Trajectories in Student Mathematics Performance

Research Partnerships Program

Final Report

L. McGarvey, O. Bulut, Q. Guo, E. Simmt, and M. Gierl

University of Alberta

In collaboration with

P. Wozny, Aurora Academic Charter School

T. Reid, Black Gold Regional Schools

K. Barber & J. Drent, Chinook's Edge School Division

C. Campbell, Edmonton Public Schools

D. Barron & A. Ortigosa, Elk Island Public Schools

C. Coyne & S. Rudakoff, St. Albert Public Schools

Abstract

This report presents the results of a retrospective longitudinal quantitative study to analyze Provincial Achievement Test (PAT) data to identify potential predictors of mathematics performance across the Grades 3, 6, and 9 exams. Using a range of data analysis techniques, we explored individual characteristics and mathematics content strands that predict future performance. Our results indicate a general trend of decline in performance from Grades 3 to 9. Within this trend, we noted that students coded in Grade 3 as English as a Second Language learners (ESL) and French Immersion, and students from large city centres had a slightly less likelihood of decline in performance at the Grade 9 level. We also identified a significant change in the predictors of performance with regard to mathematics content. For our cohorts of students who experienced the 1996 curriculum, the Number strand was the best predictor of future performance. However, for students who experienced the 2007 curriculum, Patterns & Relations appeared to be a stronger predictor of future performance.

Keywords: mathematics; provincial achievement tests, curriculum change

Table of Contents

Trajectories in Student Mathematics Performance

Introduction

Rationale for the Research

In the past few years, there have been multiple reports about the decline in student mathematics achievement in our province using evidence from Provincial Achievement Tests (PAT) and the Programme for International Student Achievement (PISA) (e.g., CBC, 2016). In addition, there is a historically consistent decline in levels of acceptable performance between the Grade 6 PAT and the Grade 9 PAT (Alberta Government, 2016a; Simmt et al., 1999). Over the past few years there have been several attempts by Alberta Education to reverse the trend by, for example, providing bursaries to preservice and inservice teachers for courses in mathematics pedagogy (Alberta Government, 2016b), offering researcher partnership grants to improve student learning in mathematics (Alberta Education, 2017a), through revisions to and clarifications of the *Mathematics K-9 Program of Studies* (Alberta Education, 2017b), and by emphasizing numeracy across the curriculum (e.g., Alberta Education, 2017c). In response to the attention towards mathematics, many Alberta school districts are now emphasizing numeracy as a district wide priority and elementary mathematics professional learning is an Alberta Regional Consortia provincial initiative (see arpdc.ab.ca).

Prior Research

Mathematics achievement has long term implications for education, career, and financial success for people in today's technology driven society (National Mathematics Advisory Panel, 2008; Ritchie & Bates, 2013). As a result, there is a growing body of research dedicated to identifying early predictors of future performance in mathematics. Understanding the factors that predict future mathematics performance is important for providing opportunities and support for children and youth to succeed, and for developing curriculum that provides the necessary foundation for learning.

Research suggests that individual mathematics performance is often relatively stable over several years, especially in comparison to other academic subjects (Duncan et al., 2007). Unfortunately, this means that children whose performance is poor in the early years will likely continue to perform poorly over time. Lee (2012) noted a related result that regardless of the range of possible predictive factors, mathematics achievement alone, from prekindergarten to high school, was a good predictor of a student’s access to and successful attainment of a postsecondary bachelor’s degree.

Other research streams attempt to identify specific content areas within mathematics that may predict future achievement. Siegler et al. (2012) investigated secondary data sets from arithmetic and fractions tests administered to 10-year-olds and also applied problems on tests taken by 15-year-olds in the United States and the United Kingdom. The findings from that study indicated that fractions and whole-number division content predicted high school mathematics achievement better than other arithmetic content, general intelligence scores, working memory, and family background. They hypothesized that knowledge of fractions and whole-number division were both necessary for solving algebra problems, and that both fractions and division are often more challenging for students and may require more advanced mathematical thinking than addition, subtraction, or multiplication. In related studies, Geary, Hoard, Nugent and Bailey, (2013) reported that kindergarteners’ number system knowledge was predictive of middle school mathematics performance. Bailey, Siegler, and Geary (2014) also found that whole-number knowledge in Grade 1 predicted fraction knowledge in Grades 7 and 8. Claessens and Engel

(2013) found that early mathematics skills in number, pattern recognition, and measurement were predictive of reading, mathematics and science achievement through to eighth grade.

Another body of research is revealing the importance of spatial skills to areas of science, technology, engineering, and mathematics (STEM). Spatial skills have been found to explain academic success in STEM subjects throughout schooling even beyond measures of verbal and quantitative scores (Mix & Cheng, 2012; Webb, Lubinski, & Benbow, 2007). Further, Wai, Lubinski and Benbow (2009) found that individuals with strong spatial skills were more likely to enter, enjoy, and succeed in STEM related careers.

The search for relevant predictors of mathematics achievement is a growing field of study. Much of this research involves retrospective longitudinal studies using large data sets of assessment scores collected at different periods of time. Very little of this longitudinal research involves Canadian students. The limited availability of standardized tests at different time periods has likely made this type of research difficult. However, we expect that implications of achievement in mathematics on graduation rates, postsecondary completion, meaningful employment, and economic independence is similar for Canadian students.

Project Background

As educators, we believe it is of significant value to understand the nature of change in student performance from Grade 3, to Grade 6, to Grade 9, and the potential indicators of early performance that are predictive of performance in Grade 9. That is, does poor performance on number computation tasks in Grade 3 predict poor performance in algebra tasks in Grade 9? Does excellent performance on recall and procedural tasks in Grade 6 predict improved or strong performance in problem solving and analysis tasks later on? Unless we have an understanding of student performance beyond overall average scores, where should we put our efforts in the classroom, in teacher education, in professional development, or in curriculum development?

Through a partnership between researchers at the University of Alberta and Alberta school authority leaders, we sought to investigate the trajectories of student performance with the hope of providing suggestions for mathematics teacher education, professional development, and curriculum within Alberta.

Research Questions

The following research questions guided the project:

- What are the trajectories of student performance in mathematics from Grade 3 to Grade 9?
- What individual identifiers or characteristics, such as gender, urban/rural location, and/or special education codes predict future performance?
- What aspects of mathematics content predict future performance?

The results of investigating these questions may provide insight into student performance over time and have implications for teacher education, professional development, and curriculum.

Practitioner – Researcher Collaboration

Research Partnership

Like many partnerships, the practitioners and researchers for this project were formed by drawing on relationships from other projects. The initial idea for the project was generated by researchers Drs. McGarvey and Simmt. As mathematics educators at the University of Alberta, they have worked on multiple projects together for over two decades. However, this project, which used PAT data, required a different skill set for analysis. Drs. Gierl and Bulut, colleagues from the Centre for Measurement and Assessment, were contacted to see if they would be willing to participate and contribute to the methodology and data analysis. They agreed and we were also fortunate that Guo, Bulut's doctoral student, joined the research team later on. Simmt had previously worked with Gierl on a previous project, but this was the first time that the other researchers had worked together.

Once the initial project idea had been formulated, Simmt and McGarvey drew on their relationships with school authority personnel to invite them to contribute to the project. For example, Elk Island and St. Albert partners, Barron and Ortigosa, and Coyne and Rudakoff respectively, had worked with McGarvey, Simmt, and other University of Alberta Centre for Mathematics Science and Technology (CMASTE) mathematics educators to develop and implement *Math Academies*, a professional development opportunity for their district teachers. In another project, CMASTE mathematics educators, were asked by Campbell, Edmonton Public Schools, to contribute to district resource development. Wozny, formerly of Black Gold and presently at Aurora Academic Charter School, had worked with Simmt on previous projects. Simmt has also been working with Chinook's Edge in a Researcher Partnership Project that they had initiated. The partnership for this project then was formed primarily by continuing relationships previously established through other projects rather than through geographic selection.

Research Partnership – Celebrations and Challenges. The project brought together practitioners with diverse leadership responsibilities within their respective schools including superintendents, learning services and supports persons, school administrators, and content consultants. The researchers were from two distinct fields of educational research including mathematics education and educational measurement. The broad range of skills allowed each person to offer a unique perspective to the project, analysis of results, and potential implications.

Although each of the school partners knew at least one of the researchers, several of them had not previously met. The project provided an opportunity for us all to extend our professional network. While the current project drew on prior relationships, the extended network has the potential to contribute to new partnerships in research and professional development in the future.

Another implicit, but key set of partners in the project were members of Alberta Education who helped us navigate some of the challenges we faced getting access to the PAT data, selecting appropriate cohorts for analysis, and anonymizing the student and school data.

Despite the advantages of a large and diverse team, it also had its challenges. Unfortunately, our team was only able to have one face-to-face meeting held in April 2018. Even the research team of five people, all working within the same building, had difficulty scheduling meeting times in which everyone could attend. As a result, Bulut, Guo, and McGarvey took primary responsibility for the research activities associated with the project. Similarly, our practitioners also had busy work schedules; the research project was described by a couple of partners as something they were interested in, but that existed on the side of their other leadership responsibilities. After our face-to-face meeting in April, we requested additional information from participants regarding the analysis presented, but we did not receive responses from every member.

Another challenge may have been the project itself. Initially, we assumed we would receive PAT data only from partner school authorities. Although we were fortunate to have received anonymized province-wide data, the results reflected students across the province, and may not have held information that was specific enough for participating school districts to invest time to the project. Certainly, a frequent suggestion from the practitioner partners was to examine or collect data from specific schools or school authorities. Unfortunately, such requests were beyond the scope of the project. Having said that, the methodology could be replicated by school districts themselves.

Research Partnerships Program Cohort

The Research Partnerships Program (RPP) provided several opportunities for us to meet with other grant holders and to share progress made. Given that our cohort responded to a specific call for proposals, there was significant overlap in the topics allowing us to learn about the multiple ways in which practitioners and researchers worked together to address complex educational problems. The challenge, of course, was coordinating schedules to ensure that our project team was represented at the meetings and could discuss progress made to date.

Lessons Learned

Overall, the project was a positive experience. It brought together educators who would not normally work together on a common project and it developed a methodology for studying trajectories in student performance which could be used by specific school districts or with other content areas. In future projects, a higher degree of collaboration of all partners on the development of the proposal and the initial stages would likely have led to a more cohesive partnership. In addition, a smaller team in which to develop and conduct the project may have helped in its coordination.

Next Steps

Although we have completed the research component of the project, we are looking forward to disseminating the results to a broader audience through presentations and publications for both professional and academic audiences. Once the project report is finalized we will share it with our partner school authorities and the researchers will be available to discuss it with district personnel upon request. In October 2019, we will be presenting the results of the project at the annual meeting of Mathematics Council of the Alberta Teachers' Association (MCATA). CMASTE will also host a symposium that will bring together teacher educators, researchers, mathematics leads, and district administration from across the province to share the results of the project and discuss possible implications for mathematics teaching and learning in Alberta. From the results and our interactions in these forums, we will prepare a paper for delta-K, a professional journal for mathematics teachers in Alberta. For academic audiences, we have the opportunity to prepare a poster presentation for the Canadian Mathematics Education Study Group's (CMESG) annual conference for spring 2020, present at the Fields Institute for Research in Mathematical Sciences at the University of Toronto, and submit a paper for the Canadian Journal for Science, Mathematics and Technology Education (CJSMTE).

Although we addressed the research questions posed for the project, we felt we merely scratched the surface in analyzing the data available to us. Our results were based on all students using the broad strands of mathematics (i.e., Number, Patterns and Relations, Shape and Space, and Statistics and Probability). An analysis at the level of specific outcomes or more refined categories of mathematics content, processes, or levels of complexity may help further identify predictors of student performance over time. Doing so may require assistance from Alberta Education for exploring equating methods for smaller items sets or an alternative methodology to compare individual student performance for Grade 3, 6, and 9.

Research Design

Methodological Framework

This large-scale retrospective longitudinal study used existing data from the PATs over three time points: Grades 3, 6, and 9. The study involved three student cohorts following two relatively different curriculum frameworks. The data allowed for both within and between cohort comparisons of individual student achievement results.

With secondary data analysis it is essential to apply theoretical knowledge and conceptual skills to leverage existing data sets to address new research questions. The following research questions guided the project:

- What are the trajectories of student performance in mathematics from Grade 3 to Grade 9?
- What individual identifiers or characteristics, such as gender, urban/rural location, and/or special education codes predict future performance?
- What aspects of mathematics content predict future performance?

The results of the study provide insight into student performance over time and have implications for teacher education, professional development, and curriculum.

A range of longitudinal data techniques described below were used to address the research questions. In particular, we were interesting in tracking and examining trends in individual student's mathematics scores rather than reviewing aggregate scores for a school, district or province.

Data Sources

The data analyzed was collected by Alberta Education through mathematics PATs. Currently, PATs are administered annually to all Alberta students in grades 6 and 9. Previously, PATs were also administered to all Alberta students in grade 3. Working with Alberta Education Research Branch, Analytics Branch, and Assessment Branch we identified the years for three student cohorts with associated PAT scores for Grades 3, 6, and 9 (see Table 1).

Table 1

Student Cohorts and Year of Writing the Grades 3, 6, and 9 Provincial Achievement Tests

	Grade 3	Grade 6	Grade 9
Cohort 1	2001-2002	2004-2005	2007-2008
Cohort 2	2002-2003	2005-2006	2008-2009
Cohort 3	2010-2011	2013-2014	2016-2017

Selecting cohorts founded on consistent conditions proved to be challenging due to a range of circumstances. First, there was a significant change in the Kindergarten to Grade 9 Mathematics Program of Studies in 2007 with staggered implementation occurring over multiple years. Second, a change to the PAT occurred in 2011 when the "Part A" or the non-calculator computation portion of the exam was eliminated. Third, in June 2013 Southern Alberta experienced a devastating flood. Schools were closed and thousands of students did not write the PAT in that year. As a consequence, we worked backwards from 2017 to identify three cohorts with sufficient coherence across all three test points. Using our general criteria, there were no other potential cohorts between the ones shown in Table 1. In general, the first two cohorts are similar being only one year apart. They both followed the 1996 K-9 Mathematics Program of Studies. The third cohort followed the 2007 K-9 Mathematics Program of Studies.

Once the cohort years were identified, we worked the Research Branch and Analytics Branch to develop codes to categorize schools to ensure the student data received was properly anonymized. We used the first three characters of the schools' postal codes to determine (1) location in the province including northern, central, or southern Alberta and, (2) the population of the community or region in which the schools were situated. Regions included: large cities distinguishing between Edmonton and Calgary; small cities from 25 000 to 100 000 people (e.g., Airdrie, Red Deer, Grande Prairie); towns under 25 000 people (e.g., Brooks, Camrose, St. Paul), and rural areas identified with zero as the second character in the school postal code (e.g., T0K). We received anonymized PAT data for every student in the province who wrote the exam for the grade and year indicated in Table 1. Using the total numbers of students for the Grade 3 exams, we received data for 142 689 students across all three cohorts.

Our data sources included individual student characteristics and their scores on each item and subdomains of the PATs. We used assessment-related variables (e.g., Skills/Knowledge questions; content strand) and student characteristics (e.g., gender, urban/rural, special education codes) as explanatory variables to help identify and account for changes in student performance over time. We used a two-level hierarchical linear growth model to predict the students' mathematics scores with time and assessment-related factors at the first level and student characteristics at the second level. The longitudinal analysis of student achievement data allows us to create longitudinal profiles of students and identify changes in students' strengths and weaknesses in mathematics over time.

Cohort characteristics. Although we received scores for more than 47 000 in each cohort, we eliminated students who did not have scores in all three grades or who had moved to other coded regions (e.g., moved from a town to a small city). After the removal of these students, Cohorts 1, 2 and 3 had 28 446, 29 214, and 29 486 students respectively.

Table 2 represents the composition of students in each of the three cohorts according to the student codes associated with each student. We chose to use only the codes recorded for the students at the time they wrote the Grade 3 PAT. This choice may have some minor implications for the results given the potential changes to student codes in Grades 6 and 9. For example, students may have changed their status associated with French Immersion or ESL over the three points in time.

The results illustrate a relatively even gender split that is consistent across all three cohorts. French Immersion and Francophone are similar between Cohorts 1 and 2, but increase slightly eight years later for Cohort 3. Students who were coded as English as a Second Language (ESL) Funded and ESL Canadian showed a more significant increase in numbers between Cohorts 1/2 and 3 with more than double the number of students with this coded characteristic. There is a reduction in Special Education coded students who wrote the test in Cohort 3 compared to Cohorts 1 and 2. The number of students who took their courses primarily online was very small with less than 1% of the cohort students.

