CONTEXTUAL AND SCHOOL FACTORS ASSOCIATED WITH ACHIEVEMENT ON A HIGH-STAKES EXAMINATION

Don A. Klinger, W. Todd Rogers, John O. Anderson, Cheryl Poth, & Ruth Calman

This study identified student and school-level factors associated with student achievement on the Ontario Secondary School Literacy Test (OSSLT), an examination that includes a student questionnaire examining home literacy practices. Linked student and school contextual data enabled the use of hierarchical linear modeling to complete the analyses and examine both student and school level effects. Fourteen student and three school level variables were found to be associated with students' reading achievement and twelve student and two school level variables were associated with writing. Significant variations between schools were also found for students with individualized education plans, English as a second language, or previous eligibility on the OSSLT.

Key words: hierarchical linear modeling, literacy testing, achievement, education factors

Dans cette étude, les auteurs ont identifié les facteurs élèves et niveaux scolaires associés au rendement scolaire dans le Test de compétences linguistiques de l'Ontario, un examen qui comprend un questionnaire pour les élèves au sujet des habitudes en matière de littératie à la maison. Des données reliées aux élèves et au contexte scolaire ont permis l'utilisation d'une modélisation linéaire hiérarchique en vue de compléter les analyses et d'étudier les effets quant aux élèves et aux niveaux scolaires. Les auteurs ont établi que quatorze variables ayant trait aux élèves et trois ayant trait aux niveaux scolaires étaient associées au rendement des élèves en matière de lecture et que douze variables ayant trait aux élèves et deux ayant trait aux niveaux scolaires étaient associées à l'écriture. Des écarts importants entre les écoles ont également été notés chez les élèves ayant des programmes pédagogiques sur mesure, inscrits en anglais langue seconde ou déjà admissibles au Test de compétences linguistiques des écoles secondaires de l'Ontario.

Mots clés : modélisation linéaire hiérarchique, test de compétences linguistiques, rendement scolaire, facteurs pédagogiques

---

Policy frameworks that include large-scale assessments have been constructed to make judgments about what students know or can do, provide appropriate student guidance, and, increasingly, measure school performance (Delandshere, 2001; Ryan, 2002). In Canada, provincial governments use the results to report on such interests and to support interventions intended to improve them. The use of such assessments for the purposes of system accountability has been a more recent trend (Firestone, Mayrowetz, & Fairman, 1998; Ryan, 2002). One expression of the accountability framework is a response to the fear of declining standards combined with a concern for inequalities in students' opportunities to learn (Mazzeo, 2001). A second expression of accountability is associated with responsibility and reform (Ontario Royal Commission on Learning, 1994). Within this framework, educational accountability requires the availability of accessible and interpretable information about program quality and efforts to improve or maintain it. Such educational accountability frameworks are predicated on the belief that assuring quality of student performance is a key priority and that accountability can be used to support educational reforms and monitor and enhance student progress and school performance.

Given the emerging accountability frameworks, and the call for data-based decision-making in education, there is a complementary desire to use large-scale assessment data to provide information to guide instructional and policy decisions leading to system improvement and growth. These data can be a rich source of information that, with the appropriate analyses and interpretation, lead to better understanding of and support for the performance and quality of schools and education in Canada.

There is a definite need to better utilize these existing data sets. However, the relationships of schooling to student achievement are complex. Factors in addition to curriculum and instruction – for example, student characteristics on entry, teacher and school traits, home and community characteristics—have been shown to have significant relationships to student achievement, and these relationships vary across grade level, school, and subject area (Fitz-Gibbon, 1998; Ma, 2001; Mandeville & Anderson, 1987; Rumberger, 1995; Willms & Kerckhoff,

1995). Canada has a long history of collecting information on student achievement of learning outcomes, as well as characteristics of students, schools, and communities; however, the anonymous and/or incomplete nature of the data have resulted in restricted analyses. Consequently, the statistical analyses have often been limited to examining correlations between educational indicators, (e.g., expenditures, school type) and test scores, with inconsistent results. Some researchers (e.g., Greenwald, Hedges & Laine, 1996a, 1996b) claim clear associations exist among various educational indicators whereas others (e.g., Hanushek, 1996) fail to find such associations.

Contextual student variables, also associated with higher achievement, are important to consider when examining student achievement in the context of large-scale assessments (Ma & Klinger, 2000; Taylor & Tubianasa, 2001; Willms, 1992). These include, but are not limited to, foreign language status, learning disabilities, family status, or special needs (e.g., Cummins, 1982; Hamp-Lyons, 1996; Ma & Klinger, 2000; Rogers, Ma, Klinger, Dawber, Hellsten, Nowicki, & Tomkowicz, 2000; Watt & Roessingh, 2001; Willms, 1992). Similarly, researchers have consistently identified indicators of wealth and education, at both the individual and school level, as important contextual variables affecting individual achievement (e.g., Caldas & Bankston, 1997; Raudenbush & Willms, 1995; Sirin, 2005; Willms, 1992). To increase educators' understanding of student and school performance and to support efforts to increase such performance, it is important to include and better understand the associations of such factors with measures of student achievement.

