
Research on Reform in Mathematics Education, 1993-2000

John A. Ross*
Douglas McDougall
Anne Hogaboam-Gray

Ontario Institute for Studies in Education
University of Toronto

*Corresponding author:

Dr. John A. Ross
Professor & Field Centre Head
OISE/UT Trent Valley Centre
Box 719, 633 Monaghan Rd
Peterborough, ON K9J 7A1
CANADA
Research on Reform in Mathematics Education, 1993-2000

Abstract

Proponents and opponents of reform of mathematics education each cite the research base in support of their positions. This article reports the results of a review of studies that contained empirical evidence of the effects of reform or the difficulty of implementing it that were published between 1993 and 2000. The studies reviewed indicate that implementation of math reform contributes to student achievement but evidence of superficial implementation and barriers to enactment abound. There are well-documented strategies for reducing these barriers, most promising strategies being in-service that simultaneously focuses on teachers’ practice and their cognitions about mathematics teaching.
Research on Reform in Mathematics Education, 1993-2000

When the California State Board of Education sought guidance from researchers about mathematics teaching and learning, E.D. Hirsch, Jr. provided an unambiguous answer: Research says that “only through intelligently directed and repeated practice, leading to fast, automatic recall of math facts, and facility in computation and algebraic manipulation can one do well at real-world problem solving” (Hirsch, cited by Becker & Jacob, 2000, p. 535). The Board greeted Hirsch’s summary with a standing ovation and included it in their rationale for dismantling reform initiatives enacted by their predecessors.

In this essay we will provide a different account of what research says by reviewing empirical studies of mathematics teaching reported in academic journals and conferences, 1993-2000. After describing our search procedures we will define math education reform, summarize the results of the review, and describe implications for educators and researchers. The main argument of the review is that the California State Board asked the wrong question. The issue is not whether reform in mathematics teaching contributes to student achievement (it does) but why implementation has been such a rare, fleeting occurrence and what can be done to support teacher efforts to change their practice.

Search Procedures

The review was the first step in a school improvement effort funded by the Ontario Ministry of Education that focused on grade 7-8 mathematics teaching (Impact Math). The central intervention strategy was the design and delivery of an in-service program to volunteer teachers in school districts across the province. The purpose of the review was to ground the in-service in empirical evidence by compiling answers from current research to three questions: 1) Does the implementation of reform in mathematics education contribute to improved student
achievement? 2) What are the barriers to implementing reform? 3) How can these barriers be overcome? The authors who conducted the review were sympathetic to the ideals of Standards-based reform (described below) and had previously conducted a number of research projects on strategies for implementing it.

Our immediate goal was to conduct a literature review that was systematic, reproducible, and explicit (Fink, 1998). We used a combination of manual and machine searches. We began by manually searching mathematics journals, general educational research journals that publish studies of mathematics learning, and academic conferences. The manual search identified keywords that were used in ERIC searches. The database was expanded through a final manual search (i.e., references cited by studies caught in the initial search).

We used three criteria to select studies for the review. First, the study had to contain empirical evidence, either quantitative or qualitative, of the effects of enacting education reform or data on implementation processes. We excluded reports that described but did not assess instruction, prescriptions for practice based solely on intuition and experience, policy statements unsupported by evidence, and theory development articles in which no original data were collected. Second, the study had to contain an overt strategy in which some aspect of reform was implemented; i.e., one or more of the ten dimensions listed on pp. 6-7 below (adapted from Ross, Hogaboam-Gray, McDougall, & Bruce, 2001-02). This excluded studies focusing on student or teacher characteristics that affect outcomes, unless the student or teacher attribute was included as a moderator of a treatment. Third, the search was limited to studies reported 1993-2000. Our rationale was that the latest round of reform began with the publication of the National Council of Teachers of Mathematics (hereafter NCTM) Standards in 1989. We estimated that that it took several years before the Standards were incorporated into field studies and subjected to rigorous
review. 134 studies met the inclusion criteria (listed in Ross, 2000); this set was expanded by 20 studies added in response to suggestions by reviewers.

