ISSN 0021-8251

Journal for Research in Mathematics Education

Volume 33
Number 2

March 2002

National Council of Teachers of Mathematics

NCTM

Editorial Board

Editor

EDWARD A. SILVER, University of Michigan

Associate Editors

KAREN D. KING, Michigan State University
VILMA MESA, University of Michigan

Assistant Editor

PATRICIA ANN KENNEY, University of Michigan

Book Review Editor

ANNA O. GRAEBER, University of Maryland

Editorial Panel

BETSY BRENNER, University of California, Santa Barbara; JINFA CAI, University of Delaware; MARTA CIVIL, University of Arizona; GUERSHON HAREL, University of California, San Diego; KATHY IVEY, Western Carolina University, *Chair;* LENA LICON KHISTY, University of Illinois at Chicago; CAROLYN KIERAN, Université du Québec à Montréal, *Board Liaison;* RICH LEHRER, University of Wisconsin—Madison; NORMA PRESMEG, Illinois State University; MARTIN SIMON, Penn State University

Monograph Series Editor

NEIL PATEMAN, University of Hawaii

Journal Staff

HARRY B. TUNIS, Director of Publications; DAVID BARNES, Director of Electronic Resources Initiative; KRISTA HOPKINS, Director of Marketing; JEAN T. CARPENTER, Publications Services Manager; TOM PEARSON, Advertising Manager; SANDRA S. BELSLEY, Advertising Assistant

NCTM Board of Directors

LEE V. STIFF, North Carolina State University, *President;* JOHNNY W. LOTT, University of Montana, *President Elect;* JAMES M. RUBILLO, *Executive Director;* MARY BUCK, Helena Middle School, Montana; CINDY CHAPMAN, Inez Science and Technology Magnet Elementary School, New Mexico; SUSAN K. EDDINS, Illinois Mathematics and Science Academy; JUDITH JACOBS, California State Polytechnic University, Pomona; CAROLYN KIERAN, Université du Québec à Montréal; FRANK K. LESTER, JR., Indiana University; CAROL E. MALLOY, University of North Carolina at Chapel Hill; BEATRICE M. MOORE-HARRIS, Bureau of Education and Research, Texas; MARK SAUL, Bronxville Schools, New York; J. MICHAEL SHAUGNESSY, Portland State University, Oregon; JUDITH SOWDER, San Diego State University; BERT K. WAITS, Ohio State University

The **mission** of the National Council of Teachers of Mathematics is to provide the vision and leadership necessary to ensure a mathematics education of the highest quality for all students.

The ***Journal for Research in Mathematics Education,*** an official journal of the National Council of Teachers of Mathematics, is devoted to the interests of teachers of mathematics and mathematics education at all levels—preschool through adult. The *JRME* is a forum for disciplined inquiry into the teaching and learning of mathematics. The editors encourage the submission of a variety of manuscripts: reports of research, including experiments, case studies, surveys, philosophical studies, and historical studies; articles about research, including literature reviews and theoretical analyses; brief reports of research; critiques of articles and books; and brief commentaries on issues pertaining to research. An **Index** for each volume appears in the November issue. The *JRME* is indexed in *Contents Pages in Education, Current Index to Journals in Education, Education Index, Psychological Abstracts, Social Sciences Citation Index*, and *Zentralblatt für Didaktik der Mathematik.*

Subscriptions: The *Journal for Research in Mathematics Education* (ISSN 0021-8251) is published five times a year, January, March, May, July, and November, at 1906 Association Drive, Reston, VA 20191-9988, by the National Council of Teachers of Mathematics. An individual NCTM membership with a *JRME* subscription, print or online, costs $90.00 (U.S.) per year. Dues include $50.00 for the *JRME* subscription and $16.00 for a 9-issue subscription to the NCTM *News Bulletin* (U.S.). Foreign subscribers add $18.00 (U.S.) for the first journal and $4.00 (U.S.) for each additional journal. Single copies cost $9.50 (U.S.). Special rates for students, institutions, bulk subscribers, and life and retired members, as well as information about the other three official journals, *Teaching Children Mathematics, Mathematics Teacher,* and *Mathematics Teaching in the Middle School,* are available from the Headquarters Office.

Correspondence: Editorial correspondence and manuscripts should be sent to Edward A. Silver, *JRME,* School of Education, University of Michigan, 610 E. University Avenue, Ann Arbor, MI 48109-1259. Review copies of new books and other publications should be sent to Anna O. Graeber, 2311 Benjamin Building, University of Maryland, College Park, MD 20742. Manuscripts for the monograph series should be sent to Neil Pateman, Wist Annex 2-223, Department of Teacher Education and Curriculum Studies, University of Hawaii, 1776 University Avenue, Honolulu, HI 96822. All other correspondence should be addressed to NCTM. Telephone: (703) 620-9840; orders only: (800) 235-7566; fax: (703) 476-2970; Fax on Demand: (800) 220-8483; e-mail: nctm@nctm.org; World Wide Web: **www.nctm.org.**

Submissions: Six copies of each manuscript should be submitted. Manuscripts should be typed double-spaced, with the author's name and affiliation on a separate page to ensure anonymity in the reviewing process. The expected length of manuscripts is 30 to 40 pages, not counting cover page, abstract, references, tables, and figures. Longer manuscripts will be reviewed at the discretion of the editor. Research reports and articles should begin with an abstract of about 100 words on a separate page. Manuscripts should conform to the style specified in the *Publication Manual of the American Psychological Association.* For additional details on the submission of manuscripts, see the *JRME* Web page at **www.nctm.org/jrme/jrme.html**.

Permission to photocopy material from the *JRME* is granted to persons who wish to distribute items individually (not in combination with other articles or works) for educational purposes, in limited quantities, and free of charge or at cost; to librarians who wish to place a limited number of copies of items on reserve; to authors of scholarly papers; and to any party wishing to make one copy for personal use. Permission must be obtained to use journal material for course packets, commercial works, advertising, or professional development purposes. Uses of journal material beyond those outlined above may violate U.S. copyright law and must be brought to the attention of the National Council of Teachers of Mathematics. For a complete statement of NCTM's copyright policy, see the NCTM Web site, **www.nctm.org.**

The publications of the National Council of Teachers of Mathematics present a variety of viewpoints. The views expressed or implied in this publication, unless otherwise noted, should not be interpreted as official positions of the Council.

© 2002 by The National Council of Teachers of Mathematics, Inc. Printed in the U.S.A. Periodicals postage paid at Herndon, Virginia, and additional mailing offices. POSTMASTER: Send address changes to the *Journal for Research in Mathematics Education,* 1906 Association Drive, Reston, VA 20191-1502.

Journal for Research in Mathematics Education

March 2002, Volume 33, Number 2

EDITORIAL

ARTICLES

Editorial

Education Research and Education Policy: Be Careful What You Wish For!

Edward A. Silver

Throughout the world, members of the education research community have long decried the lack of attention to research and scholarship that is evident in most education policies promulgated by state, provincial, and national legislatures. The Elementary and Secondary Education Act of 2001 (ESEA 2001), which was recently signed into law in the United States, is an interesting case to consider in this regard. Because ESEA 2001 states that education programs and policies should rely on "scientifically based research," it appears at first glance to offer a welcome counterexample to the longstanding tendency of policy professionals to ignore research when setting education policy. Unfortunately, I fear that this new development may be a good example of the kind of outcome my grandmother had in mind when she used to say, "Be careful what you wish for, because you just might get it!" Readers of *JRME* both inside and outside the United States may find it interesting to contrast the case of ESEA 2001 with other instances of education policy with which they are familiar.

The Elementary and Secondary Education Act of 2001 and its accompanying multibillion dollar budget will have an impact on virtually every public school system in the United States. The U.S. mathematics education community has voiced considerable concern about one aspect of ESEA 2001—the elimination of the Dwight D. Eisenhower Professional Development Program, which has for many years provided funds for the continuing education of mathematics and science teachers (Hoff, 2002; Morgan, 2002). Yet, there are other aspects of ESEA 2001 that merit our attention and concern as well.

A key feature of the legislation is a requirement that by the 2005–06 school year each child in Grades 3–8 must take annual tests in mathematics and in reading. Students' proficiency must be monitored by means of these tests, with results reported as individual scores and also as performance profiles for subgroups based on demographic indicators (e.g., race, family income). Schools where students fail to meet established standards in mathematics and in reading will receive some assistance and also be subject to some sanctions. If the pattern of failure persists for several years, more severe sanctions and "corrective actions" will be instituted. Because of these features, ESEA 2001 is called the "No Child Left Behind" Act.

The emphasis on the continuous progress of all students toward mathematics proficiency may appear on the surface to be consistent with *Principles and Standards for School Mathematics* (2000). After all, the Equity Principle articu-

Copyright © 2002 The National Council of Teachers of Mathematics, Inc. www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in any other format without written permission from NCTM.

lated in *Principles and Standards* asserts the fundamental importance of striving toward excellence in mathematics education by setting high expectations for all students and providing strong support for students to meet them. However, a closer look reveals some reasons to be less sanguine about the intent and likely impact of ESEA 2001.

Despite the stated requirement in ESEA 2001 that programs and policies be based on sound research, the framers of this legislation appear themselves to have created a set of policies that are at variance with the research-based recommendations of two scholarly panels convened by the National Research Council (NRC) of the National Academy of Sciences. One of these panels examined and synthesized what is known about mathematics learning in elementary school, and the other probed fundamental theoretical and design issues associated with educational assessment.

The recommendations of one of these NRC panels—comprising mathematicians, psychologists and mathematics education researchers—can be found in *Adding It Up* (Kilpatrick, Swafford, & Findell, 2001). A key contribution of this report is a multifaceted definition of mathematical proficiency as composed of five strands—procedural fluency, conceptual competence, strategic competence, adaptive reasoning, and productive disposition. On the basis of a thorough review of the extant research literature, the panel concluded that students are likely to develop mathematical proficiency only if instruction in elementary school mathematics attends to all strands.

Although ESEA 2001 does not explicitly prohibit the use of this research-based definition of mathematical proficiency as the basis for the assessment of achievement, it also neither requires nor suggests it. Moreover, there is little reason to think that this definition will guide the implementation of the legislation. Given the constraints of time and money that such legislation inevitably entails, how likely is it that individual states, or the testing companies that they hire, will develop annual tests in Grades 3–8 that will measure students' progress along the five strands of mathematical proficiency identified by the NRC panel? In the flood of testing mandated by ESEA 2001, *proficiency* is far more likely to be characterized in a much more limited way and to be restricted almost exclusively to the recall of factual information and the accurate performance of numerical and algebraic procedures.

The research-based recommendations of another NRC panel were also apparently ignored in the framing of the ESEA 2001 legislation. This panel—comprising experts in fields that pertain to educational assessment—reviewed a variety of psychological and psychometric theories and surveyed existing research and design work in educational assessment. The panel's report, *Knowing What Students Know* (Pellegrino, Chudowsky, & Glaser, 2001), called for greater integration between instruction and assessment. Moreover, the panel argued that meaningful educational assessment must be carefully designed and conducted, using multiple methods and measures whenever possible, and must be instructionally sensitive and guided by a model of student proficiency that is based on a solid conception of the domain being assessed. The panel found great promise in many recent approaches to educational testing, such as performance assessments, portfolios, and other attempts to develop multiple indicators of student performance and progress.

In contrast, the ESEA 2001 legislation calls for a single measure of proficiency—a score on a standardized test administered annually. According to Richard Rothstein, the senior education columnist for the *New York Times*, the shift in language from *performance*—the word used in the previous version of ESEA—to *achievement* in ESEA 2001 signals the apparent intent of some of the framers of the current legislation to abolish all forms of educational assessment except standardized achievement tests (Rothstein, 2002). Because of this narrow focus on annual testing in Grades 3–8, ESEA 2001 might be dubbed the "No Child Left Untested" Act.

Although the framers of ESEA 2001 appear to have disregarded the sage advice of these NRC panels, it may yet be possible to bring the recommendations of these reports to the foreground during the implementation of the legislation. Many details remain to be worked out in the design of procedures for carrying out the legislation.

I began by noting that ESEA 2001 calls for greater reliance on scientific research. It is regrettable that the framers of this legislation may have had in mind a rather narrow conception of education research, as suggested by the evident gaps between the legislative mandates and the research-based, scholarly recommendations of NRC panels. In recent years, some critics of education research have argued that studies with randomized experimental designs should be the only *scientific* basis for educational decisions. Is this the implicit view of ESEA 2001? If so, then even in a call for more attention to education research, ESEA 2001 would be ignoring yet another recent NRC report—*Science, Evidence, and Inference in Education* (Towne, Shavelson, & Feuer, 2001).

This document, which reports on the work of the NRC Committee on Scientific Principles in Education Research—a multidiscplinary team of experts in psychology, education, and other social sciences—offers a valuable alternative view. The committee considered the nature of research in education and asked how it could be defined and conducted in order to ensure credibility and scientific progress through the accumulation of useful and useable knowledge. The report argues that randomized experiments are not the only form of scientifically based research in education. According to the committee, sound research in education requires that researchers pose significant questions and link them to relevant theory, employ appropriate methods and tools to answer the questions, examine carefully the warrants for their claims, consider counterevidence and alternative arguments, and subject their work to scrutiny through peer review. I hope that this broader conception of education research will be adopted as ESEA 2001 is implemented.

Readers of the *Journal of Research in Mathematics Education* will recognize that this view of high-quality research is embodied in the policies, practices, and content of the journal. As educators interpret the call of ESEA 2001 for programs and policies that rely on scientifically based research, *JRME* should be viewed as a valuable source of research and information for those seeking guidance regarding mathematics teaching and learning.

ESEA 2001 also challenges us in new ways. We need to examine the experiences of researchers and policymakers in countries where large-scale education improvement efforts have been based on research. We also need to frame and disseminate research in mathematics education in ways that help policymakers understand and use research and scholarship as they shape legislation. Other challenges undoubtedly await us in the future. But for now, let's just be careful what we wish for!

REFERENCES

Hoff, D. J. (2002, January 16). Math and science could be big losers under new law. *Education Week*, pp. 20, 23.

Kilpatrick, J., Swafford, J. & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Morgan, R. (2002, January 18). Congress authorizes little money for math and science instruction. *The Chronicle of Higher Education*, p. A25.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Pellegrino, J., Chudowsky, N & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rothstein. R. (2002, January 16). The weird science of the education law. *The New York Times* (national edition), p. A18.

Towne, L., Shavelson, R. J., & Feuer, M. J. (Eds.). (2001*). Science, evidence, and inference in education*. Washington, DC: National Academy Press.

Journal for Research in Mathematics Education
2002, Vol. 33, No. 2, 78–110

The Interpretative Nature of Teachers' Assessment of Students' Mathematics: Issues for Equity

Candia Morgan
Institute of Education, University of London, United Kingdom

Anne Watson
University of Oxford, United Kingdom

This paper discusses fairness and equity in assessment of mathematics. The increased importance of teachers' interpretative judgments of students' performance in high-stakes assessments and in the classroom has prompted this exploration. Following a substantial theoretical overview of the field, the issues are illustrated by two studies that took place in the context of a reformed mathematics curriculum in England. One study is of teachers' informal classroom assessment practices; the other is of their interpretation and evaluation of students' formal written mathematical texts (i.e., responses to mathematics problems). Results from both studies found that broadly similar work could be interpreted differently by different teachers. The formation of teachers' views of students and evaluation of their mathematical attainments appeared to be influenced by surface features of students' work and behavior and by individual teachers' prior expectations. We discuss and critique some approaches to improving the quality and equity of teachers' assessment.

Key Words: Assessment; Communication; Equity/diversity; Reform in mathematics education; Social and cultural issues; Teaching practice

Reform of assessment methods and regimes is currently a concern in many countries around the world, and there is much discussion about the design of assessment tasks and systems that will provide valid and useful information for a variety of purposes, from immediate feedback to teachers and individual students to large-scale monitoring and evaluation of educational systems. In mathematics education, as in other subject areas, reformers have focused on the development of authentic or performance-based assessment of a broad spectrum of students' mathematical performance (e.g. Lesh & Lamon, 1992; Romberg, 1995; van den Heuvel-Panhuizen, 1996), often proposing substantial involvement of teachers as assessors of their own students (Black & Wiliam, 1998). The potentially positive influence of such assessment systems on the curriculum has also motivated reform (Bell, Burkhardt, & Swan, 1992; Stephens & Money, 1993). Although we are broadly in sympathy with the aims of such reform, we wish to raise some concerns about the nature of teachers' assessments of their students and the potential consequences for equity.

This paper arises from the United Kingdom[1] context in which alternative forms of assessing students' mathematical progress, including assessment by teachers, have been practiced alongside traditional methods for over twelve years. First, we review the ways that issues of equity and fairness are addressed in the literature on teachers' assessments in mathematics. We then develop the view that all assessment of mathematics is interpretative in nature and examine the classroom reality of informal assessment and the difficulties that arise in teachers' formal assessment of extended written mathematical tasks. Theoretical arguments are illustrated by reference to two independent empirical studies of teachers' assessment practices. Finally, we discuss some approaches to improving the quality of teachers' assessment practices.

Our concern with equity in assessment arises from the fact that the assessments made of students at all levels of education can have far-reaching consequences in their future lives. This is most obviously true in high-stakes assessments like the General Certificate of Secondary Education (GCSE) examinations in England. The results of such assessments are used explicitly for the differentiated allocation of curriculum tracks and further educational and employment opportunities, thus guiding students into different life paths. Similar uses are also made, however, of even the most fleeting and informal judgment made by a teacher of a student in the course of everyday work in the classroom. Although such assessments may not be formally recorded or even reflected upon, they nevertheless play a part in forming the educational future of individual students, often contributing to decisions about differentiation of the available curriculum. Such judgments affect the teacher's short-term behavior towards the student (e.g., the feedback provided, or the next task set) and influence the ways in which the teacher is likely to interpret the student's future performance (Walkerdine, 1988). Judgments that are based on unsound assessments can lead to unfair or unjust decisions and consequent inequitable treatment and opportunities for students. Such consequences of assessment operate in addition to systemic inequities in allocation of resources, an issue beyond the scope of this paper.

The issue of equity in assessment has generally been addressed from the point of view of attempting to ensure that all students are given equal opportunities to display their achievements and that assessment instruments do not have any systematic bias against particular social groups (see Gipps & Murphy, 1994, for a thorough review and discussion). Powerful critiques of some traditional methods of assessment in mathematics have identified the inequity inherent in them as well as the poor quality of the information they provide (Burton, 1994; Niss, 1993; Romberg, 1995). However, alternative assessment methods will not necessarily reduce inequity. Winfield (1995) identifies a number of aspects of alternative

[1] The curricula, assessment regimes, and approaches to teaching in the various countries that make up the United Kingdom are very similar. However, they are subject to different regulations. The studies reported in this article took place in England.

Copyright © 2002 The National Council of Teachers of Mathematics, Inc. www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in any other format without written permission from NCTM.

assessments that can affect equity, including the relationship between the assessment and the instructional conditions and possible mismatches between the expectations of assessment tasks and the ways in which learners draw on their cultural resources as they interpret them. Moreover, Baker and O'Neil (1994) report the concern of minority community groups about the potential for increased inequity with the introduction of performance assessments because of the unfamiliarity of the cultural contexts in which tasks may be set, a concern also raised by researchers in relation to social class (e.g., Cooper & Dunne, 2000). In response to these concerns, reformers have proposed multidimensional forms of assessment to allow all students to demonstrate what they know and can do as well as to ensure that the full range of mathematical objectives are addressed (de Lange, 1995; National Council of Teachers of Mathematics [NCTM], 2000). It has also been suggested that students' own teachers are in the best position to ensure that assessment is equitable because "probing what students are thinking, being sensitive to their experiences, and understanding how they perceive the assessment situation all contribute to making equitable decisions about students' learning." (NCTM, 1995, p. 15). Reformers argue that teachers can accumulate a multidimensional (and hence more valid) view of students' mathematics by assessing over time, using responses to a variety of tasks in several situations, and having knowledge of the context within which the student is working.

Our concern in this article is with inequity that arises not from the nature of assessment tasks but at the point of interpretation of student performance by teachers. This inequity may be comparative, in the sense that two students with similar mathematical understanding may be evaluated differently, or noncomparative, in the sense that a student who has achieved a particular form of mathematical understanding is not recognized to have done so. Although the focus of this view of inequity appears to be at the level of individual teachers' assessment of individual students, there is of course a strong possibility that some groups of students will be systematically disadvantaged because the behaviors they display to their teacher-assessors do not match those expected and valued as signs of achievement within the dominant culture. We present examples drawn from two empirical investigations into the practices of teachers who had been trained to make judgments in a criteria-referenced, multidimensional assessment system. These investigations lead us to question whether "reliable measures of mathematical achievement," as proposed by Baxter, Shavelson, Herman, Brown, & Valadez (1993, p. 213), are ever achievable and to identify some of the mechanisms of bona fide assessment practices that may lead to inequity.

RESEARCH ON TEACHERS' INVOLVEMENT IN ASSESSMENT

By moving away from so-called objective tests as a major form of assessment and introducing complex tasks that may be undertaken in a range of contexts, the traditional notions of reliability and objectivity may not be applicable. It is thus important to consider how fairness and equity may be conceptualized and achieved in different forms of assessment. A number of potential sources of inequity arising from teacher involvement in assessment have been identified in the literature.

Inconsistent Application of Standards

For assessment to be equitable, different assessors should apply the same standards. The difficulty in achieving this is well recognized, especially where the assessment tasks are complex. In general, the response to this concern has been to emphasize the importance of training teachers to assess consistently. (Borko, Mayfield, Marion, Itexer, & Cumbo, 1997; Camp, 1993; Clarke, 1996). There is some evidence that, with training and experience, a substantial degree of agreement about standards of students' work can be achieved among teacher-assessors. In the field of English education, for example, teachers have had considerably more experience in making qualitative judgments about students' writing and acceptable levels of agreement are regularly achieved, though doubts are still expressed about the meaning of such judgments and the methods used to achieve them (see, for example, Filer, 1993; Wyatt-Smith, 1999). In mathematics, Baxter et al. (1993) report that different assessors can reach the same conclusion when assessing complex writing about special tasks done in research situations, though they point to the need for giving and marking (i.e., scoring) a wide range of tasks to allow for students' variable performance. However, assessor reliability found in research and training situations may be explained by the artificiality of the assessment activity, including the lack of accountability and the limited knowledge of the students, as well as the effects that judgments and decisions may have on them (Gitomer, 1993). Nevertheless, following five years of experience with large-scale portfolio assessment in the United States, Koretz (1998) reports that consistency among raters of eighth-grade mathematics portfolios improved "with refinements of rubrics and training" (p. 320) from a correlation of .53 to a correlation of .89.

