Through the lens of subgroups: Learning loss during the COVID-19 Pandemic

Introduction

The 2019-2020 school year will go down in history as the year that formal in-person (face-to-face) learning was interrupted due to the COVID-19 pandemic. When schools closed in the winter of 2020 due to the pandemic, many schools did not immediately go into online mode. Even for the few that did, to a great extent it was not done with the expertise that most schools have today. The fact that some students continued with online learning and others did not have any form of instruction during the early days of the pandemic created inequalities in access to learning. By the end of the 2019-2020 school year, many government agencies and states put mechanisms in place to ensure that online learning was a means to bridge these educational inequalities. However, issues of inequalities to learning were not fully resolved as access to online learning was also a big issue due to lack of computers and/or a Wi-Fi/network to access online sessions. In addition to these problems, student attendance for online sessions was also an issue as most students needed monitoring by an adult for them to connect to online sessions. While the main goal of this document is to share the impact of the COVID-19 pandemic on student learning, we would be remiss if we shared the assessment results without bringing into context the circumstances experienced by students from the beginning of the pandemic up to this point.

By the beginning of the 2020-2021 school year, measures to mitigate some of the connectivity issues to online sessions faced earlier in the pandemic were done by providing computers and in other cases, by having Wi-Fi hubs stationed in areas where students would come and connect to internet. While some schools had online learning and others had face-to-face (in-school) learning, a majority of schools offered a hybrid model offering instruction delivery that included both in-school and online learning. In addition, the state became stricter with issues regarding online participation and attendance and provided some form of guidance to define attendance and participation requirements and reporting – a bid to ensure that all students are involved in some form of learning.

In addition to issues of inequalities in access to learning among students, educators had to reinvent the wheel and get accustomed to new delivery modes of instruction that, and in many instances, required the use of methods that they were far less familiar with. Students had to adjust to these changes too. The changes brought many challenges, including one key area which was student assessment.

By the beginning of the school year, educators wrestled with assessing how much ground students had lost, or gained, in order to tailor instruction to the students' current achievement level and start the journey toward recovery. Stated differently and in more precise terms, educators were keen to get answers to the following questions:

- What is the current achievement status for the students?
- How much learning has been lost for most (or gained, if any)?
- How do we get the students to the level where they were expected to be prior to the pandemic?

There are no easy answers to these questions since schools are still learning new strategies to counter the effects of the pandemic. It is still too early to tell whether new learning strategies are working as students are still facing both physical and emotional effects of the on-going pandemic. Nevertheless, assessing the impacts of the pandemic on student learning has become central to determining how we proceed from this point. Specifically, assessing students' performance during and before the pandemic is key to making informed decisions on how to help students make those necessary adjustments.

At this point, it is important to bring into context one key change in assessment practice that makes it difficult to interpret assessment results obtained during the pandemic – remote testing. Because most students were enrolled in online learning, they were assessed using remote testing during the pandemic whereas before the pandemic, students were assessed face-to-face in school settings with proctors designed to ensure that test administration occurred within the confines of standardized testing procedures and abiding by principles and best practices for testing. Issues with remote testing are well-documented and include - parents/siblings helping students take tests, lack of supervision while test taking, and student disengagement during test sessions. The underlying theme among serious issues with remote testing is the lack of uniform standardized settings to ensure that test administration meet the requirements in test security and that students are engaged at the same level as they could have been if tests were administered in school, with teacher proctors involved. For most schools, it was the first time to have their students involved in remote testing and proctoring by adults in their own home setting.

With the security, physical and emotional effects, and engagement and many other issues that students continue to face during the pandemic, The Center for Charter Schools at Central Michigan University has strongly recommended that assessment data collected during the pandemic should not be used for accountability purposes. Instead, the results can inform educators on how to prioritize resources to make meaningful gains for the students that will put them right back on track. Indeed, the data may be messy. However, in its messiness, policymakers can still flag students that need help and address other areas of concern to ensure that the gaps of learning endured during this period are minimal.

Preliminary fall reports by The Center for Charter Schools' Performance Data Center (PDC), NWEA (2020) and Renaissance (2020), have highlighted that student performance during the pandemic is consistent with prior years in reading but not in mathematics where there was a noticeable decline. Regardless the subject, there remains a need to extract and refine information, as well as provide razor-sharp focus on different subgroups to ensure that no groups of students are left behind due to the pandemic. For example, the percent of students meeting or surpassing the norms only looks at the average performance or the movement that is happening around grade level norms. It is possible that there is movement both below and above the 50th percentile (i.e. students moving from the 20th percentile of achievement to 45th percentile, or from 48th percentile to 17th percentile, movements from 56th to 85th achievement percentiles or from 90th to 61st achievement percentiles). The latter analyses could provide more information above and beyond what the former analyses provide. This report therefore seeks to reaffirm what has already been established but does so by using a different methodology. In addition, this analysis seeks to further explore whether there are any gaps in performance between different grade levels, performance groups, and demographic groups such as different socio-economic status (free or reduced lunch versus non-reduced lunch) and ethnicity. For continuity of what has already been established, the following questions are addressed:

1. How do student performances compare between pre-COVID-19 years (2017-2018 and 2018-2019) and during COVID-19 (2019-2020)?
2. Does student performance change due to the following factors:
    a. Socio-economic status (Free and Reduced lunch versus Non-reduced lunch)
    b. Performance level

To this end, we hope that from the ensuing information 'extracted' from the data, the results will be instrumental in driving policy and help educators on how to use any available resources from the already depleted resources that many school districts have. This will help school districts to ensure that all groups of students get the lift they need to get them across the line in their quest for academic excellence as schools move out of the pandemic. After all, when all is said and done, it is imperative that the educational needs of all students are met in order to put them back on track to the rigor needed to ensure that they are ready for college, work and life.