**Coding Studies**

Studies were coded in terms of sample (size, grade, student and teacher demographics), theoretical framework, methodology (including instructional treatment and measurement instruments), results, and implications for **Impact Math.** In coding study quality we were particularly concerned to avoid errors observed in previous research syntheses, such as unexplained selectivity, author misrepresentations of findings, and unwarranted attributions by the reviewer of study conclusions (e.g., Dunkin, 1996; Guglielmi & Tattrow, 1998; Matt & Cook, 1994). Rather than discarding studies that were flawed, we coded design quality as a study attribute that increased or decreased confidence in the findings. We constructed a rubric, shown in Table 1, containing three levels based on credibility (for qualitative designs) and internal validity (for quantitative designs). Five coders independently reviewed an initial sample of five studies. Differences in interpretation of code categories were resolved through discussion. The remaining studies were coded by a single reviewer (60% by the lead author). We created a summary of each study organized around the coding categories. These summaries were used to create a narrative review organized around the three study questions. We opted for a narrative research synthesis rather than a quantitative meta-analysis because we did not want to exclude studies that lacked statistical information required for the calculation of effect sizes. In making methodological decisions we were guided by principles for reviewing qualitative studies developed by Schreiber, Crooks, & Stern (1997).
Table 1: Rubric for Judging Study Quality

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication</td>
<td>Not refereed</td>
<td>Peer reviewed conference paper</td>
<td>Peer reviewed journal article</td>
</tr>
<tr>
<td>Quantitative Design</td>
<td>Obvious flaws, e.g., non equivalent groups</td>
<td>Minor problems, e.g., non equivalent groups with statistical adjustment</td>
<td>Few problems, e.g., equivalent groups or statistical controls</td>
</tr>
<tr>
<td>Qualitative Design</td>
<td>No overt credibility procedures*</td>
<td>2-3 overt credibility procedures*</td>
<td>4 + overt credibility procedures*</td>
</tr>
</tbody>
</table>

*Credible qualitative designs include one or more of the following techniques: triangulation of data sources (i.e., compare data using different instruments or subjects), triangulation over time (i.e., compare data collected at different times), triangulation of observers (i.e., compare data collection by different observers), triangulation of interpretations (i.e., ask others to interpret data), member checks, accurate recording (e.g., audio recording), maintaining an audit trail (tracking themes from raw data), and rich description.

Characteristics of Reform

Reform in mathematics education is motivated by the finding that traditional teaching has produced low performance on basic competency tests (Romberg, 1997), the recognition that the world into which students will graduate requires greater ability to use mathematical tools (Bossé, 1995; Heid, 1997), and by advances in pedagogy that emphasize building on student prior knowledge, peer learning, and knowledge construction (Fennema, Franke, & Carpenter, 1993). No single set of attributes characterizes all reform initiatives but we can identify central tendencies that distinguish traditional from reform approaches.

The chief characteristics of math education reform emerging from the review and NCTM policy statements (1989; 1991; 2000) are (i) broader scope (e.g., multiple math strands with increased attention on those less commonly taught such as probability, rather than an exclusive focus on numeration and operations). (ii) All students have access to all forms of mathematics,
including teaching complex mathematical ideas to less able students. (iii) Student tasks are complex, open-ended problems embedded in real life contexts; many of these problems do not afford a single solution. In traditional math, students work on routine applications of basic operations in decontextualized, single solution problems. Leighton, Rogers, and Maguire (1999) suggested that formal (traditional) tasks differ from informal (reform) tasks in that formal hold all relevant information required to solve the problem (whereas informal require the solver to bring knowledge to the problem), are self-contained, provide a single correct answer, can be solved using conventional procedures, involve solutions that are unambiguous, entail topics that are of academic interest only and do not prepare students to solve real life problems. (iv) Instruction in reform classes focuses on the construction of mathematical ideas through student talk rather than transmission through presentation, practice, feedback, and remediation. (v) The teacher’s role in reform settings is that of co-learner and creator of a mathematical community rather than sole knowledge expert. (vi) Mathematical problems are undertaken in reform classes with the aid of manipulatives and with ready access to mathematical tools (calculators and computers), support not present in traditional programs. (vii) In reform teaching the classroom is organized to encourage student-student interaction as a key learning mechanism, rather than to discourage it as an off task distraction. (viii) Assessment in the reform class is authentic (i.e., analogous to tasks undertaken by professional mathematicians), integrated with everyday events, and taps a wide variety of abilities, in contrast with end of week and unit tests of near transfer that characterize assessment in traditional programs. (ix) The teacher’s conception of mathematics in the reform class is that of a dynamic (i.e., changing) discipline rather than a fixed body of knowledge. (x) Teachers in reform settings make the development of student self-confidence in mathematics as
Mathematics Education 1993-2000

important as achievement (Pajares, 1996 reported evidence that mathematics self-efficacy in the Junior grades was a better predictor of Senior math achievement than Junior math achievement).