Agreement among teacher-assessors may be explained by the development of shared constructs during the training and rating period (Roper & McNamara, 1993). Such constructs are not necessarily articulated by either the teachers or their trainers, but are manifested in "agreed marks" (i.e., high rates of agreement), a phenomenon described by Wiliam (1994) in relation to teachers assessing written work in mathematics. As Wiliam states, "To put it crudely, it is not necessary for the raters (or anybody else) to know what they are doing, only that they do it right" (p. 60). However, the development of shared expectations of students' work may result both in "pathologizing" unusual and creative approaches to assessment tasks and in stereotyping the tasks offered to students in order to maximize their chances of meeting assessment criteria (Morgan, 1998; Wiliam, 1994; Wolf, 1990).

Systematic Bias

The sort of training described above may address some of the inconsistencies between assessors' ratings but probably does not address systematic bias such as may occur in the relationships between minority students and their teachers (Baker & O'Neil, 1994). For example, Cazden's study of Black and White teachers' responses to narratives produced by Black and White children suggests that the cultural and linguistic expectations of teachers belonging to one racial group may lead them to devalue the performance of students belonging to a different racial group (Cazden, 1988). Kraiger and Ford (1985) suggest that there are consistent differences in the ways that people assess members of different racial groups—differences that may not disappear even after training has increased general consistency. In a large-scale comparison between teacher assessments and standard tests of National Curriculum levels of students in England and Wales at Key Stage 1 (age 7), Thomas, Madaus, Raczek, and Smees (1998) found that students with special educational needs, those receiving free school meals (a surrogate measure of low socioeconomic group status), and those with English as a second language demonstrated lower achievement than others when assessed using written tests, but appeared to be even more disadvantaged when assessed by their teachers in "authentic" situations. Thomas et al. conclude that "certain aspects of how teachers judge student outcomes in the National Curriculum need to be examined in more detail. Indeed … findings suggest the possibility of systematic teacher bias" (p. 231). This finding was confirmed in a more recent study by Reeves, Boyle, & Christie (2001), in which they conclude that teachers consistently underestimate the attainment of those who have already been identified as having special educational needs. Thus, based on evidence from research studies, the presence of systematic bias demonstrates that increased reliability does not ensure equity.

Poorly Designed Tasks

The quality and consistency of teachers' informal assessment of their students are also dependent on the quality of the tasks used. This issue has been identified as a concern by several authors (e.g., Clarke, 1996). Senk, Beckman, and Thompson (1997) examined the kinds of assessment task used in 19 classrooms and found that, in general, teachers selected low-level abstract tasks that did not reflect the aims of reform curricula. Such a mismatch is likely to disadvantage those students whose achievements are stronger in reform aspects of the curriculum, restricting their opportunities to show what they can do. Advocates of recent reforms in curricula and assessment methods (e.g., Romberg & Wilson, 1995) propose that tasks designed to value students' demonstrations of what they know and can do are more in keeping with the aims of reform curricula than those that penalize them for what they do *not* know or *cannot* do. Such tasks are often practical, extended, and exploratory in nature, and are usually accompanied by specific and flexible marking (or scoring) criteria (Leder, 1992; Romberg, 1995; van den Heuvel-Panhuizen, 1996). Assessment methods based on practical or exploratory tasks have the potential to provide information that cannot be obtained from written tests. They may also serve to focus teachers' attention on specific aspects of mathematical performance rather than on general impressions, thus possibly avoiding inequitable judgments arising from teachers' preconceptions about their students.

Training teachers in the design and use of assessment tasks has also been proposed as a means of improving the quality of assessments (Clarke, 1996; Senk et al., 1997). In the United Kingdom, however, even after ten years of teaching and assessment reform and its associated training, Torrance and Pryor (1998) express concern about the methods used by teachers in formative assessment, including ambiguous questioning and tasks that are poorly focused on the subject matter.

There are also practical difficulties in giving specially designed assessment tasks to a whole class and monitoring each student's performance unless only written records are assessed (Barr & Cheong, 1995, p. 180). Administration of special tasks makes comparisons between different assessment environments difficult. These tasks do not take into account all of a student's mathematical achievements, nor are they useful for comparative purposes, because their meaning is dependent on interpretation and the specific context. None of these problems makes special tasks unjust in themselves, but such difficulties limit the uses to which the assessment outcomes can be put.

Studying Teachers' Assessment Practices

Although studies of the results of assessments can reveal inequity, they cannot tell us how this arises. Moreover, as Marshall and Thompson (1994) remark, results of assessments in research situations, with trained markers (raters), selected teachers, and in some cases artificial test situations, do not necessarily tell much about assessment in classrooms. In a more natural setting, Saxe, Gearhart, Franke, Howard, and Crockett (1999) have analyzed the changes in, and relationships between, the forms and functions of assessment used by teachers as they respond to pressures for reformed curriculum and assessment in the United States and suggest that teachers may not assess complex performance systematically. Most other studies of teachers' involvement in assessment (e.g., Gipps, Brown, McCallum, & McAlister, 1995; Senk et al., 1997; Stables, Tanner, & Parkinson, 1995) tend to be concerned with the assignment of grades as a result of formal, summative teachers' assessments, the ways teachers interpret procedures for assessing, recording and reporting, and their underlying beliefs about teaching and assessment rather than with the processes of interpreting student performance.

A notable exception is the comprehensive ethnography undertaken by Filer and Pollard (2000), which describes how different teachers formed different relationships with one student during her progress through primary school and formed correspondingly different assessments of her capabilities and achievement. In the context of secondary school mathematics, Rapaille's (1986) analysis of mathematics teachers' assessment practices suggests that the same teacher could allocate different marks to similar answers from students depending on the images of the individual students that the teacher had already formed. As Mavrommatis (1997) remarks in his discussion of classroom assessment in Greek primary schools, the implicit and "covert" nature of much teacher assessment makes it particularly difficult to research (p. 383).

THE INTERPRETATIVE NATURE OF ASSESSMENT

When teachers assess the texts produced by students, they read in an interpretative and contextualized way, relying on what Barr and Cheong (1995) term "professional judgement" (p. 177) to infer meaning in terms of the students' mathematical attainments. We are using the term "text" to refer to any verbal or nonverbal behavior that is taken by a teacher to indicate some form of mathematical attainment (or a lack of attainment). Thus, we include written and oral language in response to tasks set by the teacher or produced spontaneously by the student; we also include other signs such as drawings, gestures, and facial expressions, all of which may be read and interpreted by teachers as evidence of mathematical understanding. The teachers' professional judgment is formed not only from their knowledge of the current circumstances but also from the resources they bring to bear as they "read" the students' mathematical performance from these texts. These "reader resources" (Fairclough, 1989) arise from the teachers' personal, social, and cultural history and from their current positioning within a particular discourse. The professional enculturation of teachers seems likely to ensure a certain degree of common resource. Each individual act of assessment, however, takes place within a context that calls on teacher-assessors to make use of individual, as well as collective, resources. In teacher assessment of mathematics such resources include:

1. *Teachers' personal knowledge of mathematics and the curriculum, including affective aspects of their personal mathematics history*. For example, Watson (1999) describes a teacher who, having approached mathematics when at school in a way that had often been unusual compared to those of her classmates, was more willing than her colleagues to accept unusual methods from her students.

2. *Teachers' beliefs about the nature of mathematics, and how these relate to assessment*. Even when teachers are using the same method of assessment, its characteristics can vary substantially in practice among teachers who have different mathematical and curricular priorities (see, for example, Heid, Blume, Zbiek, & Edwards, 1999).

3. *Teachers' expectations about how mathematical knowledge can be communicated*. Individual teachers may also have particular preferences for particular modes of communication as indicators of understanding. Thus, what appears salient to one teacher may not to another (Morgan, 1998; Watson, 1999).

4. *Teachers' experience and expectations of students and classrooms in general*. Teachers' expectations about students' mathematical learning may be influenced by their existing notions of how a "good mathematics student" might behave, on the basis of evidence of nonmathematical aspects of behavior, social skills, gender, and social class background (McIntyre, Morrison, & Sutherland, 1966; Walkerdine, 1988).

5. *Teachers' experience, impressions, and expectations of individual students*. Early impressions are crucially important in the teacher's accumulation of information

about each student as these form a starting point for an ensuing cycle of interaction and interpretation (Nash, 1976).

6. *Teachers' linguistic skills and cultural background*. Mismatches between teachers' and students' language and culture are associated with lower evaluations of student performance (Bourdieu, Passeron, & Martin, 1994; Cazden, 1988; Kraiger & Ford, 1985).

Teachers may adopt a range of positions in relation to their students, other teachers and external authorities (Morgan, 1998; Morgan, Tsatsaroni, & Lerman, forthcoming). Different positionings (cf. Evans, 2000; Walkerdine, 1988) are likely to give rise to the use of different sets of reader resources and hence to different actions and judgments by different teachers or by a single teacher at different times in different circumstances.

Teachers' Construction of Knowledge About Students' Mathematics

What we have described above is a set of predispositions and experiences with which teachers approach the task of constructing their knowledge of students. The constructivist paradigm of much mathematics education research and curriculum development has led to recognition of students as active constructors of mathematical knowledge. There is also some recognition that teachers construct their knowledge of children's understanding (see Confrey, 1990; Simon, 1995) and that this does not necessarily coincide with some actual understanding. As von Glasersfeld (2000) says, "What we are talking about is but our construction of the child, and that this construction is made on the basis of our own experience and colored by our goals and expectations" (p.8). Cobb, Wood, and Yackel (1990) make clear that a teacher has to *interpret* classroom events and that the teacher is a learner, in general terms, about children's understandings. They found that teachers constructed knowledge about learning in general but made only tentative comments about individuals, treating these as working hypotheses rather than as assessments that could contribute to curriculum decisions or summative statements for an external audience. However, little research on assessment in mathematics education has focused on the teacher-assessor as an active constructor of knowledge about students' mathematical knowledge or acknowledged the essentially interpretative nature of the act of assessment.

INTRODUCTION TO THE TWO STUDIES

We present here examples from two studies of teachers' assessment practices, focusing on the question of how teachers interpret students' performance both in the classroom and in written exploratory assessment tasks. The two studies we describe and illustrate do not attempt to address the issue of systematic disadvantaging of social groups but focus on describing the mechanisms by which teachers' interpretations of student behavior are made and identifying sources of difference in those interpretations that have the potential to lead to inequity.

The two independently conceived but complementary research programs from which we draw our examples emerged from concerns about the consequences of particular methods of assessment for equity and with the quality of teachers' expertise as assessors when working with students in their own classrooms. Through close investigation of teachers' assessment practices—both formal and informal—and through interrogation of theories of interaction and interpersonal judgments, we conclude that a conventional notion of reliability is inappropriate when considering teacher assessment and, instead, we raise issues to be considered by the mathematics education community. Study A looks at teachers' assessment judgments made informally in the course of everyday teaching; Study B addresses the teacher-assessed component of a high-stakes examination. Both studies are located within interpretative paradigms on two levels: first, the teacher-assessor is conceived as interpreting the actions and utterances of the students; second, the researcher interprets the actions and utterances that the teachers make about the students (Eisenhart, 1988; Mellin-Olsen, 1993). Attention is focused on the possible meanings of the externally observable phenomena of students' mathematics and the ways in which teachers interpreted these. The researcher does not have privileged access to a so-called correct knowledge of the students.

Both studies originated within the national assessment system in England. This system is statutory for schools receiving public funding. Although the particular national context necessarily affects the detail of the teacher-assessor practices that we describe, our analyses are intended to illuminate much broader issues.

The Context of Assessment Systems in England

Teachers' roles as assessors in England are multiple; they assess in order to support the learning of their students, to help their students achieve good qualifications, to optimize summative results for their school, and to provide evidence for accountability. Though routine skills are largely assessed by written tests, teachers are also centrally involved in assessing the ways in which their students approach mathematical problems investigate and communicate mathematics, plan their work, and use and apply mathematics in nonmathematical, as well as mathematical, situations. High-stakes assessments are made by externally assessed examinations combined with teacher assessments. The results of these assessments are reported to parents and to the national Qualifications and Curriculum Authority. They are also used to compile "league tables," which are published each year and rank schools according to their assessment results.

The results of Key Stage Tests (scored by examiners employed directly by the national Qualifications and Curriculum Authority) are reported for students at ages 7, 11, and 14, alongside assessments provided by teachers, on the basis of the students' everyday work in class. The grade that students achieve at age 16 in the General Certificate of Secondary Examination (GCSE) combines scores on written examinations, assessed by examiners employed by national examination boards, and scores on reports of investigative problem solving, assessed by the students' own teachers. At all stages, teacher assessment is criteria-referenced, using descriptors of levels of attainment in each of the four areas of the National Curriculum (Using and Applying Mathematics; Number and Algebra; Shape, Space, and Measures; and Handling Data). At GCSE, the parts of the descriptors of attainment in Using and Applying Mathematics relevant to investigative problem solving tasks are elaborated in greater detail (see Appendix A), and performance indicators related to specific tasks are also provided.

Open-ended explorations, with emphases on mathematical thinking, group work, discussion and extended written tasks, have been common and expected practice in mathematics teaching in England and other parts of the United Kingdom, in both primary and secondary phases, for many years. We are not claiming that all teachers teach in these ways, but we want to avoid drawing a distinction between traditional and reform methods because these coexist, even in the same classroom. For example, a teacher might use open-ended, student-centered methods yet assess learning through pencil-and-paper tests. Likewise, a teacher may teach predominately through teacher exposition and drill and practice but also include occasional investigative tasks in his or her repertoire. Such methods are now statutory as part of the National Curriculum introduced in 1989 and have been formally assessed as part of the GCSE examination since 1988, and they are also the product of gradual development over the last 30 years.

Training in assessment was given to all practicing secondary teachers when the GCSE started and to both primary and secondary teachers when the National Curriculum was introduced. In-service training, which has been in progress since the late 1980s, continues to be available. This takes the form of exercises in interpreting criteria, exercises in grading work of various kinds, agreement trials (in which teachers grade the same pieces of work independently and then meet in order to achieve agreement), and role-playing moderation procedures. New teachers learn about the statutory instruments and criteria in their preservice courses and are inducted into assessment practices in school through regular moderation discussions, in which a group of teachers examines and discusses samples of students' work in order to reach agreement about the grades that should be awarded. Teachers' assessment practices are expected to conform to national standards and are regularly inspected by government agencies. This well-established system therefore provides an important source of experience for countries whose curriculum changes are more recent than those in the United Kingdom.

The two studies reported here each took place toward the end of a period of adaptation to the new requirements, by which time all aspects of the GCSE and National Curriculum assessment systems were fully operational and teachers had all been trained in the assessment procedures and had been teaching and assessing in relevant ways for several years. All teachers in both studies had been trained in interpretation of criteria, continuous and authentic assessment, and moderation procedures. Although some had been involved in curriculum projects or other innovative work that may have led to, or presaged, national changes, others had not and may have been assessing in these ways largely because of the statutory requirements. If the teachers had been introduced to the assessment procedures more recently, we might be inclined to conceive of any problems in their practice as the result of inexperience or inadequate training. Given the extent of the teachers' experience and training, however, we see the issues that arise from the studies as having a more general significance for systems involving assessment by teachers. For each study, we describe the methods used, provide and discuss an illustrative example, and outline further issues arising from the studies.

Study A: A Teacher's Constructions of Views of Students' Mathematics

The aim of Study A was to understand the ways in which teachers might accumulate and interpret their experience with students in mathematics classrooms, not only during formal assessment situations but also during normal day-to-day activity, and to critique the robustness of the ways in which the teachers formed their assessment judgments (Watson, 1995, 1998a). This study therefore goes some way towards dealing with the lack of research noted by Marshall and Thompson (1994) and provides an English perspective on the work of Senk et al. (1997).

In a previous study involving 30 teachers, Watson (1999) found that most of the teachers used a combination of record-keeping and personal recollection as the basis for their assessments. They spoke of "getting to know" their students as an essential contributor to assessment, and of using more formal assessment as one part of the process of "getting to know" students. They made summative assessments partly by personal recollection, a process which some of them, who were aware of the flaws of paper-and-pencil testing, regarded as a much fairer method.

Study A was designed to look in more depth at assessment processes in the classroom. Two teachers were each "getting to know" new students in Year 7, the first year of secondary education. The teachers each had in excess of 10 years' experience, both were well qualified to teach mathematics and had been using activity-based classroom approaches involving group work and discussion, extended work, and so on, for several years. Both teachers kept their repertoires up-to-date by attending out-of-school meetings and reading journals. They were fully involved in the study, knowing that its purpose was to find out more about how they developed their personal views of students and whether other views might be possible. Both teachers intended to accumulate knowledge of their students' mathematics performance over time, rather than substituting tests and special tasks for their intended holistic assessment style (Firestone, Winter, & Fitz, 2000).

In each class, the teacher and researcher together selected a small number of target students, none of whom initially appeared to represent extremes of attainment within the group. The target students were not told of their special role in the study, but both classes were aware that research was in progress about how mathematics was taught in their school.[2] The researcher observed one lesson a week with each

[2] At this time in the United Kingdom, getting informed consent to study individual students was not a requirement. The school authorities, the teacher, and the researcher agreed that it would be counterproductive to do so. It was decided that greater good would be achieved by pursuing the research this way than by informing the students and hence making the situation more artificial.

teacher during the first term with the new class and took detailed field notes. All public utterances by the target students in the observed lessons and some one-to-one interactions with them were noted; all written work produced by each student during the term was photocopied and kept as data; and behavior and actions were observed and noted using systematic observation punctuated with records of complete incidents. It was not possible to tape-record the target students, because they had to be unaware of their role, but it was possible to take written notes of all their public verbal comments and some of their other interactions. The researcher recorded what each target student was doing every few minutes. However, if one student was speaking or interacting with the teacher or doing some practical task, observations were made of the entire episode until the student returned to some continuous behavior such as doing exercises or listening to the teacher, when systematic observation would resume. After each lesson, the teacher and researcher would briefly discuss their views of incidents during the lesson, and field notes were taken during this discussion. Classroom field notes were copied for the teacher every three weeks, and there were two lengthy tape-recorded formal meetings during the term when all evidence was shared and discussed. The researcher, by virtue of watching the target students primarily, generally saw more of each one's actions, and saw these in more detail, during the observed lessons than the teacher did; however, the teacher had access to observational evidence from other lessons when the researcher was not present.

In the oral domain, the teacher's and researcher's experience overlapped but did not coincide. The researcher was aware of some utterances by students that were not noticed by the teacher, but the teacher again had access to evidence from other lessons. The students' written work was included in the discussion between teacher and researcher. In the written domain, teacher and researcher had the same evidence, including that from unobserved lessons, although in the case of the observed students, the researcher sometimes knew more about what the students had achieved but not written down, or what had preceded their written work. Both teachers in Study A were reluctant to use written work as a dominant source of evidence for students' mathematical prowess, particularly as they have to report on students' mathematical thinking, which might not be expressed—or expressible—in writing. In discussion, the teachers always contextualized written work by referring to what they knew about the circumstances of its production. The researcher also adopted this approach, accepting that the part that written work played in teachers' informal assessments appeared to be minor during the stage of the year when teachers were getting to know their students.

The regular discussions and formal meetings took the form of sharing data and impressions. At the formal meetings, participants gave a summary of their views of the students' mathematical strengths and weaknesses and described the data that they believed had led to these views. The researcher reconstructed episodes by juxtaposing the written work produced during or between the same lessons alongside records of what students said, did, and wrote. Then the complete data set was scrutinized, and alternative interpretations of the significance of different features were discussed. The aim was neither to end up with the same view nor to generate a summative view but to explore differences and to attempt to explain why they arose. Both the researcher's and teachers' views would influence each other, and this was regarded as both inevitable and ethically desirable from the point of view of the students and the research participants.

Watson (1998a) gives further details of the collection and analysis of data, but for the purposes of this article, a summary of one case study highlights the kinds of problems that emerged. The following case of one 12-year-old student, Sandra,[3] illustrates ways in which the teachers' and researcher's observations and interpretations differed. Data about Sandra's behavior and utterances were extracted chronologically from the entire data set. The explicit aim of the discussions between teacher and researcher was to identify differences between what the two of them thought about Sandra and how these views had arisen. The key incidents arising from this discussion were those that supplied data suggesting different views or that stimulated the participants to talk about what may have influenced their views. The next sections of this article illustrate what emerged about Sandra during the formal meetings, focusing on two important areas of Sandra's mathematical achievement: mental arithmetic and mathematical thinking and problem solving.

Mental arithmetic. In verbal feedback sessions on arithmetic review questions done at home, Sandra frequently called out answers, waving excitedly and noisily when she wanted to contribute. Nearly all her enthusiastic contributions arose from work done at home (possibly with help from her parents; see episode below) or, very occasionally, from discussions with the teacher. The researcher noted that Sandra regularly changed her written answers in every lesson observed. The teacher was aware that she had altered some but did not know that she did so as frequently as was recorded in the field notes. As recorded in the researcher's field notes, the following episode, which took place in a whole-class session where students were giving answers to arithmetic homework, was typical of Sandra's behavior:

Teacher: What is the product of 125 and 100?

Sandra [*calls out*]*:* 225.

[*Teacher thanks her but explains she has given the sum.*]

Sandra [*audible to the researcher but not to the teacher*]:
But my mum says ... [*then inaudible*]

Teacher [*eventually, after getting answers from others*]:
12500.

Sandra [*loudly*]: Oh! that's what I had!

For later questions, Sandra calls out about half the answers correctly, but for the others, she changes the answers in her book. Maintaining a bouncy demeanor, she calls out at the end of the feedback, "I got nearly all right." After the lesson, the

[3] All students' and teachers' names in this article are pseudonyms.

teacher commented that Sandra and one other student had done better than the others on these questions, which was his view at that time.

The researcher noted that Sandra used her fingers when doing subtractions of small numbers in situations where a competent arithmetician would be likely to have known number bonds or to have developed some patterning strategies. The teacher had not seen this at all and said that his assessment of her competence in arithmetic—initially low based on results from a paper-and-pencil test—had risen to a relatively high level as a result of her oral contributions in class. The researcher saw a pattern to her oral contributions that the teacher did not, or could not, see. It was the pattern and circumstances of the contributions, rather than their quantity or nature, that indicated a contrast between Sandra's arithmetical abilities and her desire to be good with calculations. The teacher had a view that Sandra was mainly good at mental arithmetic, changing wrong answers in order to keep her confidence up, and assessed her as relatively strong in that area of mathematics. The researcher's view was that Sandra *wanted* to appear to be good but was regularly making and hiding errors that were not appearing in her final written work. If the teacher had been able to spend one-to-one time with Sandra, a situation that was impossible in a class of 30 students, these differences might have been explored further. Meanwhile, however, he had not assessed her as needing any particular help with arithmetic.

