When children succeed: an evaluation of one and a half years of strategic intervention to close

the achievement gap

Derek J. Gaudet, M.A.

The University of New Brunswick Saint John

Abstract

This is an evaluation of the 2018-2020 attempt to close the achievement gap in Saint John priority schools. The program involved adding additional teaching staff to lower the student to teacher ratio. Priority school students in kindergarten and grade 1 were tracked over a year and a half of the program and their proficiency in literacy and numeracy was compared to that of students from non-priority schools. The results of the evaluation suggest that there was some evidence of fully closed or narrowed achievement gaps, particularly in the kindergarten cohort. Literacy and numeracy benchmarks, as well as the role of attendance in student success, was also examined within the priority schools. Limitations and suggestions for future work are also discussed.

Keywords: closing the achievement gap, literacy and numeracy

When children succeed: an evaluation of one and a half years of strategic intervention to close the achievement gap

During the 2018-2019 academic year, a strategy to close the literacy and numeracy achievement gap for schools in priority neighborhoods within the city of Saint John was implemented. Each Kindergarten, Grade 1, and Grade 2 class was provided with one additional teaching staff, at each of the seven targeted priority schools. The rationale for the extra teaching staff was that the additional support would lead to smaller student to teacher ratios, and therefore more one-on-one interaction between the teachers and each of the students. This, in turn, would facilitate learning and hopefully translate to better literacy and numeracy performance in students attending these schools.

The manner in which each additional teaching staff was used at each school to reduce the student to teacher ratio varied between schools, but two general strategies were applied. In some cases, teachers decided to follow a co-teaching model, while in other cases additional classes were created with a lesser number of students. Both strategies resulted in what was considered the primary intervention, a reduction of the student to teacher ratio.

To determine whether or not literacy and numeracy achievement gaps were closing, we needed to rely on the measures of numeracy and literacy that were readily available to us. The first analyses conducted examined report card grades that use standards-based grading. Students are evaluated by their teachers as to whether or not they are meeting certain standards set by the province and assigned a grade of 1, 2, 3, 3+, 4, or 4+, depending on where their abilities fall. A grade of "3" is considered to be "meeting standards". These grades were then compared to those of students from non-priority schools. As the achievement gap represents the difference between priority and non-priority school achievement level, the goal was to demonstrate that the priority

school children were approaching the achievement level of their non-priority school counter

parts.

As the analyses were quantitative, a continuous numerical measure of progress was

required. The provincial report card format uses a categorical system that places students in one

of the six categories. Luckily, this categorical system is ordinal in nature, so we transformed

grades of "3+" and "4+" into "3.5" and "4.5", respectively. This is because the "+" appears to

denote that the student falls somewhere in between a "3" and a "4" or is excelling beyond the

standards of a grade of "4". Though not true continuous measures, these values served as a proxy

to evaluate differences using statistical models and tests. The outcomes of these analyses allowed

us to see whether or not differences between the priority and non-priority schools were real, and

whether or not they were showing a pattern that suggested the achievement gap was closing.

Additional measures of numeracy and literacy examined here include the use of literacy

and numeracy benchmark assessments. These data have no comparison group, so while we can

assess student progress within priority schools, it is not possible to determine whether or not the

achievement gap in closing or has closed. Instead, we can examine whether or not students

appear to be making sufficient progress.

Results

The following results are separated by grade and only include those priority school

student cohorts that had received the intervention from September 2018 until March of 2020.

This allowed us to track student progress over five points of assessment. The priority school

student cohorts examined here are the 2018-2019 kindergarten class up until their second report

card assessment in grade 1 and the 2018-2019 grade 1 class up until their second report card

assessment in grade 2.

In terms of literacy, three provincial report card items were considered which represent different aspects of literacy. Reading and Viewing, Speaking and Listening, and Writing and Representing grades were all assessed separately to evaluate literacy proficiency over the five assessment points. This was done for both the kindergarten students and the grade 1 students. Numeracy, on the other hand, required collapsing the available report card data for 3 items into a single average based on whatever data was available. This was done because there was a large amount of non-randomly missing data, that can likely be explained in terms of assessment differences between schools or grades. In some schools or grade levels, one (or more) of the three grade items were missing for a large number of students. The result was one average numeracy grade based on available data.

Kindergarten Literacy

Over the year and a half program, there are a few examples of the literacy and numeracy achievement gaps closing between schools. The graphs presented below demonstrate some of these effects. We should be cautious when interpreting the results of the 2019-2020 academic year, as the pandemic likely created additional stressors at the school and within the home that likely had an effect on student performance in both priority and non-priority schools. More through and in depth analyses can be found in the APPENDIX, which include the results of all statistical tests that were conducted in this evaluation.

Reading and Viewing. This first piece of data demonstrates the achievement gap closing between priority and non-priority schools by the end of the first year. Statistical tests indicate that there is no difference between priority and non-priority schools from 2018-2019 T3 until the end of the intervention period. Before celebrating, however, we do need to consider a few other factors here. First, non-priority school students (the group to which our priority schools were

compared) Reading and Viewing grades followed a downward trend when considering all five

assessment points. It is likely that priority school students did not differ from non-priority

students in the 2019-2020 academic year simply because non-priority school students worsened

in performance and came down to the level of priority school students. Supporting this

interpretation is the fact that statistical test showed no difference in Reading and Viewing Grades

between the first and end of the program in priority school students. If, however, we isolate the

first year, this would be evidence of the achievement gap (as measured by reading and viewing
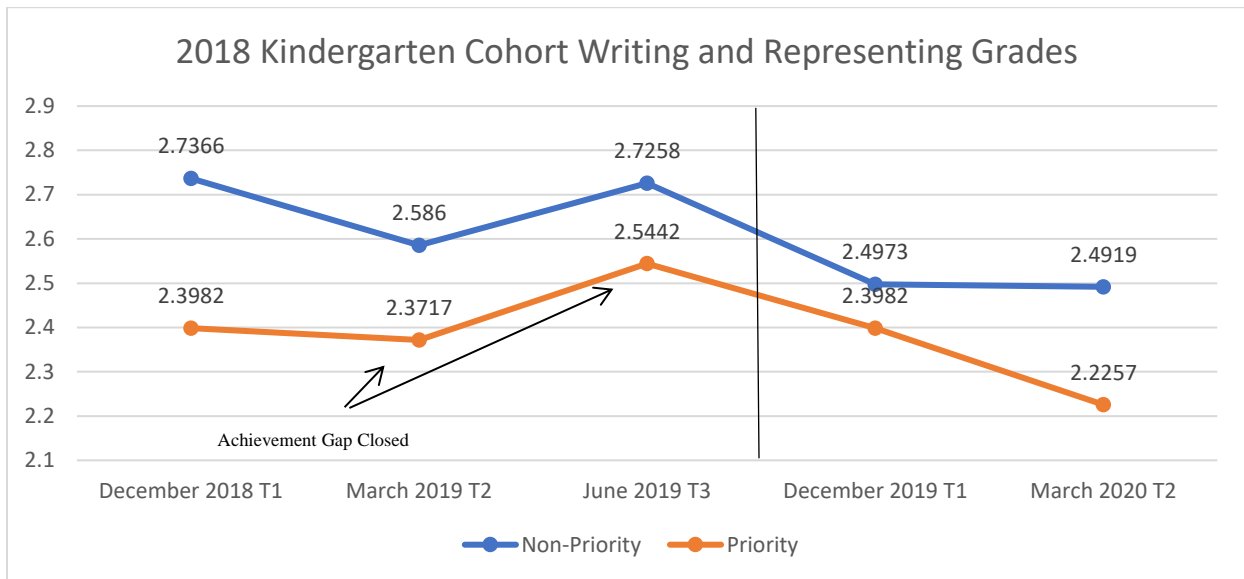
grades) closing.

