Canadian Journal of Educational Administration and Policy, Issue #96, November 9, 2009.

"THE FIRST YEAR, THEY CRIED":

HOW TEACHERS ADDRESS TEST STRESS*

Ruth A. Childs and Lina Fung, *OISE, University of Toronto*

Grade 3 teachers in Ontario must administer the Primary Assessment of Reading, Writing and Mathematics to their students. In this study, we analyze how four teachers framed and addressed the problem of the stress some of their students experienced when sitting the test. The teachers framed the problem in different ways: as related to students' perceptions of the test, students' preparation for the test, teachers' attitudes about the test, or the developmental appropriateness of the test administration instructions. Their approaches to addressing the problem were related to how they framed the problem and whether they believed the test administration instructions must be strictly followed. We believe it is important for educational policy makers and test developers to understand the problem of test stress and to engage teachers in finding ways to address it.

For three days each May, Grade 3 teachers in Ontario administer the Primary Assessment of Reading, Writing and Mathematics to their students. In this article, we use Schön's (1983) metaphor of frames or frameworks to investigate the genesis of the test and how four teachers understand and address the problem of test stress for their students. We contrast policy makers' focus on the test as a solution to the problem of low public confidence in publicly-funded education with the teachers' focus on the problem of stress caused by the test.

* Based in part on a paper presented at the annual meeting of the International Test Commission, Liverpool, UK, July 2008. We gratefully acknowledge the teachers who participated in this research and the Social Sciences and Humanities Research Council of Canada's support of this research. Thank you to Madison Aitken, Susan Elgie, Monique Herbert, Vesna Milenkovic, Anita Ram, Linda Umezawa, and Yunmei Xu for their contributions to the data collection and analysis.

Frame Analysis

In *The Reflective Practitioner*, Donald Schön (1983) describes how professionals define problems. In particular, the *frame* on which their understanding of a problem rests determines "the decision to be made, the ends to be achieved, the means which may be chosen" (p. 40). The frame affects the professional's decisions. As Schön puts it, "their frames determine their strategies of attention and thereby set the directions in which they will try to change the situation, the values which will shape their practice" (p. 309).

In his 1994 book with Rein, *Frame Reflection*, Schön explores the effects of frames. If frames are the "underlying structures of belief, perception, and appreciation" (p. 23), then when individuals or groups frame the same problem differently, agreement on acceptable actions will be particularly difficult to reach. Comparing the frames that individuals or groups use for a problem can provide invaluable insights into the reasons for different decisions. Indeed, frame analysis and its variants have been applied to problems in a wide range of disciplines (Benford & Snow, 2000). However, that frames are rarely explicit makes the comparison of frames among individuals or groups especially difficult. Often, frames must be inferred from what individuals or groups say about their beliefs and perceptions or, when that fails, inferred from their decisions.

A Test is Born

In 1993, Ontario's Lieutenant Governor established a Royal Commission on Learning to answer, among other questions, How should students be evaluated?, How should student progress be reported?, How should schools and programs be evaluated?, and To whom should progress be reported and for what purpose? Nineteen months later, the Commission issued its

report, titled "For the Love of Learning." In that report, they recommended Ontario develop large-scale assessments, citing "the public's need for some measure of basic student achievement that is applied in the same way to every student at a few points in time" (Bégin et al., 1994, vol. 2, p. 150). A Crown agency to develop such tests was created in 1996 by the Ontario Legislature's passage of the Education Quality and Accountability Office (EQAO) Act. The new agency's first test was the Grade 3 Assessment of Reading, Writing and Mathematics (now called the Primary Assessment of Reading, Writing and Mathematics). EQAO describes the goals of that test as "to measure student achievement against curriculum expectations ... to direct improvements in education at individual, school, board and provincial levels" (EQAO, 2007, p. 11).

The Grade 3/Primary Assessment contains both multiple-choice and open-response items and is administered with paper and pencil. It has been shortened several times, from two weeks of testing in 1997 to its current length of six hours spread over three days. The percentage of students in each school who achieve the provincial standard in each subject area is published on the EQAO's website and in many local newspapers. The lowest-scoring schools and schools with declining results receive extra funding and professional development support through the Ontario Focused Intervention Partnership (OFIP). The test is not intended to have direct effects for either students or teachers.

