The Alberta Journal of Educational Research *Vol. 52, No. 2, Summer 2006, 107-126*

Sonia Ben Jaafar
Ontario Institute for Studies in Education of the University of Toronto

An Alternative Approach to Measuring Opportunity-to-Learn in High School Classes

The Opportunity-to-Learn framework has provided policymakers and researchers a means to develop strategies to measure classroom practices. In particular, the measure of the delivered content has been shown to be a good predictor of student achievement on tests. The method presented in this article uses classroom artifacts as the main data source to determine the attention teachers give to various content in the curriculum. The number of treatments that address set learning outcomes was the unit of measurement employed in this method. The article illustrates how this method was used to describe content delivery and how content emphasis exposed the differences between two teachers following the same prescribed syllabus. This method is best applied at the secondary school level to measure one component of the delivered curriculum. Finally, the limitations and potential of this method are discussed for use in research and for school improvement.

Les décideurs et les chercheurs ont eu recours au cadre « Opportunity to Learn » (occasion d'apprentissage) dans le développement de stratégies pour mesurer les pratiques en salle de classe. En fait, la mesure du contenu fourni aux élèves s'est avérée être un bon prédicteur de la performance des élèves aux examens. Cet article présente une méthode s'appuyant sur les artéfacts de la salle de classe comme source principale de données et visant l'évaluation de l'attention qu'accordent les enseignants à divers contenus du programme d'études. Le nombre d'interventions visant des résultats d'apprentissage déterminés a constitué l'unité de mesure. L'article démontre, d'une part, l'emploi de cette méthode pour décrire la prestation de contenu et d'autre part, comment le fait de porter attention au contenu fait ressortir des différences entre deux enseignants qui suivent le même syllabus. Cette méthode s'applique le mieux au niveau secondaire pour mesurer une composante du programme d'études. L'article termine par une discussion des limites et du potentiel de cette méthode dans le contexte de la recherche et de l'amélioration du rendement scolaire.

The measurement of student Opportunity-to-Learn (OTL) is increasing in significance as standards, accountability, and equity interact and redefine values embedded in education systems. Standards homogenize expectations for schooling across social and physical geographies. These standards act as the pillars of accountability that hold schools responsible for student achievement to promote equity in education (Murphy & Datnow, 2003; Skrla & Scheurich, 2004). Policymakers monitor access to education by measuring student achievement to ensure that performance standards are met for various populations. But outcomes and output measures are insufficient for securing access to quality education for all. In response to this shortcoming, input variables such as OTL, which capture classroom practices that support student achievement, have become more important. The relationship between OTL and student achievement is elemental in considering the effect of school on student success.

Sonia Ben Jaafar is currently a doctoral candidate in the Department of Theory and Policy Studies. Her research involves OTL, accountability, and school improvement.

The use of student performance on standardized tests as an indicator of school effectiveness to ascertain the influence of improvement efforts is a norm in education systems (Fitz-Gibbon & Kochan, 2000; Reynolds, Teddlie, & Stringfield, 2000). Performance-based accountability systems assume "that setting high standards of achievement will inspire greater efforts on the part of students, teachers, and educational administrators" (American Educational Research Association, 2000, p. 24). However, pressuring schools to provide improved opportunities to learn is problematic (Linn, Baker, & Betebenner, 2002; Ontario Secondary Schools Teachers' Federation, 2002; People For Education, 2003; Popham, 1999; Zancanella, 1992). The discrepancy between the *intended* curriculum, what policymakers have standardized across the jurisdiction, the *delivered* curriculum, what actually gets taught in each classroom, and the *tested* curriculum has been addressed in OTL studies. Indicators of classroom practices are difficult to measure because of the variability inherent in teaching, inconsistencies in researcher and practitioner understanding of terminology, and the logistical challenges of monitoring thousands of classrooms. These practical and theoretical problems make defining, operationalizing, and measuring OTL in a valid and reliable manner a challenging endeavor.

OTL represents the school-controlled factors that influence student achievement. The concurrent promotion of standards, accountability, and equity calls for a systematic monitoring of OTL to ensure that schools support *all* students in their achievement of learning the intended curriculum. In this article, I present an alternative method for measuring OTL in secondary schools where student artifacts serve as the primary data source. This article is divided into four sections. First, I situate and define the OTL framework and the relevant operational terms. Second, I describe and contextualize the case study and the method. Third, the results and findings from the analyses are presented to illustrate the application of the method. Finally, I discuss the merits and limitations of this OTL measurement method with special attention to future work and the implications for practice.

Operationalizing Opportunity-to-Learn

OTL is a construct that measures classroom practices to account for curriculum delivery. Carroll (1963) first formalized OTL as an operational construct in the educational field in his early model of school learning. The model consisted of two variables that were external to the student: Opportunity-to-learn and quality of teaching. Time was used as the unit of measurement for OTL and was regarded as the critical factor determining degree of learning.

Degree of Learning = [time actually spent]/[time needed] (Carroll, 1963)

This model assumed that all students could succeed in a given task if given adequate time to engage actively in learning relative to their aptitude, which determines *time needed*. The potential for student success in school was represented as a ratio of the time they needed to learn over the time they were given to learn.

In his reviews of the research conducted employing OTL variables derived from his work, Carroll (1989) supported Karweit's (1983) emphasis on time-on-task by recognizing that "time as such is not what counts, but what happens in that time" (p. 27). For Husén (1967), the alignment between the content taught and the content tested operationalized OTL, a definition adopted by the IEA studies where the reported curriculum taught was found to predict student achievement. When students were given a chance to learn the concept being tested, they were more likely to answer the test question correctly (Muthen, Kao, & Berstein, 1991). Although the findings remained undisputed, this approach to OTL was later criticized based on concerns of unstable judgment about teaching adequacy (Nagy, 1996).

Time-on-task and content alignment proved insufficient for reliably operationalizing OTL as a construct subject to stable measurement. Stevens (1993) refined OTL as a multidimensional construct consisting of four components operating at the classroom-level: content coverage, exposure, emphasis, and quality of instructional delivery. Her framework characterizes OTL with the specificity necessary to circumvent the variance in OTL definitions and measurement methods that problematized claims about the effect of OTL on student achievement (Karweit & Slavin, 1982).

The method presented below adopts Stevens' (1993) three content OTL variables. The content OTL variables are defined as follows: (a) *content coverage* addresses the issue of coverage of specified topics or learning outcomes in a given grade level; (b) *content exposure* reflects the attention given to learning outcomes (Winfield, 1987); and (c) *content emphasis* reflects the relative attention given to learning outcomes throughout the course (McDonnell, Berstein, Ormseth, Catterall, & Moody, 1990). Relative attention is the attention offered each learning outcome compared with the attention given to all outcomes. Wang (1998) reviewed the literature and identified common measures for each OTL dimension. He concluded that content coverage was measured through teachers' reports and textbook analyses; content emphasis was measured through reported practices; and content exposure was measured through observations, which proved to be methodologically unstable for reliable conclusions (Karweit & Slavin, 1982).

