The Rights and Responsibility of Test Takers when Large-Scale Testing Is Used for Classroom Assessment

Christina van Barneveld
Lakehead University

Karieann Brinson
Lakehead University

Abstract

The purpose of this research was to identify conflicts in the rights and responsibility of Grade 9 test takers when some parts of a large-scale test are marked by teachers and used in the calculation of students' class marks. Data from teachers' questionnaires and students' questionnaires from a 2009–10 administration of a large-scale test of Grade 9 mathematics were analyzed using descriptive statistics. Written comments by teachers were analyzed into themes. Results were interpreted using a framework comprised of a policy document and two theories that are relevant to large-scale testing: the rights and responsibility of test takers as documented in the Standards for Educational and Psychological Testing (AERA/APA/NCME/JCSEPT, 2014), expectancy-value theory of motivation, and theory regarding the developmental stage of students in Grade 9. Several conflicts were identified. Potential solutions to conflicts were presented.

Canadian Journal of Education / Revue canadienne de l'éducation 40:1 (2017)

www.cje-rce.ca

Keywords: large-scale testing, motivation, adolescent development, standards, classroom assessment

Résumé

Cette recherche avait pour objectif d'identifier les conflits dans les droits et la responsabilité des élèves de 9e année qui sont soumis à des tests à grande échelle lorsque certaines sections de ces examens sont corrigées par des enseignants et qu'ils servent dans le calcul des notes des élèves. Des données tirées des questionnaires des enseignants et des questionnaires des élèves à la suite d'un test à grande échelle de mathématique de 9e année qu'on a fait passer en 2009-2010 ont été analysées à l'aide de la statistique descriptive. Les commentaires écrits des enseignants ont été analysés et regroupés par thème. Les résultats furent interprétés à la lumière d'un document de politique et de deux théories à propos des tests à grande échelle : les droits et la responsabilité des élèves tels qu'ils sont documentés dans *The Standards for Educational and Psychological Testing* (AERA/APA/NCME/JCSEPT, 2014), la théorie des attentes et de la valeur de la tâche et une théorie sur les stades de développement des élèves de 9e année. Plusieurs conflits ont été identifiés. Des solutions possibles sont présentées.

Mots-clés : test à grande échelle, motivation, développement de l'adolescence, normes, évaluation des élèves

Acknowledgement

This research was funded by a Social Science and Humanities Research Council (SSHRC) Standard Research Grant, and the data were provided by Ontario's Education Quality Accountability Office (EQAO). The opinions presented in this article are solely those of the authors and do not necessarily reflect the opinions of SSHRC or EQAO.

Introduction

The rights and responsibility of test takers is a key issue in educational testing and there are specific standards that outline good quality testing practices. One well-known document is the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], & Joint Committee on Standards for Educational and Psychological Testing [JCSEPT], 2014), hereafter referred to as the *Standards*. In the *Standards*, Standard 8.2 describes the rights of test takers: "Test takers should be provided, in advance, as much information about the test, the testing process, the intended test use, test scoring criteria, testing policy and confidentiality protection as is consistent with obtaining valid responses and making appropriate interpretations of test scores" (p. 134). In other words, test takers should be informed about the test provided it does not affect the validity of interpretations based on test results. The responsibility of test takers is described in the *Standards*: "The responsibility of test takers is to represent themselves fairly and accurately during the testing process" (p. 133). Examples of irresponsible test taker behaviours include those that result in an erroneously high score, such as cheating, or behaviours that result in an erroneously low score, such as low effort.

Collecting evidence of validity of interpretations based on test scores involves test users clearly stating the intended uses of test results and test score interpretations. When there are multiple uses and interpretations for the same test result, the validity of each test use and test score interpretation must be addressed individually (Kane, 2013). Multiple uses of the same test result occurs, for example, when large-scale tests are used for accountability and also for student class marks. Multiple uses of the same test occur frequently. In most Canadian provinces and some American states, teachers count a portion of the provincial (or state) large-scale assessment toward their students' classroom marks (Klinger, Deluca, & Miller, 2008; Miller, 2013; Simon, van Barneveld, King, & Nadon, 2011). The percentage used toward classroom marks and how it is used varies from province to province, state to state.