The Data

The data used was from 47 schools in Michigan for students in grades 3-8 with test scores in both fall

2019 and fall 2020 sessions. The table below shows the number of students in mathematics and reading that were used from 2017-2018, 2018-2019, and 2019-2020:

Year	Subject	Number of students
2019-2020	Mathematics	8,933
2019-2020	Reading	8,981
2018-2019	Mathematics	8,701
2018-2019	Reading	8,685
2017-2018	Mathematics	8,443
2017-2018	Reading	8,404

Methods for Analyzing Data

We use NWEA's achievement percentile descriptors (Thum & Kuhfeld, 2020) to investigate students who 'move' from one category to another (slider or gainer) or 'stay' in the same category (maintainer). This categorization is adopted from NWEA collaboration discussions (NWEA 2020)

The following are the achievement categories:

Achievement Percentile Range	Category Descriptor
81-99	High
61-80	High Average/Above Average
41-60	Average
21-40	Below Average/Low Average
1-20	Low

Who are sliders, maintainers, and gainers?

Sliders are students who have moved from a higher achievement percentile category to a lower achievement percentile category between testing sessions (e.g. moving from average to below average category or from high to average category from fall 2019 testing session to fall 2020 testing session). Sliders can 'slide' through multiple categories.

Gainers are students who have moved from a lower achievement percentile category to a higher achievement percentile category between testing sessions (e.g. moving from below average to average category or from average to high category from fall 2019 testing session to fall 2020 testing session). Gainers can also move through multiple categories.

Maintainers 'stay' in the same category from one testing session to another testing session. The following table illustrates the concept of sliders, maintainers, and gainers:

Student ID	Fall Test Percentile 2019	Fall Test Percentile 2020	Movement	Movement Category
A-200	90	61	From High to Above Average	Slider
B-210	67	24	From Above Average to Below average	Slider
C-320	17	9	From Low to Low	Maintainer
D-410	82	99	From High to High	Maintainer
E-525	63	90	From Above Average to High	Gainer
F-110	22	65	Below average to Above Average	Gainer

NWEA (2020) and Renaissance (2020) have indicated that there are declines in performance between fall 2019 and fall 2020 in mathematics but reading is consistent with prior performance. This exploratory analysis seeks to confirm that observation. In addition, the analysis seeks to investigate whether different demographic groups were equally affected by the pandemic whenever there are any effects.

More importantly, whenever categorical data is used to draw conclusions about differences in occurrences of outcomes of interest, in our case the rate of sliding, maintaining and gaining, there is a temptation to just have a visual inspection on the outcome and draw speculative conclusions on whether there are notable differences or not. What if the frequencies are so close? For example, is a 3% difference between sliders in 2018-2019 and 2019-2020 something we should be concerned about? How about 5%? At what point should we be concerned?

To obtain answers to these questions, it is important to run a Pearson Chi Square test. Chi-Square tests on categorical data show whether there is any independence/dependence among variables. For example, in our case, if we treat the rates of sliding, maintaining, and gaining as one variable and the school years of 2017-2018, 2018-2019, and 2019-2020 as the other variable, we will be able to create contingency tables where we can calculate expected values and compare them to observed values. That way, we can draw conclusions to see if the rate of sliding, maintaining, and gaining is dependent on school year. Simply put, independence of variables in this case will mean that the patterns of sliding, maintain, and gaining have nothing to do with the years – it does not matter what year you focus on, the rates of sliding, maintaining, and gaining will be similar. On the other hand, dependence will show that the patterns differ from one year to another. The value of the Chi-Square test and its associated p-value together with degrees of freedom (df) is all we will need to make these determinations. We will use the more stringent p-value of 0.01 as the threshold to determine independence of variables. Any values in p-value lower than 0.01 will indicate that the two variables are not independent while values equal to or above 0.01 will indicate that the variables are independent.

In summary, this analysis explores the differences in performance among different groups of interest and further seeks to see if there are statistical differences in performance. Specifically, the analysis seeks to address the following questions:

1. Are students sliding, maintaining, or gaining differently over the years; or is the rate of sliding, maintaining, or gaining independent of school year?
2. Are students sliding, maintaining, or gaining differently across different socio–economic backgrounds; or is the rate of sliding, maintaining, or gaining independent of socio-economic background?
3. Do low achieving students slide, maintain, or gain at the same rate as high achieving students; or is the rate of sliding, maintaining, or gaining independent of achievement level?

Results

Are students sliding, maintaining, or gaining differently over the years; or is the rate of sliding, maintaining, or gaining independent of school year?

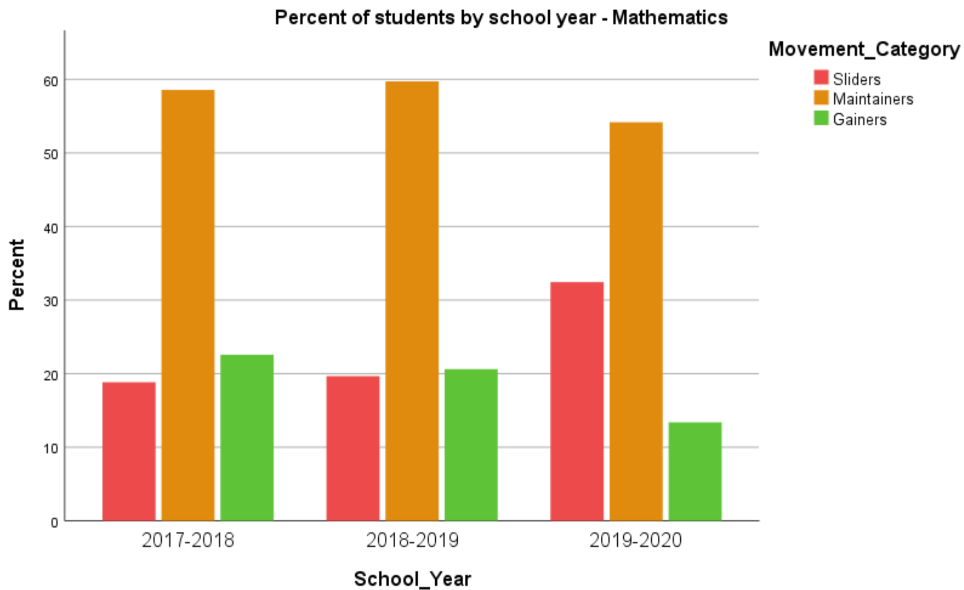

Figure 1. Percent of sliders, maintainers, and gainers by school year in mathematics

From figure 1 above, while the percentages of sliders, maintainers, and gainers are almost similar for 2017-2018 and 2018-2019, clearly, the percentages of sliders, maintainers, and gainers in 2019-2020 does not follow the same pattern as that of prior two years. We see more sliders, fewer maintainers, and fewer gainers in 2019-2020 than in prior two years. Is this change significant?