EQAO creates the testing materials and the test administration instructions that accompany those materials. School administrators and teachers are required to administer the test. The test administration instructions for 2007 included the following statements:

- Student assessment material packages must not be opened prior to the administration of the first section of the assessment.
- Teachers administering the assessment may receive the student assessment materials no earlier than on the morning of the day the administration begins.

- Once the assessment has been opened, no one may use information from the assessment to provide instruction on any concept or item being tested prior to or during the administration of the assessment.
- During the assessment, no one may explain, define or provide examples of reading vocabulary or writing or mathematics terminology to students, including those with accommodations.
- During the assessment, nothing can be said or done to influence student responses, including, but not limited to, actions such as drawing a student's attention to an unanswered question.
- At no point during or after the assessment may anything be said or done to encourage students to alter or revise their responses.
- No one may read, review or correct student work during or after the administration. This includes darkening, rewriting, editing, erasing or altering student work in any way. (Education Quality and Accountability Office, 2007, pp. 3-4)

What problem was the test intended to solve and how was that problem framed by educational policy makers? The Royal Commission on Learning focused on "the public's need for some measure ... that is applied in the same way to every student" (Bégin et al., 1994, vol. 2, p. 150). Ontario's Minister of Education, in announcing the subsequent creation of EQAO, declared that the agency's tests "will give us valuable, accurate and credible information on how students are performing in schools" (Ministry of Education News Release, December 14, 1995). These statements assumed low public confidence in Ontario's education system (indeed, the Survey of Educational Issues in Ontario in 1990 found that only 47% of adults were satisfied with Ontario's schools; Hart & Livingstone, 2007). The problem of low public confidence was framed by the Ministry of Education as a need for independent measurement of students' achievement.

Recall that, according to Schön, the way a problem is framed determines "the decision to be made, the ends to be achieved, the means which may be chosen" (1983, p. 40). The framing of the problem of low public confidence as the need for independent measurement requires an independent entity to create or select an instrument that can be administered consistently to

students across the province. While not the only possible choice, a written test administered by teachers following strict guidelines is a relatively cost-effective way of obtaining the desired data; it is also an approach that is used in many other countries.

The Test in the Classroom

In the policy makers' view, the teachers' role is simply to administer the test according to the instructions. However, not all teachers follow the instructions. A review of the Ontario College of Teachers' (OCT) Discipline Committee hearings (published by OCT both in its magazine, *Professionally Speaking*, and on its website) shows that several teachers have had their teaching certificates suspended because of violations of test administration instructions (although most have been related to the Ontario Secondary School Literacy Test, not the tests for younger students). Indeed, several studies in the United States have found that teachers vary widely in whether, when preparing their students for and administering large-scale tests, they follow the instructions they are given (Abrams, Pedulla, & Madaus, 2003; Monsaas & Engelhard, 1994; Smith, 2001; Urdan & Paris, 1994; Wodtke, Harper, Schommer, & Brunelli, 1989). In this article, we suggest a reason that teachers might choose not to follow the instructions: The teachers may not share the policy makers' focus on the test as a response to the problem of low public confidence in education; instead, some teachers focus on the immediate effects of the testing in their classrooms, especially that some students experience the testing as stressful. Depending on how they frame the problem of test stress, teachers may choose to administer the test in ways that are inconsistent with the test instructions.

Method

For examples of how teachers address test stress, we draw on four Grade 3 teachers' descriptions of how they prepare their students for and administer a large-scale assessment. These teachers were interviewed as part of a larger study of Ontario Grade 3 teachers' beliefs about test preparation and administration, in which a total of 40 teachers were interviewed (Childs & Umezawa, 2009). About 80% of the teachers were in the greater-Toronto area, with the rest distributed across the province; all but four were females. In the interviews, most of which were conducted by phone, we asked teachers what they would do in testing situations, such as if they noticed that a student was answering in only a word or two questions that asked students to explain their reasoning. Although these vignettes were presented as hypothetical, and teachers were asked what they *would do*, not what they *have done*, many of these teachers told us about their past experiences with the test.

The interviews were conducted by the first author and a graduate student who was also a member of the research team. The audiotapes and notes from the interviews were independently reviewed by research team members who were not present at the interviews and who subsequently typed summaries of the responses and transcribed relevant quotations.

The four teachers who are the focus of this article were chosen for closer analysis because, in answering general questions about the effects of the test, they identified test stress as a problem and provided detailed descriptions of how they addressed it. Because the original study did not ask teachers specifically whether test stress was a problem, we do not know what percentage of the 40 teachers believed test stress was a problem in their classrooms.

