Michigan's Fall 2021 Benchmark Assessments

APRIL 2022

AUTHORS:

Tara Kilbride, Bryant Hopkins, Katharine O. Strunk, Dongming Yu

Education Policy Innovation Collaborative

COLLEGE OF EDUCATION | MICHIGAN STATE UNIVERSITY

236 Erickson Hall, 620 Farm Lane, East Lansing, MI 48824 | (517) 884-0377 | www.EPICedpolicy.org

ACKNOWLEDGEMENTS

The authors wish to acknowledge the many people who graciously gave of their time in support of this effort. We are especially grateful to our partners for their collaboration and thoughtful feedback. In particular, we would like to thank Vic Bugni, Dr. Sue Carnell, Kate Cermak, Dr. Delsa Chapman, David Judd, Andy Middlestead, and Dr. Michael Rice from the Michigan Department of Education ; Tom Howell, Mike McGroarty, and Lauren Paluta from the Center for Educational Performance and Information; Kyle Kwaiser and Nicole Wagner from the Michigan Education Data Center at the University of Michigan; and Don Dailey from the Michigan Data Hub. At Michigan State University, we thank Emily Mohr for coordinating and facilitating the project, and Meg Turner for compiling and formatting the text for the report. Finally, we thank Bridgette Redman for her excellent copy-editing.

DISCLAIMER

The Education Policy Innovation Collaborative (EPIC) at Michigan State University is an independent, non-partisan research center that operates as the strategic research partner to the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI). EPIC conducts original research using a variety of methods that include advanced statistical modeling, representative surveys, interviews, and case study approaches. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data are modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the MDE and/or CEPI. Results, information, and opinions solely represent the author(s) and are not endorsed by, nor reflect the views or positions of, grantors, MDE and CEPI, or any employee thereof.

TABLE OF CONTENTS

Executive Summary

BACKGROUND AND PURPOSE

In order to understand student learning and progress toward educational goals during the COVID-19 pandemic, the Michigan legislature mandated new data collection and reporting requirements for local school districts beginning in the 2020-21 school year (2020 PA 149, 2021 PA 48). In response, the Education Policy Innovation Collaborative (EPIC) is releasing a series of reports in collaboration with the Michigan Department of Education (MDE), the Center for Educational Performance and Information (CEPI), the Michigan Data Hub (MDH), and the Michigan Education Data Center (MEDC). This is the third in the series, all of which will be delivered to the governor and the Senate and House standing committees responsible for education legislation in the Michigan legislature to provide insight into Michigan students' progress toward learning goals during the COVID-19 pandemic.

The first report, which was released in August 2021, showed that, on average, students made less than normal progress toward learning goals, as measured and defined by math and reading benchmark assessments, during the 2020-21 school year. The second report released in December 2021 examined how student learning outcomes differed across student groups and district types, showing substantial racial and socioeconomic achievement gaps that widened over the course of the school year. In this report, we incorporate new results from benchmark assessments that districts administered in the fall of the 2021-22 school year that enable us to update some of our earlier analyses and continue to monitor student learning trajectories, growth, and achievement gaps into the next school year.

RESEARCH QUESTIONS AND STUDY OVERVIEW

With no summative testing at all in 2019-20 and lower than typical participation in end-of-year summative assessments in 2020-21, benchmark assessments are critical to help policymakers, practitioners, and stakeholders understand how the pandemic has affected, and continues to affect, student learning across Michigan. In this report, we use K-8 mathematics and reading benchmark assessment data from fall 2021, and when available, prior achievement data from the 2020-21 school year, to investigate each of the following questions:

1. **How do recent student achievement trends for Michigan students compare to national or state trends from before the pandemic?** To better understand how much the pandemic has affected Michigan students, we examine how average scores on state-mandated benchmark assessments for K-8 students have progressed over the fall 2020, spring 2021, and fall 2021 testing periods. We then compare Michigan students' trajectories to norms from before the pandemic.

2. **Did students make progress toward or reach the appropriate growth targets?** Our previous reports showed that, on average, students' scores on benchmark assessments improved over the course of the 2020-21 school year, but at a slower rate than would typically be expected before the pandemic. However, this does not mean that all students' scores improved, or that no students met or exceeded pre-pandemic growth norms. In this report, we delve into how growth outcomes varied among Michigan students. We compare students' growth between fall 2020 and fall 2021 to pre-pandemic expectations for "typical growth" for students in the same grade level with similar prior achievement scores. We then compare the proportions of students who met or exceeded their growth targets, students who made partial progress toward their growth targets, and students who did not demonstrate any growth on their benchmark assessments from fall 2020 to fall 2021.

   It is important to note that these targets do not represent the *minimum* growth that all students are expected to achieve, but rather, the *median* growth that students with similar prior scores achieved before the pandemic. In other words, in a "typical year" we would only expect about 50% of students reach these targets. This is different from a classroom-based growth goal a teacher might set for a student, which represents what a teacher expects a student to achieve in a given period.

3. **How much and in what ways did achievement trends differ across subgroups of students?** To better understand the differential effects of the pandemic on historically underserved populations of students and whether/how longstanding achievement gaps have persisted or changed during the pandemic, we examine standardized differences in average scores across key subgroups. In addition to achievement gaps between student demographic groups, we examine achievement gaps between students whose districts offered different instructional modalities (i.e., in-person, remote, hybrid) throughout the 2020-21 school year.

> Of course there are several intersecting factors that contribute to these achievement gaps. For instance, socioeconomic differences across students of different races/ethnicities likely account for some of the achievement gaps across race/ethnicity subgroups. We focus on gaps across entire populations of students to help us understand the overall magnitude of each achievement gap, given the many complex and interrelated factors that drive these differences. However, this means that we do not adjust for additional student characteristics and that our analyses do not tell us how large achievement gaps are between otherwise-similar students from two different subgroups.

Our analyses include benchmark assessment results from about 750,000 of Michigan's 935,000 K-8 students in 735 of Michigan's 848 school districts. While these analyses help to deepen our understanding of how Michigan public school students progressed and learned between fall 2020 and fall 2021, it is important to stress that they are based on imperfect and incomplete data. For instance, the sample of students who began kindergarten in fall 2020 is especially small due to declines in kindergarten enrollment. Further, prior research has shown that the pandemic has had a greater negative effect on achievement and achievement growth for students who are Black, Latino/a/x, economically disadvantaged, or English learners, and these same student populations are underrepresented in our analysis.

Michigan's benchmark assessment legislation allows districts to choose an appropriate assessment from one of four MDE-approved providers, and thousands of students participated in assessments from each of the four. As in the 2020-21 school year, NWEA's MAP Growth assessment was Michigan's most frequently used benchmark assessment; more than 550,000 students and more than 600 districts participated in an NWEA MAP Growth assessment in fall 2021. Nearly 130,000 students, from some of the largest districts in the state, participated in Curriculum Associates' i-Ready Diagnostic assessments, over 60,000 students participated in Renaissance Learning's Star 360 assessments, and over 5,000 participated in one of DRC's assessments (the MDE K-2s or the Smarter Balanced Interim Comprehensive Assessments for grades 3-8). These assessments are all designed in slightly different ways, cover slightly different content, and tend to appeal to different types of districts. However, as we discuss in the section that follows, we find several common themes in the results from different assessments.

KEY FINDINGS

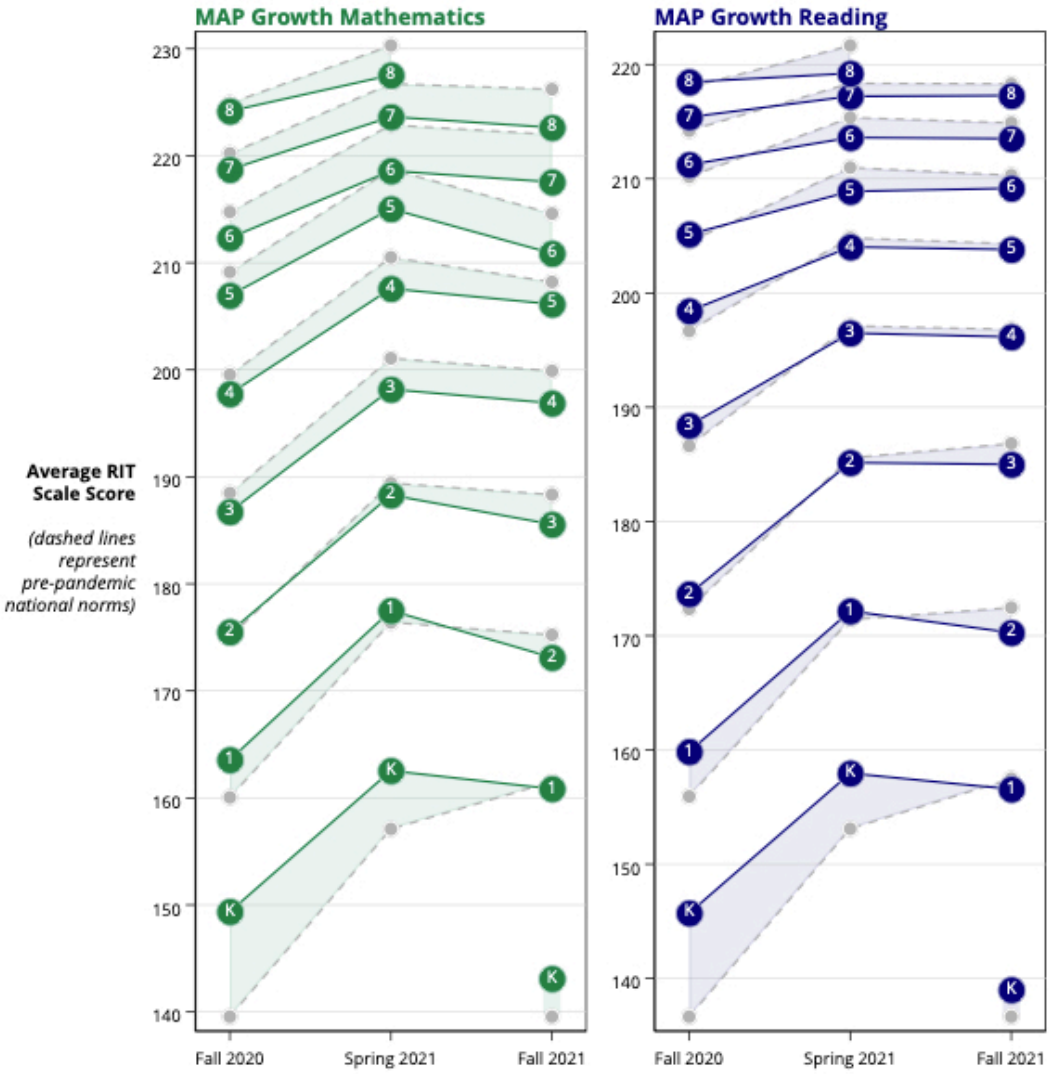
At-Home Testing Conditions in Fall 2020 Make it Difficult to Assess Younger Students' Performance

Across all our analyses, we consistently find abnormal achievement patterns for students in grades K-2, which suggest that many of these students' fall 2020 scores overstate their true knowledge or skill levels due to help they received from their parents or caregivers during the test. NWEA and Curriculum Associates identified similar patterns in their reports, indicating an "at-home advantage" for early elementary students who took their fall 2020 assessments remotely (Huff, 2020; Kuhfeld, Lewis, et al. 2020). Although spring 2021 and fall 2021 scores more accurately reflect students' true performance, it is difficult to measure or interpret growth from fall 2020 to fall 2021 for students in these early grades without knowing where students truly started. Throughout the remainder of our findings, we note whether and how this issue affects our understanding and interpretations of the data.

