TESTING AND FOURTH GRADE TEACHING

Roberta Y. Schorr
Rutgers University
New Jersey USA

Sylvia Bulgar
Rider University
New Jersey USA

William A. Firestone
Rutgers University
New Jersey USA

Advocates of testing often maintain that tests can provoke teachers to adopt new instructional practices. The purpose of this paper is share results from a large scale study designed to investigate the teaching practices of a group of fourth grade teachers who are all involved in state mandated testing programs. We found that while teachers are aware of the test and have made some instructional changes in terms of specific teaching strategies, the changes that have been made tend to focus mostly on strategies and techniques such as the use of small group instruction or manipulatives rather than changes in, for example, the nature of the discourse that takes place in the classroom

Introduction and Framework: In the last two decades, more and more states within the United States have either introduced or extended their testing of children (Editorial Projects in Education, 2001). This trend, however, is not limited to the United States (Niss, 1996; Keitel, and Kilpatrick, 1998 as cited in Abrantes, 2001; Firestone and Mayrowetz, 2000; Abrantes, 2001). Some advocates of tests consider them to be part of a broader effort to raise educational standards and make educators accountable for reaching them. They see testing as a way to use the authority of the state to improve teaching and learning and enhance equity by holding all children accountable to the same high standards (O'Day & Smith, 1993). Some maintain that a test that is well designed can prompt teachers to revise their practices because teachers will inevitably "teach to the test", and that can be good if the test is well designed. Others maintain that content that is emphasized on tests gets emphasized in class, and that untested content either falls out of the curriculum or gets put off until the end of the year (Corbett & Wilson, 1991; McNeil, 2000). The types of items that are placed on the test are also claimed to influence the types of problems teachers use in class. The argument is that by including items that require students to solve more complex types of problems, teachers will be more likely to provide students with the opportunity to do the same in class. One reason given for the great interest in various forms of performance assessment and portfolios in the 1990s was the hope that tasks requiring students to show their work and explain their answers would promote inquiry-oriented instructional approaches (Resnick & Resnick, 1992; Rothman, 1995). Currently many tests combine conventional, multiple- choice formats with other formats intended to measure higher order thinking and problem solving abilities. However, even when tests employ formats where students construct responses, some of the same risks that are typical of the more traditional tests have been found to occur (Smith, 1996; Stecher & Barron, 1999).

Some opponents believe that extensive testing will encourage measurement of less relevant skills, and reinforce traditional approaches to teaching (McNeil, 2000). There are also those who believe that the effects of state tests have been overstated and that any modest changes in teaching exist alongside what has been conventional practice (Wilson & Floden, 2001). Regardless of the format, the evidence that testing promotes instructional change remains unconvincing or inconclusive at best (Newmann, Bryk, & Nagaoka, 2001; Smith, 1996).

In this research[1], we were interested in learning more about teachers reactions to a standardized state test designed to encourage them to implement more student-centered instructional practices that reflect state and national standards (c.f. NCTM, 2000), and what kinds of practices they were actually implementing in their classrooms. Would the teachers, for instance, adopt practices that are associated with reforms and found to be more likely to help students to develop a deeper understanding of mathematics (Davis, 1984; Cobb, Wood, Yackel & McNeal,1993; NCTM, 2000; Klein and Tirosh, 2000; Schorr, 2000)? Would they adopt specific strategies, like using more manipulatives? Would they simply ignore the test and not make any changes? Most importantly, we were also interested in how the changes that they might report actually manifested themselves in the context of their actual classroom instruction. More specifically, in this research, we were interested in gaining a deeper understanding of how teachers feel that the fourth grade test is impacting their teaching, and also in learning more about their actual mathematics teaching. This paper presents data regarding the actual practices of a group of teachers and their perceptions of how the test has influenced their teaching. We looked for evidence that teachers were incorporating approaches that would provide opportunities for students to learn mathematics as they were engaged in meaningful mathematical activity. Our approach in this research was to observe a sample of 4th grade mathematics teachers from across the state, and interview them about their practice and their reactions to the test (The test that we focus on is New Jersey's fourth grade Elementary School Performance Assessment (ESPA) that has been in place since 1999.) which include any changes that the test has prompted in their teaching. This paper describes the results. The particular codes and methods of analysis will be further described below. Our data suggest that the teachers we interviewed and observed are adopting new procedures as part of their instructional practice but not changing their basic approach to teaching mathematics. For example, they tend to continue to assign tasks that reinforce a procedural view of

[1] A portion of this paper was presented, in a preliminary format, at the annual meeting of the American Educational Research Association, April 2001 and at PME NA, October 2001. Neither presentation focused on both the test data and the observation data. The work on this paper was supported by two grants from the National Science Foundation. The opinions presented here are those of the authors and are not necessarily shared by NSF, Rutgers University or Rider University.

mathematics, and classroom discourse does not tend to encourage students to defend and justify solutions.