Mathematical thinking. In contrast to his view that Sandra was reasonably good at arithmetic, the teacher's perception was that Sandra was relatively weak in her ability to think mathematically while using and applying mathematics or when tackling new ideas, such as in exploratory work designed to allow the teacher to assess mathematical processes. However, the researcher observed several occasions when Sandra appeared to use mathematical thinking skills. For example, she was initially unsuccessful in making a rectangle with pentomino jigsaw pieces when using counting squares as a strategy. The teacher attempted to help her organize her counting method, but eventually she independently constructed another approach using an appropriate width and length as constraints within which to work successfully. During this time, she worked alone except when the teacher helped her to count. Nevertheless, later in the term, he said that she "doesn't see new ways of doing stuff."

In a lesson involving algebraic generalization of patterns made from strings of colored cubes, Sandra was working with a repeating sequence of three white cubes followed by one black cube. She called the teacher to her and said, "I'm trying to find out the reason why on a 2 it would be white, on 4 it would be black, on 6 it would be white and so on. So on 10 it would be white. So white occurs every other number. I don't know how to write that." Teacher said yes and then walked away. When asked about this later, he was unable to recall the incident but supposed that he had been happy with the verbal generalization and had not thought it worth pushing her further to an algebraic representation or that other students had needed him.

There were several other incidents in which Sandra appeared to do little except when the teacher was with her, and he commented to the researcher that "she always seems to have her hand up when I go past her." Nevertheless, some observations suggested that she might be able to devise strategies, reuse strategies she had found effective in the past, describe patterns, and make conjectures resulting from pattern. The mismatch between the teacher's evaluation of Sandra's mathematical thinking as low level and dependent on his help and the researcher's interpretations that she had shown in some circumstances some features of mathematical thinking could suggest that the teacher had helped her to achieve progress in these areas but had not yet noticed this. Alternatively, it may be that he was always underestimating her thinking because the only times he had noticed it were in the context of her requests for help. It is also possible that the incidents observed by the researcher were atypical because not every lesson was observed. However, in Sandra's case, the teacher believed that discussing these incidents with the researcher had been useful, giving him further evidence for his evaluations, and he later observed and reported some similar incidents in which he, too, had noticed that she could think on a more abstract level. In some written work near the end of the term, Sandra had successfully drawn some complicated 3-D shapes constructed of interlocking cubes, a task which had involved imagery and reasoning as well as observation, and which few other students had managed. In this case, the teacher was able to say that he knew this work had been produced by Sandra on her own without direct help, but with frequent affirmation from him.

Relative to the researcher, therefore, the teacher initially appeared to overestimate Sandra's skills in mental arithmetic, the area of her mathematics achievement about which Sandra most wanted to impress him, and to underestimate her skills of reasoning, perhaps because she demonstrated less confidence about them or had less opportunity to articulate them, or perhaps because she created a negative impression by asking for help frequently. The teacher, seeing her work always in the context of what the rest of the class did and what his own expectations of her were, made a judgment that was *comparative* to what she had done before and to the rest of the class. But "what she had done before" included creating an impression in his mind; therefore, his judgments were relative to the picture already formed.[4] In this case, the teacher had already changed his mind once about Sandra because of a dramatic difference between a low entry test result and her confident demeanor in the classroom. He was slow to change his opinion again, even in the light of accumulating evidence, because what he saw could be explained in the context of his current view. Only when he had the chance to reflect on evidence presented as a sequence of similar incidents could he begin to review her strengths and weaknesses in mathematics. In general, we contend that initial impressions are likely to change only when it is not possible to explain later behavior in a way that is consistent with those impressions. Because any piece of evidence can be explained in multiple ways, the explanation most likely to be chosen is that which seems most natural in the light of existing theories and resources.

[4]The teacher has seen this analysis and accepts that it is not a criticism of his practice but rather a description of aspects of Sandra's work and the problems of informal judgment of which he was not aware.

This case illustrates several important features: the strong influence of and tendency to cling to impressions and the strong influence of obviously positive or negative behavior. To summarize, it suggests that teachers might also be influenced by: (a) seeing, or failing to see, patterns in responses and behavior, as shown in the differences between the teacher's and researcher's impressions; (b) types and patterns of behavior being overrepresented or underrepresented in the teacher's mental picture; (c) students' strong or weak social skills, such as Sandra's ability to attract the teacher's attention; (d) comparisons to early impressions, such as Sandra's enthusiasm about arithmetic; (e) time constraints that prevent a full exploration of a student's mathematics, as illustrated in the teacher's comment about having other students who needed him; (f) inability to see and use all the details that occur in classrooms, such as Sandra's finger-counting.

Similar influences were found in nearly all the ten cases of students studied in two classrooms (Watson, 1998a, 1998b), and similar results have also been reported by Ball (1997). In no sense are we suggesting that there is a true view of the student to be achieved or that the researcher is correct and the teacher is wrong. The outcomes of the study do not support these assumptions. On the contrary, the study suggests that informal assessments, including those that furnish information for summative assessments of performance, are inevitably and unavoidably influenced by a variety of factors that may have little to do with mathematical achievement. Of course, all these influences are acting also on the researcher, who is busy constructing her view of someone else's practice. So if "information about some aspects of mathematical performance as it is now conceived is best gained through observation" (Morony & Olssen, 1994, p. 397), then it is important to recognize that the partial and interpretative nature of classroom observation leads people to form legitimately different views even about standard tasks.

It may, however, be tempting to try to prevent bias in informal assessment from entering the formal process by making written work the only source of evidence. One may argue that such work, when it includes written reports of open-ended and extended problem solving and mathematical investigation, allows for the assessment of aspects of mathematical performance that cannot be assessed in timed tests (Black & Wiliam, 1998). Teachers' interpretation of written mathematics was the focus of Study B.

Study B: Teacher Assessment of Written Mathematics in a High-Stakes Context

The second study was set in the context of the GCSE examination for students aged 16+ in England. The 1988 reform of the public examination system introduced a component of *coursework*, completed in class and at home and assessed by students' own teachers, along with a more traditional timed examination assessed by external examiners employed by national examination boards. The coursework component, which at the time of its introduction could count for up to 100% of the GCSE but is now restricted to only 20%, most commonly takes the form of reports of one or more extended investigative tasks. These reports are intended to include evidence of the mathematical processes that students have used (e.g., systematizing, observing, conjecturing, generalizing, and justifying) as well as the results of their investigation. These tasks thus involve students in producing sometimes lengthy written reports. It is, however, widely accepted that many students find it difficult to produce acceptable written evidence of their mathematical achievement (see, for example, Bloomfield, 1987).

Coursework tasks are assessed by students' own teachers, using a standard set of criteria of mathematical processes (see Appendix A), which are applied to all tasks. Because the set tasks allow students to choose the methods they will use and encourage some elements of original problem posing, it is difficult to provide more task-specific criteria (see Wiliam, 1994, for a discussion of assessment methods for this type of task). In practice, some examination boards publish task-specific "Performance Indicators" that describe the features of student work at various levels, but these are only meant to be illustrative of possible approaches to the task, and teachers are advised that "the Performance Indicators are only a guide and may need to be modified by the teacher in the light of specific individual cases" (Edexcel, 1999, p.8).

The original purpose of Study B was to investigate the forms of writing that students produced when presenting reports of extended investigative work for high-stakes assessments and to consider the match or mismatch between student writing and the forms of mathematical writing valued by their teacher-assessors (Morgan, 1995, 1998). A sample of three students' texts (written reports of their investigative work) was selected for each of two coursework tasks containing a range of linguistic characteristics (see Morgan, 1995, for details of the analysis and selection of these texts). Eleven experienced teachers from five secondary schools each read and evaluated the texts for one of the tasks during individual interviews. Before the interview, they were given time to work on the task themselves and to consider what they would look for when assessing it. The teachers were all experienced in using the general criteria for assessing such tasks, and these criteria were available for them to refer to if they wished. They were then prompted to think aloud as they read the students' texts and were encouraged to explain their judgments. Although we cannot assume that such an interview setting provides an authentic example of assessment practice, it does provide insight into the resources and strategies that teachers can use to make and justify evaluations of students' work and into the range of possible variation between them. Although some of the teachers expressed a lack of certainty about the "correctness" of the grades that they eventually allocated to the students' work, all tackled the task of assessment with confidence, appearing to be engaging in a familiar activity. Analysis of the interviews explored the teachers' assessment practices, identifying the features of the texts that the teachers attended to and the values that came into play as they formed judgments about the texts and about the student writers. Different teachers' readings of some sections of text were also compared in detail.

Most of the teachers had been promoted within their schools, holding posts as head or deputy head of a mathematics department or with other specified curricular, pastoral,[5] or administrative responsibilities in addition to their teaching. Their positions and duties suggested that they were acknowledged as competent within their school communities. All had been trained in the use of the common set of criteria, were experienced in applying these criteria to their own students' work, and had participated in moderation processes both within their own schools and in relation to the decisions of external moderators employed by the examination boards. Their use of the language of the criteria during the interviews provided evidence of this familiarity.

The issue that we wish to consider here is the diversity that was discovered in the meanings and evaluations that different teachers constructed from the same texts. Given the small number of teachers and student texts involved, it is not appropriate to attempt to quantify the differences in the grades assigned to individual student texts, beyond commenting that, whereas the grades were consistent for some texts, they differed substantially for others. In one case, the same text was ranked highest (of a set of three texts) by some teachers and lowest by others and was assigned grades ranging from B to E where the possible grades ranged from A (highest) to G (lowest). We are more concerned here with how such differences may arise than with the differences themselves, and we present a single case study that illustrates one source of variation: the sense made by teachers of the mathematical content of a text. This example illustrates the ways in which teachers reading with different resources (including different prior experiences, knowledge, beliefs, and priorities) can arrive at very different judgments about the same student.

The task "Topples" involved investigating piles built of rods of increasing lengths, seeking a relationship between the length of the rod at the bottom of the pile and the length of the first rod that would make the pile topple over (see Figure 1). This task was one of a set provided by the official examination board for the

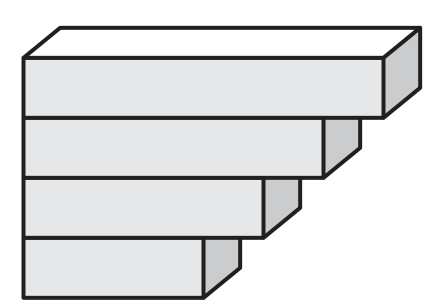


Figure 1. A pile of rods (the next rod might make the pile 'topple').

[5] *Pastoral* responsibilities include duties that American educators usually label as *guidance* or *counseling* activities. Teachers in the United Kingdom who have pastoral responsibilities may be called on to act as liaisons with parents, other caregivers, and social services; may play senior roles in disciplinary procedures within their schools; and may coordinate information from other teachers in order to monitor the overall behavior and performance of individual students.

specific purpose of allowing students to meet the general criteria for the coursework component of the GCSE (see Appendix A). The full text of the task is in Appendix B, and the performance indicators for this task are in Appendix C.

One student, Steven, completed the early parts of the task with results similar to those of many other students. He had built several piles of rods and found the length of the "topple rod" for piles of rods with bases of lengths from 1 to 10 units. He had tabulated his results and used the pattern in his empirical results to derive a formula relating the length of the "topple rod" to the length of the rod at the base of the pile. Having shown that he could use this formula to find the length of the "topple rod" for piles starting with rods longer than those he had available to build with, he then presented an alternative method for finding results for piles starting with long rods by scaling up his results for piles starting with short rods, illustrating this method by taking the result for a pile starting with a rod 10 units long and multiplying it by 10 to find the result for a pile starting with a rod 100 units long (Figure 2). The original formula and the work preceding this were similar to those produced by other students and their validity was not questioned by any of the teacher-readers. We are interested here only in the alternative method (which begins with "An alternative way to do this ..." in Figure 2), which was unique to Steven's report.

Steven had found the formula $(A + A) + \left(\frac{A}{2}\right) + b$, where A = the length of the rod at the base of the pile, and used it to calculate the length of the rod that would topple a pile that started with a rod of length 100 units:

$$(100 + 100) = 200 \qquad \left(\frac{100}{2}\right) = 50$$

$$200 + 50 = 250$$

250 would be the one at which the pile would topple.

An alternative way to do this would be to take the result of a pile starting at 10 and multiply it by 10.

$$(10 + 10) = 20 \qquad \left(\frac{10}{2}\right) = 5$$

e.g. $20 + 5 = 25$

or you could even take the basic result 1 without rounding it up, and you could multiply it by 100:

$$(1 + 1) = 2 \qquad \left(\frac{1}{2}\right) = 0.5$$

e.g. $2 + 0.5 = 2.5$

$$2.5 \times 100 = 250$$

Figure 2. Steven's alternative method.

No further justification of this method appeared in the text, which, in particular, gave no indication of how Steven had derived it. In order to evaluate Steven's work, each teacher-reader constructed his or her own understanding not only of the method itself but also of the means by which Steven might have derived it and of his level of mathematical achievement. The following extracts from interviews with three teachers (called here Charles, Grant, and Harry) who read this section of Steven's work illustrate how different these understandings can be.

Charles: Um OK, so I mean he's found the rule and he's quite successfully used it from what I can see to make predictions about what's going to happen for things that he obviously can't set up. So that shows that he understands the formula which he's come up with quite well, I think. There's also found some sort of linearity in the results whereby he can just multiply up numbers. Which again shows quite a good understanding of the problem I think.

Charles recognizes the mathematical validity of the alternative method, relating it to the linearity of the relationship between the variables. He takes this as a sign that the student has "come up with" the formula as a result of understanding the linearity of the situation. This results in a positive evaluation of Steven's mathematical understanding.

Grant: It's interesting that the next part works. I don't know if it works for everything or it just works for this, but he's spotted it and again he hasn't really looked into it any further. He's done it for one case but whether it would work for any other case is, er, I don't know. He hasn't looked into it … And he's used it in the next part, er, used th- this multiplying section in the next part, and it's just a knowledge of number that's got him there, I think, intuition whatever. He may have guessed at a few and found one that works for it.

Grant appears less confident about the mathematical validity of the alternative formula, expressing uncertainty about whether the method would work in general. Perhaps because of this uncertainty, his narrative explaining how Steven might have arrived at the method devalues this student's achievement, suggesting that the processes involved were not really mathematical: "spotting" the method, not looking into it properly, guessing, using "just a knowledge of number" or intuition. Steven is clearly not being given credit either for the result itself or for the processes that he may have gone through in order to arrive at it.

Harry: And he's got another formula here…. I don't really understand what he's done here…. So he's produced another formula where…. He's taken the result of a pile starting at ten and multiplying by ten and I don't understand what he's done there…. I would have asked him to explain a bit further. He's – the initial formula with two hundred and fifty is proved to be correct and he's trying to extend it. He's trying to look for other ways. Maybe he has realized that two hundred and fifty could be the exact answer or maybe not. So he's trying other ways to explain some of the inconsistencies that he's seen, but I think greater explanation [is] needed here.

Like Grant, Harry seems to have some difficulty making sense of the mathematics and does not appear to recognize the equivalence between the original formula and the alternative method. In spite of this, he is able to compose yet another narrative to explain the student's intentions, stressing by repetition the suggestion that Steven has been "trying" (possibly with the implication that he has not succeeded). Harry, although appearing willing to applaud Steven's perseverance, also has a low evaluation of his mathematical understanding, even questioning whether he knows that his first method had already yielded a correct answer. Moreover, Harry locates the responsibility for his own failure to understand in the inadequacies of the student's text.

In this case, the three teachers' different interpretations of Steven's level of understanding and their different hypotheses about the methods he might have used seem to be connected to their personal mathematical resources. It is Charles, expressing the clearest understanding of the mathematics of the situation, who makes the most positive evaluation of Steven's understanding, whereas Grant and Harry, apparently uncertain of the general validity of the method, construct pictures of the student working in relatively unstructured or experimental ways.

Such major differences between teachers in their interpretations and evaluations of students' texts occurred primarily where the student's text diverged from the norm in some way—either in its mathematical content or in the form in which it was expressed. In this context, the norm appeared to be a text that followed a linear route from gathering data, through pattern spotting, to forming, verifying, and applying a generalization in a standard form. Our identification of this as a norm arises both from analysis of the teachers' interview data as a whole and from the literature on investigational work in the United Kingdom—for example, Wells (1993) and Hewitt (1992). Steven's presentation of an alternative method, though likely to be valid in other contexts, appeared as a deviation from the norm, not only because its mathematical form was nonstandard but also because it was unusual to see more than one method presented.

In order to make sense of a text (in this case a report of investigative work) and to use it to evaluate the student writer's achievement, each teacher must compose an explanatory narrative, drawing on the resources available to him or her. These resources include common expectations of the general nature of mathematical problem solving and reports of such work as well as more personal mathematical understanding and experiences. When a student text diverges from the usual to the extent that it is not covered by the established common expectations, each teacher must resort to his or her more personal resources, thus creating the possibility of divergence in the narratives they compose.

There is not space in this article to provide additional detailed examples, but we will briefly indicate some of the other ways in which teachers' interpretations and approaches to evaluation were found to differ in the broader study from which this example has been taken:

1. Teachers formed different hypotheses about work the student might have done in the classroom that was not indicated in the written report, and they expressed different attitudes towards valuing such unrecorded achievement. For example, an algebraic generalization unsupported by an earlier verbal generalization might lead

one teacher to suspect that the student had cheated whereas another would be "confident that … he's done it" (Morgan, 1998, p. 161).

2. Teachers made different judgments about some factual aspects of the text—for example, whether the wording of the problem given by the student had been copied from the original source or had been paraphrased in the student's own words (Morgan, 1998, p. 164).

3. Some teachers appeared to focus on building up an overall picture of the characteristics of the student, some were interested in making mathematical sense of what the student had done, and others focused solely on finding evidence in the text to meet specific criteria. A case study exemplifying this point can be found in Morgan (1996).

4. When there were tensions between the teachers' own value systems and their perceptions of the demands of externally imposed sets of assessment criteria, some teachers resolved this by submitting to the external authority whereas others were prepared to apply their own, unofficial criteria instead (Morgan, 1998, pp. 158–159).

Even when reading the same student's text, teachers may understand it in different ways. Even when they have formed similar understandings of what the student has done, they may assign it different values. The greatest discrepancies in final evaluations occurred in those cases where the student's work, like Steven's alternative formula, was unusual in some way. In cases where the student had produced a more conventional solution in a conventional form, teachers using different approaches to the assessment process and drawing on different resources seemed more likely to coincide in their overall evaluations. Students who are creative or produce unusual work—qualities that are officially endorsed by the aims and values of the curriculum reforms—are thus at risk because the value to be placed on their work depends crucially on the idiosyncratic resources of the teacher assessing it.

APPROACHES TO IMPROVED AND MORE EQUITABLE ASSESSMENT

Both the studies described above detail aspects of assessment practices that have the potential to lead to inequitable decisions about students' futures either through assignment of particular summative grades, thus affecting employment or life choices or through missing out on particular kinds of recognition, curricular support, or other opportunities, thus affecting future mathematical learning. In this section, we consider some approaches suggested within the mathematics education community that attempt to improve the quality of assessment, and their viability in the light of issues raised by the two studies. As we have argued, assessment is essentially interpretative. This means that, when we talk of the *quality* of assessment, we do not wish to suggest that *accuracy* is an appropriate criterion to apply. We are concerned rather, on the one hand, with the match between the stated aims of the curriculum and assessment systems and the outcomes of the assessment and, on the other hand, with reducing the potential sources of inequitable judgments identified in the two case studies.

Could Specially Designed Tasks Ensure Equity?

Well-designed assessment tasks may go some way toward focusing teachers' attention on specific aspects of mathematical performance rather than the sort of general impressions found in Study A. However, they do not solve the problem of inequity in teacher assessment for the following reasons.

First, assessment still depends on interpreting particular actions of students. Some actions will be seen as significant and others not. Interpretations may not match a student's intentions but will depend, as we have seen in Study B, on the personal resources each teacher-assessor brought to the assessment. It is not merely a problem of task design, since many of the tasks used in Studies A and B were designed specifically for assessment of students' mathematical thinking and knowledge. Van den Heuvel-Panhuizen (1996) shows how tasks may be redesigned to leave less scope for alternative inferences, but this inevitably limits the opportunities for students to use higher-level problem solving processes, including specifying problems and choosing methods, that show their thinking skills.

Second, choice of method is a feature of mathematical problem solving that can be influenced by context, size of given numbers, form of presentation, available equipment, and so on. Hence, what is done is situationally specific and not necessarily generalizable for a student. Equally, absence of expected evidence may not indicate inability to generate such evidence in another task. Sandra, in Study A above, independently shifted from a low-level counting method of problem solving to one that was structural. At some other times, she remained stuck in a lower-level mode of working, even when helped. In attempting to make use of the results of such assessment tasks, teachers will interpret the meaning of the evidence (or the lack thereof) according to their preexisting views of the student.

Third, avoiding the effects of teachers' casual judgments is not possible, because the day-to-day interactions of the classroom are influenced by teachers' constructions of the students' mathematical attainment in the context of tasks designed primarily for teaching—for developing cognitive and metacognitive aspects of mathematics understanding—rather than for assessment. Reactions to these tasks influence the teacher's and students' views and expectations of what might happen in special assessment tasks. Teachers in Study B drew on their knowledge of the sorts of things that students generally did in classrooms in order to make sense of the texts they were assessing.

How Can Better Specification and Use of Assessment Criteria Avoid Inequity?

It is clear from the literature reviewed earlier that, in some circumstances, the clear specification of assessment criteria and training in the use of such criteria can improve the reliability of formal assessments made by groups of teachers. This by itself, however, is not enough to eliminate the possibility of differences among teachers' interpretations. In the Study B, for example, the three teachers constructed very different understandings of the mathematical content of Steven's writing and of the mathematical processes that he might have gone through to achieve his results. The particular form of the mathematics produced by this student (and all the possible alternative forms) could not have been predicted by those designing the task in order to provide these teachers with guidance for all possible cases. If we are to make use of open-ended problems and investigations, encouraging students to be original and creative in mathematics (as advocated by curriculum reform movements in many countries), then criteria cannot be made specific enough to ensure that all teachers understand in identical ways the potential diversity of the mathematics produced by students. The experience of assessment of investigative work in England and Wales suggests that the existence of detailed criteria drives teachers towards setting less open, more stereotyped tasks (Wolf, 1990) and that, even where teachers themselves are committed to the ideals of investigative work and have been involved in the development of criteria, they become less appreciative of students' work that strays from predictable paths (Morgan, 1998; Wiliam, 1994).

There is some evidence to suggest that when students are aware of the criteria by which their work is to be assessed and are involved in the assessment process through peer- and self-assessment, they become "acculturated" (Tanner & Jones, 1994) and produce work that is closer to their teachers' expectations. The assessment of work produced by such students may thus be more likely to receive consistent evaluations. On the other hand, Love and Shiu's investigation of students' perceptions of assessment criteria indicated that some students also had a "sceptical awareness of the routinisation of producing work for assessment" (1991, p. 356). Again, there is the risk that students and teachers will direct their efforts towards producing work that is "safe," in that it matches routine norms, rather than taking possibly creative risks (cf. Gilbert, 1989, in the context of creative writing in English).