As is shown in the table below, the Chi-Square statistic associated with the observed differences between the patterns observed in 2019-2020 to those of 2017-2018 and 2018-2019 is 675.495 and 4 degrees of freedom. This value is statistically significant ($p < 0.01$) and highlights that the rate of sliding, maintaining, or gaining is not independent of school year – student performance in mathematics for 2019-2020 differs from prior years.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	675.495	4	.000
N of Valid Cases	26077		

Are there similar patterns in reading?

Although not as clear-cut as was the case with mathematics, there are small differences in the performance of students (especially sliders and gainers) in 2019-2020 with prior years, especially 2017-2018. For example, 25.9% of the students were sliders in 2019-2020 while 22.8% and 20.0% were sliders in 2018-2019 and 2017-2018 respectively. The bar chart below in figure 2 shows this information. Is this a significant difference?

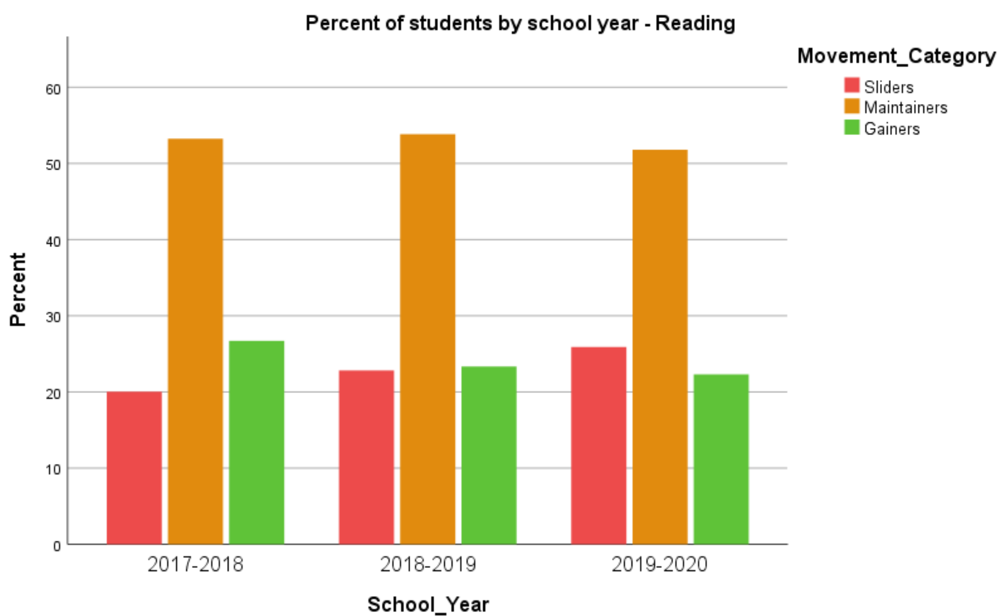

Figure 2. Percent of sliders, maintainers, and gainers by school year in reading

The table below shows a Chi-Square statistic of 106.810 and 4 degrees of freedom, which has a p-value less than 0.01. This shows that although the difference is not as pronounced as in mathematics, there is still a significant difference in performance of students across the three years – students performed lower in 2019-2020 than 2017-2018 and/or 2018-2019.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	106.810	4	.000

N of Valid Cases	26070		

Are students sliding, maintaining, or gaining differently across different socio–economic backgrounds; or is the rate of sliding, maintaining, or gaining independent of socio-economic background?

As has already been established, there are significant differences in performance for sliders, maintainers, and gainers for 2019-2020 data compared to 2018-2019 and 2017-2018 data. In 2019-2020 there were more sliders and fewer gainers and maintainers than prior years. Are Free and Reduced lunch students sliding, maintaining and gaining more than students who are not in Free and Reduced lunch programs? Using 2019-2020 data displayed in figure 9, there are about 5% more sliders for Free and Reduced lunch students than non-reduced lunch students (34.2% versus 29.5%). In addition, there are 5% fewer maintainers for Free and Reduced lunch students than non-reduced lunch students (52.2% versus 57.5%) while the percentages for gainers are about the same. Are the percentages of sliders and maintainers significantly different?

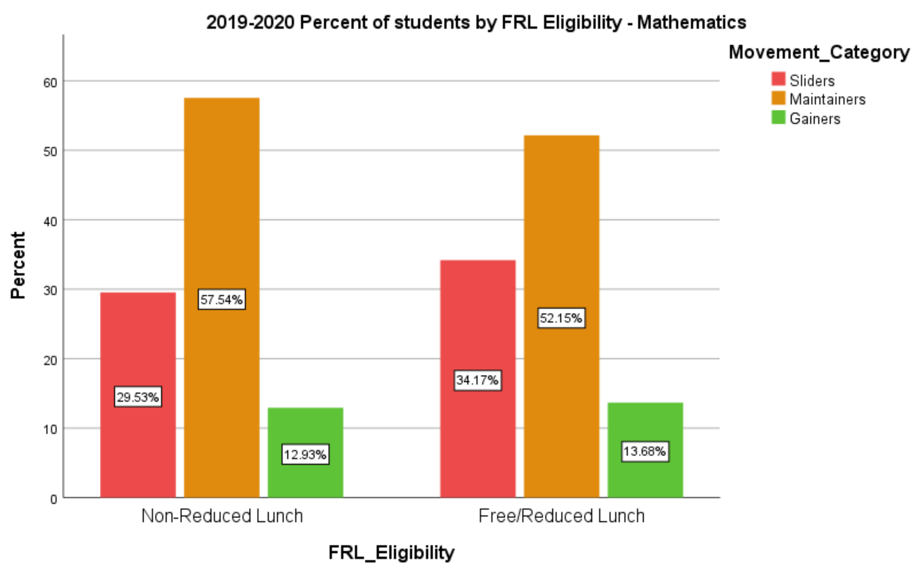

Figure 9. Percent of sliders, maintainers, and gainers by FRL eligibility in mathematics

As the table below shows, there is a significant difference between sliders and maintainers between non-reduced and Free and Reduced lunch students. With a Chi-Square statistic of 25.694, 2 degrees of freedom, and a p-value less than 0.01, it is very clear that Free and Reduced lunch students did not perform as well as non-reduced lunch students in mathematics in 2019-2020. Could it be possible that the difference we see is similar to that observed in prior years?