After selecting the four teachers, we created more complete transcripts of their interviews and reviewed those transcripts for evidence of how they framed and addressed the problem of

test stress, based on Schön's (1983) descriptions of framing. We have chosen to use each teacher's own words in order to convey not only how the teachers framed and addressed the problem, but also their frustration with the test administration instructions and their excitement about the ways they had developed to minimize their students' stress. Each teacher is identified using an initial; these initials were chosen arbitrarily and have no meaning.

Results

Ms. L: "They're eight years old!"

Ms. L has taught for eight years, seven of them with a Grade 3 class or grade-2/3 combined class. She vividly remembers the first time she administered the test and identifies that some of her students experienced the test as stressful as a problem. She is strongly motivated to avoid repeating the experience of that year:

> I really felt like I failed the kids that first year, because I didn't know what to do to get them ready. After my first experience, I really gave it some thought, and thought about what they struggled with and what I could do to make it better... One poor little boy got stomach problems so bad because he was so upset, I had to send him home. I thought we really need to do something about this to make it better, so I came up with what I do, whether it's right or wrong.

As the last sentence reveals, she recognizes that some activities may be inconsistent with the test administration instructions. However, as she went on to tell us, she chooses not to follow the instructions that she believes are not developmentally appropriate. For example, the rules stipulate, "During the assessment, nothing can be said or done to influence student responses" (Education Quality and Accountability Office, 2007, pp. 3-4). As Ms. L explains:

> These people are little human beings that we come to know very well by May, and they would not respond favorably to me following a contrived script and just [saying], "No talking at all and you can't speak to me at all." I realize there are no questions [allowed], but they are little kids, they're eight years old!

Ms. L has also developed elaborate traditions intended to make the testing days "a celebration":

> We wear pajamas and thinking caps (even I do – I wear my pajamas). We wear whatever caps, ball caps – I've had a sombrero, and a bridal veil, whatever – whatever thinking cap makes them think better. I buy special pencils. We always have a treat, and the kids decide what we have, but it has to be healthy during the test because we want our brains to work well. So we'll have bagels and cream cheese or a muffin and fruit, and the principal pays for this. I tell him he has no choice .... The fourth day is our "Grade 3 EQAO fun day," where we have a sub [sandwich] lunch, they plan all the activities for the day and we really celebrate. They feel so proud of themselves, and they ... get an "I survived EQAO testing" sticker, that kind of thing, and we really make it into a celebration, and they look forward to it. And this group that I have this year, on the first day of school ... I ask them, What is the one question they want to ask about grade 3, and several of them asked me things about EQAO: "Do we really get to wear our pajamas every day?".... We do all these kitschy things to make them enthusiastic and positive.

Ms. L frames the problem of test stress as one of developmental inappropriateness. She believes that the test administration instructions are inappropriate for Grade 3 students and so create undue stress for her students. Based on this framing, she feels justified in deviating from the instructions by, for example, answering the students' questions. She also schedules activities, such as costume days and celebrations, which she knows will be exciting for most students of this age, to coincide with the testing. Such activities are not addressed in the instructions.

Ms. T: "Yippee!"

Ms. T, with 18 years of teaching experience, six of them with third grade, is also strongly motivated to avoid repeating past negative experiences. From the start of the school year, she trains her students to cheer whenever they hear the acronym for the provincial testing agency:

> This is a weird thing I do: Whenever I say the word "EQAO", the kids have to go "Yippee!" The first time I did it, the EQAO, the kids were crying and they were lying on their desks. That's so many years ago and I thought, "Okay, well what can I do? Maybe if I get them to think positively." So, they go "Yippee!" and put their hands up in the air.

As she reiterated later in the interview, that the students did not cry during the most recent test administration is proof that her approach is working:

> The first year they had EQAO, I had a bad response to it. The kids, they were crying, and they were lying on the tables, and to me, I remember that all the time. But when I did it last year, it was great, it was just a new positive attitude.

While Ms. T acknowledges that her approach to promoting a positive attitude toward testing is unusual, she believes that this kind of test preparation is consistent with the instructions. For her, following the instructions for administering the test is a given:

> I've heard stories of how some schools do it differently, and it's something that I wouldn't do. I don't think it's ethical, and when there are guidelines, you should follow them.