Methods and Procedures: This paper focuses on observational and interview data from the first two years of a three-year multi-method study of testing and teaching in New Jersey that combines a statewide survey with a more intensive observation and interview study of a smaller sample of teachers.

Sample: The observation study focuses on 63 teachers. The sample was chosen to be representative in terms of both district wealth and geographic spread—i.e., north-south, east-west (Firestone et al., 2001). Almost all teachers taught fourth grade. That grade was chosen because it is the elementary grade tested in New Jersey.

Actual Observations: Fifty-eight teachers were observed for two math lessons and five teachers were observed once for a total of 121 classroom observations. The classroom researcher kept a running record of the events in the classroom, focusing on the activities of the teacher as well as capturing the activities of students, all problem activities and explorations, the materials used, the questions that were posed, the responses that were given—whether by students or teachers, the overall atmosphere of the classroom environment, and any other aspects of the class that they were able to gather. (For further information on sample selection, see Schorr and Firestone, 2001; Firestone, et.al. 2001)

Interviews: At the conclusion of each lesson, the teachers were asked to respond to a series of open-ended questions about the observed lesson. For example, they were asked: What were you trying to accomplish for today's lesson? What concept or ideas were you focusing on? What, if anything, would you change about today's lesson, and why? Why did you do this, or how did you feel about that (referring to a particular instance where for example, students explained mathematical ideas to each other or to the teacher, or with regard to a particular event or activity).

Teachers were also asked how state testing affected their teaching. Sample questions included: What kinds of things do you generally do to help your student get ready for the Elementary School Performance Assessment (ESPA)? Considering either the ESPA or the Content Standards, how, if at all, has that affected the topics you teach? How have you changed the teaching strategies you use in response to the ESPA and/or the Content Standards?

Coding: While observations were underway, researchers conducted detailed analyses of records of classroom observations, and adapted several pre-existing coding schemes to be used for coding the classroom data. These were based on the works of Stein, Smith, Henningsen, and Silver (2000); (Stigler & Hiebert, 1997, 1999); and Davis, Wagner, and Shafer, 1997). These codes were selected because they reflected ideas about effective mathematics instruction as indicated in national and state

standards. They included attention to the mathematical discourse that emerged, the opportunity for conceptual understanding to take place, the nature of student conjectures, the opportunities students had to share ideas and defend and justify solutions, etc. They were also chosen because we felt that they would supply information on the nature and use of reported strategies (i.e. manipulatives, small group instruction, use of different types of problems and activities, questioning strategies, etc.).

A preliminary coding scheme was tried out on approximately six observations before being agreed upon. A sheet of code definitions was created and a training session was held for coders involved in the activity. Ultimately, a coding instrument was developed which incorporated 18 dimensions, along with detailed descriptors of each coding category.

Two individuals independently coded each observation—at least one coder was an experienced mathematics education researcher. The other coder also had extensive experience in elementary education. After independent coding, raters sought to reconcile their differences and were successful in all but 2 of the 108 cases. In those two cases, another mathematics education researcher discussed differences with the raters and helped them to reach agreement.

Interview data were transcribed and entered into a qualitative data analysis software package. Interviews were sorted by question. Responses were analyzed in clusters, as there was considerable overlap in responses given to individual questions. Within each cluster, responses to specific questions on test preparation practices were reviewed and coded according to emergent themes. Responses were counted within each code. Interviews from 58 of the 63 teachers were available for analysis.

Results and Discussion: Some teachers reported that they liked the changes they had to make in response to ESPA. Nine made general comments to the effect that the test is forcing teachers "to evaluate their teaching style". Several made more specific comments that the presence of ESPA was encouraging them to use alternative teaching methodologies like manipulatives or have children respond to more open-ended questions. Many teachers reported that ESPA is encouraging them to implement more inquiry-oriented instructional practice. One teacher explained, "It's become my philosophy to teach them the concepts before, just, you know, ramming these rote facts down their throats."