In Fall 2020, Michigan Students Were Scoring Close to Pre-Pandemic Norms

Figure I shows trends in average scores of students who took the MAP Growth assessments in fall 2020, spring 2021, and fall 2021. In the fall of 2020, students in upper elementary and middle school grades tended to score very close to pre-pandemic norms. This suggests that, although students experienced substantial disruptions to their instruction during the initial phase of the pandemic in late spring 2020, they still performed about as well in fall 2020 as students from a nationally representative sample who took the same tests before the pandemic. Notably, fall 2020 scores for early elementary students were often above pre-pandemic national or state norms, especially for kindergarteners. The new cohort of students who began kindergarten in fall 2021 scored much closer to pre-pandemic norms, suggesting that the higher scores in fall 2020 were likely due to an "at-home advantage," and are not a true reflection of where students started the 2020-21 school year.

Figure I. Trends in Average Scale Scores on NWEA's MAP Growth Mathematics and Reading Assessments



Notes: This figure includes all Michigan students who participated in comparable MAP Growth assessments in all possible test periods. The comparison points represent the median growth for students who began at the 50th percentile. The y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: Districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

By Fall 2021, Michigan Students Tended to Score Below Pre-Pandemic Grade-Level Norms

Figure I shows that, across most grade levels, average scale scores fell below national or state norms by fall 2021; this pattern was consistent across all assessment providers with growth data and was generally more pronounced in mathematics than in reading. Michigan students are far from alone in this, as recent studies from across the U.S. find similar district-, state-, and national-level trends (e.g., Jack et al., 2021; Kuhfeld et al., 2022; North Carolina State Board of Education, 2022; Regional Educational Laboratory, 2022).

Three-Quarters of Michigan Students Demonstrated Growth From Fall 2020 to Fall 2021, But Only About 40% Reached Their Growth Targets for the Year

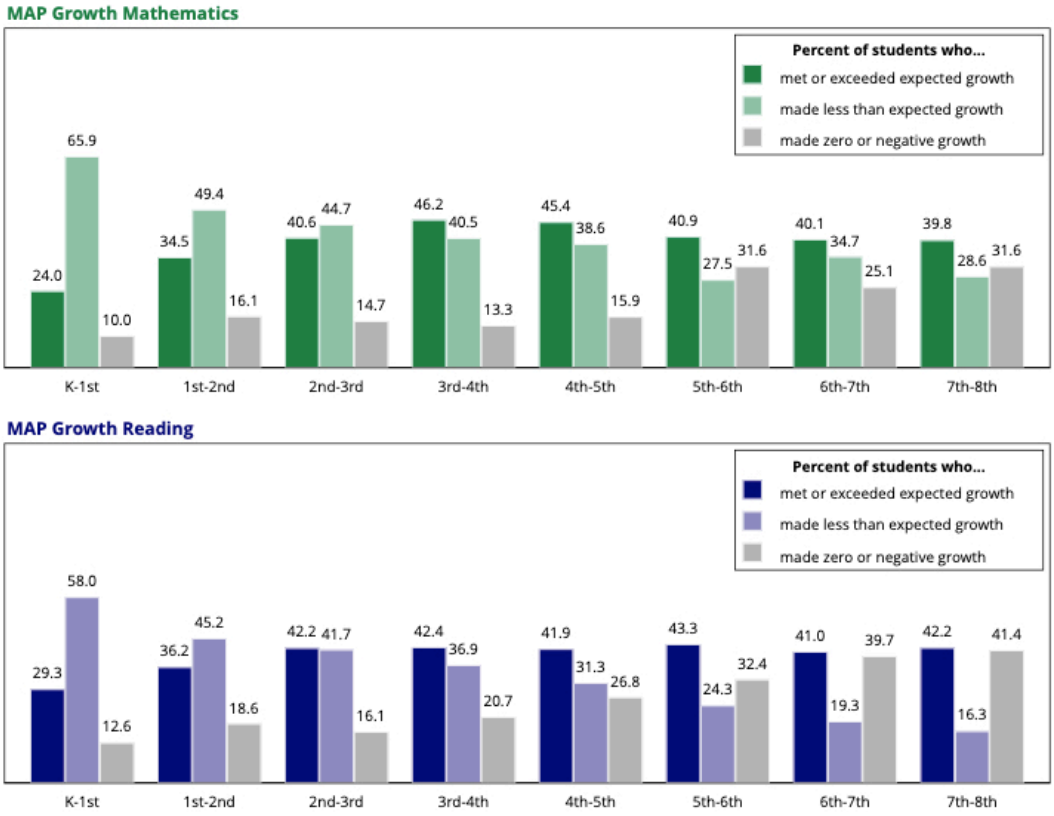
On average, scale scores increased from fall 2020 to fall 2021, but at slower rates than before the pandemic. However, this was not universally true for all students. Across all grade levels, subjects, and assessment providers, 41% of students met or exceeded their growth targets, while 35% made partial progress toward (but did not reach) their growth targets, and 24% did not demonstrate any growth, meaning that their scale scores decreased or did not change from fall 2020 to fall 2021.

Figure II shows that most early elementary students made partial progress toward their growth targets, while upper elementary students were more likely to reach or exceed their targets. Outcomes for middle school students were split with substantial proportions who did not demonstrate any growth at all and substantial proportions who met or exceeded their targets. On average across subjects and assessment providers, 38% of early elementary students, 43% of upper elementary students, and 41% of middle school students reached or exceeded their growth targets, while 18% of early elementary students, 22% of upper elementary students, and 36% of middle school students did not demonstrate any growth.

Across all grade levels, students with the highest initial scores were the least likely to demonstrate growth from fall 2020 to fall 2021. Upper elementary and middle school students in the highest prior achievement group were about 33% more likely to demonstrate no fall-to-fall growth. This pattern was starker in early elementary grades; K-2 students in the highest prior achievement group were more than twice as likely as those in the second-highest prior achievement group to demonstrate no fall-to-fall growth. One reason for this is that students' growth targets are based on their initial (fall 2020) scores, so students whose initial scores did not accurately reflect their

knowledge or skill levels may have been assigned growth targets that were not realistic without the same "at-home advantage."

Figure II. Fall-to-Fall Growth Outcomes on NWEA's MAP Growth Mathematics and Reading Assessments by Grade



Notes: "Expected growth" is based on the median growth for students with the same initial percentile rank (Thum & Kuhfeld, 2020). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

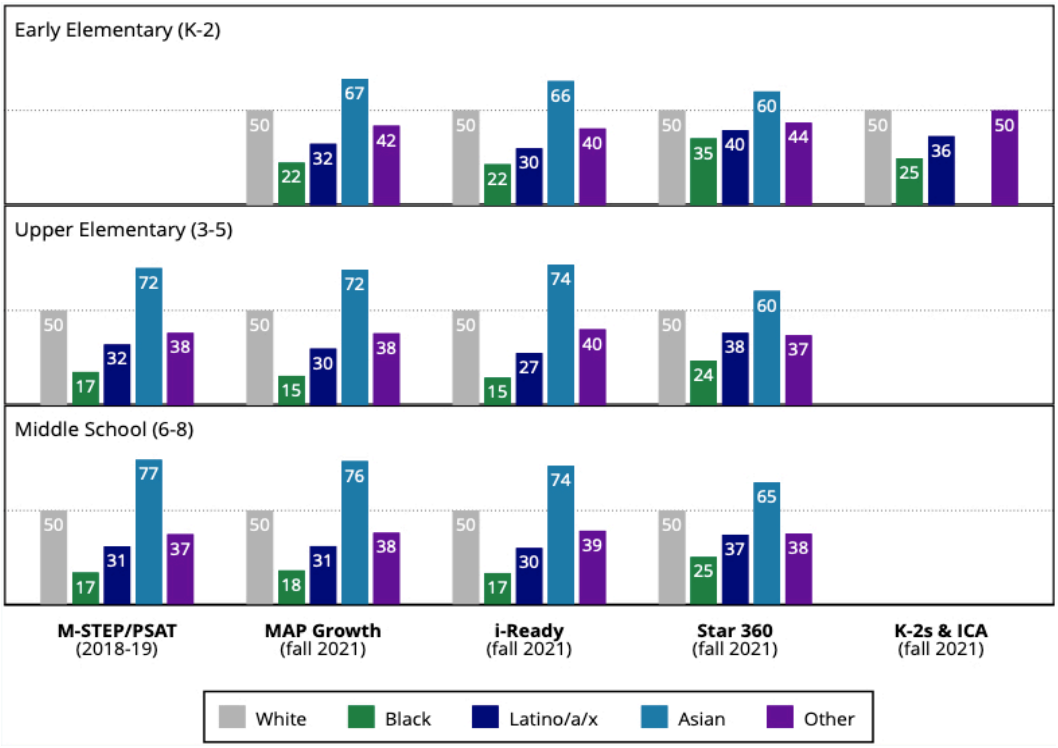
Longstanding Racial and Socioeconomic Achievement Gaps Persisted into the 2021-22 School Year But Did Not Appear to Grow Over the Course of the Pandemic

In fall 2021, Black-White achievement gaps were consistently large and negative. For example, the green bars in Figure III show that average MAP Growth scores for Black students were between the 15th and 22nd percentiles in math, relative to the distributions of scores for White students. Average math scores for Latino/a/x students were between the 30th and 32nd percentiles, while those for Asian students

were between the 67th and 76th percentiles. Math achievement gaps for the other benchmark assessments (shown in Figure III) mirror these same trends. Though not shown in the figure, disparities in reading achievement by race/ethnicity were similar to those for math, but slightly smaller in magnitude.

These achievement gaps are a serious problem, but not a new one; as Figure III shows, achievement gaps between race/ethnicity subgroups on the fall 2021 benchmark assessments are comparable in size to the achievement gaps on Michigan's 2018-19 summative (M-STEP and PSAT) assessments.

