Investigating Math Achievement Patterns over Time Among Ontario Elementary School Students with Different Language and Literacy Characteristics

Ellen Fesseha
Ontario Institute for Studies in Education, University of Toronto

Hanna Wickstrom
Ontario Institute for Studies in Education, University of Toronto

Eunice Eunhee Jang
Ontario Institute for Studies in Education, University of Toronto

Canadian Journal of Education / Revue canadienne de l'éducation 43:2 (2020)

www.cje-rce.ca

Abstract

Downward trends in Ontario's math achievement have raised concerns about the need for effective identification of students struggling to meet math achievement standards. This study identified latent profiles from math achievement patterns of students from Grades 3 to 6, examining how specific language and literacy characteristics predicted profile membership. Participants' test scores were collected from longitudinal cohort data of provincial math assessments. Latent class analysis identified two achievement pattern profiles: consistent and declining achievement groups. Subsequent logistic regression analyses revealed English Language Learners' (ELL) as likelier to decline in math achievement, suggesting the need for greater instructional support for ELL students.

Keywords: math achievement patterns, English Language Learners, standardized assessments, test-driven accountability, latent class analysis

Résumé

La tendance à la baisse des résultats en mathématiques en Ontario soulève des inquiétudes quant à la nécessité d'identifier efficacement les élèves ayant du mal à atteindre les standards de réussite en mathématiques. Cette étude identifie des profils latents, à partir de tendances de réussite en mathématiques d'étudiants de troisième à sixième année, en examinant comment des caractéristiques spécifiques de langage et d'alphabétisation prédisent l'appartenance à un profil. Les données analysées proviennent des résultats longitudinaux obtenus par les cohortes de participants aux évaluations provinciales en mathématiques. L'analyse des classes latentes a permis d'identifier deux profils de tendance de la réussite : un groupe dont les résultats sont constants et un autre dont les résultats sont en déclin. Les analyses de régression logistique ultérieures ont révélé que les apprenants de langue anglaise (ELL) sont plus susceptibles de voir leurs résultats en mathématiques diminuer, ce qui laisse présager un besoin de soutien pédagogique plus important pour ces élèves.

Mots-clés : tendances de réussite en mathématiques, apprenants de langue anglaise, évaluations standardisées, responsabilité fondée sur les tests, analyse de classes latentes

Introduction

The importance of mathematics learning in the elementary grades is widely agreed upon. Early math performance is a known predictor of later academic success and career choices (Claessens & Engel, 2013; Duncan et al., 2007; Ritchie & Bates, 2013), and the predictable nature of above-average numeracy and reading achievement toward careers in science, technology, engineering, and mathematics is well-substantiated (Holmes et al., 2018). Globally, math achievement is measured as an indicator of adequate progress toward educational standards (Loveless, 2007). Similarly, in some Canadian provinces, elementary grades have maintained a heightened focus on math, given what some describe as a concerning decline in students' math achievement (Education Quality and Accountability Office [EQAO], 2017a). A common and widely favoured method for collecting information regarding students' patterns of achievement, including in Canada, is high-stakes standardized assessments that are embedded in wider test-driven systems of accountability (Anderson et al., 2006). One of the most significant challenges with such standardized test-driven accountability practices lies in whether or not aggregated test scores provide useful information about the levels of achievement across diverse subgroups of students with different learning needs, such as English Language Learners (ELLs) or students with exceptionalities.

The concerns around test-driven systems of accountability are of particular immediacy for Ontario educators and other stakeholders who face increases in diversity in classrooms in many ways and have observed declines in math achievement in the elementary years. Of yet, however, little empirical evidence has been made available to account for these math declines. Considering these challenges are arising from a lack of empirical evidence to substantiate the nature of these serious declines in Ontario's math achievement, the purpose of the present study was to examine the characteristics of math achievement patterns on Ontario's standardized Education Quality and Accountability Office (EQAO) mathematics assessment as students progress from Grade 3 to Grade 6, and the likelihood of achievement patterns when taking into account students' home linguistic backgrounds and literacy achievement levels. Specifically of interest was the patterns of achievement from Grade 3 to Grade 6 of students typically identified as ELL students. Given the apparent association between language and math, a primary research question guiding the present study was: Are the declining math achievement patterns associated with students' linguistic background? This study applied a two-step latent class analysis (LCA) approach to provincial longitudinal cohort data to support understanding of the nature of these math achievement patterns and their associations with students' linguistic backgrounds.

Language and Literacy in Math Achievement

Previous research has examined various factors influencing math achievement, such as language proficiency (Abedi & Dietel, 2004), parental math anxiety (Soni & Kumari, 2017), socio-economic status (Merolla, 2017), and gender (Lee et al., 2011), among others. Research has shown that literacy is an effective and reciprocal predictor of students' math and overall academic achievement (Lerkkanen et al., 2005; Romano et al., 2010; Zhang et al., 2017). There is well-supported evidence to suggest that early literacy skills, such as print knowledge and phonological awareness, are closely related to early math and numeracy skills (Purpura et al., 2011). Similar studies have shown significant correlation between students' drawings, an early writing convention, and ability to print letters of the alphabet, and mathematical sorting in kindergarten (Steffani & Selvester, 2009). Recently, Zhang et al. (2017) investigated the impact of early language abilities on both formal (e.g., calculations including operations) and informal (e.g., basic number knowledge) math skills (e.g., Ginsburg & Baroody, 2003; Jordan et al., 2009) and found a strong predictive relationship between the early language and math skills of children from kindergarten to Grade 3.

The relationship between literacy and math achievement is particularly relevant to students receiving instruction in a language they are also still learning. In English education programs (English Second Language [ESL] Programs), ESL students with fewer than three to five years of residence tend to perform worse than their English-proficient peers in a variety of academic subjects, including math (Abedi, 2002; Dearing et al., 2016). Studies have demonstrated that for this population of students, math proficiency is directly related to language proficiency (Vukovic & Lesaux, 2013). Researchers de Araujo, Roberts, Willey, and Zahner (2018) noted that, from an information-processing perspective, the influence of language and literacy on math achievement is explained as the result of higher cognitive-processing loads; that is, students must process the language in which the math question is asked, as well as simultaneously activate the relevant math knowledge to then answer the question. Attempting to balance learning in two subject areas that interact so closely could presumably result in great difficulty for assessment of ELL students' math knowledge.

Test-Driven Accountability

Test-driven accountability systems in education are hardly new (Jang & Ryan, 2003). Standardized large-scale tests are widely used to provide information for evaluating how well students and schools perform with reference to a pre-determined set of curricular expectations. Well-known standardized assessment systems include the Programme for International Student Assessment (Organisation for Economic Co-operation and Development, 2016), the Trends in International Mathematics and Science Study (Provasnik et al., 2016), and in Canada, Ontario's EQAO (Jang & Sinclair, 2018). Accountability in these systems refers to who is held accountable for what and to what extent (Lee, 2008; Wiliam, 2010). The purpose of test-driven accountability has been described as providing students, teachers, and administrators with incentive to "work harder" and help those at risk of failure (Jacob, 2005).