What Is the Role of Professional Dialogue in Improving Assessment Practice?

Studies of teachers working together over time suggest that, through the development of shared constructs, teachers can achieve greater reliability in the standards that they apply to examples of students' work. It has been suggested that such professional dialogue is the key to ensuring quality (and, by implication, equity) in teachers' assessments (Clarke, 1996). During the introduction of the National Curriculum in England and Wales, group moderation by teachers both within and between schools was recommended as a means of achieving reliability and as professional development for the teachers involved (Task Group on Assessment and Testing, 1987).[6] Moderation both during training sessions and as part of each

[6] In practice, the complex model of group moderation recommended by the National Curriculum Task Group on Assessment and Testing has never been implemented. It is necessarily time consuming and expensive—and hence unattractive to those who are eager for simple solutions to educational problems.

school's regular assessment practice is common for formal assessments such as GCSE coursework. The achievement of shared constructs and reliability in application of criteria and standards, however, does not address all the issues raised in the two studies we have presented here.

The three teachers in Study B had been trained and had participated in moderation of similar work in the past. That experience can be said to have been successful to the extent that they did not generally differ substantially in the final grades they allocated to most of the student texts. Their differences of opinion over Steven's work arose not from differences in the standards they were applying but from differences in their reading and interpretation of the idiosyncratic mathematical content of his text. Such differences might have been resolved if the teachers had had the opportunity to share and negotiate their interpretations, but it is surely not feasible to make every student text subject to such negotiated assessment.

The teachers in Study A were carrying out their assessment in the relative privacy of their own classrooms. The issue here is not so much the standards that the teachers were applying but the ways in which they became aware of and made use of the evidence available to them. The opportunity to share evidence and discuss impressions with the researcher may be seen as a form of professional dialogue. In order to examine one's own perceptions critically, it is necessary to become aware of possible alternative perceptions and interpretations and to engage with these.

One purpose of professional dialogue is to establish shared language with which to think about students' achievement as well as to communicate about it to others. Such a language should provide specialized resources particularly suited to the task of interpreting students' mathematical behavior, while helping to avoid the reliance on general resources such as the notion of ability that characterizes much classroom assessment practice (Dunne, 1999; Ruthven, 1987). However, professional dialogue and teacher education should not be seen simply as vehicles for teachers to acquire and apply similar understandings of criteria and standards. Indeed, the notion that there is one correct way of assessing a student or a piece of work underlies many of the potential problems we have identified. In contrast, we would suggest that one of the main benefits of professional dialogue among teachers is that it can raise doubts in teachers' minds about the certainty of their own interpretations of their students' behaviors. Resolving differences of opinion in assessment contexts by reaching a consensus is not necessarily the best course. Fully recognizing such differences may be a more useful approach, leading to an awareness of the possibility of alternative interpretations, of the inadequacy of the evidence available, and of the need to seek other perspectives. Dialogue with colleagues can thus provide teachers with questions with which to interrogate their own judgments.

CONCLUSION

The two case studies that we have presented not only illustrate the interpretative nature of assessment in both informal and formal contexts but also provide some insight into the details of how different interpretations of students' achievements can occur. There are potential sources of inequity deeply embedded in traditional and reform processes for formative and summative assessment. Though closed-answer assessment tasks are known to pose problems in relation to equal access for students of different gender, ethnic group, and class, whenever evaluation of students relies on observation and interpretation of their behavior and their oral or written production, it will be influenced by the resources individual teachers bring to the assessment task. Such evaluation brings with it the possible sources of difference between teachers and differences in the ways individual students may be evaluated that we have illustrated in the two studies presented. The scope of the two studies has not allowed us to address directly the issue of systematic bias in relation to the assessment of students belonging to different social groups, but interpretation and evaluation of behavior are likely to be influenced by cultural expectations (Dunne, 1999; Walkerdine, 1988), and mismatches between the cultural and linguistic resources of teachers and students are likely to lead to evaluations that disadvantage students from nondominant social groups (Bourdieu, Passeron, & Martin, 1994; Cazden, 1988). We have argued that tighter specification of tasks or of criteria cannot remove such sources of inequity entirely and may indeed introduce other undesirable consequences for the quality of assessment. Interpretation is an essential characteristic of assessment activity and cannot be eliminated. However, insight into the details of how differences in interpretation occur may enable teachers to engage in critical reflection on their practice and, through such critical awareness, to lessen the likelihood of resultant inequity.

In day-to-day classroom interactions, we have identified the necessary incompleteness of teachers' awareness of their students' behavior, the potential for alternative interpretations of what is observed, the ways in which early judgments about individual students may influence subsequent interpretations of their behavior, and the consequent potential for inequitable evaluations. We certainly do not wish to question the integrity or competence of teachers as they make these judgments about their students, though raised awareness of processes of formative assessment (as described, for example, by Clarke & Clarke, 1998, and Torrance & Pryor, 1998) may at least increase the sources of evidence on which teachers base their judgments. Rather, the potential for inequity is a necessary consequence of the interpretative nature of assessment taking place in human interaction.

The concern for reliability in summative assessments has led to calls for tighter specification of criteria and training of teachers to develop shared ways of applying the criteria. We have argued that reliability is not a concept that can be applied simplistically to assessments of students' mathematics. Indeed, where students are given opportunities to respond to assessment tasks in open-ended ways, reliability may be an impossible goal. Moreover, it seems likely that tighter specification of criteria will lead to stereotyped responses from both teachers and students – in opposition to the value ascribed to creativity, openness and authenticity within 'reform' discourse.

Nevertheless, tentative judgments have to be made so that teachers can make pedagogical decisions. Teachers' own recognition that these determinations are situated and temporary indicates that such judgments should not be used as a basis for discriminatory action or high-stakes decisions. The issue of which teachers are less aware—the effects of their own beliefs and practices on equity—is harder to tackle, given that it affects students on many levels. Leaving the discussion of possible bias until summative judgments are made is too late; the model offered above of self-doubt and regular critical collegial discussion is rare, but it could be an important step towards equity.

REFERENCES

Baker, E. L., & O'Neil, H. F. (1994). Performance assessment and equity: A view from the USA. *Assessment in Education*: *Principles, Policy, and Practice, 1*(1), 11–26.

Ball, D. L. (1997). From the general to the particular: Knowing our own students as learners of mathematics. *Mathematics Teacher*, *90*, 732–737

Barr, M. A., & Cheong, J. (1995). Achieving equity: Counting on the classroom. In M. T. Nettles and A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 161–184). Dordrecht, Netherlands: Kluwer Academic.

Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. *Journal for Research in Mathematics Education*, *24*, 190–216

Bell, A., Burkhardt, H., & Swan, M. (1992). Balanced assessment of mathematical performance. In R. Lesh & S. J. Lamon (Eds.), *Assessment of authentic performance in school mathematics* (pp. 119–144). Washington DC: American Association for the Advancement of Science.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–74.

Bloomfield, A. (1987). Assessing investigations. *Mathematics Teaching*, *118*, 48–49.

Borko, H., Mayfield, V., Marion, S., Itexer, R., & Cumbo, K. (1997) Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, *13*, 259–78.

Bourdieu, P., Passeron, J. C., & de Saint Martin, M. (1994). *Academic discourse: Linguistic misunderstanding and professorial power* (R. Teese, Trans.). Cambridge, UK: Polity Press.

Burton, L. (Ed.). (1994). *Who counts? Assessing mathematics in Europe*. Stoke-on-Trent, UK: Trentham Books.

Camp, R. (1993). The place of portfolios in our changing views of writing assessment. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 183–212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Clarke, D. (1996). Assessment. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education* (Vol. 1, pp. 327–370). Dordrecht, Netherlands: Kluwer Academic.

Clarke, D., & Clarke, B. (1998). Rich assessment tasks for Years 9 and 10. In J. Gough & J. Mousley (Eds.), *Mathematics: Exploring all angles* (pp. 93–96). Brunswick, Victoria, Australia: Mathematics Association of Victoria.

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 125–146). Reston, VA: National Council of Teachers of Mathematics.

Confrey, J. (1990). What constructivism implies for teaching. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 107–122). Reston, VA: National Council of Teachers of Mathematics.

Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex, and problem-solving.* Buckingham, UK: Open University Press.

de Lange, J. (1995). Assessment: No change without problems. In T. A. Romberg (Ed.), *Reform in school mathematics and authentic assessment* (pp. 87–172). New York: SUNY Press.

Dunne, M. (1999). Positioned neutrality: Mathematics teachers and the cultural politics of their classrooms. *Educational Review, 51*, 117–128.

Edexcel (1999). *Mathematics 1385 and 1386: The assessment of MA1 1999*. London: Author.

Eisenhart, M. A. (1988). The ethnographic tradition and mathematics education research. *Journal for Research in Mathematics Education, 19,* 99–114.

Evans, J. (2000). *Adults' mathematical thinking and emotions: A study of numerate practices*. London: Routledge.

Fairclough, N. (1989). *Language and power*. Harlow, UK: Longman.

Filer, A. (1993). The assessment of classroom language: Challenging the rhetoric of 'objectivity.' *International Studies in Sociology of Education, 3,* 193–212.

Filer, A., & Pollard, A. (2000) *The social world of pupil assessment*. London: Continuum.

Firestone, W. A., Winter, J., & Fitz, J. (2000). Different assessments, common practice? *Assessment in Education: Principles, Policy, and Practice, 7*(1), 13–37.

Gilbert, P. (1989). *Writing, schooling, and deconstruction: From voice to text in the classroom*. London: Routledge.

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence? Teachers and National Assessment of seven-year-olds*. Buckingham, UK: Open University Press.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement, and equity*. Buckingham, UK: Open University Press.

Gitomer, D. H. (1993). Performance assessment and educational measurement. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 241–264). Hillsdale, NJ: Lawrence Erlbaum Associates.

Heid, M. K., Blume, G. W., Zbiek, R. M., & Edwards, B. S. (1999). Factors that influence teachers learning to do interviews to understand students' mathematical understanding. *Educational Studies in Mathematics, 37,* 223–49.

Hewitt, D. (1992). Train spotters' paradise. *Mathematics Teaching, 140*, 6–8.

Koretz, D. (1998). Large-scale portfolio assessments in the U.S.: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy, and Practice, 5,* 309–334.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70,* 56–65.

Leder, G. (Ed.) (1992). *Assessment and learning mathematics*. Victoria, Australia: Australian Council for Educational Research.

Lesh, R., & Lamon, S. J. (Eds.). (1992). *Assessment of authentic performance in school mathematics*. Washington, DC: American Association for the Advancement of Science.

London East Anglian Group. (1991). *Mathematics coursework tasks and performance indicators (1988–1991)*. London: Author.

Love, E., & Shiu, C. (1991). Students' perceptions of assessment criteria in an innovative mathematics project. In F. Furinghetti (Ed.), *Proceedings of the 15th Conference of the International Group for the Psychology of Mathematics Education* (Vol. II, pp. 350–357). Assisi, Italy: Program Committee of the 15th PME Conference.

Marshall, S. P., & Thompson, A. G. (1994). Assessment: What's new—and not so new—A review of six recent books. *Journal for Research in Mathematics Education, 25,* 209–218.

Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process—The assessment episode. *Assessment in Education: Principles, Policy, and Practice, 4,* 381–399.

McIntyre, D., Morrison, A., & Sutherland, J. (1966). Social and educational variables relating to teachers' assessments of primary students. *British Journal of Educational Psychology, 36,* 272–279.

Mellin-Olsen, S. (1993). A critical view of assessment in mathematics education: Where is the student as a subject? In M. Niss (Ed.), *Investigations into assessment in mathematics education: An ICMI study* (pp. 143–156). Dordrecht, Netherlands: Kluwer Academic.

Morgan, C. (1995). *An analysis of the discourse of written reports of investigative work in GCSE mathematics*. Unpublished doctoral dissertation, Institute of Education, University of London, London.

Morgan, C. (1996). Teacher as examiner: The case of mathematics coursework. *Assessment in Education: Principles, Policy, and Practice, 3*, 353–375.

Morgan, C. (1998). *Writing mathematically: The discourse of investigation*. London: Falmer Press.

Morgan, C., Tsatsaroni, A., & Lerman, S. (forthcoming). Mathematics teachers' positions and practices in discourses of assessment. *British Journal of Sociology of Education*.

Morony, W., & Olssen. K. (1994). Support for informal assessment in mathematics in the context of standards referenced reporting. *Educational Studies in Mathematics*, *27*, 387–399.

Nash, R. (1976). *Teacher expectations and student learning*. London: Routledge and Kegan Paul.

National Council of Teachers of Mathematics (NCTM). (1995). *Assessment standards for school mathematics*. Reston VA: NCTM.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Niss, M. (Ed.) (1993). *Investigations into assessment in mathematics education*: *An ICME Study*. Dordecht, Netherlands: Kluwer Academic.

Rapaille, J. P. (1986). Research on assessment process in 'natural' conditions. In M. Ben-Peretz, R. Bromme, & R. Halkes (Eds.), *Advances of research on teacher thinking* (pp. 122–132). Lisse, Netherlands: Swets and Zeitlinger.

Reeves, D. J., Boyle, W. K., & Christie, T. (2001). The relationship between teacher assessment and pupil attainments in standard test tasks at Key Stage 2, 1996–8. *British Educational Research Journal*, *27*, 141–160.

Romberg, T. A. (Ed.). (1995). *Reform in school mathematics and authentic assessment*. New York: SUNY Press.

Romberg, T. A. & Wilson, L. D. (1995). Issues related to the development of an authentic assessment system for school mathematics. In T. A. Romberg (Ed.). *Reform in school mathematics and authentic assessment* (pp. 1–18). New York: SUNY Press.

Roper, T., & MacNamara, A. (1993). Teacher assessment of mathematics Attainment Target 1 (MA1). *British Journal of Curriculum and Assessment*, *4*(1), 16–19.

Ruthven, K. (1987). Ability stereotyping in mathematics. *Educational Studies in Mathematics*, *18*, 243–253.

Saxe, G. B., Gearhart, M., Franke, M. L., Howard, S., & Crockett, M. (1999). Teachers' shifting assessment practices in the context of educational reform in mathematics. *Teaching and Teacher Education*, *15*, 85–105.

Senk, S. L., Beckmann, C. B., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education*, *28*, 187–215.

Simon, M. (1995) Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*, 114–145.

Stables, A., Tanner, H., & Parkinson, J. (1995). Teacher assessment at Key Stage 3: A case study of teachers' responses to imposed curricular change. *Welsh Journal of Education*, *4*(2), 69–80.

Stephens, M., & Money, R. (1993). New developments in senior secondary assessment in Australia. In M. Niss (Ed.), *Cases of assessment in mathematics education: An ICMI Study* (pp. 155-171). Dordrecht, Netherlands: Kluwer Academic.

Tanner, H., & Jones, S. (1994). Using peer and self-assessment to develop modeling skills with students aged 11 to 16: A socio-constructivist view. *Educational Studies in Mathematics*, *27*(4), 413–431.

Task Group on Assessment and Testing. (1987). Report of the task group on assessment and testing. London: Department of Education and Science and the Welsh Office.

Thomas, S., Madaus, G. F., Raczek, A. E., & Smees, R. (1998). Comparing teacher assessment and the standard task results in England: The relationship between students' characteristics and attainment. *Assessment in Education: Principles, Policy, and Practice*, *5*, 213–254.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment*. Buckingham, UK: Open University Press.

van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht, Netherlands: CD-ß Press.

von Glasersfeld, E. (2000). Problems of constructivism. In L. P. Steffe & P. W. Thompson (Eds.) *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld* (pp. 3–9). London: Routledge.

Walkerdine, V. (1988). *The mastery of reason*. London: Routledge.

Watson, A. (1995). Evidence for students' mathematical achievements. *For the Learning of Mathematics*, *15*(1), 16–20.

Watson, A. (1998a). *An investigation into how teachers make judgements about what students know and can do in mathematics*. Unpublished doctoral dissertation, University of Oxford, Oxford.

Watson, A. (1998b). What makes a mathematical performance noteworthy in informal teacher assessment? In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 169–176). Stellenbosch, South Africa: University of Stellenbosch.

Watson, A. (1999). Paradigmatic conflicts in informal mathematics assessment as sources of social inequity. *Educational Review*, *51*, 105–115.

Wells, D. (1993). *Problem solving and investigations* (3rd ed.). Bristol, UK: Rain Press.

Wiliam, D. (1994). Assessing authentic tasks: Alternatives to mark-schemes. *Nordic Studies in Mathematics Education*, *2*(1), 48–68.

Winfield, L. F. (1995). Performance-based assessments: Contributor or detractor to equity? In M. T. Nettles & A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 221–241). Dordrecht, Netherlands: Kluwer Academic.

Wolf, A. (1990). Testing investigations. In P. Dowling & R. Noss (Eds.), *Mathematics versus the National Curriculum* (pp. 137–153). London: Falmer Press.

Wyatt-Smith, C. (1999). Reading for assessment: How teachers ascribe meaning and value to student writing. *Assessment in Education: Principles, Policy, and Practice*, *6*, 195–223.

Authors

Candia Morgan, Mathematical Sciences, Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, United Kingdom; c_morgan@ioe.ac.uk

Anne Watson, University of Oxford Department of Educational Studies, 15 Norham Gardens, Oxford OX2 6PY, United Kingdom; anne.watson@educational-studies.oxford.ac.uk

APPENDIX A:
General criteria for GCSE coursework

Assessment Criteria for *Using and Applying Mathematics*

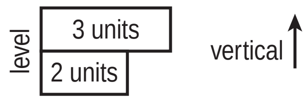
	Strand (i): Making and monitoring decisions to solve problems	**Strand (ii): Communicating mathematically**	**Strand (iii): Developing skills of mathematical reasoning**
1	Candidates try different approaches and find ways of overcoming difficulties that arise when they are solving problems. They are beginning to organise their work and check results.	Candidates discuss their mathematical work and are beginning to explain their thinking. They use and interpret mathematical symbols and diagrams.	Candidates show that they understand a general statement by finding particular examples that match it.
2	Candidates are developing their own strategies for solving problems and are using these strategies both in working within mathematics and in applying mathematics to practical contexts.	Candidates present information and results in a clear and organised way, explaining the reasons for their presentation.	Candidates search for a pattern by trying out ideas of their own.
3	In order to carry through tasks and solve mathematical problems, candidates identify and obtain necessary information; they check their results, considering whether these are sensible.	Candidates show understanding of situations by describing them mathematically using symbols, words and diagrams.	Candidates make general statements of their own, based on evidence they have produced, and give an explanation of their reasoning.
4	Candidates carry through substantial tasks and solve quite complex problems by breaking them down into smaller, more manageable tasks.	Candidates interpret, discuss and synthesise information presented in a variety of mathematical forms. Their writing explains and informs their use of diagrams.	Candidates are beginning to give a mathematical justification for their generalisations; they test them by checking particular cases.
5	Starting from problems or contexts that have been presented to them, candidates introduce questions of their own, which generate fuller solutions.	Candidates examine critically and justify their choice of mathematical presentation, considering alternative approaches and explaining improvements they have made.	Candidates justify their generalisations or solutions, showing some insight into the mathematical structure of the situation being investigated. They appreciate the difference between mathematical explanation and experimental evidence.
6	Candidates develop and follow alternative approaches. They reflect on their own lines of enquiry when exploring mathematical tasks; in doing so they introduce and use a range of mathematical techniques.	Candidates convey mathematical meaning through consistent use of symbols.	Candidates examine generalisations or solutions reached in an activity, commenting constructively on the reasoning and logic employed, and make further progress in the activity as a result.
7	Candidates analyse alternative approaches to problems involving a number of features or variables. They give detailed reasons for following or rejecting particular lines of enquiry.	Candidates use mathematical language and symbols accurately in presenting a convincing reasoned argument.	Candidates' reports include mathematical justifications, explaining their solutions to problems involving a number of features or variables.
8	Candidates consider and evaluate a number of approaches to a substantial task. They explore extensively a context or area of mathematics with which they are unfamiliar. They apply independently a range of appropriate mathematical techniques.	Candidates use mathematical language and symbols efficiently in presenting a concise reasoned argument.	Candidates provide a mathematically rigorous justification or proof of their solution to a complex problem, considering the conditions under which it remains valid.

APPENDIX B:
The "Topples" Task

TOPPLES

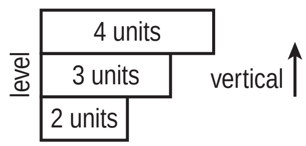
In this task you will be asked to balance some rods of different lengths on top of each other, until the pile topples.

The diagrams below are given as illustrations.



We start the pile with the 2 unit rod on the bottom and balance the three unit one on top of it, being careful that the left hand edges are level.

Then we balance the 4 unit rod on top of the three unit rod.



We continue building the pile, progressing through the sequence of rods, until the pile topples.

You should find that this pile of rods topples when we get to the 5 unit rod.

So the pile that starts with the 2 unit rod at the base eventually topples when we get to the 5 unit rod.

Your task is to investigate the relationship between the length of the rod at the bottom of the pile and the rod which first makes the pile topple.

1. Starting with rods of different lengths at the base, build up your piles until each one topples. Make sure that the rods increase by one unit of length at a time.
   (a) Record the length of the rod at the base and the length of the rod that makes the pile topple.
   (b) Tabulate your results.
   (c) Make any observations that you can.
   (d) GENERALISE.
   (e) Explain your result. (Well argued explanations based on intuition and insight will gain at least as much credit as those based on the principles of Physics.)
2. Imagine that you start with a rod of length 100 units and build up the pile using rods of lengths 101, 102, 103, … units.
   What will be the length of the rod that first makes the pile topple?
3. A pile topples when we place a rod of length 50 units on the top.
   (a) What will be the length of the rod on the bottom of the pile?
   (b) Explain your working.

OPTIONAL EXTENSION

Extend this investigation in any way of your own choosing.

APPENDIX C:

Performance Indicators for "Topples"

[From LEAG (1991). Mathematics coursework tasks and performance indicators (1988-1991). London: London East Anglian Group.]

The generalisation for this task is well within the syllabus at intermediate and higher level; it is a simple linear function. We would therefore expect to see an algebraic (symbolic) representation for this generalisation from candidates at grade C and above. From the candidates at grades B and A we would expect to see use of this algebraic form in the two specific cases given and to offer, certainly at grade A, some explanation of why the pile topples.

For the award of a grade D, candidates should be expected to do much of that for a grade C but to lack an element of sophistication. They might, for instance, fail to generalise in an algebraic form but be able to state generalisation in words or through a specific examples. They might even, at the lower levels of grade D or top grade E, answer the specific examples by extending their table of results.