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	25.694	2	.000
N of Valid Cases	8866		

Further analysis on prior years mathematics data (2017-2018 and 2018-2019) indicates that the percentages of sliders, maintainers, and gainers are about the same between non-reduced lunch students and Free and Reduced lunch students, especially in 2017-2018 (see figures 10 and 11 below). Precisely, for 2017-2018 mathematics data, with a Chi-Square statistic of .813, 2 degrees of freedom, and p-value of 0.66 (which is greater than the threshold of 0.01), it is very clear that there is no significant difference between different socio-economic backgrounds – Free and Reduced lunch students performed similar to non-reduced lunch students. However, 2018-2019 data shows some differences between Free and Reduced lunch and non-reduced lunch students (Chi-Square statistic = 10.949, 2 degrees of freedom, and a p-value less than 0.01), which is still lower than that of 2019-2020. Therefore, there is a concerning growing gap in mathematics performance between Free and Reduced lunch and non-reduced lunch students that we see in 2019-2020 data, to the extent that policymakers should not ignore the difference.

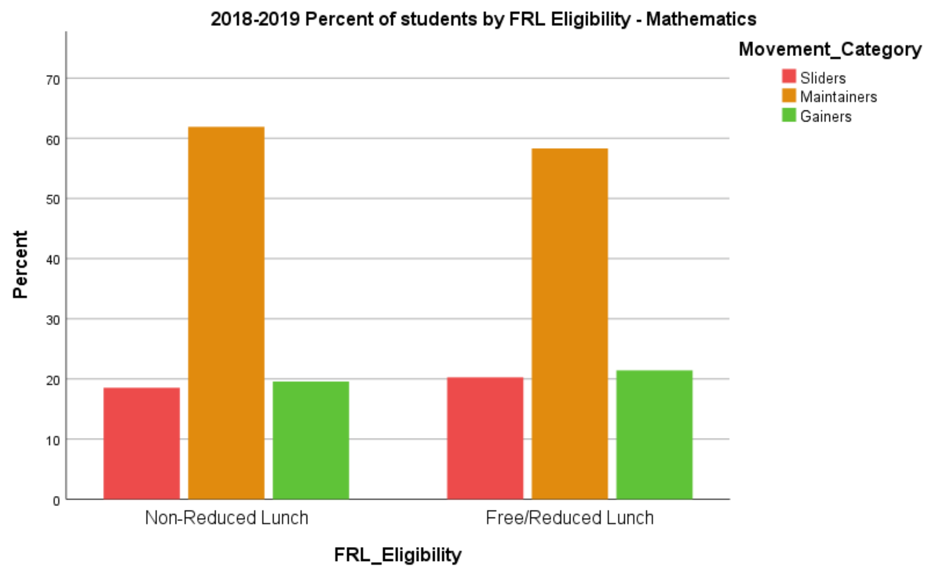

Figure 10. Percent of sliders, maintainers, and gainers by FRL eligibility in mathematics

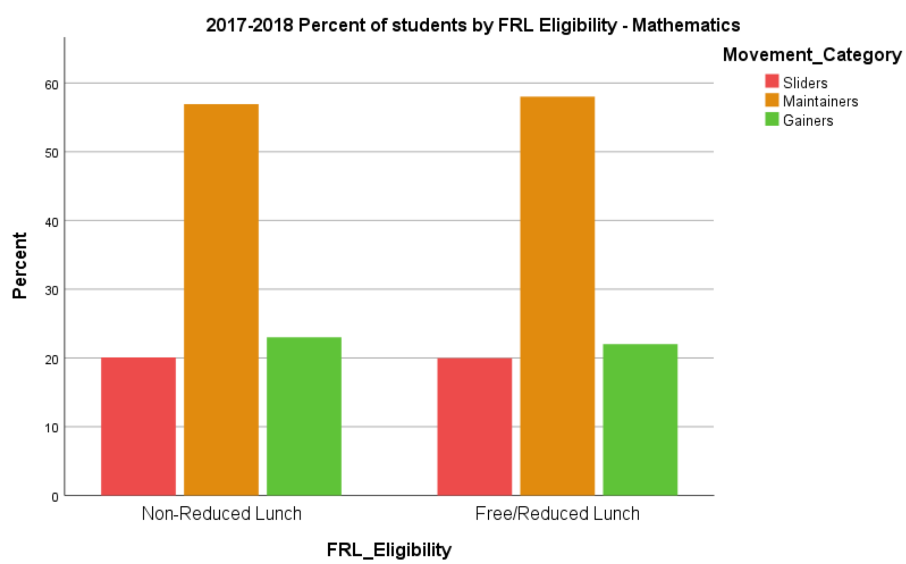

Figure 11. Percent of sliders, maintainers, and gainers by FRL eligibility in mathematics

A similar pattern as seen in mathematics data is observed in reading. As is seen in figure 12 below, using 2019-2020 data, the percentage of sliders is about 7% more for Free and Reduced lunch students than it is for non-reduced lunch students (28.4% versus 21.7%). In addition, there is about 8% fewer maintainers for Free and Reduced lunch students than non-reduced lunch students (48.9% versus 56.6%) while the percentages for gainers differ by only one percentage point. Are these percentages significantly different?