When large-scale test results are used for more than one purpose—e.g., as part of an educational accountability program and part of students' class marks—there is potential conflict between the rights and responsibility of test takers for each test use (Koch, 2013). Observing the rights and responsibility of test takers for one use of the test might

interfere with, or contradict, the rights and responsibility of test takers for the second use of the test. These conflicts, if unresolved, may negatively affect the validity arguments associated with test uses.

The purpose of this study was to identify the potential conflicts in the rights and responsibility of test takers when a large-scale assessment is used for two purposes—as part of an educational accountability program and as part of students' class marks—and identify potential solutions to the conflicts. We analyzed questionnaire data from Grade 9 teachers and students who participated in a large-scale educational test. Because Grade 9 students are adolescents who vary in their motivation to perform tasks and also who are at an important developmental stage, we used two theoretical frameworks to interpret and discuss our results, expectancy-value theory of motivation and the development of decision-making autonomy in adolescents.

Expectancy-Value Theory of Motivation

For this study, we focused on a theory of motivation that integrates expectancy and value constructs, entitled *expectancy-value theory* (Atkinson, 1964; Cole, Bergin, & Whittaker, 2008; Eccles & Wigfield, 2002; Pintrich, 2004; Wigfield & Eccles, 2000). We chose expectancy-value theory because it is the system of ideas guiding much of the current research on motivation and large-scale testing (Eklof, 2006; Sundre & Moore, 2002; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995), although there are other motivation theories that also apply to large-scale testing (Ryan, Ryan, Arbuthnot, & Samuels, 2007).

Expectancy-value theory links achievement performance, persistence, and choice directly to individuals' expectancy-related and task-value beliefs. Expectancy-related beliefs refer to individuals' beliefs about how well they will do on an upcoming task, either in the immediate or longer-term future (Eccles & Wigfield, 2002). Task-value beliefs are defined by four components: (a) attainment value—the personal importance of doing well on a task; (b) intrinsic value—the enjoyment the individual gets from performing the task; (c) utility value—how well the task relates to current and future goals, such as career goals; and (d) cost—negative aspects of engaging in the task, such as fear of failure. Eccles and Wigfield (2002) describe the theory:

> In this model, choices are assumed to be influenced by both negative and positive task characteristics, and all choices are assumed to have costs associated with them precisely because one choice often eliminates other options. Consequently, the relative value and probability of success of various options are key determinants of choice… Expectancies and values are assumed to directly influence performance, persistence, and task choice. (p. 118)

Applied to large-scale testing, expectancy-value theory states that a test taker's motivation to engage in activities related to large-scale testing depends on their belief about experiencing success on the test and the value that they place on the content, process, and/or outcomes of the test. That is, if a test taker believes they will experience success on the large-scale test and they value it, they are more likely to be motivated and engage with the tasks to the best of their ability.

Development of Decision-Making Autonomy of Students in Grade 9

For this study, we defined the development of decision-making autonomy of students as the development of the students' ability to think independently (Beckert, 2007) and to self-regulate their behaviour as part of the process of developing independence and self-guided action (Bandura, 1977; Collins, Gleason, & Sesma, 1997; Feldman & Wood, 1994). There were two reasons why we focused on students' development of decision-making autonomy. First, becoming more autonomous is a central developmental task for adolescents (Erikson, 1968; Wray-Lake, Crouter, & McHale, 2010). Second, adolescents make decisions about their behaviour that involve thinking about interpersonal relationships, social systems (Kohlberg, 1963, 1969), multiple alternatives, different consequences, and assessing the probability of positive and negative outcomes (Boyer, 2007; Inhelder & Piaget, 1958), even though they may not have the judgement, reasoning, or mental maturity most often associated with adult cognition (Brown, Tapert, Granholm, & Delis, 2000; Caskey & Ruben, 2003; Reyna, 2004; Spano, 2003).

Decision-making autonomy of students applies to large-scale testing in that, ultimately, students decide whether they will fully engage in the tasks associated with a large-scale educational assessment. In making this decision, they may consider multiple alternatives (e.g., their other assignments), the consequences of their behaviours (e.g., failing/passing), and the probability of positive or negative outcomes. For example, when

students are aware that assessment results don't count for marks, promotion, or graduation, students may decide to disengage with the tasks associated with the large-scale testing, like preparing for the assessment or responding to test items to the best of their abilities (DeMars, 2000; Setzer, Wise, van den Heuvel, & Ling, 2013; Wise, 2015; Wise & Demars, 2006; Wise & Kingsbury, 2016; Wolf & Smith, 1995). They may decide that their time and energy are better spent elsewhere.