Not surprisingly, researchers have found achievement related variables to be associated with subsequent achievement. Previous achievement, an involvement in activities related to the measure of interest, and access to relevant materials are all associated with higher achievement (e.g., Rogers, et al., 2000, Willms, 1992). For literacy, the quantity and types of reading and writing activities have been linked to higher achievement although the majority of the researchers have examined early literacy development (e.g., Leseman & de Jong, 1998; Payne, Whitehurst, & Angell, 1994; Saracho, 1997; Scarborough, Dobrich, & Hager, 1991). Relevant computer use has also been linked to higher

achievement; however, these associations are much less definitive (e.g., Attewell & Battle, 1999; Miller & McInerney, 1994; Ravitz, Mergendoller, & Rush, 2002). These associations have been found at both the individual student level and the school level. Hence students who attend schools in which more students engage in supporting activities will also be more likely to have higher achievement (e.g., Ma & Klinger, 2000; Willms, 1992). Such research indicates that achievement is related not only to the efforts and activities of individual students but also the efforts and activities of schools and their staffs.

However, it is not clear where research efforts should be directed. Previously used analytic procedures have been unable to address such issues. For example, evidence suggests that the level at which the system traits are aggregated influences the predictability of results. Yair (1997) pointed out that within-school variation in student achievement is generally greater than between-school variation. The between-school variation is further complicated by the nested structure of education because students are nested within classrooms (and teachers) that, in turn, are nested within schools. Traditional statistical procedures can be negatively affected by such nested data because the classroom and school effects differentially impact student performance. These nested data lend themselves well to multi-level or hierarchical linear modeling (HLM). HLM facilitates the examination of relations occurring at each level, across levels (specify how variables at one level influence relations occurring at another), and assesses the amount of variation at each level (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Multi-level analyses offer a useful alternative for nested data because they better account for the non-independence of observations within groups and are more aligned with the tradition of model-based inferences (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

The analyses of data arising from large-scale assessment programs have been limited either by the nature of the examination, the data-collection design, or the lack of contextual student data. Many large-scale assessment programs used in previous research combine a low-stakes assessment program (e.g., PISA, TIMMS, SAIP) with student and, at times, teacher and/or school surveys. Because of the low-stakes of the test, it can be argued that students are less motivated to produce their

best work, teachers are not motivated to encourage maximum student performance, and not all survey items are answered, leading to problems with data quality or missing data. Research using data from these assessment programs is further hampered by data collection methods. Although these assessment programs use a large randomly selected sample of students, the number of students chosen from a particular school is generally quite small. Hence it is difficult to identify specific school effects associated with achievement. In contrast, several educational jurisdictions use high-stakes assessment programs. For example, in Canada, British Columbia, Alberta, and Québec use high-school provincial examination programs to help determine high-school grades and graduation. The majority of the eligible students within a school complete these assessments. However, these assessment programs have not included student, teacher, or principal surveys.

PURPOSE

In contrast to other large-scale testing programs, the Ontario Secondary School Literacy Test (OSSLT) program includes a high-stakes test along with contextual and achievement related student data obtained from the school through the student information form and a brief student survey. As well, potentially relevant school information can be obtained through amalgamation of the student level data and from other sources, for example, the Educational Quality Indicator Framework (EQI). The purpose of the current study was to identify what student and available school-level factors were associated with student achievement on the OSSLT, a high-stakes examination of student literacy. Given the nested structure of the data (students nested within schools), HLM was used to avoid the problems associated with simple correlational analyses and multiple regression techniques.

The Ontario Secondary School Literacy Test

The OSSLT[1] is a graduation requirement for high-school students in Ontario. Implemented in 2002, the OSSLT is the first high-stakes examination in Ontario schools that includes relevant student contextual information and a student survey about student literacy activities. The purpose of the OSSLT is to ensure that students have acquired the

essential reading and writing skills for understanding reading selections and communicating in a variety of writing forms taught up to the end of grade 9 and that apply to all subject areas in the Ontario curriculum (Education Quality and Accountability Office [EQAO], 2005a, 2005c). As stated by EQAO (2005b), the OSSLT determines "whether individual students have met the literacy graduation requirement" (p. 3). In fulfilling its accountability mandate, a student's success on the OSSLT gives "the public confidence that the student has acquired the basic reading and writing skills expected by the end of grade 9" (EQAO 2005c, p. iv). As part of its mandate, EQAO also promotes the use of the OSSLT results to support classroom instruction and raise student achievement.

All students in public and private Ontario secondary schools who are working toward an Ontario Secondary School Diploma are required to successfully complete the OSSLT or the corresponding Ontario Secondary School Literacy Course (OSSLC) prior to graduation (Ontario Ministry of Education, 2003). Students registered in their second year of high school (typically grade 10) are eligible to write the test, although they may be deferred from writing the assessment until a subsequent administration. The OSSLT can be written in either English or French depending on the school board in which the student is enrolled. The test design from its inception in 2002 until the 2004 administration consisted of two 2.5-hour components[2] – reading and writing in a print context.

The following information, associated with the 2003 October administration, was used in the current study.[3] In the reading component, students completed a total of 12 reading selections which were classified as information (e.g., written explanation, opinion; worth 50%), graphic (e.g. graph, schedule, instructions; worth 25%), and narrative (e.g., story, dialogue; worth 25%). Three different question formats were used to measure students' comprehension and understanding of these selections: multiple-choice (MC) (40 questions; worth 40%), questions requiring a short constructed response (CR) (35 questions; worth 35%), and questions requiring a constructed response with an explanation (CRE) (25 questions; worth 25%). In the writing component, students completed four tasks: a summary, a series of paragraphs expressing an opinion, a news report, and an information paragraph. For each task, students were instructed to organize their

ideas, write complete sentences, and use standard spelling, grammar, and punctuation. The test provided students with the purpose and audience for each task and guidelines for the length and structure of their writing samples.