Wang (1998) found that content exposure was the most important predictor of student achievement on written tests. His findings corroborated past findings that optimizing time-on-task was one of the most important factors in improved student achievement when written tests were used (Hawley, Rosenholtz, Goodstein, & Hasselbring, 1984). However, quality of instructional delivery was the OTL predictor for student achievement in hands-on tests (Wang). This distinction highlights the importance of achievement measures given claims about OTL effect. If the operational constructs of OTL lack detail, the distinction is lost. Hence this level of detail is needed to describe the link between teaching and student achievement measures (Traub & Wolfe, 1981).

Issues of validity and reliability of content exposure are exacerbated by insufficient detail in operationalizing OTL (Lapointe, Mead, & Phillips, 1989). Porter (2002) addressed these criticisms with a method that reliably measured the alignment of content with curriculum standards and state tests. His method measures content coverage using teacher-reported time spent on a topic. Berstein et al. (1995) support the validity of this measure as long as the content measure remains at a general level and reform terminology are not employed.

In this article, I present an alternative method for measuring content OTL variables that does not rely on time as the unit of measurement. I use the construct of Attention to Learning Outcome (ATLO) to represent content delivery. ATLO is based on the number of treatments given to a single learning outcome in the curriculum during the course of a class. I am not suggesting that time is not an important element of classroom practice, but it may not serve as an accurate proxy for content OTL variables for valid interpretations of curriculum implementation.

The Study Site

The data for this study were collected in 2001 in two classes in one school in southern Ontario. The school, housed in a large urban board in southern Ontario, was selected because of the above average success rate of the students on the grade 9 provincial assessment. A high-achieving school was selected to reflect the presumption that high assessment scores represent an alignment between the *delivered* and *intended* curricula. Two mathematics teachers participated in this study. Both were teaching the grade 9 *Principles of Mathematics* course, had been teaching at the school for over three years, and had taught the new curriculum in the school since its introduction. The data collected were linked to each teacher as the person responsible for the delivered curriculum and the unit of analysis.

At the time of data collection, Ontario schools were implementing a series of changes to the education system, which were a part of the then conservative government's "common sense" reform strategies that began in 1995. One core change was a new outcomes-based curriculum for all courses and the creation of the Education Quality and Accountability Office (EQAO, 1996), an independent organization responsible for ensuring greater accountability and quality of schooling in the province through "assessments and reviews based on objective, reliable and relevant information, and the timely public release of that information along with recommendations for system improvement." One year after the introduction of the new grade 9 mathematics curriculum, the EQAO institutionalized a standardized province-wide assessment to be administered at the end of all grade 9 mathematics courses. The testing is intended to evaluate the success of the implementation process.

Although the EQAO articulated that the received curriculum would be measured, it is rarely

> possible to measure what students learn with sufficient accuracy to lead to unequivocal conclusions about the effectiveness of a new curriculum. The assessment scores of students depend on much more than the curriculum itself, and there are also numerous unanticipated consequences and unknown side effects of any curriculum, which test scores do not begin to get at. (Marsh & Willis, 1999)

Because assessment scores cannot adequately measure curriculum implementation, they cannot serve as evaluative tools for its success. Yet the EQAO stated that this was an aim of the assessment (Ontario Ministry of Education, 2000). To address this limitation, this study regarded the delivered curriculum as the intermediary stage to ascertain students' opportunity to learn the intended curriculum in the classroom.

Method: Measuring the Delivered Curriculum

This study examined the effects of the grade 9 EQAO assessment on the implementation process in grade 9 mathematics classrooms. Measuring the alignment of the delivered curriculum to the intended one was essential to examine the implementation of the new curriculum. The intended curriculum was established through a document analysis of relevant Ministry of Education policies. The delivered curriculum was established through grade 9 mathematics classroom artifact analysis triangulated with classroom observations and teacher interviews (Mathison, 1988; Merriam, 1998).

Data collection and analysis were done in two sequenced states so that the delivered curriculum could be juxtaposed with the intended one as a point of reference. First, a document analysis of the Ontario Ministry of Education policy statements established the constructs defining curriculum and its intentions for classroom practices. These constructs were fitted to Stevens' Opportunity-to-Learn framework to shape the empirical data collection for a "*yield* of mathematics instruction" (Travers, Westbury, & International Association for the Evaluation of Educational Achievement, 1989). The analytical results from the provincial guides detailing the intended curriculum were used as a coding scheme for stage two data. This approach facilitated the analysis as data representing classroom practices were coded using a scheme drawn from the intended curriculum (see Figure 1).

The stage two data collection strategy was developed to capture the teacher-intended content delivery associated with the implemented curriculum (Travers et al., 1989). All activities, interactions, exchanges, and communications between the students and the teacher's intentions were considered noteworthy. The data collected from each classroom were examined individually to define the curricular practices of each teacher. These sets of practices were then compared, treating each teacher as a single unit.

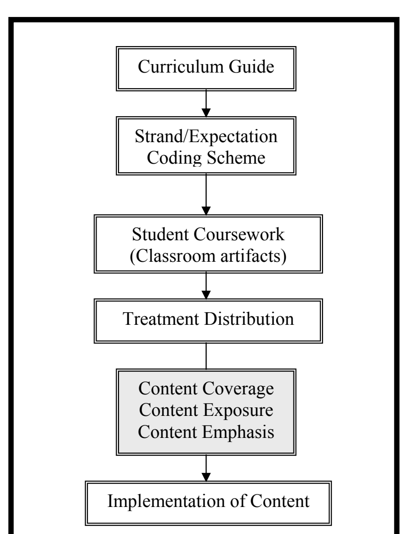

Figure 1. Content analysis framework.

Classroom Data Collection

Data collection started six weeks before the grade 9 assessment (April) and continued until the end of the scholastic year (June). Multiple instruments and data-types were used to check the validity of findings between data-types and expose the limitations concealed in the data (Smithson & Porter, 1994). Data from various sources were collected for an in-depth holistic treatment of the delivered curriculum (Cohen, Manion, & Morrison, 2000). The data collected were:

1. Documents and artifacts (ministry documents, curricular resources, student materials, and teacher materials).
2. Daily field notes of informal exchanges.
3. Six observations of each grade 9 class (total of 12 75-minute observational sessions). The first two observational sessions for each class were unstructured narrative accounts of the classroom activities. The remaining sessions were structured using an observation tool developed from the findings in stage one of the study.
4. Twelve post-observation open-ended interviews (Brown & McIntyre, 1993) with the teacher ranging from six to 17 minutes.