Some argue that this purpose has backfired, though, with accountability instead providing stronger incentives for teachers and schools to focus on those nearer to successful achievement of standards, and thereby pulling attention away from students who are either less likely to achieve or have surpassed achievement of standards (Neal & Whitmore Schanzenbach, 2010). Teacher attitudes toward test-driven accountability reflect a rejection of accountability as a positive incentive. Khalvandi and Chenari (2012) found that while teachers hold some positive attitudes toward test-driven accountability structures, they prefer a more comprehensive approach to assessment that includes space for student feedback and formative assessment, among other features. Wiliam (2010) notes that the notion of accountability is both interesting and troubling, as the literal definition of accountability suggests expectations of positive outcomes being placed on someone or something, as well as proof of actions made toward meeting those expectations. In Ontario, students meeting academic standards on the EQAO assessments could be considered the positive outcome, with accountability and expectation of meeting those standards tending to fall on teachers, arguably most especially those of elementary students, as it is in the elementary grades where much of the decline in math achievement has been found.

In many cases, and recently in Ontario, teachers' own ability to teach math and their assessment competency has been questioned (Khalvandi & Chenari, 2012; Ontario Ministry of Education [OME], 2019).

An additional element contributing to accountability and incentive is the stakes involved in such testing; that is, the consequences attached to performance on tests (Steedle & Grochowalski, 2017). High-stakes testing focuses less on a minimum competency and more on a proficiency of skills (Lee, 2008). Low-stakes testing is much less common as research has shown that when stakes are increased, students' performance and motivation to perform increases (Steedle & Grochowalski, 2017; Wise & DeMars, 2005). There are drawbacks to the use of higher stakes in testing, however; high-stakes testing can increase students' experiences of anxiety around certain curricular subjects, and teaching practices can be skewed toward "teach[ing] to the test" (Giouroukakis & Honigsfeld, 2010; Shepard, 1990; Steedle & Grochowalski, 2017). Teaching to the test risks only teaching students how to succeed on the tests themselves, impeding overall academic learning (Holmstrom & Milgrom, 1991; Neal & Whitmore Schanzenbach, 2010). Questions may also be raised about whether the results collected from such tests can provide information that would be useful in identifying students requiring additional supports, especially supports that are beyond the scope of overall achievement indicators used at individual schools and wider school board levels.

Ontario's Test-Driven Accountability System

Where societies are culturally, ethnically, and linguistically diverse, the complexities of understanding appropriate academic and assessment needs and techniques for ELL students calls for consideration of students' individual characteristics, such as their home language backgrounds (Jang et al., 2013). In Canada, and particularly Ontario, students who are still developing English language proficiency experience difficulty closing the achievement gap in classrooms that include increasingly diverse linguistic populations but may not provide instructional diversity. The need for consideration of ELL students' achievement is particularly notable as difficulty meeting math standards is well-documented in existing literature (DiGisi & Fleming, 2005; Moschkovich, 2015), and over the last decade the achievement of provincial math standards on the EQAO provincial assessment has consistently declined across student bodies, resulting in what has been unofficially deemed Ontario's "math problem."

EQAO assessments take place every spring across Ontario in Grade 3, 6, and 9 classrooms. These assessments target literacy (reading and writing) and math skills. The EQAO focus on math and literacy performance is unsurprising, as early math achievement can predict later overall academic achievement (Duncan et al., 2007). Students in Grades 3 and 6 complete assessments in both literacy and math, while Grade 9 students receive an assessment of math skills only. In Grade 10, the Ontario Secondary School Literacy Test (OSSLT) is used to focus exclusively on literacy skills (EQAO, 2017b). Students spend two weeks between May and June completing tests that closely align with the provincial curriculum expectations for their respective grades. EQAO has identified their number one legislative objective as "evaluat[ing] the quality and effectiveness of elementary and secondary school education" (EQAO, 2003, p. 2).

Since EQAO's inception in 1996, trends in meeting academic standards have generally been positive for each grade. One exception to this trend, however, was found for math achievement from Grades 3 to 6. As of 2003, EQAO has reported yearly on trends in student achievement (Figure 1). The most recently published reports have detailed consistent gains in the percentage of students meeting standards on reading and writing assessments in Grades 3 and 6 (74% and 81%, and 73% and 79%, respectively). However, the opposite trend has persisted for meeting math standards in Grades 3 and 6 (62% and 50%) (EQAO, 2018). This trend in declining scores, they report, has consistently been the case over the last six years of assessments. The same drastic decline in scores over time has not been revealed by results on academic and applied student performance in Grade 9.

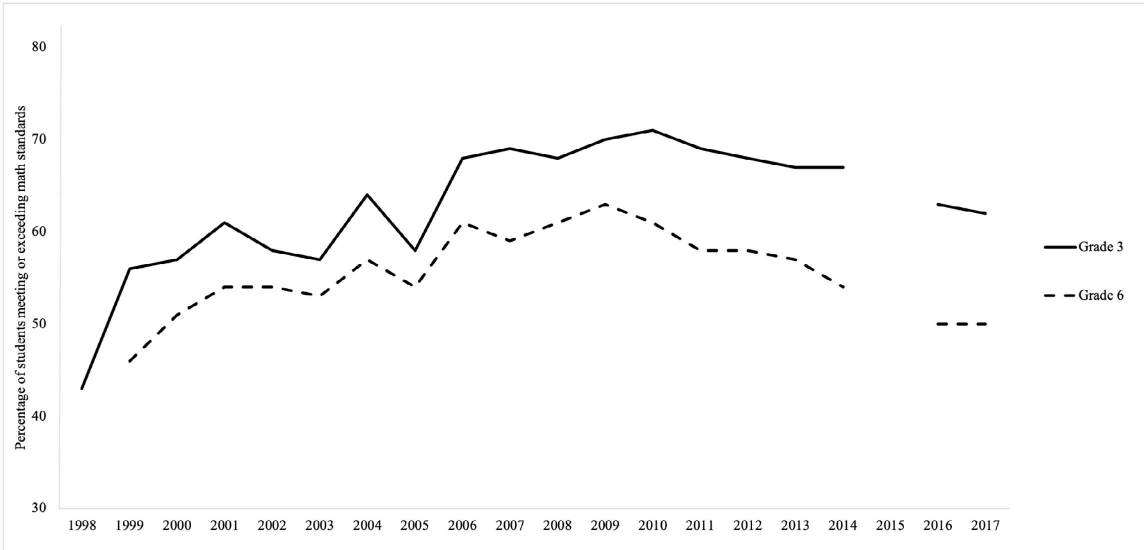

Figure 1. Trends in meeting or exceeding EQAO math standards from 1998 to 2017. Spaces represent years in which no assessments were administered (EQAO, 2017a).

In 2016, EQAO publicly declared that in Ontario "half of all grade six students [are not meeting] the provincial math standard" (EQAO, 2017c, p. 9), yet failed to provide any details on whether the declining trends are uniform across all student bodies and, if not, which subgroups are represented in the declining trends. Various remedies to improving teachers' math competencies have been implemented to "fix the math problem" with little empirical evidence (EQAO, 2019a; OME, 2019). This study presents a first attempt at a more nuanced understanding of Ontario's "math problem." The purpose of this study was twofold: first, to uncover any patterns in EQAO math achievement scores of a cohort of students' in Grade 3 and Grade 6 . Our research question supporting this purpose was: What are the trends in patterns of math achievement across two elementary grades? Research suggests there is a close relationship between math and literacy. Given the need to characterize Ontario's achievement patterns, and a diversifying landscape, this study's second purpose was to explore the predictive nature of language and literacy characteristics on math achievement trends. This purpose was supported by the following research question: Do language and literacy characteristics affect declining EQAO math achievement patterns?