The completed tests were sent to EQAO and centrally marked by trained teachers and markers. The MC and CR items on the reading component were scored on a 2-point (0, 2) scale and the CRE items were scored using item specific scoring rubrics using a 3-point scale (2 points for correct, 1 point for partly correct, or 0 for incorrect). An equated scaled score of 125/200 represented the minimum passing score for reading. Scores between 115 and 124 were rescored. Each writing task was scored using a 4-point (10, 20, 35 and 45) scoring scale having specific performance descriptors. A score of 0 was given for responses classified as "blank/illegible and irrelevant content/off-task." Students who obtained at least 100/180 received a "pass" on the writing component. A separate but single rater scored a student's four writing tasks. However, students who failed the writing test but were within 10 points (e.g., between 90 and 100) of the passing score were rescored by a second set of raters. For these students, the highest set of scores was used for the writing score.

Students who wrote the OSSLT were informed that they had either successfully or unsuccessfully completed the test. Unsuccessful students also received an Individual Student Report (ISR) that provided details about their performance with the various types of reading materials, specific reading skills, and the four writing tasks.

At the same time students completed the OSSLT, they also completed a student survey that asked for information regarding home literacy practices. The survey was designed to be completed with little extra effort or time, using largely dichotomous items.

METHOD

Sample

The sample for the current study consisted of students who completed the English version of the OSSLT in October 2003. Separate analyses were completed for the reading and writing components. Data were obtained for an initial sample of 218,242 students. Of these, 192,457

students completed the English reading and writing components and 180,670 students completed the accompanying survey. List-wise deletion of missing data reduced the student sample to 177,638 students with complete student data. The sample consisted of students who were writing the test for the first time as well as those who either previously failed to successfully complete or were deferred from the assessment. Schools with fewer than 10 students or those schools that did not have relevant EQI data were removed from the sample resulting in a final sample of 160,491 students nested within 611 public and catholic schools (EQI data was not available for private schools).

Data

The data for the analyses came from two sources. Student level achievement scores, demographic information, and literacy activity data came from the OSSLT and accompanying student survey. Through identification codes, it was possible to link individual student OSSLT scores and survey responses. Students were also linked to their individual schools through school identification codes. These codes determined average school values for the student survey items. The EQI database was used to obtain an estimate of the average income of families in each school. This information was based on the 2001 census data using the postal codes of families in the school rather than in the neighborhood around the school. Hence it likely provided a more accurate measure of the average socio-economic status of the students within the schools. The number of first-time eligible students writing the assessments was used as an estimate of school size. The school board names were used to classify schools as either public or catholic schools. Classification of urban or rural location was not available in the datasets and was not included in the analysis. Attempts were made to include an education variable using the EQI database; however, the correlations between the income and the education variable (0.93) indicated that a single indicator would be appropriate. Given that the income variable was most associated with the school population, that variable was used.

DATA ANALYSIS AND RESULTS

Factor Analysis

Student level. To reduce the number of variables included in the analyses, exploratory factor analyses were conducted to identify sets of variables that could be used to form composite factors. Principal component analysis with varimax rotation of the components with eigen-values greater than or equal to one was used in the analysis. Using only those items with single loadings of at least 0.30, two factors were created. The scores for these factors were calculated using the sum of the items within each factor (Morris, 1979). The first factor, entitled creative literacy activities, included four questions from the student survey that asked about literacy related activities that the students did outside of school: read letters (31% of students responding yes), read poetry (41% responding yes), write letters (36% responding yes), and write songs (32% responding yes). The second factor, home materials, included the four questions that asked students to identify which of the following reading/literacy materials they had at home: books (89%), magazines (85%), dictionaries (86%), and newspapers (85%). The reliabilities (Cronbach's α) of these two scales were 0.56 and 0.68, respectively. These factors were used rather than the individual items in the HLM analyses. Because of the lack of clear factor loadings, the other literacy related variables were maintained as discrete predictors in the analyses.

School level. Factor analysis was also used to determine if any of the school average student survey responses could be reduced into interpretable school level factors. Based on this analysis, three factors were created. The first, school average home materials, consisted of four variables: percentage of students reporting having books, magazines, dictionaries, and newspapers at home. The internal consistency (Cronbach's α) of this factor was 0.86. The average creative literacy activities factor also contained four variables: percentage of students who read letters, read poetry, write letters, and write songs outside of school. This factor had an internal consistency of 0.83. The two average internet use variables (read on the internet and reading email) formed a doublet factor, average internet use. This factor had a Cronbach's alpha of 0.91.

Descriptive Results

Table 1 provides the descriptive statistics for the data used in the analyses. For aggregated student variables, the school and student level variables had the same observed means. Consequently, only the standard deviation for schools is reported in the final column of Table 1. In the case of dichotomous variables, the standard deviations are not provided for the student level variables. Unless otherwise noted, the student status variables were dichotomously coded (0, 1), with "1" signifying designation of each variable. The status variables were official provincial designations for specific students as identified by the school. In Ontario, English as Second Language and English Literacy Development (ESL/ELD) learners are combined as a subgroup of recently immigrated students. ESL students can read and write in a first language (other than English) and mostly have had continued schooling before arriving in Canada. ELD students are those who may not read and/or write in their first language (L1) and may have missed years of schooling. They could come from countries where Standard English is the official dialect but where other dialects of English are in common use (Ontario Ministry of Education and Training, 1999).