This does not mean that she wholeheartedly supports the testing:

> I have mixed feelings about it.... I think it's how we interpret the results. If we use it to say our school is better than yours, then I don't want to do it. If we use it so that we can help the teachers program better for the kids, then that is more useful as a tool.

Although Ms. T has, like Ms. L, had very negative experiences with the testing in the past and identifies the stress some students experience from the testing as a problem, she frames it differently: as caused by students' negative perceptions of the test. She therefore develops activities to help the students develop positive associations with the testing. Because such activities are not addressed in the instructions, she does not worry about violating the instructions, which she believes must be strictly followed.

Ms. P: "I'm not telling answers"

Ms. P has taught for 15 years, six of them with third grade. She, too, sees test stress as a problem. She believes it is important to follow the test instructions; however, she believes that anything that is not explicitly forbidden is acceptable. As she puts it:

> I have found some creative ways of getting around the instructions. I'm not telling [the students] answers ... the test scores are not being affected: That is where I draw my moral line.

She goes on to describe some of these "creative ways":

> If you're sitting with a kid and the kid is visibly upset and distressed, you're not going to ignore the kid and say, "I'm sorry it's a test and I can't talk to you." It's cruel. Just sit down with the kid and have a pep talk with them, and let them know they are working very hard, and, "Let's go up and get a drink and have a snack. And let's go back to it later."

Ms. P believes that teaching her students test-taking skills and strategies will help them to be less anxious during the test and she works to make the preparation itself fun:

> [T]he kids love highlighting things. So you say "we're going to be word detectives now, let's look for the mathematic words that teach us what to do in the question" And they start to highlight madly away and you have to try to narrow their focus: "What words in the question tell you what to do?"

She has developed ways to subtly remind the students of these strategies during the testing, while also providing general encouragement. For example, she may make an announcement like the following during the test:

> "Oh, wasn't everybody working really hard? Let's make sure we put down all the ideas that we have in our heads, because, remember, I am not able to mark the test and the people that mark the test have never met you. So, you need to tell them how intelligent you are. I know that you are terribly intelligent, everybody here knows you are terribly intelligent, but [the markers have] never met you. So, you make sure you put all your ideas down!"

Like the other teachers, Ms. P is motivated by wanting to minimize the students' stress, based on her previous experiences:

> I certainly have had kids cry on me because you couldn't help them, which is ridiculous, so you work through that.... I mean, you have to explain to a new teacher, "This is the way it is and this is what we are going to do to make it as easy as possible on the kids."

Ms. P frames the problem in two ways: as students not being sufficiently prepared for the test and as the instructions restricting normal teacher-student interactions. She works to

familiarize her students with the test ahead of time and she carefully analyzes the instructions to figure out how she can minimize her students' anxiety without violating the letter of the instructions. Ms. P prides herself in finding creative ways to interpret the test administration instructions that permit her to continue to have supportive interactions with her students.

Ms. S: "Wonderful, you're great!"

Ms. S has taught for 20 years, eight of them with third grade. Like the other teachers, she sees test stress as a problem. However, she frames the problem of test stress as largely one of the students taking cues from their teachers. She is focused on making the testing a positive experience for the students:

> I want to prepare my student so they can feel comfortable during those three days.... I talk it down, rather than talk it up. After we do the first piece of the test, I always will do something fun, we'll have a little game or go out for an early recess, and I'll say to them, "Was that hard?," and they'll always say, "No," and I'll say, "Isn't it just like the things we've been doing? What about that question? Do you remember doing another one like that?" It makes them just think, "Okay, I can do this" … Even those kids that weren't doing very well thought they were, because of the things I was saying to them: "Wonderful, you're great!"

As she sees it, the Grade 3 students do not yet have enough experience with testing to know that they should be anxious:

> I don't think, generally speaking, eight-year-olds are inherently anxious. I mean, what I see in my students is they're excited about it. It's new, they run their hands over the booklets, it's different.

Consequently, in preparing the students for the test, she focuses on modeling a positive attitude and relevant skills:

> I don't ever call it a test. Usually I use the word "assessment" or "EQAO activities" and make it sound like it's language we use in the classroom on a day to day basis.... If you treat it like it's not something to get worked up about, then they probably will follow suit.