Table 2

Characteristics of Student across the Three Cohorts

Characteristics	Cohort 1		Cohort 2		Cohort 3	
Female	14 200	(49.9%)	14 507	(49.7%)	14 686	(49.8%)
Male	14 246	(50.1%)	14 707	(50.3%)	14 800	(50.2%)
French Immersion	1 841	(6.5%)	1 933	(6.6%)	2 451	(8.3%)
Francophone	232	(0.8%)	197	(0.7%)	373	(1.3%)
ESL Funded	540	(1.9%)	659	(2.3%)	1 876	(6.4%)
ESL Not Funded	22	(0.08%)	15	(0.05%)	42	(0.14%)
ESL Canadian	1 071	(3.8%)	1 271	(4.4%)	2 668	(9.0%)
Special Education	2 538	(8.9%)	2 445	(8.4%)	1 780	(6.0%)
Online Program	22	(0.08%)	36	(0.12%)	14	(0.05%)
Totals	28 446		29 214		29 486	

Note. The coded characteristic is based on a student's Grade 3 Provincial Achievement Test. The majority of students do not have a coded characteristic and some students may have more than one coded characteristic.

The size of each cohort is relatively the same as is the gender division, and the majority of students did not have an additional code associated with their Grade 3 exam. However, Cohort 3 represents a more diverse sample. While approximately 22% of students in Cohorts 1 and 2 had an additional code for French Immersion, Francophone, ESL, Special Education or Online Program, over 41% of students in Cohort 3 were coded in some way.

Table 3 shows the numbers of cohort students in each of the five regions we developed for the study. The numbers of students in each of the areas is relatively consistent across all three cohorts. More than half of the students lived in either Calgary or Edmonton; approximately one-third lived in a small city or town; the remainder lived in rural communities.

Table 3

Distribution of Students in the Provincial Regions

Region	Cohort 1	Cohort 2	Cohort 3
Large Cities (>100K)			
Calgary	8435 (29.7%)	8779 (30.1%)	9029 (30.6%)
Edmonton	5997 (21.1%)	6175 (21.1%)	6329 (21.5%)
Small Cities (25K - 100K)	5361 (18.8%)	5520 (18.9%)	6027 (20.4%)
Towns (<24K)	3790 (13.3%)	3840 (13.1%)	3724 (12.6%)
Rural	4863 (17.1%)	4900 (16.8%)	4377 (14.8%)

Note. Regions were determined using the first three characters of a school's postal code.

Data Analysis and Results

Trajectories of Student Performance

The first research question for the study was, 'What are the trajectories of student performance in mathematics from Grade 3 to Grade 9?' To gain insight into this question based on student scores we conducted a growth curve analysis.

Growth curve analysis. In order to conduct growth curve analyses, students' Grade 3, 6 and 9 mathematics scores needed to be equated. However, since the tests were not developed based on any vertical scaling design and there were no common items across the three grades, traditional equating could not be used. Instead, we chose to centre each grade's mathematics score based on the proficiency cut scores provided by Alberta Education, and then divided the centred scores by their standard deviation. The resulting standardized scores represent how many standard deviations a student is above or below the proficiency cut score.

Each question on the PATs was classified as either "knowledge" or "skills." For Cohorts 1 and 2, growth curve analyses were conducted for knowledge and skills questions separately. Knowledge questions involved aspects such as recalling facts and terminology, recognizing place value, and knowing procedures for computation, measurements, and geometric constructions (see Alberta Education Subject Bulletins for Mathematics prior to 2009). Skill questions included the application of concepts, judging the reasonableness of solutions, justifying answers, and demonstrating relationships in numbers and space. After 2009, proficiency cut scores were not available for knowledge and skills questions. Therefore, for Cohort 3 the growth curve analysis was conducted using the total score only.

To determine what type of knowledge and skills growth curves existed for Cohorts 1 and 2 (e.g., linear or nonlinear), we conducted our analysis based on gender and region. Examples of growth curve plots based on Cohort 1 are presented in Figures 1 to 4. They include knowledge-gender (Figure 1), knowledge-region (Figure 2), skills-gender (Figure 3), and skills-region (Figure 4).

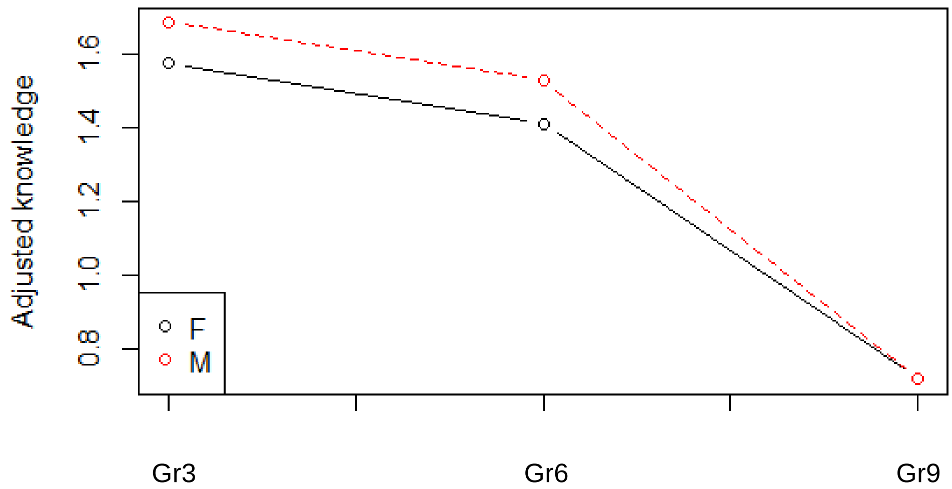

Figure 1. Cohort 1 Knowledge growth curve by gender. Female-black; Male-red.

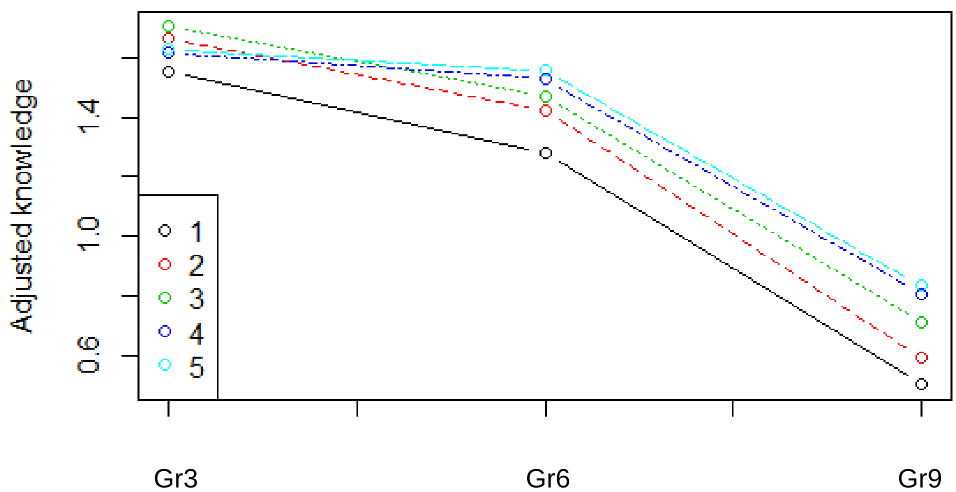

Figure 2. Cohort 1 Knowledge growth curve by regions in Alberta. Rural-black; Town-red; Small City-green; Edmonton-dark blue; and Calgary-light blue.

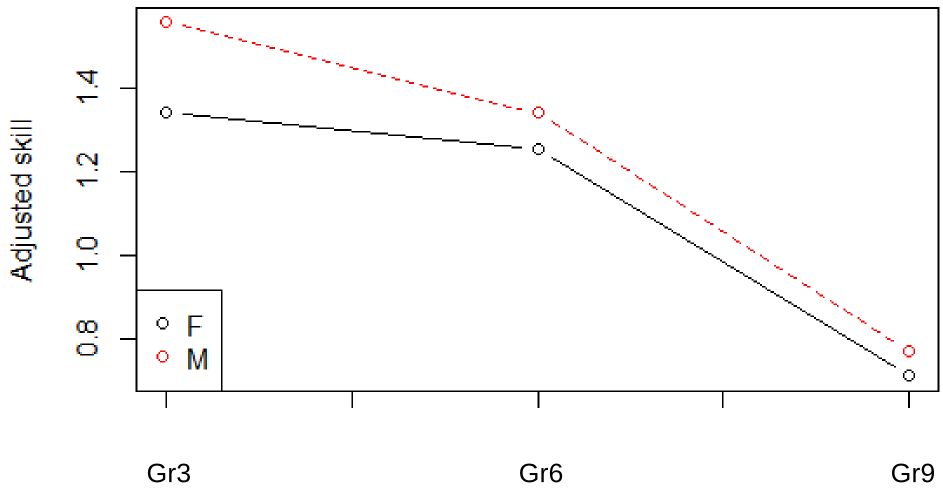

Figure 3. Cohort 1 Skills growth curve by gender. Female-black; Male-red.

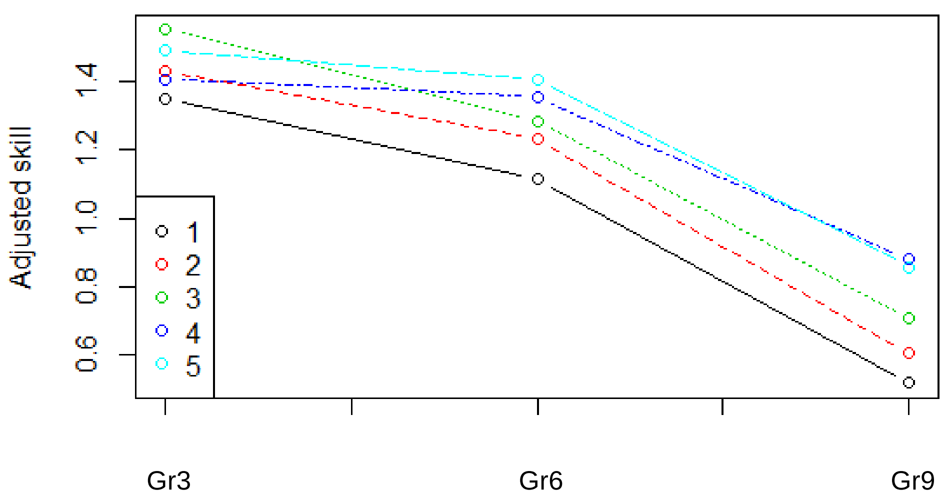

Figure 4. Cohort 1 Skills growth curve by regions in Alberta. Rural-black; Town-red; Small City-green; Edmonton-dark blue; and Calgary-light blue.

We show the results for Cohort 1 only because the other two cohorts presented very similar results with downward trending growth curves regardless of the type of question (i.e., knowledge, skills) or code used (i.e., gender, region). As shown in the figures, the growth curves tended to be nonlinear and downward trending for both knowledge and skills questions. In general, there was a minor decrease in the proficiency from Grade 3 to Grade 6, but a much larger decrease in proficiency from Grade 6 to Grade 9.

Within the figures, we noted some additional trends. For both knowledge and skills questions, male students tended to do slightly better than female students at Grade 3 and Grade 6, but this difference diminished at Grade 9. For regions in Alberta, schools located in communities with larger populations tended to do better than schools in regions with smaller populations. An exception to this finding is that Edmonton and Calgary students' Grade 3 performance were slightly lower than other smaller cities. Overall, these the growth curve plots suggest that nonlinear growth curve analysis may be more suitable for the datasets. Since there were only three time points (i.e., Grade 3, 6, and 9), nonlinear growth curve analysis was equivalent to examining what factors predict the performance at Grade 3, what factors predict the differences in scores between Grades 3 and 6, and what factors predict the differences in scores between Grades 6 and 9.

Cohort 1 growth curve analysis results. Growth curve analyses were conducted separately for mathematics knowledge and skills. For knowledge, gender significantly predicted students' performances at Grade 3 ($b=.136$, $p<.001$), the change in performance between Grade 3 and Grade 6 ($b=.087$, $p<.001$), and the change in performance between Grade 6 and Grade 9 ($b=-.060$, $p<.001$). This means that male students did better in mathematics than female students at Grade 3, they also showed less decrease between Grade 3 and Grade 6; however, between Grade 6 and Grade 9, male students' performance dropped more than female students.

Region in Alberta was another significant predictor. At Grade 3, Calgary students appeared to perform better than other regions in Alberta, with two exceptions. Performance of students from Alberta towns (i.e., population less than 25 000) was not significantly different from Calgary, and students from Northern Alberta cities outperformed Calgary (b=.133, p<.001). However, all the other regions showed significantly more decrease in performance than Calgary between Grade 3 to 6, and between Grade 6 to 9. A third important predictor was the special education code. As expected, students who had a special education code tended to perform relatively poorly at Grade 3 (b=-.519, p<.001); they showed a larger decrease in performance between Grade 3 and Grade 6 (b=-.379, p<.001), and also from Grade 6 to Grade 9 (b=-.289, p<.001).

The growth curve analysis pointed to a number of other key factors related to predicting student performance over time. Students who changed schools, but stayed within the coded region significantly predicted change in performance. Students who changed schools between Grades 3 and 6 tended to show a larger decrease in performance (b=-.098, p<.001) than students who did not change schools. Similarly, students who changed schools between Grade 6 and 9 also showed a larger decrease in performance (b=-.103, p<.001). French Immersion students tended to perform worse than other students at Grade 3 (b=-.617, p=.001). However, French Immersion did not significantly predict the changes over time. ESL students who were funded tended to perform more poorly than other students at Grade 3 (b=-.4469, p<.001). However, their change in performance between Grade 3 and 6 was not significantly different from other students. Between Grades 6 and 9, ESL students who were funded tended to show less decrease than other students (b=.203, p<.001). For students who attended online programs, their performances were not significantly different from other students at Grade 3, nor did they show more change than other students between Grades 3 and 6. However, online program students' performance dropped significantly more than other students between Grade 6 and 9 (b = -.310).

For skills growth curve, most variables function similarly to knowledge growth curve with a few minor exceptions. For example, previously, we showed that Calgary had significantly less decrease in knowledge performance than all other regions. This was mostly true for skills growth curve with the exception that Southern Alberta small cities had significantly less decrease in skills than Calgary. Overall, the changes were minor.

Cohorts 2 and 3 growth curve analysis results. The growth curve analyses for Cohort 2 and 3 presented very similar patterns as the Cohort 1 analysis. The relationship between predictors and mathematics growth mostly remained the same with minor changes in the strength of the coefficients. For simplicity, the detailed regression coefficients will not be discussed further.

Characteristics Predicting Future Performance

The previous discussion illustrates that the majority of students declined in mathematics performance from Grades 3 to 9. Yet, we recognized that not every student experienced a decline in performance or followed the same growth curve. Our second research question was, “What individual identifiers or characteristics, such as gender, urban/rural location, and/or special education codes predict future performance?” In order to address this question we needed to understand what variations in growth trajectories existed and then the demographic characteristics of students whose performance followed the different trajectories.

To explore the different types of trajectories, we wanted to distinguish between students whose performance did not decline from students who did. An indicator variable was first constructed for each cohort to determine whether each student’s standardized mathematics skills score did or did not decline between Grade 3 and Grade 9. Table 4 shows the number of students who maintained or improved performance and those who declined in performance.

Table 4

Number of students who maintained or improved performance

Performance	Cohort 1	Cohort 2	Cohort 3
Maintained/Improved	6035 (21.2%)	5296 (18.1%)	9404 (31.2%)
Declined	22 411 (78.8%)	23 918 (81.9%)	20 082 (68.1%)
Totals	28 446	29 214	29 486

A decision tree algorithm was used to classify the indicator variable constructed in the previous step based on students' demographic characteristics. Decision tree algorithm was used because given that all the demographic variables were categorical, decision tree provides decision rules that help researchers understand the differences between students who declined and students who did not. Further, *k*-mean cluster analysis was used to cluster students' grade 3, 6, and 9 standardized mathematics skills scores. *K*-mean cluster analysis was chosen because of its computational efficiency for large sample size. The number of clusters were determined based using the elbow method. After the cluster analysis, the demographic characteristics for each cluster were visualized using bar charts. All the analyses were conducted using R (R Core Team, 2013). The decision tree was conducted using the *rpart* package; *k*-mean cluster analysis was conducted using the *NbClust* package. The bar charts (below) were constructed using the *ggplot2* package.

For cohort 1, decision tree identified 8 decision rules (i.e., the tree had 8 branches).

- Rule 1. For students who were not ESL Canadian, and who were not from metropolitan areas, the probability to decline was 82.9%.

- Rule 2. For students who were not ESL Canadian, who were from metropolitan area, and who were not ESL funded, the probability to decline was 78.8%.
- Rule 3. For students who were not ESL Canadian, who were from metropolitan area, who were ESL funded, and who were from special education, the probability to decline was 83.3%.
- Rule 4. For students who were not ESL Canadian, who were from metropolitan area, who were ESL funded, who were not a special education student, and who were not French Immersion students, the probability to decline was 56.4%.
- Rule 5. For students who were not ESL Canadian, who were from metropolitan area, who were ESL funded, who were not a special education student, and who were French Immersion students, the probability to decline was 46.2%.
- Rule 6. For students who were ESL Canadian, who were male, and who were not French Immersion, the probability to decline was 61.7%.
- Rule 7. For students who were ESL Canadian, who were male, and who were French Immersion, the probability to decline was 37.5%.
- Rule 8. For female students who were ESL Canadian, the probability to decline was 54.5%.