These elements can be found in provincial guidelines. For example, the intended curriculum in Ontario (Ontario, 1997) includes all ten of these features, while omitting some elements often included in reform initiatives such as having students invent algorithms (e.g., Ball, 1993; Carroll, 1996). The Ontario curriculum also includes a feature not usually associated with math reform—detailed grade level expectations.

This list of reform characteristics is not an unorganized set of disembodied teaching behaviours. The dimensions overlap and constitute an orientation to instruction that differs fundamentally from traditional practice. Each teacher enacts these dimensions in unique ways. Yet there are patterns of excellence across teachers that make it possible to talk about a central tendency of reform teaching that is consistent within the subject and congruent with reform in other subjects (Sternberg & Horvath, 1995). For example, Borko and Putnam (1995) suggested that expert teachers have a cognitive mediational view of learning (i.e., that students relate incoming information to existing knowledge, impose meaning on experience, and monitor their learning processes) that translates into a constructivist approach to teaching. Although expertise in teaching in a given subject shares characteristics of excellence in teaching other subjects, subject-specific enactments differ, as demonstrated by studies examining the effects of out of subject assignments on teacher practice and cognitions about that practice (Ross, Cousins, Gadalla, & Hannay, 1999).

Research on the Effects of Reform on Students and Teachers

Research on math reform 1993-2000 took two paths. The first consisted of a relatively small number of studies of the effects of reform on student achievement. When reform was
implemented these effects were positive. Some of the most convincing evidence comes from qualitative studies that tracked teachers over several years as they elicited rich talk as students solved rich, meaningful problems in mathematical communities created in their classrooms. For example, Fennema et al. (1993) tracked a teacher over four years as she implemented Cognitively Guided Instruction, a program that focused on helping students construct deep understanding of mathematical concepts and strategies for solving problems embedded in their everyday experiences. Fennema et al. found that this exemplary teacher with a deep understanding of the structure of mathematics and children’s mathematical thinking had a profound effect on her grade 1 students. They solved more complex problems than other grade 1 pupils, used higher-level strategies, and adapted their procedures in response to problem requirements. They were knowledgeable about what they knew, had positive affect for the subject, persisted in problem solving when confronting by obstacles, and were fluent in describing their thinking.

Jo Boaler (1993; 1994; 1997; 1998) conducted an extensive longitudinal study of two schools in Britain, tracking students from age 12 to 16. In Phoenix, a school characterized by a commitment to math education reform, students worked in cooperative groups on 3-week projects, asked their teacher when they wanted input on math concepts (i.e., concepts were only introduced when needed), and classroom talk emphasized construction of student thinking. In contrast in the other school, Amber Hill, the program emphasized individual workbooks and textbooks. Classrooms were characterized by a search for correct answers rather than understanding, competition, individual work, and the transmission of algorithms and procedures. When given open-ended tasks, Phoenix students outperformed students in Amber Hill. Phoenix students were willing to derive meaning from the problem and they were able to select an appropriate procedure or adapt one to fit a new situation. In contrast the knowledge of Amber
Hill students was inert—they could not apply their knowledge. Boaler concluded that in Phoenix students learned how to use their knowledge (Boaler, 1998). Phoenix students performed more consistently, (i.e., they tended to use intuitive methods on all problems) and were enabled rather than distracted by contextual features. In contrast Amber Hill students were negatively influenced by superficial problem features and used standard school algorithms, regardless of their appropriateness—they were unable to transfer their knowledge (Boaler, 1993; 1994). Boaler also noted that student attitudes toward mathematics were consistently better in Phoenix, especially for females who also enjoyed a reduction in the gender achievement gap (Boaler, 1997).

These rigorously conducted qualitative studies are persuasive because of the readers’ knowledge of the effects of traditional mathematics programs. There is overwhelming evidence that such programs lead to mastery of basic algorithms without conceptual understanding. As Hiebert (1999) notes, old math is a proven failure. When we encounter evidence that Standards-based programs promote deep understanding it has an inter-ocular effect—it hits you between the eyes.