Figure III. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 State Summative Assessments by Race/Ethnicity (Math)

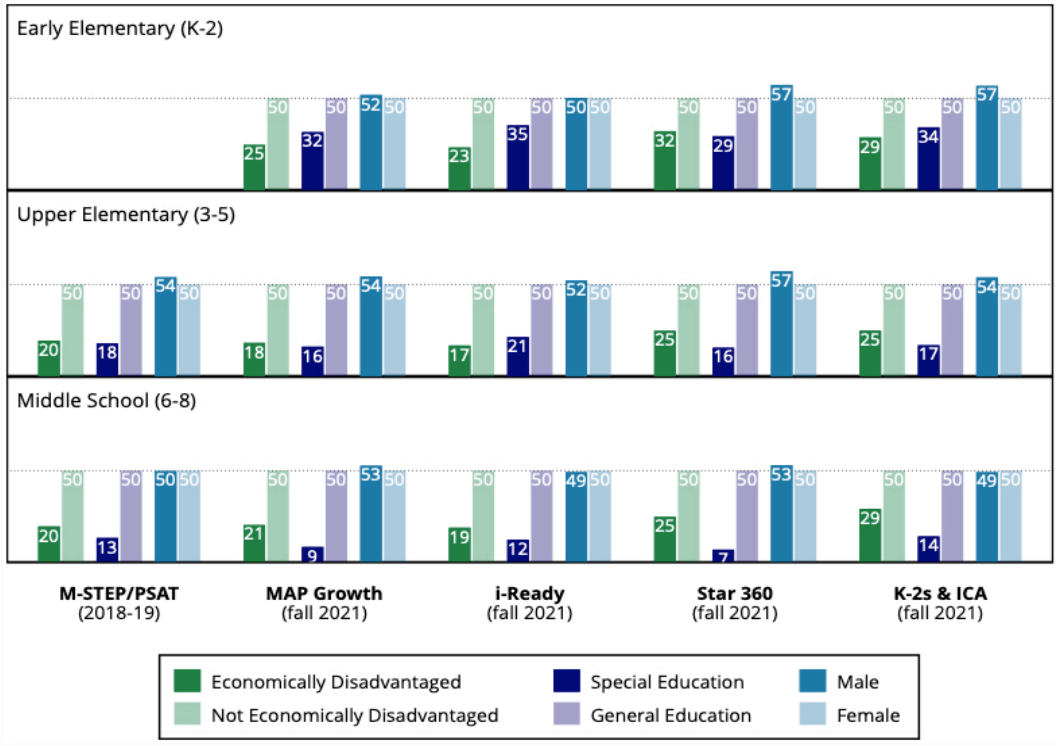


Notes: The numbers in this figure represent the percentile rank of the average score for each subgroup, relative to the distribution of scores for the largest subgroup (White students). We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. There were too few Asian students who participated in the K-2s, and too few students in any non-White subgroup who participated in the Smarter Balanced ICAs, to calculate relative percentiles. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Figure IV shows that average MAP Growth math scores for economically disadvantaged students were between the 18th and 25th percentiles, relative to the

distributions of scores for students who are not economically disadvantaged, while average scores for special education students we between the 12th and 35th percentiles for general education students, and those for male students were between the 52nd and 54th percentiles for female students. Once again, these patterns are consistent across benchmark assessments and similar to the pre-pandemic achievement gaps on the 2018-19 M-STEP and PSAT assessments. Gaps in reading scores (not shown in the figure) were slightly smaller in magnitude but followed similar patterns. The one exception to this is the gender gap, as female students tend to score slightly higher than male students in reading, while male students tend to score slightly higher in math.

Figure IV. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 State Summative Assessments by Subgroup (Math)

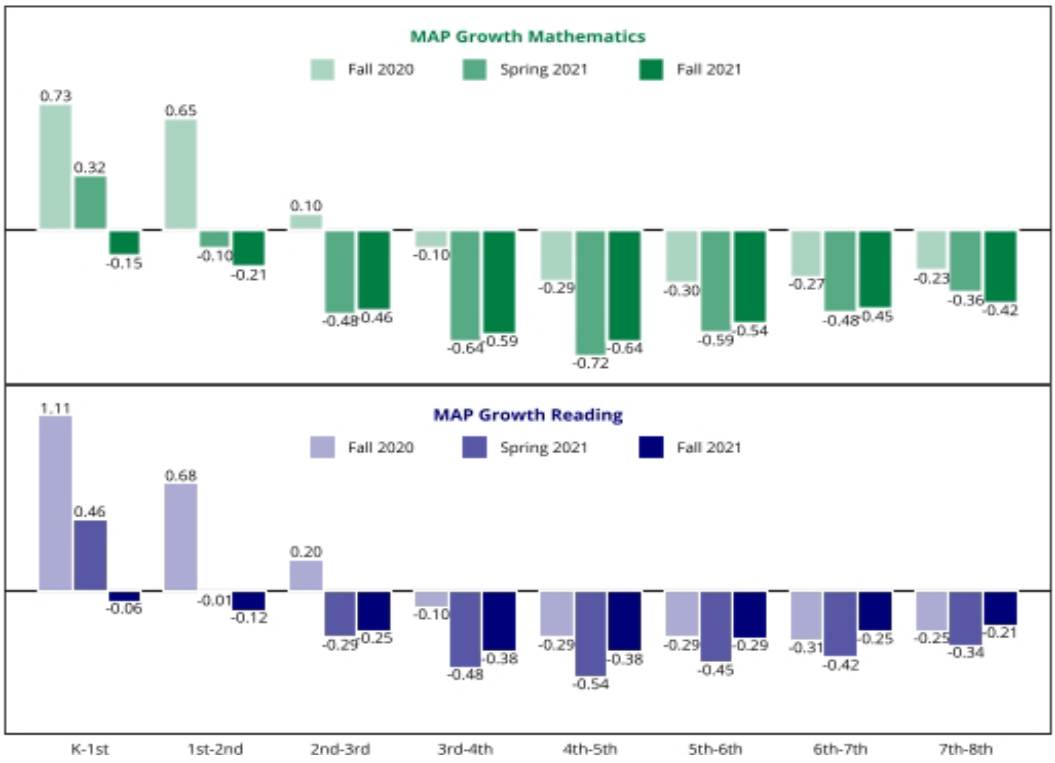


Notes: The numbers in this figure represent the percentile rank of the average score for a subgroup (economically disadvantaged students, special education students, or male students), relative to the distribution of scores for students who are not part of the subgroup. We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Gaps Between Districts That Were Fully Remote in 2020-21 and Those That Offered In-Person Instruction Grew During the 2020-21 School Year and Improved Slightly Over the Summer of 2021

Figure V shows math and reading achievement gaps on the MAP Growth assessments between students in districts that were fully remote throughout the 2020-21 school year and those whose districts offered in-person instruction all year, across all students who were tested in the fall 2020, spring 2021, and fall 2021 semesters. In upper elementary and middle school grade levels, these gaps are consistently negative, indicating that students in remote-only districts performed lower than students in districts that offered fully in-person instruction. This is true even at the beginning of the 2020-21 school year, suggesting that students in fully remote districts (which were largely urban, charter, and lower-income districts) started the 2020-21 school year behind their peers enrolled in districts that offered in-person instruction for the entire school year.

Figure V. Remote All Year-In-Person All Year Standardized Achievement Gaps on NWEA's MAP Growth Mathematics and Reading Assessments



Notes: The gaps in this figure are based on all Michigan students with valid test scores on the MAP Growth assessments in the same subject for fall 2020, spring 2021, and fall 2021. "Remote All Year" districts offered

remote instruction for their students during both the fall 2020 and spring 2021 benchmark assessment periods. We calculate achievement gaps by subtracting the average score for in-person districts from the average score for remote districts, then dividing by the standard deviation of scale scores for students in in-person districts. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

For early elementary grade levels, fall 2020 gaps were consistently *positive*. This seems to substantiate the "at-home advantage" for younger students who completed their benchmark assessments remotely. Kindergarteners had the largest fall 2020 gaps and were the only students for whom these gaps remained positive in spring 2021 (although they became much smaller). By fall 2021, the remote/in-person gaps were negative across all grade levels.

Students With Access to Some In-Person Instruction in 2020-21 Were Less Affected Than Those Whose Districts Offered Only Fully Remote Instruction

Mathematics and reading gaps for students in districts that were fully in-person for part of the year or hybrid for part of the year were typically smaller than gaps for students in districts that were remote all year. This suggests that access to even a limited amount in-person instruction was beneficial for student achievement. Many of these gaps increased during the 2020-21 school year, suggesting that the effects of disrupted instruction grew over time. However, decreases in gaps from spring 2021 to fall 2021 may indicate that students who received less in-person instruction in 2020-21 are starting to catch up to their peers now that most of their districts have resumed in-person instruction.

SUMMARY

This report shows that Michigan students were, on average, testing at about pre-pandemic levels in fall 2020. However, achievement growth slowed over the course of the 2020-21 school year, with only about 40% of Michigan students reaching expected growth targets by fall 2021. At the beginning of the 2021-22 school year, students were achieving, on average, below pre-pandemic grade level norms. This is not unique to Michigan, but rather echoes findings from other research in districts and states across the U.S. that consistently show lower average achievement (e.g., Jack et al., 2021; Kuhfeld et al., 2022; North Carolina State Board of Education, 2022) and less than typical achievement growth (e.g., Regional Educational Laboratory, 2022) than would have been expected for similar students before the pandemic.

While earlier analyses found evidence of substantial gaps in student achievement on benchmark assessments between students of color and White students and between economically disadvantaged students and their wealthier peers, results here show that these gaps—while still alarming—did not appear to grow over the course of the pandemic. However, we do find evidence that students enrolled in districts that operated fully remotely over the 2020-21 school year experienced less achievement growth than their peers in in-person districts, although these gaps decreased slightly over the summer of 2021.

All these results must be placed in the context of the imperfect data available to analyze student learning growth during the pandemic. Not only were participation rates lower than for a typical end-of-year summative assessment, but the resulting analytic samples are not entirely reflective of Michigan's larger student population. There are lower assessment participation rates amongst Black, Latino/a/x, and economically disadvantaged students, which may drive some of the findings regarding consistent rather than widening achievement gaps over the course of the 2020-21 school year and into the fall of 2021. In addition, we document a likely "at-home advantage" for early elementary students in fall 2020, which makes assessing growth and adherence to expected growth trajectories difficult for these cohorts of students. Nonetheless, the results presented herein provide important information for policymakers, educators, and stakeholders as we continue to grapple with the academic effects of the pandemic on Michigan's students.

Section One: Introduction

There is mounting evidence that students across the country and around the world missed important opportunities to learn during the pandemic. Early estimates of unfinished learning from state and national assessments suggest that students experienced much lower learning gains during the 2020-21 school year relative to previous years. This is particularly the case for students without sufficient access to parent or teacher supports (Kuhfeld, Soland et al., 2020); for low-income, Black, and Hispanic or Latino/a/x students (Azevedo et al., 2020; Baisley et al., 2021; Gross & Lake, 2021; Dorn et al., 2020a, b; Kogan & Lavertu, 2021; Kuhfeld & Tarasawa, 2020), and for those learning remotely (Gross & Lake, 2021; Kogan & Lavertu, 2021; Sass & Goldring, 2021).

To understand student learning and progress toward educational goals during the pandemic, the Michigan legislature enacted new data collection and reporting requirements for local school districts during the 2020-21 (2020 PA 149), 2021-22, and 2022-23 school years (2021 PA 48). This report is the third in a series that we will provide to the Michigan Department of Education (MDE), the governor and the House and Senate standing committees responsible for education legislation to give insight into Michigan students' progress toward learning goals for the 2020-21, 2021-22, and 2022-23 school years. The Education Policy Innovation Collaborative (EPIC) at Michigan State University prepared this report in collaboration with MDE, the Center for Educational Performance and Information (CEPI), the Michigan Data Hub (MDH), and the Michigan Education Data Center (MEDC) at the University of Michigan as a summary of the student academic growth from the fall of the 2020-21 school year through the fall of the 2021-22 school year.