Overall, the decision tree suggested that students who were not ESL and who were not from metropolitan area were more likely to decline. Conversely, students who were ESL (as coded in Grade 3) and lived in Edmonton and Calgary tended to maintain or improve their performance over time.

For Cohort 2, decision tree identified four rules.

- Rule 1. For non-ESL Canadian students, the probability to decline was 82.7%.

- Rule 2. For ESL Canadian students who are from non-metropolitan areas, the probability to decline was 75.6%.
- Rule 3. For ESL Canadian students who were from metropolitan areas, and who were not French Immersion, the probability to decline was 60.8%.
- Rule 4. For ESL Canadian students who were from metropolitan areas, and who were French Immersion, the probability to decline was 47.6%.

Overall, this decision tree highlighted the importance of demographic factors such as ESL Canadian, metropolitan area and French Immersion.

For cohort 3, decision tree identified 11 rules.

- Rule 1. For students who were from non-metropolitan areas, and who were male, the probability to decline was 75.4%.
- Rule 2. For students who were not from metropolitan areas, who were female, not ESL funded, not French Immersion, and not ESL Canadian, the probability to decline was 72.0%.
- Rule 3. For students who were not from metropolitan areas, who were female, not ESL funded, not French Immersion, who were ESL Canadian but not in special education, the probability to decline was 59.1%.
- Rule 4. For students who were not from metropolitan areas, who were female, not ESL funded, not French Immersion, who were ESL Canadian and were in special education, the probability to decline was 37.5%.
- Rule 5. For students who were not from metropolitan areas, who were female, not ESL funded, who were French Immersion but not in special education, the probability to decline was 64.9%.

- Rule 6. For students who were not from metropolitan areas, who were female, not ESL funded, who were French Immersion and in special education, the probability to decline was 30.8%.
- Rule 7. For students who were not from metropolitan areas, who were female, who were ESL funded, the probability to decline was 53.2%.
- Rule 8. For students who were from metropolitan areas and who were not ESL funded, the probability to decline was 64.9%.
- Rule 9. For students who were from metropolitan areas, who were ESL funded but not French Immersion, the probability to decline was 53.4%.
- Rule 10. For students who were from metropolitan areas, who were ESL funded, French Immersion, and male, the probability to decline was 53.3%.
- Rule 11. For students who were from metropolitan areas, who were ESL funded, French Immersion, and female, the probability to decline was 35.3%.

Overall, the decision tree for cohort 3 was quite different from Cohorts 1 and 2. The most significant difference was that students from non-metropolitan areas seemed to have less probability to decline than in Cohorts 1 and 2.

Cluster Analysis. To continue our analysis in understanding predictors for students who maintained or improved their performance, we conducted a cluster analysis. For Cohort 1, the elbow method for the scree plot suggested that there were five clusters of performance trajectories for these students (approximately 21% of Cohort 1). The sample sizes for each cluster were respectively: 913, 1583, 958, 1080, 1501. The trajectory plot for the five clusters are shown in Figure 5.1. As shown, cluster 2 (red), 4 (dark blue) and 5 (light blue) were approximately parallel indicating that performance was maintained over time. Although these three clusters had different starting scores in Grade 3, all showed improvement between Grade 3 and 6, and maintained performance between grade 6 and 9. In contrast, cluster 1 (black) trajectory seemed to decline slightly between grade 3 and 6, and improved between grade 6 and 9. Lastly, cluster 3 (green) trajectory started relatively low but increased at grade 6, and slightly decreased at grade 9.

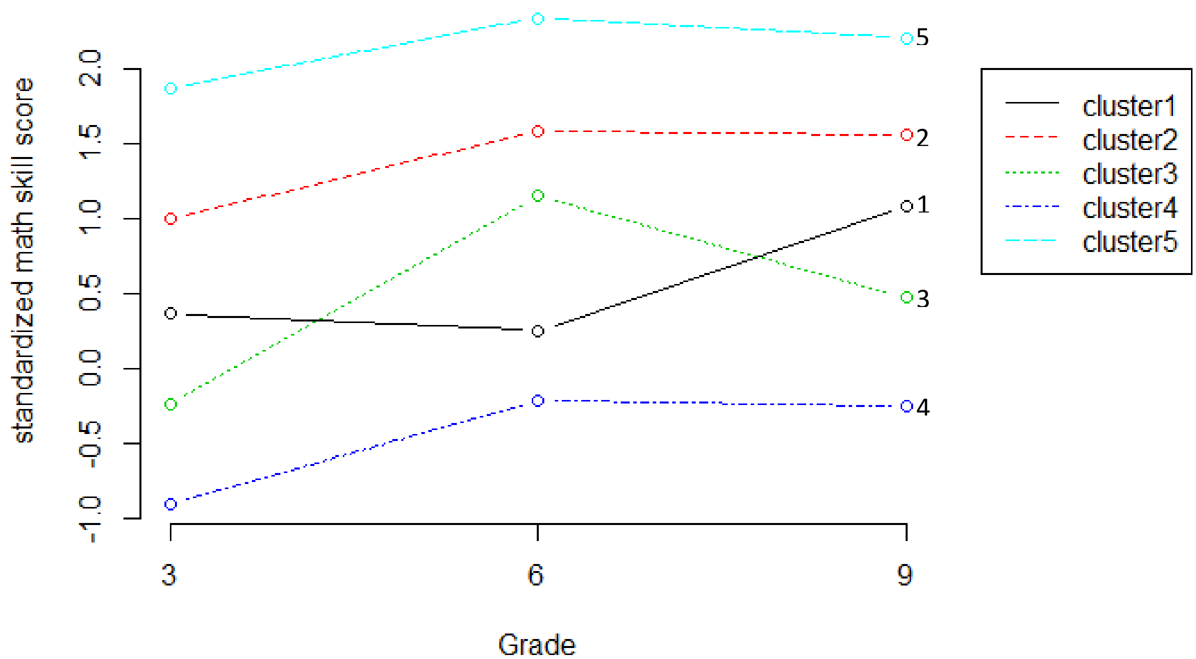

Figure 5.1. Clusters of students in Cohort 1 whose performance improved between the Grade 3 and Grade 9 PAT. Each cluster represents a subset of students with similar trajectories.

Bar charts were constructed to visualize the differences in demographic characteristics available to us among the five performance trajectory clusters. We also include students whose performance declined from grade 3 to 9 as cluster 0. Six bar charts for Cohort 1 are shown in Figures 5.2 to 5.7. Each bar chart helps visualize the demographic distribution of six characteristics: gender, French Immersion, ESL funded, ESL Canadian, Special Education, and

Region. Using cluster 0 as a reference point, it is useful to look for differences in characteristics across the clusters. For example, in Figure 5.2 cluster 0, students whose performance declined between grades 3 and 9, shows a fairly even split between males and females. In contrast, there were more females in cluster 3 (students with strong increase in performance from grade 3 to 6, and then declined performance from grades 6 to 9) and 4 (cluster with lowest starting score). The remaining clusters were similar to cluster 0 with a fairly evenly split between males and females. The remaining bar charts for Cohort 1 can be interpreted in a similar way for the remaining coded characteristics within the data. Cluster charts and characteristic bar graphs for Cohorts 2 and 3 are also shown below.

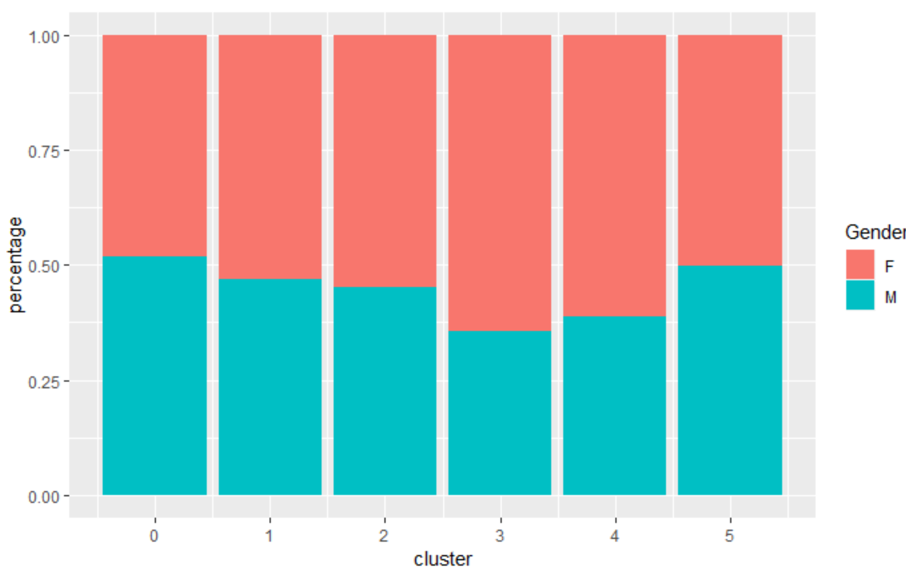

Figure 5.2. Gender bar chart for Cohort 1.

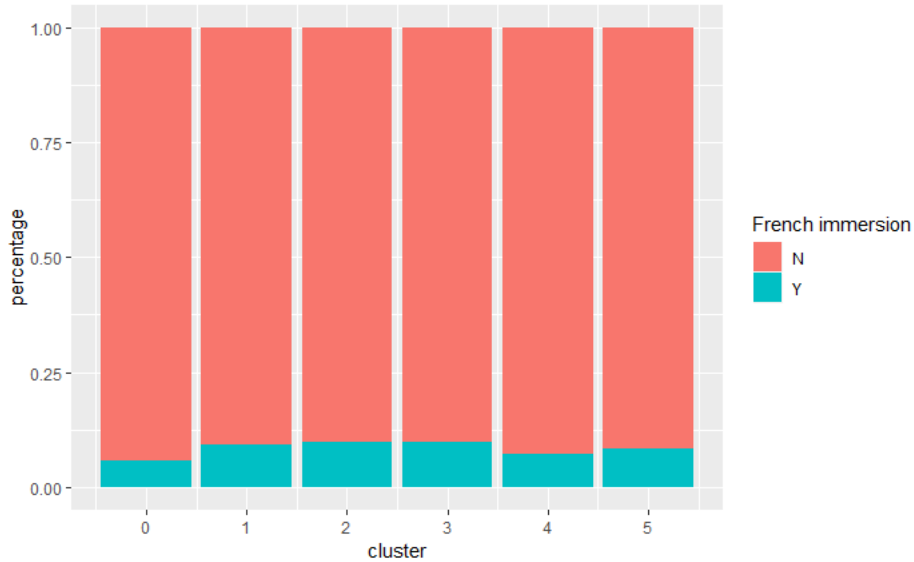

Figure 5.3. French Immersion bar chart for Cohort 1.

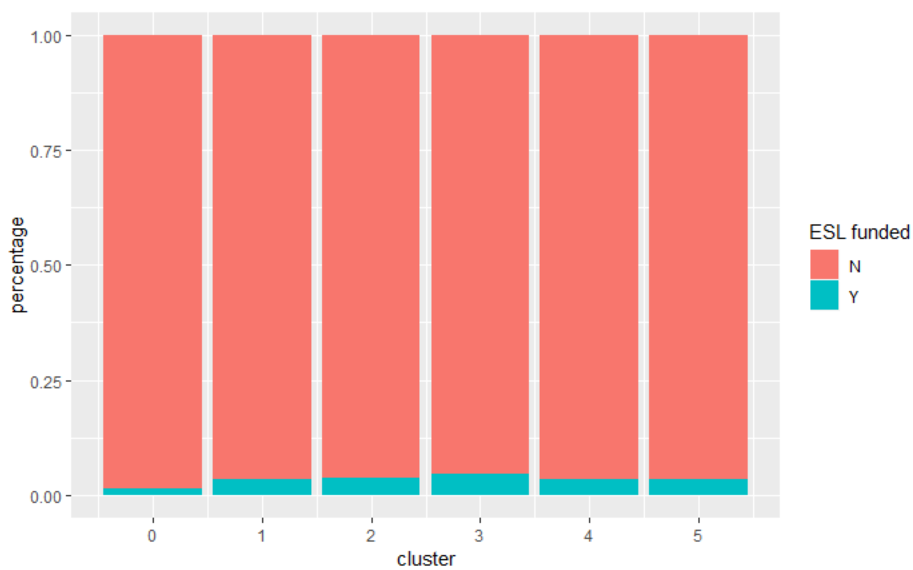

Figure 5.4. ESL funded bar chart for Cohort 1.

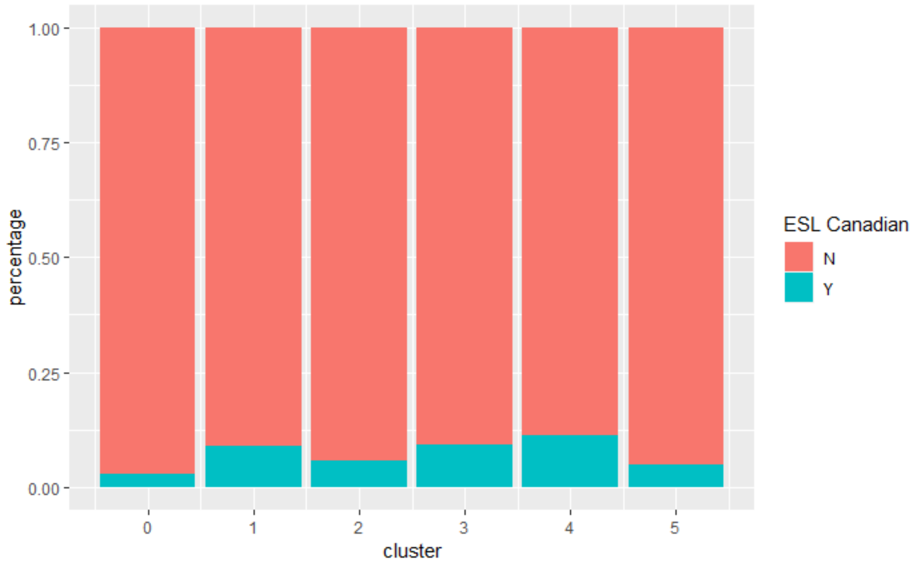

Figure 5.5. ESL Canadian bar chart for Cohort 1.

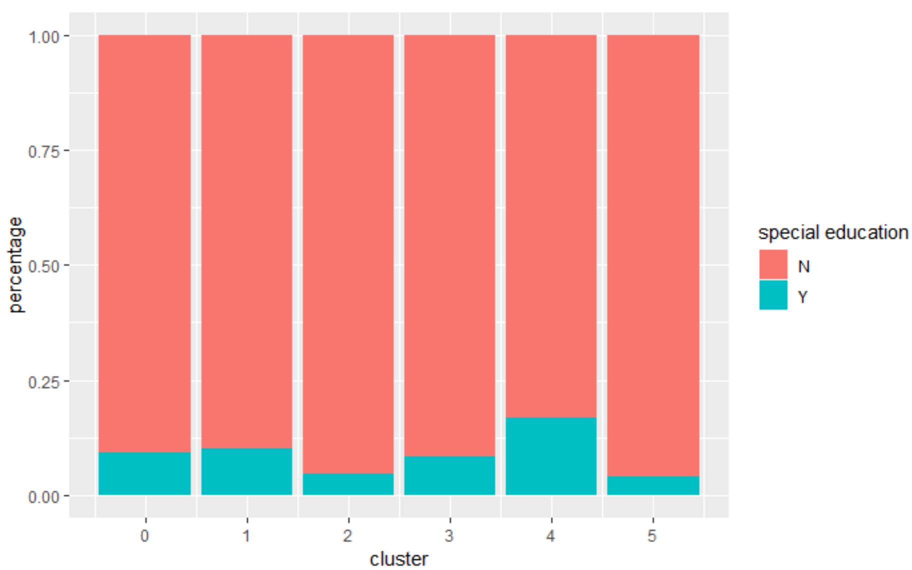

Figure 5.6. Special education bar chart for Cohort 1.

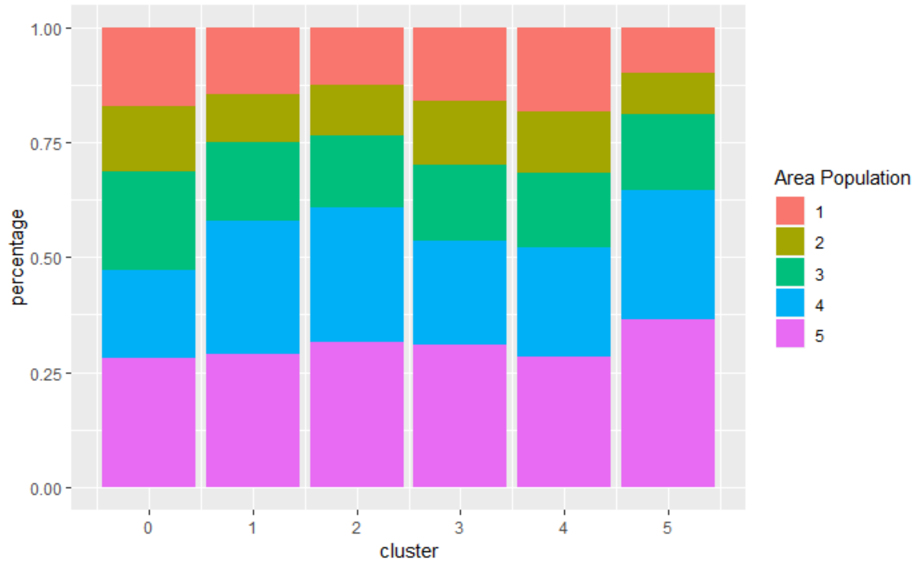

Figure 5.7. Area population bar chart for Cohort 1. Rural – 1; Small Town – 2; Small City – 3; Edmonton – 4; Calgary – 5.