Like some of the other teachers, she is motivated at least in part by past negative experiences. In fact, Ms. S teacher administered the test in its first year when it took two weeks, instead of three days:

> When it was two weeks, it was ridiculous. They were crying, I was crying. I have never in my life felt as cruel as I did in those early years.

She is explicit in describing the dilemma that the testing presents for her:

> I think, too, we have to remember these are children, we are not assessing adults … even to walk around and say to a child, "Look at this word, it says 'explain,' and you haven't really explained." In the course of my classroom teaching, I see nothing wrong with that. I see they are there, they are learners, it is an on-going process. I'm not telling them what to write, I don't see anything wrong with that. Having said that, I won't do it because they tell you not do.

In other words, Ms. S sees the test administration instructions as non-negotiable – that is, the decisions about how to administer the test are not hers to make. However, she believes she can influence the students' attitudes about and experiences of testing through how she talks about the testing and by modeling a positive attitude.

Discussion and Conclusion

Although the four teachers in this study shared a concern about test stress, they differed in how they framed the problem: as caused by developmentally inappropriate instructions, students' negative perceptions of the test, students' lack of preparation for the test, or teachers conveying to the students that the test should be a source of stress. They also varied in whether they believed they had to follow the instructions strictly, could interpret them creatively, or could ignore them when they believed that was appropriate. How each teacher could address the problem of test stress was constrained by how she framed the problem and whether she believed the instructions must be strictly followed. Evident in their responses were the teachers'

frustration with the test administration instructions, the thought they put into developing ways to minimize their students' stress, and their pride when their efforts succeeded.

This study has several limitations. These teachers were chosen, not because they were representative of all Grade 3 teachers, but because they shared a focus on test stress and provided detailed descriptions of how they framed that problem and what they did to minimize the students' stress. The four teachers had seen students in past years cry or otherwise express distress during the testing; because of these experiences, they were highly motivated to minimize the stress of the testing for future students. Teachers without these experiences might not view test stress as a problem and so might be less motivated to minimize test stress. It is also likely that teachers at other grade levels would have different views of large-scale assessments.

What do these results mean for policy makers and for the agencies, such as EQAO, that develop tests? Both should bear in mind that the teachers who are being asked to administer tests are much less concerned about tests as solutions to problems of public confidence than they are about the immediate effect of the testing on their students. If teachers believe, in particular, that the test instructions are developmentally inappropriate, they may feel justified in not following the test administration instructions. Even those teachers who believe they must follow the instructions may vary in their activities surrounding the testing. We urge educational policy makers and testing agencies to acknowledge that test stress is an important concern for some teachers, especially those in early elementary grades, and to engage teachers in finding ways to address it, including ensuring that the test administration instructions are developmentally appropriate. To ignore teachers' concerns is to risk overestimating the comparability of test administration across classrooms and, in the end, the potential usefulness of such tests as independent measures of student achievement.

References

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42,* 18-29.

Bégin, M., Caplan, G., Bharti, M., Glaze, A., Murphy, D., & DiCecco, R. (1994). *For the love of learning: Report of the Royal Commission on Learning*. Toronto: Queen's Printer.

Benford, R. D., & Snow, D. A. (2000). Frame processes and social movements: An overview and assessment. *Annual Review of Sociology, 26,* 611-639.

Childs, R. A., & Umezawa, L. (2009). When the teacher is the test proctor. *Canadian Journal of Education, 32,* 618-651.

Education Quality and Accountability Office (EQAO). (2007, Spring). *Administering the Assessments of Reading, Writing and Mathematics. Primary Division (Grades 1-3) and Junior Division (Grades 4-6).* Toronto: Author.

Hart, D., & Livingstone, D. W. (2007). *Public attitudes towards education in Ontario 2007: The 16th OISE Survey.* Toronto: Ontario Institute for Studies in Education, University of Toronto.

Monsaas, J., & Engelhard, G., Jr. (1994). Teachers' attitudes toward testing practices. *Journal of Psychology, 128,* 469-477.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Schön, D. A., & Rein, M. (1994). *Frame reflection: Toward the resolution of intractable policy controversies.* New York: Basic Books.

Smith, J. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28,* 521-542.

Urdan, T. C., & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy, 8,* 137-156.

Wodtke, K. H., Harper, F., Schommer, M., & Brunelli, P. (1989). How standardized is school testing? An exploratory observational study of standardized group testing in kindergarten. *Educational Evaluation and Policy Analysis, 11,* 223-235.