Students with an individual education plan (IEP) due to special needs or required learning support or with an identification placement and review committee (IPRC) designation other than gifted were classified under the category of IEP for the analyses. These students included those with learning disabilities, physical disabilities, or behavioural problems. Gifted students were identified via the IPRC and given an "IPRC-gifted" designation. Finally, students who had previously been deferred from writing the OSSLT, were absent, or unsuccessful on a previous attempt were considered "previously eligible." Two other variables that described student status related to home language were included in the student survey questionnaire. In the first question, students were asked to identify if the first language they learned at home was English or not. These students were considered to be those for whom English was a foreign language (EFL). The second question asked students to describe the language they spoke at home, only or mostly English, another language as often as English, or

Table 1: Descriptive Statistics for Study Variables and Factors

	Students (n = 160,491)	Schools (n = 611)
Reading score	143.33 (31.75)	
Writing score	116.02 (41.19)	
Student Status (% of students)		
Gender (female = 1)	48	(9)
ESL status	3	(6)
IEP status	15	(15)
IPRC-gifted	1	(2)
Previously eligible	26	(15)
English as a foreign language (EFL)	22	(20)
Foreign home language	23	(20)
Home Literacy Survey Items (% of students)		
Read comics	32	(6)
Read non fiction	24	(5)
Read on the internet	77	(8)
Read novels	52	(9)
Read newspapers	46	(8)
Read manuals	21	(4)
Read magazines	73	(6)
Read religious materials	14	(6)
Writing email	88	(7)
Writing notes	34	(5)
Writing stories	17	(6)
Writing work	36	(6)
Scaled Survey Items and Factors		
Reading hours per week	2.15 (1.04)	(0.18)
Writing hours per week	1.99(1.01)	(0.12)
Home computer use for school work.	3.74 (1.08)	(0.35)
Home literacy materials	3.39 (1.01)	(0.23)
Creative literacy activities	1.42 (1.34)	(0.20)
Variables and Factors reported at the school level only		
Number of Students writing the		263 (121)

OSSLT	
School Type	24
Students' average Internet use	1.62 (0.13)
Mean Family Income	59341 (11776)

Notes: The values in parentheses are standard deviations.

The % of students and average school % of students are equal.

only or mostly another language. Because of the small percentage of students choosing options 2 or 3, 14.4 per cent and 6.9 per cent respectively, this "Foreign home language" variable was transformed into a dichotomous variable (0, 1) in which a score of 1 indicated that the language commonly spoken at home was not English.

The student home literacy dichotomous survey variables (0, 1) were coded such that a "no" response was given a score of zero and a "yes" response was given a score of one. Lastly, two of the three scaled survey times – reading hours per week and writing hours per week – were measured on a 4-point scale from 1 to 4, while the third – home computer use for school work – was measured on a 5-point scale (1 to 5). The last two student variables were the two factors identified through factor analysis.

Comparison of the samples used in the present study to the overall student sample indicated that the reduced, analyzed student sample was very similar to the actual full sample. The reading and writing scores were slightly higher in the analyzed sample (less than 2 score points) and the percentages of ESL and IEP students were slightly lower. All four of these variables also had slightly smaller standard deviations in the final sample.

The four remaining variables reported at the bottom of Table 1 were collected solely at the school level. The number of students within a school who were eligible for the first time and completed the OSSLT during the 2003 administration of the test is listed first. The second variable, "school type" was used to distinguish between public and catholic schools, both of which are fully funded in Ontario. A total of 24 per cent of the schools included in the analyses were considered to be catholic schools.

HLM Analysis

Before the analyses of the set of student and school level variables, a "null" model was examined containing no explanatory variables. This model indicates the amount of variability at each level of analysis. With respect to reading, 86.5 per cent of the variability in reading scores was found to exist between students and just over 13.5 per cent of the variability in reading scores occurred between schools. The between student variability was even higher for writing scores (90.3%), with the variability between schools being 9.7 per cent. Because class effects were not measured, it was not possible to separate student and class effects.

HLM models were then examined to determine which constellation of student and school level variables best predicted reading and writing achievement on the OSSLT. All the Level 1 variables identified in Table 1 were included in the HLM analyses. Student level variables were also examined to determine if the effects were constant across schools (fixed slope effects) or different across schools (identified by HLM as random [differential] slope effects). Fixed slope effects indicated a relationship between a student level explanatory variable and achievement but the relationship was constant within and across schools. Student level explanatory variables were not fixed in the analyses if the between school variances (*tau)* because the variables were considered potentially meaningful based on the criterion that the variance was significant ($p < 0.01$) and greater than 0.05. Such variables do not have consistent associations with the dependent variable between schools. Differential (random) school effects can be important from a research and policy perspective because they demonstrate that an association may be susceptible to modification among schools. For student level variables having differential school effects, school level variables may help explain why a "gap" is present in some schools but not in others. Although generally providing a superior explanatory model, the disadvantage of the inclusion of such differential effects is the difficulty in determining the amount of variance accounted for in the final model. However, in the presence of differential effects, Snijders and Bosker (1999) note that an estimate of the accounted variance can be found by running the final model as a fixed model. All non-dichotomous student scores were transformed into *z*-scores for the purposes of HLM analyses.

Consequently, because all the scores were on a 0/1 metric for the dichotomous variables and a *z*-score for non-dichotomous variables, the coefficients indicated the relative change in the dependent variable that was associated with a unit change in the explanatory (independent) variable.