Table 1
Coding Scheme for Grade 9 Overall Learning Outcomes

Code	*Overall Expectations*
NSA	*Number Sense and Algebra*
NSA1	solve multi-step problems requiring numerical answers, using a variety of strategies and tools
NSA2	demonstrate understanding of the basic exponent rules and apply them to simplify expressions
NSA3	manipulate first-degree polynomial expressions to solve first-degree equations
NSA4	solve problems, using the strategy of algebraic modeling
RTL	*Relationships*
RLT1	determine relationships between two variables by collecting and analysing data
RLT2	compare the graphs and formulas of linear and nonlinear relations
RLT3	describe the connections between various representations of relations
AG	*Analytical Geometry*
AG1	determine the relationships between the form of an equation and the shape of its graph with respect to linearity and nonlinearity
AG2	determine the properties of the slope and y-intercept of a linear relation
AG3	solve problems, using the properties of linear relations
MG	*Measurement and Geometry*
MG1	determine the optimal value of various measurements facilitated, where appropriate, by the use of concrete materials, diagrams, and calculators or computer software
MG2	solve problems involving the surface area and the volume of three-dimensional objects
MG3	formulate conjectures and generalizations about geometric relationships involving two-dimensional figures facilitated by dynamic geometry software, where appropriate
EXT	*Extensions*
EQAO	*EQAO Assessment*

5. Three 45-minute semistructured interviews with each teacher. The last interview was a member check.
6. Sixty-minute interviews with three departmental leaders identified by the teacher participants as influential in creating and maintaining local policy.
7. Follow-up member-check interviews with teacher-identified leaders.

Field notes were made of all meetings and observations, and all interviews were tape-recorded and transcribed (one of the department leaders refused to be audiotaped, so I scripted the interview responses). The transcribed interviews and detailed field notes were uploaded to Atlas.ti. for analysis, and the documents were filed and analyzed manually. Although the data collection was extensive and analyses were completed on all dimensions of curriculum using an identical coding scheme (Table 1), in this article I focus exclusively on the content portion of the study. Again, this restriction is intended to maintain the focus of this article on presenting the method for examining OTL content variables using Attention-to-Learning Outcomes (ATLO) as the constructed measure.

The classroom data consisted primarily of student artifacts. In each class, two complete sets of student work were collected from two students identified as successful by their high class marks and the teacher. These were thought to contain the most comprehensive representation of the delivered curriculum. The notebooks of the selected students, according to the teacher, best captured in- and out-of-class work. In other words, the teacher identified these students as the best recordkeepers of the delivered curriculum in their classroom.

Analysis of Content Opportunity-to-Learn

The delivered content was described by determining how much ATLO each teacher chose for his or her course. The sets of student work were examined to determine the number of treatments given each intended learning outcome. A treatment was defined as any attention a teacher offered a learning outcome. For example, when a teacher gave notes on an outcome, it was one treatment of that outcome. Any teaching addressing the learning outcome was considered one treatment (e.g., homework assignment, worksheet, in-class activity). The sum of the number of treatments for a learning outcome represented the ATLO in the delivered curriculum.

The student work sets that were corroborated with teacher materials and textbook sections were used to construct the delivered content map. Coursework, once noted as "the most tangible feature of curriculum at the high school level" (McDonnell et al., 1990), was central to describing teacher practices in this study. Each artifact was coded using a scheme generated from the curriculum guides. The content map was compared with classroom observation data to identify the classroom practices reflected in the content of the artifacts collected.

The grade 9 curriculum consisted of four strands subdivided into overall expectations. As in the coding of the data, text indicating specific teaching strategies were removed from the overall expectations to reflect local policymakers' interpretations. In addition, two codes emerged from the document analysis, *Extensions* and *EQAO*. *Extensions* included all mathematical concepts evidenced in the work but absent from the Ministry's curriculum guides. *EQAO* included all treatments specifically preparing the students for the grade 9 provincial assessment. The original coding scheme did not include these two categories because they were not identified as part of the grade 9 curriculum.

To minimize the influence of my interpretation of the work while coding, the headers in teacher notes, textbook titles and subtitles, question wording, and worksheets and assessment labels were used to indicate the curricular expectations in the artifacts. Each set of work from Students A and B in each class was coded separately with a 48-hour wait time between coding sessions. When all the work from both students in the same class was coded, the two sets were compared for congruence. All items found in only one set were assumed to be a part of the course. These few items were input to the database as they were originally coded. The rest of the items that were found in both sets of data were compared for coder reliability.

The data were summarized to the number of times a specific expectation was addressed with respect to instructional strategy. The total number of treatments for an overall expectation was summed, which served as the ATLO. The ATLO was then examined to yield a content map for content coverage, exposure, and emphasis based on the number of treatments addressing a learning outcome.

The *content coverage* was defined by the presence or absence of a treatment of any specific expectation. However, when determining if full coverage was realized, the overall expectations were used to guide the analysis reflecting the organization of the curriculum documents (Ontario Ministry of Education, 1999). Content coverage offered a crude portrait of what concepts were delivered and subsequently contained little constructive information. It was only when content exposure was resolved that the delivery of each course began to come into focus. Although the literature suggests that the decisive indicator for content exposure is time spent on a topic (McDonnell et al., 1990; Winfield, 1987), this was predetermined by a compulsory departmental course outline in this school. Hence mapping the number of lessons or time spent attending to a topic would have failed to capture the variation between classes. However, when the ATLO was used as the unit for establishing content exposure, the differences surfaced. The sum of treatments for each overall expectation was calculated, divided by the total number of treatments for the course, and multiplied by 100 to yield a percent distribution of treatments of all the expectations.

Finally, *content emphasis* was established by determining the distribution of treatments for expectations in a single strand. This is distinct from the distribution of treatments of expectations overall, and allowed for a refined image of which parts of the curriculum were emphasized. It is noteworthy that the method in earlier studies using the number of treatments for a single topic to determine content emphasis (Porter & Associates, 1994) was not used. This decision was based on the belief that any concept may be treated in little depth a number of times using different strategies. Moreover, content emphasis represents the teacher's selection of certain parts of content over others, which may not manifest itself in varying teaching strategies. To ascertain the content the teachers chose to emphasize, each strand was treated as a whole. The treatment distribution was calculated for each strand individually and the two sets of results were compared to expose the choice teachers exhibited in their curriculum delivery.