The candidates who score a grade B will certainly have tried to undertake the task in an ordered, strategic manner, looking at well selected specific cases. They will also obtain a good table of results.

At the two lower levels, we would expect to see some reasonable attempts at the investigation but not handled in a any strategic fashion. The results at grade G are likely to be flimsy, few and not particularly accurate. For a grade F we would expect to see at least a couple, and preferably a trio, of correct results. The grade F and G candidates will not have been able to handle either of the specific examples.

As before, the optional extension should be used to enhance grades at all levels.

Journal for Research in Mathematics Education
2002, Vol. 33, No. 2, 111–144

Initial Fraction Learning by Fourth- and Fifth-Grade Students: A Comparison of the Effects of Using Commercial Curricula With the Effects of Using the Rational Number Project Curriculum

Kathleen A. Cramer, Thomas R. Post, and Robert C. delMas
University of Minnesota

This study contrasted the achievement of students using either commercial curricula (CC) for initial fraction learning with the achievement of students using the Rational Number Project (RNP) fraction curriculum. The RNP curriculum placed particular emphasis on the use of multiple physical models and translations within and between modes of representation—pictorial, manipulative, verbal, real-world, and symbolic. The instructional program lasted 28–30 days and involved over 1600 fourth and fifth graders in 66 classrooms that were randomly assigned to treatment groups. Students using RNP project materials had statistically higher mean scores on the posttest and retention test and on four (of six) subscales: concepts, order, transfer, and estimation. Interview data showed differences in the quality of students' thinking as they solved order and estimation tasks involving fractions. RNP students approached such tasks conceptually by building on their constructed mental images of fractions, whereas CC students relied more often on standard, often rote, procedures when solving identical fraction tasks. These results are consistent with earlier RNP work with smaller numbers of students in several teaching experiment settings.

Key Words: Conceptual knowledge; Elementary, K–8; Fractions; Instructional intervention; Large scale studies; Manipulatives; Rational number/representations

Teaching and learning about fractions have traditionally been problematic. Results from large-scale assessments such as the National Assessment of Educational Progress (NAEP) reveal that fourth-grade students have limited understandings of fractions (Kouba, Zawojewski, & Strutchens, 1997). For example, students' understanding of the fundamental concept of equivalent fractions should

This research was supported in part by the National Science Foundation under grant No. NSF-ESI-9454371 (Rational Number Project). Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors are indebted to Lisa Jeffers Ruff and Apryl Henry for their assistance in this study while they were in graduate school at the University of Minnesota.

Copyright © 2002 The National Council of Teachers of Mathematics, Inc. www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in any other format without written permission from NCTM.

reflect more than just a knowledge of a procedure for generating equal fractions but should be rich in connections among symbols, models, pictures, and context. Yet, only 42% of fourth graders in the NAEP sample could choose a picture that represented a fraction equivalent to a given fraction, and only 18% could shade a rectangular region to produce a representation of a given fraction.

Another NAEP item showed that fourth graders also had difficulty with a problem that assessed their understanding of the importance of the unit in determining the quantity associated with a fraction. The item first told students that each of two children had eaten $^1/_2$ of a pizza, and then it asked the students to decide if it was possible that one child had eaten more than the other. Students had to show their work and provide an explanation. Fourth-grade students had difficulty with this task: Only 24% gave a satisfactory or extended response; 20% gave a partial or minimal response that showed some fraction ideas; 49% gave an incorrect response that did not use ideas about fractions at all; and 7% did not try to answer the question.

Mathematics educators have examined instructional issues surrounding why students have difficulty learning about fractions. Summarizing issues raised in several studies, Moss and Case (1999) suggested that these difficulties are related in part to teaching practices that emphasize syntactic knowledge (rules) over semantic knowledge (meaning) and discourage children from spontaneous attempts to make sense of rational numbers. Traditionally, mathematics instruction in Grade 4 devotes little time to developing an understanding of the meaning of fractions beyond simple part-whole shading tasks. Too often, students have few experiences in comparing different concrete models for fractions and have little opportunity to communicate their solutions to problems. Equivalence is taught as a procedure disconnected from representation, and the role of the unit is addressed inadequately, if at all.

The *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) and the *Principles and Standards for School Mathematics* (NCTM, 2000) provide teachers with suggestions for improving initial fraction instruction in Grades 3–5. Developing an understanding of the meaning of the symbols, examining relationships, and building initial concepts of order and equivalence should be the focus of instruction. Conceptual understanding should be developed before computational fluency, since fluency in rational-number computation will be a major focus in grades 6–8. Learning activities for students should involve fractions that are easily modeled, and the importance of the unit and its subdivision into equal parts should be emphasized. Because many children experience difficulty constructing these ideas, the *Standards* documents also suggest that instruction should use physical objects, diagrams, and real-world situations and that instruction should help students make connections from these representations to oral language and symbols. With these basic understandings, students should be able to develop their own strategies for computing with simple fractions. The *Standards* documents advise that emphasizing these basic ideas in the elementary grades would reduce the amount of time that teachers in the middle grades need to spend addressing student misunderstandings and procedural difficulties.

The National Science Foundation–sponsored Rational Number Project (RNP) has investigated students' learning of fraction ideas for several years (Behr, Lesh, Post, & Silver, 1983; Post, Behr, & Lesh, 1982). One product of the RNP research is a curriculum that helps students develop initial ideas about fractions by working with multiple physical models and other representations and by translating between and within these various representations (Cramer, Behr, Lesh, & Post, 1997a; Cramer, Behr, Lesh, & Post, 1997b). Students using this curriculum develop an understanding of the meaning of fractions, and order and equivalence before developing symbolic procedures for fraction operations.

The study reported here examined achievement in initial fraction learning of fourth and fifth graders and contrasted the results of using commercial curricula (CC) with the results of using the RNP curriculum. Fourth and fifth graders in 66 classrooms participated in this 6-week study, which examined student achievement patterns as well as differences in students' thinking. This study is important and timely because of the need for effective ways to help large numbers of classroom teachers implement curricular materials that establish a solid conceptual base on which to build students' conceptual fluency. Previous research by the RNP supports the use of curricula that involve children with multiple representations, and in particular, multiple manipulative models, as a way of developing their conceptual understanding of mathematical ideas (Bezuk & Cramer, 1989; Post et al., 1982; Post, Cramer, Behr, Lesh, & Harel, 1992). This study further documents the influence of multiple representations on children's initial learning of fraction ideas.

BACKGROUND OF THE RATIONAL NUMBER PROJECT

The Rational Numbers Project (RNP) has been funded by the National Science Foundation since 1979 and has over 80 publications on its Web site (RNP, 2001), many of which report on several investigations concerning the teaching and learning of fractions among fourth- and fifth-graders (Bezuk & Cramer, 1989; Behr, Wachsmuth, Post, & Lesh, 1984; Post, Wachsmuth, Lesh, & Behr, 1985). The curriculum used in the study reported in this article emanated from our previous research involving 12-, 18-, and 30-week teaching experiments, with the longest experiment involving students in fourth (and then fifth) grade. The purpose of these teaching experiments was to document students' thinking as they interacted with fraction ideas over an extended period of time. The curriculum created for the teaching experiments reflected the following beliefs: (a) Children learn best through active involvement with multiple concrete models (Dienes, 1969); (b) physical aids are just one component in the acquisition of concepts—verbal, pictorial, symbolic and realistic representations also are important (Lesh, 1979); (c) children should have opportunities to talk together and with their teacher about mathematical ideas (Johnson & Johnson, 1989); and (d) curriculum must focus on the development of conceptual knowledge prior to formal work with symbols and algorithms (Hiebert, 1994).

The RNP curriculum lessons covered the following topics: (a) a part-whole model for fractions, (b) a flexible concept of unit, (c) concepts of order and equivalence,

(d) estimation of addition and subtraction with fractions, and (e) finding exact answers to fraction addition and subtraction problems at the concrete level. The curriculum did not emphasize symbolic procedures for ordering fractions, fraction equivalence, and operating on fractions but instead stressed the development of a quantitative sense of fraction. For example, in an interview, a fourth grader using the RNP curriculum was asked to estimate the sum of 2/3 and 1/6 by placing an *x* on a number line partitioned to show these intervals: 0, 1/2, 1, 1 1/2, and 2. She placed the *x* between 1/2 and 1 and reasoned as follows: "2/3 is more than 1/2; then you add 1/6; it is not bigger than one; 1/6 is less than 1/3." This student was able to judge the relative sizes of 2/3 (2/3 > 1/2) and 1/6 (1/6 < 1/3) mentally without relying on any formal procedures and then to reason that if 2/3 plus 1/3 equals one, then because1/6 is less than 1/3, the answer must be less than one. Thus, she demonstrated an ability to coordinate all the information needed to provide a reasonable estimate of the answer to the problem. This type of reasoning was a goal of the RNP curriculum.

An important finding of the RNP fraction-teaching experiments was the identification of four student-constructed ordering strategies (Behr et al., 1984): *same numerator*, *same denominator*, *transitive*, and *residual*. These strategies do not depend on the paper-and-pencil methods of least common denominator or the cross-product algorithm commonly suggested in commercial curricula. Instead, they rely on students' mental imagery of fractions and reflect a more conceptual, as opposed to procedural, understanding of fractions. These four strategies closely parallel students' actions with manipulative models and contrast with the paper-and-pencil procedures that require rewriting both fractions with common denominators or calculating cross products. RNP lessons supported the development of these student-constructed strategies.

The "same numerator" and "same denominator" strategies can be demonstrated in the following examples. When students compare 2/3 and 2/6 (fractions with the same numerator), they can conclude that 2/3 is the larger fraction because thirds are larger than sixths and two larger pieces must be bigger than two smaller pieces. This strategy involves understanding that an inverse relationship exists between the number of parts into which a unit is partitioned and the size of each part. The same-denominator strategy applies to fraction pairs like 3/8 and 2/8. In this case, the same denominator implies that one is comparing parts of the unit that are the same size; three parts must be greater than two parts when all the parts are the same size. This strategy is based on the number of pieces, with the size of each piece remaining constant. In the same-numerator situation, the deciding factor is the size of each piece, whereas in the same denominator example, the deciding factor is the number of pieces. Deciding *which* is important and *when* are important parts of the rational number understanding, and this interplay between size of piece and number of pieces underlies an understanding of the part-whole subconstruct.

The transitive strategy can be modeled by comparing 3/7 and 5/9. When making this comparison, a student can conclude that 3/7 is less than 5/9 because 3/7 is less than 1/2, whereas 5/9 is greater than 1/2. This strategy is labeled as transitive because both fractions are compared to an external value. One can use the residual strategy when comparing 3/4 and 5/6, for example. A student using this strategy can reflect that both fractions are one "piece" away from the whole unit, and because the missing amount 1/6 is less than the missing amount 1/4, 5/6 must be closer to the whole and therefore the bigger fraction. Thus, this strategy involves thinking about a residual—the part of the fraction that is "left over."

The fraction curriculum used in earlier investigations was revised and extended (Cramer et al., 1997a; Cramer et al., 1997b). The goal for this revision was to reorganize lessons from the 30-week teaching experiment into two levels of materials that could be used by fourth- and fifth- grade teachers. The revisions reflected our understandings about the complexity inherent in teaching beginning concepts about fractions to 10- to 12-year olds. The revised curriculum incorporated examples of earlier student thinking captured during the teaching experiments and was formatted so that classroom teachers who had limited in-service opportunities would be able to implement the curriculum successfully with children.

THE STUDY

Origins

Prior to the study we report here, a mathematics curriculum coordinator from a large suburban school district in a northern Midwestern state participated in a Rational Number Project (RNP) workshop where research on fractions was presented and sample lessons from the RNP curriculum were shared. She contacted the RNP staff to ask if her district could use the RNP fraction curriculum because she believed that the commercial curricula (CC) currently in use might not develop the type of fraction understanding in the elementary grades envisioned by the NCTM *Standards*. The RNP curriculum appeared to her to be better equipped to do that. Conversations with her led to this study. Her district would pilot the RNP curriculum within a structured study to evaluate whether this curriculum was better suited to the district's goals for fractions than either of the two commercial series used within the district.

An examination of the RNP curriculum and commercial curricula by project staff suggested that the former had a better alignment with the *Standards*' fraction goals than the latter. This study would determine whether that was in fact the case as well as identify differences between learners using these two different approaches. Although both the commercial and RNP curricula covered the same topics, the emphasis on the topics varied. Our examination of the curricula showed that students using the CC would devote less time than RNP students to developing an understanding of the meaning of fractions at the concrete level and more time to developing operational skill and procedures for ordering fractions and finding fraction equivalence. Students using the RNP curriculum would devote less time than CC students to developing symbolic computational skills and more time to building an understanding of the meaning of these new numbers.

Questions

To what extent can large numbers of classroom teachers accustomed to textbook-based instruction effectively implement a research-based curriculum that involves students in working with multiple concrete models, emphasizes translations among and within multiple representations (manipulative, pictorial, verbal, symbolic, and real-world), and has students regularly interacting with one another in group situations? We were interested to see if this could be done with limited staff development time for teachers. More specifically, the RNP was interested in investigating the following two research questions: (a) Do fraction-related differences in student achievement exist when fourth- and fifth-grade students using the conceptually oriented RNP curriculum are compared with students using district-adopted commercial mathematics curricula? and (b) if differences in student achievement do exist, what is the nature of the differences in students' thinking and understanding?

METHOD

Research Design

This study used a posttest-only control group design referred to by Campbell and Stanley (1963) as a Type 6 design. Campbell and Stanley categorize this type of design as a true experimental design because it incorporates random assignment of experimental units (in this case classrooms) to treatments and thereby eliminates some forms of initial bias contained in studies without randomization. In this setting, a pretest was deemed to be inappropriate and likely to interact with the treatments. Type 6 designs inherently control for a number of sources of internal validity: history, maturation, testing, instrumentation, regression, selection, and mortality (Campbell & Stanley, 1963, p. 8). In addition, the design controls for one of the sources of external invalidity—the interaction of the pretesting process with the treatments.

This study integrated two types of research methodologies. Quantitative analytical methodologies involving statistical methods were used to examine information gathered from large numbers of students. The aim to generalize findings to a larger population was accomplished by using normally suggested methods of randomization and multivariate techniques. We also incorporated what has been referred to as the humanistic perspective (Brown, Cooney, & Jones, 1990). Project staff collected interview data from 20 students on three or four different occasions over the course of this investigation. Classroom teachers also interviewed two or three students at the end of the instruction. These interviews provided information on differences between CC students and RNP students in how they thought about and operated on fractions.

Lastly, we have attempted to use triangulation in reviewing and making sense of these data (Schoenfeld, 1992). That is, the study employed RNP curriculum materials that were revised editions of earlier versions that we used in our own previous studies relating to children's learning of rational number. As part of RNP, we have rather extensive interview information on the thinking of student while they interact with rational number concepts, and our current research built on this knowledge base. We examined whether the results obtained in our earlier teaching experiments with small numbers of students and relatively small pupil-teacher ratios could be replicated in classrooms taught by teachers with little exposure to recent research on the teaching and learning of rational number concepts. We were able therefore to examine data from three related perspectives: (a) earlier RNP teaching experiment research, (b) large-scale data in the current study, and (c) student interviews modeled on those in our earlier work. The validity of our findings was improved as a result of this integrated perspective.

Data Sources

To investigate the two questions mentioned above, RNP personnel relied on several different data sources. Written posttest and retention test results from 66 classrooms provided data to address the first research question. Student interviews conducted by project staff during instruction and by classroom teachers at the end of instruction provided the foundation for describing differences in student thinking between the two groups, and these descriptions provided the data to address the second research question. These interview results supplemented the written test results in addressing the question of differences in the nature of student thinking**.**

Sample

This study was conducted in 66 fourth- and fifth-grade classrooms in a suburban school district south of Minneapolis. The school district had approximately 27,000 students and 18 elementary schools. One percent of the student population received free or reduced lunch. All 200 fourth- and fifth-grade teachers were contacted in the fall of 1994 to assess their interest in participating in this study. Sixty-six teachers from 17 schools chose to participate, agreeing to delay all fraction instruction until the time of the study, in mid-January. These teachers received one graduate quarter credit from the University of Minnesota and a $75 stipend, or a $100 stipend without graduate credit for their participation. In addition, for participating in the study, the teachers using the RNP curriculum received a prototype version of that curriculum (Cramer et al., 1997a; Cramer et al., 1997b) and classroom sets of manipulative materials; the teachers using the commercial curricula received the RNP curriculum and manipulatives at the end of the study.

The teachers were randomly assigned to experimental (Rational Number Project, or RNP group) or control (commercial curriculum, or CC group) conditions. Of the 33 teachers in the CC group, 27 used the 1989 Addison-Wesley series; 6 teachers used the 1992 Harcourt-Brace series. Teachers were not assigned to treatment groups by schools; rather they were assigned on a district-wide basis to "randomize out" any school effects. This is important because we were not interested in school-by-school differences but intended to generalize findings to the district level. Table 1 shows the number of classrooms and students by treatment and grade level.

Table 1
Number of Classrooms and Students by Grade Level and Treatment Group

	RNP Group	CC Group	Totals
Grade 4	19 classrooms (470 students)	19 classrooms (483 students)	38 classrooms (953 students)
Grade 5	14 classrooms (369 students)	14 classrooms (344 students)	28 classroom (713 students)
Totals	33 classrooms (839 students)	33 classrooms (827 students)	66 classrooms (1666 students)

Treatments

RNP curriculum. The RNP curriculum reflects cognitive psychological principles as suggested by Piaget (1960), Bruner (1966), and Dienes, (1969). Lesh (1979), elaborating on Bruner's ideas, developed a translation model for instruction that suggests that learning is enhanced when children have opportunities to explore mathematical ideas in multiple ways— using manipulatives, pictures, written symbols, verbal symbols, and real-life contexts—and by making translations between and within these modes of representation. Contending that these translations make ideas meaningful to children, the RNP project staff used the Lesh model to guide the development of the RNP fraction curriculum. Students using this curriculum had opportunities to explore fraction concepts, order and equivalence ideas, and addition and subtraction of fractions using fraction circles, chips, pictures, story problems, and written symbols. They were asked to make connections systematically between different representations, and were given opportunities to discuss mathematical ideas before attaching symbols to them.

The manipulative model used first and more often than any other was the fraction circle. Although the circle model was the first model introduced, students also used paper folding and sets of chips. Fraction circles and paper folding use an area interpretation of fractions; the chip model is based on a discrete counting approach to the part-whole construct for rational number (Kieren, 1976; Post et al., 1982). Each new manipulative model was introduced as a translation from one model to another. For example, the teacher would show 2/3 or some other fraction with fraction circles and ask students to show that same fraction with chips. The students then described how both models represented the fraction, noting similarities and differences between the models. Strengths and limitations of each model were discussed with students, and in later lessons, students had opportunities to choose specific models for a variety of problem situations and to justify their choices.

The RNP curriculum asked students to vary the unit when modeling fractions. With fraction circles, the whole circle was not always the unit; at times 1/2 a circle or 1/4 of a circle would be named as the unit. Students reinterpreted the value of fraction circle pieces each time a new unit was identified. Such flexibility with the unit was also emphasized with the chip model.

Moss and Case (1999) listed limitations of the circle model as one of the instructional issues behind difficulties that students have with fractions. Summarizing research by others, they suggested that using a circle model may inadvertently reinforce whole number thinking, since children need only to count, for example, three parts out of four, to name the fraction 3/4 without consciously focusing on the relationships between three and four. This caution is appropriate for most manipulative models for fractions. Though the circle model was the primary manipulative in this study, concerns that this model reinforced whole number thinking did not materialize. Such problems may occur when students are limited to using a single model and do not examine the model while the unit is continually and systematically varied as in this study.

Although two levels of RNP materials have been developed, all students in the RNP treatments used Level 1 materials regardless of grade level. Level 2 materials were designed to build on the conceptual understandings developed through the Level 1 materials. Fifth graders assigned to RNP group would not have had the conceptual understandings needed to benefit fully from Level 2 materials and therefore used the same level of the RNP curriculum as did fourth graders. An outline of the RNP instructional scope and sequence for Level 1 is given in Appendix A.

The format of the lessons was similar to that of the lessons developed in the "Math and the Mind's Eye" project available from the Math Learning Center (1988). Each lesson included an overview of the mathematical idea to be developed. Materials needed by teachers and students for each lesson were noted, and examples of student thinking derived from earlier work also were included. The lessons reflected a classroom organization that valued the role that teachers play in student learning, as well as the need for students to work cooperatively, to talk about ideas, and to use manipulative models to represent various rational number concepts. Each lesson began with the development of a mathematical idea or an extension of an idea previously introduced. Teachers were provided with problems and questions to help generate the initial discussion and target the subsequent exploration. Nearly all the lessons required interactions with manipulative materials in large and small groups. In the small groups, students worked with manipulatives and activity pages that helped them pursue ideas introduced in the large group setting.

An important part of each lesson in the RNP curriculum is the Comments section. Here insights into student thinking captured from the initial RNP teaching experiments were communicated to teachers. These comments also clarified a wide variety of other issues, such as why mastery at the symbolic level is not the primary objective for many of the earlier lessons, why developing visual imagery is important, and how to use the concrete materials to model fractions. The comments also presented numerous examples from our earlier work of students' misunderstandings and anecdotes of students' thinking for teachers' reflection.

The translation model used to guide the development of these lessons reflects our view of how problem solving can be integrated into structured learning materials. Building a rich and flexible representational system is a problem-solving experience. As children experience mathematical ideas in multiple ways and make connections between the different modes of representation, they come to understand a need for multiple representations, among which symbols are just one way to represent mathematical ideas. In the RNP lessons, the manipulative materials were the central focus, with other representations playing more of a supporting role. Symbols were used primarily to record what students observed, discussed, and acted out with various manipulatives.

Commercial curricula. The majority of CC classrooms used *Addison-Wesley Mathematics* (Addison-Wesley, 1989a; Addison-Wesley, 1989b); six of the 33 CC classrooms piloted *Mathematics Plus* (Harcourt, Brace, & Jovanavich, 1992a; Harcourt, Brace, & Jovanich, 1992b). The scope and sequence of each textbook series is given in Appendix B. Although all RNP groups used the same Level I curriculum, CC classrooms used the Grade 4 textbooks with fourth graders and the Grade 5 textbooks with fifth graders. The content in the Addison-Wesley (AW) series was covered in two chapters in Grade 4 and three chapters in Grade 5; the content was covered in two chapters in both the Grade 4 and Grade 5 Harcourt Brace Javonovich (HBJ) textbooks.