Overall, the results were consistent with the decision tree analyses in that factors such as region, ESL, and French Immersion were important demographic characteristics that distinguish the clusters representing improved or maintained performance between Grades 3 and 9.

For Cohort 2, the elbow method also identified five clusters. The sample sizes for each cluster were respectively: 761, 881, 1392, 1372, and 890. The trajectory plot is shown in Figure 6.1. As expected, the trajectories are very similar to those in Cohort 1 with three clusters (3, 4 and 5) being relatively parallel with slight increases from Grades 3 to 6 and maintaining or slightly improving performance in Grade 9. Clusters 1 and 2 start and finish in relatively the same place, but cluster 1 shows performance increasing in Grade 6 and decreasing in Grade 9 (but still above Grade 3 performance) while cluster 2 students decrease performance in Grade 6 and increase in Grade 9. The demographic factor bar charts are shown in Figures 6.2 to 6.7.

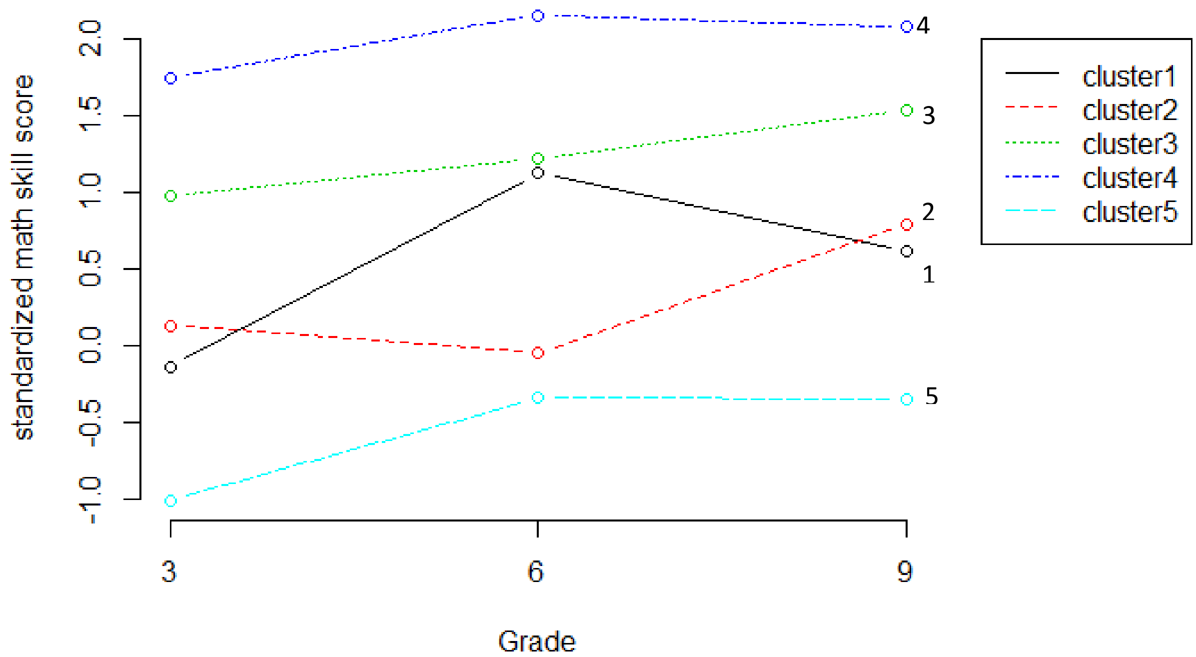

Figure 6.1. Mathematics skill trajectory plot for Cohort 2.

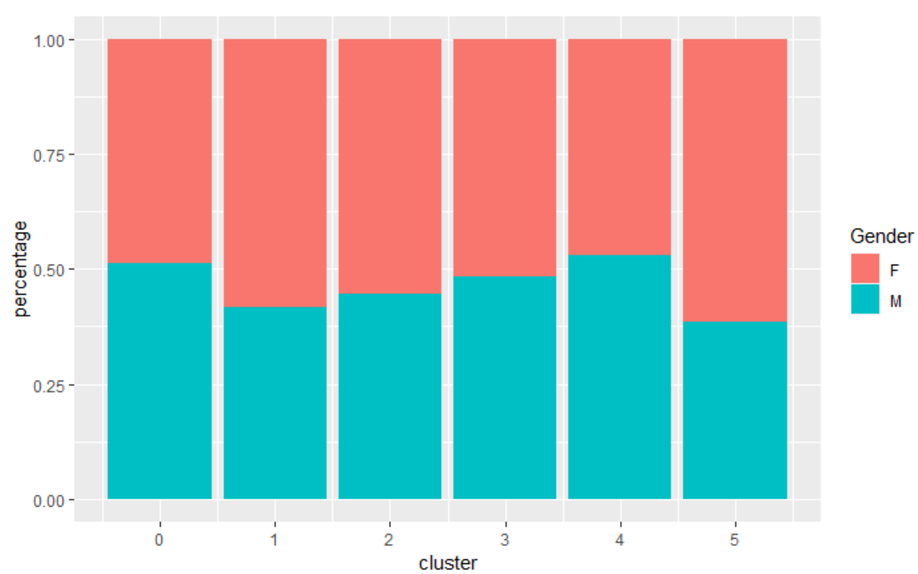

Figure 6.2. Gender bar chart for Cohort 2.

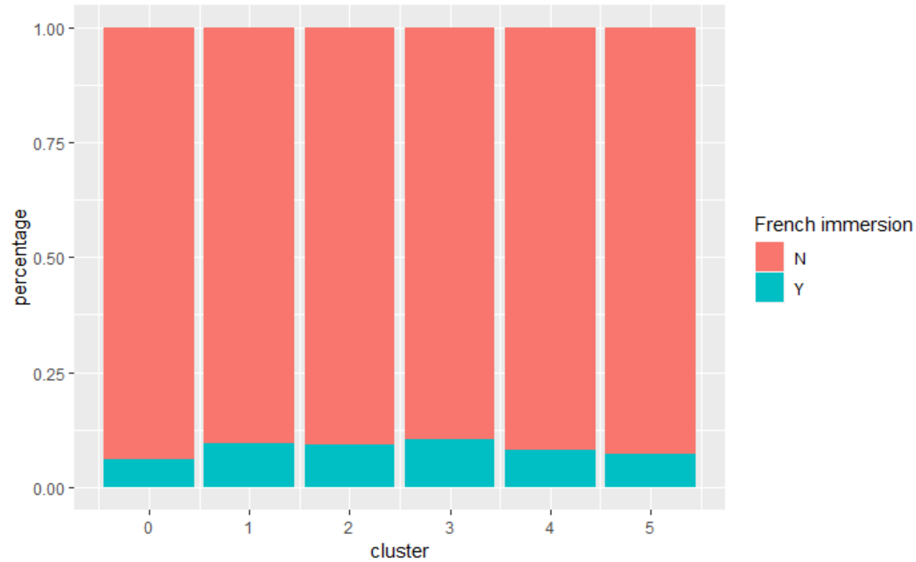

Figure 6.3. French Immersion bar chart for Cohort 2.

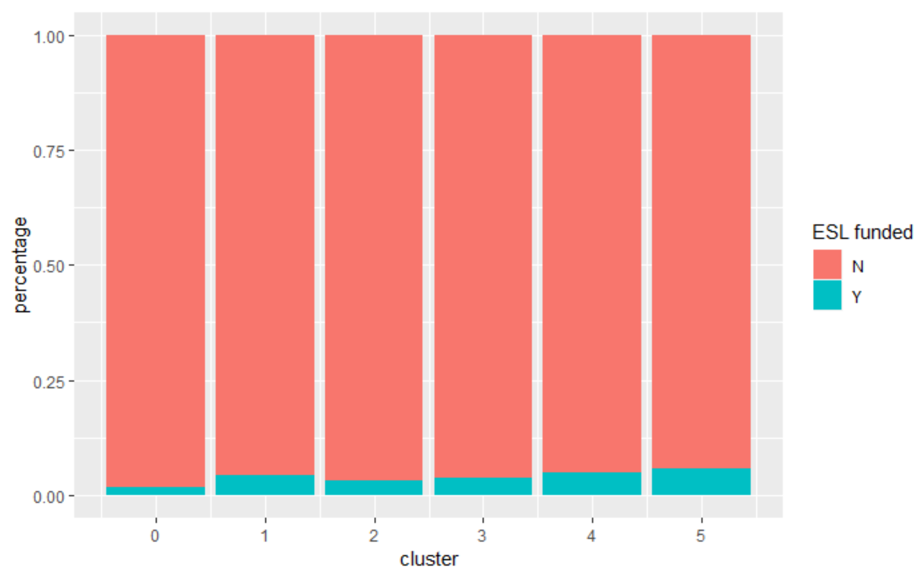

Figure 6.4. ESL funded bar chart for Cohort 2.

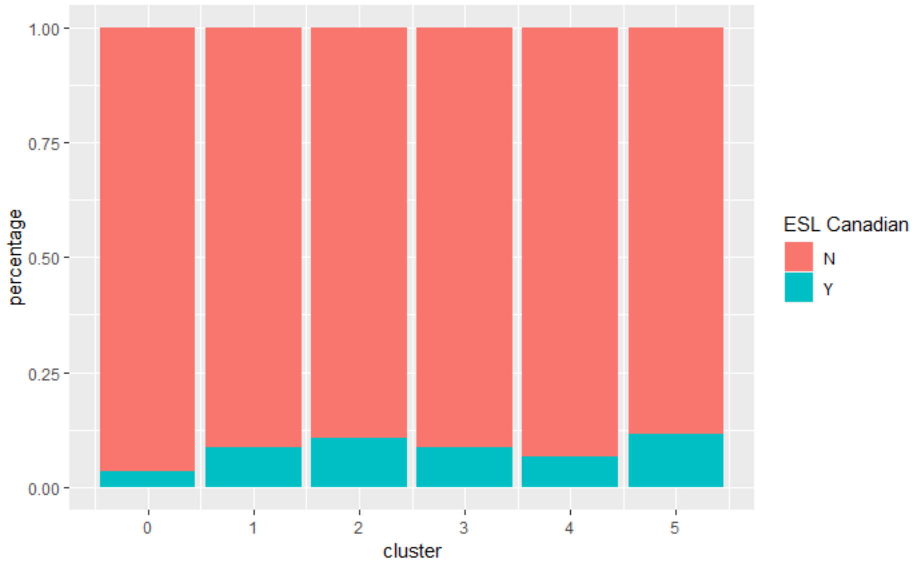

Figure 6.5. ESL Canadian bar chart for Cohort 2.

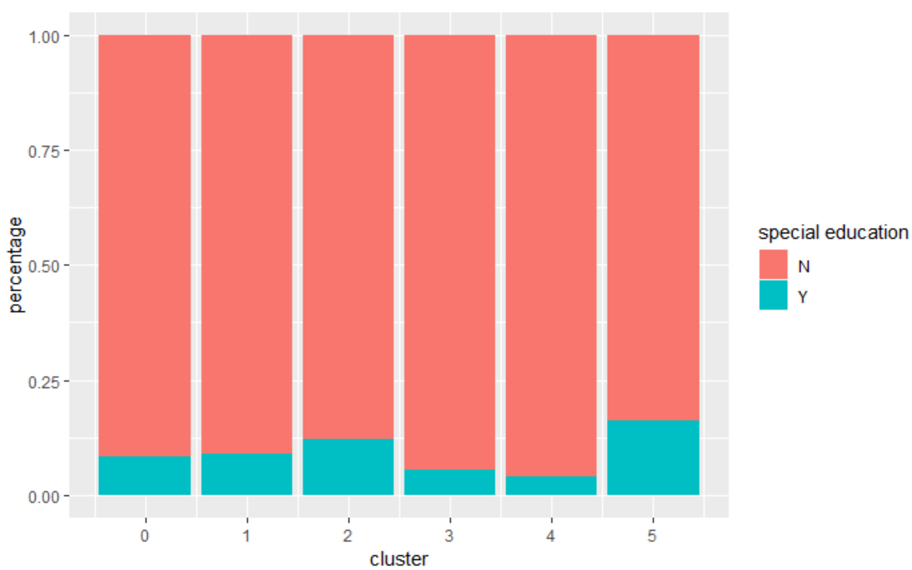

Figure 6.6. Special education bar chart for Cohort 2.

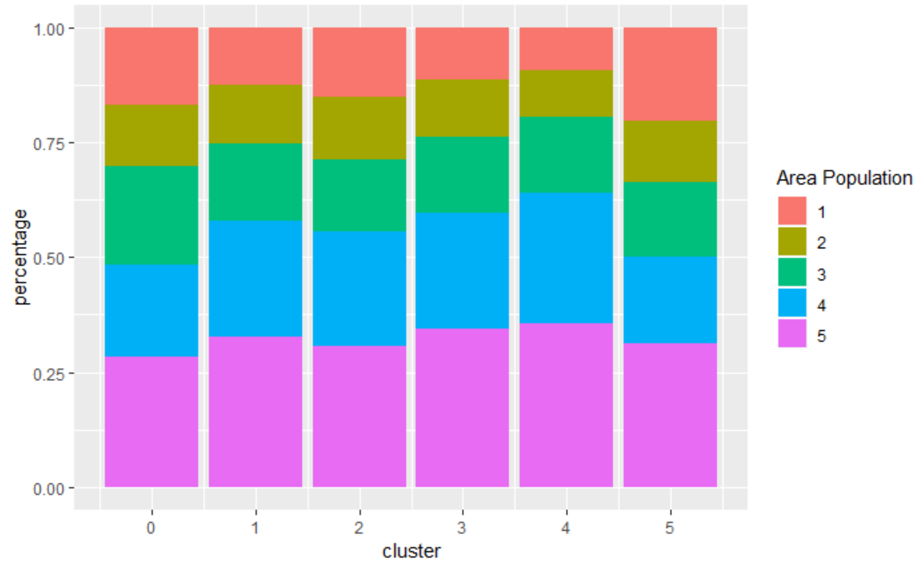

Figure 6.7. Area population bar chart for Cohort 2. Rural – 1; Small Town – 2; Small City – 3; Edmonton – 4; Calgary – 5.

Overall, the student characteristics related to performance trends are very similar in Cohorts 1 and 2.

For Cohort 3, the elbow method identified 4 clusters. The sample sizes for each clusters were 1623, 2677, 2436, and 2658. The trajectory plot is shown in Figure 7.1. For this cohort, all four clusters have parallel trajectories. All four clusters show an increase in performance from Grades 3 to 6. Clusters 1 and 4 show a slight decline between Grades 6 and 9, while Clusters 2 and 3 continue a slight upward trend between Grades 6 and 9. The demographic factor bar charts are shown in Figures 7.2 to 7.7. Compared with Cohorts 1 and 2, the differences between the clusters in gender proportion for Cohort 3 were smaller and all other demographic factors were similar.