Results

In this section, I present the results of the delivered content for both teachers. The content coverage is not presented below because it was complete for both teachers, which did not add anything to the findings. Content coverage was established for each of the learning outcomes by presence or absence in the artifacts. Content exposure and emphasis proved to be more refined measures of the delivered curriculum. These constructs are described for both teachers, followed by a comparative analysis of each content variable between the two teachers.

Teacher 1

Teacher 1 exposed the students to all the overall expectations in each strand. As seen in Table 2, the content exposure by strand was covered in the following proportions: 47.7% on Number Sense and Algebra (NSA), 21.9% on Analytical Geometry (AG), 18.1% Measurement and Geometry (MG), and 7.6% on Relationships (RTL).

The content exposure map favoured Number Sense and Algebra (NSA) and disfavoured Relationships (RLT). Disproportionate attention was given to certain overall expectations, as seen in the content emphasis results. The content emphasis results of each strand are presented in Tables 3 and 4 to illustrate the relative percent treatment in each strand.

Almost half the treatments for this course attended to NSA, and almost half the emphasis was on manipulating first-degree polynomial expressions to solve first-degree equations (45.1%). There was minimal emphasis on solving

Table 2
Content Exposure by Strand for Teacher 1

Strand	*Percent of Treatments*
Number Sense and Algebra	47.7
Relationships	7.6
Analytical Geometry	21.9
Measurement and Geometry	18.1
Additional Content	4.7

Table 3
Content Emphasis by Overall Expectation for NSA and RLT for Teacher 1

Overall Expectation	*Relative Percent of Treatments*
NSA1	7.0
NSA2	27.4
NSA3	45.1
NSA4	20.4
RLT1	16.6
RLT2	22.2
RLT3	61.1

Table 4
Content Emphasis by Overall Expectation for AG and MG for Teacher 1

Overall Expectation	*Relative Percent of Treatments*
AG1	3.8
AG2	50.0
AG3	46.1
MG1	14.0
MG2	20.9
MG3	65.1

multi-step problems requiring numerical answers, using a variety of strategies and tools (7.0%). RLT was the least emphasized in the delivered content, representing only 7.6% of the content exposure. The emphasis in this strand demonstrated that describing the connections between various representations was the most prominent expectation (61.1%) addressed. Teacher 1 gave less attention to the other two learning outcomes. Comparing graphs and formulas was emphasized slightly more (22.2%) than determining the relationship between two variables through experimental analysis (16.6%).

Teacher 1 mostly attended to two overall expectations in Analytical Geometry (AG). Attention was almost equally divided between determining the properties of the slope and y-intercept of a linear relation (50.0%), and solving problems using the properties of linear relations (46.1%). There was a minor treatment on establishing the relationship between the form of an equation and the shape of its graph (3.8%). The distribution of treatments in Measurement and Geometry (MG) showed that the formulation of conjectures and generalizations about geometric relationships involving two-dimensional figures was the most emphasized expectation (65.1%). The measurement portion when students solved problems involving surface area and volume (20.9%) and the determination of the optimal value of various measurements (14.0%) were comparatively minimized during content delivery.

In interpreting content emphasis results, the differential emphasis of the strands in the curriculum was taken into consideration. The distribution of treatments in the strands facilitated determining the emphasis given each expectation in respective strands. For a true sense of content emphasis, the prominence of each strand in the whole course was noted. Teacher 1 emphasized the second and third overall expectations in NSA above all others. Then the second and third expectations in AG, and finally the third expectation in MG were stressed in the course. Conversely, the first expectation in each strand was minimized, as well as the second expectation in both RLT and MG.

Teacher 2

Teacher 2 exposed the students to all the overall expectations in each strand. As presented in Table 5, content exposure by strand was covered in the following proportions: 46.7% on NSA, 24.2% on AG, 14.2% on MG, and 8.2% on RTL.

Content exposure had identical trends as that of Teacher 1. The differences in content delivery between Teachers 1 and 2 were exposed with the content emphasis analysis.

Table 5
Content Exposure by Strand for Teacher 2

Strand	*Percent of Treatments*
Number Sense and Algebra	46.4
Relationships	8.1
Analytical Geometry	24.5
Measurement and Geometry	14.2
Additional Content	6.7

Table 6
Content Emphasis by Overall Expectation for NSA and RLT for Teacher 2

Overall Expectation	*Relative Percent of Treatments*
NSA1	24.7
NSA2	21.4
NSA3	39.0
NSA4	14.9
RLT1	22.2
RLT2	40.7
RLT3	37.0

The expectation most emphasized in NSA was solving the first-degree polynomial expressions in order to solve similar equations (39.0%). However, solving multi-step problems requiring numerical answers using a variety of strategies and tools (24.7%), and demonstrating an understanding of the three basic exponent rules as well as applying them to simplify expressions (21.4%) were almost equally attended to by the teachers. Finally, solving problems using algebraic modeling was given the least attention in this strand (14.3%). RLT was the strand least attended to in the course (8.2%). In this strand, determining relationships between two variables by collecting and analyzing data was the least emphasized (22.2%), and the other expectations were almost equally emphasized.

Solving problems using the properties of linear relations (62.5%) was emphasized in AG. The remaining attention was on determining the properties of the slope and y-intercept of a linear relation (28.8%), so that determining the relationship between the form of an equation and the shape of its graph with

Table 7
Content Emphasis by Overall Expectation for AG and MG for Teacher 2

Overall Expectation	*Relative Percent of Treatments*
AG1	8.8
AG2	28.8
AG3	62.5
MG1	2.1
MG2	29.8
MG3	68.1

respect to linearity and nonlinearity (8.8%) was deemphasized. The distribution of treatments in MG demonstrated a significant emphasis on the formulation of conjectures and generalizations about geometric relationships involving two-dimensional figures (68.1%). Solving problems involving the surface area and the volume of three-dimensional objects (29.8%) received substantial attention, whereas determining the optimal value of various measurements was negligibly treated (2.1%).

Comparing Content Exposure

The content maps for the grade 9 courses were created using a proportional model where the percentages of ATLO established were compared. This analysis indicated that the ATLO percentage of each strand was similar for the two teachers.

No striking differences were found between the ATLO percentages of each strand between the two courses. The greatest difference was in the treatment of MG, where Teacher 2 offered 3.9% more attention to it than did Teacher 1. The smallest difference was the EQAO treatment percentage. The insignificant 0.1% difference resulted from the different treatment totals in each course, which affected the calculated percentage treatment although the actual exposure was identical. Although the other strands received slightly different percent treatments, the differences were always less than 3% and were considered insignificant.