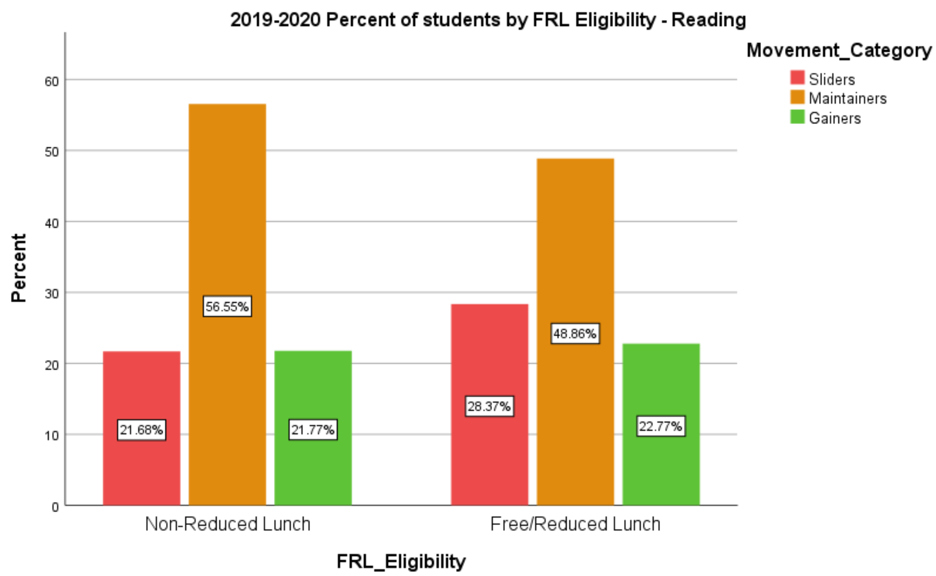

Figure 12. Percent of sliders, maintainers, and gainers by FRL eligibility in reading

Just like with mathematics data, these differences are statistically significant. A Chi-Square value of 60.561, 2 degrees of freedom, and a p-value less than 0.01 highlights that Free and Reduced lunch students are not performing as well as non-reduced lunch students in reading for 2019-2020 data (see table below). Is this finding consistent with 2018- 2019 and 2017-2018 data?

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	60.561	2	.000
N of Valid Cases	8914		

In addition, prior years reading data (2017-2018 and 2018-2019) indicate that the percentages of sliders, maintainers, and gainers are about the same between non-reduced lunch students and Free and Reduced lunch students (see figures 13 and 14 below). A Chi-Square statistic of 2.931, 2 degrees of freedom, and a p-value of 0.231 (greater than the threshold of 0.01) for 2017 -2018 reading data underscores the fact that the rate of sliding, maintaining, and gaining was not dependent on Free and Reduced lunch eligibility, i.e. the differences we see in reading performance between Free and Reduced lunch students and non-reduced lunch students for 2019-2020 reading data were not there in 2017-2018. Although 2018-2019 reading data shows some differences in performance between Free and Reduced lunch and non-reduced lunch students (Chi-Square statistic = 12.916, 2 degrees of freedom, and p-value = 0.002 (less than 0.01), it is still lower than that of 2019-2020. Therefore, the growing gap in performance that we find between Free and Reduced lunch and non-reduced lunch students for 2019-2020 reading data also call for our immediate attention.

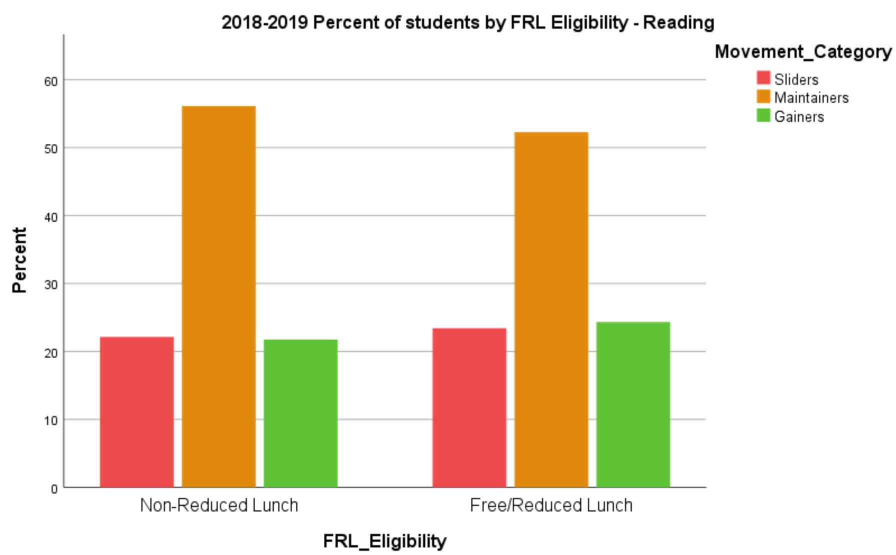

Figure 13. Percent of sliders, maintainers, and gainers by FRL eligibility in reading

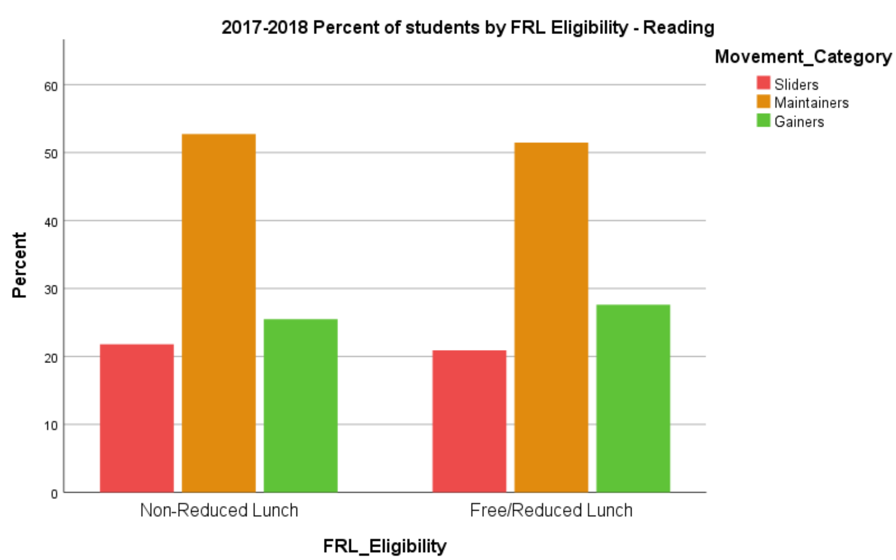

Figure 14. Percent of sliders, maintainers, and gainers by FRL eligibility in reading

Do low achieving students slide, maintain, or gain at the same rate as high achieving students; or is the rate of sliding, maintaining, or gaining independent of achievement level?

In analyzing the data using achievement levels, achievement percentiles of 50 or greater were classified as high achievement percentiles whereas all achievement percentiles lower than 50 were classified as low achieving. For 2019-2020 mathematics performance displayed in figure 15 below, the percentage of sliders is about 11% more for low achieving students than it is for high achieving students (36.8% versus 25.7%). In addition, there is just about 3% fewer maintainers for low achieving students than there are for high achieving students (53.1% versus 55.8%). The percentages for gainers are also 8% lower for low achieving students than it is for high achieving students (10.1% versus 18.5%). Are these percentages significantly different?