2018 Kindergarten Cohort Reading and Viewing Grades
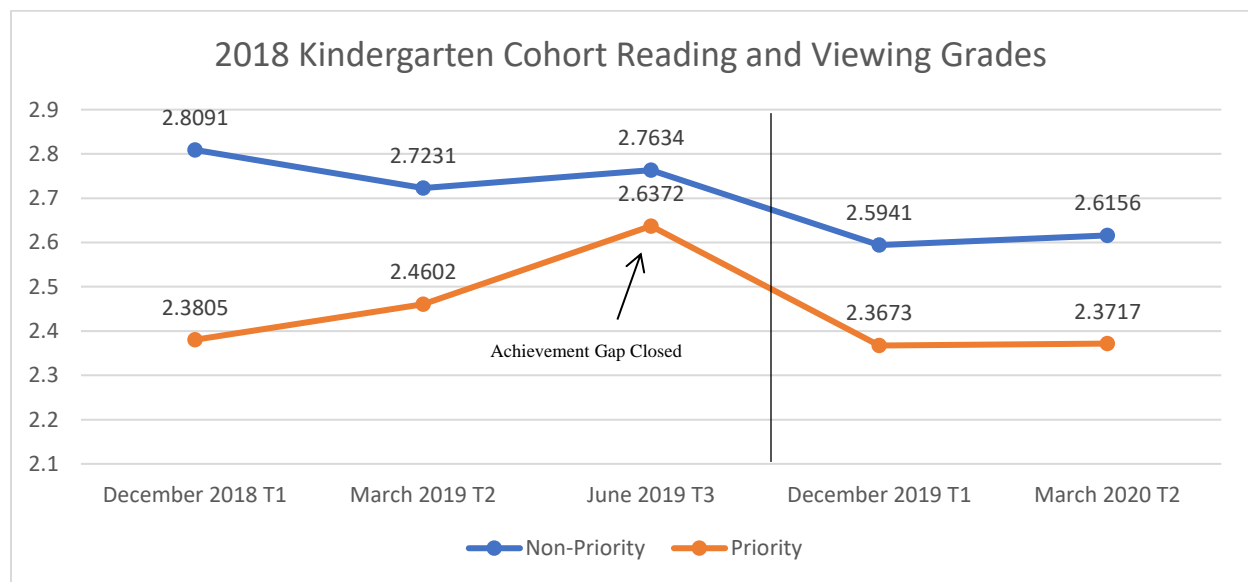
Speaking and Listening. According to the results of statistical tests, the achievement gap

in Speaking and Listening grades did not close throughout the intervention period. If we examine

the graph, however, we see a pattern that suggests the gap was narrowing. Priority school

students also made statistically significant gains from the beginning of the intervention to the

end. It is possible that a further closure of the achievement gap in speaking and listening grades

may have resulted had the program continued. Despite the presence of a gap at all five

assessment points, the pattern suggests and upward trend in priority school kindergarten student literacy proficiency level.
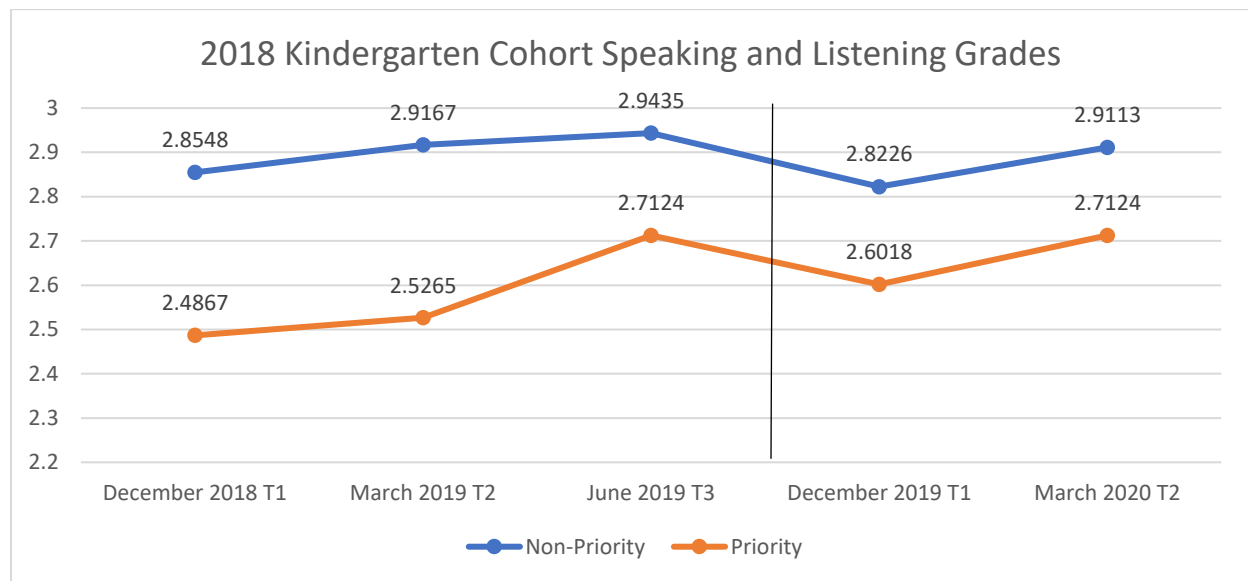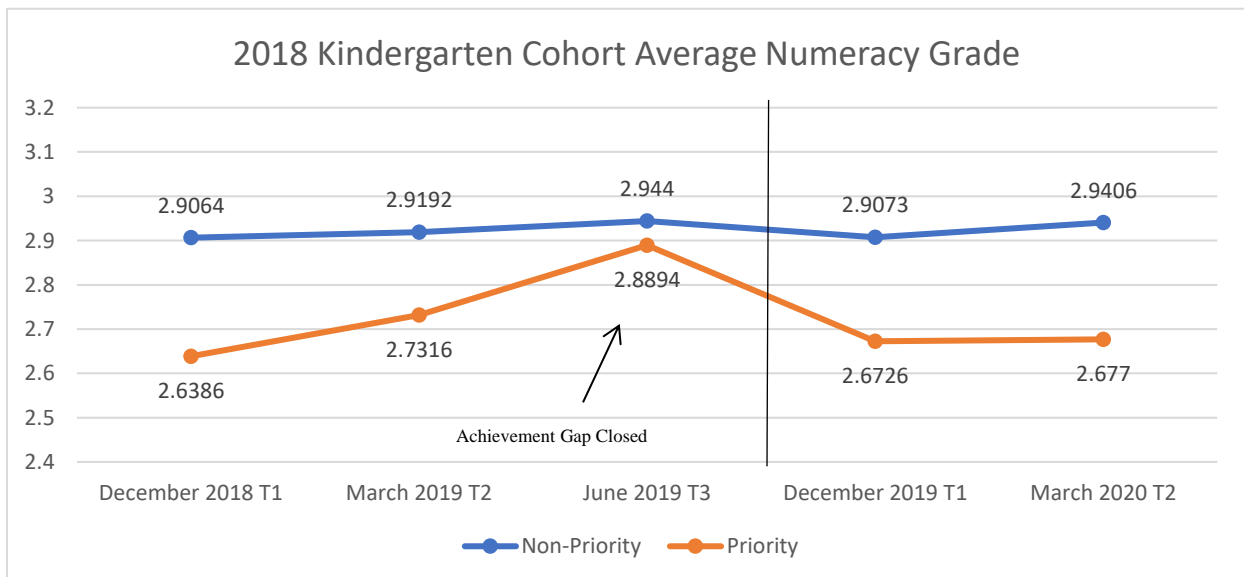