The content exposures of the classes were similar because of a locally developed prescribed timeline directing topic coverage for all grade 9 classes in the department. The outline detailed the number of lessons that were to be spent on each topic and the time of year when the topic should be addressed.

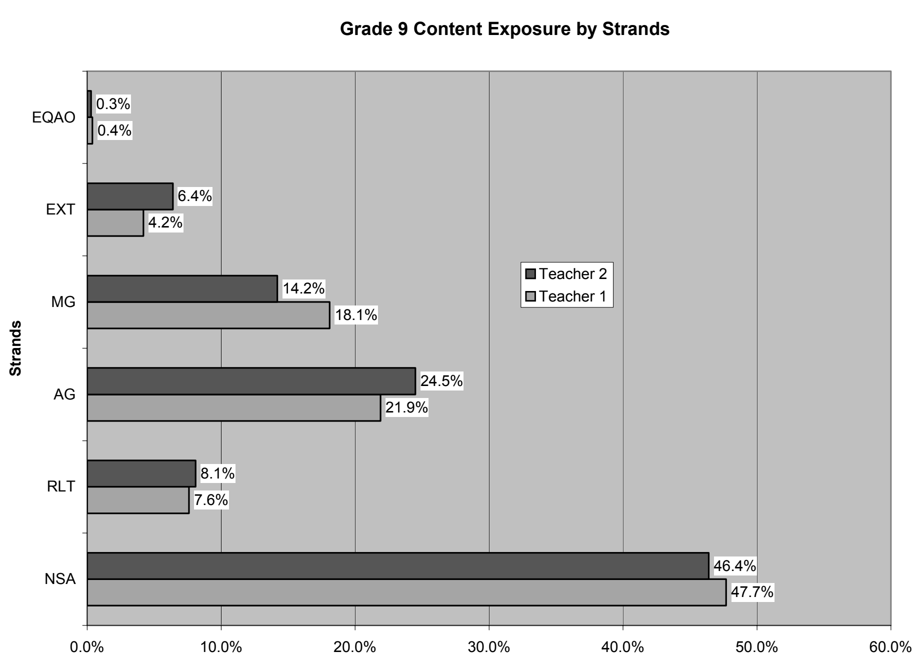

Figure 2. Grade 9 exposure by strand for teachers 1 and 2.

Although there was some flexibility in these guidelines, the course leader consistently communicated in writing and orally with both teachers, reminding them of where they should be on the timeline throughout the year. The finding that local policy standardized content exposure was validated in member check interviews. Both teachers subscribed to maintaining uniform content delivery throughout all the grade 9 classes in the department.

Comparing Content Emphasis

Although the content exposure was similar in both cases, the content emphasis maps for the individual teachers suggested substantial differences in the content delivery of each course. These differences surfaced when the proportions of the treatments in each strand were compared.

The content exposure map indicated a minimal difference in the coverage of NSA between the two courses. However, as Figure 3 illustrates, Teacher 1 minimized the first expectation to represent only 7.0% of the strand's ATLO, whereas Teacher 2 emphasized this expectation so that it represented a quarter of the strand's ATLO. Teacher 2 believed that these numeracy skills were essential for grade 9 students and that the students entered her course with weak algebra skills. The emphasis was a purposeful attempt to elevate their numeracy skills. However, when Teacher 2 decided to emphasize NSA1, she had to reduce attention to the other NSA learning outcomes. Although the difference in the proportional percentage of treatments for the other three expectations did not exceed 6.2% in each case, Teacher 2 had to deemphasize them in comparison with Teacher 1. Noteworthy in these results is that this particular difference surfaced only when content emphasis was examined.

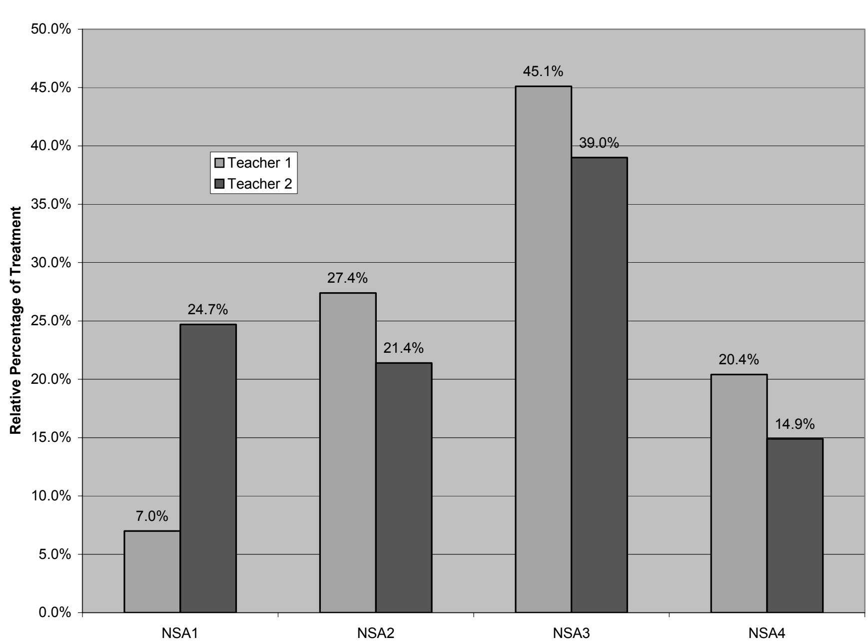

Figure 3. Grade 9 content emphasis: number sense and algebra strand.

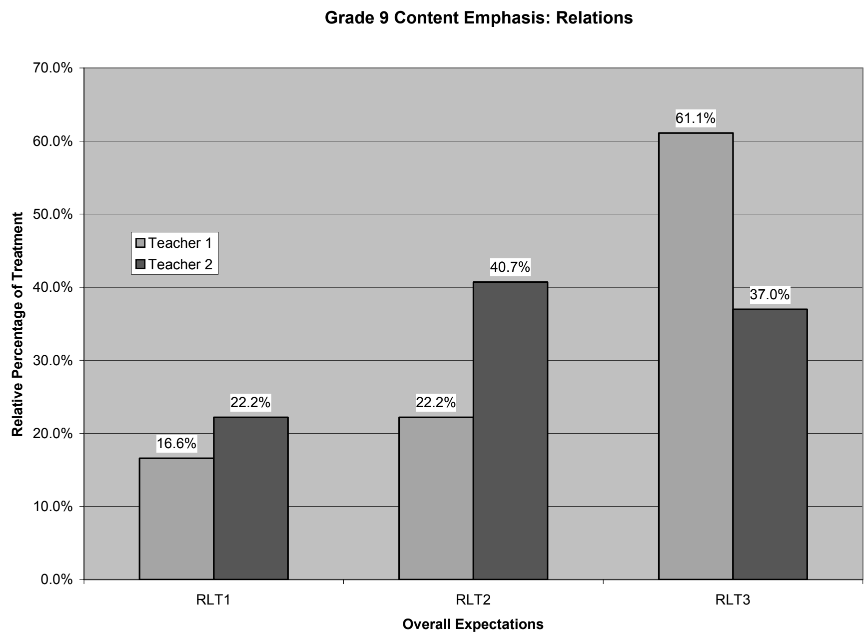

Figure 4. Grade 9 emphasis: relationship strand.