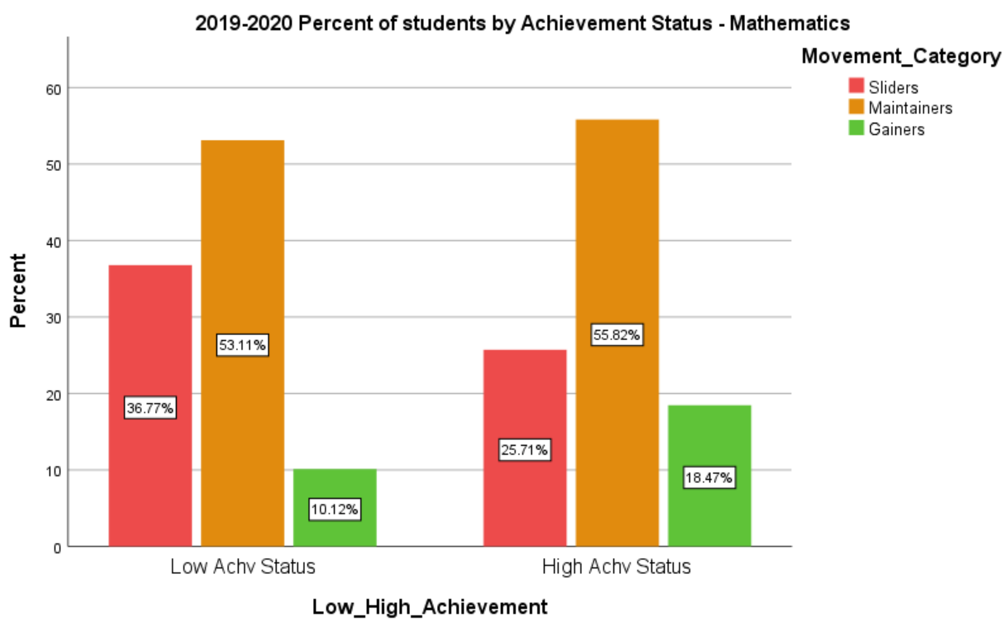

Figure 15. Percent of sliders, maintainers, and gainers by achievement level in mathematics

In the table shown below, the rates of sliding, maintaining, and gaining for high achieving students are significantly different from those of low achieving students (Chi-Square statistic = 194.165, 2 degrees of freedom, and a p-value less than 0.01 shown in the table below). High performing students are sliding, maintaining, and gaining better than low achieving students. An in-depth analysis of whether the pattern is consistent with 2018-2019 and 2017-2018 data reveals that the significant differences we see in 2019-2020 data are not new. High achieving students and lower achieving students tend to have different sliding, maintaining, and gaining tendencies even using prior data. What is more striking is that across the three years of data analyzed for mathematics, the Chi-Square statistics are about the same for 2017-2018 and 2018-2019 (192.165 and a p-value less than 0.01 in 2018-2019, and 192.165 and a p-value less than 0.01 in 2017-2018) and only 2 points higher in 2019-2020 (194.165 and a p-value less than 0.01) and 2 degrees of freedom in all cases. This brings home the point that, although we see higher percentages of sliders between 2018-2019 and 2019-2020, those declines in performance occurred in similar proportions across years and across sliders. For example, we already pointed out that the percentage of sliders is about 11% more for low achieving students than it is for high achieving students (36.8% versus 25.7%). The differences are consistent with data from prior years – 10 % (23.6% versus 13.9%) for 2017-2018 data and 10% (24.3% versus 14.5%) for 2018-2019 (see figures 15, 16 and 17). Indeed, for 2019-2020 mathematics data, both low achieving and high achieving students are not performing as well as in prior years. There is clearly an increase in sliders, a decrease in maintainers and a decrease in gainers from prior years when we focus on 2019-2020 data. Notably, the increase in sliders, decrease in maintainers and decrease in gainers from prior year to 2019-2020 mathematics data affected both low achieving students and in equal amounts. The ratio of change between the two groups across all three years is fairly constant – 11%: 10%: 10%. Therefore, the data does not support the notion that one group was affected more than the other across years. Simply stated, there is a gap between low achieving and high achieving students that we notice every year in mathematics. That gap has not widened due to the pandemic.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	194.165	2	.000
N of Valid Cases	8933		

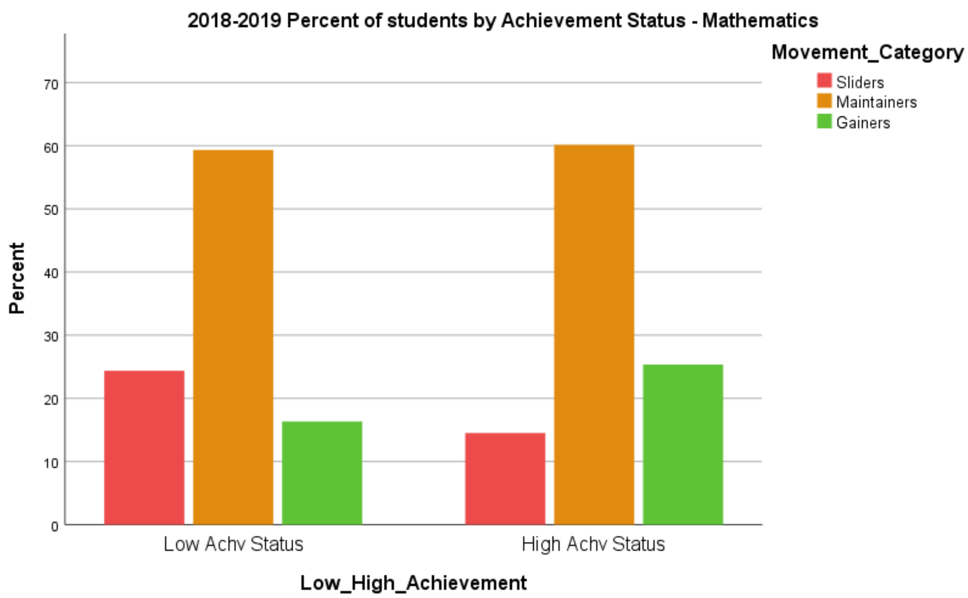

Figure 16. Percent of sliders, maintainers, and gainers by achievement level in mathematics