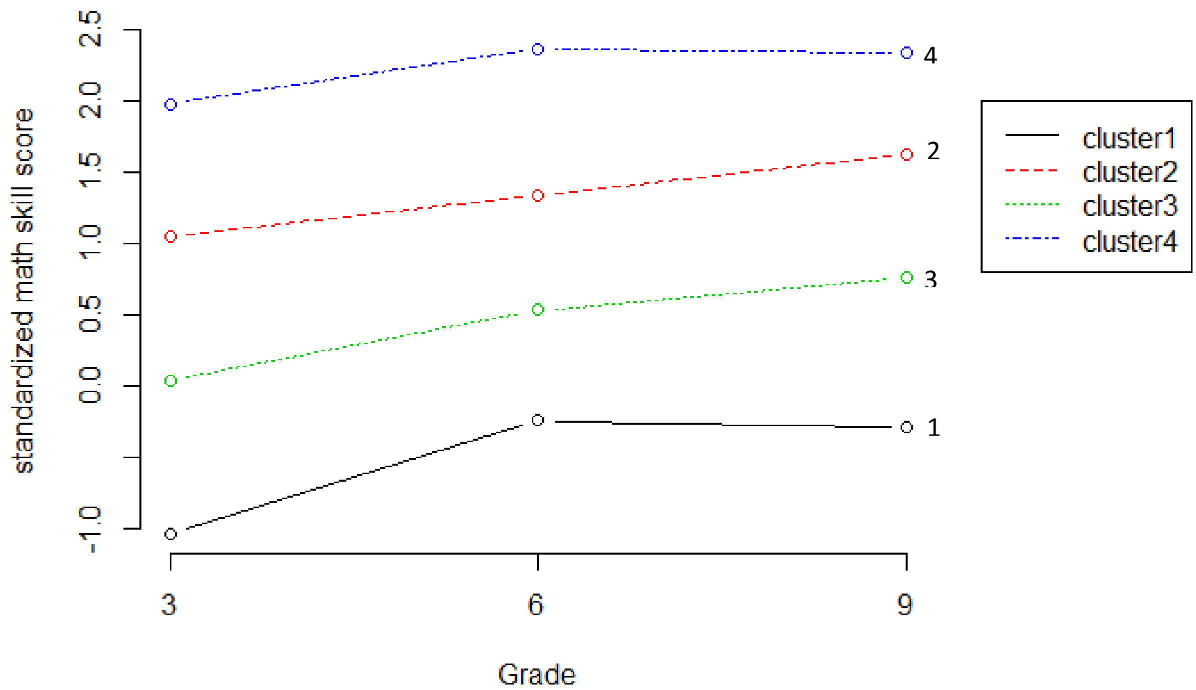

Figure 7.1. Mathematics skill score trajectory plot for Cohort 3.

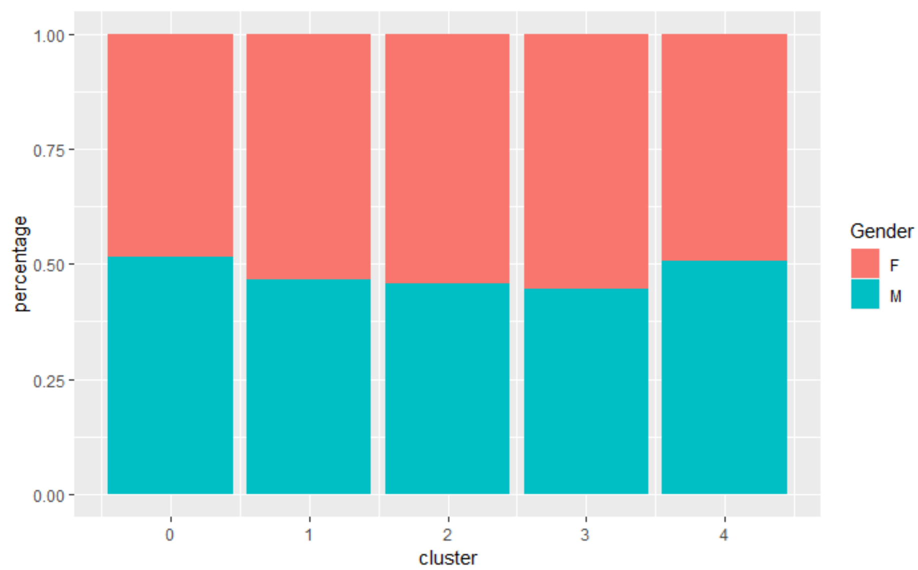

Figure 7.2. Gender bar chart for Cohort 3.

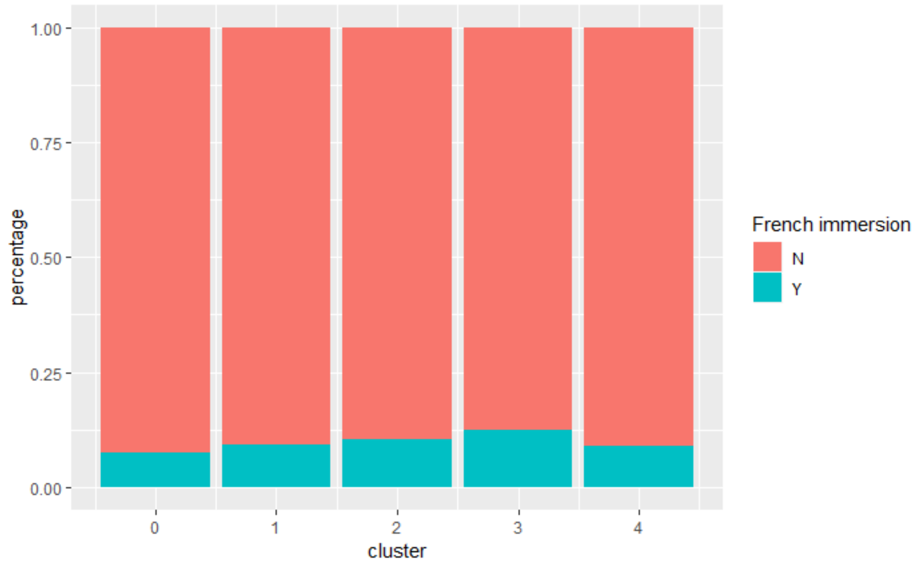

Figure 7.3. French Immersion bar chart for Cohort 3.

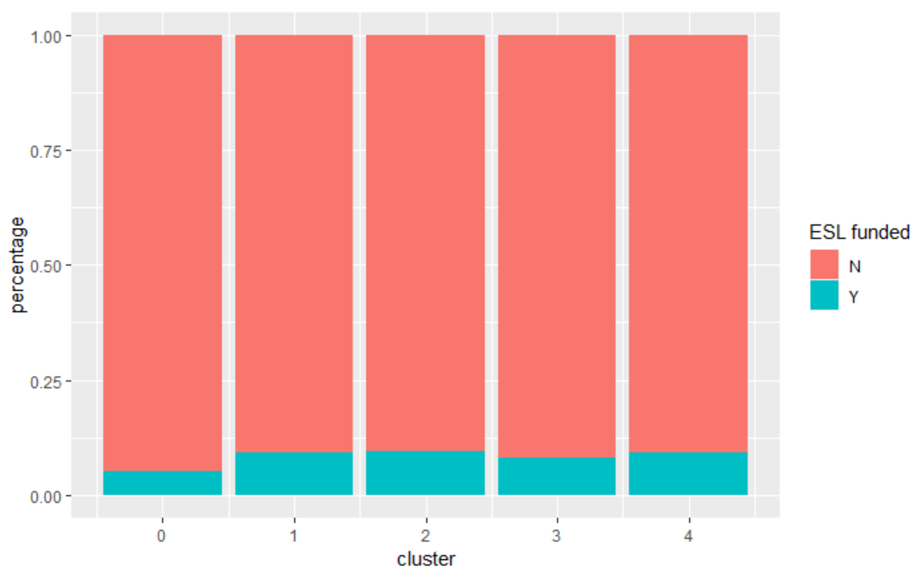

Figure 7.4. ESL funded bar chart for Cohort 3.

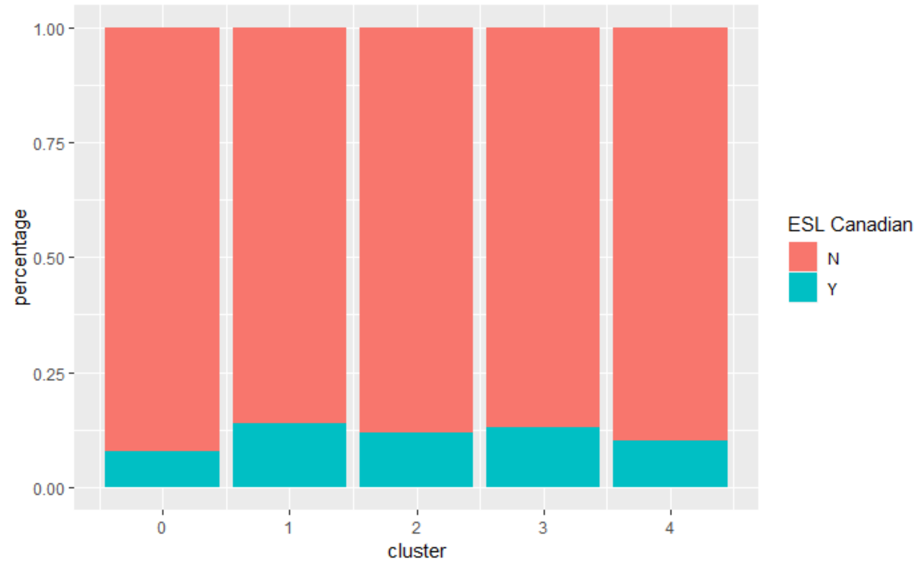

Figure 7.5. ESL Canadian bar chart for Cohort 3.

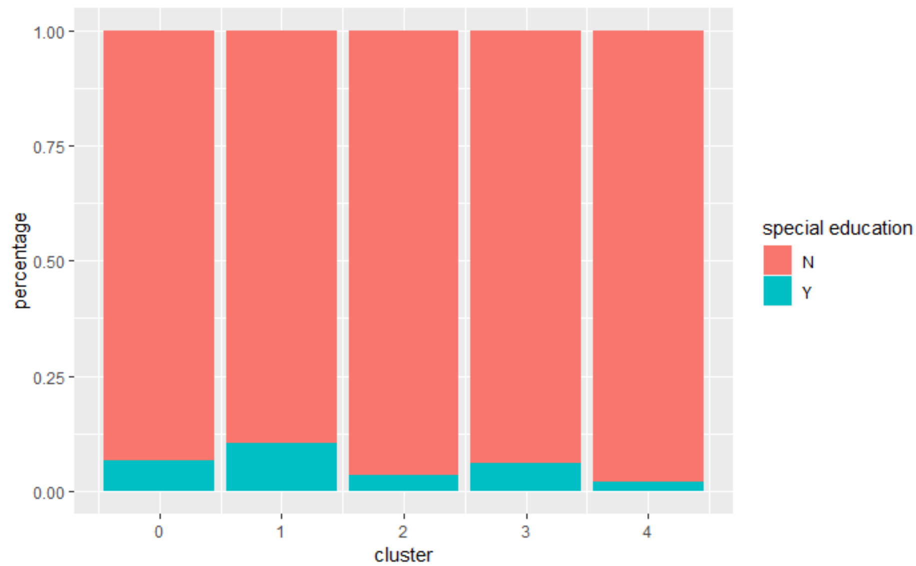

Figure 7.6. Special education bar chart for Cohort 3.

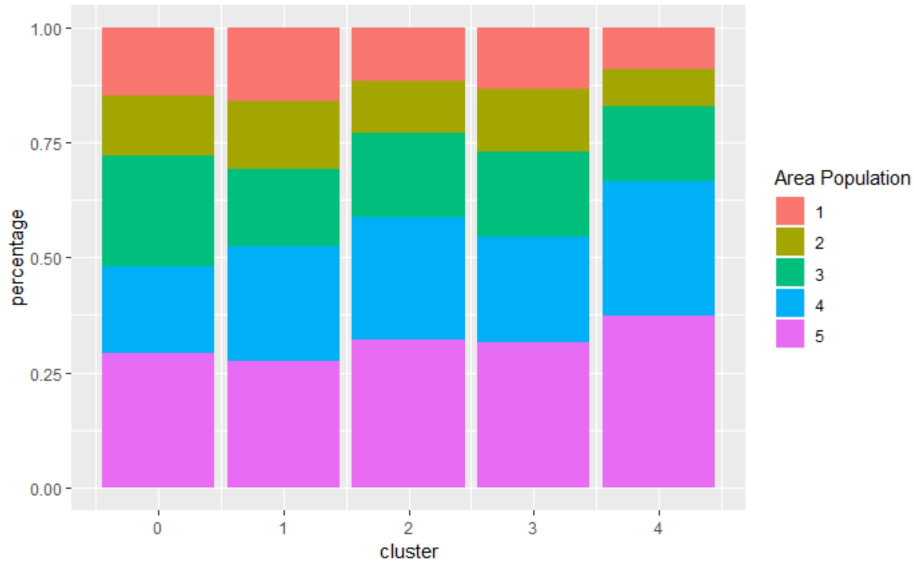

Figure 7.7. Area population bar chart for Cohort 3. Rural – 1; Small Town – 2; Small City – 3; Edmonton – 4; Calgary – 5.

Predictions of Future Performance Based on Mathematics Content

The third research question of our study was, “What aspects of mathematics content predict future performance?” In this section we use a Panel Path analysis and a Confirmatory Factor analysis to gain insight into what strands within the Mathematics Program of Studies predict longitudinal performance on the PATs.

Panel Path Analysis. To examine how different areas of mathematics influenced each other over time, a panel path analysis was conducted for each cohort using Mplus 7, a statistical modelling program. One challenge of this analysis is that with large sample sizes, any small effects could be detected as statistically significant making the model unnecessarily complex and difficult to interpret. To overcome this problem, three criteria were used: 1) the model had to be theoretically meaningful, 2) only standardized coefficients larger than .05 were retained in the model, and 3) when examining model fit, *Root Mean Squared Error of Approximation*

(RMSEA), *Comparative Fit Index* (CFI) and *Standardized Root Mean Residual* (SRMR) were emphasized over the Chi square test, which is very sensitive to large sample size. The resulting models for each cohort are presented in Figures 8 to 10. The higher the coefficients between the strands at each grade, the larger the significance and greater the relationship in terms of student performance on the questions.

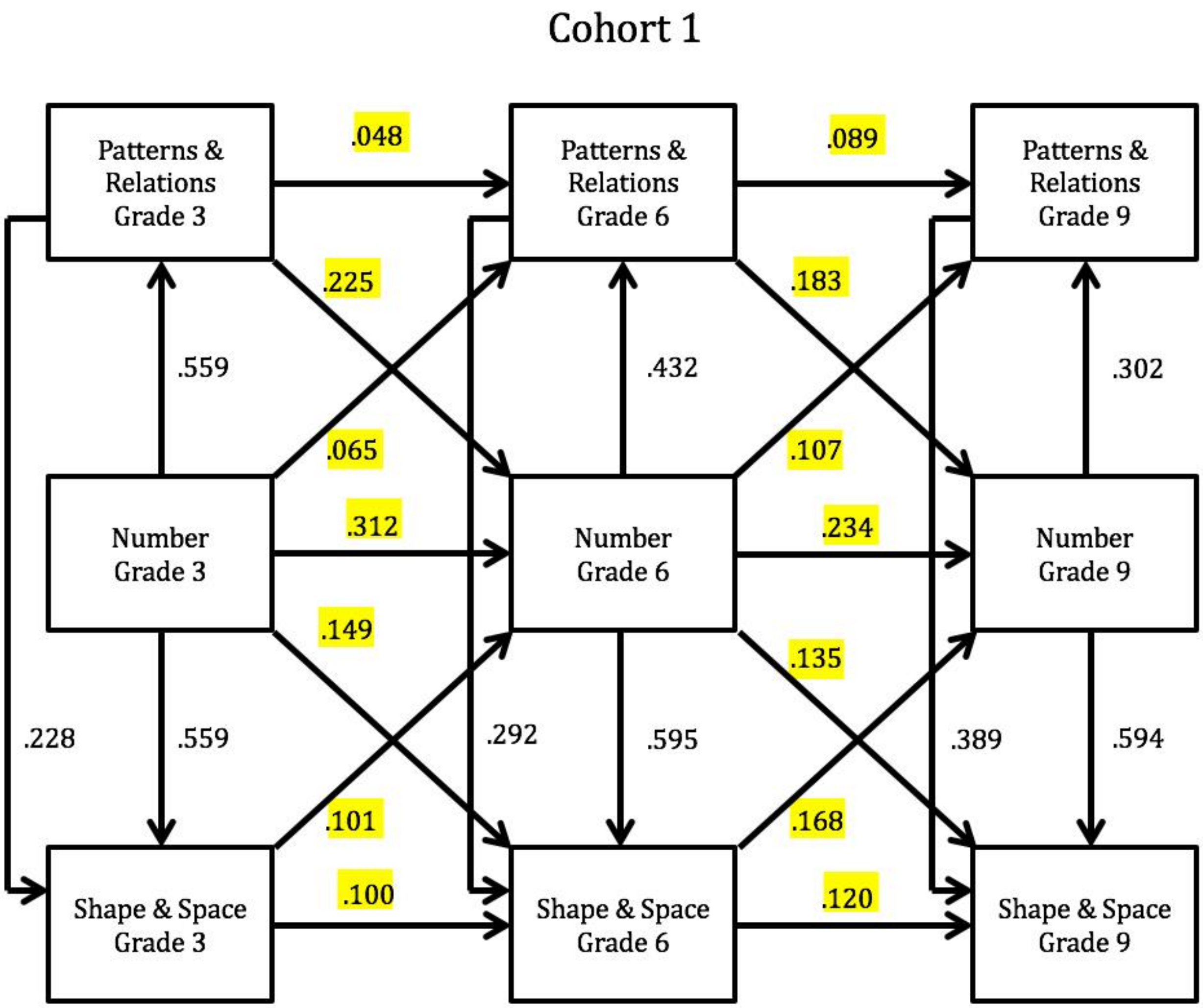

Figure 8. Panel path analysis for Cohort 1 with standardized coefficients (all significant).