Methods

Participants

The participants of this study (N = 123,773) were in a cohort of students in Grades 3 and 6 in the years 2014 and 2017, respectively. These data were made available to the study's researchers by EQAO. As a part of the EQAO assessment itself, detailed demographic information on all students completing the assessment is recorded by the classroom teacher through a student questionnaire. These questionnaires cover such areas as gender identification, place of birth, years lived in Canada, years enrolled in the school board and school at the time of testing, first language, and language spoken at home. Table 1 provides a summary of descriptive statistics for the cohort of students included in analyses. Gender was equally represented, with 50.7% of participants identifying as male, and 49.3% identifying as female. Over half of the students (61.7%) learned English as their first language, while 22.4% learned another language. The majority of these students were not enrolled in an ESL program (87.0%). In their home environments, most students spoke only English (44.6%) or mostly English (23.8%). Approximately 17.5% of the participants spoke both English and another language(s) at home, while 7.1% indicated speaking mostly another language(s). Less than five percent of participants indicated speaking exclusively another language at home (4.4%). All students were enrolled in English language programs.

Table 1. Descriptive statistics in Grade 3

	n	%		*N*	%
Gender			**Home Language**		
Male	62,714	50.7	Only English	55,182	44.6
Female	61,059	49.3	Mostly English	29,405	23.8
English Second Language Program			English and another language(s)	21,704	17.5
Not enrolled in ESL program	107,631	87.0	Mostly another language(s)	8,828	7.1
Enrolled in ESL program	16,142	13.0	Only another language(s)	5,401	4.4

	n	%		*N*	%
	First Language				
English	76,385	61.7			
Another language	27,673	22.4			

Note. Participant totals *n* for select variables do not equal total sample due to incomplete or missing data.

In 2014, the majority of these participants (70.2%) were successful in meeting or exceeding Grade 3 EQAO math standards, obtaining a level 3 or 4 on the tests. Just under one third of Grade 3 students (29.8%), however, were unable to meet those same standards in that year, obtaining a level 1 or 2 on the test. In 2017, the proportion of students meeting or exceeding Grade 6 math standards decreased to approximately 54%, while the proportion of students failing to meet those same standards increased to just over 46% (Table 2).

Table 2. EQAO achievement levels in Grades 3 and 6

Grade 3 (2014)	*n*	%	Grade 6 (2017)	*n*	%
Level 1	4,036	3.3	Level 1	18,946	15.3
Level 2	32,846	26.5	Level 2	38,185	30.9
Level 3	69,904	56.5	Level 3	48,200	38.9
Level 4	16,987	13.7	Level 4	18,442	14.9

Note. Participant totals *n* do not equal total sample due to incomplete or missing data.

Measures

Math. The EQAO elementary math assessment tests students' achievement outlined by Ontario's math curriculum expectations in five math strands: number sense and numeration, measurement, geometry and spatial reasoning, patterning and algebra, and data management and probability (OME, 2005). Demonstration of strength in three math skills are targeted by the test: knowledge and understanding, application, and thinking (EQAO, 2017a). The Grade 3 test consisted of 36 items with a Cronbach's alpha of .89, rising to a Cronbach's alpha of .91 for the Grade 6 test, suggesting acceptable internal consistency among the test items (EQAO, 2014, 2019b). Items on both the Grade 3 and Grade 6 tests include two item formats: multiple-choice response items and constructed response items. Scores are calculated based on a modified 3PL IRT (Item Response Theory) model with a guessing parameter estimate fixed at .2 for multiple-choice items (EQAO, 2017b). Students are assigned to one of four achievement levels based on the application of cut-score points to IRT theta values. As noted above, Levels 1 and 2 indicate standards that have not been met, and Levels 3 and 4 indicate standards that have been met or exceeded (see Table 2 for cohort distribution at Grades 3 and 6). Item writing, field trial, and scoring involve classroom teachers to a great extent. Both literacy and math test items are intentionally aligned with grade-specific curriculum expectations (EQAO, 2017d). EQAO claims that all test items are evidence-based and age- and grade-appropriate, and that scoring rubrics for writing prompts for open-response items are "clearly related to the writing prompts and effectively capture the range of student responses" (EQAO, 2013, p. 13). Permittance of math tools, such as calculators, is clearly marked for specific questions as students move through the test. Unless otherwise noted, all students are required to complete the assessments. Factors, such as cognitive development and conferencing with parents and/or teachers are considered when determining the eligibility or disqualification of students to complete the EQAO assessment.

Literacy. EQAO reading and writing assessment scores were taken from Grade 3 literacy assessments. Similar to the math test, the literacy assessment is used to evaluate the overall literacy achievement levels in three reading strands including reading for meaning, understanding form and style, and reading with fluency, and three writing strands including developing content, organizing content, and using knowledge of form and style in writing. Demonstration of three reading skills (e.g., understanding explicitly-stated information and ideas, making inferences, and making connections between information and ideas in a reading selection and the reader's personal knowledge and experiences) and two writing skills (e.g., topic development and conventions) are also included in score calculation (OME, 2006).

Variables

Math assessment score. Overall math achievement levels at each of the two grades, 3 and 6, were used as outcome variables. The four achievement levels were recoded into a binary variable. Levels 1 and 2 (i.e., *did not* meet the standard, or fail) were grouped and recoded as 0, and Levels 3 and 4 (i.e., *did* meet the standard, or pass) were grouped and recoded as 1. The result was a binary pass/fail indicator to be used specifically for the identification of class membership in LCA. Students for whom data were missing as a result of, for example, exemption, withholding, or missing data, were removed from subsequent analyses.

Reading and writing assessment scores. Overall reading scores and overall writing scores in Grade 3 were used as predictor variables. Similar to math assessment scores, the overall achievement levels provided by EQAO, originally Levels 1 to 4, were recoded into binary variables. Levels 1 and 2 (i.e., *did not* meet the standard, or fail) were grouped and recoded as 0, and Levels 3 and 4 (i.e., *did* meet the standard, or pass) were grouped and recoded as 1 for each of the overall reading scores and overall writing scores. Historically, EQAO literacy assessment scores have maintained consistent results meeting standards (Hinton, 2014).

Language background. We identified students as English as first language (EL1) or ELLs based on whether a student's first language was English or another language was used. The survey question "*Did the student learn English as a first language at home?*" was dummy coded such that ELLs were coded as 1, and EL1s were coded as 0. This variable identifies whether students' first language was or was not English, and is not to be confused with "ESL Program" (see below), which refers to the language support program offered in schools mostly to those who recently migrated to Canada.

ESL program. In addition to language profiles, students' enrolment in school ESL programs in Grade 3 was included in regression analysis to predict latent class membership. The Ontario Ministry of Education (2007) described the goals and curriculum expectations of their ESL programming as a provision "to help English language learners develop the skills they need to develop proficiency in everyday English and, most especially, proficiency in academic English that will allow them to integrate successfully into the mainstream school program" (p. 3).

Gender. Gender was used as a covariate to control for any differences in achievement patterns as a function of students' gender. This is a commonly used variable of interest in studies of academic achievement, and there exists a rich body of research to support the identification of gender as a factor of students' math achievement in particular (Lee et al., 2011; Ganley & Lubienski, 2016).