MICHIGAN'S BENCHMARK ASSESSMENT LEGISLATION

On August 20, 2020, Michigan Governor Gretchen Whitmer signed a series of three "Return to Learn" bills intended to provide districts with flexibility to adapt their programs as necessary to safely provide instruction during the pandemic (2020 PA 147, 2020 PA 148, 2020 PA 149). For the 2020-21 school year only, the state legislature

waived many instructional requirements, including minimum number of days and hours and what learning activities count toward the attendance and enrollment calculations that determine state aid allocations. Along with this increased flexibility, the "Return to Learn" legislation outlined a new set of requirements for the 2020-21 school year to ensure that districts continued to adequately meet students' needs without the same instructional requirements in place.

As a condition for receiving state aid for the year, the legislation required each district to develop an extended COVID-19 learning plan describing how it would deliver instruction and establishing educational goals for the 2020-21 school year. These educational goals were to include increased student achievement or growth as measured using one or more benchmark assessments, overall and for all subgroups of students. Districts were required to select and administer appropriate benchmark assessments to all K-8 students at the beginning and end of the school year to determine whether students made meaningful progress toward mastery of state standards in reading and mathematics.

The "Return to Learn" legislation provided districts the option to choose one of four state-approved benchmark assessments or one or more benchmark assessments that contain progress monitoring and enhanced diagnostics in reading and/or mathematics. Alternately or in addition, districts were allowed to use a locally developed benchmark assessment that met the same requirements. While the legislation prohibited the use of these data for accountability purposes, districts that elected to use an approved provider's benchmark assessment were required to compile and report their results through the MDH network for use in a statewide aggregate report for the governor and the House and Senate standing committees responsible for education legislation in the Michigan legislature.

To continue tracking academic progress, the Michigan legislature again passed legislation in summer 2021 that required districts to administer benchmark assessments throughout the 2021-22 and 2022-23 school years (2021 PA 48). The new legislation provided districts with the same flexibility to choose among benchmark assessments. Like requirements for the 2020-21 school year, districts must administer benchmark assessments to all K-8 students in both fall and spring semesters of the 2021-22 and 2022-23 school years.

PURPOSE OF THIS REPORT

MDE, CEPI, and MDH have worked with two university research partners—EPIC at Michigan State University and MEDC at the University of Michigan—for more than a year to compile the 2020-21 and 2021-22 benchmark assessment data districts

provided under the "Return to Learn" legislation. The first report in this series, released in August 2021, found that students across the state missed critical opportunities to learn during the 2020-21 school year; regardless of assessment vendor, subject, or grade level, a substantial set of students scored "significantly behind grade level" on both the fall and spring assessments, with slower rates of benchmark achievement growth than in a typical school year. Further, our second report showed that disruptions to the 2020-21 school year may have exacerbated many pre-existing achievement gaps. In particular, Black, Latino/a/x, economically disadvantaged, and special education students were more likely to be "significantly behind grade level" than their peers at both the beginning and end of the year, and these gaps typically grew larger in magnitude over the course of the year. Moreover, students learning remotely for longer periods during the 2020-21 school year had slower rates of achievement growth than those learning in person.

This third report extends our assessment of student progress toward learning goals during and beyond the 2020-21 school year. Specifically, in this analysis we examine performance on benchmark assessments between the fall 2020 and fall 2021 semesters and assess differences in performance across student subgroups (i.e., race/ethnicity, economically disadvantaged, and disability status) and districts using various instructional modalities during the 2020-21 school year (i.e., fully in-person, hybrid, or fully remote during each test administration period).

In what follows, we first discuss the data and methods we use in this report. Section Three provides results from our analyses. We conclude in Section Four with a brief discussion of the implications of our findings for Michigan K-12 education as we progress through the 2021-22 school year.

Section Two: Data and Methods

Each year, millions of K-12 students across the country participate in benchmark assessments. Benchmark assessments are designed to help educators and administrators track students' progress toward grade-level standards and learning goals, and to provide feedback that helps drive future instruction.

Under Michigan's benchmark assessment legislation, districts must administer a benchmark assessment from the MDE-approved provider list, an assessment that provides adequate progress monitoring, or a local benchmark assessment to all K-8 students at the beginning and end of each school year starting in 2020-21. Districts that choose an assessment from one of the four approved providers are required to provide aggregate data regarding the results of these assessments through the MDH. The MDH is designed to collect student-level data, and districts were encouraged to submit student-level data rather than aggregating the data themselves. Doing so allows MEDC and EPIC to complete all necessary aggregations in a consistent manner across districts, while still ensuring that state agencies maintain access only to aggregate data, as stipulated in Michigan's benchmark assessment legislation (2020 PA 149 and 2021 PA 48).

In this section, we describe the indicators of academic performance from the benchmark assessment data and the analytic samples we use in this report. For a full description of the unique characteristics of each MDE-approved benchmark assessment, please see the first report in this series.

INDICATORS OF ACADEMIC PERFORMANCE ON BENCHMARK ASSESSMENTS

To meet the reporting requirements in Michigan's benchmark assessment legislation, we include analyses about student performance in fall 2021 and, when available, growth from fall 2020 to fall 2021. Below, we provide details about the benchmark

assessment data that districts submitted to the MDH and explain how we use these data to derive indicators of average achievement and fall-to-fall growth.

Trends in Average Scale Scores

This report's first set of analyses examines trends in average scale scores across the fall 2020, spring 2021, and fall 2021 testing periods. The MAP Growth, i-Ready, Star 360, and Smarter Balanced ICA benchmark assessments are scored on vertical scales that are consistent within vendors across all grade levels, so we can compare scores for the same group of students on the same assessment across two different school years. However, each benchmark assessment has its own unique scale and scale scores are not comparable across assessments (e.g., MAP Growth scores range between 100 and 350, while i-Ready scores range between 0 and 800), so we present cohort-specific trends in average scale scores separately for each provider.

As comparison points to help us interpret these overall trends, we plot each trend line alongside each assessment provider's established pre-pandemic national norms. While we use the pre-pandemic national medians as comparison points for all assessments, not all providers calculate or present this information in the exact same ways. For the MAP Growth assessments, we use NWEA's conditional growth distributions to identify comparison points based on the 50th percentiles of initial fall scores, fall-to-spring growth, and fall-to-fall growth (Thum & Kuhfeld, 2020). For Curriculum Associates and Renaissance Learning, we use the fall and spring distributions of scale scores for the norming sample to identify fall and spring medians as pre-pandemic comparison points for the i-Ready and Star 360 assessments, respectively (Curriculum Associates, 2020; Renaissance Learning, 2021b, c, d). Finally, for the Smarter Balanced ICAs, we compare fall scores to the 50th percentile for the norming sample from the prior grade level and compare spring scores to the 50th percentile of the norming sample for the current grade level (Smarter Balanced Validity Research, 2020).

Percent of Students Making Expected Fall-to-Fall Growth

Although we can compare average scale scores across student groups and grades, it is important to note that expectations for test score growth over the course of a school year often differ by grade level, subject, and initial achievement levels. To account for these differences, we also compare changes in students' scale scores to their "expected growth," based on pre-determined growth norms for each assessment provider, subject, grade level, and initial (fall 2020) score range.

The growth norms for each assessment are defined in slightly different ways and have slightly different meanings. For students who completed MAP Growth assessments, we use as a growth norm the 50th percentile of the fall-to-fall conditional growth distribution for students with the same initial (fall 2020) percentile rank (Thum & Kuhfeld, 2020). For Curriculum Associates, we use *typical growth* targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). However, we note that the i-Ready growth targets are designed to measure growth from fall to spring and may not perfectly represent growth norms from fall to fall (as students may have experienced a "summer slide" between spring and fall tests [e.g., McEachin & Atteberry, 2017]). For the Star 360 and Smarter Balanced ICA assessments, we use pre-pandemic scale score distributions to identify "expected growth" as the change in scale scores necessary for a student to maintain the same percentile rank as they did in the previous year (Renaissance Learning, 2020b, c, d; Smarter Balanced Validity Research, 2020). These measures represent the increase in scale scores necessary for a Star 360 or Smarter Balanced ICA student who scored, for example, in the 25th percentile in fall 2020 to also score in the 25th percentile on their fall 2021 benchmark assessment.

While these benchmarks help us gain a better understanding of academic growth among Michigan students during the pandemic, it is important to note that we are using summary tables released by each assessment provider to assign growth norms to groups of students. Each assessment provider uses sophisticated student-level models to derive growth measures and we are unable to perfectly replicate those measures from just the summary tables and the aggregate district-level data and analyses made possible under the Return to Learn legislation. For example, in their growth calculations, most assessment providers account for the amount of instruction a student received between two testing occasions, based on the test dates relative to the district's instructional calendar. For our aggregate, statewide analyses, we cannot account for the exact amount of instructional time between each student's fall 2020 and fall 2021 assessments and assume that each student's fall tests were taken one year apart for the purpose of assigning growth norms.

To compare students' actual growth to their "expected growth," we first calculate the difference between each student's fall 2021 and fall 2020 scale scores, then divide this fall-to-fall change by the appropriate growth norm (i.e., the *expected scale score increase* based on the assessment provider, grade level, subject, and the student's initial achievement level). The resulting percentage indicates how much of the "expected growth" a student achieved; this may be greater than 100% if a student's actual scale score growth exceeds the expected growth and may be less than 0% if a student's fall 2021 scale score was lower than the fall 2020 scale score. However, on

average, we would expect these percentages to fall between 0% and 100% given that our December 2021 report shows that average scores increased over the last year, but at slower rates than typically expected before the pandemic.

Before aggregating the data to the district level, we group students into three categories that describe their fall-to-fall growth. The categories are students who achieved:

- no growth (i.e., their percent of expected growth is 0% or less, meaning that their scale scores remained the same or decreased from fall 2020 to fall 2021),
- partial growth (i.e., their percent of expected growth is between 1% and 99%, meaning that their scale scores increased by an amount less than their expected growth), or
- full growth (i.e., their percent of expected growth is 100% or greater, meaning that their scale scores changed by at least their expected growth).

Standardized Achievement Gaps

We also examine achievement gaps, or differences between the average scale scores for two subgroups of students, across multiple student and district characteristics. We present grade-specific achievement gaps by race/ethnicity, economically disadvantaged status, and special education status. In addition, we calculate gaps across districts that offered different modes of instruction throughout the 2020-21 school year. We focus on 2020-21 instructional modality in this report, rather than 2021-22, as most student learning between fall 2020 and fall 2021 took place during the 2020-21 school year. In future analyses, we will work to incorporate new student-level instructional modality data from the 2021-22 school year. We calculate and present standardized achievement gaps separately for each assessment provider just as we do for the analysis of average scale score trends.

As we discussed in our second report, we can judge the relative size of an achievement gap by comparing it to the standard deviation of scale scores for students in the reference group (e.g., for the gap between Black students and White students, we would divide the difference between average scale scores for Black and White students by the standard deviation of scale scores for White students). This allows us to compare achievement gaps with those from other studies or other assessments. For instance, Black-White gaps on the 2019 M-STEP ranged from -0.75 to -0.87 standard deviation units in ELA and -0.90 to -0.99 in math, depending on the grade level. This method also allows us to convert standardized gaps to percentiles and compare achievement for the focal group relative to the distribution of scores for the reference group. For instance, the standardized Black-White gaps on the 2019 M-STEP indicate that the average score for Black students is at the 19th to 23rd percentile for

White students in ELA and the 16th to 18th percentile for White students in math. Because we are comparing the distributions of the two subgroups at a particular point in time, we use the standard deviation for the reference group at that same point in time to calculate the standardized gaps. However, this means that some changes in gaps over time may be due to a change in the reference group's standard deviation, rather than a change in the difference between the average scores of the reference group and focal group.