Several quantitative studies report that classrooms that provide rich tasks embedded in the real life experiences of children, with rich discourse about mathematical ideas, and a focus on children’s thinking, contribute to deeper understanding. For example, Cardelle-Elawar (1995) found that such a program contributed to superior grade 3-8 student performance on a problem solving measure. Brenner et al. (1997) reported that a program with similar features for grade 7-8 had positive effects on problem solving outcomes valued by reformers (such as the ability to represent mathematical relationships in multiple ways). The Core Plus Math Project that embodied Standards principles in curriculum materials increased secondary school student skills emphasized in reform agenda such as interpreting charts and tables and promoted deep
understanding of algebra and geometry concepts (Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000; Schoen, Fey, Hirsch, & Coxford, 1999). Silver and Stein (1996) found that a program that provided high-level problems to students produced growth in student understanding, reasoning and problem solving. Gains were larger in classes that implemented the tasks as intended (Stein, Gover, Henningsen, 1996). In contrast Ganter (1994) found that providing rich tasks to college students and structuring collaborative development of solutions had no effect on conceptual understanding, although there was a positive effect on student attitudes to mathematics.

Villasenor and Kepner (1993) found that children in classrooms that fully implemented math reform were also more successful on traditional math tasks, a finding reflected in international comparisons reporting higher achievement in countries that have adopted reform practices, such as Japan (Reys, Reys, & Koyama, 1996; Romberg, 1997). Schoen et al. (1999) also found a significant advantage for Core Plus Math students on a traditional algebra test, although the advantage did not endure beyond the first year of the study. Brenner et al., (1997) found no effects of Standards-based programming on a measure of traditional math objectives. Mayer (1998) found that on a traditional, multiple choice algebra test, Standards-based programs had a positive effect but only for higher ability students. Although the results of these studies are mixed with some reporting no significant advantage for Standards-based programs, there are no studies that show such programs producing results lower than those attained through traditional mathematics instruction.

There is also evidence that reform contributes to the achievement of disadvantaged students (Silver & Stein, 1996), as well as those of average ability, although there is limited
evidence that lower status students might be disadvantaged in reform classrooms (Lubienski, 2000).

In summary, students in classrooms characterized by mathematics education reform have higher achievement on achievement measures emphasized by reformers such as problem solving and conceptual understanding, have more positive attitudes toward the subject, and at least have no less achievement on objectives emphasized by traditional programs such as computational efficiency.\(^4\)

But math reform is difficult to implement. The second set of studies, larger than the first, focused on evidence of non-implementation and barriers to enactment. Even teachers chosen as exemplars of reform practice regress from the ideal, displaying the height of reform one day but regressing to traditional methods the next (Senger, 1998). Some elements of reform are more difficult to implement than others. The most challenging is the management of student talk about mathematical reasoning—including finding the right balance between encouraging student constructions without leaving them floundering (Ball, 1993; Ross & Cousins, 1995a; 1995b; Ross, Haimes, & Hogaboam-Gray, 1996; Smith, 2000; Williams & Baxter, 1996).

The catalogue of barriers to reform is a lengthy one. Among the most important are: Teachers must be agents of a change they did not experience as students (Anderson & Piazza, 1996). The pedagogy is not only different but also harder to learn. For example, in traditional math there is a generic script that guides each day’s lesson through a manageable body of content. In reform math the day is governed by unpredictable student responses to real life problems. Teachers, especially elementary generalists, tend to lack the disciplinary knowledge required to make full use of rich problems (Henningsen & Stein, 1997; Lloyd & Wilson, 1998; Mandeville & Liu, 1997; Monk, 1994; Phillip, Flores, Sowder, & Schappelle, 1994; Stein et al.,
Mathematics Education 1993-2000

1996; Spillane, 2000) and texts cannot prescribe universally applicable courses of action (Remillard, 2000). Adoption of reform math can leave teachers feeling less efficacious because their contribution to student learning is less visible than in traditional teaching (Ross, McKeiver, & Hogaboam-Gray, 1997; Smith, 1996). Teacher beliefs about mathematics (i.e., a rigid set of algorithms, not understandable by most students, that must be approached in an inflexible sequence) conflict with reform conceptions of math as a fluid, dynamic set of conceptual tools that can be used by all (Gregg, 1995; Prawat & Jennings, 1997). Reform does not meet parental expectations about how math should be taught and tested (Graue & Smith, 1996; Lehrer & Shumow, 1997). Reform conceptions of mathematics conflict with mandated assessment programs that measure computational speed and accuracy (Firestone, Mayrowetz, & Fairman, 1998). Time to cover the curriculum is a major challenge. Keiser and Lambdin (1996) found that student constructions took longer than lecture-recitations, novel problems increased time taken for discussion of homework, and students with poor motor skills took longer to use manipulatives than anticipated.