Context of the Study

In Ontario, Canada, the Education Quality and Accountability Office (EQAO) is an independent provincial agency funded by the Government of Ontario. EQAO's mandate is to conduct province-wide tests in both official languages (i.e., English and French) at key points in every student's primary, junior, and secondary education, and report the results to educators, parents, and the public. The province-wide test results are used—along with other data like student classroom work and other assessment information—to account to the public about the education system and to inform the improvement of student learning strategies by helping to identify areas of learning that may need additional support. Province-wide test results are not used to evaluate teachers or determine school closures.

One of Ontario's provincial tests is the Grade 9 Assessment of Mathematics. This yearly assessment provides individual and system data on students' knowledge and skills, based on the expectations for students in Grade 9 applied and academic courses as documented in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (Ontario Ministry of Education, 2005). As defined in this curriculum document,

> Academic courses develop students' knowledge and skills through the study of theory and abstract problems. These courses focus on the essential concepts of a subject and explore related concepts as well. They incorporate practical applications as appropriate. Applied courses focus on the essential concepts of a subject, and develop students' knowledge and skills through practical applications and concrete examples. Familiar situations are used to illustrate ideas, and students are given more opportunities to experience hands-on applications of the concepts and theories they study. (p. 6)

In 2009–10, 101,268 (99%) English-language students in the academic courses and 47,566 (95%) English-language students in the applied courses participated in the Grade 9 Assessment of Mathematics (Education Quality and Accountability Office, 2010).

The EQAO Grade 9 Assessment of Mathematics, 2009–10 administration, provided teachers the option to use student responses on the test for classroom assessment. As stated in the 2017 *Grade 9 Assessment of Mathematics: Administration Guide*, "All or some components of assessment may be marked on completion of the assessment and prior to the return of materials to EQAO" (EQAO, 2017, p. 14). Marks may be used for course grades only according to Ontario Ministry of Education guidelines. The selection of items to mark, if any, and their weight toward students' grades, is a teacher, school, or board decision. Also, students' questionnaires and teachers' questionnaires included questions about the use of the assessment in students' class marks. The availability of these large-scale, empirical data allowed us the opportunity to explore the issue of rights and responsibility of test takers when large-scale testing is used for classroom assessment.

Method

Data

English data from the Grade 9 Assessment of Mathematics Student Questionnaire and Teacher Questionnaire for the 2009–10 test administration were provided to us by EQAO. The Teacher Questionnaire contained 25 questions, seven of these questions asked teachers about their use of the Grade 9 Assessment of Mathematics in students' class marks. The Student Questionnaire contained 12 questions, three of these questions asked students about the use of the assessment in students' class marks. We selected three questions from the Student Questionnaire (SQ) and three questions from the Teacher Questionnaire (TQ) as the focus of this study because these questions were sufficiently similar to justify a comparison of responses from teachers and students. The questions we selected are listed in Table 1. Please note that for TQ 22, there were four lines available for teachers to provide a written comment.

Table 1. Selected questions from the EQAO Grade 9 Assessment of Mathematics student questionnaire and teacher questionnaire

Student Questionnaire	Teacher Questionnaire
SQ 10. Will your teacher count some or all parts of the Grade 9 Assessment of Mathematics as part of your class mark? (Response options: Yes, No, or Don't know)	TQ 19(a). Do some or all components of the Grade 9 Assessment of Mathematics count as part of your students' class marks? (Response options: Yes or No)
IF YES…	IF YES…
SQ 11(a). Were you told how much the assessment will count as part of your class mark? (Response options: Yes or No)	TQ 21. Before writing the Grade 9 Assessment of Mathematics, were students informed about the weight it would be given in the calculation of their class mark? (Response options: Yes or No)
SQ 12. Does counting the Grade 9 Assessment of Mathematics as part of your class mark motivate you to take the assessment more seriously? (Response options: Yes, No, or Undecided)	TQ 22. In your opinion, does counting some or all components of the Grade 9 Assessment of Mathematics as part of class marks motivate students to take the assessment more seriously? (Response options: Yes, No, or Undecided) Please comment:

Merging Student and Teacher Questionnaires

The TQ data were matched to the SQ data to create one file that identified teachers' responses and their students' responses to questionnaire items. We used the SQL merging strategy described by Sarwar, Zerpa, van Barneveld, Simon, and Brinson (2013).