The CC teachers were encouraged to use the resources suggested in the teacher's guides. Prior to the first teacher workshop, RNP staff examined both series in detail to identify resources presented in each commercial series. We looked through all pages of the teacher's guide related to fraction instruction as well as the student textbooks, and counted the number of lessons and pages allocated to topics covered. We found that the AW textbook series used fraction bars and pictures of fraction bars as the suggested models, whereas the HBJ series used a wider variety of manipulative materials, including counters, paper folding, fraction circles, and fraction bars made from paper strips. The examination of the textbooks showed that although the AW textbook series did not integrate manipulatives into student versions, the HBJ series did, to some extent. In each case, however, manipulative materials played only a cursory role in the development of fraction ideas, with the primary goal in both the AW and HBJ series being to develop student competence at the symbolic level.

Instruments

The assessment instruments developed for this study reflected the goals articulated by the *Standards* and matched those identified by the district. We describe the various instruments—both written and oral—in this section.

Written tests. We developed two tests to be used in this study: a posttest and a retention test. The tests, which were developed as parallel forms, each contained 34 items, most of which were taken from previous RNP teaching experiments. The items assessed student learning in six strands, which were those used in our earlier RNP teaching experiments: (a) fraction concepts, (b) fraction equivalence, (c) fraction order (d) concept of unit ideas, (e) operations (+, –) and estimation, and (f) transfer. We believe that basic concepts and ideas about order, equivalence, and flexibility of unit are fundamental to any significant initial understanding of fractions (Behr et al., 1984; Mack, 1993). These items were designed to measure students' understanding of the part-whole model for fraction and the relative size of fractions. The operation and estimation items measure students' ability to apply this knowledge to estimation and fraction arithmetic tasks. The transfer items included division and multiplication tasks that are embedded in story problems as well as number line questions that ask students to transfer conceptual understanding to a new model. These strands, which represent the major areas in the RNP curriculum, are also topics covered in the commercial curricula. The approach and degree of emphasis, however, were different in the two treatment groups. Nevertheless, the six strands were discussed with *all* classroom teachers, who were encouraged to submit items for the final version of the tests. The single item submitted was included on the posttest and a parallel item was created and included on the retention test.

To insure comparability of the parallel forms, pairs of items were identified that were similar in structure and content. Figure 1 depicts the breakdown of the 34 items by strand, along with sample test items. The purpose of the tests was to assess both conceptual and procedural knowledge. All students had the option to use manipulative materials of their choice on the addition and subtraction problems, and this information was gathered in questions 26 and 30 on the test. We allowed students to use manipulatives because mastery at the symbolic level was not the primary goal of the RNP lessons.

Student interviews. The purpose of the interviews with RNP and CC students was to identify differences in students' thinking about fractions. The interview questions used in this study were based on the RNP interviews used in our earlier fraction-teaching experiments and covered the same topics as the written tests. The interview questions reflected our vision of the knowledge and understandings that students beginning to study fraction should have and that were also compatible with the vision presented in both NCTM *Standards* documents. Specifically, we believed that students should be able to model fraction concepts concretely and in multiple ways and that their knowledge should include a wealth of mental images connected to their experiences with manipulative aids. Students should be able to use these mental images to determine the relative sizes of fractions and to estimate reasonable answers for fraction addition and subtraction. Students' rich conceptual understandings should facilitate the learning of fraction operations and enable them to apply ideas in new situations.

RNP staff interviewed a randomly selected group of 10 RNP and 10 CC students from seven different schools three to four times over the course of the study. All classroom teachers interviewed two or three students whom they randomly selected

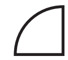
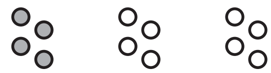
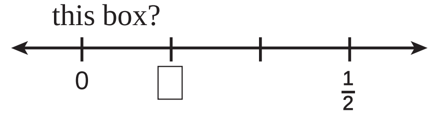
Subscales	Number of Items	Test Item Number	Sample Items
A. Fraction concepts story problems/ pictures	 7	 #4a & b #5–#9	 This is the unit. What name can I give two of these pieces? 
B. Fraction Equivalence pictorial symbolic	 3 1	 #13–#15 #16	Give two names for the fraction amount shaded: 
C. Fraction Order	7	#17–#23	Circle the larger fraction: 6/14 5/9
D. Concept of Unit	1	#10	 is 3-fourths of some length. Draw the whole length.
E. Operations (+, –) & Estimation estimation story problems symbolic	 3 2 3	 #1–#3 #24, #25 #27–#29	• Annie and Josie receive the same allowance. Josie spent 4/9 of hers on tapes. Ann spent 1/3 repairing her bicycle. Josie spent how much mpre of her allowance than Annie? • Estimate by writing in the box the whole number you think is close to: 3/8 + 5/12 • You may use your fraction circles or any other manipulative to do this problem: 1/3 + 2/6
F. Transfer items story problems • division • multiplication number line • concepts • concept of unit	 2 1 2 2	 #11, #12 #31 #32, #33 #34, #35	• Mary had 1-half yard of rope. She cuts it into 3 equal-sized parts. What fraction of a yard is each piece? • What number should go in this box? 

Figure 1. Breakdown of posttest and retention test by category.

from their classes at the end of the study. They used a subset of questions from the interviews that RNP staff conducted during the study. Each interview question asked the children to give verbal explanations or demonstrations of their responses. To obtain accurate information on students' responses to the interview questions, teachers and RNP staff audiotaped the interviews and took detailed notes at the same time. The teachers transcribed their own interviews and returned these transcriptions to the RNP staff for analysis at the end of the study.

Using the notes and the audiotapes, we obtained information on children's responses to interview questions. Individual RNP staff members coded responses for students in schools to which the staff members were assigned. First, the students' responses were coded as correct or incorrect, and then they were further coded by type of response—either procedural or conceptual. These analyses were subsequently shared and discussed in project meetings to ensure consistency across RNP staff members' codings and analyzes of the data.

Timeline and Teacher Activities

The study, designed to last for 30 days, began in mid-January, 1994, with the first of two 2-hour teacher workshops during which teachers were informed of their assignments (RNP or CC treatments). For the first hour, all teachers heard a presentation that dealt with these topics: a history of the RNP, the structure of the study, and the research on student learning of fractions. During the second hour, teachers separated into two groups by treatment. The RNP group received the RNP curriculum and manipulatives, reviewed the RNP goals and scope and sequence, and worked through sample activities from the lessons. The CC group reviewed the goals and the scope and sequence of the textbook fraction-related chapters, discussed the suggestions that the teacher's edition presented to enhance instruction, and worked with paper folding and several board-like games as described in the teacher's editions.

Between mid-January and mid-March, the teachers taught fractions in their classrooms, with instruction lasting between 28 and 30 days, within a 50-minute class period. Teachers in both groups kept a daily log. In the RNP group, teachers completed a form for each lesson, on which they documented the amount of time allocated to the lesson, changes made, and recommendations for improving that lesson. The CC teachers, who provided more information in their logs because they had more flexibility in how they offered instruction, also completed a form asking for the following information: objectives, page numbers covered in the text, and time spent on whole-class instruction, seatwork, and group work. These teachers were asked to list the materials that they and the students used, to describe how the lesson was developed, and to give examples of reteaching and enrichment activities.

A second 2-hour workshop session was held in late February. In the first hour, all teachers participated in a discussion on assessing fraction learning to prepare them for interviewing three of their own students at a later date. They viewed and discussed the Marilyn Burns assessment video (1993) depicting a fraction-related interview with a fifth grader. During the second hour, the RNP teachers worked through activities with manipulatives to model fraction addition and subtraction. The CC group considered several fraction enrichment activities, including one activity that used ancient Egyptian methods for calculating with unit fractions.

The teachers in both treatment groups kept a daily log documenting classroom activities. Teacher logs showed that teachers followed through on their responsibilities. Sixty-two of the 66 teachers submitted their logs, 31 from each group. Examination of the RNP logs showed that, with three exceptions, the RNP teachers completed all 23 lessons. The RNP teachers made few adaptations to the lessons, with the most common being the inclusion of similar problems in the whole group part of the lesson or the addition of more story problems to the last three lessons dealing with the operations. All fourth-grade CC teachers completed both fraction chapters, and all fifth-grade classrooms covered sections dealing with concepts, order, equivalence, addition and subtraction. Some fifth-grader teachers did not complete sections of the chapters dealing with multiplication and division of fractions, but these topics were not addressed in RNP curriculum, so their exclusion was not important.

Although we asked teachers to allocate 50 minutes to each class, some teachers in both groups averaged less time. Average class times for RNP teachers ranged from 33 minutes to 54 minutes, with a median of 48 minutes; 25 of 31 RNP teachers averaged 45 minutes or more. Average times for CC teachers ranged from 26 minutes to 60 minutes, with a median of 49 minutes; 23 of 31 CC teachers averaged 45 minutes or more.

For this study, teachers completed a series of tasks. They administered posttests in the first week of March and retention tests 4 weeks later, scored both tests, and recorded data on a sheet provided by project staff. Complete sets of posttests from two teachers from each school were randomly selected for rescoring to estimate the reliability of both teacher scoring and their recording of posttest data. A similar reliability check was performed for the retention tests, but only one teacher from each school was selected. The percentage of agreement was 99% for post and retention tests.

Overview of Analytic Procedures

Both quantitative and qualitative analyses were conducted. The quantitative analyses involved conducting a factor analysis to assess the viability of the original test subscales, calculating reliabilities, and using MANOVAs to determine whether treatment differences on total test and revised subscales exist. The qualitative data involved interviews given by project staff and classroom teachers.

RESULTS

Quantitative Analyses

Factor analysis. Student achievement based on the written tests was analyzed by total test score and by scores on clusters of items based on the original strands. We adjusted the original six strands shown in Figure 1 to five strands (subscales) at the time of the data analysis because there was only one concept-of-unit question; this item was included with the other concept questions. The test item numbers for each of the five subscales—concepts, equivalence, order, operations (+, –) and estimation, and transfer—are presented in Table 2. Based on RNP research studies, the five subscales used in this study identify important aspects of students' initial learning of fractions. In particular, earlier RNP teaching experiments identified order and equivalence (Behr et al., 1984; Post et al., 1985), estimation (Behr et al., 1986) and concept of unit (Behr et al., 1992) as important ideas for instruction and assessment. We built on this knowledge base and experience to develop the test strands for this study.

Table 2
Test Items for Original Subscales Prior to Factor Analysis

Original Subscale	Test Item Numbers
Concepts	4a, 4b, 5–10
Equivalence	13–16
Order	17–23
Operations (+, –) and Estimation	1–3; 24, 25; 27–29
Transfer	11,12, 31–35
Total Test	1–25; 27–29; 31–35

Note. Items 26 and 30 were not included in analyses, because they were yes/no inquiries about the use of manipulative materials.

Because the test items and original subscales have never been used as a separate instrument, it was important to assess the degree to which the items within each subscale were statistically verifiable. For this purpose, we conducted a principal components factor analysis with a varimax rotation of the factor matrix using SPSSX FACTOR program, a part of SPSS 4.0. All subsequent analyses were conducted with this program. The principal components analysis extracted eight factors for the posttest and seven factors for the retention test using a criterion of an eigenvalue of at least 1.0. In both the posttest and retention test, the identified factors accounted for 50% of the total variance.

Table 3 illustrates the congruence between our original subscales and the extracted factors. On the basis of a table provided by Stevens (1996, p. 371), a conservative critical value of .16 was used to determine the significance of factor loadings at $p < .01$. Table 3 also identifies variables that share16% or more variance in common with a factor (Stevens, 1996, p. 372). Five of the factors extracted on the posttest and retention test aligned well with our original subscales. This is noted in the table by the test item numbers marked with an asterisk (*) and a dagger (†). The majority of items in each subscale had statistically significant loadings on the same factor, and most share a significant amount of variance in common with the same factor. The one exception is the set of items 1, 2, and 3 from the Operations and Estimation subscale. The factor analyses for both the posttest and retention test showed that these three items uniquely identified a separate factor. As a result, we decided to separate the estimation items from the other operation items and use them to define a sixth subscale, called Estimation. The remaining factors extracted on the posttest and retention test accounted for small proportions of total variance (3% or less) and were defined by only one or two test items that were already associated with one of the original subscales.

Table 3
Congruence Between Original Subscales and Extracted Factors

Original Subscale	Test Item Numbers							
Concepts	4a	4b	5	6	7	8	9	10
Posttest	*†	*†		*†	*†	*	*†	*
Retention Test	*		*	*†		*†	*†	*
Equivalence	13	14	15	16				
Posttest	*†	*†	*†	*				
Retention Test	*†	*†	*†	*				
Ordering	17	18	19	20	21	22	23	
Posttest	*†	*†	*†			*†	*†	
Retention Test		*†	*	*†	*†	*	*	
Operations (+, –) and Estimation	1	2	3	24	25	27	28	29
Posttest				*†	*†	*†	*†	*†
Retention Test				*†	*†	*†	*†	*†
Transfer	11	12	31	32	33	34	35	
Posttest	*	*	*	*†	*†	*†	*†	
Retention Test	*†	*†	*	*†	*†	*†	*†	
New Subscale								
Estimation	1	2	3					
Posttest	*†	*†	*†					
Retention Test	*†	*†	*†					

Note. † designates a factor loading ≥ .40 (i.e., item shares at least 16% of the variance with corresponding factor).
* $p < .01$.

Reliability analysis of subscales. Reliability analyses of internal consistency were conducted on the total set of items and on each of the six revised subscales on both the posttest and retention test. The initial analyses indicated that in general the reliability coefficients (Cronbach's alpha) could not be improved significantly for the total test or subscales with one exception. When the two ordering items designed to encourage students to use a residual strategy were dropped from the ordering subscale, the reliability coefficient for that subscale increased by .05 and .09 on the posttest and retention test, respectively. The removal of any other items on this scale would have resulted in a decrease in reliability. Reliability coefficients for the posttest and retention test were .88 and .90, respectively. Subscale reliabilities for the post and retention tests ranged from .60 to .76, with the majority of them exceeding .72.

Description of MANOVA analyses. SPSS 4.0 multivariate analysis of variance (MANOVA) with classroom as the experimental unit was used to conduct subsequent analyses. Using classrooms was appropriate because classrooms (not individuals) were assigned randomly to treatments. Significance was set at $p \leq .0083$ ($.05 \div 6$) to reflect a Bonferroni adjustment to control for test-wise error. The MANOVA design consisted of one within-subject factor (Posttest vs. Retention Test) and two between factors, (RNP vs. CC and Grade 4 vs. Grade 5). The dependent variables for the first MANOVA were total scores on the posttest and retention test. A separate MANOVA also was run for each of the six revised subscales. The design was identical to that of the first MANOVA except that scores on the subscales, rather than the total test scores, were used as dependent variables.

Table 4 contains the summary of the MANOVA results for grade level and treatment comparisons. No significant grade level effects were found. The significant F-values shown for RNP vs. Text (see column 3, table 4) suggest substantial treatment effects favoring the RNP treatment on the total test and on four of the six subscales: concepts, order, transfer, and estimation. A significant interaction was found between treatment and grade-level for the scores on the equivalence subscale, $F(1,62) = 8.74$, $p = .004$. A simple effects test was performed as a post hoc procedure to determine the source of the interaction (Howell, 1997). When grade-level was held constant, a significant difference between treatments was found for Grade 5, $F(1, 62) = 11.63$, $p = .004$, whereas the equivalence subscale difference was not significant for Grade 4, $F(1, 62) = .33$, $p = .571$. In Grade 5, the RNP group performed better than the CC group, whereas in Grade 4, the performance on equivalence items was similar between the two groups.

Table 4
Summary of MANOVA Results for Total Test and Revised Subscales

	Grade 4 vs. Grade 5 Between Subjects F	RNP vs. Text Between Subjects F	Significant Interactions
Total Test	.6	15.5*	None
Subscales			
Concepts	.6	24.6*	None
Equivalence	1.2	4.9	Grade by Treatment
Order	.3	13.83*	None
Operations (+, –)	2.2	.3	None
Transfer	.87	18.9*	None
Estimation	1.9	10.7*	None

Note. $df = (1, 62)$ for grade level and treatment comparisons. All significance levels tested at $\alpha = .0083$ ($.05 \div 6$) to control for test-wise error using the Bonferroni Procedure.
* $p < .0083$.

Effect sizes. In the absence of a significant grade-level effect on the total test and each of the six subscales, the Grade 4 and Grade 5 data were pooled. The subsequent calculation of means and standard deviations was conducted on RNP (Grade 4 and Grade 5) vs. CC (Grade 4 and Grade 5) for the total test and for the six revised subscales on the posttest and retention test (See Table 5). Effect sizes and power estimates were calculated for the differences between RNP and CC groups. The SPSS MANOVA POWER subcommand was used to determine estimated power values and to calculate partial eta squared as an index of effect size. According to Cohen (1977), an eta squared of .01 is a small effect, .06 is a medium effect, and

.14 is a large effect. The eta-squared values for the total test and 4 of the 6 subscales (concepts, order, transfer, and estimation) were greater than .14, and thus the power values were quite high for the total test and for these four subscales. The large effect sizes and power values indicate that the differences between RNP and CC students are both meaningful and reliable.

Table 5
Means and Standard Deviations for Treatment Groups by Posttest and Retention Test Along With Effect Size and Power Estimates

	RNP Posttest	CC Posttest	RNP Retention Test	CC Retention Test	Effect Size[a]	Power
Total Test	63 (9)	53 (11)	58 (10)	49 (13)	.205	.972
Subscales						
Concepts	56 (14)	40 (15)	52 (13)	38 (15)	.284	.998
Equivalence	64 (14)	58 (17)	63 (17)	56 (19)	.073	.584
Order	80 (11)	70 (11)	75 (9)	68 (11)	.182	.955
Operations (+, –)	78 (8)	78 (13)	63 (12)	59 (17)	.005	.047
Transfer	43 (10)	30 (11)	45 (13)	33 (13)	.240	.990
Estimation	69 (21)	50 (23)	63 (21)	48 (25)	.149	.895

Note. Means and standard deviations (in parentheses) are depicted as percents.
[a]Effect size calculated as eta-squared by SPSS MANOVA POWER command. A large effect would be a value >.14.

Summary of quantitative analysis. From students' performance on the written tests, we conclude that there were significant differences in student achievement between students using the RNP curriculum and students using a commercial curriculum. The significant MANOVA *F* for total test scores suggests that RNP students significantly outperformed students using the commercial curriculum on both the overall posttest and retention test results. Subscale analyses on the posttest and retention test identified four significant differences, all favoring RNP students. These subscales show that the RNP students had a stronger conceptual understanding of fractions, were better able to judge the relative sizes of two fractions, used this knowledge to estimate sums or differences, and were better able to transfer their understanding of fractions to tasks not directly taught to them. No differences were found between RNP and CC students on equivalence items or items dealing with symbolic addition and subtraction. Given that the CC group spent much more instructional time on the operations, this was somewhat surprising. Only 5 of the 23 RNP lessons dealt with addition and subtraction.

Qualitative Analysis

Interview data results. Analyses of the interviews document the nature of the differences in thinking between the RNP group and the CC group. To highlight these differences, we have presented four examples: responses to two ordering questions (same numerator and residual) and to two fraction estimation tasks (addition and subtraction). These types of questions were selected to illustrate differences in the groups' number sense for fractions. Students need a well-internalized concept of the "bigness" of rational numbers; without this conceptual foundation, they cannot operate on fractions in a meaningful way. The ordering questions show how students judge the relative sizes of fractions, and those on estimation provide students with the opportunity to use their understanding of the relative sizes of fractions to construct a reasonable answer to a fraction operation task.

Table 6 presents a summary of results from the classroom teachers' interviews of fourth and fifth graders at the end of the study on a fraction comparison question**.** Students interviewed were asked to select the larger of two fractions, 4/15 or 4/10. Fourth and fifth graders who have developed fraction concepts using manipulative models should select 4/15 as the smaller fraction without relying on procedures like cross products or finding least common denominators. For example, one RNP fourth grader reasoned as follows: "4/15 is less because tenths are bigger. Fifteenths are smaller. It takes more fifteenths to cover the circle; you have equal top numbers."

Table 6
Results From Fourth- and Fifth-Grade Students' Performance on a Problem Comparing Two Fractions With the Same Numerator: Which Is Larger? 4/15 or 4/10?

	RNP Group (n = 95)	CC Group (n = 99)
Correct		
Conceptual	65 (68%)	34 (34%)
Procedural	7 (7%)	23 (23%)
No Explanation	0	3 (3%)
Manipulatives	3 (3%)	2 (2%)
Unclear	2 (2%)	2 (2%)
Wrong Reasoning	2 (2%)	4 (4%)
Totals	79 (83%)	68 (69%)
Incorrect		
Conceptual	3 (3%)	3 (3%)
Procedural	9 (10%)	21 (21%)
No Explanation	1 (1%)	5 (5%)
Manipulatives	3 (3%)	1 (1%)
Unclear	0	1 (1%)
Totals	16 (17%)	31 (31%)

Note. Percentages may not add to 100 because of rounding.

This answer shows the student understood the compensatory relationship between the number of equal parts of a whole and the size of each part. This student made a connection between symbols and a concrete model and coordinated the relationship between the numerator and denominator, and such a response was coded as "conceptual." A procedural response reflected a reiteration of a rule with no reflection on the fractions as numbers. For example, a CC student gave this explanation for why 4/15 is less than 4/10: "You need to multiply 15×4 and 10×4. Whichever one is larger is bigger; whichever one is less will be smaller." This student's reasoning was coded as procedural because it involved the application of a rote algorithm that did not consider the numbers in the problem as fractions.

As shown in Table 6, RNP students were not only more successful than CC students in ordering these two fractions (83% vs. 69%), but they relied on a conceptual ordering strategy more frequently than did the CC students (68% vs. 34%). Although 44 out of the 99 CC students attempted a procedural strategy, 21 did so incorrectly. Sample responses to the ordering question for selected categories are shown in Figure 2.

In examining student responses, we found that RNP students relied on their mental images of fraction circles in their responses—judging the size of each fraction by thinking about the relationship between the numerator and denominator. CC students often applied a common denominator procedure taught to them that did not require consideration of a fraction as a single entity. As a result of disconnecting the fraction symbols from a physical embodiment, students do not discern the relationship between the numerator and denominator or how this relationship affects the size of the number. Students then make errors that are based on whole-number ordering strategies, as shown by their incorrect procedural responses. A larger percentage of CC students than of RNP students made this type of error (21% vs. 10%).

Table 7 presents data from classroom teachers' interviews of fourth graders on another fraction question: Which is larger? 4/5 or 11/12? This question uses a fraction pair that invites a residual ordering strategy, one that focuses on what is left over, or the difference between the fractional part and one unit. To solve this problem without an algorithm, students would need to rely on their mental pictures of the two fractions, deciding on their relative sizes by making a judgment about the amount that each fraction is away from the whole. Our previous work found that 40% to 50% of the fourth graders constructed this residual strategy. Problems of this type were not explicitly addressed in the RNP Level 1 curriculum, but opportunities to construct this strategy are offered to students in the Level 2 book. We included this question to see if any students in this study could construct this strategy in an interview situation prior to instruction.