Reducing Barriers to Implementation

There have been many suggestions for increasing classroom applications of reform ideals (e.g., policy development, preservice training, materials development, alignment of assessment with instruction, etc.).

The most powerful method for increasing implementation is in-service. In reviewing similarities between the current round of math reform and the New Math movement of the 1950/60s that ultimately failed to influence teacher practice, Bossé (1995) noted that inattention to teacher in-service was the key deficiency of both movements. The Standards of Practice (NCTM, 1989) anticipated that teachers would be able to develop materials and practices to enact
the vision of reform with little support. Experience since then demonstrates that it is essential to provide ongoing professional development, particularly focused on providing teachers with examples of constructivist teaching (Bitter & Hatfield, 1994) and explicitly addressing their beliefs about mathematics as a teachable subject (Grant, Peterson, & Shojgreen-Downer, 1996). The delivery sequence (i.e., changes in beliefs before adjustment of practices or vice versa) appears not to matter: what is essential is that in-service contain both components (Borko, Davinroy, Bliem, & Cumbo, 2000). Especially important is public and private reflection. Sharing professional experiences is such an essential element of professional growth that it has become axiomatic that in-service events should provide opportunities for participants to describe their experiences, reflect on the meanings of personal practice, and exchange interpretations with colleagues (e.g., Fullan & Connelly, 1990; Grimmett & Erickson, 1988; Kemessis, 1987). Evidence of the positive effects of such in-service on teacher implementation of math education reform and student achievement is accumulating (e.g., Knapp & Peterson, 1995; Pligge, Kent, & Spence, 2000; Schifter & Simon, 1992; Smith, 2000). There is also evidence that provision of new curriculum materials, in the absence of sustained in-service, has little impact on teacher implementation (Price, Ball, & Luks, 1995; Roulet, 1998).

Another promising approach to reducing barriers to implementation of reform in mathematics education is through integration with technology. There is ample correlational evidence that teachers who are more frequent users of technology (calculators, computers) are more likely to adopt even the most difficult dimensions of reform such as constructivist teaching (Becker, 1998; Waxman & Huang, 1996). Provision of software within a reform curriculum contributes to teacher implementation of the Standards (Huetinck, Munshin, & Murray-Ward, 1995; Ross, Hogaboam-Gray, & McDougall, 2000). Student achievement increases when
calculators (Hembree & Dessart, 1992) and computers (Christmann, Badgett, & Lucking, 1997; Heid, 1997) are used.

What is less clear about the integration of computers is how it contributes to reform. The relationship may be spurious: good teachers tend to adopt the innovations of the day, in this case technology integration and math reform (Becker, 1998). More likely is that technology enables teachers to implement their constructivist beliefs by relieving students of the tedium of calculation and providing them with visual representations to support dialogue about mathematical ideas. Some researchers (e.g., Sandholtz, Ringstaff, & Dwyer, 1997) have argued that technology demands that teachers change to a constructivist orientation because teachers have to share control with students in a computer-based learning environment. The contribution of technology to math reform is not automatic. Providing computers and software to teachers without appropriate in-service has minimal effect on teacher practice (Robertson, Calder, Fung, Jones, O'Shea, & Lambrechts, 1996).

A less promising strategy, frequently advocated, is curriculum alignment—of assessment and curriculum or curriculum integration across subjects. It is argued that schools will improve if governments set clear standards for students and teachers, assess the extent to which standards are met using curriculum aligned tests, and provide schools with feedback (Teddlie & Reynolds, 2000). The credibility of the argument is threatened when raw test scores are used. The only Canadian study to report the effect of school population factors on provincial test scores (Lytton & Pyryt, 1998) found that none of the variance in Grade 3 and Grade 6 mathematics scores could be attributed to instructional factors. Mandated testing programs provide some support for reform in mathematics teaching when the tests reflect reform learning goals. State tests influence teachers’ choice of content, although not their instructional strategies (Firestone et al., 1998) and
have only modest impact on achievement (Shepard et al., 1996). Positive effects have been observed when mathematics and science curricula have been aligned around problem-based units (Austin, Hirstein, & Walen, 1997; Ross & Hogaboam-Gray, 1998) or around the structure of mathematics (Isaacs, Wagreich, & Gartzman, 1997).