Writing and Representing. The results of statistical tests suggest that the achievement gap for Writing and Representing grades had closed between 2018-2019 T2 and 2019-2020 T1 but opened again at 2019-2020 T3. We again need to be cautious of the fact that the non-priority school writing and representing grades demonstrate a decrease in proficiency over time. It is possible that the gap first closed at 2018-2019 T2 because of the decrease in proficiency seen in the non-priority school students. The fact that non-priority school students show a subsequent rebound in proficiency at 2018-2019 T3, and that this coincides with an increase in the priority school students writing and representing grades suggest that a closing of the achievement gap likely occurred in writing and representing grades. The decrease in writing and representing proficiency seen in the non-priority school children during the program may be the result of progression to a new grade with a new teacher, or because the 2019-2020 academic year had numerous other complications to account for. Isolating the first year suggests a pattern of a

closing achievement gap between priority and non-priority schools in terms of Writing and
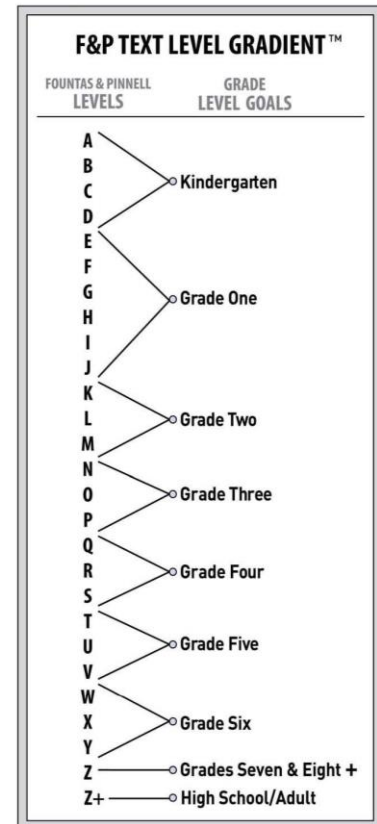
Representing grades.

2018 Kindergarten Cohort Writing and Representing Grades

Non-Priority values: December 2018 T1: 2.7366; March 2019 T2: 2.586; June 2019 T3: 2.7258; December 2019 T1: 2.4973; March 2020 T2: 2.4919

Priority values: December 2018 T1: 2.3982; March 2019 T2: 2.3717; June 2019 T3: 2.5442; December 2019 T1: 2.3982; March 2020 T2: 2.2257

Achievement Gap Closed

Kindergarten Numeracy

Average Numeracy Grade. Examining the achievement gap in numeracy, there was

evidence that the achievement gap had closed by the end of the 2018-2019 academic year but

opened back up again when the kindergarten student progressed to grade 1. Examination of the

graph below shows a clear closure of the numeracy achievement gap, in terms of an average

numeracy grade, by 2018-2019 T3. Though there is no evidence of it here, it is entirely possible

that this pattern may have repeated itself if the 2019-2020 academic year had finished normally.

2018 Kindergarten Cohort Average Numeracy Grade

Non-Priority values: December 2018 T1: 2.9064; March 2019 T2: 2.9192; June 2019 T3: 2.944; December 2019 T1: 2.9073; March 2020 T2: 2.9406

Priority values: December 2018 T1: 2.6386; March 2019 T2: 2.7316; June 2019 T3: 2.8894; December 2019 T1: 2.6726; March 2020 T2: 2.677

Achievement Gap Closed

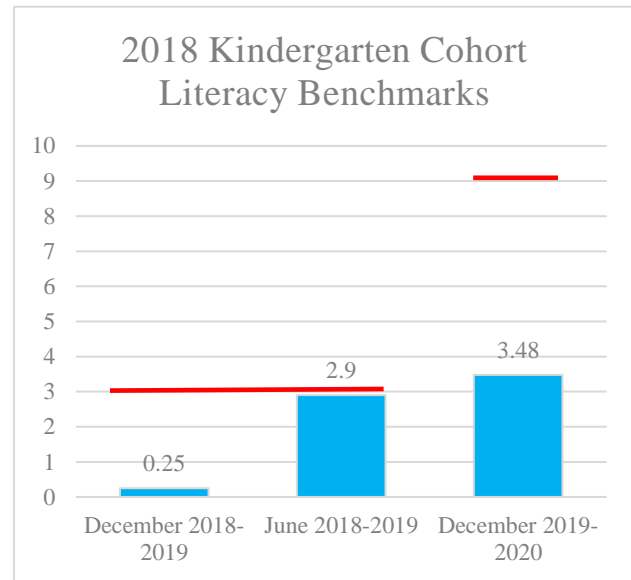**Kindergarten literacy and numeracy benchmarks**

> *Kindergarten Literacy Benchmark.* Literacy
proficiency was also assessed using the Fountas and Pinnell
reading levels. These levels range from "Pre-A" (which is
indicative of no reading ability) to "Z+", which is
considered to be high school or adult reading level. A
figure depicting expect reading levels for each grade can be
found below. Children in kindergarten are expected to
achieve an "A", "B", "C", or "D" level of reading
proficiency. Children in Grade 1 are expected to achieve an
"E", "F", "G", "H", "I", or "J" level of reading proficiency.
Children in Grade 2 are expected to achieve a "K", "L", or
"M" level of reading proficiency. The reading level is
determined by having children read various small books corresponding to each reading level.
The evaluator then determines where the student falls in terms of the Fountas and Pinnell reading
levels. For the purpose of statistical analyses, the Fountas and Pinnell reading levels were
assigned a numerical value as a proxy. As a student progresses through the letters, they are
considered to have a better grasp on literacy. This means that the values are ordinal in nature.
The letter values were substituted for numerical values in the following way. "Pre-A" was
assigned a value of "0", "A" a value of "1", "B" a value of 2, ..., and "Z", a value of "26". When
examining the graphs below, we would round to the nearest whole number and assign the letter
score corresponding to the numerical value.

F&P TEXT LEVEL GRADIENT™

FOUNTAS & PINNELL LEVELS   GRADE LEVEL GOALS

A
B
C — Kindergarten
D
E
F
G — Grade One
H
I
J
K
L — Grade Two
M
N
O — Grade Three
P
Q
R — Grade Four
S
T
U — Grade Five
V
W
X — Grade Six
Y
Z — Grades Seven & Eight +
Z+ — High School/Adult

The grade level goals on the F&P Text Level Gradient™ are intended to provide general guidelines, which should be adjusted based on school/district requirements and professional teacher judgement.

© 2012 Irene C. Fountas and Gay Su Pinnell

Here, our 2018-2019 kindergarten cohort was shown to make continuous progress throughout the program. By the end of the kindergarten year, children were, on average, at a "C" level of proficiency, which meets Fountas and Pinnell's suggested standards for kindergarten. During their grade 1 year, only 1 assessment was made. Students

2018 Kindergarten Cohort Literacy Benchmarks

continued to be at a "C" level, on average, which falls below the suggested standards. The red lines in the graph represent the expected standards for kindergarten and grade 1. It is possible that students may have met criteria by the end of the year during their second assessment, but there is no way to know this for sure as the assessment could not be completed.