In the interviews, teachers mentioned four general changes: having students explain their thought processes, using manipulatives, problem solving and working on students' writing. Forty three percent talked about trying to get students to explain their thinking in more detail. According to one teacher, the part "that I guess I really didn't do a lot of before is really get the students to start to learn how to

explain their thinking, to explain what they were doing. Sometimes they do it in writing; sometimes they do it to a partner; sometimes they do it to me." One strategy to encourage student explanation is the use of more open-ended questions on tests and in class, and was mentioned by 33%. Fourteen percent talked about using more "how" and "why" questions in their whole-group teaching. One described this as working on "critical thinking skills" instead of "feeding them the answer". Several also talked about using small-group instruction so students would explain their work to each other.

Another theme involved using manipulatives, and was mentioned by 45%. The ESPA has questions that involve at least written or pictorial descriptions of manipulatives. Many teachers felt that students who are more familiar with some of the current manipulatives could therefore better respond to those questions.

A third theme was a greater emphasis on problem solving (mentioned by 38%), though the actual meaning of "problem solving" was not always clear. For example, some teachers noted that they actually give students a set of strategies (i.e. draw a picture, think of a simpler problem, work backwards). Another teacher said, "We do a lot of work with problem-solving skills, just the basic skills of how you read a problem, how do you find the question, how do you find the information that you need, how do you check to see whether your solution is logical and can solve it a couple of different ways." This emphasis on word problems reflects in part the use of open-ended problems on the ESPA.

Finally, 40% said that they emphasize "writing" to prepare their students for the ESPA. One teacher said that she now had her students "write all the time for all subjects." Some teachers used "writing" to have their students explain their line of thinking in mathematics. In fact 12 teachers said they had students keep journals in math as well as other subjects.

The observations confirm that teachers are making some changes. Manipulatives were used in about 60% of all observed lessons. Similarly, students worked in groups for at least a portion of the time, in almost 65% of all observed lessons and in almost half of all cases, teachers made an effort to connect the lessons to the students' real life experiences.

The adoption of specific strategies was not necessarily accompanied by a change in overall approach to teaching mathematics, however. For example, while manipulatives were used extensively, they were used in a non-algorithmic manner in less than 19% of all observed lessons. This essentially means that the manipulatives were used in ways that did not foster the development of conceptual understanding. In fact, in almost two thirds of the lessons where manipulatives were used, they were used in a very procedural manner, where the teacher generally told the students exactly what to do with the materials, and the students did it as best they could.

Other times, teachers used manipulatives to demonstrate a particular procedure to the class. In many of the lessons—while many teachers had students physically touch concrete manipulatives, there often was little or no opportunity for the students to develop their own solutions to the problem or consider the relationship between the problem activity and the concrete (or alternative) representations.

Beyond looking at the use of specific practices and materials, we also examined the mathematical tasks students were asked to perform. We categorized tasks as memorization only, doing procedures where the focus was on producing correct answers rather than developing mathematical understanding, doing procedures to develop a deeper understanding of mathematical concepts or ideas, or doing a mathematical task that requires complex and non-algorithmic thinking (Stein and Smith, 1998). Only 3% of all observed lessons involved situations where students were required to do non-algorithmic thinking.

We also examined whether tasks involved practice or non-practice activities. With practice tasks, the teacher demonstrates or develops a procedure, such as long division, and then assigns a number of similar problems on which students are to repeat the same procedure (Stigler, 1999). Alternatively, in a non-practice task the student may be required to invent a new solution method, analyze a mathematical situation, or generate a proof. Practice tasks predominated, constituting almost 80 % of the observed lessons.

We also found that classroom discourse did not foster substantive conversations amongst students. Many teachers reported that they were interested in having students explain their reasoning. They also said they were interested in having students find and understand multiple strategies for solving problems, however, they rarely insisted on such activity. For instance, one code documented whether or not the teacher encouraged students to reflect on the reasonableness of their responses. In almost 80% of all cases the teacher rarely asked students whether their answers were reasonable. If a student gave an incorrect response, another student provided, or was asked to provide, a correct answer, but there was little discussion of an appropriate strategy to solve the problem. In an additional 15% of all cases, the teacher may have asked students if they checked whether their answers were reasonable, but did not promote discussion that emphasized conceptual understanding.

When students *were* provided with opportunities to talk about their answers or strategies, they usually simply stated answers to problems, and did not elaborate on their solutions. When a student was asked to share his solution, often he would respond with a numerical answer such as "5" or a procedure such as "you should add". Students were rarely asked to explain how they got their answer, or how they arrived at their particular strategy. In fact, students only explained their responses or

solution strategies in a way that went beyond the execution of procedures in six percent of the observed classes. Sometimes teachers would ask for an explanation for using a particular operation but would not encourage students to expand upon their answers, or move beyond simplistic responses.