DATA AGGREGATION AND ANALYSIS

Before aggregating the student-level benchmark assessment data provided through the MDH, we restricted the sample to exclude: 1) districts that were not required to report data under Michigan's benchmark assessment legislation (i.e., districts that did not use products from an MDE-approved assessment provider and districts that opened after or closed before the official fall student count date); 2) students who are not in grades K-8; 3) results from assessments in subject areas other than math and reading/ELA or that only cover a narrow sub-topic within math or reading/ELA (i.e., the Smarter Balanced Interim Assessment Blocks, or IABs, which do not cover as broad a range of topics as the ICAs); and 4) results from assessments that are not normed for the grade level of the assessed student (i.e., results from Star Early Literacy assessments for students above grade 3 and results from Star Math assessments for students in kindergarten).

We merged the benchmark assessment data with student characteristics from the Michigan Student Data System (MSDS) fall 2021 General Collection for the purpose of identifying student subgroups based on their race/ethnicity, economically disadvantaged status, and special education status. We excluded students from subgroup breakdowns if they were missing the necessary demographic data to determine whether they belonged to the subgroup of interest. We also merged these data with district-reported information about the mode(s) of instruction (fully in-person, hybrid, or fully remote) offered each month of the 2020-21 school year.

To construct the final aggregate data files used for the analysis, we calculated the average and standard deviation of scale scores, and the percentages of students who achieved no growth, achieved less than "expected growth," and met or exceeded "expected growth" from fall 2020 to fall 2021 across all students in the same subgroup and grade level who completed an assessment from the same provider. We calculated each of these aggregate measures both by district and for the whole state. We then combined the resulting aggregate datasets with data from individual districts that prepared their own aggregate data files in lieu of submitting student-level data through the MDH. Using our final state-level aggregate dataset, we calculated

standardized gaps in average scale scores for subgroups by subtracting the average scale scores for each "reference group" from the average scale scores for the corresponding "focal group" and dividing this difference by the standard deviation of scale scores for the reference group. We completed this process separately for three analytic samples (described in the next sub-section) to create aggregate measures appropriate for examining fall 2021 performance, growth from fall 2020 to fall 2021, and trends across the fall 2020, spring 2021, and fall 2021 testing periods. This report's results are aggregated to the state level. To prevent identification of any individual students from very small subgroups, we suppress results for any cells that represent fewer than ten students.

District Participation

Under Michigan's benchmark assessment legislation, school districts serving K-8 students are expected to submit benchmark assessment data in some form. For this analysis, CEPI identified districts of interest as those that served students in at least one grade level within the K-8 range and were open before the official fall student count date for the 2021-22 school year (October 6th, 2021). The analysis that follows represents 735 of the 848 Michigan school districts that meet all these criteria.

In total, 746 districts provided some form of benchmark assessment data through the MDH. Of these, 724 provided student-level data, 21 provided aggregate files that they prepared themselves, and one provided both student-level and aggregated data. We omitted 11 of these districts from our analyses because all the assessment results they provided were from time periods, grade levels, or subject areas that are not within the scope of this report. The remaining 735 districts (714 that provided student-level data, 21 that provided aggregate data, and one that provided both) are represented in our analyses. This includes 618 districts using NWEA's MAP Growth, 66 using Curriculum Associates' i-Ready assessments, 70 using Renaissance Learning's Star 360 assessments, and 29 using DRC's ICAs and MDE's K-2s. Forty-six of these districts administered assessments from two different providers and one used assessments from three providers. These 735 districts teach 817,560 K-8 students, or 87.4% of the population of K-8 students in Michigan.

The "Return to Learn" legislation specifies a few options for districts as alternatives to the four approved benchmark assessment providers. Of the 102 districts that did not provide any data through the MDH, 67 indicated through the Grant Electronic Monitoring System/Michigan Administrative Review System (GEMS/MARS; 2020 PA 149) that they selected an alternate vendor or locally developed assessment, 14 indicated that they did not plan to submit any benchmark assessment data, 5 did not provide the necessary authorization for MEDC and EPIC to access their data in the MDH, and the

remaining 16 provided the authorization but did not have any student benchmark assessment data in the MDH by the deadline for us to include them in this report.

Sample Characteristics and Representativeness

Our full analytic sample includes data from 750,789 Michigan students, representing 80% of K-8 students in the state. This includes 679,555 students with fall 2021 benchmark assessment scores for at least one subject in the student-level data that districts submitted to the MDH and 71,234 students who are represented in the aggregate datasets that districts prepared themselves. To understand how the analytic sample compares to the population of Michigan K-8 students, Table 2.1 presents average characteristics for the statewide population of K-8 students and the samples of students with fall 2021 data from each of the four approved assessment providers. Because the districts that prepared their own aggregate datasets did not indicate how many students were tested in both subjects, we can only account for the lowest possible number of tested students from these districts (i.e., the larger of the two subject-specific counts). This means that the actual numbers of students who are represented in at least one subject area may be slightly larger than the total counts in Table 2.1.

As seen in the first column of Table 2.1, across the entire state, about 64% of K-8 students are White and 18% are Black. More than half of Michigan's K-8 students are economically disadvantaged, 13% qualify for special education services, and 7% percent are eligible for English learner services. In general, students who took the NWEA MAP Growth assessment in the fall 2021 analytic sample are relatively similar to the statewide population of K-8 students. This is not surprising given that districts that administered the NWEA MAP Growth assessment educate more than 50% of all Michigan K-8 students. Students in i-Ready districts are less representative of the full population of K-8 students than MAP growth districts. Specifically, students in i-Ready districts with at least one valid assessment score in fall 2021 were substantially more likely to be Black, less likely to be White, and somewhat more likely to be Asian compared to Michigan's full K-8 population. Compared to the rest of the state, students who participated in the Star 360 assessments are overwhelmingly White and considerably less likely to be eligible for English learner services. Students who participated in either the Star 360 or DRC assessments (K-2s and ICAs) are the least likely to be economically disadvantaged. Students who participated in the DRC assessments are also least likely to be eligible for special education services.

Table 2.1. Summary Statistics of K-8 Students in All Michigan Districts and by Assessment Provider

Demographics (%)	All MI	MAP Growth	i-Ready	Star 360	K-2/ICA
Female	48.6	48.8	48.4	48.4	49.0
Asian	3.6	2.9	4.7	1.5	0.3
Black	18.2	16.4	35.6	7.7	17.5
Hispanic or Latino/a/x	8.7	8.5	8.6	9.2	3.7
White	63.5	65.9	46.5	74.9	75.2
Econ. Disadvantaged	53.9	53.0	59.5	51.9	51.6
Special education	13.3	12.4	11.7	13.4	9.9
English learner	7.2	6.7	9.2	4.0	1.7
N students	934,969	566,547	128,629	61,241	5,066
% of all MI K-8 students	100	60.6	13.8	6.6	0.5

Notes: The "All MI" column includes the full population of K-8 students across Michigan. Each vendor-specific column includes all students who both attended a district offering MAP Growth, i-Ready, Star 360, or K-2/ICA assessments and had at least one valid test score during the fall 2021 semester, respectively. Sources: School districts submitted information regarding the assessment offered directly to the MDH. These data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. Student enrollment data is from the fall 2021 Michigan Student Data System collection.

Research exploring trends in academic achievement over the past two years, including EPIC's second report using Michigan benchmark data, makes clear that the effects of the COVID-19 pandemic on students varied across student populations and the pandemic has had a greater and more negative effect on the achievement and achievement growth of economically disadvantaged, Black, and Hispanic or Latino/a/x students, as well as English learners (e.g., Amplify Education, 2021; Dorn et al., 2020a, b; Kilbride et al., 2021; Kogan & Lavertu, 2021; Pier et al., 2021; Sass & Goldring, 2021). Given that these specific student populations are underrepresented in the analytic samples for some of the benchmark assessment providers, it is likely that our overall results overstate any academic growth observed throughout the 2020-21 school year.

Restricted Samples

For some of our analyses, we imposed additional sample restrictions to ensure that comparisons of aggregate measures over time reflect changes in student performance as opposed to changes in the populations of students tested. Table 2.2 shows the total number of students and districts for whom we received student-level data through the MDH (i.e., the "fall 2021 sample") and the subsets of these students who we can and cannot include in our aggregate measures. We have two sets of restricted analytic samples: the fall-to-fall sample (shown in the top panel of Table 2.2) and the fall-spring-fall sample (bottom panel).

Each panel of Table 2.2 summarizes the reasons why subsets of districts and students were not included in the more restrictive samples. For analyses of fall-to-fall growth, we remove a total of 123 districts and 259,594 students from the fall 2021 sample and include only the 427,961 students who participated in comparable benchmark assessments, in the same district, in consecutive grade levels in fall 2020 and fall 2021. The most common reason for exclusion of districts and students from our fall-to-fall sample was simply that districts did not provide fall 2020 data for a particular student, or often for any students at all. With new funding available for districts to implement benchmark assessment systems (2021 PA 48), many districts that did not use approved benchmark assessments in 2020-21 started doing so for the first time in 2021-22. While we include these districts in our fall 2021 analyses, we cannot yet measure their growth because they do not have data from the prior year.

The fall-spring-fall sample is the most restrictive sample in that it only includes the 416,882 students who participated in comparable benchmark assessments in fall 2020, spring 2021, *and* fall 2021 in the same district, in the same grade level in fall 2020 and spring 2021, and in the next consecutive grade level in fall 2021. About 97% of students and 99% of districts in the fall-to-fall sample were also in the fall-to-spring sample.

Table 2.2. Restricted Analytic Samples and Reasons for Exclusions

Exclusions	Districts		Students	
	N	%	*N*	%
Fall-to-Fall Exclusion Restrictions				
Fall 2021 sample	714	100.0	679,555	100.0
No fall 2020 data: new kindergarten cohort	-0	-0.0	-65,714	-9.7
No fall 2020 data: returning 1st-8th graders	-102	-14.3	-143,377	-21.1
Different district in fall 2020	-10	-1.4	-29,444	-4.3
Different test in fall 2020	-11	-1.5	-13,059	-1.9
Fall-to-fall sample	591	82.8	427,961	63.0
Fall-Spring-Fall Exclusion Restrictions				
Fall-to-fall sample	591	82.8	427,961	63.0
No spring 2021 data	-8	-1.1	-10,161	-1.5
Different district in spring 2021	-0	-0.0	-224	-0.0
Different test in spring 2021	-0	-0.0	-694	-0.1
Fall-spring-fall sample	583	81.7	416,882	61.3

Notes: The counts and percentages in this table do not include data from districts that prepared their own aggregate datasets. Source: School districts submitted benchmark assessment data to the MDH and EPIC received the data through a collaboration between EPIC, MEDC, and MDE.