In RLT, which comprised the smallest strand of the content exposure map for both teachers, there were striking differences in the emphasis of each course. The third overall learning outcome received 24.1% more ATLO in the strand in the course offered by Teacher 1 than in that offered by Teacher 2. The increased attention to this learning outcome from Teacher 1 was paralleled with reduced attention to the other two outcomes in the strand as compared with Teacher 2. Although Teacher 2 had not emphasized the third expectation to the same degree as Teacher 1, she did stress the second expectation more than the first. So Teacher 2 attended to the second outcome 18.5% more than did Teacher 1 and gave the first expectation almost the same emphasis. Noteworthy is that both teachers chose to deemphasize the first expectation, which was determining the relationships between two variables by collecting and analyzing data. This may be because of the nature of the expectation that explicitly calls for experimentation where students collect and analyze data. This expectation would have required instructional strategies with which both teachers reported discomfort in implementing due to time constraints (Ben Jaafar, 2002). Noteworthy in the analysis results of the content emphasis for RLT is the spotlight it sheds on shortcomings of content delivery that have the potential to serve as points of discussion with teachers about their professional decisions.

In AG, which represented 21.4% and 24.5% of the content exposure maps for Teacher 1 and 2 respectively, there were differences in the content emphasis. The greatest difference in emphasis was evident in the second expectation where it represented half of Teacher 1's attention in this strand, whereas in the course delivered by Teacher 2, it represented 28.8% of the ATLO. Teacher 2 mostly emphasized the third expectation that received 62.5% of the ATLO in the strand. Although Teacher 1's emphasis of the third expectation was less

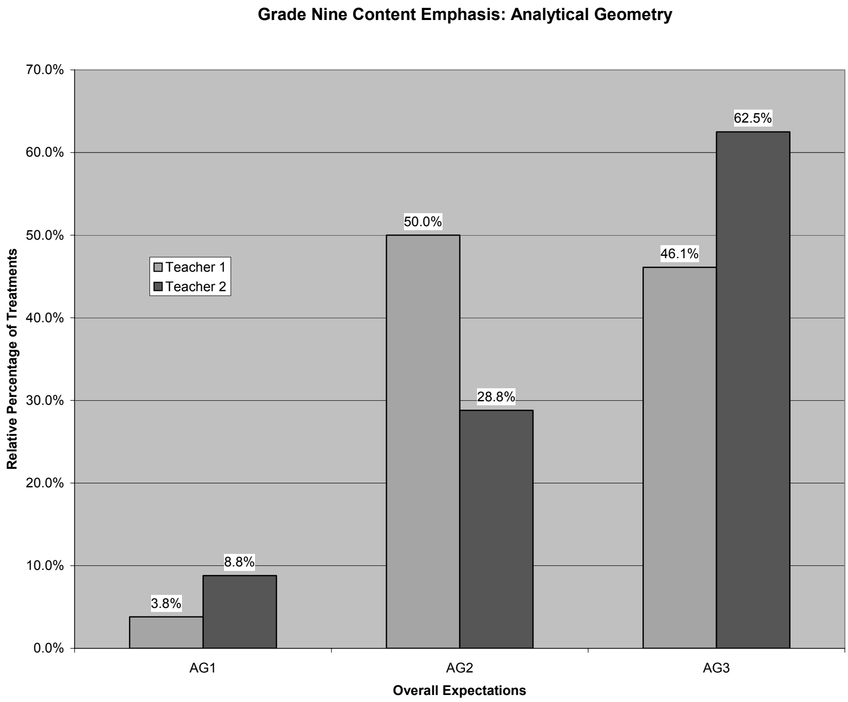

Figure 5. Grade 9 content emphasis for analytical geometry.

than that in the course delivered by Teacher 2, this expectation was treated with almost the same emphasis as the second expectation in that same course. Hence the different emphasis between the two expectations did not alter the trend observed. This case is the same as that of RLT where both teachers opted to minimize the same learning outcome. Although the determination of relationships between the form of an equation and the shape of its graph with respect to linearity and nonlinearity was especially minimized in the course delivered by Teacher 1, it was still comparatively deemphasized in the course delivered by Teacher 2. A potential reason for this choice could be the nature of the expectation. The skills encompassed in this expectation are necessary for the other expectations in the strand. It is possible that although the expectation was not overtly treated in the course, it was embedded in the other two expectations. The coding scheme would not identify the application of these skills if it were not stated in the question, or if it were a necessary element of another expectation.

In MG, which represented 18.1% and 14.2% of the content exposure maps for Teachers 1 and 2 respectively, fewer differences were observed in content emphasis between the two teachers. The third expectation received over 60% of the strand's treatments in both cases. The remaining content emphasis was mostly dedicated to the second strand, although Teacher 2 did give the learning outcomes 8.9% more attention than did Teacher 1.

Finally, the first expectation was minimized in the course that was delivered by Teacher 2, whereas Teacher 1 addressed it so that it represented 14% of the strand's treatments. In fact this was the strand where the teachers' emphasis diverged the most. Despite this difference, content emphasis was most similar

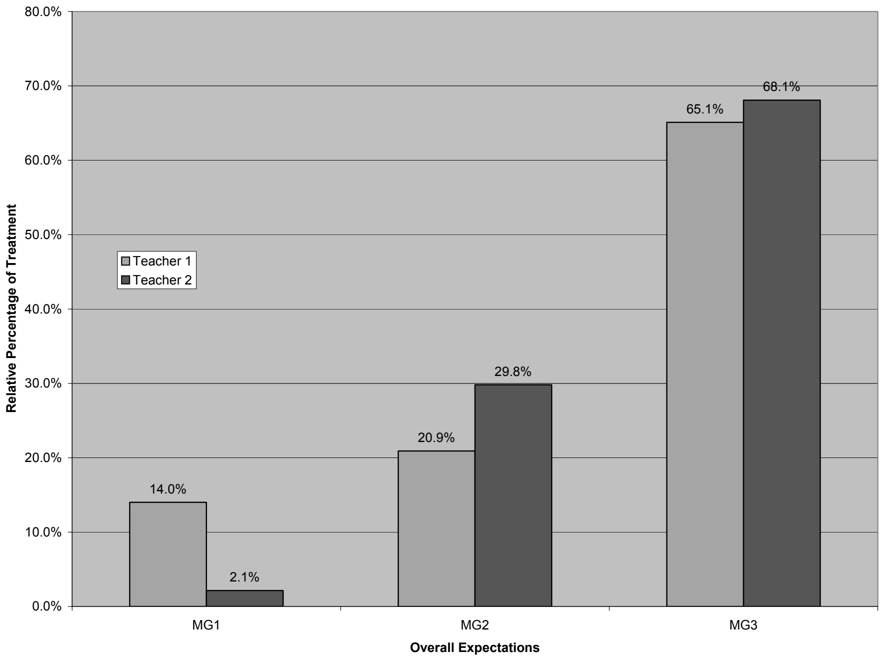

Figure 6. Content emphasis for grade 9 measurement and geometry.