Several studies investigated the impact of restructuring on implementation of reform mathematics. Changes in teaching strategy to improve student-student communication and female achievement in mathematics have been reported when single gender classes or schools were created (e.g., Parker & Rennie, 1997) but most of these studies were methodologically flawed (Mael, 1998). Teacher knowledge and use of reform practices increase when schools established partnerships with outside agencies, particularly NCTM (e.g., Mills & Garet, 1996 found that department head membership in NCTM increased reform implementation), universities that provide in-service on reform practices (e.g., Ross, 1995a; Borko, 1997; Brahier, 1998), and school networks (Hernandez-Gantes & Brendefur, 1997). Site-based management, in which instructional decision making at the school level is shared with teachers, also contributes to curricular change (Wagstaff, 1995). Each of these studies identified restructuring as a key change element. A more likely explanation is that restructuring stimulated collaboration among teachers that led to instructional innovation. Ross et al. (1997) found that teacher collaboration contributed to implementation of mathematics education reform. The least experienced teacher in this qualitative study benefited the most from collaboration because it reduced her workload, clarified expectations for content coverage, and set the pace of instruction. Other teachers benefited from peer emotional support, identification of new teaching strategies, and workload sharing. The contribution of collaboration to implementation of math reform was observed by in
other studies (Feikes, 1998; Moreira & Noss, 1995; Ponte, Matos, Guimaraes, Leal, & Canavarro, 1994).

A large number of investigations reported during the time period of the review addressed new ways of teaching mathematics within the reform classroom. Most of these instructional improvement approaches displayed a high degree of rigor, although the degree of reform implementation could not always be determined. For example, several studies focused on demonstrating that cooperative learning techniques contribute to achievement of reform ideals (Kiczek & Maher, 1998; Mulryan, 1995; Slavin & Madden, 1999; Whicker, Bol, & Nunnery, 1997). However, some researchers found evidence in mixed ability groups of passivity on the part of lower ability students in mathematical discussions and dysfunctional responses to their learning needs on the part of higher ability students (King, 1993; Ross, 1995a; 1995b). A major theme in this literature was the search for ways to make cooperative groups more effective. Some studies focused on grouping strategies, finding that in mathematics class, homogenous ability grouping is preferable for complex problems (Fuchs, Fuchs, Hamlett, & Karns, 1998) but only if students have different bodies of knowledge to draw upon to solve problems (Mevarech & Kremarski, 1997). Other studies focused on procedures for improving the quality of discourse in student groups. Researchers found that training students how to give explanations had a positive effect on mathematics achievement (Fuchs, Fuchs, Karns, Hamlett, Katzaroff, & Dutka, 1997; Hoek, van den Eeden, & Terwel, 1999; Nattiv, 1994) especially when the training was focused specifically on how to give mathematical explanations (Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997).
Implications of Research on Mathematics Education Reform

Research reported in 1993-2000 found, first, that reform in mathematics education contributed to higher student achievement. Although the number of studies that have investigated achievement effects is relatively small, the studies reviewed in the achievement section of this review were of high quality with all of them reaching at least level 2 in the quality rubric provided in Table 1 and the majority reaching level 3. These positive results were attained only when there was substantial implementation of reform, a rare event. There was evidence of unintended variation in treatments within districts, schools, and teachers and some reform elements were more difficult to implement than others. The key implication for math educators is to recognize that, despite the outbreak of the "mathematics wars" in many countries, the research base encourages teachers and schools to implement the Standards. We found no evidence that warrants a regression to past practices as implied by the Hirsch quotation at the beginning of the review. The research also indicates that progress toward implementing reform ideals will be incremental, with advances occurring on a broken front with many backward steps.

Measuring teacher change will be problematic. There is no consistent image of what reform should look like in the classroom and even less consensus around how it should be measured. The Standards are an accumulation of the visions of its writers. Although the philosophy behind the Standards is appropriately described as constructivist, this was a label assigned by Thomas Romberg to gain political support after the Standards were written (Bossé, 1995). It is unclear how the dimensions of reform should be weighted. Although there have been attempts to describe levels of implementation on particular reform dimensions (Bright, Bowman, & Vace, 1998; Franke, Carpenter, Levi and Fennema, 2001; Gabriele et al., 1999; Hall, Alquist, Hendrickson, & George, 1999; Lambdin & Preston, 1995; McDougall et al., 2000; Nolder &
Johnson, 1995; Ross et al., 2000; Ross, Hogaboam-Gray, & McDougall, 2001; Slavit, 1996; Spillane and Zeuli, 1999) and suggestions for conceptual tools (such as discourse analysis) for distinguishing levels of classroom practice (Blanton, Berenson, & Norwood, 2001), no overall rubric has been created that has broad approval. What researchers need to provide is: (i) a rubric for guiding the generation of instruments, (ii) a self-report survey for tracking the progress of large groups of teachers, and (iii) coding schemes for observing classrooms for use in qualitative studies.