Data Analysis

Latent class analysis (LCA). Using Mplus version 8 (Muthén & Muthén, 2017), LCA was used to identify the latent profiles of students with the distinct patterns of math achievement on the EQAO assessment across the two grades, 3 and 6. LCA is a mixture modelling method used to identify latent classes of individuals who share distinct characteristics (Collins & Lanza, 2013). It is a person-oriented latent profiling approach that characterizes naturally occurring differences among groups, rejecting the assumption of homogeneity within those groups (Hickendorff et al., 2018; Swanson et al., 2018). LCA use also assumes there are underlying and unobserved classes that divide a population into multiple mutually exclusive groups (i.e., latent classes) (Lanza & Rhoades, 2013). Traditional linear modelling approaches, which are variable-oriented and use variables as the unit of analysis rather than individuals, present limitations in their characterization of heterogeneity and non-linear patterns (Sinclair et al., 2018; Hickendorff et al., 2018). LCA in literacy research has been used to create intervention programs that effectively address the individual needs of students (Sinclair et al., 2018). For the present study, LCA was used to identify more nuanced math achievement patterns beyond average percentages of students at four achievement levels.

No prior theoretical argument has been made for appropriate class membership for elementary school math achievement patterns; therefore, initial model analyses began at two classes. Various model-fit evaluation criteria were considered. They included entropy for overall classification accuracy, Akaike information criterion (AIC), Bayesian information criterion (BIC), Adjusted Bayesian information criterion (aBIC), Vuong-Lo-Mendell-Rubin Likelihood Ratio (VLMR), and Lo-Mendell-Rubin Adjusted (LMRA) LRT tests of significance of model-fit (Nylund-Gibson et al., 2007). A 3-class model showed no significant improvement over a 2-class model as shown by statistically insignificant results from VLMR and LMRA LRT tests. As a result, no other model was tested. Two LCA models were compared. As shown in Table 3, a 2-class model was determined to be the best-fitting model for class membership of math achievement patterns across Grades 3 to 6. Although its AIC, BIC, and aBIC values were slightly higher than those for the 3-class model, the 2-class model showed a higher entropy value and better interpretability.

The same analysis was run using students' EQAO literacy assessment scores. Students' 4-level literacy scores were recoded into a binary score: those who did not meet the standard (Levels 1 and 2), or 0, and those who did meet the standard (Levels 3 and 4), or 1. This binary variable was used in the LCA. Trends in Ontario's literacy scores had historically not experienced the same downward trend seen with math scores. Rather, a high majority of students continued to achieve average or above average scores on EQAO literacy assessments across Grades 3 to 6 (EQAO, 2018). Latent Class Analysis confirmed this trend in literacy scores, showing two distinct latent profiles of achievement that remained steady over time or increased in Grade 6. Therefore, one can conclude that the majority of students were able to achieve literacy standards in Grade 3 and continued to maintain that achievement in Grade 6.

Table 3. Model-fit evaluation

Criteria	2-class model	3-class model
Akaike (AIC)	**287199.57**	287205.57
Bayesian (BIC)	**287248.20**	287283.38
Adjusted BIC (aBIC)	**287232.31**	287257.95
Log-likelihood	**-143594.78**	-143594.78
Entropy	**.96**	.60
Vuong-Lo-Mendell-Rubin Likelihood Ratio Test For 1 (H0) Versus 2 Classes	**.00**	.92
Lo-Mendell-Rubin Adjusted LRT Test	**.00**	.92
Proportion in Class 1	**66,642**	66,642
Proportion in Class 2	**57,131**	57,131
Proportion in Class 3	---	0

Note. Bolded values indicate the best-fitting model.

Binary logistic regression. Using the two latent classes of math achievement patterns from the LCA analyses as the outcome variable, a hierarchical binary logistic regression analysis was conducted using SPSS version 24 to examine the predictive relationship between students' language and literacy characteristics and latent classes. Independent variables related to language and literacy included overall reading and writing scores, enrolment in ESL programs, language background, and gender. Because gender was used as a covariate, it was applied to the first level in the regression analysis (Step 1). Overall reading and writing scores were inputted into the second level (Step 2). Because the earlier LCA showed a consistent trend toward maintaining average or exceeding above average achievement, we concluded that students' literacy scores did not fluctuate enough to have a negative impact, or confound, results of the logistic regression using latent math profiles, and was therefore used as a predictor variable in the analysis. The remaining two variables, language background and ESL program enrolment, were applied to a third level (Step 3), with the preceding steps' variables controlled.

The dependent variable, now latent classes, was coded as "0" (class 1) and "1" (class 2). Because the focus of the study was on class 2, which was the declining math achievement pattern group, regression coefficients for class 2 were estimated and compared with reference to class 1. Any participants with missing values for any of the variables of interest were removed from analyses, resulting in a sample of 98,486 participants included in the logistic regression.

Results

Math Achievement Patterns

Latent class analysis resulted in the 2-class model as the best-fitting model. Of the students included in the 2-class analysis, 66,642 (54%) were categorized into class 1, and 57,131 (46%) into class 2, as shown in Table 4.

Table 4. Class distribution and results in probabilities

		Counts	Proportions	Estimate	S. E.	Est./S. E.	P-Value
Latent Class 1		66,642	.54 (54%)				
Grade 3	Fail			.08	.001	76.22	.00
	Pass			.91	.001	864.70	.00
Grade 6	Fail			.01	.003	2.91	.00
	Pass			.99	.003	379.40	.00
Latent class 2		57,131	.46 (46%)				
Grade 3	Fail			.55	.003	217.10	.00

		Counts	Proportions	Estimate	S. E.	Est./S. E.	P-Value
	Pass			.45	.003	174.44	.00
Grade 6	Fail			.998	.000	0.000	1.00
	Pass			.002	.000	0.000	1.00

Note. Bolded values indicate probabilities of meeting the standard across the two grades.

These results confirm that there are two distinct patterns of achievement on the EQAO math assessment from Grades 3 to 6. The differences between these classes were unmistakeable; class 1 showed an approximately 92% chance of meeting or exceeding math standards in Grade 3, increasing to just under a 100% chance of doing the same in Grade 6. This class was identified as the "consistent achievement group." Students in class 2, though having an approximately 45% chance of meeting standards in Grade 3, had significantly less success meeting standards again later, with their chance of continuing to meet standards decreasing to well below 1% in Grade 6 (Table 4). Class 2 was therefore identified as the "declining achievement group" (Figure 2).