As previously described, the schools did not implement the additional teaching staff in a standardized way, so there is always the possibility that not all priority schools showed equal gains. Breaking down the literacy benchmarks by school, in the graph presented below, we can see some cases where schools outperformed others. A continuing pattern through all three assessments was that Hazen-White St. Francis was underperforming the other schools. Statistical tests, however, revealed that this observed difference was only statistically significant (i.e., a real difference) when comparing Hazen White St. Francis to either Glenn Falls or Prince Charles. Another point of note is that Hazen white St. Francis fell below the grand mean (overall average) of literacy benchmarks (represented by the black horizontal line) at every assessment. Additionally, by the June 2018-2019 assessment, a minimum of "C" (numerical value = 3) is expected. Here, we can see that most of the schools were only just meeting criteria.
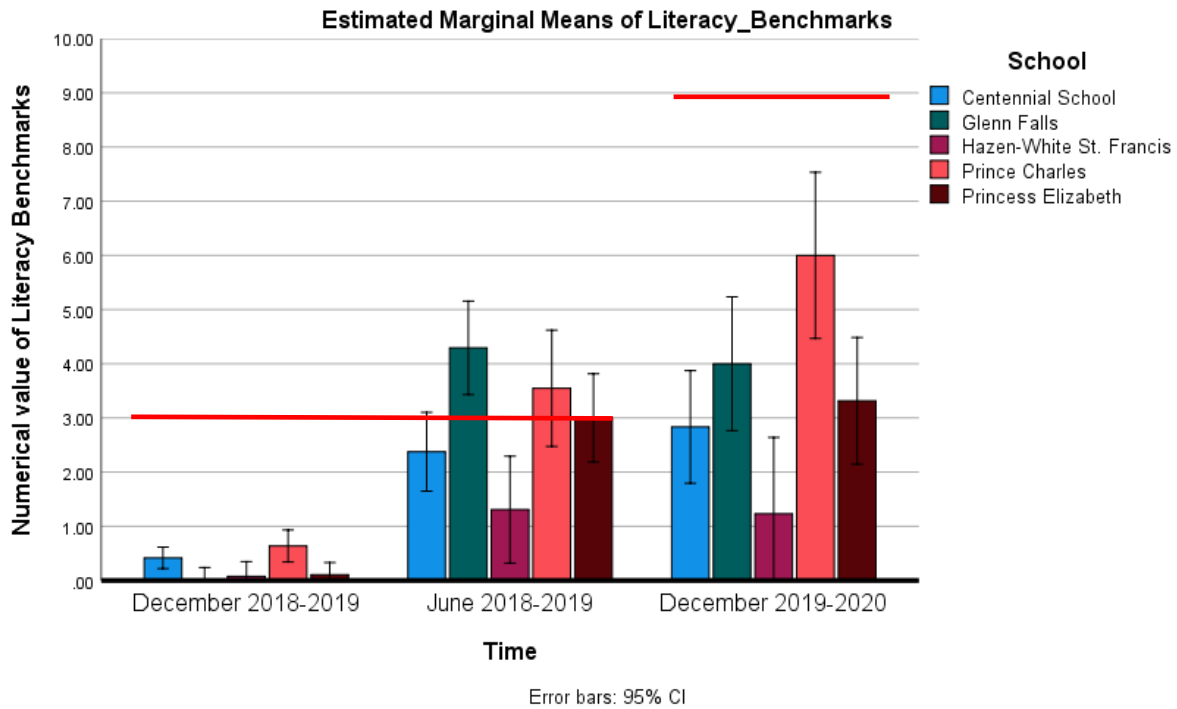
Numeracy Benchmarks. Numeracy benchmarks were assessed by having students respond to several math questions and then calculating the proportion that they had answered correctly. The initial assessment was used to determine where students were in term of their math skills prior to being taught. We see here that kindergarten students answered
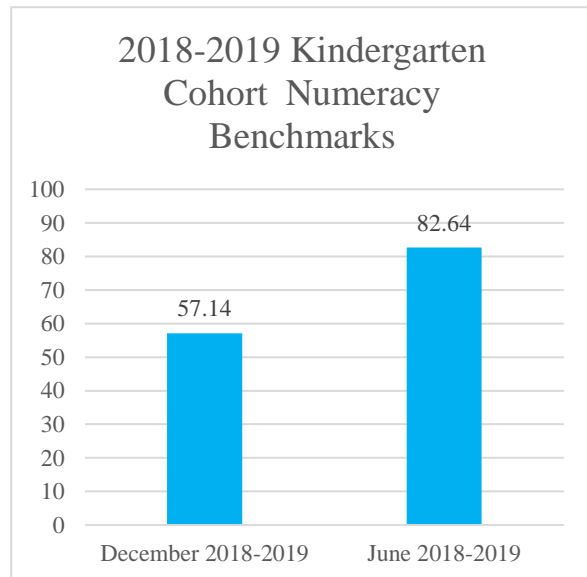
57.14% of the questions correctly when collecting this baseline data. By their second assessment, which took place at the end of the year, students had improved considerably, answering 82.64% of the questions correctly. This is a gain of 25.5%. Students are supposed to be able to achieve a score of 100% correct on this assessment, but these gains are not trivial.

Again, we can break these benchmarks down by school. The differences between schools, in this case, are not quite so dramatic. In fact, only one of the comparisons showed a statistically significant difference between schools (Princess Elizabeth Vs. Centennial), but overall the graph suggests minimal differences between schools, and that most schools improved numeracy proficiency by the second assessment.

Estimated Marginal Means of Numeracy_Benchmarks

Error bars: 95% CI
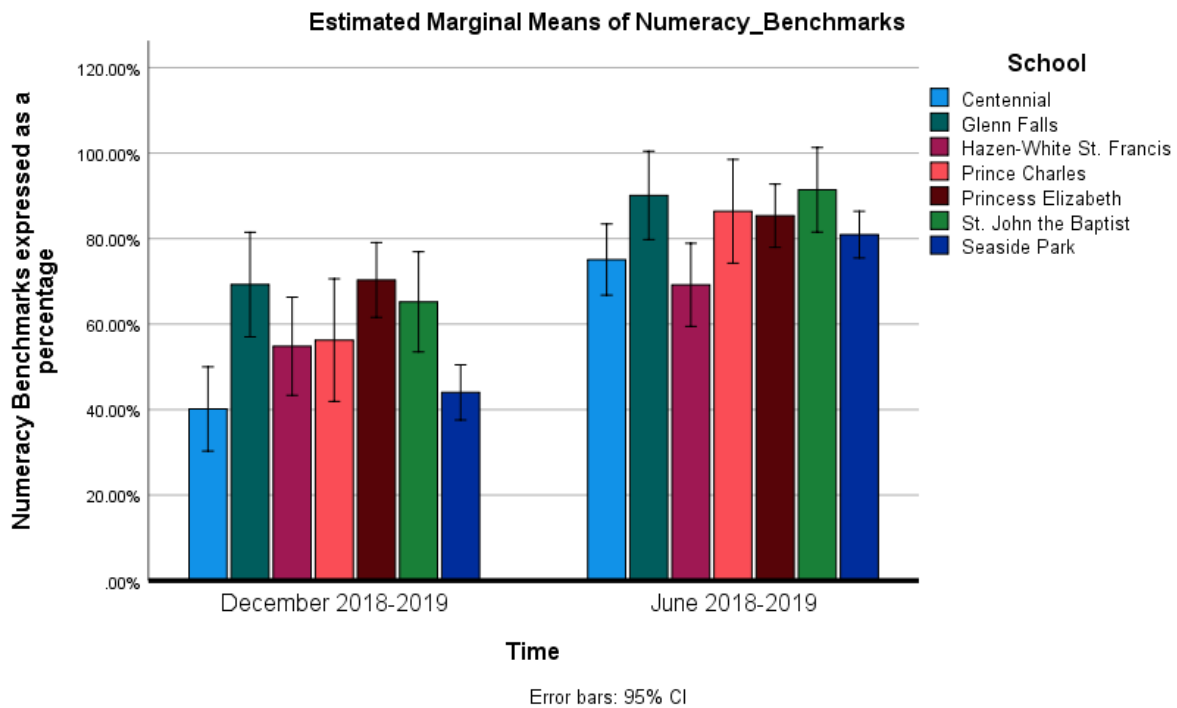
The importance of attendance in Kindergarten

It is reasonable to assume that absenteeism would have an effect on the learning process in young children. Afterall, if the students are not attending school, they are not able to fully benefit from school. What we are less sure about, is the extent to which attendance is important in learning during the kindergarten (and grade 1) academic years. To address some aspects of this question, we ran Pearson's correlations between the number of days absent and each of the literacy and numeracy benchmarks for each academic year. Pearson's correlations measure the strength and direction of a relationship between two variables (literacy/numeracy benchmarks

and attendance). We also need to note whether a correlation is positive or negative. If the correlation is positive, this means that as one variable increases, so does the other. If a correlation is negative, it means as one variable increases, the other decreases. Correlations can also give us an idea of one variable's (attendance) overall contribution to another variable (numeracy/literacy proficiency). This is done by squaring the correlation coefficient (the "*r*" value in the table). When the *r* value is squared, it tells us the proportion of variation in values accounted for by the other variable. The table below shows the results of these correlations.