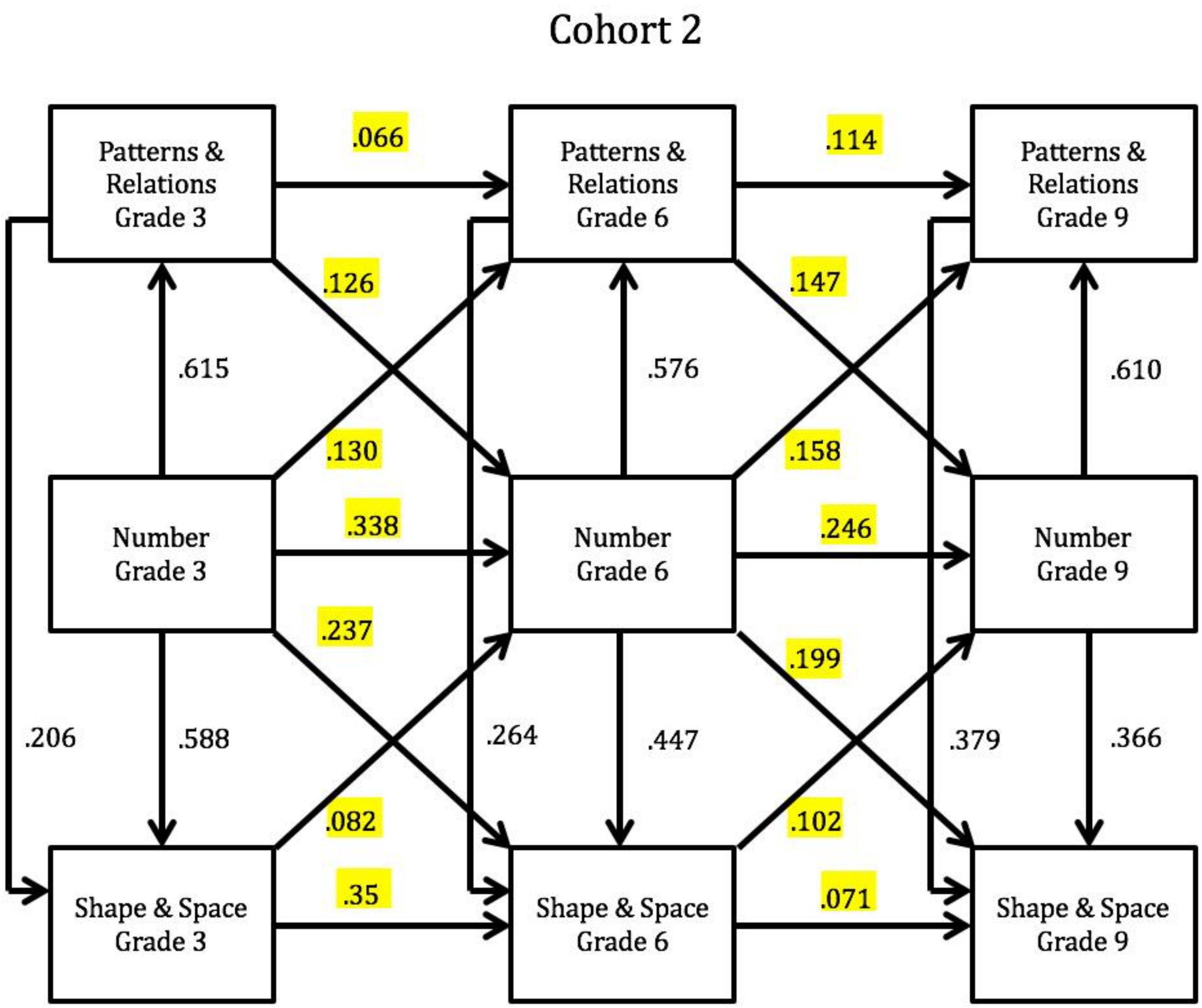

Figure 9. Panel path analysis for Cohort 2 with standardized coefficients (all significant).

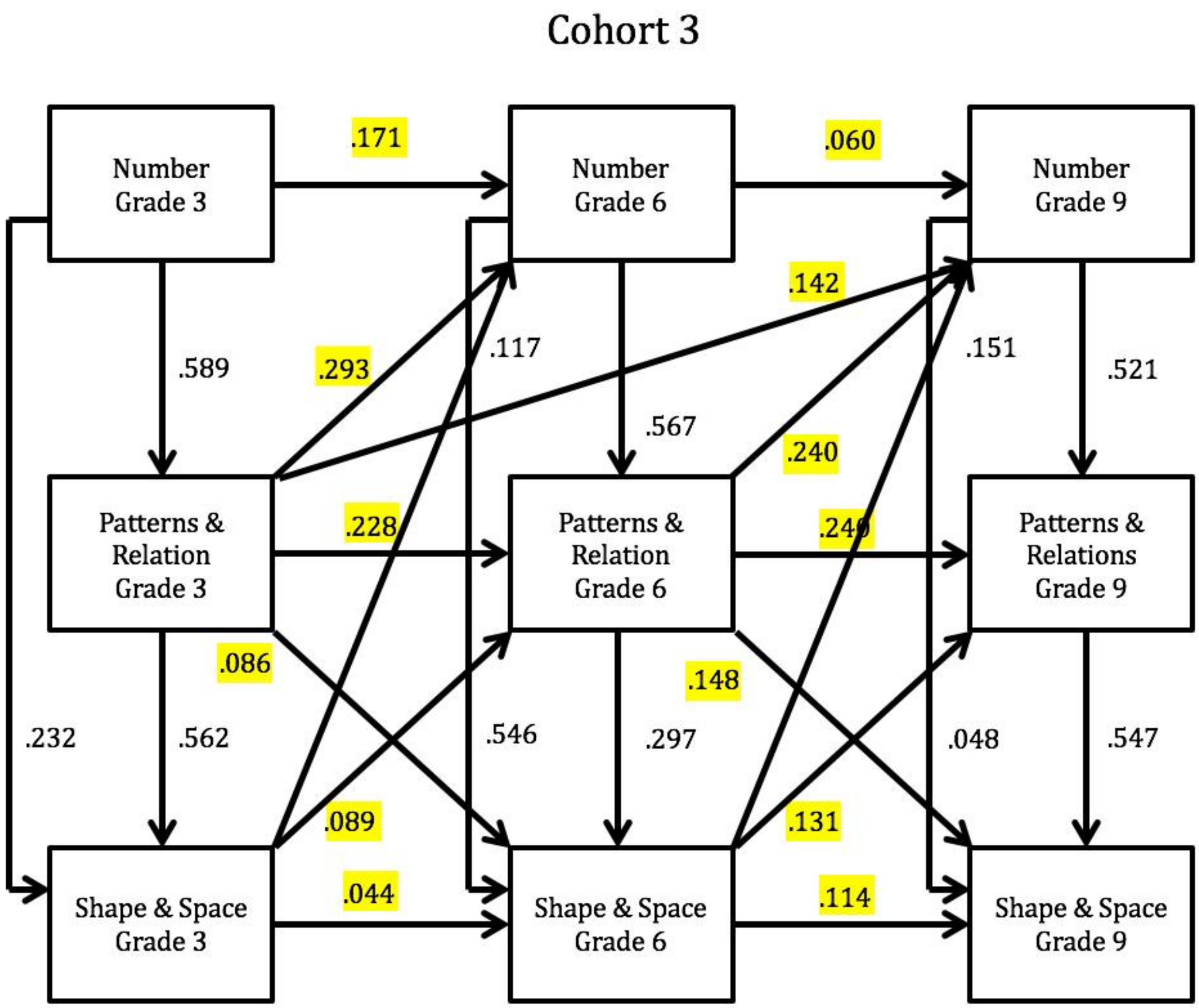

Figure 10. Panel path analysis for Cohort 3 with standardized coefficients (all significant).

As shown in the figures, panel models for Cohorts 1 and 2 had the same structure with minor differences in path coefficients. The model fitted the data well for Cohort 1: RMSEA=.05, CFI=.995, and SRMR=.017; and for Cohort 2: RMSEA=.046, CFI=.996, SRMR=.013. All the coefficients shown were positive, standardized and statistically significant. To simplify the model we used three of the four content strands including Number, Patterns & Relations, and Shape & Space.

The coefficients for the panel path models show a range of small and large coefficients throughout the figure—all of which are considered statistically significant. Theoretically, the more important paths in the model are the cross links highlighted in yellow. That is, the links between grades. For example, the cross link between Number at Grade 3 and Patterns & Relations at Grade 6 has a coefficient of 0.065 in Cohort 1. Since it is impossible for a future event to cause a past event, the relationship between Number at Grade 3 and Patterns & Relations at Grade 6 is significant after controlling for Patterns & Relations at Grade 3, and we have strong, though not perfect, evidence to suggest that Number at Grade 3 influenced Patterns & Relations at Grade 6.

Not surprisingly, the cross links with high coefficients occur between Grade 3 Number and Grade 6 Number (i.e., coefficients of 0.312 and 0.338 for Cohorts 1 and 2 respectively) and Grade 6 to Grade 9 Number (i.e., coefficients of 0.234 and 0.246 for Cohorts 1 and 2 respectively). This suggests that performance on Grade 3 Number predicted Grade 6 Number, which in turn predicted Grade 9 Number.

The model in Cohorts 1 and 2 suggests that Number is the mathematics strand that is central to the model. This pattern is particularly noticeable in Cohort 2. That is, Grade 3 Number strongly predicted all strands at Grade 3 and Grade 6 Number strongly predicted all strands at Grade. 9.

In addition, the model showed that Patterns & Relations and Shape & Space also predicted future performance in Number. This trend has stronger coefficients in Cohort 1. For example, Grade 3 Patterns & Relations appear to be a strong predictors of Grade 6 Number, although less strong than Grade 3 Number. Grade 6 to Grade 9 shows the same pattern.

Interestingly, Grade 3 Patterns & Relations was not as strong a predictor of Grade 6 Patterns & Relations as Number. This trend is repeated from Grade 6 to Grade 9.

Finally, the model also considers the relationship between different mathematics strands within the same time point. For example, at Grade 3, Number strongly predicted Patterns & Relations and Shape & Space, and Patterns & Relations also independently predicted Shape & Space. It is important to note that unlike cross links, we are less certain about the directions of the causal links within the same time point. Therefore, the direction was determined based on theoretical considerations, but should be considered correlational, rather than predictive.

The same model was fitted to Cohort 3. However, model fit indices suggested the fit was not as strong as before with RMSEA greater than .08, and many path coefficients were smaller than .05. Consequently, a new model was developed for Cohort 3 based on the modification of indices and theoretical interpretation. The resulting model is shown in Figure 10. It fitted the data well with RMSEA=.039, CFI=.997, SRMR=.011. While the new model had similar basic links and within time point links, the cross links were very different from the models in Cohort 1 and 2. The most striking difference is that Number no longer predicted future performances in Patterns & Relations, and Shape & Space. However, Patterns & Relations and Shape & Space strongly predicted future performances in Number, and these strands were stronger predictors of future Number than past Number. Another difference was that in Cohort 3 model, Patterns & Relations and Shape & Space predicted each other’s future performances. These findings highlight the most prominent aspects of the path model analysis.

Confirmatory Factor Analyses (CFA) for Mathematics Strands. The panel path model results illustrated very different predictors for performance for Cohorts 1/2 and Cohort 3. These results suggested that the third cohort, based on the 2007 curriculum framework, was different from the results of the first two cohorts who experienced the 1996 curriculum framework. In an attempt to provide additional information and evidence for the differences in performance based on strands in the mathematics curriculum, we undertook a multi-group Confirmatory Factor Analyses (CFA) using Mplus 7.11 (Muthén & Muthén, 1998-2015). To evaluate the model data fit, the following criteria were used: Root Mean Square Error of Approximation (RMSEA) < .08 (Browne & Cudeck, 1993), Standardized Root Mean Square Residual (SRMR) < .08, (Hu & Bentler, 1998), and Comparative Fit Index (CFI) > .90 (Bentler, 1990). Chi square test was not used as it is well known that it is too sensitive under large sample sizes. In these CFA models, we assumed there was a general factor underlying all four strands of the mathematics program of studies (i.e., Number, Patterns & Relations, Shape & Space, and Statistics & Probability). The strand of mathematics having the strongest standardized factor loading is considered to be most central to the general mathematics factor. Longitudinal CFAs were conducted for each cohort. The results are shown in Figures 11-13.

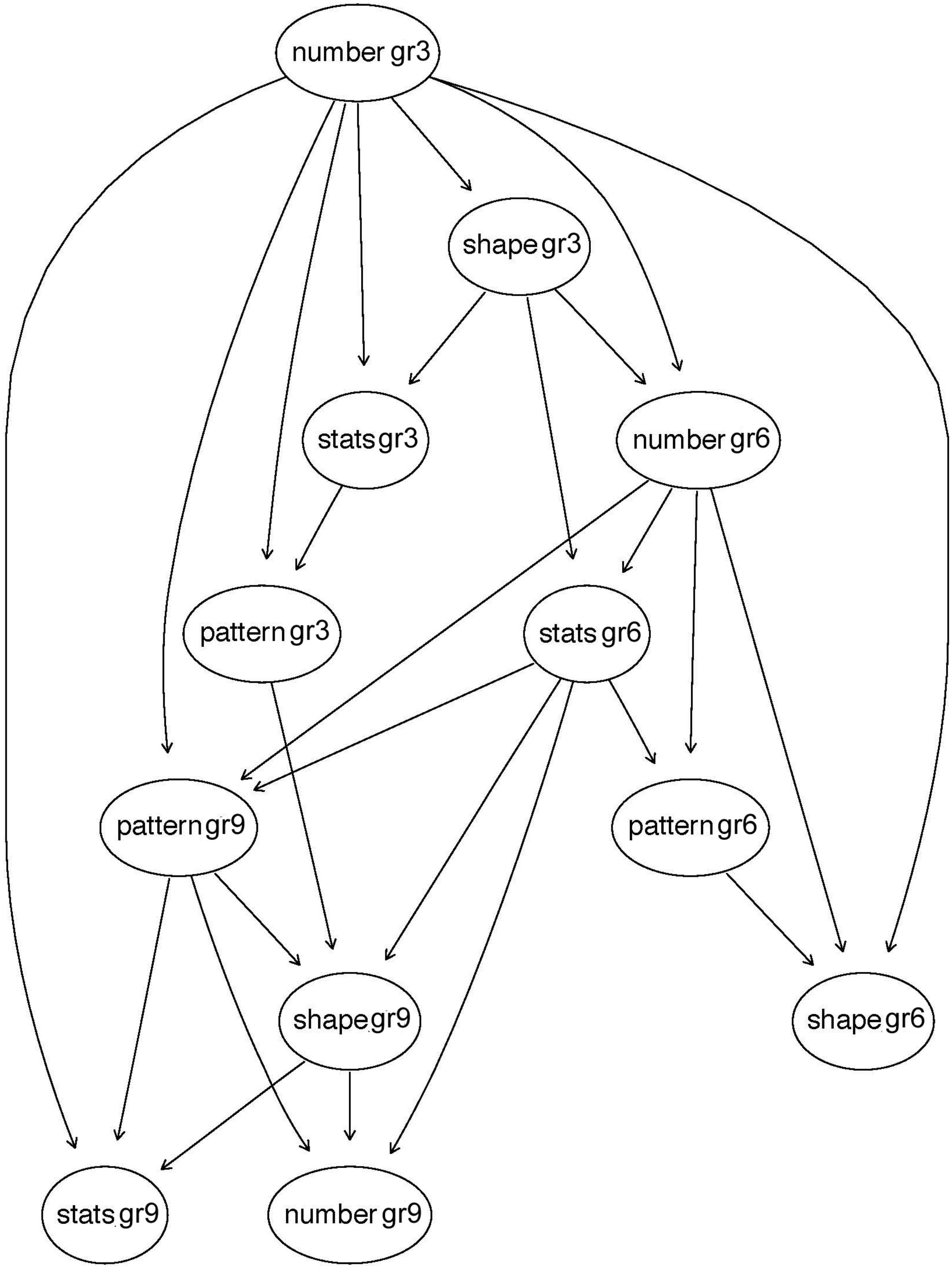

Figure 11. Cohort 1 Confirmatory Factor Analyses (CFA)