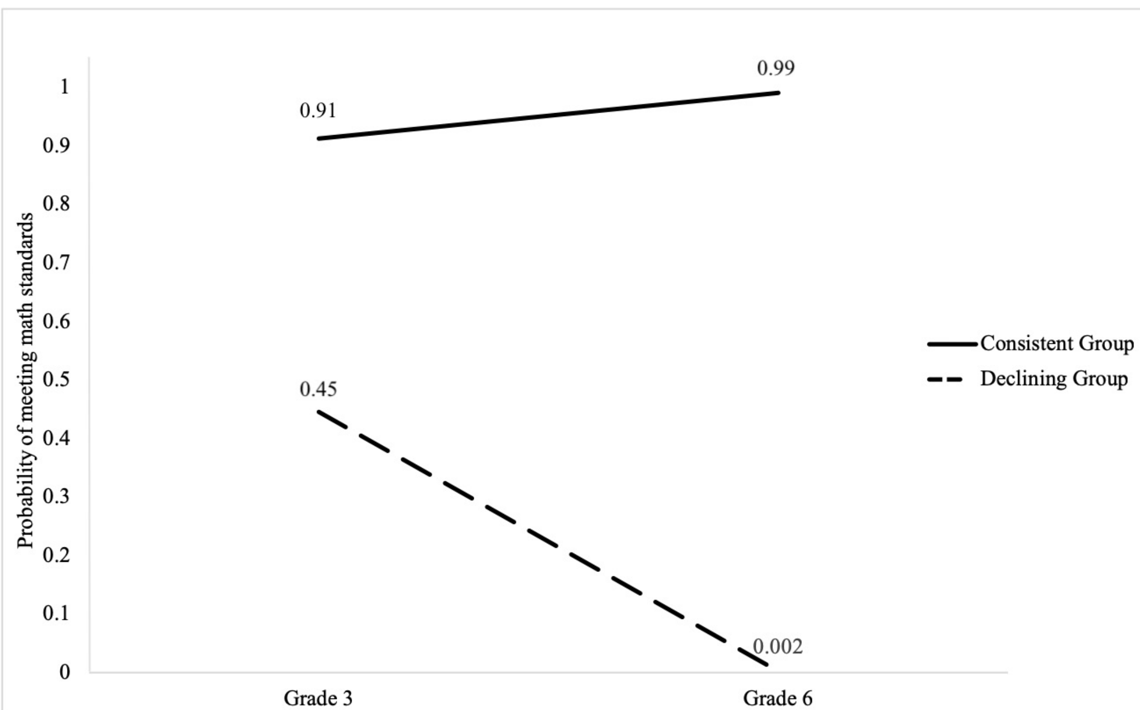

Figure 2. Probabilities of class membership based on meeting or exceeding math standards

Predicting Latent Class Membership

To determine the predictive nature of language and literacy on the math achievement latent class membership, a hierarchical binary logistic regression was conducted. A test of the full model with independent variables, overall reading score, overall writing score, ESL program enrolment , language background, and gender as predictors was statistically significant, $\chi^2(5) = 16{,}619.60$, $p < .001$, indicating that the math achievement latent class membership can be predicted by language and literacy variables.

All variables were significant predictors of latent class membership, as shown in Table 5. Specifically, the odd ratios of two literacy variables, reading and writing scores, were .18 and .39 each. As indicated by the negative beta coefficients for reading and writing scores, the higher reading and writing test scores, the less likely a student was to belong to the declining achievement pattern group.

Table 5. Binary logistic regression predicting membership to Class 2 based on language and literacy predictors

Step	Predictor	B	S. E.	Wald	Sig.	Exp(B)
1	Gender	.03	.02	323.82	.00	1.33
2	Overall Reading Score	-1.72	.02	7,317.37	.00	.18
	Overall Writing Score	-.94	.02	1,549.72	.00	.39
3	Language Background	.69	.02	1,019.77	.00	1.99
	ESL Program	.23	.03	71.43	.00	1.26
	Constant	.99	.04	768.74	.00	2.68
	Model χ^2			16,619.60 p<.001		
	-2 Log likelihood			97,927.69		
	N			98,487		

On the other hand, the beta coefficient for the language background variable (1 for ELL students) was positive and statistically significant. Its odd ratio (1.99) strongly suggested that ELL1 students would be almost two times more likely to belong to the declining math achievement pattern group. The ESL program variable showed a similar result, indicating that those who are in ESL programs had a greater likelihood of experiencing declines in their math achievement in Grade 6.

Discussion

This study's findings uniquely contribute to the study of EQAO test achievement by exploring the effect of students' English language and literacy characteristics on patterns of achievement across two elementary grades, with emphasis on the achievement of ELL populations. Latent class analyses revealed two distinct patterns of EQAO achievement for students from Grades 3 and 6. Class 1 included students who met and continued to meet standards in both Grade 3 and Grade 6 (consistent achievement group), while class 2 included students who met standards in Grade 3, but were unable to do so in Grade 6 (declining achievement group). While initial claims stated that half of Ontario's students were failing to meet math standards, these results show that claim was not entirely accurate. In actuality, less than half of students who initially were able to meet expectations in Grade 3, were unable to do so in Grade 6. Although the percentage of students unable to meet math standards was just under 50% (specifically 46%), these results show the existence of more heterogeneous achievement patterns of students, and therefore suggests a need for more targeted identification and assessment of those students who failed to achieve a passing mark on the EQAO math assessments.

Subsequent logistic regression analyses revealed a significant predictive relationship between literacy and language characteristics and math achievement. Those who struggled with literacy proficiency were more likely to belong to class 2 (declining achievement group), highlighting the importance of literacy skills on math achievement. These results are consistent with other studies, such as Rutherford-Becker and Vanderwood (2009), who similarly found a predictive relationship between the literacy skill of reading comprehension and math computation (e.g., addition and subtraction). Jordan, Kaplan, and Hanich (2002; see also Jordan, Hanich, and Kaplan, 2003) also found that children in early elementary grades who experience math difficulties, yet are good readers, are likely to perform better on math tests than their peers who are poor readers and experience math difficulties. These findings also indicated that children who live and learn in environments with different dominant languages (i.e., whose first language is not English in the context of this study), and who would be identified as ELL students, experience significant difficulty meeting math standards when compared to their English-proficient peers. The predictability of students' math achievement as a function of their language background was also consistent with studies that have found that language proficiency can explain differences in math achievement between ELL students and their English peers (Jang & Sinclair, 2018; Zheng et al., 2007). These difficulties have been explained by some researchers as relating to the specific language used to communicate math problems (Lager, 2006), which could in turn bring into question broader issues of differential item functioning (DIF) in accountability testing (Martiniello, 2009) and the biasing of questions toward certain populations of students (see Kim & Jang, 2009, for an investigation of DIF in EQAO math assessments). Though there exists a number of possible explanations for the differences in ELL achievement patterns, both generally and in the context of math learning, this study highlights and contributes to first steps in rectifying "Ontario's math problem" and closing the achievement gap for ELL students by identifying ELL students as a specific group failing to meet math standards. Identification is the first step in supporting students who struggle with EQAO math assessments; steps to follow should identify needs and possible explanations.

ELL Students and Academic Achievement

Instances of ELL students performing below their domestic counterparts on tests of academic achievement are well documented in the literature (Abedi, 2002). In standardized test accountability settings similar to EQAO, the needs of ELL students are distinctly different from those of their peers (Honigsfeld & Giouroukakis, 2011). In line with the literature, an achievement gap was found for ELL students in the current study. A possible explanation for this phenomenon could be that the language complexity and proficiency that was demanded by the test in Grade 6, though manageable in Grade 3, was higher than the language proficiency students had been able to develop by the time of the test (Abedi & Lord, 2001). Failing to reach the level of language proficiency necessary to complete standardized tests has been known to create barriers to successful achievement, especially for ELL students (Martiniello, 2009). Language, then, has the potential to serve as a barrier to successful math achievement. For these students, language may have become something of a confounding factor in their successful achievement of academic standards (Abedi, 2002).

These students showed no difficulty meeting standards across the two grades in EQAO's literacy assessment. However, it is important to note here that when referring to the language proficiency needed to comprehend and respond to math questions appropriately, we are referring to specific math language and cultural and contextual references used in math questions. Mathematical literacy, as implied by the name, involves understanding and using key terminology (DiGisi & Fleming, 2005; Roberts, 2009). This concept, however, also requires that students recognize how language is used to internalize and disseminate math knowledge in a way that can be understood by an audience, which necessitates having some familiarity with cultural context and language. Roberts (2009) suggests that this in turn means that, as a discipline, math is tied not only to specific mathematical vocabulary and "talk," but equally to societal constructs being represented by those words. Moschkovich (2015) argued that the confounding effect of language on math achievement is due to certain complex linguistic skills that are needed when responding to math questions, including defining, explaining, and justifying.