Although many teachers in Ontario teach both academic and applied math courses, teachers were asked to complete only one questionnaire for the type of course that included the majority of their students. When matching the teachers' and students' data, those students whose teachers did not fill out the questionnaire for their course were dropped. In addition, students were dropped because their teachers did not answer the selected question from the questionnaire. The numbers of student responses per question are found in the tables in the results section.

Analysis

We used four steps to analyze the questionnaire data. First, we calculated frequencies and percentages for teachers' responses to selected questions from the TQ data file

for academic and applied courses. Missing or ambiguous data for each question were excluded from the analysis. Second, we calculated frequencies and percentages to student responses to selected questions from the SQ data file for academic and applied courses. Again, missing or ambiguous data for each question were excluded from the analysis. Third, we calculated frequencies and percentages of matched and mismatched responses between students and their teachers using data from the merged file. For this analysis, we included only those data where "Yes" or "No" responses were provided. Matches were defined as follows: If a student responded "Yes" to a question and their teacher responded "Yes" to the similar question on the TQ, then we considered that a Yes-Yes match. If a student answered "No" and their teacher answered "No" we considered that a No-No match. Mismatches were either Yes-No, indicating that a student responded "Yes" and their teacher responded "No," or No-Yes, indicating that a student responded "No" and their teacher responded "Yes." The percentages were calculated as the ratio of the frequency (e.g., frequency of Yes-Yes) over the sum of the frequencies of all possible matches or mismatched response pairs (i.e., the sum of the frequencies for Yes-Yes, No-No, No-Yes, and Yes-No). Fourth, we scanned the written comments of teachers and we selected those with any reference to the rights and responsibility of test takers. We analyzed the content of these comments and formulated meanings.

Results

Table 2 contains the frequencies and row percentages of teachers' responses to selected questions from the TQ for academic and applied courses. Almost all the teachers reported counting some or all of the Grade 9 Assessment of Mathematics as part of students' class marks and reported informing the students of the weight it would be given in the calculation of their class mark. The majority of teachers—83% for applied courses and 89% for academic courses—opined that counting the test from some or all of students' class marks motivated the students to take the test more seriously.

Table 2. The number (and row %) of teacher responses to selected questions from the Grade 9 EQAO Mathematics teacher questionnaire

Question	Courses	Teacher Response			
		Yes	No	Undecided	Total
Do some or all of the Grade 9 Assessment of Mathematics count as part of your students' class marks?	Applied	1,853 (99)	12 (1)	–	1,856
	Academic	2,718 (99)	16 (1)	–	2,734
IF YES…					
Before writing the Grade 9 Assessment of Mathematics, were students informed about the weight it would be given in the calculation of their class mark (e.g., 5%, 10%)?	Applied	1,784 (96)	69 (4)	–	1,853
	Academic	2,647 (97)	71 (3)	–	2,718
In your opinion, does counting some or all components of the Grade 9 Assessment of Mathematics as part of class marks motivate students to take the assessment more seriously?	Applied	1,538 (83)	121 (7)	194 (10)	1,853
	Academic	2,430 (89)	86 (3)	202 (8)	2,718

Table 3 contains frequencies and row percentages of student responses to selected questions from the SQ for academic and applied courses. More than half the students in applied courses (58%) reported *not* knowing if some or all of the Grade 9 Assessment of Mathematics counted toward their class marks; the majority of students in academic courses (66%) reported that it did count toward their class marks.

To explore potential explanations for why more than half the students in applied courses might not know if some or all of the assessment counted toward their class mark, we calculated the frequency of student self-reported lateness and absenteeism from math class; these data were obtained from the SQ. Students in applied courses were twice as likely as those in academic courses to be absent 10 or more times (6% and 3%,

respectively) and three times as likely to be late five or more times (6% and 2%, respectively). While these percentages are not high, they may explain, in part, the difference in results for the students in applied and academic courses.

Students reported that they were told about the weight of the test toward class marks; students in academic courses reported this more frequently than those in applied courses (91% and 86%, respectively). The majority of students reported that counting the assessment motivated them to take it more seriously; students in academic courses reported this slightly more frequently than students in applied courses (73% and 70%, respectively).