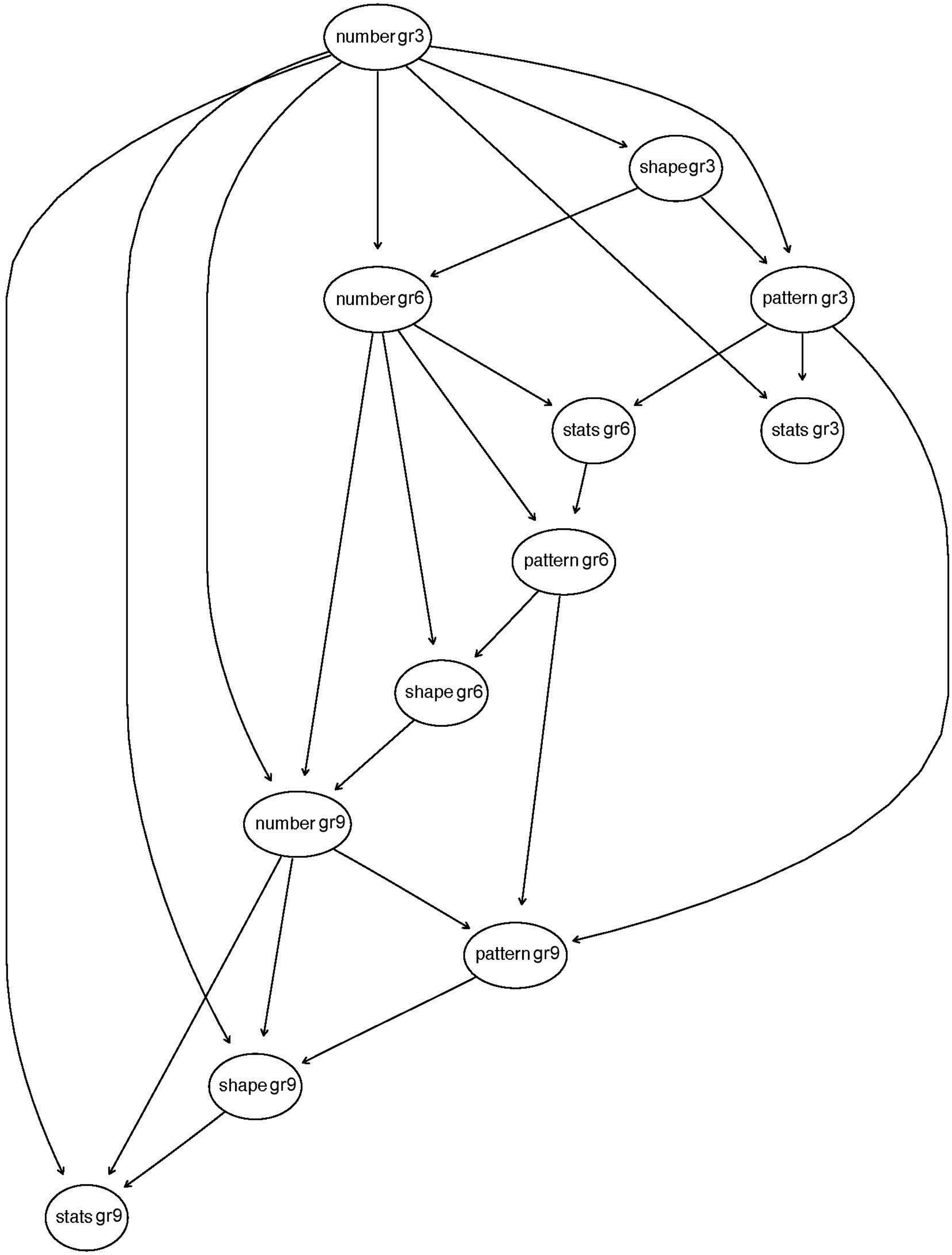

Figure 12. Cohort 2 Confirmatory Factor Analyses (CFA)

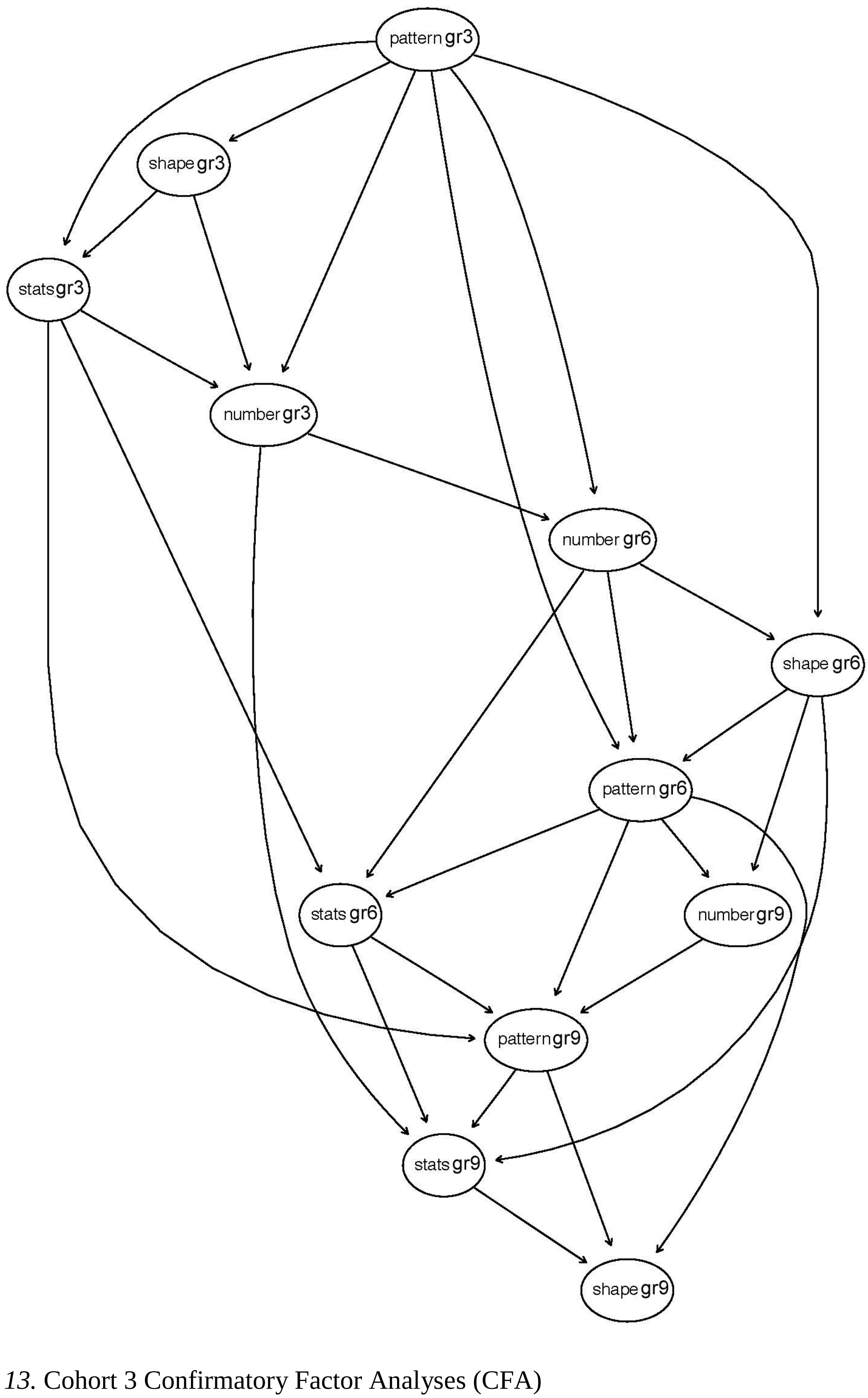

Figure 13. Cohort 3 Confirmatory Factor Analyses (CFA)

Fit indices suggested that all models fitted the data well: Cohort 1 CFA model (RMSEA = .018, CFI=.998, SRMR=.009), Cohort 2 CFA model (RMSEA = .018, CFI=.998, SRMR=.009), and Cohort 3 CFA model (RMSEA = .074, CFI=.971, SRMR=.022). Upon closer examination of the factor loadings, once again, Number seemed to have the strongest standardized factor loadings for Cohorts 1 and 2 for all three exams with the minor exception of Cohort 1 at Grade 9 in which Patterns & Relations had the highest factor loading. In contrast, Patterns & Relations seemed to have the strongest standardized factor loading in Cohort 3, with the exception of the Grade 6 exam, in which Number had the strongest standardized factor loading. Overall, Number and Patterns & Relations appeared to be the most central concepts for the general mathematics factor. In Cohort 1 and 2, Number played a more important role. In Cohort 3, Patterns & Relations was more important. These results are consistent with those in the panel model in which Number tended to be the strongest predictor of performance in Cohorts 1 and 2, not only for Number in subsequent grades, but for performance in the other strands. In Cohort 3, performance on Patterns & Relations questions was the stronger predictor.

Confirmatory Factor Analyses (CFA) for Longitudinal Performance. As shown in our growth curve analysis presented previously, there was a clear downward trend for students overall across the three time points. As noted, there was generally a slight decline from Grade 3 to Grade 6, and a substantial decline from Grade 6 to Grade 9. This downward trend appeared regardless question type (i.e., knowledge, skills), gender, or region. In a final effort to determine what contributed to student performance over the three time points, we were interested in examining the mathematics factor structure for students who did not decline in performance across the three examination points compared to the students who did.

A multi-group longitudinal confirmatory factor analyses were conducted to examine whether the factor structures for students who declined in mathematics performance were different from students who did not decline. Factor loadings were held to be constant for both the decline group and the non-decline group. The results revealed that there were only minor differences between the two groups in terms of factor structures, as the constrained CFA models showed mostly acceptable fit: Cohort 1 (RMSEA= .041, CFI=.986, SRMR=.137), Cohort 2 (RMSEA =.034, CFI=.991, SRMR=.074), and Cohort 3 (RMSEA =.080, CFI=.958, SRMR=.147). The model fit indices satisfied criteria for RMSEA and CFI, but not for SRMR, indicating borderline fit. While removing the constraint improved the SRMR fit, the overall patterns of factor loadings stayed the same. Therefore, we conclude there were only minor structural differences between the decline and non-decline groups. Figures 14 to 19 provide a visual means to compare the CFA results for students whose performance declined vs. students who performance did not decline for each cohort.

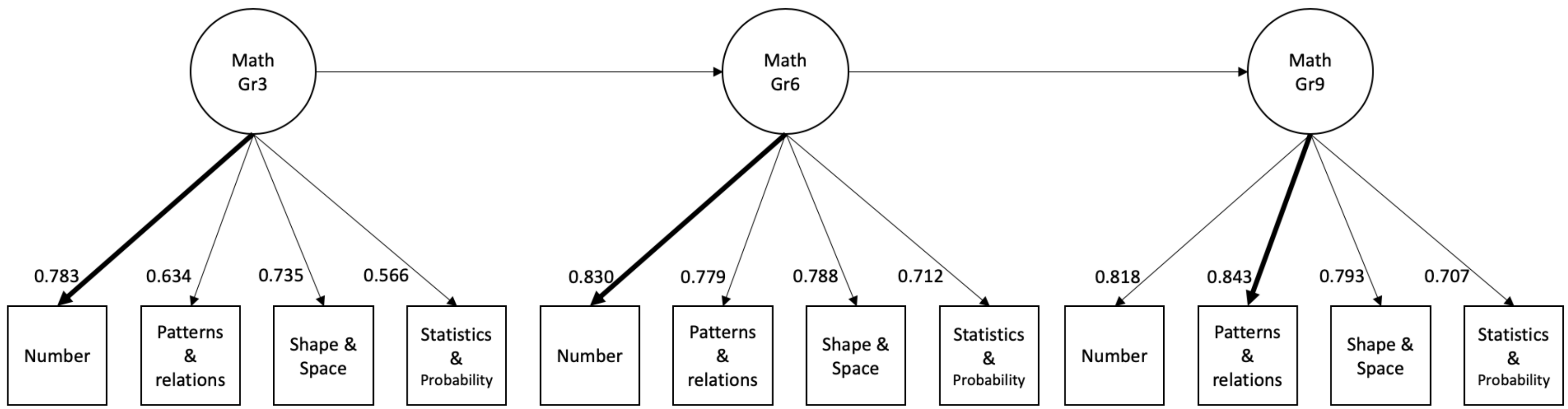

Figure 14. Cohort 1 Longitudinal confirmatory factor analysis for students whose performance declined between Grades 3 and 9. RMSEA = .017, CFI=.998, SRMR=.007.

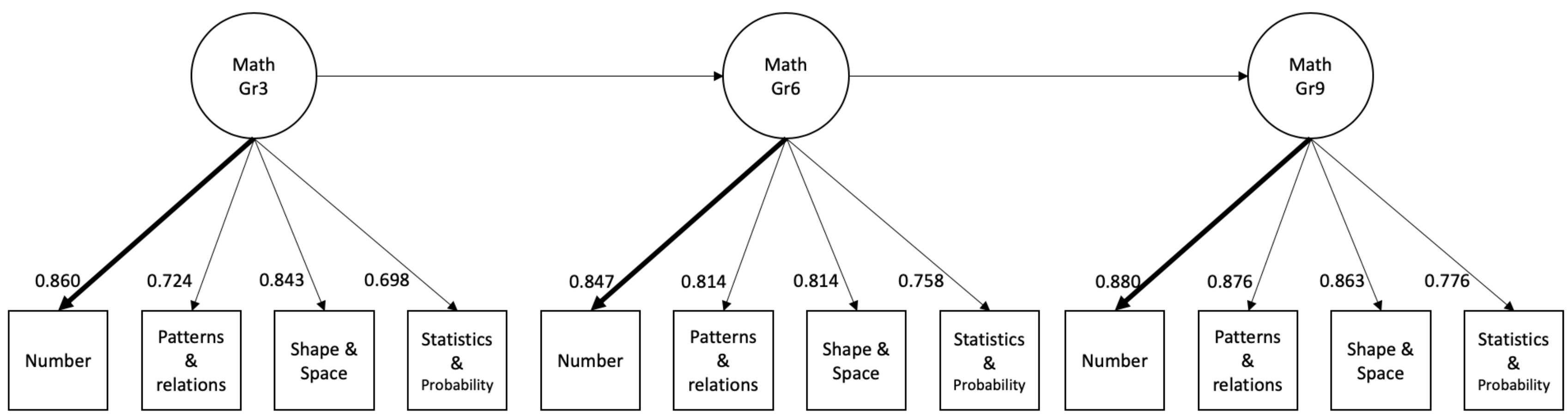

Figure 15. Cohort 1 Longitudinal confirmatory factor analysis for students whose performance did not decline between Grades 3 and 9. RMSEA = .017, CFI=.998, SRMR=.007.

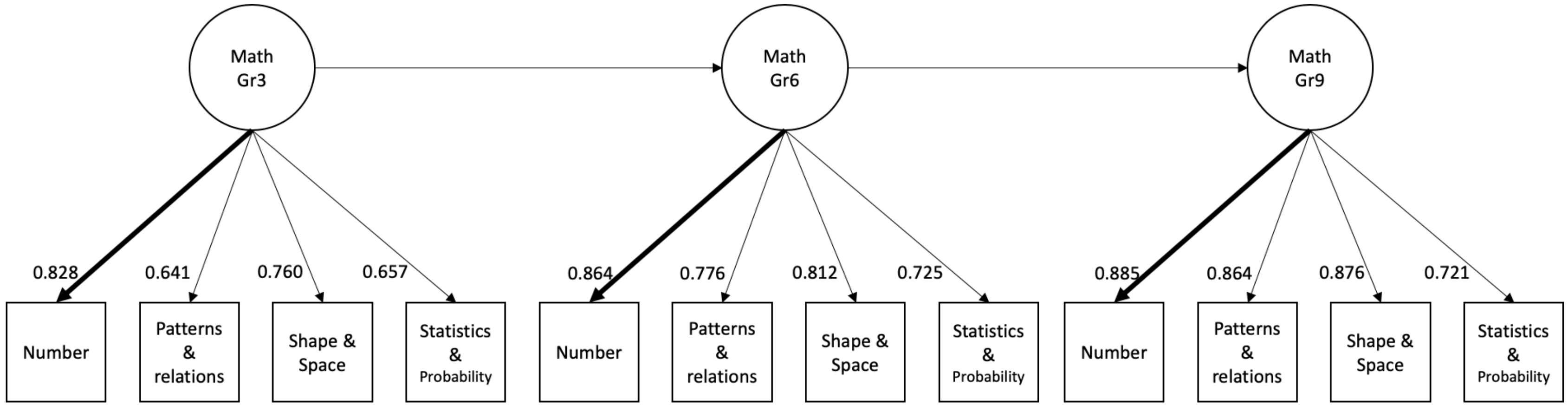

Figure 16. Cohort 2 Longitudinal confirmatory factor analysis for students whose performance declined between Grades 3 and 9. RMSEA = .020, CFI=.997, SRMR=.009.