When possible, we also include data from the 22 districts that prepared their own aggregate files in analyses of the fall 2021 and fall-to-fall samples. We also note that we cannot show trends in average scale scores for the MDE K-2 assessments due to a change in the test score scale starting in fall 2021. For the same reason, we cannot assess growth from fall 2020 to fall 2021 on these assessments. Although the Smarter Balanced ICA assessments use a consistent scale across all three test administration periods, there are too few districts that administered these assessments in both school years, and too few students in those districts with scores from both years, for us to assess trends or growth in a meaningful way. Thus, we only present and discuss results from the MAP Growth, i-Ready, and Star 360 assessments in Section Three.

Table 2.3 presents grade-specific enrollment counts and percentages of enrolled students who are represented in each of our analytic samples. The denominator for each inclusion rate is the aggregate enrollment count across all districts offering a particular benchmark assessment for a particular grade level (e.g., a district may use MAP Growth for some grade levels and a locally developed assessment for others). Since grade-specific enrollment counts and inclusion rates were relatively consistent across our reading and mathematics samples, we provide figures for the percentage of students with valid test scores in at least one subject area. We note, the percentages in this table do not include students from districts that submitted their own aggregate data. Districts reported mathematics and reading outcomes separately but did not indicate how many students participated in benchmark testing for both subjects.

As seen in the table, inclusion rates vary across grade levels and analytic samples. Across all three samples, inclusion rates among elementary school students increase with grade level while rates among middle school students decrease for each grade level. Additionally, inclusion rates across all grade levels are consistently higher among students in the fall 2021 sample since the sample includes any student with at least one valid test score in fall 2021. The largest differences in inclusion rates between the fall 2021, fall-to-fall, and fall-spring-fall samples are for the students who were in kindergarten or 1st grade in fall 2020. Compared to higher grade levels, a larger percentage of students who were in kindergarten or 1st grade in fall 2020 were excluded from the fall-to-fall and fall-spring-fall samples because they did not have testing data in fall 2020 (i.e., 26% of kindergarten and 18% of 1st-grade students in the fall 2021 sample did not have testing data in fall 2020 while only 11 to 13% of students in other grade levels were missing the same data). We find similarly low fall 2021 participation rates for the incoming 2021-22 kindergarten cohort.

Table 2.3. Percent of Enrolled Students Included in Analytic Samples, All Four Assessment Providers					
2020-21 Grade	2021-22 Grade	Enrollment	Inclusion Rates by Analytic Sample		
			Fall 2021	Fall-to-Fall	Fall-Spring-Fall
---	K	82,596	78.8%		
K	1st	75,852	90.7%	51.9%	51.1%
1st	2nd	80,712	90.0%	58.9%	58.1%
2nd	3rd	81,711	92.5%	65.1%	63.7%
3rd	4th	81,916	93.0%	68.4%	66.6%
4th	5th	82,949	93.9%	69.5%	67.7%
5th	6th	83,651	93.9%	68.1%	66.3%
6th	7th	85,258	92.9%	68.0%	65.5%
7th	8th	87,327	92.0%	67.9%	65.4%

Notes: The "Enrollment" column represents the total number of students from a specific K-8 grade level enrolled in districts that offered an NWEA MAP Growth, Curriculum Associates i-Ready, Renaissance Learning Star 360, or MDE K-2/Smarter Balanced ICA benchmark assessment during the fall 2021 semester and provided student-level data to the MDH. The "Fall 2021," "Fall-to-Fall," and "Fall-Spring-Fall" columns represent the percentage of students from each grade with valid mathematics or reading benchmark assessment scores in fall 2021; fall 2020 and fall 2021; or fall 2020, spring 2021, and fall 2021, respectively. Student enrollment data is from the fall 2021 Michigan Student Data System collection.

SUMMARY

Overall, the descriptive statistics and participation rates discussed earlier make clear that the following analyses are based on incomplete, although still meaningful, data. Only 87% of districts (735 of the 848 total districts) and 80% of students in the state (750,789 of the 934,969 total students) are represented in analyses using the least restrictive fall 2021 sample, and those who are represented may not be reflective of those who are not included.

While it is important to keep in mind these limitations when interpreting results, the report nonetheless helps deepen our understanding of how Michigan public school students progressed academically throughout the fall 2020, spring 2021, and fall 2021 semesters. Importantly, the analyses presented in Section Three continue to expand on the descriptive results presented in our previous reports, providing a more refined estimate of academic growth by incorporating another semester of assessment data and comparing academic trajectories of Michigan students to pre-pandemic trajectories of students from across the country.

Section Three: Results

In this section, we summarize Michigan student outcomes on benchmark assessments that districts administered between fall 2020 and fall 2021. First, we show changes in average scale scores across the fall 2020, spring 2021, and fall 2021 test administration periods. Next, we assess how Michigan students' test score growth from fall 2020 to fall 2021 compares to pre-determined growth norms that each assessment provider established before the pandemic. Finally, we examine gaps in average test scores across subgroups of students and changes in these gaps over time. As noted earlier, before interpreting these results, it is important to remember that each benchmark assessment has its own unique scale and scale scores are not comparable across assessments. We therefore analyze data from each assessment provider separately. However, we compare general patterns and some standardized metrics across the different assessments to identify commonalities and points of divergence in the results.

For most of the results that follow, we focus on MAP Growth, i-Ready, and Star 360 assessments. As noted in Section Two, we cannot include the MDE K-2 or Smarter Balanced ICA assessments in our longitudinal analyses, but we do discuss outcomes for students who participated in these assessments in our fall 2021 analyses.

TRENDS IN AVERAGE SCALE SCORES

Background

The results in this section show changes in average scale scores between fall 2020 and fall 2021 for the fall-spring-fall sample. In the appendix, we provide additional figures that repeat this analysis using the full fall 2021 sample and provide tables of values used to create the figures in both the main report and the appendix (see Appendix Figures A1.1-A1.4 and Appendix Tables A1.1-A1.6). We present trends separately for each grade and subject and show each trend line alongside pre-pandemic norms from each assessment provider to help contextualize the overall trends in Michigan.

Typical-year growth trajectories should show increases in average scale scores over time as students receive more instruction and progress academically, and slight decreases between spring 2021 and fall 2021 as students experience "summer slide" (e.g., see McEachin & Atteberry, 2017; Quinn et al., 2016; von Hippel & Hamrock, 2019). However, the 2020-21 and 2021-22 school years have not been "typical." For instance, prior research—including our own—shows that student learning has taken place at a slower rate during the pandemic than would have been expected in a typical year (e.g.,

Amplify Education, 2021; Dorn et al., 2020a; Kilbride et al., 2021; Kogan & Lavertu, 2021; Sass & Goldring, 2021). Thus, we might expect to see trends in average scale scores fall further below pre-pandemic norms over time.

It is also important to note that testing environments for some students changed between the fall 2020 and spring 2021 assessment periods. In particular, students who were tested remotely—which was more likely to occur in fall 2020—may have had access to additional resources (e.g., assistance from parents/guardians) which could have led to inflated fall scale scores. To be clear, we are not suggesting that parents actively helped their children "cheat" during remote testing, however, it is possible that the assistance some parents provided to their children during the testing process exceeded what is normally acceptable during in-school testing (e.g., answering a clarifying question versus helping a student complete a calculation fundamental to answering a larger question on the assessment). In their own reports based on fall 2020 data, NWEA and Curriculum Associates identified an "at-home advantage" for some early grade-level students. For example, Curriculum Associates estimated that a significantly smaller share of 2nd graders who tested outside of school performed two or more grade levels below their peers in fall 2020 compared to historical trends. Similarly, NWEA found that achievement trends between fall 2019 and fall 2020 looked very different between remote and in-person testers; students who tested remotely in the 1st and 2nd grade in fall 2020 showed large increases in their percentile ranks compared to the previous fall, while students who tested in-person showed patterns more consistent with students in higher grade levels (Huff, 2020; Kuhfeld, Lewis, et al. 2020; Renaissance Learning, 2021a).

While we do not have the ability to identify which students completed benchmark assessments at home or in the classroom, NWEA, Curriculum Associates, and Renaissance Learning collected data on students' testing location for at least a portion of the 2020-21 school year and have reported national findings related to modality of assessment. Renaissance Learning determined that, nationally, about 20% of Star 360 tests from the spring 2021 testing period were completed remotely. Curriculum Associates reported that, nationally, approximately 40% of students who took an i-Ready assessment in fall 2020 self-reported taking the assessments remotely, outside of a school building, while only 28% of students reported the same testing environment in the spring 2021 semester. Similarly, the incidence of remote testing decreased over time among students taking MAP Growth assessments in 2020-21, as 54% were tested remotely in winter and 46% in the spring. As a result, we might expect to see average scores *above* pre-pandemic national norms in fall 2020 when more school districts and students were learning remotely (Hopkins, Kilbride, & Strunk, 2021), particularly in younger grade levels, but this may be more indicative of irregular testing conditions than a reflection of students' abilities at that time.

Interpreting Results Figures

Figure 3.1.1 shows trends in average mathematics and reading scale scores for students in the fall-spring-fall sample who completed a MAP Growth benchmark assessment. As a reminder, we use NWEA's conditional growth distributions to identify comparison points based on the 50th percentiles of initial fall scores, fall-to-spring growth, and fall-to-fall growth (Thum & Kuhfeld, 2020). Within Figure 3.1.1, as well as the other two figures in this subsection, the solid green (mathematics) and blue (reading) points and lines represent average scale score trajectories for the cohorts of Michigan students completing a MAP Growth benchmark assessment in fall 2020, spring 2021, and fall 2021. Each solid line represents a different cohort of students, and each point on these lines is labeled to indicate the grade level of students in the cohort during a particular time period. The dashed gray lines represent pre-pandemic norms for a particular cohort of students. The area between pairs of solid and dashed lines is shaded to indicate that both lines correspond to the same cohort of students. When a solid line is above the corresponding dashed gray line, it means average scale scores for that cohort were above the national pre-pandemic norm. Conversely, when a solid line is below the corresponding gray line, average scale scores for that cohort were less than the norm. In each of these figures, the highest and lowest values on the y-axis scales represent the highest and lowest grade-level norms for a particular assessment. These norms differ across subject areas, and as a result, the y-axis scales are different for each subject.

Results

We find that, on average and across grade levels and assessment providers, students experienced achievement growth over time, with slight declines ("summer slide") between spring 2021 and fall 2021. However, these average growth trajectories differed from pre-pandemic norms. In general, growth was slower during the 2020-21 school year and into fall 2021 than would be expected in a typical pre-pandemic year.

Figure 3.1.1 shows that, as expected, average fall 2020 scale scores for kindergarten and 1st-grade students who took the MAP Growth assessments were above mathematics and reading pre-pandemic norms; this is likely due to the "at-home advantage" discussed earlier. This pattern does not hold for students in later grades; for 2nd-8th grade, fall 2020 average mathematics scale scores were at or below the norm while initial average reading scale scores were at or slightly above the norm.