Table 3. The frequency (and row %) of student responses to selected questions from the EQAO Grade 9 Assessment of Mathematics student questionnaire

Question	Courses	Number of student responses (%)			
		Yes	No	Don't know/ Undecided[a]	Total
Will your teacher count some or all parts of the Grade 9 Assessment of Mathematics as part of your class mark?	Applied	12,420 (39)	1,060 (3)	18,604 (58)	32,084
	Academic	45,392 (66)	1,340 (2)	21,769 (32)	68,501
IF YES…					
Were you told how much the assessment will count as part of your class mark (e.g., 0%, 5%, 10%)?	Applied	10,586 (86)	1,653 (14)	–	12,239
	Academic	41,100 (91)	3,974 (9)	–	45,074
Does counting the Grade 9 Assessment of Mathematics as part of your class mark motivate you to take the assessment more seriously?	Applied	8,502 (70)	1,628 (13)	2,079 (17)	12,209
	Academic	32,467 (73)	6,044 (13)	6,185 (14)	44,696

[a] "Don't know" is a response option to the question "Will your teacher count some or all parts of the Grade 9 Assessment of Mathematics as part of your class mark?" "Undecided" is a response option to the question "Does counting the Grade 9 Assessment of Mathematics as part of your class mark motivate you to take the assessment more seriously?"

Table 4 contains frequencies and percentages of student–teacher matches and mismatches to selected questions from the questionnaires. The response of almost all the students in applied courses (94%) and academic courses (98%) matched the responses of their teachers when asked if the teacher will count some or all parts of the Grade 9 assessment of mathematics as part of their class mark. The matches were somewhat lower for the question regarding the weight that the assessment will count as part of the students' class mark. For this question, 86% of responses from students in applied courses and 91% of responses from students in academic courses matched those of their teachers. The third question, which asked about whether counting the assessment as part of student class marks motivates them to take the test more seriously, had the lowest number of student–teacher response matches. For students in applied courses, 79% of their responses matched those of their teachers. For students in academic courses, 82% of their responses matched those of their teachers.

For all of the questions, when there was a mismatch between students' responses and those of their teachers, it tended that the students reported "No" and the teachers reported "Yes." Response mismatches appeared to be higher for students in the applied courses than for those in the academic courses. Response mismatches appeared to be higher for the third question, which asked about the impact of counting some or the entire test on the motivation of students to take the assessment more seriously.

Table 4. The frequency (and row %) of student–teacher response matches and mismatches to selected questions from the EQAO Grade 9 Assessment of Mathematics student questionnaire

Question	Courses	Number of student–teacher response matches or mismatches (%)				
		Yes-Yes	No-No	No-Yes	Yes-No	Total
Will your teacher count some or all parts of the Grade 9 Assessment of Mathematics as part of your class mark?	Applied	8,364 (94)	6 (<1)	473 (5)	15 (<1)	8,858
	Academic	33,269 (98)	7 (<1)	532 (2)	119 (<1)	33,927
IF YES-YES…						
Were you told how much the assessment will count as part of your class mark (e.g., 0%, 5%, 10%)?	Applied	7,231 (86)	99 (1)	883 (11)	151 (2)	8,364
	Academic	30,230 (91)	406 (1)	2,223 (7)	410 (1)	33,269

Does counting the Grade 9 Assessment of Mathematics as part of your class mark motivate you to take the assessment more seriously?	Applied	6,639 (79)	84 (1)	1,253 (15)	388 (5)	8,364
	Academic	27,312 (82)	186 (1)	4,921 (15)	850 (3)	33,269

Table 5 contains samples of written comments of teachers and formulated meanings, grouped into two themes: the rights of test takers and the responsibility of test takers. Teachers' comments regarding the rights of test takers tended to focus on the impact of informing the students about which test items counted toward their class grades. Specifically, teachers' comments suggested that some teachers do not fully inform the students about which test items count because they want to motivate their students to apply their abilities to the whole test. Other teachers, who do fully inform their students about which items count, report concerns regarding the effort students make on those items that do not count for class marks. With respect to the responsibility of test takers, teachers' comments suggested that students assess the value of participating in the EQAO test, sometimes against their other commitments (e.g., school assignments) or interests (e.g., World Cup Soccer). Student behaviours related to preparing for and representing themselves honestly on the test appear to vary depending on the value they assign to the test and the course (i.e., academic or applied).