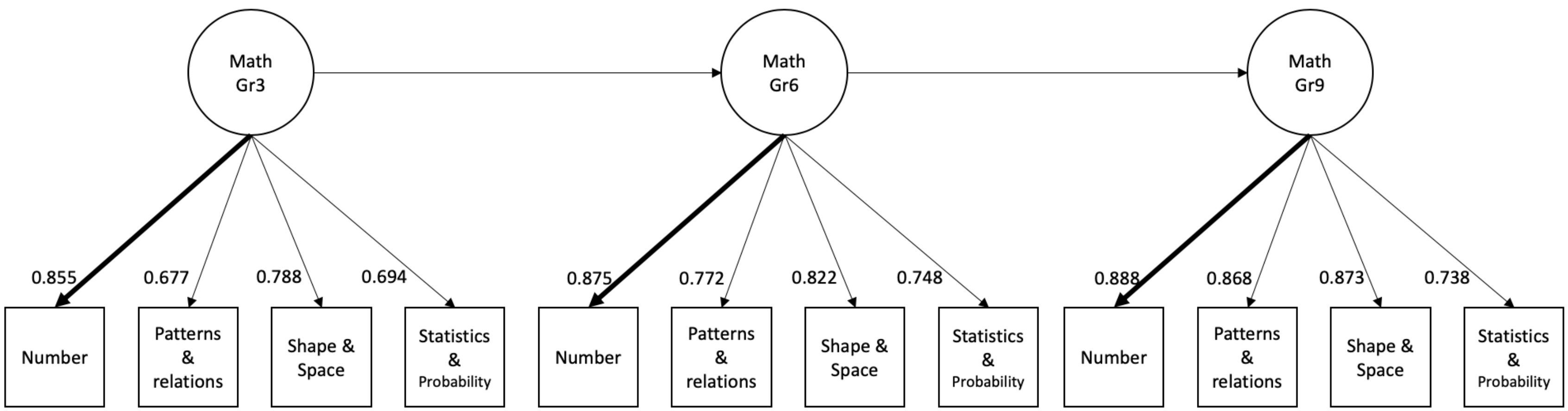

Figure 17. Cohort 2 Longitudinal confirmatory factor analysis for students whose performance did not decline between Grades 3 and 9. RMSEA = .020, CFI=.997, SRMR=.009.

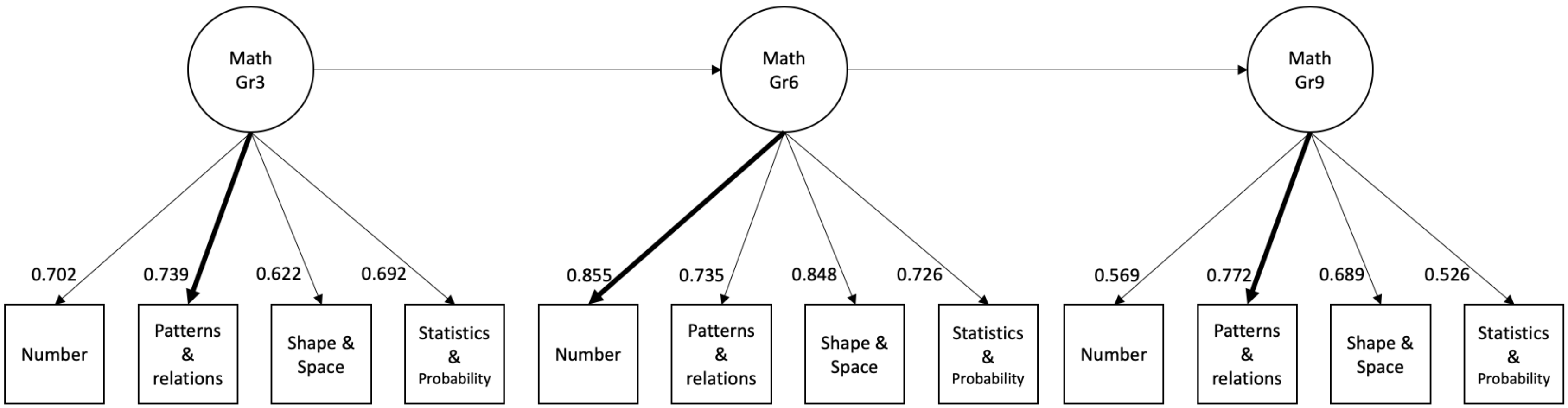

Figure 18. Cohort 1 Longitudinal confirmatory factor analysis for students whose performance declined between Grades 3 and 9. RMSEA = .078, CFI=.964, SRMR=.0030.

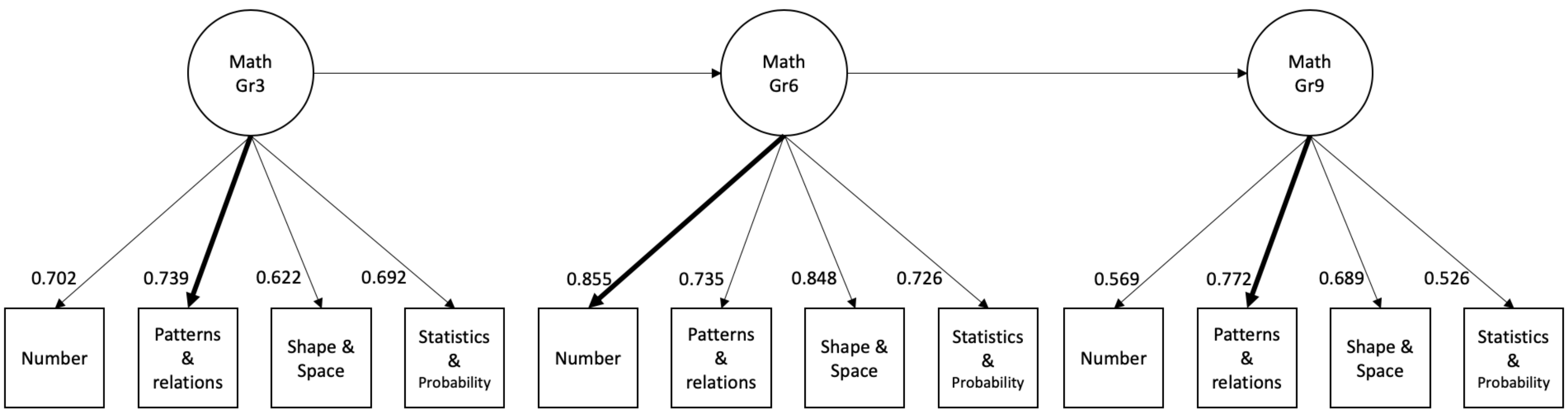

Figure 19. Cohort 1 Longitudinal confirmatory factor analysis for students whose performance did not decline between Grades 3 and 9. RMSEA = .078, CFI=.964, SRMR=.0030..

The confirmatory factor analyses was our third attempt to understanding the predictive capacity of different strands on student performance across the three time points. Overall, we see similar patterns again across the three cohorts, with a few minor differences. For Cohorts 1 and 2, Number was the dominant structure for both the students whose performance declined across the three time points and the students whose performance did not decline. A minor exception occurs in Cohort 1 for students whose performance declined where Patterns & Relations in Grade 9 plays a more prominent role, but only slightly. In Cohort 3, students whose performance declined and those who did not shows the same pattern. That is, performance on Patterns & Relations is the dominant structure, except in Grade 6 where Number plays a more important role.

These differences are relatively minor and all three of our attempts to understand the role of performance across the different strands showed the same results. That is, Number was a more dominant predictor overall for students in Cohorts 1 and 2, while performance on Patterns & Relations played a more central role for students in Cohort 3.

Discussion and Conclusion

Discussion of Findings

This retrospective longitudinal study involved three cohorts with nearly 30 000 students in each cohort. The three cohorts wrote the Grade 3 PAT in June of 2002, 2003, and 2011 respectively. As a whole, the characteristics of the majority of students within each cohort were relatively similar as was the distribution of students across large and small cities, towns, and rural communities; however, Cohort 3 appeared to be more diverse in that approximately 40% of students were coded in Grade 3 as French Immersion, Francophone, ESL, Special Education or as participating in an Online program. This percentage almost doubles the number of students with an additional code from Cohorts 1 and 2. The change in characteristics may suggest that classrooms are becoming more diverse over time, more opportunities for alternative schooling exist, or it may indicate changes to policies regarding how students are coded.

In our analysis of the first research question investigating the trajectories of students' mathematics achievement from Grade 3 to Grade 6 to Grade 9, we saw a very stable pattern across all three cohorts regardless of gender, regions in Alberta, or whether students responded to knowledge or skills questions on the PAT. Across all broad categories we saw a small decrease in performance from Grade 3 to Grade 6, followed by a much greater decrease between Grade 6 and Grade 9. The possible reasons for the declining performance may be due to the change in mathematics curriculum content between elementary and junior high school with increased attention to generalized over computational processes, an increase in the complexity of problems in the upper grades, or factors associated with the potential accumulation and increasing connectedness of skills and knowledge emphasized in mathematics over time.

Although the downward trend in performance was the predominant pattern, our results did reveal that students located in either Edmonton or Calgary performed somewhat better on average than students in other regions, and students in rural areas and towns performing lower on average than students in the other regions. The possible rationale for the difference may be due to potential differences in resources available within large centres compared to smaller areas, and the likelihood that there are mathematics specialists specifically at the Grade 9 level within cities. In smaller schools, teachers may be expected to teach multiple subjects and possibly even split grades making the teaching of mathematics somewhat more challenging and which may have a detrimental impact on student performance.

Our second research question was focused on investigating individual identifiers or characteristics that might be indicative of future performance. Although we had a small number of students who had a coded identifier in Grade 3, we noted that ESL students and French Immersion students were more likely to maintain or improve performance over time. In both student groups we assumed that the initial language barrier may have had a detrimental impact on student performance in Grade 3. However, once students became more fluent in the language of instruction, they were more likely to improve their performance.

The most interesting results stemmed from our third research question, “What aspects of mathematics content predict future performance?” The three techniques we used to investigate this question all revealed very similar results. That is, performance on Number was the best predictor of future performance for Cohorts 1 and 2, and performance on Patterns & Relations was generally a better predictor of future performance for Cohort 3. There was a range of hypotheses as to why these results occurred. Certainly, the change in curriculum may have had an impact. In particular, the increased emphasis on Patterns & Relations between the 1996 and 2007 curriculum documents may be the primary contributing factor. However, this notion was debated. The curriculum frameworks maintained the same structure across all cohorts, so perhaps an alternative rationale for the differences between Cohorts 1/2 and Cohort 3 was simply because it may have taken at least a decade for teachers, particularly at the Elementary level, to become familiar with the content in the Patterns & Relations strand and make it a prominent aspect of their teaching. Similarly, the increased emphasis on the processes, particularly problem solving, may have made Patterns & Relations more prominent in the post-2007 curriculum. A more detailed comparative analysis of the two programs of studies may reveal other dimensions to this debate.

Implications for Practice

The results of the study provide insight into student performance over time and have implications for teacher education, professional development, and curriculum. The downward trend in performance from Grade 3 to 6 to 9 is concerning and does not appear to be ameliorated by small to moderate changes in mathematics curriculum content. By Grade 9 there is a smaller percentage of students who are performing at the acceptable level of performance, particularly in comparison with other subjects including science (Alberta Education, 2018). Based on the literature reviewed previously in the paper, mathematics achievement has wide ranging implications as it is a good predictor for achievement across all academic subjects, acceptance into and completion of postsecondary education, and future earning potential for individuals.

Despite the relatively stable, albeit concerning, pattern in student trajectories, the results of the study revealed a distinction in predictive factors affecting future performance. In the post-2007 curriculum, Patterns and Relations became a stronger predictor of future performance than Number. Given that there was not an apparent change in student performance trajectories, we cannot suggest that increased attention to Patterns and Relations would necessarily improve student performance over time. However, it does suggest that returning to pedagogy centred on Number in the elementary grades will also not improve student performance in Grade 9. The ongoing tug-of-war as to what should be the foundation of mathematics curriculum seems somewhat futile given these results. In fact, it may be detrimental to student performance. A provincial achievement test assesses performance based on curricular outcomes, so pedagogy, resources, and family support needs to be aligned.

Scholarly and Educational Benefits

The study offers a range of both scholarly and educational benefits. It brought together researchers and practitioners to collaborate on a common project. This form of collaboration strengthens the network of educators across the province. The study required the creative utilization of research methods and analytic tools applied to the data from the PATs. Our multiple forms of analyses allowed us to gain insight into our research questions from different perspectives. Doing so shed light on some of the ongoing debates in Alberta regarding past and current mathematics curriculum and pedagogy. By sharing these findings with teachers, administrators, and researchers across Alberta and Canada, we can enrich the public discourse around mathematics teaching and learning and shift the conversation from speculation to data-driven responses.

Recommendations for Future Research

Our study offered a broad overview of possible predictors for student performance on the Provincial Achievement Tests. In addition to our discussion of the results of the three research questions, our project team raised many other questions related to appropriate assessment practices, performance across districts, performance prior to 1990, trends in performance for diploma exams, performance outside of Alberta, and trends in performance across subject areas. These questions were beyond the scope of the available for the project, but could form the basis for future research.

As mentioned at the beginning of the paper, we investigated content using the broad strands of mathematics (i.e., Number, Patterns and Relations, Shape and Space, and Statistics and Probability). A more detailed study focused on more refined categories of mathematics content, processes, or levels of complexity may help further identify predictors of student performance over time, and further information to understand the ongoing decline in student performance over time.

References

Alberta Education (n.d.). Research partnership program. Retrieved from https://www.alberta.ca/alberta-research-network.aspx#toc-1

Alberta Education (2018). Multiyear results (provincial). Retrieved from https://www.alberta.ca/provincial-achievement-tests.aspx#toc-2

Alberta Government (2017). Numeracy progressions. Retrieved from https://education.alberta.ca/media/3402196/num-progressions.pdf

Alberta Education (2016). Summary of clarifications to the Alberta Mathematics Kindergarten to Grade 9 Program of Studies. Retrieved from https://education.alberta.ca/media/3115246/2016_sum_of_clar_to_k_to_9_math_pos.pdf

Alberta Government (2016a, Dec.). Provincial achievement test multiyear reports. Retrieved from https://education.alberta.ca/provincial-achievement-tests/pat-results/

Alberta Government (2016b). New supports to help students and teachers with math. Government of Alberta Press Release. Retrieved from https://www.alberta.ca/release.cfm?xID=4496751C95F9A-C376-BDD7-5D226ED2F103D97E

Bailey, D. H., Siegler, R. S., Geary, & D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775-85. doi:10.1111/desc.12155

Bentler, P. M. (1990). Comparative fit indexes in structural equation models. *Psychological Bulletin, 107*, 238-246.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of testing model fit. In K. A. Bollen & J. S. Long (eds.), *Testing structural equation models* (pp. 445–455). Newbury Park, CA: Sage.

CBC News (2016, Oct 7). Low Edmonton Grade 6 math results = sour note for Alberta education minister. Retrieved from http://www.cbc.ca/news/canada/edmonton/low-edmonton-grade-6-math-results-sour-note-for-alberta-education-minister-1.3796383

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.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE*, 8, e54651. doi:10.1371/journal.pone.0054651.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 10,* 333-351.

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher*, *41*(2), 43-55.

Mix, K.,&Cheng, Y. (2012). The relation between space and math: Developmental and educational implications. In J. B. Benson (Ed.), *Advances in child development and behaviour* (Vol. 42, pp. 197–243). Burlington: Academic Press, Elsevier Inc.

Muthén, L. K. and Muthén, B. O. (1998-2015). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritchie, S. J., & Bates, T. C. (2013). Enduring Links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*(7), 1301–1308. doi:10.1177/0956797612466268

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science 23*(7), 691-697.

Simmt, E., F. Glanfield, M. Gierl, M. Hauk, R. Johnson, K. McCabe, J. Mgombelo, E. Mowat, E. Pinco, B. Quinn& S. Sookochoff. (1999). The teaching practices project: Research into teaching practices in Alberta schools that have a history of students exceeding expectations on Grade 9 provincial achievement tests in mathematics. A report to Alberta Learning, Edmonton, AB. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.4089&rep=rep1&type=pdf

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420.

Appendix A

Table A1

Current Budget

Items	Total Funds Used to Date	Remaining Funds
Collaborator discussion		$1 500.00
Graduate Research Assistant	$7372.19	$8227.81
Technology and Media Design Support		$4 500.00
Dissemination • Mileage • Photocopying • Food/catering		$2 500.00
TOTALS	$7 372.19	$16 727.81

Note. The remaining funds will be used to support knowledge mobilization including travel, technology, and graduate assistantship for presentations for partner school authorities, a provincial symposium, and academic presentations and publications.

Appendix B

Table B1

Research Project Timeline and Knowledge Mobilization Plan

Dates	Description
May 2017 - Nov 2017	Preliminary tasks: • Data request • Ethics application • Research plan • Access to data
Dec 2017 - April 2018	Data: • Identifying students • Initial data modelling Interim Report: • Due to Alberta Education on or before March 1, 2018 Collaborator discussion (April 10, 2018): • Overview of data available • Questions to ask of data
May 2018 - April 2019	Data: • Complete data analysis Collaborator discussion: • Share results through email with opportunity for feedback

May 2019 – June 2019	Results ● Prepare draft of final report Draft of Final Report: ● Submitted June 3, 2019 Final Report: ● Submitted June 27, 2019
June, 2019 – June, 2020	Knowledge Mobilization: ● Presentation of project at the Annual Meeting of the Mathematics Council of the Alberta Teachers Association. ● Provincial symposium hosted by CMASTE ● Presentations at participating school authorities ● Academic conferences and publications