Table 3. Correlations between absenteeism data and benchmark data for kindergarten

	December 2018-2019 Literacy Benchmark	June 2018-2019 Literacy Benchmark	December 2018-2019 Numeracy Benchmark	June 2018-2019 Numeracy Benchmark
2018-2019 Absenteeism	$r = -.545$ $p = .002$ $N=30$	$r = -.618$ $p < .001$ $N=30$	$r = -.502$ $p = .005$ $N=30$	$r = -.657$ $p < .001$ $N=30$
	December 2019-2020 Literacy Benchmark	**December 2019-2020 Numeracy Benchmark**		
2019-2020 Absenteeism	$r = -.226$ $p = .01$ $N=115$	$r = -.292$ $p = .002$ $N=115$		

In the table above, the first thing we must note is that all of the relationships are statistically significant, which means that they are likely not to be "false positive findings". This is because the *p* - value is less than, or equal to, .01 in all cases. The second thing to note, is the direction of the relationship. All *r* values are negative, which means the variables (benchmarks and attendance) share a negative relationship. In the context of the current analysis, this means that ***higher levels of absenteeism are associated with lower benchmark scores*** in kindergarten students. These findings are true for these students both in their kindergarten academic year and their grade one academic year.

The next thing to consider here, is the size of the relationship. For this, we look at the r value which can range between 0 (no relationship) and 1(perfect relationship). As a rule of thumb, values between .1 and .3 are considered small effects, those between .3 and .5 are considered medium effects, and those larger than .5 are considered large effects. Looking first at the 2018-2019 academic year (the first rows in the table above), All of the relationships are large. Squaring the r values tells us that attendance alone accounted for between 25.20% and 43.20% of the variance in literacy and numeracy benchmark scores. This is a large amount of variation, and suggests that during the kindergarten year, attendance is quite important. When we look at these students grade 1 year, we see a considerable dip in the r value's size. This means that the strength of the relationship between attendance and benchmarks went from large in the kindergarten year to small in their grade 1 year. Despite this drop, attendance is still a relatively important contributor to learning. Attendance accounted for 5.12% of literacy benchmark scores and 8.53% of numeracy benchmark scores. Although the contribution of merely attending class to learning may drop in grade 1, it is still a substantial part of the learning process and should not be ignored.
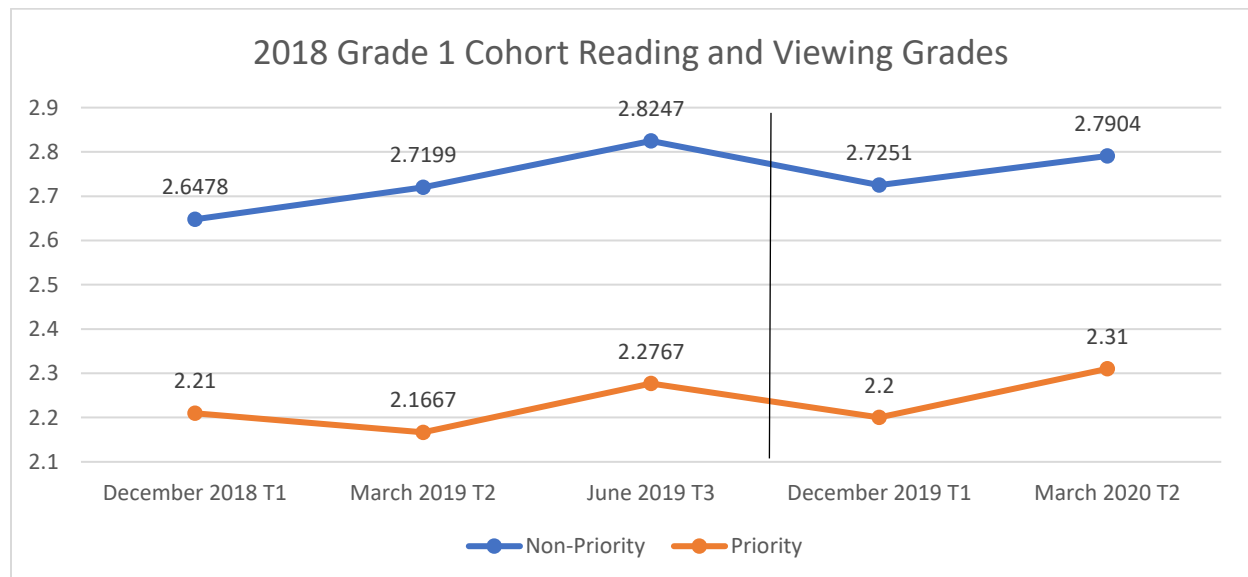
Grade 1 Literacy

Although kindergarten showed several examples of narrowing or completely closed achievement gaps, the results for grade 1 students are less promising. There was only a single example of the achievement gap having closed, and extremely limited evidence that the gap was narrowing with the other grade items. Though this is disheartening, it is also important to consider the potential reasons that the findings did not parallel those of the kindergarten students. It is possible that the results indicate that early intervention (in kindergarten) is necessary for the full benefits of the intervention to be realized. Alternatively, it is possible that the program just

does not work past the kindergarten level. A third, but also possible interpretation, is that the results are not the product of the program but are completely random. As the intervention was not well controlled, it is difficult to conclude with a high degree of certainty that the extra teaching staff (reduced student-teacher ratio) is responsible for the improvements seen here. More research would need to take place to determine which of these interpretations are correct.

Here we will examine more closely the findings from following the 2018-2019 grade 1 cohort through to their grade 2 year.

Reading and viewing. In terms of Reading and Viewing grades, the 2018-2019 grade 1 cohort showed no signs of a narrowing achievement gap. While grade 1 students in non-priority schools demonstrated a statistically significant increase in Reading and Viewing grades by the end of the intervention period, the priority school students did not. Grades of priority school students were consistently lower than their non-priority school counterparts at every assessment point and are considered to be "Approaching learning goals".

2018 Grade 1 Cohort Reading and Viewing Grades

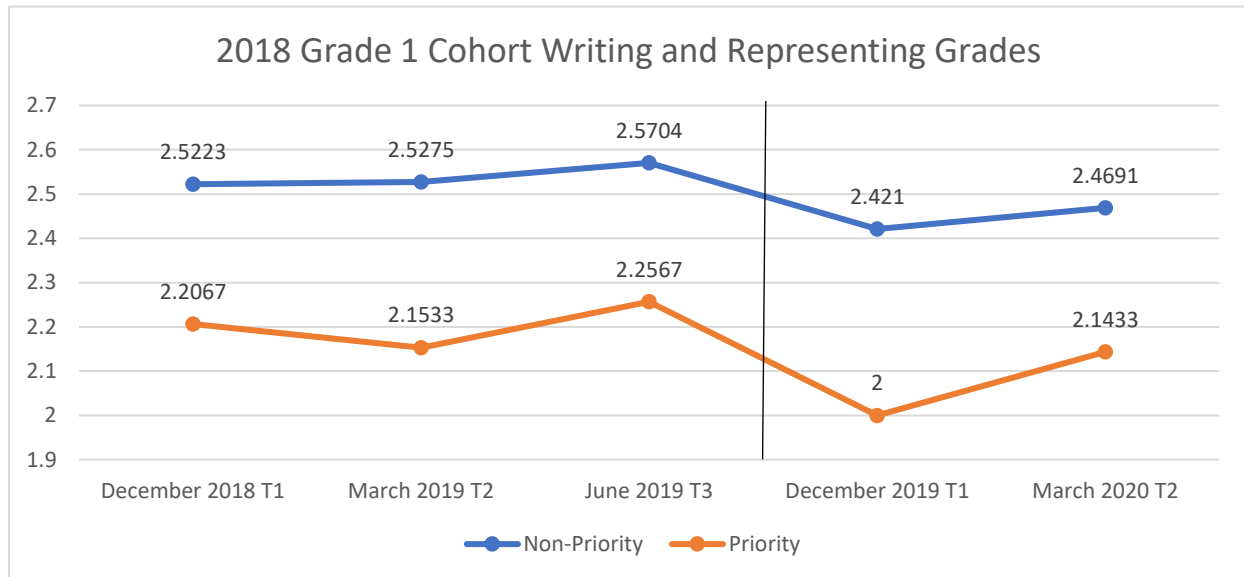| Assessment	Non-Priority	Priority
December 2018 T1	2.6478	2.21
March 2019 T2	2.7199	2.1667
June 2019 T3	2.8247	2.2767
December 2019 T1	2.7251	2.2
March 2020 T2	2.7904	2.31