School level variables were examined after including the set of significant student level variables. School level variables indicated the impact of the average school effect on students' dependent variable scores within that school after accounting for relevant student factors. In the case of student level variables having differential (random) slopes, significant school level variables were those that accounted for a portion of the identified variance in the relationship between a student-level explanatory variable and the dependent variable across schools. Because school effects provide a measure of the overall impact across all students in a school, they are generally smaller than student-level effects. With the exception of reading novels, the individual dichotomous survey items aggregated to the school level were not included in the HLM analyses. This decision was made because of the low variability of these variables at the school level and their lack of impact at Level 1 in the presence of other student level explanatory variables. All school level variables were transformed into *z*-scores. Hence the coefficients indicated the relative change in the average achievement score for a school (intercept) for every standard deviation change in the explanatory variable.

The results of the HLM analyses are provided in Table 2 for the student-level variables and Table 3 for the school-level variables. Only those variables that were significantly associated with achievement and had a coefficient of at least 0.05 in absolute value were included in the final model. This would eliminate any variables that may have a slight association with student achievement but whose effect would be too small to have a meaningful impact on student achievement. The results for reading are presented in the second column of each table and the results for writing are presented in the third column. The coefficients reported for the predictive variables included in the model are adjusted for the presence of the other predictive variables included in the model. For example, the coefficient for ESL status for reading achievement, -0.43, has been adjusted for the other predictor variables included in the

model. This value suggests that after controlling for the remaining variables, ESL students on average had a reading score approximately 0.43 of a standard deviation below the mean reading score for the full sample of students.

Table 2: Coefficients for the Student Level Variables

Student Variables	Reading	Writing
Intercept	0.02 (<0.01)	0.08 (<0.01)
Student Status		
Gender	0.05 (<0.01)	0.09 (<0.01)
ESL status	-0.43 (0.02)	-0.07 (0.03)
IEP status	-0.40 (0.01)	-0.28 (0.01)
IPRC-gifted	0.52 (0.01)	0.37 (0.02)
Previously Eligible	-0.51 (0.01)	-0.97 (0.02)
EFL	-0.05 (<0.01)	
Foreign home language	-0.12 (<0.01)	-0.07 (<0.01)
Home Literacy		
Read on the internet	0.16 (<0.01)	0.10 (<0.01)
Read novels	0.18 (<0.01)	0.13 (<0.01)
Scaled Survey Items and Factors		
Reading hours/week	0.13 (<0.01)	0.09 (<0.01)
Home computer use for school work.	0.05 (<0.01)	
Home materials	0.12 (<0.01)	0.07 (<0.01)
Creative literacy activities	-0.08 (<0.01)	-0.06 (<0.01)

Note: The numbers in parentheses represent the standard error of estimate.

Reading results. As shown in Table 2, the seven student status variables, two of the 12 home literacy survey items, and four of the five scaled survey items/factors were included in the model for reading at the student level. With one exception, the associations between these variables and reading achievement were in the expected direction. Four of the seven variables included in the set of student status variables had the largest relative associations with reading scores. The relatively large

negative associations of ESL status, IEP status, and previously eligible status with reading achievement indicated that these students' reading scores were, respectively, on average 0.43, 0.40, and 0.51 of a standard deviation lower than the average reading score. In contrast, IPRC-gifted students had reading scores that were approximately a half of a standard deviation (0.52) higher than average reading score. The associations of the remaining two student status variables were small (-0.05 and –0.12). Two dichotomously scored home literacy variables, reading on the internet and reading novels, had relatively small but positive associations with reading achievement, 0.16 and 0.18 respectively. The associations of the four significant scaled variables were relatively small. Of these variables, the number of reading hours per week and home literacy materials had the highest positive associations, 0.12 and 0.13, respectively. Interestingly, the association between engagement in creative literacy activities was negatively associated (-0.08) with reading test performance.

The school level results for the reading model are presented in the second column of Table 3. Panel A contains the school-level relationships with the school mean (intercept). Panel B contains the results for the three student-status variables – ESL (*tau* = 0.11), IEP (*tau* = 0.06), and previously eligible (*tau* = 0.09) – that had significant between school variance, suggesting that the relative differences in reading scores obtained by these designated students as compared to the other students in the school were not consistent across schools. That is, the gap between the average performance of ESL and non-ESL students was larger in some schools than in others. Similarly, these gaps varied across schools for IEP and previously eligible students. Hence relative differences in reading performance could be attributed, at least in part, to the schools that these students attended.

Three Level 2 variables were found to have significant associations with reading achievement in the final model (Panel A, Table 3). The coefficients were all weak, ranging from 0.05 (average reading hours per week) to 0.12 (average home literacy materials). As the average use of home literacy materials, hours spent reading, and use of the internet increased, school reading performance increased.

Table 3: Coefficients for the School Effects

Variable	Reading	Writing
Panel A: School Mean		
Average reading hrs/wk	0.05 (<0.01)	
Average home computer use for school work		0.10 (0.01)
Average home materials	0.12 (0.01)	0.09 (<0.01)
Average Internet use	0.06 (<0.01)	
Panel B: Differential Slopes		
ESL slope		
School IEP	0.11 (0.04)	
Number of students	-0.07 (0.02)	-0.06 (0.02)
Previously eligible		0.05 (0.02)
IEP slope		
Previously Eligible slope		
School IEP	0.05 (0.01)	
Average home materials	-0.09 (0.01)	-0.10 (0.02)

Note: The numbers in parentheses represent the standard error of estimate.