Table 5. Selected examples of significant comments by teachers and related formulated meanings

Theme	Significant Teacher Comment	Formulated meaning
Rights of test takers	"We don't tell the students which parts are going to count. Knowing that some of it will count definitely motivates them to try harder and take it more seriously." "Students heard rumours that only mc (multiple choice) were going to count and asked if they could just leave open response blank (We later told them one open response question was going to count)." "Not informed until after EQAO completed (not informed as to what to count)."	If students don't know which items count for marks, they may be motivated to make an effort on the whole test.

Theme	Significant Teacher Comment	Formulated meaning
	"Students spent most of their time on the multiple choice questions and left the open response questions blank or incomplete when they were told only the MC will count as part of their mark." "Some students know we only count the multiple choice and take the open questions less seriously."	Students will select which portions of the test to apply their efforts, based on their understanding of which items "count."
	"We mark multiple choice as 5% of the final mark. Students know this weight and it helps them to be motivated to succeed."	When informed about the weight, students may be motivated to take some or all parts of the test more seriously
Responsibility of test takers	"They do not take it seriously if they don't feel they are 'getting something' for it." "Many students have commented 'what is the point' of EQAO? With all of their courses getting busy with final assignments, they are not concerned about doing well on something that does not count as marks."	Students assess the value of the test.
	"If you don't count it, students either skip or do not take the time to answer the questions." "Many students only do work when they are being marked. If the test is not worth anything, they will not try during the test or during the preparation." "Attendance is a problem in my class, as is punctuality. Assigning the EQAO a mark made my students all show up on time, otherwise half would stay home to watch World Cup Soccer."	Student behaviours vary by test stakes.
	"It does help, but by the end of the semester so many applied students have stopped working."	The effectiveness of the strategy depends on the timing of the assessment and the program of the student.

Discussion

There are two potential conflicts in the rights and responsibility of test takers when large-scale assessment is used as part of an educational accountability program and also as part of students' class marks. The first conflict relates to the rights of test takers to be informed about the test. The second conflict relates to their responsibility as test takers to represent themselves honestly on the test.

Test Takers' Right to Be Informed vs. Obtaining Valid Responses

There is a conflict between test takers' right to be informed about the uses of the test and obtaining valid responses to test questions. For example, half the students in applied math courses and 23% in academic math courses reported not knowing if their teachers were going to count some or all parts of the large-scale assessment as part of their class mark. Student absenteeism or lateness does not seem to explain this result. There appears to be a communication problem (or a reporting problem). If the test *does* count toward class marks, then this result is inconsistent with Standard 8.2 (AERA/APA/NCME/JCSEPT, 2014) because some students report that they were not informed about the test stakes. Since most students reported that counting the test for class marks motivated them to take it more seriously, informing them that it counts might result in more valid responses. To be consistent with Standard 8.2, the communications about the test could be reviewed and enhanced. Test takers have the right to be informed about this use of the test and steps should be taken to ensure that they received relevant information. If the test, however, *does not* count for class marks then the results of this study are consistent with Standard 8.2 if the teachers are withholding information about the test in an attempt to create testing conditions which will produce valid responses.

The conflict is salient when informing students about the details of the test. For example, some students reported not being told about the weight that the assessment counted toward their class mark. Teachers' comments highlighted the conflict between test takers' right to be informed about the test and obtaining valid responses; some teachers suggested that providing only partial information to students (e.g., informing them that the test counts but not informing them which items count) might increase the likelihood that students would provide valid responses to test questions. When students were informed about which items counted, they did not take the other items as seriously. Some teachers reported that they do not inform the students about which items count because they think it will reduce the validity of responses. Does not informing the students about which items count constitute a violation of the students' rights to be informed about the test? According to Standard 8.2 (AERA/APA/NCME/JCSEPT, 2014), no, because the standard explicitly states that informing the student should be consistent with obtaining valid student responses. Not informing the students may be problematic in that students may perceive their teachers as manipulating the extent to which they can assess the value

of the assessment, withholding information, or reducing students' decision-making autonomy. Not informing the students may weaken the relationship between students and their teachers, and/or may result in protest-related test-taking behaviors on the part of students, such as non-participation in the assessment.

Some teachers may be put in a difficult position when it comes to informing students about the assessment. The tension for teachers is that either they do not fully inform the student about the test—which may involve some level of deception—and promote valid test responses *or* they fully inform the students—no deception—and increase risk that the student will produce invalid responses for some or all of the test. Teachers might experience significant professional tension with either of these options as this dichotomy challenges the balance between their commitment to implementing the large-scale test so that it produces valid student scores, their desire to respect the rights of their students, and their desire for a good relationship with their students. We need to find a solution where the rights of the students are respected, teachers' integrity is preserved, and the test results are valid.

