

Would You Explain the Instructions? Ontario Teachers' Responses to Common Testing Dilemmas

Ruth A. Childs, Linda Umezawa, Monique Herbert, Yunmei Xu, and Saad Chahine

Ontario Institute for Studies in Education of the University of Toronto

Paper presented at the June 2007 workshop
on "Large-scale assessments of achievement and the issue of official languages"
at the University of Ottawa Research Unit on Measurement, Evaluation and Assessment

Abstract

Elementary teachers in French- and English-language schools across Ontario were surveyed using an on-line questionnaire about how they would respond to common dilemmas in administering large-scale assessments and why. Results show that teachers' beliefs about what they should do when administering the tests and what they would do vary widely. The relationship between beliefs and actions is complex and the motivation for deviating from the instructions in test administration is often not consistent with that typically associated with "cheating." The implications of different teacher practices for the interpretation of test results are explored.

Introduction

Teachers play critically important roles in large-scale assessments by preparing students to write the assessments, administering the assessments, and interpreting the results. However, these roles – and whether they complement or conflict with teachers' primary roles – are rarely acknowledged when decisions about large-scale assessment programs are made.

In many jurisdictions, elementary teachers are responsible not only for preparing their students to take large-scale assessments, but also for administering the assessment to their own students. The literature on test administration has focused on the importance of test security and strict adherence to the administration instructions and has characterized any deviations from the instructions as cheating (e.g., Cizek, 2003). Largely absent from the literature (see, however, Anders & Richardson, 1992; Smith, 1991) is a recognition that elementary teachers who are responsible for administering large-scale assessments to their own students may experience conflicts between their professional identities as teachers and their temporary roles as test administrators.

The purpose of this study is to investigate elementary teachers' beliefs about their professional roles and about the assessments, their responses to common testing dilemmas and the relationship between their beliefs and their responses. This paper will focus particularly on teachers' beliefs about what their roles in administering a large-scale assessment should involve – whether they should strictly adhere to the policies, procedures and instructions, as instructed in the administration guide, or use their professional judgement in deciding how to administer the test – and whether these beliefs are related to what they say they would do when faced with test administration dilemmas.

Test Preparation and Administration Practices

Over the past two decades, much has been written about the appropriateness of test preparation practices (Cohen & Hyman, 1991; Haladyna, Nolen & Hass, 1991; Mehrens, 1991; Mehrens & Kaminski, 1989; Mehrens, Popham & Ryan, 1998; Popham, 1991; Smith, 1991). Mehrens (1991), in particular, suggests that "the most general...principle is that a teacher should not engage in any type of instruction that attenuates the ability to infer from the test score to the domain of knowledge/skill/...ability of interest" (p. 4). While agreement about which practices are appropriate is not unanimous (indeed, Cohen & Hyman, 1991, disagreed with Mehrens), the

majority of measurement theorists seem to endorse practices that they believe will support test validity and oppose practices that might undermine it.

Less has been written about the appropriateness of test administration practices. Indeed, most of those who have addressed test administration have started from an assumption that any deviation from the test administration instructions is not only inappropriate, but unethical (see, however, Wiggins, 1994). For example, Cizek (2003) writes, “When we expect that testing irregularities may have occurred as a result of human intervention... then our sense of ethical behaviour and fairness is violated as are, in many cases, legal or administrative guidelines” (p.368). Cizek assumes that teachers who do not follow the test instructions are motivated by the desire to artificially inflate students’ scores – hence, his labelling of such behaviour as cheating.

A few researchers have considered alternative motivations for deviating from test administration instructions. For example, Wellhousen and Martin (1995) surveyed 63 preservice teachers about whether they would deviate from test administration instructions. While 20% said they would if they thought “the test or test items were inappropriate,” 36% would “to benefit the children.” Giving clues or hints during the test was considered acceptable by 24% of the preservice teachers and 6% would reword test directions or test questions.

Others have questioned whether the emphasis on strict adherence to test administration instructions is reasonable. Wiggins (1994), in particular, argues that the emphasis in standardizing test administration in order to ensure “that each student have equal opportunity to answer a question correctly...hardly justifies the regular practice of forbidding almost all human interaction and the use of contextually appropriate resources, particularly if our aim is to make tests educative and more authentic” (p. 173). He further asserts that, in forbidding students access to resources during a test, “the technical constraints requiring the standardizing of test conditions invariably seem to inappropriately outweigh the pedagogical ones” (p. 173).