Conclusions: The teachers involved in this research have indicated that they have been motivated to change their styles of teaching as a result of the ESPA test. Indeed, our observations confirm that they do incorporate many of the strategies and techniques that they reported in our interviews (such as small group instruction and the use of manipulatives). This research does not and cannot document just when these strategies first became part of their practice; we can only note that the teachers attribute the implementation of many of them to the test. This study provides evidence that the teaching practices that we noted in our observations, however, are not focused on the more conceptually oriented aspects of instruction. Perhaps with appropriate support, teachers who are ready and willing to make changes in their teaching will be able to incorporate practices that will enable children to have access to mathematical instruction that fosters the growth of mathematical thinking.

References:

Abrantes, Paulo. (2001). Revisiting the goals and the nature of mathematics for all in the context of a national curriculum. In M. vandenHeuvel-Panhuizen (Ed.). Proceedings of the 25th Conference of the International Group for the Psychology of Mathemtaics Education. Pp.25-40.

Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform, and rebellion.* Norwood, NJ: Ablex.

Cobb, P., Wood T., Yackel, E & McNeal, E.(1993) Mathematics as procedural instructions and mathematics as meaningful activity: The reality of teaching for understanding. In R.B. Davis & C.A. Maher (eds.) *Schools, Mathematics, and the World of Reality.* (pp.119-133).MA: Allyn and Bacon.

Davis, R.B. (1984*). Learning mathematics: The cognitive science approach to mathematics education.* Norwood New Jersey: Ablex Publishing Co.

Davis, J., Wagner, L. R., & Shafer, M. C. (1997). Classroom observation scale. Madison, WI: University of Wisconsin.

Editorial Projects in Education. (2001). *A better balance: Standards, tests, and the tools to succeed: Quality Counts 2001.* Bethesda, MD: Editorial Projects in Education.

Firestone, W. A., & Mayrowetz, D. (2000). Rethinking "high stakes:" Lessons from the United States and England and Wales. *Teachers College Record, 102*(4), 724-749.

Firestone, W. A., Monfils, L., Camilli, G., Schorr, R. Y., Hicks, J., & Mayrowetz, D. (2001, April 11, 2001). *The ambiguity of test preparation: A multimethod study of one state.* Paper presented at the Annual conference of the American Educational Research Association, Seattle,WA.

Keitel, C. & Kilpatrick. J. (1998). Rationality and irrationality of international comparative studies. In G. Kaiser, E. Luna & I. Huntley (eds.) International comparisons in mathematics edcuation. London: Falmer Press, 242-257.

Klein, R. & Tirosh, D. (2000). Does a research based teacher development program affect teachers' lesson plans? In T. Nakahara & M. Koyama (Eds.) *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education.* Hiroshima, Japan.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

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

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001*). Authentic intellectual work: Conflict or coexistence*. Chicago: Consortium on Chicago School Research.

Niss, M. (1996). Goals of mathematics teaching. In A. bishop et al. (eds), International Handbook of Mathematics Education. Dordrecht: Kluwer Academic Publishers, pp. 11-47.

O'Day, J., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing coherent educational policy* (pp. 250-312). San Franciscy: Jossey-Bass.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco: Jossey-Bass.

Schorr, R.Y. (2000). Impact at the student level. *Journal of Mathematical Behavior* 19 209-231.

Schorr R.Y. & Firestone, W.A. (2001). Changing mathematics teaching in response to a state testing program: A fine-grained analysis. Paper presented at the Annual meeting of the American Educational Research Association, Seattle, WA.

Smith, M. L. (1996). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program*. Tempe, AZ: Southwest Educational Policy Studies, Arizona State University.

Stecher, B. M., & Barron, S. I. (1999). *Quadrennial milepost accountability testing in Kentucky*. Los Angeles: CRESST.

Stein, K.S., Smith, M.S., Henningsen, M.A. & Silver, E.A. (2000). *Implementing standards-base mathematics instruction: A casebook for professional development*. New York: Teachers College Press

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction. *Phi Delta Kappan, 79*(1), 14-21.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

Wilson, S. M., & Floden, R. E. (2001). Hedging bets: Standards-based reform classrooms. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 193-216). Chicago: University of Chicago Press.