These results demonstrate how language development, when not appropriately supported, can hinder achievement. The language of testing may have affected students' abilities to meet standards, as the test was in a language that they were still learning. Further, these results provoke important questions about the appropriateness of these tests for diverse classroom contexts, how to conceptualize the learning needs of ELL students, and how to identify effective assessment tools and strategies to serve their learning needs. These are questions that are reflective of the current discourse around culturally and linguistically responsive practices and ELL testing considerations (Abedi & Dietel, 2004; Sinclair et al., 2018; Honigsfeld & Giouroukakis, 2011; Ladson-Billings, 1995, 2014).

There are a number of other notable contributions from this study related to ELL discourse, education, and testing. Firstly, this study puts an urgently necessary spotlight on the issues faced by all ELL students in math education, especially in Ontario where students are required to participate in standardized testing four different times throughout their elementary and secondary education, particularly in a linguistically and culturally diversifying Canada. ELL students continue to experience systemic and institutional barriers to their successful academic achievement. This study followed a traditional dichotomous view of ELL students that is commonly used in research about language learning and academic achievement. Dichotomizing ELL students in such a way forces others' assumptions of their language needs and abilities, and risks disregarding the unique lived experiences of students outside of school hours, and the influence after-school experiences and the home environment have on their overall learning (Honigsfeld & Giouroukakis, 2011). To give prominence to students' unique identities outside the classroom context, a more nuanced and multidimensional view of ELL students, one that considers other facets of their language development such as home language and immigration status, was used by Sinclair, Jang, and Vincett (2018). Their findings showed the significant and favourable impact these other facets of language background may have on academic achievement, and further underscore the need to consider students' lived experiences outside of the classroom as important factors in academic achievement.

As an additional (and only superficially explored) finding of the study, these results also call attention to the possibility that ESL programming may not be reaching as many students as may have needed such supports. While just over 20% of the students in this study were identified as ELL, who were in large part also born within Canada, it also showed that just over 10% of students overall, and only 58% of ELLs specifically, received supports from an ESL program. This may be due to the limiting and antiquated definitions of who is an ELL. Canadian-born multilingual students are less likely to be identified as ELL students than those born outside of Canada, and as such may not be adequately identified as in need of language supports or placed into language support programmes, such as ESL and English Literacy Development (ELD) programmes (Mcgloin, 2011). Canadian-born ELLs' high communicative or conversational competence in English can confuse or mask the need for more formal language supports. Previous studies have found that Canadian-born ELLs show lower rates of meeting the curricular expectations compared to domestic ELLs (Jang et al., 2015), and Canadian-born ELLs with low literacy in their first language are at the highest risk of not succeeding in the OSSLT (Cheng & Sun, 2015). At the teacher training level, Cummins (2014) argues that Canadian schools have failed to equip teachers with the knowledge to appropriately identify students in need of ESL supports and develop necessary pedagogical skills to teach them effectively once identification has been made. Indeed, he notes that many teachers and administrators "have not had opportunities to access the knowledge base regarding effective instruction for [students of immigrant backgrounds], nor have they had opportunities for pre-service or in-service professional development regarding effective instructional practices" (p. 1). Overall, the results of this study call attention to the complex nature of Ontario school students' linguistic backgrounds in order to provide effective accommodations for ELL students' math learning.

Cultural and Linguistic Sensitivity in Math Assessment

The study findings suggest that students' math ability is highly correlated with their literacy ability. They further suggest that ELL students' math test performance may be hindered by their limited English language proficiency. The study findings are consistent with other studies that have pointed to the structural and content features of math assessments as contributing to disproportionate ELL achievement. For example, DiGisi and Fleming (2005) suggest that US state math tests require additional reading skills before one could even begin to answer questions, further linking math test performance to language and literacy. Honingsfeld and Giouroukakis (2011) argued for the consideration of concepts and ideas highlighted by test questions and their relevance to, or familiarity with, diverse cultural contexts. Test accommodation strategies include translating tests into the primary language of test-takers, however, even when tests are translated to primary languages, ELL students may still achieve scores lower than their peers (Abedi & Dietel, 2004). Abedi and Dietel (2004) argue that though translation allows for a moving away from ascribing poor achievement on tests to the tests themselves, it also underscores the need to develop better understanding of the effect of language skills on achievement outcomes.

Abedi and Lord (2001) report the consistent performance improvements exhibited by ELL students when test questions are modified or simplified while maintaining rigour as justification to provide linguistically modified, rather than translated, versions of tests. Honigsfeld and Giouroukakis (2011) argued that, for ELL students, these types of assessments create greater challenges due to decreased familiarity with the cultural contexts of questions. These students, with different cultural and linguistic skillsets, take tests that were designed to give preference to English-speaking students and include western cultural references (Honigsfeld & Giouroukakis, 2011). Providing questions that capitalize on students' lived experiences allows for greater contextual understanding of content (Chval & Chavéz, 2011; Honigsfeld & Giouroukakis, 2011) and a decrease in cognitive-processing loads related to complex linguistic skills required to decode meaning from questions (Moschkovich, 2015). Chval and Chavéz (2011) proposed, from a teacher methods perspective, that teachers give consideration for the speed at which complex language skills and proficiencies develop (when compared to math skills), suggesting that making deliberate pedagogical decisions to support language development alongside math can result in the creation of optimal learning environments for ELL students.

The language of test administration also presents certain challenges that Coltrane (2002, as cited in Honigsfeld and Giouroukakis, 2011) would argue leave these students at a disadvantage. The sustainable use of scores on standardized math tests as indicators of overall student achievement is also brought into question. These findings call attention to the possibility that EQAO, while successful in other subject areas, may not accurately capture the learning of Ontario's increasingly heterogeneous population of students. As such, the use of standardized achievement measures for accountability purposes in current policy structures, particularly in Ontario, requires critical reflection. Although these issues were not directly addressed in the present study, the study results strongly suggest further research to examine the extent to which EQAO math tests are culturally and linguistically sensitive to students from diverse backgrounds.

Limitations and Next Steps

This study was affected by a few potential limiting factors. The absence of definitive information on participants' socio-economic status (SES) presented an important limitation, especially as studies have shown that SES can significantly affect students' successful academic achievement (Barr, 2015; Merolla, 2017; Sucuoğlu, 2018). This has continued to hold true in the context of math learning, as marked differences in SES have presented across math strands beginning as early as, and in some cases earlier than, kindergarten, and often not diminishing as students moved through grades (Elliott & Bachman, 2018; Galindo & Sonnenschein, 2015; Klein et al., 2008). Therefore, it is suitable to question whether these results would have been affected by the inclusion of SES demographic information. Additionally, the use of a single data source and only two timepoints as representative of overall elementary achievement may have resulted in findings that were not easily applied to all elementary grades. Further research, using more timepoints and multiple sources, may help support the generalizability of these findings.