As noted above, three student variables were found to have significant between school differences: ESL status, IEP status, and previously eligible status. The average coefficients (slope) for these variables were negative (see Column 2, Table 2). Thus students with these designations had lower reading scores than students without these designations. As shown in Panel B, Table 3, the coefficients were again all weak ($<|0.12|$). Analyses were then conducted to see if the school level variables considered in the present study accounted for these school differences. In the case of the differential ESL school effect (slope), the negative coefficient (-0.07) for number of students indicates that the difference (gap) in reading scores between ESL and non-ESL students was larger in larger schools than in smaller schools. This suggests that ESL students had relatively poorer success in larger schools than in smaller schools. The positive school IEP coefficient (0.11) indicates that the scores of ESL and non-ESL students were closer in schools having more IEP students than in schools having fewer IEP students. None of

the school level variables was found to be associated with the differential IEP slope. Two school-level variables were found to affect the previously eligible slope. Schools with more IEP students had more similar scores between first time and previously eligible students than schools with fewer IEP students. In contrast, the gap between first time writers and previously eligible students was larger in schools in which students had, on average, more home literacy materials.

Writing results. As shown in Table 2, six student status variables, two of the 12 home literacy survey items, and three of the five scaled survey items/factors were included in the model for writing at the student level. With two exceptions, the values of the coefficients for writing were less than the corresponding values of the coefficients for reading. The one large exception was previously eligible status, which had a much stronger negative association with writing scores (-0.97) than that found for reading scores (-0.51). However, like reading, the associations between the student variables and writing performance were in the expected direction with the same one exception (creative literacy activities). Similarly, three of the student-status variables had the largest associations with reading scores. The relatively large negative associations of IEP status and, especially, previously eligible status with writing performance indicated that these students' writing scores were, respectively, on average 0.28 and 0.97 of a standard deviation lower than the average writing score. In contrast, IPRC-gifted students had writing scores that were approximately 0.37 of a standard deviation higher than the average writing score. The associations of the remaining four student-status variables and the scaled items/factors associated with literacy activities were smaller ($< \left| 0.10 \right|$). Lastly, as pointed out, creative literacy activities had a small negative association with writing performance.

The school level results for the writing model are presented in the third column of Table 3. Panel A contains the school-level relationships with the mean writing achievement in schools (intercept). Two Level 2 variables – average home computer use for school work (0.10) and average home literacy materials (0.09) – were significantly associated with writing achievement in the final model (Panel B, Table 3). As the average student use of a home computer for school work and the

average student use of home literacy materials increased, the average writing performance in schools increased.

Panel B contains the results for the three student status variables – ESL (*tau* = 0.09), IEP (*tau* = 0.06), and previously eligible (*tau* = 0.31) – that had significant between school variance, suggesting that the differences in writing scores obtained by these designated students could be attributed, at least in part, to the schools that they attended. The same three student variables found for reading had significant between school differences for writing. In the case of the ESL slope, the negative coefficient (-0.06) for number of students indicated that the difference (gap) in writing scores between ESL and non-ESL students was larger in bigger schools than in smaller schools. Again, as for reading, this finding suggests that ESL students had relatively poorer success in larger schools than in smaller schools. Again, none of the school-level variables was found to be associated with the IEP slope. Lastly, the gap between first time writers and previously eligible students was larger in schools in which students had, on average, more home literacy materials than in schools in which the students had, on average, fewer home literacy materials.

Explained variability. The proportion of the student and school variability explained by the set of variables included at the two levels of the model for reading and the model for writing were calculated. First, as noted previously, most of the initial variation in achievement was among students within classes: 86.5 per cent for reading and 90.0 per cent for writing. The initial variation at the school level was 13.5 per cent for reading and 10.0 per cent for writing. Not unexpectedly, the amounts of variance to be accounted for at the student level are considerably greater than the amounts to be explained at the school level (see also Rogers, Wentzel, & Ndalichako, 1997; Yair, 1997). The 13 student-level and 8 school-level variables retained in the final model for reading accounted for 24.3 per cent of the initial variability among students and 87.0 per cent of the initial variability among schools. The corresponding percentages for writing were 27.7 per cent and 76.3 per cent. Although the variances at the two levels of the final models for reading and writing were considerably less than the initial values found for both

subjects, there was still unexplained variability, particularly at the student level.

DISCUSSION

The Ontario Secondary Schools Literacy Test (OSSLT) provides a unique opportunity to investigate the associations between student- and school-level factors with achievement as measured on a high-stakes examination. To the extent that high-stakes examinations, in this case a graduation requirement, result in increased student effort and more accurate achievement scores, the results of the OSSLT can be viewed as better indicators of achievement than data obtained from low-stakes examinations. The results of the analyses provide an indication of relevant student- and school-level variables that are associated with reading and writing achievement as measured by a high-stakes achievement test. Through the identification of such associations, it becomes possible to examine policy implications.

As found in previous research, the majority of the variability for both the reading and writing achievement scores was between students rather than between schools (e.g., Ma & Klinger, 2001; Raudenbush & Bryk, 2002; Rogers, et al., 2000; Rogers, Wentzel, & Ndalichako, 1997; Yair, 1997). Several student-level variables were associated with higher achievement, but the amount of unexplained student-level variance remained substantial. Less than 30 per cent of the student-level variability in achievement was accounted for by a set of available student variables. Student status (ESL, IEP, gifted, and previously eligible) had the largest associations with the achievement variables. Literacy related variables had small associations with both reading and writing. Such findings are discouraging from a policy perspective because they indicate that student contextual variables continue to have the most identified influence on achievement. However, because the majority of student-level variability was not accounted for in the final models, the results also indicate that these contextual variables are not sufficient to explain the variability in student performance. As noted previously, with one exception, the associations tended to be smaller between the student variables and writing performance than between the student variables and reading performance. The one notable exception was the previously

eligible variable that had a much stronger negative association with writing performance than with reading performance (-0.97 vs. -0.51). This suggests that previously eligible students had relatively more difficulty with writing than with reading as measured by the 2003 OSSLT. It would be worthwhile to examine if this trend continues across future administrations.