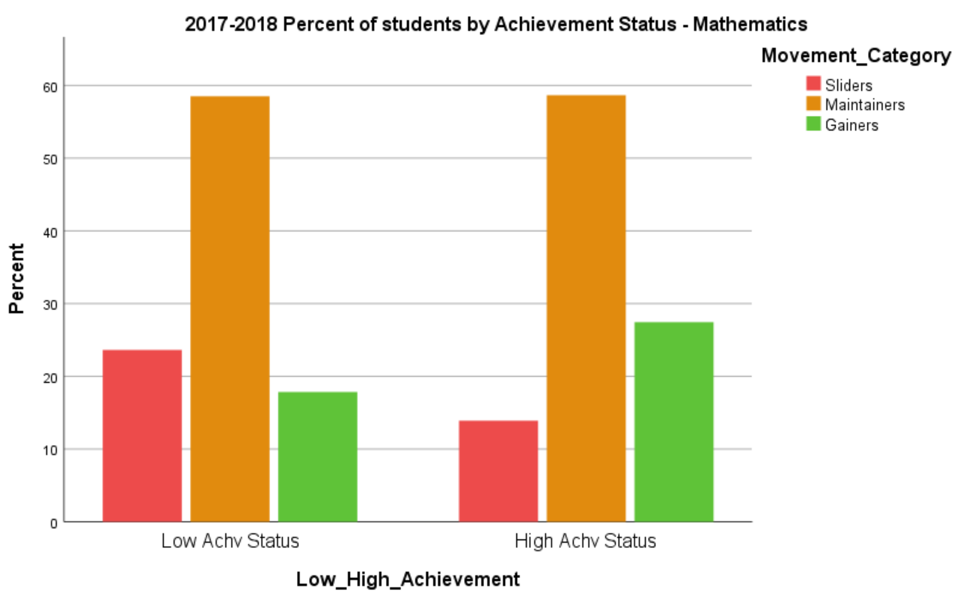

Figure 17. Percent of sliders, maintainers, and gainers by achievement level in mathematics

As figures 18, 19, and20 show, in reading, there is clearly an increase in sliders, a decrease in maintainers, and a decrease in gainers from prior years to 2019-2020 data. Notably, the changes are not as pronounced as was the case with mathematics. The high Chi-Square statistic of 397.614, 2 degrees of freedom and a p-value less than 0.01 that is shown in the table below confirms that students slide, maintain, and gain differently across achievement levels for 2019-2020 reading data.

This is consistent with what we observed with mathematics data. As was the case with mathematics data, further analysis on reading data from prior years highlights that this is not new. In 2018-2019, low achieving students performed significantly worse than high achieving students (Chi-Square statistic = 201.875, 2 degrees of freedom and p-value less than 0.01). Similarly, in 2017-2018, low achieving students performed significantly worse than high achieving students (Chi-Square statistic = 172.695, 2 degrees of freedom and p-value less than 0.01). The high value of Chi-Square statistic in 2019-2020 reading, which is very different from the other two values from data from prior years, is a sign of trouble.

Upon closer inspection, an area of concern that needs to be addressed in reading is that there is evidence to suggest that low achievement students are more affected using 2019-2020 reading data than we see using prior data. For example, from figures 18-20, we notice a 17% difference between low achieving students and high achieving students in sliders for 2019-2020 reading data (34.5% versus 17.7%). In 2018-2019 and 2017-2018, the differences in percentages of sliders were about 11% (28.7% versus 17.6%) in 2018-2019 and about 11% (25.9% versus 15.2%) in 2017-2018. In other words, the gap between low achieving and high achieving students widened up in reading for 2019-2020. This calls for immediate attention from policymakers.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	397.614	2	.000
N of Valid Cases	8981		

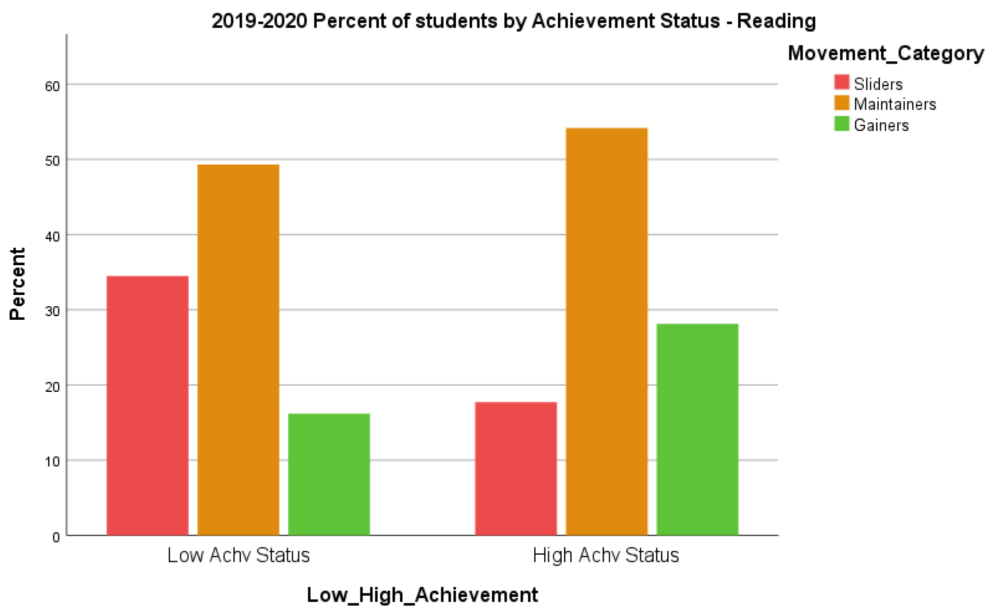

Figure 18. Percent of sliders, maintainers, and gainers by achievement level in reading