Because RNP students were not taught in class a least common denominator or cross-product procedure for ordering fractions, they would have to rely on their mental images of fractions and to coordinate both size-of-piece and number-of-piece aspects of those mental images to order this pair. In this example, the one part that each fraction is away from the whole tells something about the relative sizes of the two fractions. In this case, the fraction 4/5 is 1/5 away from the whole and is smaller than the fraction 11/12, which is 1/12 away from the whole. Results show that close

RNP Correct—Conceptual

- The green fraction piece [fifteenths] is smaller than the purple piece [tenths]. You have 4 of each. So since the greens are smaller than the purple then you know it is smaller.
- Because if you divide it into 15 pieces it would be smaller than if you divide it into 10 pieces.

RNP Correct—Procedural

- [Student finds cross products].
- Higher number on bottom of 4/15 and the tops are the same. [No reference to concrete model of to idea of fifteenths.]

RNP Incorrect—Conceptual

- Fifteenths are smaller than tenths. [But still identified 4/10 as smaller fraction.]
- Because 4/10 has more *groups* than 4/5. [4/5 was given instead of 4/15.]

RNP Incorrect—Procedural

- Because 10 is less than 15.
- If you take 4 from 15 it's 11; if you take 4 from 10 that's 6. 6 is way smaller than 11.

CC Correct—Conceptual

- Because 4/15 the pieces have to be smaller because there are more pieces and 4/10, the pieces are bigger.
- The two top numbers are the same so that tells you get the same number of pieces but the bottom numbers tell you which have the bigger pieces. Look which denominator is smaller and that's the bigger pieces. So 4/10 is the bigger fraction.

CC Correct—Procedural

- I got a common denominator. Since $15 \times 2 = 30$ and $10 \times 3 = 30$. Then $15 \times 2= 30$ so $4 \times 2 = 8$ and $10 \times 3 = 30$ so $3 \times 4 = 12$.
- I can reduce 4/10 to 2/5 then change to fifteenths—6/15; so 4/15 is less.

CC Incorrect—Conceptual

- The shaded parts are equal.
- 4/10 is less because 4/15 is in a circle graph—the pie pieces are smaller and in 10 the pieces are bigger. [Inconsistency between answer and reasoning.]

CC Incorrect—Procedural

- Because if you have 4 out of 15, it leaves 11 left. And if you have 4 out of 10, it leaves 6 left.
- Since the both have the same numerator, I just looked at the denominator and picked one that was less.

Figure 2. Sample responses to a question comparing fractions with the same numerator: Which is larger? 4/15 or 4/10?

Note. Authors comments appear in brackets.

Table 7
Results From Fourth-Grade Students' Performance on a Problem Comparing Two Fractions Using a Residual Strategy: Which Is Larger? 4/5 or 11/12?

	RNP Group (n = 53)	CC Group (n = 57)
Correct		
Conceptual	21 (40%)	9 (16%)
Procedural	3 (6%)	18 (32%)
No Explanation	0	3 (5%)
Pictures or Manipulatives	8 (15%)	3 (5%)
Unclear	1 (2%)	5 (9%)
Totals	33 (62%)	38 (67%)
Incorrect		
Conceptual	13 (25%)	5 (9%)
Procedural	1 (2%)	7 (12%)
No Explanation	1 (2%)	2 (4%)
Pictures or Manipulatives	2 (4%)	2 (4%)
Unclear	3 (6%)	3 (5%)
Totals	20 (38%)	19 (33%)

Note. Percentages may not add to 100 because of rounding.

to 40% of the fourth graders interviewed from the RNP group constructed a residual strategy to order these two fraction pairs. Fewer CC students (16%) did this. Although the overall percentages of correct and incorrect responses were close for the two groups (62% RNP; 67% CC), how students thought about the problem differed markedly. RNP students' responses were more conceptual, whereas CC students' responses were more procedural. More RNP students than CC students were able to verbalize their thinking, as noted in the number of responses categorized as no explanation or unclear. Sample student responses to this question are shown in Figure 3.

Table 8 summarizes data from an item on the final interview given by RNP staff to 17 fourth graders.[1] Students were asked to respond to an addition estimation question (2/3 + 3/12) that was embedded in a story problem; the correct answer was judged to be greater than 1/2 but less than 1. Estimation questions offer students an opportunity to use their understanding of the relative sizes of fractions to judge what would be a reasonable result to operating on two fractions. Responses that relied on mental images to determine the relative size of a fraction were coded as conceptual, whereas responses that relied solely on symbolic procedures with no reference to the relative sizes of the two fractions were coded as procedural. Estimation after finding an exact answer also was coded as procedural.

Although the small sample size precludes any definitive conclusions, the results in Table 8 show that the RNP students were more likely than the CC students to

[1] Of the ten RNP students originally selected to be interviewed four times, two were tracked into a low mathematics group and did not finish the lessons. These two students were not given the final interview; therefore, they were interviewed only three times. One CC student did not take the final interview.

RNP Correct—Conceptual

- 4/5 takes a bigger piece to get 1 whole. 11/12 takes a smaller piece to get 1 whole.
- 4/5 has bigger pieces but one piece is left; 11/12 has smaller pieces but only 1 piece left. Fifth piece that is missing is bigger

RNP Correct—Procedural

- [Student finds the cross product.]
- Bottom number is lower than the other one.

RNP Incorrect—Conceptual

- Equal because both 4/5 and 11/12 are one away from the whole.
- 11/12 is less. Twelfths are smaller than fifths. So it would take more pieces to fill a whole circle.

RNP Incorrect—Procedural

- 11/12 is less. Because if you triple 4 it would equal 12 but not 11, so they are not equal. The size of the pieces makes a difference.

CC Correct—Conceptual

- If you have 12 pieces and 12/12 equal 1 whole. Less pieces would be a little smaller. The 12 pieces are so small that when you take one away it's not going to be that big of a gap. But in 4/5 the pieces are bigger, so there would be a bigger gap.
- 4/5 pieces are bigger than 11/12 pieces. So there is a smaller piece left over with 11/12.

CC Correct – Procedural

- [Student finds the cross product.]
- [Student changes to denominators of 60: 48/60 < 55/60.]

CC Incorrect—Conceptual

- Because if you had little piece and you only picked 11 and if you had 5 large pieces and you picked 4 there would be more in 4/5 than 11/12.
- 11/12 is less. Because the equal amount of pieces is 12 and they have to be small. The 4/5 pieces have to be bigger, there are fewer of the big pieces but because they are small it's more.

CC Incorrect—Procedural

- They are equal because 5 is one more than 4 and 12 is 1 more than 11.
- 11/12 is less. I took $5 \times 11 = 55$ and $4 \times 12 = 48$; $55 > 48$.

Figure 3. Sample responses to a question about the relative sizes of fractions: Which is larger? 4/5 or 11/12?

Note. Authors comments appear in brackets.

Table 8
Results From Fourth-Grade Students' Performance on an Addition Estimation Question Given During the Final Interview

Marty was making two types of cookies. He used 3/12 cup of flour for one recipe and 2/3 cup for the other. How much flour did he use altogether? Without working out the exact answer, give an estimate that is reasonable. [If necessary, interviewers asked, Is it greater than 1/2 or less than 1/2? Greater than 1 or less than 1?]

	Correct; Conceptual	Correct; Procedural	Correct; No explanation	Incorrect	Missing data
RNP ($n = 8$)	6	0	0	1	1
CC ($n = 9$)	0	2	2	4	1

use a conceptual strategy and to answer correctly. The CC students gave more correct procedural responses or correct responses without explanations than their RNP counterparts. Some sample responses that illustrate the coding categories appear below:

RNP Correct—Conceptual

- It would be more than 1/2. It would be less than a whole. If you had 2-thirds, that's more than half and then you put 3-twelfths to add to it, it would not be a whole. When asked how she knew it wasn't going to be a whole she said: 3-twelfths isn't very big so you'd add a little more.
- About one. 1/3 is bigger than 1/12. Then 3/12 wouldn't equal 1/3. And you need 2 more thirds to equal a whole.
- Greater than 1/2. It takes 3 reds [twelfths] to cover one blue [fourths] so it probably takes 4 reds to cover a brown [thirds]. So there's only 2 of 3 [browns]. There's a gap when you fill with 3 reds.

RNP Incorrect

- About one. 3-twelfths equals 1-third; 2-thirds plus 1-third equals one.

CC Correct—Procedural

- Greater than 1/2 and less than one. I know how many times this could go into 12 is four and you go four to get the denominator. And it was four times three, you take three times four equal twelve and then two times four equals eight and then you get 8-twelfths. Then you go 8-twelfths plus 3-twelfths equals 11-twelfths and then it's more than 1/2 and less than one.
- Greater than 1/2; less than one. I am just guessing.

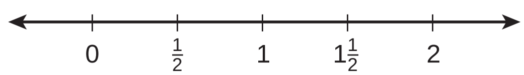
CC Incorrect

- Less than 1/2. [Unable to explain reasoning.]
- More than one. I don't know. It just seems high.

RNP students' thinking depended on mental images of fractions; these images related directly to fraction circles, the manipulative most often used in the instruction. Their thinking modeled the type of number sense that mathematics educators advocate and differed greatly from that of their CC counterparts. Differences in students' abilities to verbalize were evident, with more RNP students than CC students demonstrating an ability to describe their thought processes. But then, RNP lessons emphasized student discussion of ideas and translations to and from the verbal mode of representation. The manipulatives themselves became the focal point for student discussion, with students talking about their actions with manipulatives. The extended use of physical models may have fostered the development of students' verbal skills.

Data from a subtraction estimation question (11/12 – 4/6) given to fourth graders by the classroom teachers is presented in Table 9. The correct answer was determined to be less than 1/2, unless the estimation strategy that the student used rounded 11/12 to one and 4/6 to 1/2; in this case, equal to 1/2 was deemed to be correct. Responses that reflected an understanding of the relative sizes of the fractions were coded as conceptual; responses where students found an exact answer and then estimated were coded as procedural. Responses that applied whole-number counting strategies to parts of the fractions also were coded as procedural.

Table 9
Results From Fourth-Grade Students' Performance on a Subtraction Estimation Problem: Tell Me About Where 11/12 – 4/6 Would Be On This Number Line



	RNP Group (n = 49)	CC Group (n = 34)
Correct		
Conceptual	20 (41%)	2 (6%)
Procedural	5 (10%)	14 (41%)
No Explanation	2 (4%)	0
Pictures or Manipulatives	8 (16%)	0
Unclear	2 (4%)	0
Totals	37 (76%)	16 (47%)
Incorrect		
Conceptual	7 (14%)	3 (9%)
Procedural	2 (4%)	6 (18%)
No Explanation	3 (6%)	6 (18%)
Pictures or Manipulatives	0	0
Unclear	0	3 (9%)
Totals	12 (24%)	18 (53%)
Missing Data	4	23

Note. Missing data for the RNP group represents those students who did not complete all lessons and therefore did not respond to this question. Missing data for the CC group represents teacher error. In particular, teachers asked the wrong question. For unknown reasons, teachers posed the question 11/12 – 4/6 as "identify 11/12 and 4/6 on this number line." Percentages may not add to 100 because of rounding.

Estimation should precede calculation. By using one's number sense for fractions, a reasonable estimate of 11/12 – 4/6 would be less than 1/2. Some students reasoned as follows: The fraction 11/12 is not quite 1, whereas 4/6 is greater than 1/2. An amount just less than 1 minus an amount greater than 1/2 must be less than 1/2. To reason in this manner, students needed to call up images of 11/12 and 4/6 and then operate on these images mentally. The results in Table 9 show that 75% of RNP fourth graders as compared to 47% of CC fourth graders were able to provide a reasonable estimate to this problem, but differences in strategies used by students in the two groups highlight differences in their thinking about how to reach an estimate. Twenty of the 49 RNP responses reflected the conceptual kind of thinking noted above. Only 2 out of 34 CC students estimated by reflecting on the relative sizes of the fractions. In general, CC students estimated after finding the exact answer. Examples of the kind of reasoning and strategies just described and other examples used by the RNP and CC students can be seen in Figure 4.

Summary of Interview Data

Analyses of the interview data showed important differences in student thinking. Percentages of conceptually oriented responses were higher for RNP students than for CC students. Results showed that a large percentage of the students in classrooms where teachers used the RNP lessons developed a strong conceptual understanding of fractions. These students had mental representations for the symbols and used these representations to determine the relative sizes of fractions. They consolidated their understandings to estimate sums or differences of two fractions by judging the fractions' relative sizes and then mentally operating on them. In addition, data from interviews showed that most of the RNP incorrect responses were conceptually based, suggesting that many RNP students were well on their way to developing these understandings.

The percentages of conceptually oriented responses are considerably lower for the CC group and reflect the method of instruction. Instruction provided by the commercial curricula emphasized procedural skill and offered students limited opportunities with manipulative models. Results from the interview task shown in Table 6 are especially revealing. Only 34% of the CC students ordered 4/10 and 4/15 conceptually as compared to 68% of the RNP students. Given that CC students did not construct a well-internalized concept of the size of fractions, it is not surprising that as a group they did not apply number sense to find an estimate to fraction operation tasks but instead relied on procedures that did not include judging the relative sizes of the numbers.

DISCUSSION

The purpose of this study was measure and contrast the impacts of two different curricula on the achievement of initial fraction ideas by fourth and fifth graders. One

RNP Correct—Conceptual

- 11/12 is more than 4/6. I know that 11/12 is nearly one whole thing and 4/6 is way over 1/2. So when I take 4/6 away from 11/12 I will get way less than 1/2.
- I was thinking about reds for twelfths and 11 is really close to a whole and for sixths, four sixths is almost one whole. I tried to take away sixths and I thought it would be close to, less than 1/2.

RNP Correct—Procedural

- I changed 4/6 to 8/12 and 11 – 8 = 3 so 3/12 is less than 1/2. You need three more pieces to equal 1/12.

RNP Incorrect—Conceptual

- I took 1/12 and thought of 4/6. 3/6 is 1/2 of 6 and take away 1/2 and then add on 11/12.
- 11/12 is one away from the whole, 4/6 is two away from a whole so I think it will be a little over 1/2.

RNP Incorrect—Procedural

- 11/12 is almost a whole. 4/6 changes into 3/12 and subtract from 11/12 equals 9/12.

CC Correct—Conceptual

- 11/12 is close to 1 whole and 4/6 is close to 1/2 so you can subtract 1/2 from one. So the answer is about 1/2.
- 11/12 is about a whole. 4/6 is about 1/2 so I put the “X” by 1/2.

CC Correct—Procedural

- The answer is 3/12 and that’s the same as 1/4. 1/4 is less than 1/2.
- Common denominator is 12. 11 – 8 = 3; that is 3/12 = 1/4.

CC Incorrect—Conceptual

- 11/12 is close to one and 4/6 is more than half which means the answer is more than 1/2.

CC Incorrect—Procedural

- I changed 4/6 to 8/12 and then subtracted from 11/12 and I get 3/12. [Marked answer as greater than 1/2 but less than 1.]
- I took 11 – 4 = 7. I added 12 and 6 and got 18 so I got 7/18 and I thought 11/18 is between 1/2 and 1.

Figure 4. Sample responses to a question about the location of 11/12 – 4/6 on a number line
Note. Authors comments appear in brackets.

group used one of two district-supported commercial curricula for the learning of fractions, whereas the other group used the RNP fraction curriculum, which emphasized the extended use of multiple physical models and translations within and between other modes of representation-pictures, written symbols, verbal symbols, and real-world contexts. Written tests and interview tasks were used to measure student achievement and describe differences in students' thinking on fraction tasks.

The results show the importance of providing students with instruction that involves multiple representations—particularly multiple manipulative models—over extended periods to help them develop initial fraction ideas. With just 4 hours of staff development, classroom teachers were able to implement the RNP fraction curriculum effectively. Students in these classrooms demonstrated a number sense for fractions that was similar to understandings documented in earlier RNP teaching experiments in small group and classroom settings. In particular, students using RNP project materials had significantly higher mean scores on the posttest and retention test and had higher mean scores on four (of six) subscales: concepts, order, transfer and estimation. No significant differences were found on equivalence items or symbolic operation tasks. We actually expected RNP students to do less well than CC students on the operation tasks asking for exact answers, because CC students spent considerably more time on the topic. We conclude that the conceptually focused experiences gave students the foundation to tackle procedural tasks successfully even though they spent less classroom time learning the procedural skills. RNP students' development of procedural knowledge was apparently not impeded despite their having devoted very limited classroom time to it.

Interview data showed differences between the two groups in students' thinking about fractions. RNP students, in general, approached order and estimation tasks conceptually by building on their mental images of fractions; these mental images most often described the fraction circle model. Compared to RNP students, CC students relied on procedures more often to solve order and estimation tasks.

The program of study used by the CC students did not include a wide variety of materials or regular use of hands-on manipulative experiences but focused instead on pictorial and symbolic modes of representation. Students learned fraction ordering by finding common denominators or cross products, and equivalence was treated as a symbolic manipulation involving multiplying by n/n. Developing a flexible concept of unit and a quantitative concept of fraction was not seriously addressed in either of the CC programs. There were substantial differences in the amounts of time devoted to various subtopics, as well. For example, in the RNP group, a large amount of time was devoted to developing an understanding of the meaning of the fraction symbol by making connections between the symbols and multiple physical models. The RNP students developed order and equivalence ideas concretely. Of the 23 lessons, only 5 lessons dealt with fraction operations.

In both commercial series, the major concern was the development of proficiency in operations with fractions—finding equivalent fractions, adding, subtracting, finding common denominators. There also were substantial differences in the presentations of these topics. For example, in the RNP group, symbols and their manipulation were closely tied to physical objects and physical transformations on these objects. Students also made connections among symbols, pictures, and contexts, with language facilitating translations among different representations.

In the CC group, symbol manipulation seemed to be an end in itself, independent of context and physical models.

Goals for fraction instruction can be set that reflect the more conceptual framework espoused by the NCTM *Standards* than the procedurally-oriented goals set by more conventional approaches as exemplified in these two commercial curricula. The *Principles and Standards for School Mathematics* (NCTM, 2000) supports the notion that Grades 3–5 are the critical years for developing a solid conceptual framework for work with rational numbers. This foundation is intended to undergird the development of computational fluency with rational numbers, which begins in Grades 3–5 but should be the major focus in Grades 6-8. Teachers in Grades 3–5 can reach more conceptually oriented goals by using a curriculum that provides children with learning experiences that offer multiple modes of representations and with multiple concrete models.

As shown in this study, teachers may not always need extensive preparation to use such a new curriculum effectively. Recall, though, that the 66 teachers volunteered to participate in this study. This may represent a biased sample of teachers who were looking for ways to teach fractions differently and more effectively. The random assignment procedures that the study used, however, essentially eliminated any treatment, classroom (teacher), or school bias that may have existed.

The assessment instrument was more closely aligned with the RNP curriculum. Although the assessment accurately reflected district goals and those goals presented in the NCTM *Curriculum and Evaluation Standards* (1989), the tests did not measure procedural skill to the same extent that they measured conceptual knowledge. Although our assessments contained items that could be solved from either a conceptual or procedural perspective, additional items assessing procedural skill may have shown greater strength in this area among the CC students than among the RNP students.

What was the nature of the materials that underlie these results? First of all, the RNP curriculum is organized so that teachers had little extra work to do in order to implement lessons. RNP lessons explain in a great deal of detail how to introduce, develop, and reinforce ideas. A second important aspect of the RNP lessons is the integrated set of extended comments for the teacher in the margins regarding anticipated student-learning problems and suggestions for resolving these issues. Many of these problem areas were first illuminated in our own earlier work with children (Behr et al., 1984; Bezuk et al., 1989; Post et al., 1985; Post et al., 1992) and discussed in several of the RNP publications. Other issues were gleaned from the relatively extensive literature base in the domain (Brown, 1993; Kieren, 1976; Mack, 1993; Sowder, Bezuk, & Sowder, 1993). Our interview notes indicated that teachers appreciated, profited from, and used these insights and suggestions.

Fractions represent a new number system that is based on multiplication, rather than addition. New rules and relationships exist, and students need time to sort through these ideas using a variety of physical representations. RNP lessons actively involved children over extended periods of time with concrete models for fractions and provided extended opportunities for students to talk to one another and to their teacher about their evolving understandings. The RNP materials incorporated regular and sustained attention to translations within and among five modes of representation as suggested by Lesh (1979). This translation model extends the ideas of Bruner (1966) and adds verbal and real-world modes of representation to Bruner's original enactive, iconic, and symbolic modes. The mathematical ideas developed were represented in all five modes with focus on the manipulation of physical models and determining similarities and differences among these models as suggested originally by Dienes (1969) in his mathematical and perceptual variability principles.

Of equal importance is the interactive nature of the five modes and in particular the translations within and between them. Thus, a child might be given a fraction circle showing 2/3 and asked to show "that" fraction with chips and to reflect on how the two models are alike and different. Translating from one manipulative model to another in this way is a qualitatively different and intellectually more demanding task than showing 2/3 using a single model. Similarly, a child might be asked to model a story problem with chips and to explain in his or her own words how and why chips can be used to model the story problem. This task involves translating from a real-world to a manipulative to a verbal model. A basic tenet of earlier RNP work was the belief that it is the translations within and between modes of representation that make ideas meaningful for children (Behr et al., 1984; Post et al., 1985). We believe that some degree of cognitive disequilibrium is necessary for concept development within the domain of rational number (Behr et al., 1994) and that translations provide students with an opportunity to overcome aspects of their own cognitive disequilibrium as related to these new ideas. This study suggests students do profit from these translation activities, and that materials can be written in such a way that classroom teachers can effectively implement lessons based on a translation model.

In summary, we believe RNP teachers were able to be effective because of the careful structure of the materials and the integration of information on student thinking into each lesson. We believe these students learned disproportionately well because they spent their time interacting with fraction ideas in multiple ways and were provided extended periods of time to develop an understanding of the meaning of symbols. Investing time in fostering students' understanding of rational number concepts was shown to be an effective method for developing a quantitative sense for fractions in large numbers of fourth and fifth graders. The theoretical framework and lesson structure employed here can be applied to other mathematical topics and shows promise in enabling large numbers of teachers to participate in the curriculum regeneration process and to improve student understandings.

REFERENCES

Addison-Wesley Publishing Company (1989a). *Addison-Wesley mathematics: Grade 4*. Menlo Park, CA: Author.

Addison-Wesley Publishing Company (1989b). *Addison-Wesley mathematics: Grade 5*. Menlo Park, CA: Author.

Behr, M., Harel, G., Post, T., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296–333). New York: Macmillan.