Although the relatively low reliability of the creative literacy activities factor indicates caution, the small but consistent negative associations between creative literacy activities and reading and writing are also worthy of further study. This finding may suggest that students who focus more of their attention on engaging in more creative literacy activities may have more difficulty on the literacy examination. It is possible that these students represent a unique sub-group who do not perform well on a traditional measure of literacy.

Although most of the variability in achievement was found to be between students rather than between schools, school-level associations were found. For example, schools with students who made more use of the computer at home to assist with their school work or had increased amounts of literacy related home materials also had students with higher writing scores. Of interest was the lack of impact of the school-level SES measure on achievement. In the presence of other school-level literacy factors and demographic variables, SES did not appear to have a significant association with either reading or writing performance. It may be that the presence of variables such as average home computer use or the average home materials may have diminished the association between SES and achievement (see Rogers et al., 2000). However, the SES variable did not have a large association with either reading or writing performance when considered alone. This latter finding is somewhat surprising, and may be attributed to the way SES was measured. Unfortunately, direct and indirect student-level SES information was not available, thereby making it more difficult to determine the impact of this factor on achievement on this high-stakes literacy test. Lastly, the finding that the relationship of ESL status, IEP status, and previously eligible status with reading and with writing depended on the school the students attended deserves further examination. What factors other than

school size and percentages of special status students can account for the differential levels of association?

The results of the present study are somewhat disappointing in that the amount of variance explained was not large, particularly at the student level where most of the variability resides. Part of the problem in the present study may be due to the dichotomous nature of the student survey data, which served to constrain the relationships modeled. Further, the relatively small number of items used in the survey coupled with the "yes" or "no" response format made it difficult to identify stable relevant factors. The moderately low reliabilities of the two student-level factors, creative literacy activities and home literacy materials, illustrate this problem. A second problem may be the inability to include class as a level in the multi-level model. Rogers et al. (2000), for example, found that the variability among classes was greater than the variability among schools. The OSSLT is designed to represent literacy achievement across subject areas and is not viewed as being connected to specific courses or teachers. Hence it may be argued that teacher-level effects are not relevant. However, because data are collected only at the student and the school levels, it is not possible to determine any significant class-/teacher-level effects associated with student achievement. A third problem, which is related to the second, is the absence of other variables that may influence performance: internal and external funding, school organizational structure and operation, school leadership, staff qualifications, nature of instructional support, quality of professional development and professional development opportunities, parent involvement and support, and, importantly, classroom processes. Inclusion of these variables may very well account for a substantial proportion of the variance in achievement measures.

Given the similarity in the results of the current study with other previous studies, the information obtained through the use of high-stakes examinations may not be substantially different than that obtained using low-stakes test programs (e.g., Ma & Klinger, 2001; Rogers et al, 2000). The amount of score variability was not all that different from that found in other research in which low-stakes examinations were used. Further research is needed to determine if similar results occur across years of the examination program and with

similar assessment programs having lower associated test consequences for students. Such studies should provide evidence regarding the associated stakes of an assessment program required to obtain accurate associations between context, practice, and achievement.

CONCLUSION

The quality of public education is a key contributor to the well being of Canadian society. The relationships between school system traits and schooling outcomes are of basic interest and significance to the educational policy community and educational stakeholders. The foundation for this interest lies in examining those variables and factors that can be influenced by policy makers and influence the key outcomes of schooling (Kennedy, 1999a). The most commonly used indicator of educational performance will continue to be student achievement on centrally administered assessments of learning outcomes. It is equally important to collect other relevant information that is associated with this achievement if learning is to be enhanced and the current gaps between sub-groups of students closed. Unfortunately, the hope for explanatory models of complex systems such as education remains elusive (Lindblom, 1990). At best, such models are likely to be enlightening, allowing incrementally increasing understandings of such dynamic systems (Kennedy, 1999b). For these models to expand understandings of educational systems, the analyses and models must be school-relevant and conducted over the long term. It seems unlikely that one model will suffice. Thus the research most appropriate to this field has to be of a complex, dynamic, and sustainable nature, and findings have to be fully communicated and critically analyzed by the policy community to feed further iterations of research. The findings of the present study, although limited, illustrate the analyses needed and point to the need to expand the nature of influential variables in the data collection design.

ACKNOWLEDGEMENT

The authors thank the Education Quality and Accountability Office for support of the research.

NOTES

[1] Students who have been unsuccessful or have been deferred from writing the OSSLT may complete the Ontario Secondary School Literacy Course (OSSLC) to meet this graduation requirement.

[2] In 2005 - 2006, the two components were combined to form a single literacy score. The test is now administered in late March.

[3] The format and procedures associated with the OSSLT have changed slightly over time.

REFERENCES

Attewell, P., & Battle, J. (1999). Home computers and school performance. *Information Society,* 15(1), 1-10.

Caldas, S. J., & Bankston, C., III (1997). Effects of school population socioeconomic status on individual academic achievement. *Journal of Educational Research,* 90(1), 269-277.