in this strand. This relative uniformity was probably a consequence of the local policy that allocated all teachers and their respective classes time to work in the computer lab for one third of the second semester. In addition, the concepts that were addressed during these lab sessions were specified in the departmental outline and matched the expectations for the third overall expectation of this strand. Although the computer lab allocation was intended to ensure that all the students had the opportunity to use Geometer Sketchpad, it served as an organizational mechanism that standardized the content delivery.

Discussion

The results presented in this study suggest that the use of time as a proxy for measuring content OTL variables may not be sufficiently sensitive to identify differences related to content-delivery in a course. The combination of local policies, departmental culture, and teacher beliefs interacted to produce what would appear to be relatively uniform content delivery across classes. The departmental course syllabus directed all the teachers to spend the same amount of time on given topics, and the teachers felt bound by this unifying timeline.

However, when the number of treatments addressing each learning outcome in the curriculum was examined, the subtle differences embedded in the coursework of each classroom surfaced. This method for measuring the delivered content may address the concerns raised by Burnstein et al. (1995) with respect to the loss of validity in measuring OTL variables using teacher-reported time when curriculum specificity is increased. Using the ATLO approach is potentially more refined than the use of time spent on topics because it is not limited by teacher interpretation of terminology. The tentative language employed in this discussion is purposeful because this case study represents its first use, and more work is needed using this method of testing. However, its use in this study shows promise of its potential. In addition, ATLO could serve to validate the differences found in studies where time is used as a proxy to determine delivered content. Currently no alternative method is available that would allow for time-dependent measures of content OTLs to be validated without intense resource allocation for frequent and consistent observations.

The promise of using ATLO is counterbalanced by its limitations of relying on coursework. First, the margin of error cannot be absolutely established because of the case study nature of this methodology. The source of the coded data was primarily student-derived data, and although precautions were taken to ensure that the whole course was represented in the artifacts selected, it is possible that not all curricular occurrences were documented or preserved. Moreover, the documents collected would not appropriately represent in-depth discussions to understand ideas and challenge misconceptions. The only solution for assuring complete records would be to conduct classroom observations, which raises issues of restricted funding and time when considering larger-scale research. Second, this method is better suited to the secondary level where students are required to keep records of their learning. Unlike those in elementary, high school students are responsible for retaining their notes and work. Third, there is an argument that the student records selected may be biased because the "best" students who keep copious records according to the teacher may not be representative of the classroom population. However, given that the measure is of the delivered content, this problem may be less significant as the "best" record of coursework is desired. Fourth, if an expectation is not overtly addressed in the materials, it is not coded. This crude portion of the approach means that skills embedded in higher-order assignments would not be detected unless explicitly expressed by the teacher or the text. Finally, the time required to employ this kind of methodology might preclude its use in large-scale research without substantial funding.

Given the limitations of this method and the purpose for which it can be used, I submit that Burnstein et al. (1995) offered a means to think about how to use time-consuming methods to validate the measure of indicators in large-scale research. ATLO is not being presented as a stand-alone approach for large-scale examination of the delivered content, nor could it replace the survey data being collected to investigate alignment between curriculum standards, the delivered curriculum, and assessment that is increasing in popularity in the United States (Chief State Schools Officers, 2003). However, it can be used to validate the survey data being analyzed and help to answer questions investigating the degree of difference in implementing standardized curriculum across units of interest. The curriculum documents that are being published are increasingly prescriptive in content. This offers a common reference to identify the differences and congruencies between teachers, schools, and districts. Moreover, as we move to more uniform curricula across provinces, states, and nations, we need methods that traverse geographic divisions and draw on tangible data.

The measure of classroom practices such as delivered content is important in the light of increased calls for educational accountability. In order to account for education, school practices need to be described, explained, and/or justified (Leithwood & Earl, 2000). If these practices are to be elucidated in this manner, then they need to be defined and measured. Moreover, if they are to be measured on a large scale, they need to be quantified, or valid quantifiable proxies need to be identified and developed. Despite efforts and progress, there are problems regarding the delivered content as a means to identify curricular alignment to predict student achievement. Valid representations of classroom content coverage, exposure, and emphasis will not suffice. Teaching encompasses more than delivering content efficiently. The historic notion that teaching can be reduced to the efficient delivery of content (Taylor, 1911) across all classes homogenizes classroom activity and loses sight of differentiated instruction. Teachers need to respond to the needs of the students in their classes in the context of their communities and schools. This adaptation is a fundamental characteristic of the profession that should be reflected in research methods because the results guide evaluative decisions. Hence whether time or ATLO is used, content variables are only one indicator. As we develop other ways to measure classroom practices and attach importance to them through accountability mechanisms in the system, caution should be taken to avoid overemphasis of those indicators that are relatively easier to measure such as content. Before we evaluate teachers on their content delivery, it is important to understand the rationale for their decisions. Previous knowledge of the students in the class, teacher capacity, resources, dependence on homework, and local policies are only a handful of mediating factors that alter decisions and affect content delivery. It is, therefore, imperative to find a balance and refrain from reducing teaching to readily measurable units while continuing to develop valid indicators of classroom practice.

Finally, the precautions of using OTL variables may be respected and the risks diminished if the results of measuring delivered content are used for school improvement efforts. ATLO in particular could provide local data to support teachers in enquiries of their own teaching practices. As seen in the results, teachers emphasized certain learning outcomes over others. The interview data offered explanations for these differences; teachers were responding to their students' needs or their own comfort level with certain teaching approaches. Content delivery findings can provide teachers with information on how well they attended to various parts of the curriculum. It can subsequently serve as a mechanism for them to consider explicitly the rationale for their professional decisions, promoting the reflective practice that improves teaching.