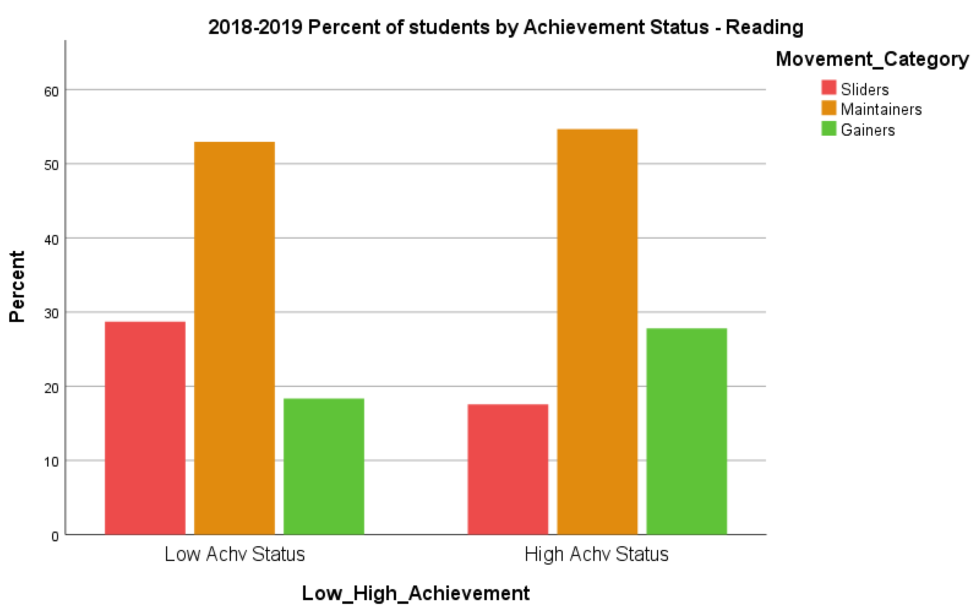

Figure 19. Percent of sliders, maintainers, and gainers by achievement level in reading

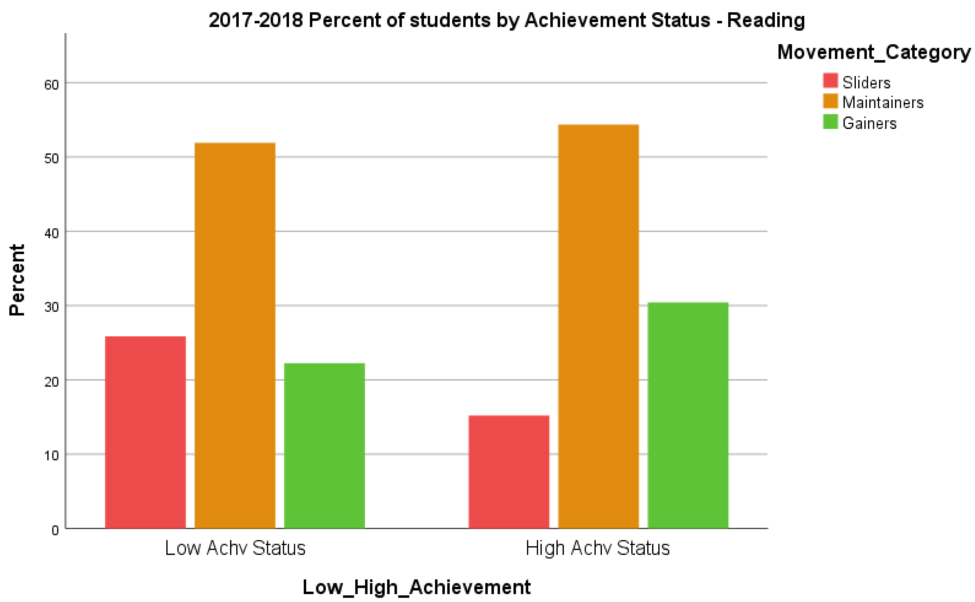

Figure 20. Percent of sliders, maintainers, and gainers by achievement level in reading

Conclusions from the Analysis

In summary, these results highlight the issue of leaving out subgroups when analyzing data using all students lens only. Because the analysis tuned into the performance of different subgroups of interest and examined the performance of students on multiple categories (e.g. sliders, maintainers, gainers across five performance levels) rather than forcing the data on a dichotomy, meet or did not meet the norms, more information was obtained. As such, there were declines that were noted for both mathematics and reading data that couldn't be revealed through all students lens. More importantly, policy makers need to be aware that when conclusions are drawn based only from the performance of all students and not subgroups, some subgroups will perform below what is reported for all students. A perfect example in our analysis is the socio-economic factor, where Free and Reduced lunch students performed differently from non-reduced lunch students during the pandemic in both mathematics and reading – a departure from their nearly consistent performance pre-pandemic. Could the growing gap in performance during the pandemic between students from different socio-economic backgrounds be attributed to availability of resources and support during this time of need?

Policymakers may want to consider what types of resources are needed for students of different backgrounds, where possible, that can help bridge the gap in performance between subgroups.

The data clearly shows that student performance during the pre-pandemic period was generally better than the performance during the pandemic for both low achieving and high achieving students. Unfortunately, there were instances where mathematics and reading results were different. A case in point is when the level of performance is taken into account, there is no evidence to conclude that low achieving students and high achieving students performed differently in mathematics during the pandemic. Conversely, there is evidence to suggest that low achieving students were more affected using 2019-2020 reading data than we see when we use data from prior years – the gap in performance between low achieving and high achieving students increased in reading during the pandemic but remained the same for mathematics.

Finally, this analysis contains many moving parts; however, solving even one of the factors identified may produce a domino effect that could put students back on a trajectory that prepares them for success in college, work, and life.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Howell, C. H. (2007). Statistical Methods for Psychology. Thomson Wadsworth.

Kuhfeld, M., & Tarasawa, B. (April, 2020). The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement.

NWEA. (December, 2020). Fall Data Results. Data Presentations from NWEA research team representatives.

Renaissance Learning (2020). *How Kids are Performing: Tracking the impact of COVID-19 on Reading and Mathematics Achievement.* Special Report Series, Fall 2020 Edition (https://www.renaissance.com/how-kids-are-performing/)

Thum, Y. M., & Kuhfeld, M. (2020). NWEA 2020 MAP Growth Achievement Status and Growth Norms for Students and Schools. NWEA Research Report. Portland, OR: NWEA