Average mathematics and reading scale scores across all grade levels increased between fall 2020 and spring 2021. However, as can be seen by comparing the slopes of the solid lines to the slopes of the grey dashed lines, the increases realized by

students at each grade level consistently trailed the pre-pandemic norms. As a result, spring 2021 average mathematics scale scores for kindergarten and 1st-grade students were much closer to the pre-pandemic norms, 2nd-grade scores dropped below the norm, and students in 3rd through 8th grade fell even further below the norms than they were in fall 2020. Similarly, spring 2021 average reading scale scores for kindergarten and 1st-grade students were much closer to the norm and 2nd- through 8th-grade average scale scores dropped below pre-pandemic norms.

Between spring 2021 and fall 2021, average mathematics and reading scale scores for most grade levels dropped slightly, as expected due to "summer slide"; only students in 5th or 7th grade in fall 2020 posted increases in average reading scale scores over the summer 2021 months, outperforming pre-pandemic norms over the summer. Additionally, by fall 2021, all average mathematics and reading scale scores for each grade level trailed their respective norms. Nearly all the grade-specific decreases in average mathematics scale scores between spring 2021 and fall 2021, as well as the K-3 decreases in reading, increased the gaps between Michigan student outcomes and the corresponding norms. In other words, the "summer slides" experienced by most MAP Growth students in mathematics and early elementary students in reading brought these students even further below national pre-pandemic norms by the fall of 2021.

Figure 3.1.2 shows trends in average mathematics or reading scale scores for students in the fall-spring-fall sample who completed an i-Ready benchmark assessment. Given the vast differences in the characteristics of students from i-Ready districts relative to the state average (see Section Two of this report), we use the distributions of scale scores from Michigan districts that used the i-Ready assessments in 2018-19 to identify comparison points that better reflect the population of students studied in this report than the national distributions from Curriculum Associates' norming sample. Appendix Figure A1.5 provides trends in Michigan students' average i-Ready scale scores relative to *national* pre-pandemic norms.

Again, as expected given younger students' "at-home advantage," we find that initial average mathematics and reading scale scores for kindergarten, 1st-, and 2nd-grade students in districts that administered the i-Ready assessments were higher than pre-pandemic Michigan norms. Average fall 2020 mathematics and reading scale scores for all other grade levels were below the norms.

Between fall 2020 and spring 2021, in both subjects, all grade levels saw increases in average scale scores. However, as with students in districts offering MAP Growth assessments, the fall-to-spring increases realized by students at each grade level were consistently less than the pre-pandemic norm, resulting in spring 2021 average mathematics and reading scale scores for kindergarten and 1st-grade students that were

much closer to the Michigan norms, 2nd-grade average scale scores that dropped below the norm, and increased disparities for 3rd- through 8th-grade students. Between spring 2021 and fall 2021, average mathematics and reading scale scores for all grade levels dropped slightly (except for 6th-grade reading); however, the decreases in average mathematics and reading scores for students in 2nd through 6th grade were all smaller than spring-to-fall decreases in the pre-pandemic norms. This result is slightly different from the trends seen for MAP Growth districts, where nearly all students fell further below national pre-pandemic norms in mathematics by the fall of 2021.

Finally, Figure 3.1.3 shows trends in average math, reading, and early literacy scale scores for students in the fall-spring-fall sample who completed a Star 360 benchmark assessment. We also use Renaissance Learning's fall and spring distributions of scale scores from their norming samples to provide fall and spring medians as pre-pandemic comparison points. As a reminder, students in grades K-3 can take either the Star reading or early literacy assessments, so we show results for the assessment that is most common for each grade level within this range (early literacy for grades K-1 and reading for 2nd through 3rd grade). Because students typically transition from the Star early literacy to the Star reading assessment between 1st and 2nd grade and these assessments have different scales, we do not show trends for the 1st grade cohort of students.

We find that math, reading, and early literacy scale scores for elementary grade levels were very close to the corresponding pre-pandemic norms across the 2020-21 school year. Fall-to-spring growth in average scale scores for these same grade levels also tended to mirror national norms, and in some cases exceeded the norms. In math, fall-to-spring growth for students in 1st, 3rd, 4th, and 5th grades exceeded national norms, as did fall-to-spring reading growth for 2nd- through 5th-grade students. However, fall-to-spring early literacy growth for both kindergarteners and 1st-grade students was slightly below the norm.

For middle school grade levels, initial math and reading scores were all below pre-pandemic norms. In some cases, these discrepancies were quite large. For instance, 8th-grade students in Star 360 districts started the 2020-21 school year at approximately the 7th-grade norm in both math and reading based on the pre-pandemic norming sample, and 7th-grade students scored, on average, at the 6th-grade norm in math and just above the 6th-grade norm in reading. In addition, fall-to-spring growth for middle school students in Star districts lagged pre-pandemic norms.

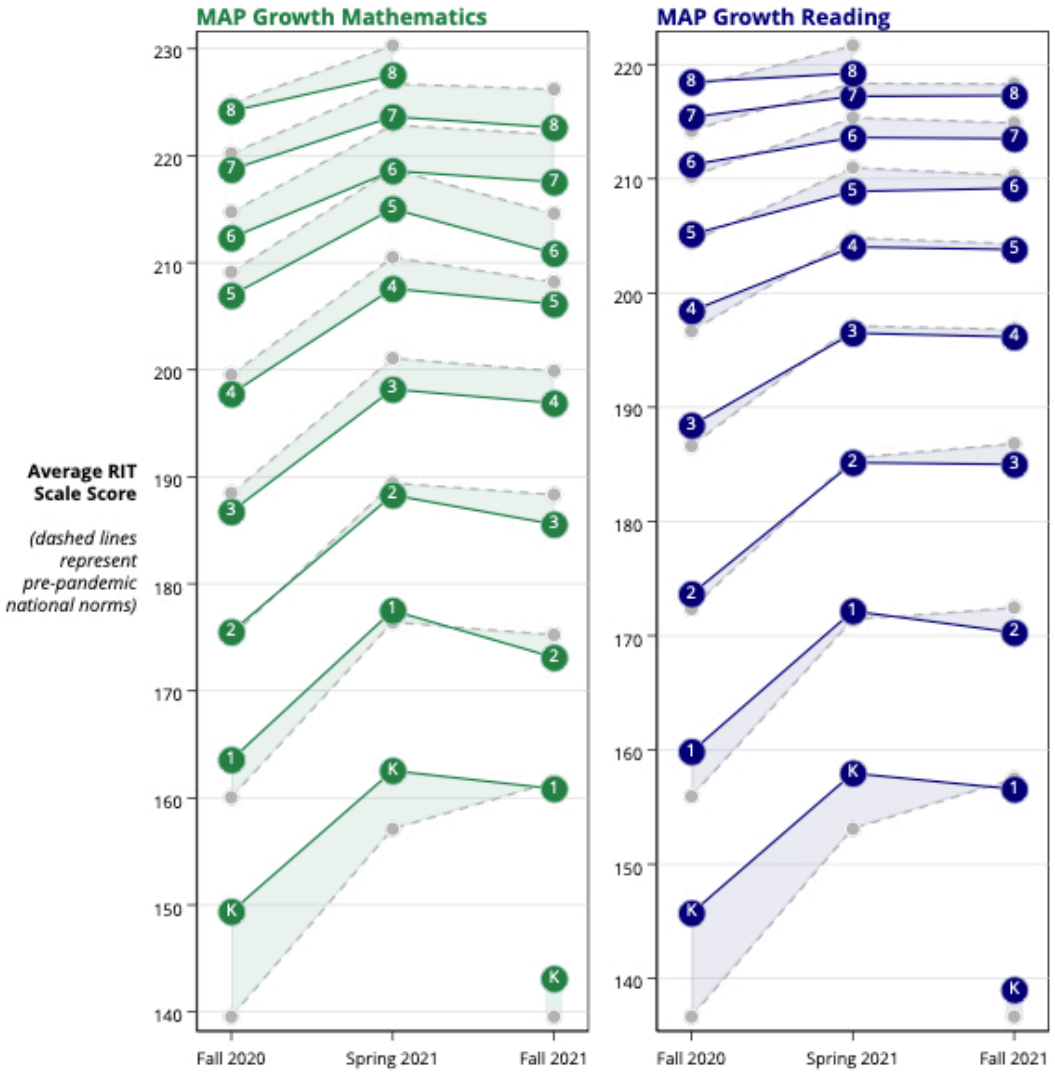
We also saw evidence of summer slide (changes between spring 2021 and fall 2021) that exceeded national norms in districts administering the Star 360 assessments.

Except for reading scores among the cohort of student transitioning from 7th to 8th grade, decreases in average math, reading, and early literacy scale scores were all larger than changes in their respective norms.

Overall, the results in this section show that, on average, students across all grade levels and assessment providers experienced growth in mathematics and reading achievement between fall 2020 and fall 2021. For students in MAP Growth and i-Ready districts, fall 2020 average scores were below comparable norms in mathematics and above norms in reading, fall-to-spring growth in both subjects lagged changes in each respective norming sample, and fall 2021 achievement consistently trailed pre-pandemic norms. For students in Star 360 districts, K-5 fall 2020 achievement was relatively on-pace with the assessment provider's norming sample before slowing over the next two semesters, while middle school achievement lagged norms throughout the sample period.

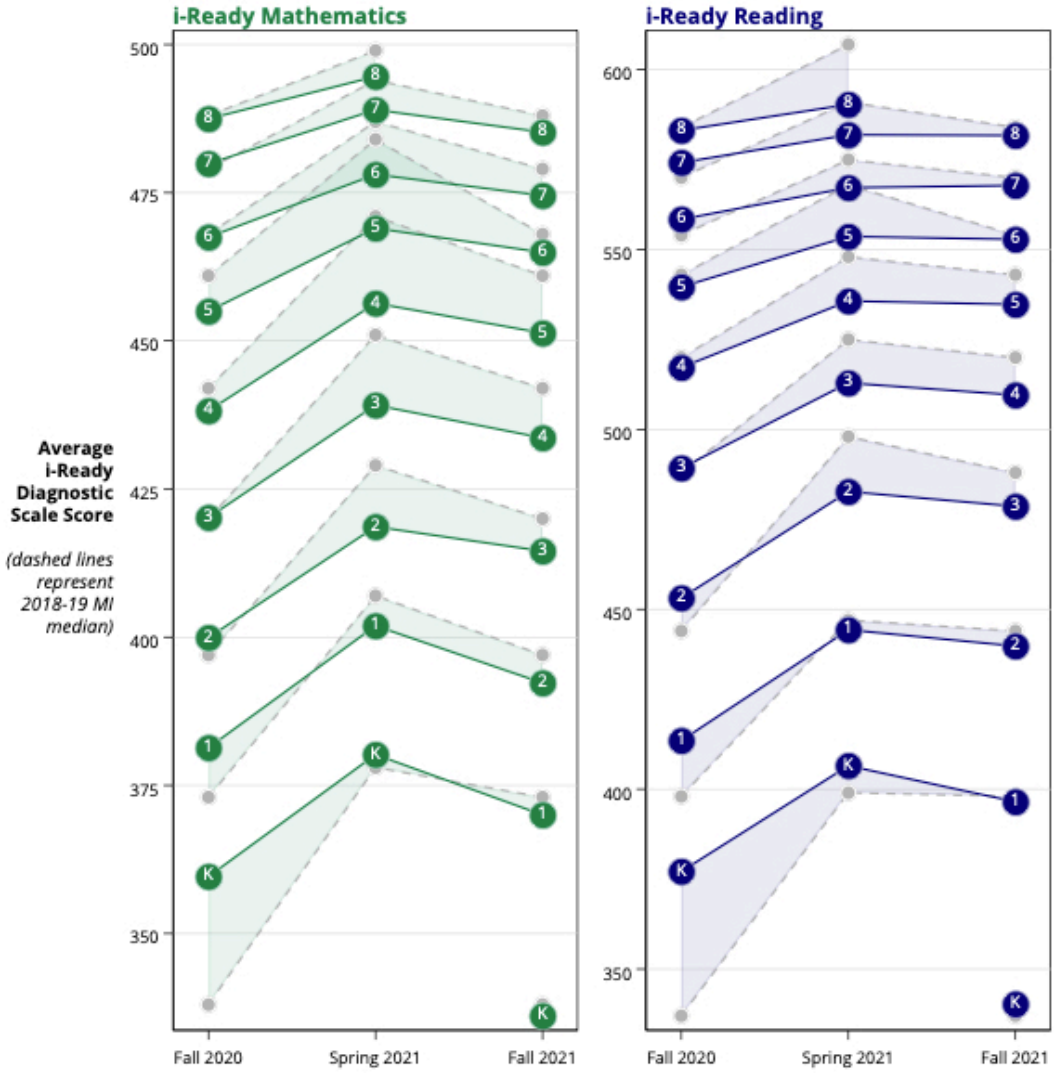
These findings are not unique to Michigan, but rather, mirror many of the district-, state-, and national-level trends in recent studies about student achievement and growth during the pandemic. For instance, using data from more than 5 million 3rd-through 8th-grade students across the country, Kuhfeld, Soland, and Lewis (2022) found that average fall 2021 math and reading test scores in grades three through eight were 0.20 to 0.27 and 0.09 to 0.18 standard deviations below students in the same grade levels in fall 2019, respectively. Similarly, Jack, Halloran, Okun, and Oster (2021) found that, across end-of-year and end-of-course summative assessments in 11 states (Colorado, Connecticut, Massachusetts, Minnesota, Mississippi, Ohio, Rhode Island, Virginia), average pass rates declined by 13 percentage points in math and 6.5 percentage points in ELA from spring 2019 to spring 2021.