The other advantage of using the ATLO approach is that it does not require teachers to monitor their content delivery throughout the year. This is attractive for the research agenda because there is no requirement from the teacher except to identify those students whom they regard as good recordkeepers in their course. This means that once teachers have made this identification, they are no longer involved in the data collection regarding content delivery. Thus the risk of skewing the results because of heightened awareness when teachers keep records of their practice for researchers is minimized. For the school improvement agenda, this method is attractive because it requires minimal input from teachers who feel time constraints and complain of overload. They do not have to do any additional work to receive a breakdown of how they delivered the curriculum. This advantage extends to teachers using the time they would have spent on data collection to reflect on what the results mean for their teaching and if they feel they can justify their professional decisions.

The potential for the use of ATLO is situated in both research and practice. As such, this is a method that needs to be further examined in both these capacities and validated in larger studies. This modest beginning provides an alternative to using time and offers a choice when measuring OTL content variables.

References

American Educational Research Association. (2000). Position statement of the American Educational Research Association concerning high-stakes testing in pre-K-12 education. *Educational Researcher, 29*(8), 24-25.

Ben Jaafar, S. (2002, October). *The effects of testing on classroom practices*. Poster presented at the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, GA.

Brown, S., & McIntyre, D. (1993). *Making sense of teaching*. Milton Keynes, UK; Philadelphia, PA: Open University Press.

Berstein, L., McDonnell, L.M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators (MR-658-NSF)*. Santa Monica, CA: Rand Corporation.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Carroll, J.B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher, 18*(1), 11-15.

Chief State Schools Officers. (2003, June 9). Surveys of enacted curriculum. Retrieved April 20, 2004, from: http://www.ccsso.org/projects/Surveys_of_Enacted_Curriculum/

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London; New York: Routledge/Falmer.

Education and Quality Accountability Office (EQAO). (1996, February 6, 2003.). *EQAO Madate.* Retrieved February 9, 2003, from: http://www.eqao.com/eqao/home_page/01e/1_2e.html

Fitz-Gibbon, & Kochan, S. (2000). School effectiveness and educational indicators. In D. Reynolds (Ed.), *The international handbook of school effectiveness research* (pp. 257-282). London: Falmer Press.

Hawley, W.D., Rosenholtz, S., Goodstein, H., & Hasselbring, T. (1984). Good schools: What research says about improving student achievement. *Peabody Journal of Education, 61*(4), 1-178.

Husén, T. (1967). *International study of achievement in mathematics: A comparison of twelve countries.* New York: John Wiley and Sons.

Karweit, N.L. (1983). *Time-on-task: A research review* (Report 332). Baltimore, MD: Center for Social Organization of Schools, Johns Hopkins University.

Karweit, N.L., & Slavin, R.E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology, 74*, 844-851.

Lapointe, A.E., Mead, N.A., & Phillips, G.W. (1989). *A world of difference: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Service.

Leithwood, K., & Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education, 75*(4), 1-18.

Linn, R.L., Baker, E.L., & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3-16.

Marsh, C.J., & Willis, G. (1999). *Curriculum: Alternative approaches, ongoing issues* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Mathison, S. (1988). Why triangulation? *Educational Researcher, 17*, 13-17.

McDonnell, L., Berstein, L., Ormseth, T., Catterall, J., & Moody, D. (1990). *Discovering what schools really teach: Designing improved coursework indicators* (Report R3886). Santa Monica, CA: Rand Corporation.

Merriam, S.B. (1998). *Qualitative research and case study applications in education* (2nd ed.). San Francisco, CA: Jossey-Bass.

Murphy, J., & Datnow, A. (2003). The development of comprehensive school reform. In A. Datnow (Ed.), *Leadership lessons from comprehensive school reforms* (pp. 3-18). Thousand Oaks, CA: Corwin Press.
Muthen, B.O., Kao, C., & Berstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28*(1), 1-22.
Nagy, P. (1996). International comparisons of student achievement in mathematics and science: A Canadian perspective. *Canadian Journal of Education, 21*, 396-413.
Ontario Ministry of Education. (1999). *The Ontario curriculum, grades 9 and 10 mathematics*. Toronto, ON: Queen's Printer for Ontario.
Ontario Ministry of Education. (2000). *Curriculum update: Elementary/secondary implementation—Issue 4*. Retrieved November 12, 2001, from: http://www.edu.gov.on.ca/eng/document/curricul/update/winter00/dec00.html#real
Ontario Secondary Schools Teachers' Federation. (2002, September 30). *Grade 10 literacy test results disastrous*. Retrieved January 4, 2003, from: http://www.osstf.on.ca/www/pub/pressrel/sept02_aug03/sept30-102.html
People for Education. (2003, April 24). *Literacy test results show thousands of secondary students continue to struggle* [media release]. Retrieved May 12, 2003, from: http://www.peopleforeducation.com/releases/2003/apr24_03.PDF
Popham, W.J. (1999). Where large scale educational assessment is heading and why it shouldn't. *Educational Measurement: Issues and Practice, 18*(3), 13-17.
Porter, A.C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14.
Porter, A.C., & Associates. (1994). *Reform of high school mathematics and science and opportunity to learn*. CPRE policy brief (Research ED 377 037). Washington, DC: Consortium for Policy Research in Education.
Reynolds, D., Teddlie, C., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In D. Reynolds (Ed.), *The international handbook of school effectiveness research* (pp. 206-231). London: Falmer Press.
Skrla, L., & Scheurich, J.J. (2004). *Educational equity and accountability: Paradigms, policies, and politics*. New York: Routledge Falmer.
Smithson, J.L., & Porter, A.C. (1994). *Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum—The reform up close study* (Research/Technical). Washington, DC: Consortium for Policy Research and Improvement.
Stevens, F.I. (1993). Opportunity to learn and other social contextual issues: Addressing the low academic achievement of African American students. *Journal of Negro Education, 62*(3), 227-231.
Taylor, F.W. (1911). *The principles of scientific management*. New York: London: Harper.
Traub, R.E., & Wolfe, R.G. (1981). Latent trait theories and the assessment of education achievement. In D.C. Berliner (Ed.), *Review of research in education* (pp. 377-435). Washington, DC: American Educational Research Association.
Travers, K.J., Westbury, I., & International Association for the Evaluation of Educational Achievement. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula.* Oxford, UK; New York: Pergamon Press.
Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Education Evaluation and Policy Analysis, 20*(3), 137-156.
Winfield, L.F. (1987). Teachers' estimates of test content covered in first grade reading and achievement. *Elementary School Journal, 87*, 437-454.
Zancanella, D. (1992). The influence of state-mandated testing on teachers of literature. *Educational Evaluation and Policy Analysis, 14*(3), 283-295.