Speaking and listening. In terms of speaking and listening grades, there is evidence of the achievement gap having closed by the end of the program. At the T2 2019-2020 assessment

period, the difference between speaking and listening grades in priority and non-priority schools

was not statistically significant, meaning that there was no difference in proficiency detected by

our statistical tests. Both groups of grade 1 students (priority and non-priority school students)

showed a statistically significant increase in grades by the end of the period. These two bits of

evidence suggest that the achievement gap, as measured by speaking and listening grades, had
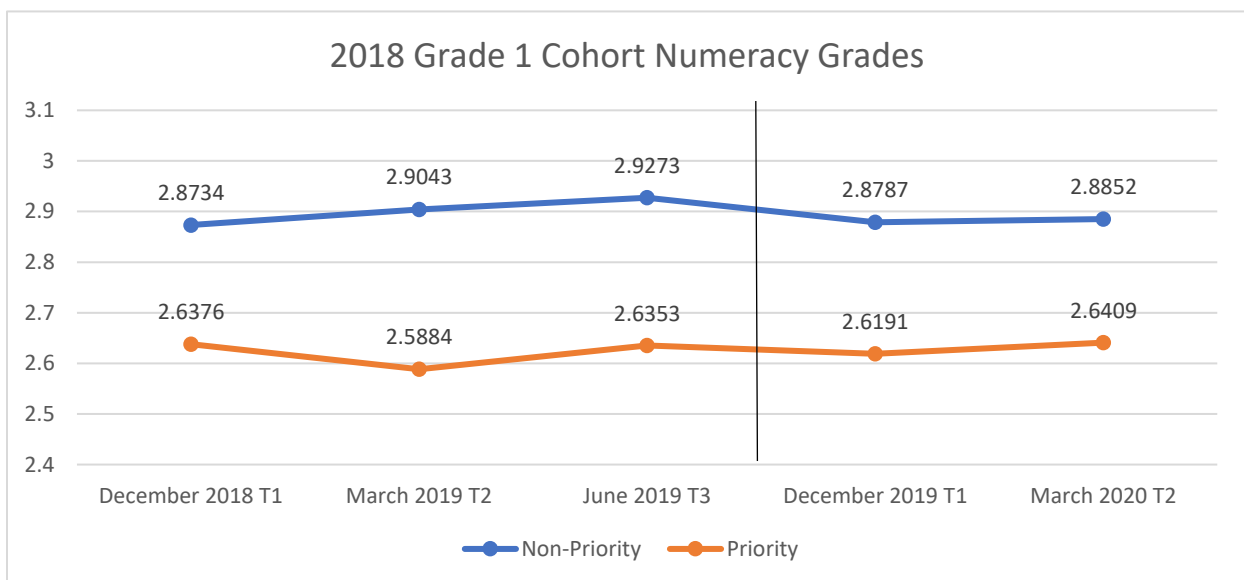
closed at the end of the program.
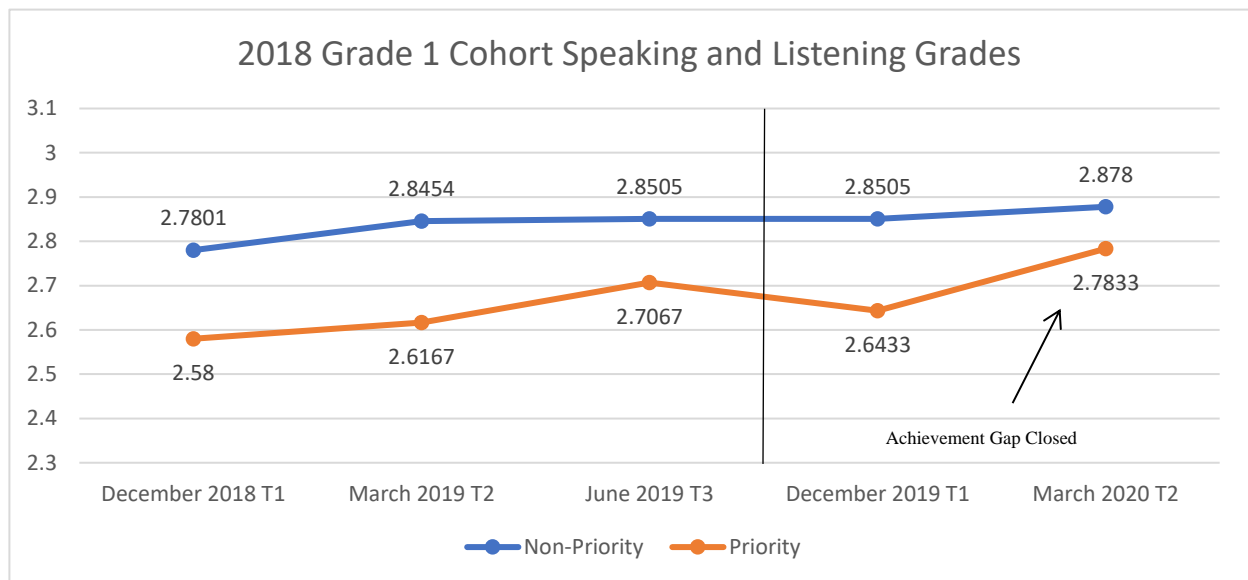
Writing and representing. Again, no evidence was found that the achievement gap had

closed for grade 1 students in priority schools during the intervention program. In fact, neither

school showed a statistically significant change in Writing and Representing grades from the

beginning to end of the program. Overall, the grades showed little signs of change throughout

the duration of the program. This was true in both our Priority and non-Priority school groups.

2018 Grade 1 Cohort Writing and Representing Grades

Non-Priority: 2.5223 (December 2018 T1), 2.5275 (March 2019 T2), 2.5704 (June 2019 T3), 2.421 (December 2019 T1), 2.4691 (March 2020 T2)

Priority: 2.2067 (December 2018 T1), 2.1533 (March 2019 T2), 2.2567 (June 2019 T3), 2 (December 2019 T1), 2.1433 (March 2020 T2)

Grade 1 Numeracy

Average numeracy grades. As was the case with the majority of literacy grade items examined, there was also no evidence that the achievement gap had closed for numeracy during the program intervention. Priority school students under-performed their non-priority school counterparts at every assessment point and neither group showed a statistically significant change in grades from the beginning of the program until the end.

2018 Grade 1 Cohort Numeracy Grades

Non-Priority: 2.8734 (December 2018 T1), 2.9043 (March 2019 T2), 2.9273 (June 2019 T3), 2.8787 (December 2019 T1), 2.8852 (March 2020 T2)

Priority: 2.6376 (December 2018 T1), 2.5884 (March 2019 T2), 2.6353 (June 2019 T3), 2.6191 (December 2019 T1), 2.6409 (March 2020 T2)

Grade 1 literacy and numeracy benchmarks

Note that everything that was previously explained about the literacy and numeracy benchmarks applies here as well. In terms of literacy benchmarks, grade 1 students are simply expected to be at a higher reading level than kindergarten students at their first assessment. Numeracy benchmarks, however, are not required to be at a higher level. Although Literacy benchmarks can be tracked from one year to the next, with the expectation that the levels will increase with grade level as literacy skills increase, numeracy benchmarks test the students each year on material that is specific to the requirements of that particular grade. For this reason, kindergarten literacy can be compared to grade 1 literacy, then again to grade 2 and so on, but kindergarten numeracy cannot be compared to grade 1 numeracy because the students are tested on unrelated materials.