As explained above, students in this study did not experience similar difficulties meeting literacy standards on the EQAO assessments. This finding allowed for the achievement scores to be used as a variable of interest in our analysis, though we did not make any inferences as to why there was not a similar decline in literacy achievement as it was beyond the scope of the present study. Some possible explanations for the existence of this decline in math but not literacy have been noted here, such as the possible cognitive load of decoding language at the same time as working out mathematical calculations that is not necessary in literacy applications alone. Future studies exploring the trends in EQAO math and literacy achievement will be done to unpack this dichotomy in achievement and try to understand why language background does not similarly affect literacy achievement. In math and literacy achievement, males tend to outperform females in math, while females outperform males in literacy. In this study, the relationship of gender to latent class membership was only explored superficially as a covariate. Though these results were consistent with the literature on males' greater success in math achievement, in order to unpack the findings and further identify the individual characteristics of students failing to meet math standards, next steps could explore the interaction of gender and language profile on math achievement to determine whether the outperforming of males to females was a function of their language learning. Finally, while these results provided important information on the characteristics of students most likely to experience difficulty meeting EQAO math standards, they did not definitively explain why almost half of these students met the standards in Grade 3 yet were unable to do so again in Grade 6. Future studies should work to unpack this problem further to determine what changes between Grades 3 and 6 could have resulted in a decline in achievement.

Conclusion

This study shed light onto Ontario's declining math achievement patterns by identifying two distinct achievement trends and further determining the association of students' linguistic factors with the declining math achievement pattern. The study shows that interpreting math test results based on overall percentages of students meeting the standards is misleading and of little use. Not all students exhibit declines in math achievement. The study suggests that declines may be associated with specific subgroups of students, especially those who need more support in schools. The current study calls first for a reassessment of the declining patterns of achievement through more detailed identification of students who are struggling to meet standards and their individual characteristics that differentiate them from those who do not struggle to meet the same standards.

We call for further research to examine the extent to which EQAO math tests are culturally and linguistically sensitive to students from diverse backgrounds. Finally, current discourse around what policy makers have referred to as "Ontario's math problems" requires more systematic empirical evidence to determine its cause and effective intervention strategies. Any policy implementation, such as the new teacher-candidate math test mandate, and resource allocations without empirical evidence, may not adequately address the math problem with sustainable effects.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, *8*(3), 231–257. https://doi.org/10.1207/S15326977EA0803_02

Abedi, J., & Dietel, R. (2004). Challenges in the No Child Left Behind act for English-Language learners. *Phi Delta Kappan*, *85*(10), 782–785. https://doi.org/10.1177/003172170408501015

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234. http://dx.doi.org/10.1207/S15324818AME1403_2

Anderson, J. O., Rogers, W. T., Klinger, D. A., Ungerleider, C., Glickman, V., & Anderson, B. (2006). Student and school correlates of mathematics achievement: Models of school performance based on pan-Canadian student assessment. *Canadian Journal of Education*, *29*(3), 706–730.

de Araujo, Z., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, *88*(6), 879–919. https://doi.org/10.3102/0034654318798093

Barr, A. B. (2015). Family socioeconomic status, family health, and changes in students' math achievement across high school: A mediational model. *Social Science & Medicine*, *140*, 27–34. https://doi.org/10.1016/j.socscimed.2015.06.028

Cheng, L., & Sun, Y. (2015). Interpreting the impact of the Ontario secondary school literacy test on second language students within an argument-based validation framework. *Language Assessment Quarterly*, *12*, 50–66. https://doi.org/10.1080/15434303.2014.981334

Chval, K. B., & Chavéz, O. (2011). Designing math lessons for English Language Learners. *Mathematics Teaching in the Middle School*, *17*(5), 261–265. https://doi.org/10.5951/mathteacmiddscho.17.5.0261

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1–29.

Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons.

Cummins, J. (2014). To what extent are Canadian second language policies evidence-based? Reflections on the intersections of research and policy. *Frontiers in Psychology*, *5*(358), 1–10. https://doi.org/10.3389/fpsyg.2014.00358

Dearing, E., Walsh, M. E., Sibley, E., Lee-St. John, T., Foley, C., & Raczek, A. E. (2016). Can community and school-based supports improve the achievement of first-generation immigrant children attending high-poverty schools? *Child Development*, *87*(3), 883–897. https://doi.org/10.1111/cdev.12507

DiGisi, L. L., & Fleming, D. (2005). Literacy specialists in math class! Closing the achievement gap on state math assessments. *Voices from the Middle*, *13*(1), 48–52.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Education Quality and Accountability Office. (2003). *Annual Report: 2002-2003*. Author. Queen's Printer for Ontario.

Education Quality and Accountability Office. (2013). *What Is the Quality of EQAO Assessments? Excellent!* Queen's Printer for Ontario.

Education Quality and Accountability Office. (2014). *EQAO's technical report for the 2013-2014 assessments*. Queen's Printer for Ontario.

Education Quality and Accountability Office. (2017a). *Annual Reports: 2016-2017*. Author. Queen's Printer for Ontario.

Education Quality and Accountability Office. (2017b). *The Assessments*. http://www.eqao.com/en/assessments

Education Quality and Accountability Office. (2017c). *Annual Report: 2015-2016*. Author. Queen's Printer for Ontario.

Education Quality and Accountability Office. (2017d). *About the agency*. http://www.eqao.com/en/about_eqao/about_the_agency/Pages/about-the-agency.aspx

Education Quality and Accountability Office. (2018). *Annual Report: 2017-2018*. Author. Queen's Printer for Ontario.

Education Quality and Accountability Office. (2019a). *Math Proficiency Test*. https://eqao.wpengine.com/?page_id=45

Education Quality and Accountability Office. (2019b). *EQAO's technical report for the 2016-2017 assessments*. Queen's Printer for Ontario.

Elliott, L., & Bachman, H. J. (2018). How do parents foster young children's math skills? *Child Development Perspectives*, *12*(1), 16–21. https://doi.org/10.1111/cdep.12249

Galindo, C., & Sonnenschein, S. (2015). Decreasing the SES math achievement gap: Initial math proficiency and home learning environments. *Contemporary Educational Psychology*, *43*, 25–38. https://doi.org/10.1016/j.cedpsych.2015.08.003

Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*(12), 182–193. https://doi.org/10.1016/j.lindif.2016.01.00

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Pro-Ed.

Giouroukakis, V., & Honigsfeld, A. (2010). High-stakes testing and English language learners: Using culturally and linguistically responsive literacy practices in the high school English classroom. *TESOL Journal*, *1*, 470–499. https://doi.org/10.5054/tj.2010.240193

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, *66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hinton, A. (2014). *Tracking student achievement in mathematics over time in English-Language Schools: Grade 3 (2006) to grade 6 (2009) to grade 9 (2012) cohort*. Queen's Printer for Ontario.

Holmes, K., Gore, J., Smith, M., & Lloyd, A. (2018). An integrated analysis of school students' aspirations for stem careers: Which student and school factors are most predictive? *International Journal of Science and Mathematics Education*, *16*(4), 655–675. https://doi.org/10.1007/s10763-016-9793-z

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization*, *7*, 24–52. https://www.jstor.org/stable/764957

Honigsfeld, A., & Giouroukakis, V. (2011). High-stakes assessments and English Language Learners. *The Delta Kappa Gamma Bulletin*, *77*(4), 6–10.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, *89*(5-6), 761–796. https://doi.org/10.1016/j.jpubeco.2004.08.004

Jang, E. E., & Sinclair, J. (2018). Assessment in education. *Principles, Policy & Practice*, *25*(6), 655–677. https://doi.org/10.1080/0969594X.2017.1329705

Jang, E. E., Cummins, J., Wagner, M., Stille, S., Dunlop, M., & Starkey, J. (2015). Assessing school-aged English language learners in Ontario curriculum learning contexts using Steps to English Proficiency. *Language Assessment Quarterly*, *12*(1), 87–109.