Test Takers' Responsibility to Represent Themselves Honestly on the Test vs. the Value of the Test

There is conflict between test takers' responsibility to represent themselves honestly on the test and the value of the test for students. The students' perceptions about the value of the test are important. Students and teachers reported that counting the assessment for class marks motivated students to take it more seriously, suggesting that having it count for class marks added some value to the assessment for students. This result is consistent with the "value" part of the expectancy-value theory of motivation, in that the greater the perceived value of the task the more motivated one may be to engage in it. For this study, evidence suggested that most students value their class marks although we acknowledge that there is a small number of students who are not motivated by marks.

Students considered the value of the test when making decisions about their engagement with the assessment. If a student considers the value of engaging with the tasks of the large-scale test against their other commitments (e.g., completing other school assignments which count for marks) and their desires (e.g., recreation time) and then judges the test to be a low priority, is that irresponsible or responsible behaviour? According to

the *Standards* (AERA/APA/NCME/JCSEPT, 2014), it is irresponsible because, for the test, they are not applying themselves to the best of their ability.[1] When we consider the theoretical frameworks for decision-making autonomy of students in Grade 9, we have a different interpretation. The prioritization of tasks and behaving according to those priorities is consistent with students' developing autonomy as decisions-makers. From this perspective, the students may be considered "responsible" or at least "practicing responsibility" consistent with their developmental stage. Some students' decision to *not* engage with the assessment may be a reasonable and rational decision. Indeed, given the same circumstances, many mature, reasoning adults might make the same decision. We need a solution where the test is valuable to students such that engaging with it to the best of their abilities is meaningful.

Potential Solutions

The potential solutions to the conflicts can be grouped in two categories: solutions that improve communications about the assessment and solutions that encourage students to represent themselves honestly on the assessment. Solutions that improve communications about the assessment include:

1. Review the communication plan for informing students about the test. Consider current, new, and multiple communication strategies, make modifications as necessary, and evaluate the effectiveness of modifications to the communication strategies.
2. Teachers select items to mark for class grades *after* the administration of the test but *before* they see student responses. The teachers would not look at the test ahead of time. Once the test administration begins, teachers may mark some or all items that students had the best opportunity to learn. Teachers could honestly inform the students that they do not know in advance which items will count for the class mark and that students should respond to all items to the best of their ability. There is no deception in this teacher statement, it is fully informative to the student, it is consistent with producing a valid student response, and may even strengthen the alliance between the student and teacher.

1 It is possible that the students are not aware of their responsibility as documented in the *Standards*.

Solutions that encourage students to represent themselves honestly on the assessment focus on raising the value of the assessment for students. Solutions may include:

1. The large-scale assessment counts toward a meaningful weight of class marks. The determination of "a meaningful weight" depends on the students, contextual factors, and may be determined by a collaboration of teachers, schools, school boards, and testing organizations and other experts.
2. The large-scale assessment is the final exam for the math class. The teacher marks all parts of the test. This solution may reduce the end-of-year assessment load for students and marking time for teachers.
3. Have the large-scale assessment count as a high school graduation requirement. This solution represents the highest stakes assessment condition for students. It would not reduce their assessment load, and would increase motivation but would also increase their stress related to the large-scale assessment.

While these potential solutions may help to resolve the conflicts described in this article, we acknowledge that there are other potential solutions that are not directly related to the evidence in this research. One might argue, for example, that a potential solution to these conflicts is to improve student effort on math-related tasks by making the learning and evaluation of math more intrinsically enjoyable and meaningful to more students. This solution is consistent with the theoretical frameworks of expectancy-value theory, decision-making autonomy, and would improve the validity of large-scale math assessments. Others might argue that we should terminate the use of large-scale provincial tests altogether. Since our research did not produce evidence that directly informs these other solutions, we do not discuss them at length in this article; however, we acknowledge that they are important topics.

This research highlights important practices in large-scale assessment and proposes suggestions for improvement of educational assessment practices in Canada. The solutions we propose in this article should be considered in relation to the specific contextual variables in large-scale test administrations, evaluated, and then used to inform edits to the standards and policies on rights and responsibility of test takers.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Beckert, T. E. (2007). Cognitive autonomy and self-evaluation in adolescence: A conceptual investigation and instrument development. *North American Journal of Psychology*, *9*(3), 579–594.