*Literacy benchmarks.* In the graph below, we see gains being made at each of the assessments. Each of these gains were considered to be statistically significant and therefore "real" changes in literacy proficiency. We see that the students started at approximately a "C" level and then progressed to a "F" level by their next assessment. Their third assessment could be considered a "G"

2018-2019 Grade 1 Cohort Literacy Benchmarks

level, should we decide to round up to the next level. By many standards, these scores are below expectations, but the important thing to consider is that there is consistently progress being made in literacy proficiency. There is also much debate over the reliability and validity of the Fountas
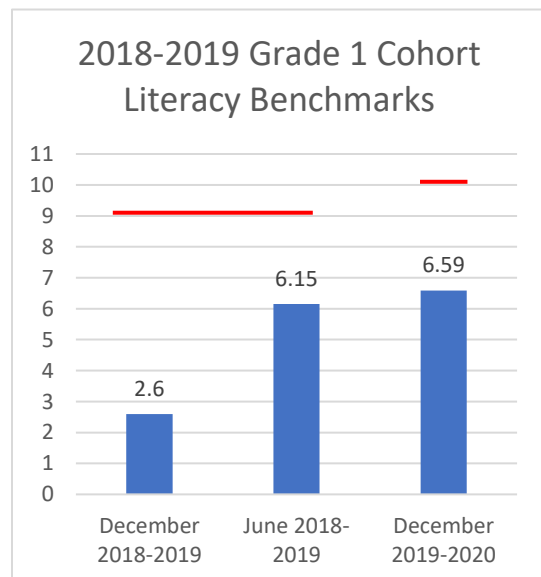
an Pinnell reading levels. This should all be taken into consideration when considering these

results. What is clearly demonstrated is that the children are progressing in their reading ability.

When considering potential differences in literacy benchmarks across the priority schools

included in this program, Hazen White St. Francis and St. John the Baptist are the two schools

that stand out as underperforming the others. They continuously fall below the grand mean

(overall average) literacy benchmarks. The other schools appear to be more or less similar in

their overall literacy proficiency. Again, we see examples of many schools just meeting the

minimum standards for the end of grade 1.
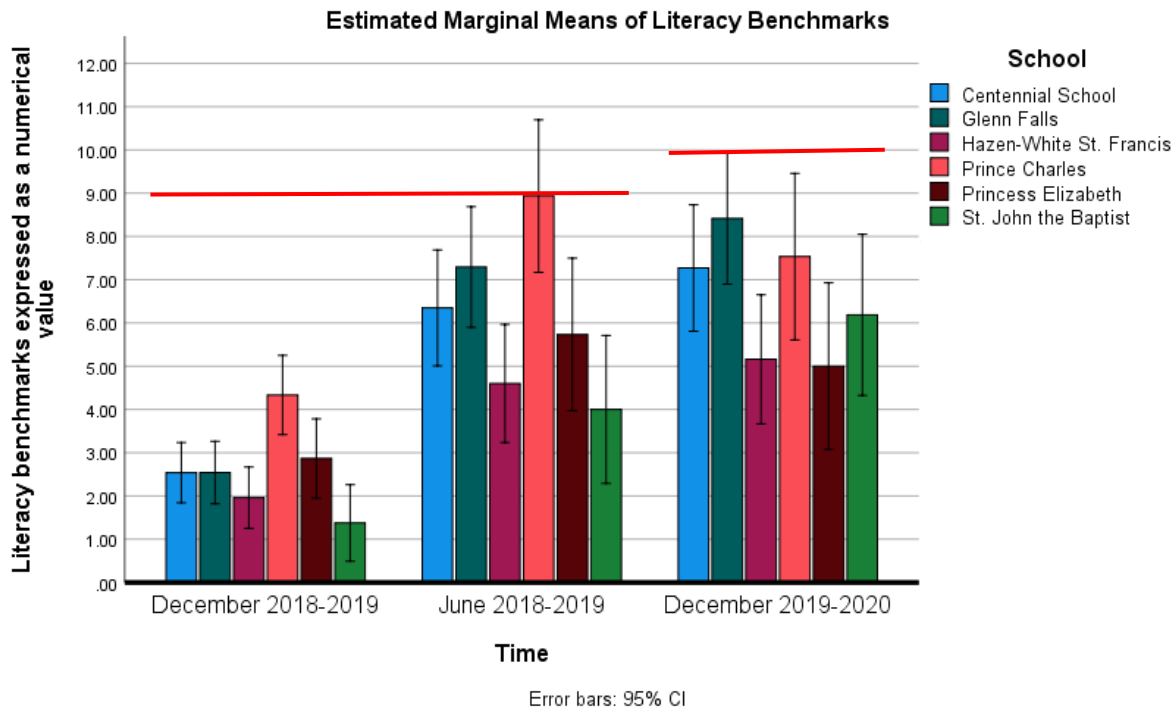
Numeracy benchmarks. Again, numeracy benchmarks represent the proportion of math

questions answered correctly during each of the assessment periods. Here we can see that

students went from answering only 37.78% correctly to 66.04%. Although there is still room for

improvement, this represents a gain of 28.26 percentage points by the end of the year. As

students are expected to answer all questions correctly, this data might suggest that there is still a

substantial amount of work to do, but we must also

consider whether or not students from these priority

schools should be expected to perform at those

particular levels of proficiency. The closer the

student's are to that 100% correct standard, the more

likely it is that the numeracy gap is closing. We must

also remember that this cohort of students was not

exposed to this program in their kindergarten year,

2018-2019 Grade 1 Cohort Numeracy Benchmarks

which may have had the possibility of bolstering their scores even further.
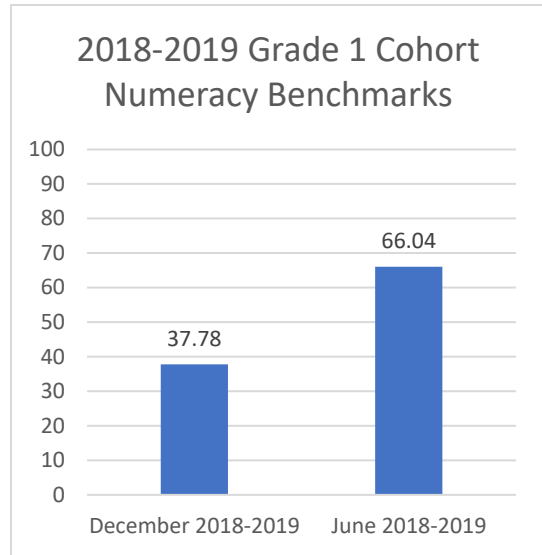
When examining differences between numeracy benchmarks between priority schools,

we can again see that Hazen White St. Francis is underperforming when compared to at least

some of the schools. Overall, however, most of the schools appear to be fairly similar in their

average numeracy benchmark scores.