Behr, M., Lesh, R., Post, T., & Silver, E. A. (1983). Rational number concepts. In R. Lesh & M. Landau (Eds.), *Acquisitions of mathematics concepts and processes* (pp. 92–126). New York: Academic Press.

Behr, M., & Post, T. (1986). Estimation and children's concept of rational number size. In H. Schoen & M. Zweng (Eds.), *Estimation and Mental Computation: 1986 yearbook* (pp. 103–111). Reston, VA: National Council of Teachers of Mathematics.

Behr, M., Wachsmuth, I., Post, T., & Lesh, R. (1984). Order and equivalence of rational numbers: A clinical teaching experiment. *Journal for Research in Mathematics Education*, *15*, 323–341.

Bezuk, N., & Cramer, K. (1989). Teaching about fractions: What, when and how? In P. Trafton (Ed.), *New directions for elementary school mathematics: 1989 yearbook* (pp. 156–167). Reston, VA: National Council of Teachers of Mathematics.

Brown, C. (1993). A critical analysis of teaching rational numbers. In T. Carpenter, E. Fennema, & T. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 197–218). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, S. I., Cooney, T. J., & Jones, D. (1990). Mathematics teacher education. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 639–656). New York: Macmillan.

Bruner, J. (1966). *Toward a theory of learning*. New York: Norton.

Burns, M. (1993). *Mathematics: Assessing understanding* [Videotape]. White Plains, NY: Cuisenaire Company of America.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cramer, K., Behr, M., Post, T., & Lesh, R. (1997a). *The Rational Number Project: Fraction lessons for the middle grades, Level 1*. Dubuque, IA: Kendall/Hunt Publishing Co.

Cramer, K., Behr, M., Post, T., & Lesh, R. (1997b). *The Rational Number Project: Fraction lessons for the middle grades, Level 2*. Dubuque, IA: Kendall/Hunt Publishing Co.

Dienes, Z. (1969). *Building up mathematics* (Rev. ed). London: Hutchinson Educational.

Harcourt, Brace, & Jovanavich (1992a). *Mathematics plus: Grade 4*. Orlando, FL: Author.

Harcourt, Brace, & Jovanavich (1992b). *Mathematics plus: Grade 5*. Orlando, FL: Author.

Hiebert, J. (1994). A theory of developing competence with written mathematical symbols. *Educational Studies in Mathematics*, *19*, 333–355.

Howell, D. C. (1997). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Books.

Kieren, T. (1976). On the mathematical, cognitive, and instructional foundations of rational numbers. In R. Lesh & D. Bradbard (Eds.), *Number and measurement: Papers from a research workshop* (pp. 101–144.). Columbus, OH: ERIC/SMEAC.

Kouba, V., Zawojewski, J., & Strutchens, M. (1997). What do students know about numbers and operations? In P. A. Kenney & E. A. Silver (Eds.), *Results from the sixth mathematics assessment of the National Assessment of Educational Progress* (pp. 87–140). Reston, VA: National Council of Teachers of Mathematics.

Lesh, R. (1979). Mathematical learning disabilities: Considerations for identification, diagnosis, and remediation. In R. Lesh, D. Mierkiewicz, & M. G. Kantowski (Eds.), *Applied mathematical problem solving* (pp. 111–180). Columbus, OH: ERIC/SMEAC.

Mack, N. K. (1993). Learning rational numbers with understanding: The case of informal knowledge. In T. Carpenter, E. Fennema, & T. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 85–105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Math Learning Center. (1988). *Math and the mind's eye*. Salem, OR: Author.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and experimental curriculum. *Journal for Research in Mathematics Education, 30*, 122–147.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Piaget, J. (1960). *The psychology of intelligence*. Patterson, NJ: Littlefield, Adams.

Post, T., Behr, M., & Lesh, R. (1982). Interpretations of rational number. In L. Silvey (Ed.), *Mathematics for the middle grades (5–9): 1982 yearbook* (pp. 59–172). Reston, VA: National Council of Teachers of Mathematics.

Post, T., Cramer, K, Behr, M., Lesh, R., & Harel, G., (1992). Curricular implications of research on the teaching and learning of rational numbers concepts. In T. Carpenter, T., E. Fennema, & T. Romberg (Eds.), *Research on the teaching, learning and assessing of rational number concepts* (pp. 327–362). Hillsdale NJ: Lawrence Erlbaum Associates.

Post, T., Wachsmuth, I., Lesh, R., & Behr, M. (1985). Order and equivalence of rational numbers: A cognitive analysis. *Journal for Research in Mathematics Education*, *16*, 18–37.

Rational Number Project. (2001). [On-line]. Available: http://rationalnumberproject.education.umn.edu/

Schoenfeld, A. H. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of videotapes. *Journal of the Learning Sciences*, *2*(2), 179–214.

Sowder, J. T., Bezuk, N., & Sowder, L. K. (1993). Using principles from cognitive psychology to guide rational number instruction for prospective teachers. In T. Carpenter, E. Fennema, & T. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 239–259). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Authors

Kathleen A. Cramer, 252 Peik Hall, 159 Pillsbury Drive, University of Minnesota, Minneapolis, MN 55455; crame013@tc.umn.edu

Thomas R. Post, 240 Peik Hall, 159 Pillsbury Drive, University of Minnesota, Minneapolis, MN 55455; postx001@tc.umn.edu

Robert C. delMas, 345 Appleby Hall, 128 Pleasant Street, SE, University of Minnesota, Minneapolis, MN 55455; delma001@tc.umn.edu

APPENDIX A

Scope and Sequence for RNP Curriculum—Level 1

LESSON	MANIPULATIVE	TOPICS
1	Fraction Circles	Exploration with circles
2	Fraction Circles	Model and verbally name 1-half, 1-third, 1-fourth
3	Fraction Circles	Model and verbally name unit fractions with denominators greater than 4
4	Paper Folding	Compare paper folding to fraction circles. Model and name (verbally and with written words) unit and non-unit fractions

LESSON	MANIPULATIVE	TOPICS
5	Fraction Circles	Model fractions and record with symbols a/b
6	Fraction Circles	Model the concept that the greater number of parts a unit is divided into the smaller is each part
7	Paper Folding	Reinforce the concept that the greater number of parts a unit is divided into the smaller is each part
8	Fraction Circles	Fraction equivalence
9	Fraction Circles/Pictures	Fraction equivalence
10	Paper Folding	Fraction equivalence
11	Fraction Circles	Order fractions by comparing to 1/2
12	Chips	Introduce new model for fraction < 1 by connecting to familiar model
13	Chips	Model fractions using several units for same fraction
14	Chips	Model fractions; determine fractions that can be shown given a set of chips
15	Chips	Fraction equivalence
16	Fraction Circles	Reconstruct the unit given the fraction part
17	Fraction Circles	Model fractions > 1 using mixed and improper notation
18	Fraction Circles	Fraction equivalence for 1/2 based on number pattern
19	Fraction Circles	Estimate sum of two fractions within story contexts
20	Fraction Circles	Find sum of two fractions using fraction circles
21	Fraction Circles	Estimate and solve concretely fraction subtraction using “take-away” and “difference” contexts
22	Fraction Circles	Estimate and solve fraction subtraction using “difference” and “how many more” contexts
23		Summary activities to tie together students’ quantitative understanding of symbols and addition and subtraction

APPENDIX B

Scope and Sequence for the Commercial Curricula Used in the Study

Addison-Wesley: Grade 4		Holt: Grade 4	
Topic	Pages	Topic	Pages
Parts of a region	3	Parts of a region	4
Parts of a set	2	Parts of a set	2
Equivalent fractions	4	Equivalent fractions	2
Lowest term fractions	2	Lowest term fractions	2
Comparing fractions	2	Comparing fractions	2
Fraction of a number	4	Fraction of a number	2
Mixed numbers	2	Mixed numbers	2
Addition and subtraction		Addition and subtraction	
Like denominators	5	Like denominators	4
Unlike denominators	4	Unlike denominators	4
Mixed numbers	2	Mixed numbers	2
Problem solving	8	Estimation	2
Probability	1	Problem solving	11
Review/chapter test	6	Probability	2
		Fractions & measurement	2
		Review/chapter test	7

AW: Grade 5		HBJ Grade 5	
Topic	Pages	Topic	Pages
Parts of region/set	2	Parts of region/set	2
Equivalent fractions	2	Equivalent fractions	4
GCF/LCM	4	GCF/LCM	4
Lowest term fractions	2	Lowest term fractions	2
Comparing fractions	2	Comparing fractions	4
Mixed/improper nos.	4	Mixed/improper nos.	9
Addition and subtraction		Addition and subtraction	
Like denominators	2	Like denominators	2
Unlike denominators	4	Unlike denominators	8
Mixed numbers:	10	Mixed numbers	6
(some renaming)		(some renaming)	
Problem Solving	9	Problem solving	11
Review/chapter test	5	Estimation	2
		Fraction of number	2
		Fraction multiplication	4
		Fraction division	2
		Review/chapter test	6

Telegraphic Reviews

Becoming a Reflective Mathematics Teacher. (2002). Alice F. Artzt and Eleanor Armour-Thomas. Mahwah, NJ: Lawrence Erlbaum Associates. 255 pp. ISBN 0-8058-3036-7 $59.95 (hb.); 0-8058-3037-5 $27.50 (pb.).

Analysis of one's own and others' teaching is an important component of secondary mathematics methods courses and most student teaching experiences. Artzt and Armour-Thomas's book uses research-based frameworks to establish foci for observations of and reflections on teaching. The authors indicate the impact of teachers' knowledge, beliefs, and goals on their planning, monitoring and regulating, and evaluating and revising of instruction. The lesson dimensions of task, discourse, and learning environment are carefully explained, and citations of relevant research are included for each dimension. Although the authors advocate and describe a student-centered teacher model, their observation schemes are useful in describing a spectrum of models, from strongly student-centered to strongly teacher-centered. Early chapters explain and illustrate the frameworks, and later chapters provide detailed assignments for observing a specific aspect of teaching (e.g., nature of the content, nature of the discourse, homework). Related appendices present helpful descriptions of user-friendly data collection methods. After the aspects are treated individually, a section outlines an assignment involving the observation of all of them. The concluding chapters discuss how to use the frameworks for analyzing one's own teaching and how such analysis can lead to accounts for a portfolio. Although there is a certain amount of vocabulary that students will be likely to find unique to the text (e.g., *preactive*, *postactive*), it would not seem likely to interfere with students' understanding. As the title suggests, the book is written specifically for mathematics education and is rich in relevant examples. The observation assignments will be helpful for instructors whose preservice methods courses include observations or who have seminars in conjunction with student teaching. The book could also be used to engage inservice teachers in viewing and reflecting on their practice.

The Teaching and Learning of Mathematics at the University Level: An ICMI Study. (2001). Derek Holton (Ed.). Boston: Kluwer Academic. 560 pp. ISBN 0-7923-7191-7 $190 (hb.).

Numerous books have been written about changes that have taken place or might take place in elementary and secondary mathematics education. Although some books describe research on the learning of specific content at the college level, fewer describe reform efforts in teaching at that level. This book, a report from the *ICMI Study on the Teaching and Learning of Mathematics at the University Level*, includes descriptions of teaching initiatives in many countries. Editor Holton notes that the larger and more diverse student population now entering universities has not resulted in a greater number of mathematics majors. He believes, therefore, that university mathematicians "are more likely to take an interest in mathematics education and what it offers" (p. v). Thus, the study was intended to provide a forum for mathematicians and mathematics educators to exchange and discuss pedagogical ideas. The book acquaints mathematicians with methods of teaching other than those that they themselves experienced. The two pieces comprising the book's introduction argue the need for such change in teaching methods. Other sections describe different teaching practices, the use of technology, types of assessment, and teacher education. A section on research provides an introduction to the types of questions asked and the methodologies employed in mathematics education research. Examples of research in the learning of linear algebra and calculus help to illustrate research issues, findings, and implications.

Copyright © 2002 The National Council of Teachers of Mathematics, Inc. www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in any other format without written permission from NCTM.

Unlocking the Clubhouse: Women in Computing (2002). Jane Margolis and Allan Fisher. Cambridge, MA: MIT Press. 180 pp. ISBN 0-262-13398-9 $24.95 (hb.).

This book, like *The Teaching and Learning of Mathematics at the University Level*, should be read by college instructors in mathematics and mathematics education alike. Jane Margolis, a social scientist, and Allan Fisher, an associate dean of computer science, report on a multi-year study aimed at understanding women's experiences in computer science education. Although the book focuses on women's college experiences, specifically at Carnegie Mellon University, it also gives attention to claims about the roles of nature and nurture in children's early preferences, as well as of environments in the home and school. The authors document the ways in which Carnegie Mellon used this research to increase the enrollment and persistence of women in the university's computer science major. These efforts included changes in curriculum, teaching, recruitment, and departmental culture. Carnegie Mellon's training of teachers for high school courses in advanced placement computer science involved instruction in both computer language and research on factors related to girls' participation in computer science. The authors report on the ways that these high school teachers found to foster girls' enrollment. Because there are parallels between women's experiences and beliefs related to computer science and their experiences and beliefs related to mathematics, the suggestions that the book offers to those engaged in the teaching of computer science to women in college and high school may also be valuable to those committed to increasing the participation and success of women in mathematics.

Why Schools Matter: A Cross-National Comparison of Curriculum and Learning. (2001). William H. Schmidt, Curtis C. McKnight, Richard T. Houang, HsingChi Wang, David E. Wiley, Leland S. Cogan, and Richard G. Wolfe. San Francisco: Jossey-Bass. 424 pp. ISBN 0-7879-5684-8 $27.00 (hb.).

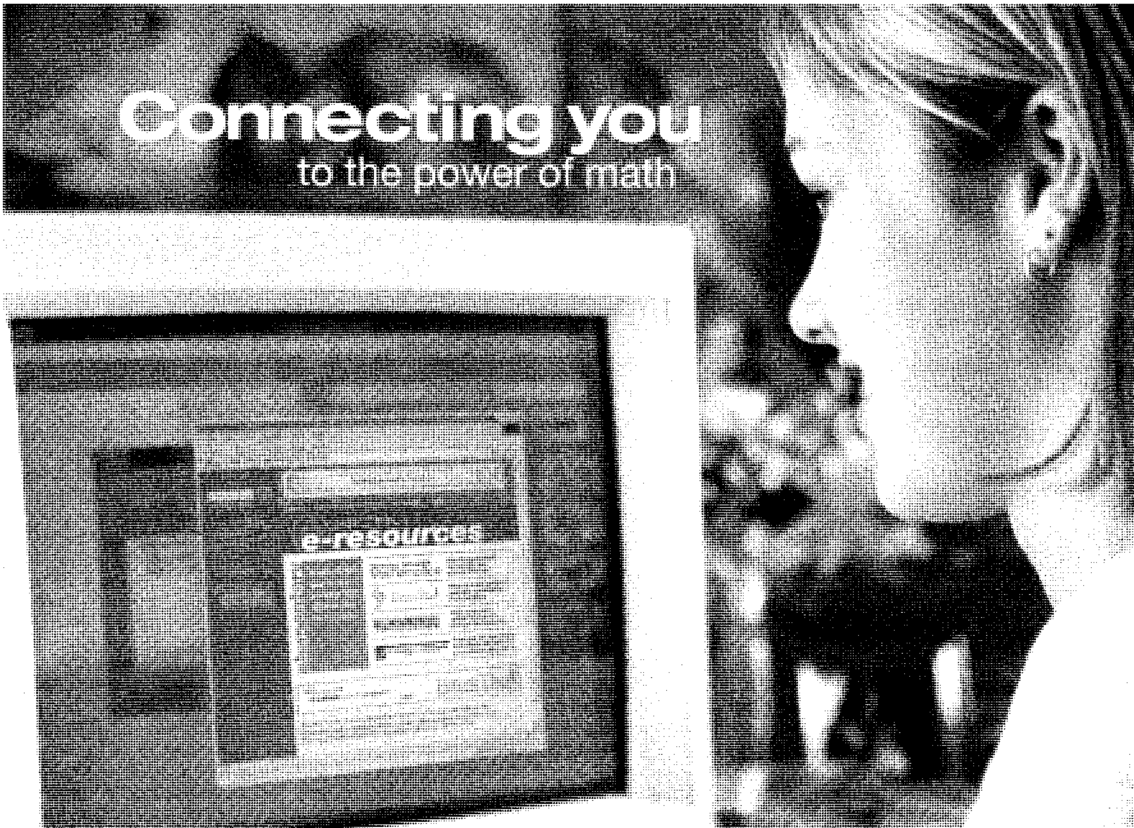
Using data from the Third International Mathematics and Science Study (TIMSS), the authors argue that schools matter because the curricular learning opportunities that they provide have a profound impact on what students actually learn. Using data on content standards, textbooks, teachers' goals, and the amount of time that teachers devoted to topics, the authors present numerous interesting cross-national comparisons of the mathematics and science that 13-year-olds (for the United States, seventh and eighth graders) are expected to learn. However, the heart of the argument about the influence of curriculum and textbooks makes use of the achievement gains between the two adjacent grade levels within countries (e.g., the difference between United States eighth graders' and seventh graders' scores on a particular topic). Examining the relationship between achievement gains and the allocation of curriculum resources, both across countries and within countries, leads the authors to a number of conclusions about factors on which the schools can have an impact. Though the authors acknowledge that other cultural and systemic factors are also likely to shape student achievement, they argue that textbooks have a strong influence on what is taught, that curricula should be organized to take advantage of the discipline's logic, and that both quantity and quality (level of cognitive demand of tasks and type of instructional activity) of coverage are important. The authors also suggest that a set of national priorities in content standards can be advantageous and not be equivalent to national control of a school system. Given the proposed increase in achievement testing in the United States, it is surprising that the book's thorough discussion of curriculum includes very little about the role of external examinations. The study claims significant contributions to cross-national comparative studies in that it focuses on achievement rather than status and employs new methods for studying key aspects of the curriculum.

Learning Policy: When State Education Reform Works. (2001). David K. Cohen and Heather C. Hill. New Haven: Yale University Press. 224 pp. ISBN 0-300-08947-3 (hb.) $30.00.

Reform in mathematics education is widely attempted but less frequently studied. The 1985 *Mathematics Framework for California Public Schools* outlined a curriculum that was designed to be responsive to calls for more attention to knowledge of the discipline and to students' thinking. In addition, it called for classroom practices that differed from conventional ones. Cohen and Hill analyzed the 1994 survey responses of some 595 California teachers as well as the 1994 scores of fourth graders from the California Learning Assessment System as a means of discerning the impact of different instruments of policy on teachers' knowledge of reform, teachers' practice, and students' achievement. Focusing on teachers' opportunity to learn, the authors conclude that "teachers' opportunities to learn about reform do affect their teaching, when those opportunities are situated in curriculum and assessments designed to be consistent with the reforms, and which their students study or use.… When the assessment of students' performance is consistent with the student and teacher curriculum, teachers' opportunities to learn pay off in students' math performance" (p. 148). The authors note that in addition to affecting teachers' opportunities to learn specific content directly related to their instructional materials, policy influenced their practice only if two other conditions applied. First, there was coherence, not only across state frameworks and assessment but also within the classroom, and between the curriculum and teachers' knowledge of how students might approach the curriculum. Second, there were effort and risk taking on the part of "some professionals and professional organizations" to mobilize "temporary communities to develop policy instruments and otherwise support change in practice" (p.186). The authors acknowledge the limitations of their data, indicate where their results are or are not consistent with results of other studies, and share their understandings of how policy was or was not translated into practice. Thus, the book is far more than a history of California's mathematics curriculum frameworks, replacement units, professional development activities, and assessments. The authors cite three challenges facing those who wish to improve instructional policies and research on them: (a) "coordination of government and professional action," (b) "learning from efforts to improve instruction," and (c) "creating better professional development and knowledge about it" (pp. 187–188). They argue that reformers and educators must work in a more responsible manner and encourage independent research on teacher learning and professional development activities, or they will see the demise of professional development.

Learning and Teaching Number Theory (*Mathematics, Learning, and Cognition*). (2002). Stephen R. Campbell and Rina Zazkis (Vol. Eds.) and Carolyn Maher and Robert Speiser (Series Eds.). Westport, CN: Ablex. 245 pp. ISBN 1-56750-652-6 (hb.) $75.00; ISBN 1-56750-653-4 (pb.) $39.50.

Each of the chapters in this volume describes how students understand ideas in elementary (domain of integers) number theory. Using data from preservice teachers and undergraduate students, the authors describe either the conceptions and methods of students who exhibit a range of understandings about a variety of concepts (including prime factorization, divisibility, inductive proof) or teaching strategies that are based on students' understanding of some of these number theory concepts. Campbell and Zazkis note that all the studies reflect a constructivist tradition. Although the editors found Dubinsky's Action-Process-Object schema theory (APOS) helpful in interpreting some aspects of students' understandings in number theory, they also note limitations that they encountered in using this theory. They suggest that considering number theory as a conceptual field (in the sense of Vergnaud's notion of a conceptual field) might be helpful to sense making about it in the long run. In the final chapter, Annie and John Selden identify and expand on some common threads in the chapters: the potential that topics in number theory offer for teaching problem solving, reasoning, and proof; the role of the language of divisibility and the concept images evoked by that language; implications for teaching; the influence of philosophical positions; and the APOS framework. The Seldens suggest some ways in which a constructivist perspective may limit the types of questions that have been pursued within (and outside) the area of number theory. They argue that the term *entity*, encompassing objects, properties, activities and relationships, might be considered as a replacement for *object* in the APOS framework. And they introduce a notion of efficacy that may or may not be characteristic of an individual's action, process, or object construction.



NCTM membership is going online with *e-resources*—designed specifically with you (our members) in mind. It provides you with access to our database of award-winning journals, NCTM's *Principles and Standards*, online mathematics resources, and more. Plus, you will be able to update your membership profile, from address changes to renewing online—it's all at your fingertips.

e-resources offers Individual members access to:

- NCTM journal articles for which you subscribe—*Teaching Children Mathematics*, *Mathematics Teaching in the Middle School Mathematics Teacher*, and *Journal for Research in Mathematics Education*. Search for articles by key words, title, and author.
- *Student Math Notes*—ready-to-use mathematics activities for student's grades 5–10
- Your membership profile—update your address, interests, renew and more
- Direct links to additional Teacher Resources including:
  - Figure This!—math challenges for middles grades
  - Illuminations—projects, lessons and Web sites that illuminate the *Standards*
  - *Principles and Standards*—full searchable text and electronic examples
- Future Member Services under development:
  - ON-Math—an online journal for math teachers and educators with interactive applets and more
  - Plus, Personalized entry pages that will allow you instant access to information that interests you!

National Council of Teachers of Mathematics

(800) 235-7566 · (703) 620-9840 · www.nctm.org

It's membership made easy—www.nctm.org/eresources

In This Issue