Boyer, T. W. (2007). Decision-making processes: Sensitivity to sequentially experienced outcome probabilities. *Journal of Experimental Child Psychology*, *97*, 28–43.

Brown, S. A., Tapert, S. F., Granholm, E., & Delis, D. C. (2000). Neurocognitive functioning of adolescents: Effects of protracted alcohol use. *Journal of the Research Society on Alcoholism*, *24*, 164–171.

Caskey, M. M., & Ruben, B. (2003). Research for awakening adolescent learning. *The Education Digest*, *69*, 36–38.

Collins, W. A., Gleason, T., & Sesma, A. (1997). Internalization, autonomy, and relationships: Development during adolescence. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 78–99). New York, NY: Wiley.

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, *33*(4), 609–624.

Demars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education*, *13*(1), 55–77.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.

Eklof, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, *66*(4), 643–656.

Erikson, E. H. (1968). *Identity: Youth and crisis*. New York, NY: W. W. Norton.

Education Quality and Accountability Office (EQAO). (2010). *EQAO's provincial secondary school report on the results of the Grade 9 assessment of mathematics and the Ontario secondary school literacy test, 2009–2010*. Retrieved from EQAO website: http://www.eqao.com/en/assessments/results/assessment-docs-secondary/provincial-report-secondary-2010.pdf

Education Quality and Accountability Office (EQAO). (2017). *Grade 9 Assessment of Mathematics: Administration Guide 2017*. Retrieved from EQAO website: http://www.eqao.com/en/assessments/grade-9-math/assessment-docs/administration-guide-g9.pdf

Feldman, S. S., & Wood, D. N. (1994). Parents' expectations for preadolescent sons' behavioral autonomy: A longitudinal study of correlates and outcomes. *Journal of Research on Adolescence*, *4*, 45–70.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73.

Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, *76*, 1–34.

Koch, M. J. (2013). The multiple-use of accountability assessments: Implications for the process of validation. *Educational Measurement: Issues and Practice*, *32*(4), 2–15.

Kohlberg, L. (1963). The development of children's orientations toward a moral order. *Vita Humana*, *6*, 11–33.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago, IL: Rand McNally.

Miller, T. (2013). Using large-scale assessments to determine student grades. *Canadian Journal of Education*, *36*(3), 317–353.

Ontario Ministry of Education. (2005). *The Ontario Curriculum, Grades 9 and 10: Mathematics*. Toronto, ON: Ontario Ministry of Education.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407.

Reyna, V. F. (2004). How people make decisions that involve risk: A dual process approach. *Current Directions in Psychological Science*, *13*, 60–66.

Ryan, K. E., Ryan, A. M., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, *36*(1), 5–13.

Sarwar, G. S., Zerpa, C., van Barneveld, C., Simon, M., & Brinson, K. (2013). Merging large-scale assessment data for secondary analysis: Experiences with EQAO's data. *Journal of Education and Learning*, *2*(2), 44–54.

Setzer, J. C., Wise, S. L., van den Heuvel, J. R., & Ling, G. M. (2013). An investigation of examinee test-taking effort on a large-scale assessment. *Applied Measurement in Education*, *26*(1), 34–49.

Simon, M., van Barneveld, C., King, S., & Nadon, C. (2011, April). *Teachers' Use of Large-scale Assessment Results: Grading Issues and Policies*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Spano, S. (2003). Adolescent brain. *Youth Studies Australia*, *22*, 36–39.

Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update*, *14*(1), 8–9.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.

Wise, S. L. (2015). Effort analysis: Individual score validation of achievement test data. *Applied Measurement in Education*, *28*(3), 237–252.

Wise, S. L., & Demars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement*, *43*(1), 19–38.

Wise, S. L., & Kingsbury, C. G. (2016). Modeling student test-taking motivation in the context of an adaptive achievement test. *Journal of Educational Measurement*, *53*(1), 86–105.

Wolf, L. F., & Smith, J. F. (1995). The consequence of consequence—motivation, anxiety, and test-performance. *Applied Measurement in Education*, *8*(3), 227–242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education*, *8*(4), 341–351.

Wray-Lake, L., Crouter, A. C., & McHale, S. M. (2010). Developmental patterns in decision-making autonomy across middle childhood and adolescence: European American parents' perspectives. *Child Development*, *81*(2), 636–651.