Center for Improving Student Learning and Achievement in Mathematics and Science.
- Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499-531.
- Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294-343.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ: Erlbaum.
- Farmer, J. D., Gerretson, H., & Lassak, M. (2003). What teachers take from professional development: Cases and implications. *Journal of Mathematics Teacher Education*, 6 (4), 331-360.
- Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American*

Educational Research Journal, 38(3), 653-690.

Haladyna, T., Nolan, S., & Haas, N. (1991). Raising standardized achievement scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.

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.

Huffman, D., & Thomas, K. (2003). Relationship between professional development, teachers' instructional practices, and the achievement of students in Science and Mathematics. *School Science and Mathematics*, 103(8), 378-387.

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

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368.

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.

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 24(4), 29-36.

Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., Anderman, E., & Roeser, R. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology*, 23, 113-131.

Moreira, C. Q. (1997). Between the academic mathematics and the mathematics education

- worlds. *European Journal of Teacher Education*, 20(2), 171-189.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Remillard, J. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new textbook. *Elementary School Journal*, 100(4), 331-350.
- Reys, B. J., Reys, R. E., Barnes, D., Beem, J., & Papick, I. (1997). Collaborative curriculum investigation as a vehicle for teacher enhancement and mathematics curriculum reform. *School Science and Mathematics*, 87(5), 253-259.
- 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.
- Ross, J. A. (1995). Students explaining solutions in student-directed groups: Cooperative learning and reform in mathematics education. *School Science and Mathematics*, 95(8), 411-416.
- Ross, J.A. (2004). *Student and teacher effects of professional development for grade 6 mathematics teachers*. Interim report to the Ontario Ministry of Education and Training. Peterborough, ON: OISE/UT Trent Valley Centre.
- Ross, J. A., & Bruce, C. (in press). Self-assessment and professional growth: The case of a grade 8 mathematics teacher. *Teaching and Teacher Education*.
- 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.

Student Achievement Effects of Teacher PD

- 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.
- Saxe, G. B., Gearhart, M., & Nasir, N. S. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4(1), 55-79.
- Schafer, W. D., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14(2), 151-170.
- 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.
- Simon, M. A., & Schifter, D. (1993). Toward a constructivist perspective: The impact of a mathematics teacher inservice program on students. *Educational Studies in Mathematics*, 25, 331-340.
- Steinberg, R. M., Empson, S. B., & Carpenter, T. P. (2004). Inquiry into children's mathematical thinking as a means to teacher change. *Journal of Mathematics Teacher Education*, 7(3), 237-267.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved March 1, 2004 from <http://PAREonline.net/getvn.asp?v=9&n=4>
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37(9), 963-980.
- 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.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12). Downloaded from <http://epaa.asu.edu/epaa/v10n12/> December 6, 2005.

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.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad, & P. D. Pearson (Eds.), *Review of Research in Education*. Vol. 24 (pp. 173-210). Washington: American Educational Research Association.

Zaslavsky, O., & Leikin, R. (2004). Professional development of mathematics teacher educators: Growth through practice. *Journal of Mathematics Teacher Education*, 7(1), 5-32.

Student Achievement Effects of Teacher PD

Appendix *Chronology of the Study*

<i>What</i>	<i>When</i>	<i>Who</i>
1. Assignment of teachers to conditions	June 2003	Researchers
2. Administration of achievement pretests	September 2003	Classroom teachers
3. Administration of student surveys	September 2003	Classroom teachers
4. In-service for treatment teachers	Sept - Dec 2003	In-service team
5. Administration of achievement posttests	December 2003	Classroom teachers
6. Marking of pre- and post-tests	February 2004	Marking team
7. In-service for control teachers	Jan – May 2004	In-service team
8. Administration of EQAO assessments	May 2004	Classroom teachers
9. Marking of EQAO assessments	July 2004	EQAO assessors

Student Achievement Effects of Teacher PD

Table 1

Pretest Equivalence: Student Motivation Variables

	Group	N	Mean	SD	Alpha	t	df	p
(i) Math Self-Efficacy	Sept	406	4.39	0.85				
	January	310	4.50	0.88	0.84	-1.640	714	0.101
(ii) Task Goal Orientation	Sept	408	4.41	1.13				
	January	309	4.38	1.18	0.86	0.450	715	0.730
(iii) Ability-Approach Goal Orientation	Sept	408	3.39	1.30				
	January	309	3.44	1.29	0.85	-0.454	715	0.650
(iv) Ability-Avoid Goal Orientation	Sept	408	2.85	1.28				
	January	309	2.87	1.28	0.84	-0.239	715	0.811
(v) Negative Affect for Failure	Sept	406	3.02	1.16				
	January	308	2.98	1.16	0.82	0.461	712	0.645
(vi) Classroom Mastery Goal Structure	Sept	408	5.01	0.68				
	January	309	4.96	0.76	0.67	0.992	715	0.322
(vii) Classroom Performance Goal Structure	Sept	408	2.51	1.11				
	January	309	2.38	1.03	0.70	1.566	715	0.118
(viii) Effort in Math Class	Sept	407	4.79	0.81				
	January	309	4.71	0.89	0.86	1.242	714	0.215

Student Achievement Effects of Teacher PD

Table 2

Univariate Effects of PD on Student Achievement

		<i>F</i>	<i>df</i>	<i>p</i>	<i>eta</i> ²
Corrected Model	pattern2	48.19	2,103	<.001	0.483
	numsen2	47.91	2,103	<.001	0.482
	commun2	51.47	2,103	<.001	0.500
	other2	51.48	2,103	<.001	0.500
Intercept	pattern2	26.27	1,103	<.001	0.203
	numsen2	58.58	1,103	<.001	0.363
	commun2	23.01	1,103	<.001	0.183
	other2	51.83	1,103	<.001	0.335
achieve1	pattern2	96.30	1,103	<.001	0.483
	numsen2	94.87	1,103	<.001	0.479
	commun2	102.64	1,103	<.001	0.499
	other2	101.82	1,103	<.001	0.497
group	pattern2	0.56	1,103	0.455	0.005
	numsen2	0.00	1,103	0.970	0.000
	commun2	2.58	1,103	0.111	0.024
	other2	0.00	1,103	0.987	0.000

Adjusted Posttest Means

	Group	Mean	St Error
patterns 2	Sept	3.044	0.070
	Jan	3.044	0.076
numsen2	Sept	3.263	0.065
	Jan	3.259	0.070
commun2	Sept	3.099	0.070
	Jan	2.932	0.076
other2	Sept	3.172	0.064
	Jan	3.173	0.069

Endnotes

¹ The research was funded by the Ontario Ministry of Education and Training, the Social Sciences and Humanities Research Council, and an Ontario school district. The views expressed in the article do not necessarily represent the views of the Ministry, Council or school district.

² Students who were unable to read the booklet were given accommodations by the classroom teacher; e.g., the teacher read the booklet to the child. All children given accommodations were excluded from the data set. Reading skills are an unexamined source of student variance in the study. However, because of random assignment it is unlikely that there were differences between the experimental and control condition on reading skills.

³ Some of the teachers were in the project. However, they had no advantage because they did not mark their own classes, there were procedures to ensure inter-rater consistency, and the marking occurred after the posttests had been completed (i.e., we marked the pre and post tests at the same time).