Jang, E. E., Dunlop, M., Wagner, M., Kim, Y.-H., & Gu, Z. (2013). Elementary school ELLs' reading skill profiles using cognitive diagnosis modeling: Roles of length of residence and home language environment. *Language Learning*, *63*(3), 400–436.

Jang, E. E., & Ryan, K. (2003). Bridging gaps among curriculum, teaching and learning, and assessment. *Journal of Curriculum Studies*, *35*(4), 499–512.

Jordan, N., Hanich, L., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*(3), 834–850. https://doi.org/10.1111/1467-8624.00571

Jordan, N., Kaplan, D., & Hanich, L. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology*, *94*(3), 586–597. https://doi.org/10.1037/0022-0663.94.3.586

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867. https://doi.org/10.1037/a0014939

Khalvandi, F., & Chenari, M. (2012). Teachers' attitudes to test-driven accountability; Developing a comprehensive approach to assessment. *Procedia - Social and Behavioral Sciences*, *69*, 1548–1554. https://doi.org/10.1016/j.sbspro.2012.12.097

Kim, Y.-H., & Jang, E. E. (2009). Differential functioning of reading subskills on the OSSLT for L1 and ELL students: A multidimensionality model-based DBF/DIF approach. *Language Learning*, *59*(4), 825–865.

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*(3), 155–178. https://doi.org/10.1080/19345740802114533

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research*, *32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: A.k.a. the remix. *Harvard Educational Review*, *84*(1), 74–84. https://doi.org/10.17763/haer.84.1.p2rj131485484751

Lager, C. A. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology*, *27*(2-3), 165–204. https://doi.org/10.1080/02702710600642475

Lanza, S., & Rhoades, B. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science*, *14*(2), 157–168. https://doi.org/10.1007/s11121-011-0201-1

Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, *78*(3), 608–644. https://doi.org/10.3102/0034654308324427

Lee, J., Moon, S., & Hegar, R. (2011). Mathematics skills in early childhood: Exploring gender and ethnic patterns. *Child Indicators Research*, *4*(3), 353–368. https://doi.org/10.1007/s12187-0109088-9

Lerkkanen, M. K., Rasku-Puttonen, H., Aunola, K., Nurmi, J. E. (2005). Mathematical performance predicts progress in reading comprehension among 7-year olds. *European Journal of Psychology of Education*, *20*(2), 121–137. https://doi.org/10.1007/BF03173503

Loveless, T. (Ed.) (2007). *Lessons learned: What international assessments tell us about math achievement*. Brookings Institution Press.

Martiniello, M. (2009). Language and the performance of English-Language Learners in math word problems. *Harvard Educational Review*, *78*(2), 333–368. https://doi.org/10.17763/haer.78.2.70783570r1111t32

Mcgloin, M. (2011). *An achievement gap revealed: A mixed method research investigation of Canadian-born English language learners* [Unpublished master's thesis]. University of Toronto.

Merolla, D. M. (2017). Self-efficacy and academic achievement. *Sociological Perspectives*, *60*(2), 378–393. https://doi.org/10.1177/0731121416629993

Moschkovich, J. N. (2015). Scaffolding student participation in mathematical practices. *ZDM: International Journal on Mathematics Education*, *47*(7), 1067–1078. https://doi.org/10.1007/s11858-015-0730-3

Muthén, L. K., & Muthén, B. O. (2017). *Mplus: The comprehensive modeling program for applied researchers (version 8)*. Muthén & Muthén.

Neal, D., & Whitmore Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, *92*(2), 263–283. https://doi.org/10.1162/rest.2010.12318

Nylund-Gibson, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy*. OECD.

Ontario Ministry of Education. (2005). *The Ontario Curriculum Grades 1-8: Mathematics*. Queen's Printer for Ontario.

Ontario Ministry of Education. (2006). *The Ontario Curriculum Grades 1-8: Language*. Queen's Printer for Ontario.

Ontario Ministry of Education. (2019). First year investment of Ontario's four-year math strategy announced: Major commitment to math improvement will promote student success and teaching excellence by going back to basics. *Government of Ontario: Newsroom*. https://news.ontario.ca/edu/en/2019/08/first-year-investment-of-ontarios-four-year-math-strategy-announced.html?_ga=2.253946894.784931949.1567625852-2025985994.1567625852

Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., & Tang, J. H. (2016). *Highlights from TIMSS and TIMSS Advanced 2015: Mathematics and Science Achievement of U.S. Students in Grades 4 and 8 and in Advanced Courses at the End of High School in an International Context* (U.S. Department of Education Publication NCES 2017-002). National Center for Education Statistics.

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*, 1301–1308. https://doi.org/10.1177/0956797612466268

Roberts, S. A. (2009). Supporting English language learners' development of mathematical literacy. *Democracy & Education*, *18*(3), 29–36.

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, *46*(5), 995–1007. https://doi.org/10.1037/a0018880

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. *California School Psychologist*, *14*, 23–34. https://doi.org/10.1007/BF03340948

Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, *9*(3), 15–22. https://doi.org/10.1111/j.1745-3992.1990.tb00374.x

Sinclair, J., Jang, E. E., & Vincett, M. (2018). Investigating linguistically diverse adolescents' literacy trajectories using latent transition modeling. *Reading Research Quarterly*, *54*(1), 81–107. https://doi.org/10.1002/rrq.220

Soni, A., & Kumari, S. (2017). The role of parental math anxiety and math attitude in their children's math achievement. *International Journal of Science and Math Education*, *15*(2), 331–347. https://doi.org/10.1007/s10763-015-9687-5

Steedle, J. T., & Grochowalski, J. (2017). The effect of stakes on accountability test scores and pass rates. *Educational Assessment*, *22*(2), 111–123. https://doi.org/10.1080/10627197.2017.1309276

Steffani, S., & Selvester, P. M. (2009). The relationship of drawing, writing, literacy, and math in kindergarten children. *Reading Horizons*, *49*(2), 125–142.

Sucuoğlu, E. (2018). Economic status, self-efficacy and academic achievement: The case study of undergraduate students. *Quality & Quantity*, *52*(S1), 851–861. https://doi.org/10.1007/s11135-018-0692-y

Swanson, H. L., Olide, A. F., & Kong, J. E. (2018). Latent class analysis of children with math difficulties and/or math learning disabilities: Are there cognitive differences? *Journal of Educational Psychology*, *110*(7), 931–951. http://dx.doi.org/10.1037/edu0000252

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, *115*(2), 227–244. https://doi.org/10.1016/j.jecp.2013.02.002

Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, *45*(2), 107–122. https://doi.org/10.1080/00461521003703060

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Zhang, J., Fan, X., Cheung, M. K., Meng, Y., Cai, Z., & Hu, B. Y. (2017). The role of early language abilities on math skills among Chinese children. *PLoS ONE*, *12*(7), 1–14. https://doi.org/10.1371/journal.pone.0181074

Zheng, Y., Cheng, L., & Klinger, D. A. (2007). Do test formats in reading comprehension affect second-language students' test performance differently? *TESL Canada Journal*, *25*(1), 65–80. https://doi.org/10.18806/tesl.v25i1.108