Second, the studies provided consistent evidence of the barriers to reform. Although the quality of the studies reviewed in this section of the essay was more variable, study quality was randomly distributed over the topics reviewed (i.e., none of the claims made in this section was based on a corpus of studies that was so consistently flawed as to introduce systemic bias). The most important obstacle is that teacher beliefs and prior experiences of mathematics and mathematics teaching are not congruent with the assumptions of the Standards. Teachers mostly support the goals of reform but overestimate the extent to which their practices approach these goals. The innovation is ambiguous and difficult to implement. The lack of accessible examples impedes the development of local visions, implementation is costly in terms of classroom and teacher preparation time, substantive change in practice threaten teachers' beliefs about their efficacy, and the complexity of student tasks is prone to diminution. Realization of reform ideals is also thwarted by policy misalignments, the most important being competition with other innovations and conflicts with mandated student assessment programs. The key implication for reformers is to encourage modesty in expectations about impact and to anticipate widespread variation in the use of reform ideas.
Third, the research identified promising strategies for overcoming barriers to reform. The most powerful mechanism is professional development. What is needed is sustained interaction of classroom teachers with professional development leaders external to the school and the provision of local support such as district/school consultants. There needs to be a dual emphasis on new classroom strategies while attending to teacher cognitions about their existing practice. There should also be a dual focus on developing teachers' disciplinary knowledge (knowing the subject) and their pedagogical content knowledge (i.e., knowing how to present mathematical content to students and being able to anticipate and respond to student misconceptions about the material to be learned). In making the argument for a professional development focus, we are mindful that in-service alone is insufficient to bring about teacher change, even though research to date indicates it is the most powerful. Schoenfield (2001), for example, suggests four other essential conditions: high quality curriculum aligned with the Standards; a stable, knowledgeable and professional teaching community; high quality assessment aligned with Standards; and a fine balance between stability and mechanisms for evolution.

Finally, a variety of successful instructional experiments have been reported. The most promising involve strategies for teaching students how to talk about mathematics in cooperative learning settings. As experiments flourish, researchers in partnership with teachers will begin to realize the promise of reform by addressing the unresolved teaching issues identified by Gutstein and Romberg (1995), such as how algorithms can be taught in a meaningful way while maintaining a commitment to student inventions.
REFERENCES


Pligge, M. A., Kent, L. B., and Spence, M. S. (2000). *Examining teacher change within the*


Reference Notes

1 This review was prepared for the *Impact Math* project funded by the Ontario Ministry of Education and Training. Ann Kjander and Alex Lawson reviewed some of the studies. A longer version of the review is available in Ross et al., 2000.

2 The keywords in the first search, focused on implementation issues, were: mathematics with educational change, educational innovation, professional development, program implementation, reform efforts, and large scale programs. In the second search, focused on effects of reform on achievement, the keywords were: mathematics with achievement, education, instruction, and skills.

3 We also coded the studies in terms of intended outcomes (categorized in terms of the four criteria in the provincial mathematics rubric) and domain of mathematics (the five strands specified in provincial curriculum guidelines) but insufficient information was provided in most studies so these codes were not used.

4 All of the studies reviewed in this section received high scores on the rubric for study quality. A few studies with less rigorous procedures produced comparable results. For example, Simon & Schifter (1993) found that students exposed to a *Standards*-based program had deeper understanding, greater facility in communicating mathematical ideas, more positive attitudes to the subject, but there were no gains on standardized test scores. However, this study was level 1 quality. It provided no reliability or validity information on the measures used, data from different standardized tests were pooled, there was a heavy reliance on teacher self-reports, it employed a pre-post cohort design without controls. The study did not report descriptive data such as means or statistical procedures, used grade-equivalent scores in statistics rather than raw scores, and analyzed each survey item separately without Bonferroni adjustments for multiple comparisons.