Method

Data for this study were collected in two phases. In Phase 1, instructors in a large teacher education program were asked to nominate Grade 3 teachers and elementary principals to take part in structured interviews about test preparation for Ontario’s Grade 3 Assessment of Reading, Writing, and Mathematics. All teachers and principals who were nominated were contacted and agreed to participate. Most of the participants – 8 Grade 3 teachers, and 3 elementary principals or vice principals – were interviewed in the spring of 2005; an additional teacher who had formerly taught Grade 3 was interviewed in the winter of 2006. All participants worked in public or separate (i.e., Catholic) school systems in the greater Toronto area.

In Phase 2, responses from Phase 1 were used to create brief vignettes describing dilemmas Grade 3 teachers may face in preparing their students for and administering the assessment. Ten open-response vignettes (e.g., “You work in a school where a question on the test asks students to estimate the volume of a box. You were planning to teach about volume after the test. There’s a week before you will be administering the test. What would you do?” “During the test, you notice that one of your strongest students is spending too much time on the first few questions. You are worried that she may run out of time later in the test. What would you do?”) and 20 Likert-type items on teacher’s beliefs about their professional roles and about the assessment were posted in both English and French in an on-line survey. Elementary teachers were recruited principally through announcements in their professional association newsletter.

By June 2007, 50 complete responses (17 French and 33 English) had been received and were the focus of this paper.

The interview transcripts and notes from Phase 1 were reviewed for themes and coded. A recursive approach was taken, so that once themes were identified in one interview, other interviews were re-examined for possible evidence of the same themes. The 50 written responses received so far in Phase 2 have been classified according to the actions described and the motivations given. This paper will focus on the Phase 2 data.

Results

Beliefs

Using information gathered from the Phase 2 teacher survey, a comparison of the beliefs of teachers in English- and French-language schools about what their roles in administering the test should involve – whether they believe teachers should strictly adhere to the policies, procedures, and instructions, as instructed in the administration guide, or use their professional judgement in deciding how to administer the test – is presented in Table 1.

Table 1

Comparison of English and French Teachers' Beliefs Concerning Test Administration

All teachers should administer the test in “strict adherence to the policies, procedures and instructions,” as instructed in the administration guide./ Tout le personnel enseignant devraient administrer le test en suivant strictement les politiques, les procédures et les directives telles que décrites dan le guide d’administration.	Teachers should use their professional judgement in deciding how to administer the test./ Le personnel enseignant devrait utiliser leur jugement professionnel lors de la prise de decision concernant l’administration du test.	Teachers in English-language Schools (N=33)	Teachers in French-language Schools (N=17)
Agree/ En accord	Disagree/En accord	58%	18%
Agree/En accord	Agree/En accord	9%	47%
Disagree/En accord	Agree/En accord	21%	12%
	Unsure or Missing	12%	24%

While the sample so far is small, the patterns suggest interesting differences. Almost half (47%) of the teachers in French-language schools favoured a combination of both strictly adhering to the test administration procedures and using their professional judgement during test administration. Only 18% favoured strict adherence to the test administration procedures,

without using professional judgment; 12% endorsed using professional judgement, but not strictly adhering to the procedures.

The picture differs somewhat for the English teachers. The majority (58%) of teachers in English-language schools indicated that they should strictly adhere to the test administration procedures, while the second largest group (21%) agreed that using one's professional judgement was the most appropriate way to approach the test administration. In contrast to the teachers in the French-language schools, only 9% endorsed both strict adherence and professional judgement.

Responses to Administration Dilemmas

Based on the analyses of the Phase 1 data and preliminary analyses of the Phase 2 data, the teachers in this study report a wide variety of test administration actions, with more than two-thirds reporting knowingly deviating from the instructions. For example, in response to the vignette describing a teacher seeing that the assessment includes a question on content they have not yet taught, many teachers reported that they feel it is unfair to test students on material they have not yet been taught ("we would never do that to an adult") and so would teach a quick unit on the general content. All the teachers agreed that teaching the specific item would be wrong, but many believed that failing to teach the content would result in an underestimate of their students' knowledge.

While there was some relationship between teachers' beliefs about the need for strict adherence or professional judgement and their responses to the dilemmas, the beliefs left much of the variability unexplained. Of those teachers (whether in the French- or English-language schools) who believed in the need for professional judgement, but not strict adherence, 80% said they would teach volume before the test. Teachers who endorsed strict adherence but not professional judgement were less likely to teach about volume – only about half would do so.

Regarding the strong student who was taking too much time on the first few items, most teachers said they would remind the student about the time or that they would make a general announcement to the class about the importance of not spending too much time on any one question. Of the teachers who endorsed professional judgement, but not strict adherence, more than two-thirds would point to or mention the instructions to the student; of the teachers who believed in strict adherence without professional judgement, only one-third would do so. Similar percentages – about 20% – of both groups would make an announcement reminding the class of the need to read the instructions carefully and to try to fill the space provided for an answer.