Cummins, J. (1982). *Tests, achievement, and bilingual students.* Arlington, VA: National Clearinghouse for Bilingual Education.

Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education*, 8(2), 113-133.

Education Quality and Accountability Office. (June, 2005). *Educator home page*. Retrieved January 12, 2006, from http://www.eqao.com/Educators/educator.aspx?Lang=E

Education Quality and Accountability Office. (November, 2005). *Ontario Secondary School Literacy Test: Framework.* Retrieved January 12, 2006, from http://www.eqao.com/pdf_e/05/05P044e.pdf

Education Quality and Accountability Office. (2005). *Ontario Secondary School Literacy Test 2004-2005: Provincial Report.* Retrieved January 12, 2006, from http://www.eqao.com/pdf_e/05/05P018e.pdf

Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis,* 20(2), 95-113.

Fitz-Gibbon, C. (1998). Indicator systems for schools: Fire fighting it is! In D. Shorrocks-Taylor (Ed.), *Directions in Educational Psychology* (pp. 189-205). London, UK: Whurr Publications Ltd.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996a). The effect of school resources on student achievement. *Review of Educational Research,* 66, 361-396.

Greenwald, R., Hedges, L. V. & Laine, R. D. (1996b). Interpreting research on school resources on student achievement: A rejoinder to Hanushek. *Review of Educational Research*, 66, 411-415.

Hamp-Lyons, L. (1996). The challenges of second language writing assessment. In E. White, W. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing: Policies, politics, practice* (pp. 226-240). New York: Modern Language Association.

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66, 397-410.

Kennedy, M. M. (1999a). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis*, 21(4), 345-363.

Kennedy, M. M. (1999b). Infusing educational decision making with research. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 53-79). New York: Academic Press.

Leseman, P. P. M., & de Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation and social–emotional quality predicting early reading achievement. *Reading Research Quarterly*, 33(3), 294–318.

Lindblom, C. E. (1990). *Inquiry and change: The troubled attempt to understand and shape society*. New Haven, CT, & New York: Yale University Press & Russell Sage Foundation.

Ma, X. (2001). Stability of school academic performance across subject areas. *Journal of Educational Measurement*, 38(1), 1-18.

Ma, X., & Klinger, D. A. (2000). Hierarchical linear modeling of student and school effects on academic achievement. *Canadian Journal of Education*, 24(1), 41-55.

Mandeville, G. K., & Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. *Journal of Educational Measurement*, 24(3), 203-216.

Mazzeo, C. (2001) Frameworks of state: Assessment policy in historical perspective. *Teachers College Record,* 103(3), 367-397.

Miller, M., & McInerney, W. (1994). Effects of a home/school computer project. *Journal of Research on Computing in Education*, 27(2), 198-210.

Morris, J. D. (1979). A comparison of regression prediction accuracy on several types of factor scores. *American Educational Research Journal*, 16, 17-24.

Ontario Royal Commission on Learning. (1994). *For the love of learning*. Toronto: Queen's Printer for Ontario.

Ontario Ministry of Education (2003). *The Ontario Curriculum. English: The Ontario secondary school literacy course, grade 12.* Retrieved May 15, 2005, from www.edu.gov.on.ca/eng/document/curricul/eng12.pdf

Ontario Ministry of Education and Training. (1999). *The Ontario Curriculum grade 9 to 12 English as a second language and English literacy development*. Toronto: Queen's Printer for Ontario.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, 9(3-4), 427–440.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307-335.

Ravitz, J., Mergendoller, J., & Rush, W. (2003 April). *What's school got to do with it? Cautionary tales about correlations between student computer use and academic achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Rogers, W. T., Wentzel, C., & Ndalichako, J. (1997). *Examination of the influence of selected factors on performance on Alberta Education Achievement Tests within Edmonton Public Schools* (Tech. Rep.). Edmonton, AB: University of Alberta, Centre for Research in Applied Measurement and Evaluation.

Rogers, W. T., Ma, X., Klinger, D. A., Dawber, T. E., Hellsten, L. M., Nowicki, D., & Tomkowicz, J. (2000 April). *Examination of the influence of selected factors on the academic and affective performance of grade 6 students.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Ryan, K. (2002). Assessment validation in the context of high-stakes assessment. *Educational Measurement: Issues and Practice*, 21(1), 7-15.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32, 583-625.

Saracho, O. N. (1997). Perspectives on family literacy. *Early Child Development and Care, 127*–128, 3-11.

Scarborough, H. S., Dobrich, W., & Hager, M. (1991). Preschool literacy experience and later reading achievement. *Journal of Learning Disabilities,* 24, 508–511.

Sirin, S. R. (2005). Socioeconomic status and achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.

Snijders, T. A. B., & Bosker, R. J., (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London, UK: Sage

Taylor, A. R., & Tubianasa, T. (2001). *Student assessment in Canada: Improving the learning environment through effective evaluation.* Kelowna, BC: Society for the Advancement of Excellence in Education.

Watt, D., & Roessingh, H. (2001). The dynamics of ESL drop-out: Plus ça change… *The Canadian Modern Language Review,* 58, 203-22.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators.* Washington, DC: The Falmer Press.

Willms, J. D. & Kerckhoff, A. C. (1995). The challenge of developing new educational indicators. *Educational Evaluation and Policy Analysis*, 17(1), 113-131.

Yair, G. (1997). When classrooms matter: Implications of between classroom variability for educational policy in Israel. *Assessment in Education,* 4(2), 225-249.