Estimated Marginal Means of Numeracy_Benchmarks

Error bars: 95% CI
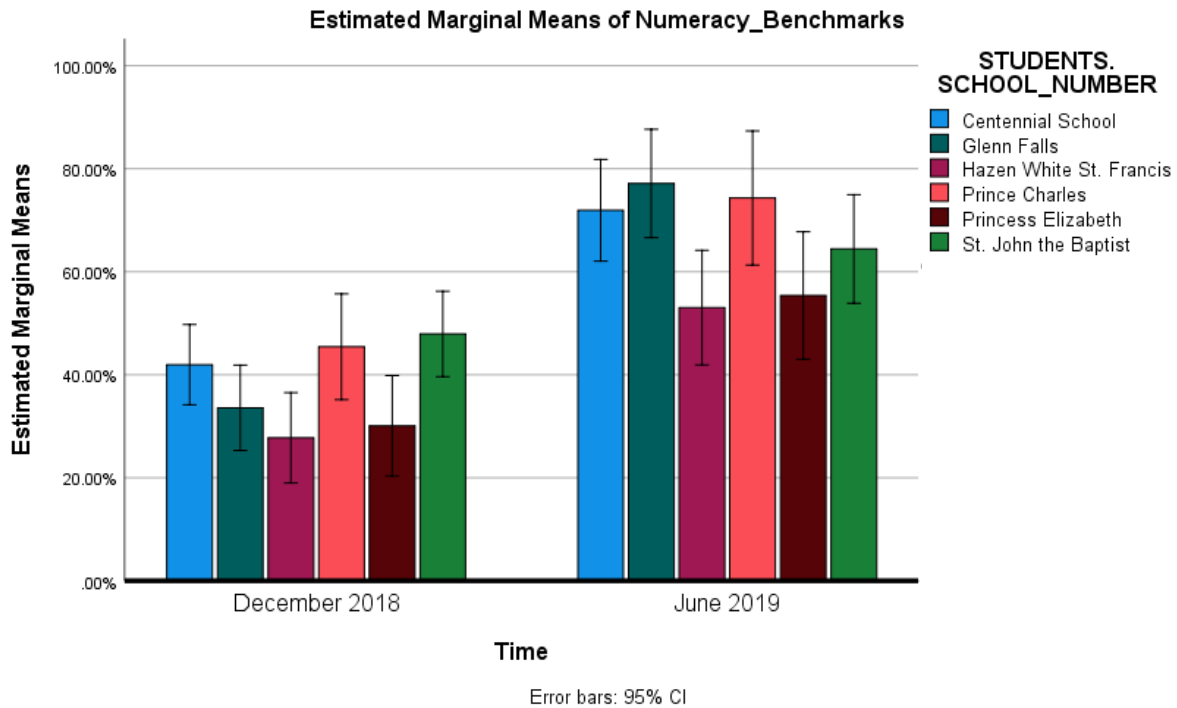
The importance of attendance in Grade 1. Unlike kindergarten, we were unable to find a relationship between attendance and the benchmark data. This could be because the attendance data was not found to be normally distributed, and therefore not ideal for correlational analyses. Furthermore, the final sample sizes in these analyses were quite small.

In attempting to understand why no relationships were found in the grade 1 cohort data, we could consider the pattern of results from the kindergarten cohort. We saw that there was a large drop in the strength of the relationship between attendance and benchmarks when those children graduated to grade 1. It is entirely possible that the lack of correlational findings for the 2018-2019 grade 1 cohort simply reflects this drop in attendance's overall contribution to learning. It is entirely possible that other factors become far more important as the children progress through the grades. I however, strongly suspect that the lack of findings has more to do with the shape of the distribution of this dataset. It seems far more likely that attendance would at least contribute a little to student success. Further research would be needed to determine why there was no relationship found between attendance and benchmarks in the grade 1 cohort data. It would also be worthwhile to examine other contributing factors to student success. This was not possible here, as little additional data concerning such factors was collected or available for analysis.

Overall Summary

Although the achievement gap was not shown to close in every case examined here, there is some evidence that the achievement gap between priority and non-priority schools in Saint John, New Brunswick was closing when using the metric of provincial standard grades. Kindergarten, specifically, demonstrated many examples of a closing (or closed) achievement gap. Grade 1, however, did only demonstrated one instance of a closed achievement gap. As the intervention

(addition of a second teaching for each class) was not standardized across all classrooms, it is difficult to exclusively link the improvement in performance to the intervention.

In terms of the student's literacy benchmarks, students were meeting the minimum required reading level by the end of the first year. To be ready for the next year, students should really be above this minimum level. For instance, if students are at a "C" level at the end of kindergarten, they need to advance two more levels to meet the minimum level in grade 1. This suggests that literacy levels may still be lower than they should be, and more focus might need to be place on reading. Numeracy benchmarks showed considerable increases at the second assessment for both kindergarten and grade 1 students, suggesting the children had retained some of what they had learned. Grade 1 students did show a considerable increase in numeracy benchmarks, but they also started with considerably lower scores than the kindergarten cohort. The grade one children showed gains, but still have room for improvement in numeracy proficiency.

Attendance was shown to be quite important during kindergarten, accounting for between 25% and 43% of the variation in numeracy and literacy benchmark scores. Although the strength of this relationship dropped when they graduated to their grade 1 year, attendance still accounted for between 5% and 8% of their literacy and numeracy scores. It appears that the fewer days the children missed, the better they tended to do on the benchmark assessments. Of particular concern for this kindergarten cohort is the range of absenteeism. Our 2018-2019 kindergarten cohort absenteeism ranged from 0 to 24 days during the 2018-2019 academic year and 0 to 49 days during the 2019-2020 academic year. Thirty percent ($n = 30$) of these children missed more than 10 days in their kindergarten year and approximately 23% ($n = 171$) missed more than 10 days in their grade 1 year.

Grade 1 showed no evidence of a relationship between absenteeism and benchmark scores. However, the pattern of absenteeism was very similar to the kindergarten group. The range of absenteeism during their grade one year ranged from 0 to 38 days and the range of absenteeism in their grade 2 year ranged from 0 to 47 days. During their Grade 1 year, 20% ($n = 30$) missed more than 10 days. During their grade 2 year, 21% ($n = 194$) missed more than 10 days.

It is important to note that the second year took place during the 2020 pandemic, which might account for an increase in the range of absenteeism seen on both cohorts during their second year (2019-2020) if parents felt it necessary to pull their children out of school early. Whatever the reason, absenteeism appears to be a problem within these priority school populations.

Evaluation limitations and future directions. There are a few limitations to the findings presented here. First, we must recognize that because the additional teaching staff was not implemented in a systematic way across schools and classrooms, we cannot draw strong conclusions about the effect of the additional staff. Some students attended a smaller classroom with one teacher, and others attended a larger classroom with two teachers. This likely affected the teaching methodologies and styles used among the different classrooms. Furthermore, though we were interested in assessing the achievement gap as it pertains to literacy and numeracy, the data available to assess these areas of learning may not be the best measures of numeracy and literacy proficiency. Provincial standard grades for math and language arts served as a measure of numeracy and literacy, respectively. The manner in which teachers assign students to one of the standard grade levels could vary among schools and teachers, which means that it lacks standardization. This limitation applies to both math and language arts. The provincial standard grades are also categorical in nature, rather than being a true continuous measure of proficiency.

This means that they are less well suited for the models of inferential statistics used here to assess the achievement gap.

The benchmark assessments also raise questions concerning validity. Although Fountas and Pinnell Reading Levels are widely used, there are questions concerning the validity and reliability of the assessment. Psychologists have questioned its use as a "gold standard" as late as 2018, suggesting that the assessment needs to be tested in more research published in peer-reviewed journals. Admittedly, I've less information available concerning what was used as a numeracy benchmark beyond knowing that the data represented the number of items answered correctly during each assessment. This makes it difficult for me to support or caution against the assessment's reliability and validity.

These are the primary limitations that need to be considered when assessing the program's success. We do need to accept the fact that we used the best information that was available to us. The limitations do, however, highlight potential opportunities going forward. First, there is little commentary regarding the teacher's perspective. It would be worthwhile to pursue a well-developed and thorough qualitative interview in which as many of the participating teachers as possible could provide their own feedback about the program. This would allow us to identify themes that correspond to things the teachers liked and things that they considered challenges. Such information could lead to better developed intervention plans aimed at closing the education gap.

Should we decide to try this intervention again in the future, it would be necessary to develop a standardized protocol for implementing the additional teachers. This would mean that all schools receive the same intervention so that the program can be evaluated with more confidence. In addition to such a protocol, we would also need to identify standardized measures

of numeracy and literacy that have been evaluated in terms of their reliability and validity. These

assessments could be done a numerous point during the year and would be implemented in all

priority schools and a random sample of non-priority schools that serve as a comparison group.