Many of the teachers reasoned that the test would underestimate that student's knowledge if they did not intervene. In general, these teachers emphasized the importance of the test accurately measuring the students' knowledge and believed that they had a duty as teachers to modify the test administration conditions where those conditions might prevent students from fully demonstrating their knowledge and skills.

In a third vignette, we asked the teachers what they would do if, in collecting the multiple-choice answer sheets, they noticed that one student had coloured two bubbles on several questions. About 50% of those favouring professional judgement and 40% of those endorsing strict adherence said they would return the sheet to the student with a reminder that only one

answer may be selected. The reasoning of most of these teachers was summarized well by a teacher who wrote, “It is not a reflection of what they know if they mess up the directions. Is the EQAO a test of directions or knowledge?”

In all three vignettes, teachers who said they would do nothing (and therefore not deviate from the administration procedures) emphasized either the professional risk of not following the instructions or the importance of having test results that were obtained under identical conditions across classrooms so that comparisons would be fair.

Discussion

As mentioned earlier, the results from the online survey of Phase 2 indicate that a large majority of teachers reported that they knowingly deviated from the instructions, by reporting modifications of the test administration guidelines such as teaching material that they knew would be on the test and alerting students to elapsed time during the test. Whether teachers believed that strict adherence to the test administration procedures was important or believed that teachers should use their professional judgement made some difference in their probability of following the procedures, but many who supported strict adherence nevertheless reported they would not follow the procedures.

In understanding the teachers’ choices, we must look beyond the strict adherence/professional judgement dichotomy to other explanations offered by the teachers. In explaining their choices, many teachers reasoned that the test would underestimate that students’ knowledge if they did not intervene. These teachers believed that they had a duty as teachers to modify the test administration conditions where those conditions might prevent students from fully demonstrating their knowledge and skills. That is, unlike teachers who are motivated to change the test administration in order to artificially inflate their students’ results and so make the results less accurate (the motivation typically associated with cheating), these teachers are motivated by the desire to make the results more accurate (and thus, in a way, support the validity of the results). They see this as an important distinction and do not view themselves as cheaters.

Some of the teachers who did not deviate from the test administration instructions in their responses to the vignettes emphasized the importance of the test as a tool for comparing schools and believed that maintaining the standardization of the administration (and so the reliability of the results) was paramount, even if they, as teachers found the test administration instructions frustrating. It is difficult to interpret the differences in responses to the belief questions by the teachers in French- and English-language schools. When a larger sample of responses is available, we may be able to confirm and better interpret the different perspectives.

Conclusion

The results from Phase 1 and 2 of this investigation into teacher beliefs and practices related to preparation for and administration of large scale assessments make it clear that further research is needed. Phase 3 of the study, in which we are interviewing 50 teachers from across Ontario, is underway and will be completed this fall. In addition, information will continue to be gathered from the on-line survey. With this, we are working to refine our understanding of the factors that influence teachers’ test preparation and administration practices.

Although many of the teachers in this study believed that they were increasing the validity of the test interpretation by modifying the test administration, the variation in the teachers' practices may pose a threat to the test's reliability and, in particular, compromise its usefulness for comparison across schools and across years. Based on our research, we believe that efforts to prevent nonstandard administration practices by threatening to prosecute teachers for cheating are likely to be ineffective because most teachers do not see themselves as cheaters. If comparisons of results are an important purpose of the test, then teachers' reasons for not following the test administration procedures need to be understood and addressed.

References

- Anders, P., & Richardson, V. (1992). Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions, and consequences of accountability. *Teachers College Record, 94*, 382-396.
- Cizek, G.J. (1999). *Cheating on tests: How to do it, detect it and prevent it*. Mahwah, NJ: Erlbaum.
- Cohen, S.A., & Hyman, J.S. (1991). Can fantasies become facts? *Educational Measurement: Issues and Practice, 10*(1), 20-23.
- Haladyna, T. M., Nolen, S. B., & Hass, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*, 2-7.
- Mehrens, W.A. (1991, April). *Defensible/undefensible instructional preparation for high stakes achievement tests: An exploratory dialogue*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Mehrens, W.A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent. *Educational Measurement: Issues and Practice, 8*(1), 14-22.
- Mehrens, W.A., Popham, W.J., & Ryan, J.M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice, 17*(1), 18-22.
- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12-15.
- Smith, J. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28*, 521-542.
- Wellhousen, K., & Martin, N.K. (1995). Preservice teachers and standardized test administration: Their behavioural predictions regarding cheating. *Research in the Schools, 2*(2), 47-49.