Figure 3.1.1. Trends in Average Scale Scores on NWEA's MAP Growth Mathematics and Reading Assessments



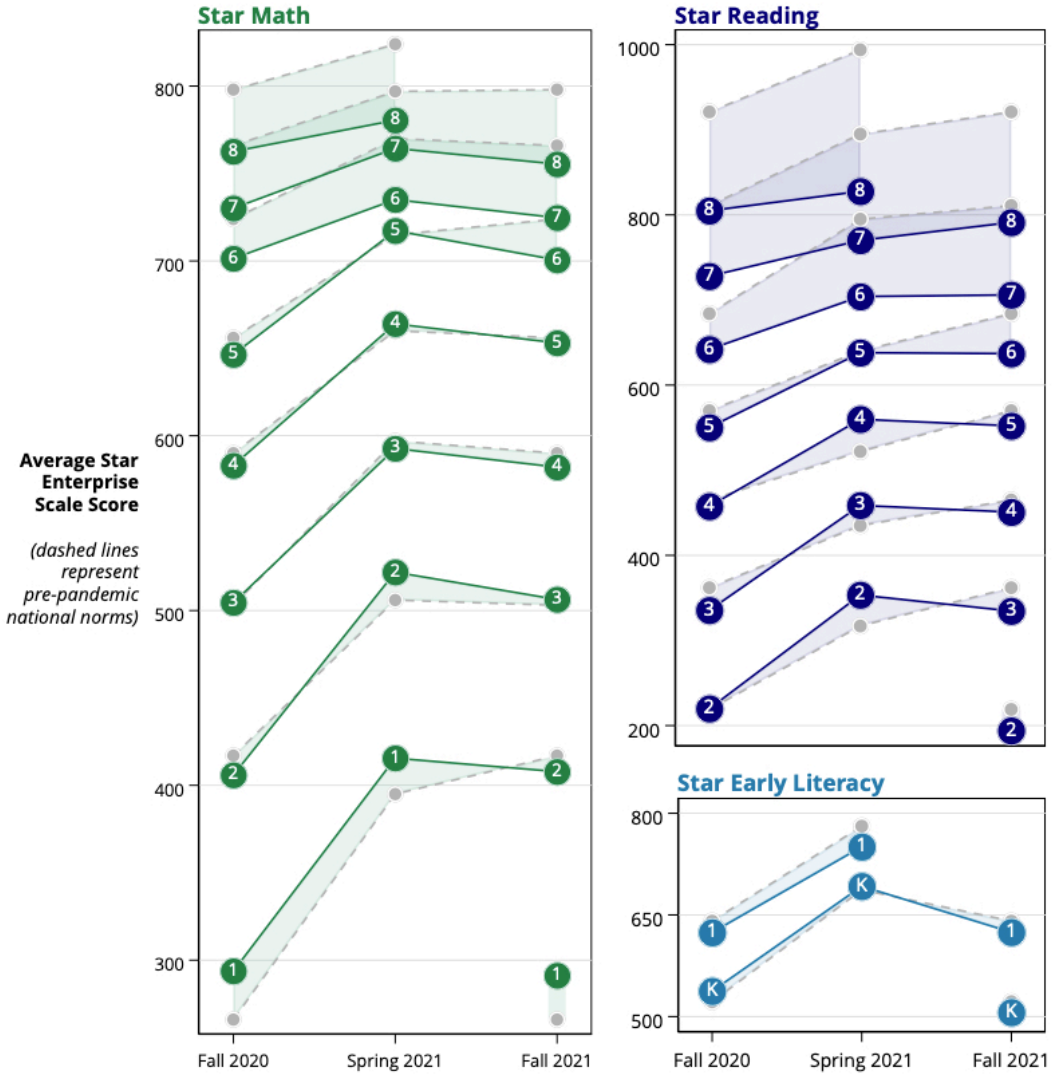
Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth Mathematics or Reading assessments in fall 2020, spring 2021, and fall 2021. The comparison points in the figure represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. These data can be found here: https://teach.mapnwea.org/impl/normsResearchStudy.pdf. NWEA's MAP Growth assessments are scored on a vertical scale which ranges from 100 to 350, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.1.2. Trends in Average Scale Scores on Curriculum Associates' i-Ready Mathematics and Reading Assessments



Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2020, spring 2021, and fall 2021. The Curriculum Associates comparison points represent median scale scores derived from the fall and spring distributions of scale scores unique to the group of Michigan students who completed an i-Ready benchmark assessment during the 2018-19 school year. Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.1.3. Trends in Average Scale Scores on Renaissance Learning's Star Math, Reading, and Early Literacy Assessments



Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Mathematics, Reading, or Early Literacy assessments in fall 2020, spring 2021, and fall 2021. We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Renaissance Learning's Star Math and Reading assessments are scored on a vertical scale which ranges from 0 to 1400, and the Early Literacy assessment is scored from 300 to 900. The y-axis scales in this figure differ across the math, reading, and literacy assessments to reflect differences in the scales and grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

PERCENT OF STUDENTS MAKING EXPECTED FALL-TO-FALL GROWTH

Background

While the previous analyses allow us to compare average scale scores across student groups and grades over the three pandemic test administrations to date, it remains unclear the extent to which growth outcomes, relative to expectations for a particular grade, subject, and initial achievement level, differed from pre-pandemic norms. Given findings both in Michigan (shown in the previous analyses as well as in prior reports) and other states that student learning has, on average, been slower than in non-pandemic school years, we may expect to see a large proportion of students making less than expected growth between fall 2020 and fall 2021. Moreover, the prominence of remote testing in at least the fall of the 2020-21 school year paired with some "at-home advantage" implies that our analyses may show a substantial proportion of students who initially received high scale scores making less than expected growth across the sample period, especially in early grade levels.

To understand the proportion of Michigan students that made expected growth between fall 2020 and fall 2021 relative to pre-pandemic norms, we examine how Michigan students' benchmark assessment score growth from fall 2020 to fall 2021 for the fall-to-fall sample compares to pre-determined growth norms that each assessment provider established before the pandemic. We examine the distributions of Michigan students across three categories of growth outcomes between fall 2020 and fall 2021 relative to assessment providers' expectations of growth in a non-pandemic school year. In particular, we show the percentages of students who:

- made zero or negative growth (i.e., their scale scores either did not change at all or decreased from fall 2020 to fall 2021 [gray bars]);
- made less than expected growth (i.e., their scale scores increased, but by less than the pre-pandemic growth norm for students in their grade level with similar prior achievement scores [light green and blue bars]); or
- met or exceeded expected growth (i.e., the increases to their scale met or exceeded the growth norm [dark green and blue bars]).

Interpreting Results Figures

Figure 3.2.1 shows the percentages of students who made zero or negative growth, made less than expected growth, and met or exceeded expected growth between fall 2020 and fall 2021 on NWEA's MAP Growth mathematics and reading assessments, by

grade level. Figure 3.2.2 shows differences by fall 2020 performance for the same group of students (those in districts administering the MAP Growth assessment) but this time split by their percentile range of performance on the fall 2020 assessment. Figure 3.2.3 and Figure 3.2.4 combine the previous two analyses and show differences in the percentages of students who made zero or negative, made less than expected, and met or exceeded expected growth by both grade and fall 2020 performance levels. We show similar grade-level analyses for i-Ready and Star 360 districts in Figure 3.2.5 and Figure 3.2.7, respectively, and the corresponding analyses by fall 2020 performance level in Figure 3.2.6 and Figure 3.2.8. The growth patterns in districts that administered the i-Ready or Star 360 assessments are largely consistent with the MAP Growth analyses. We therefore provide grade-specific analyses by fall 2020 performance level for i-Ready and Star 360 districts in Appendix Figures A2.1 through A2.4.

For the MAP Growth and i-Ready assessments, a student's "expected growth" is the median growth across students with similar baseline scores from the pre-pandemic norming samples for each assessment. Thus, in a typical year, we would expect to see about 50% of students meet or exceed their "expected growth." We can use 50% as a comparison point to assess how Michigan students' growth outcomes in fall 2021 diverge from pre-pandemic expectations. In other words, if fewer than 50% of Michigan students reached their growth targets in fall 2021, this would indicate that Michigan students were less likely than students in the pre-pandemic norming sample to meet or exceed their expected growth. We note that the terminology of "expected growth" or "growth targets" used in this report differs in meaning from the way practitioners use similar terms, such as "growth goals," in the classroom. In classroom contexts, teachers likely set student growth goals or targets that represent what the teacher expects an individual student to achieve in a given period. This is different from the targets for "typical growth" that we use in this report, which indicate the median growth that students with similar prior scores achieved before the pandemic.

For the MAP Growth assessments, we can also use NWEA's conditional fall-to-fall growth distributions (Thum & Kuhfeld, 2020) to get an approximate sense of how many students would be expected to make zero or negative growth before the pandemic. As Table 3.2.1 shows, the expected percentages vary by grade, subject, and students' initial percentile ranks. For example, across all initial achievement levels, we would expect fewer than 1% of kindergarteners to make zero or negative growth in math, whereas for 1st-grade reading, about 5% of students who started at the 10th percentile are expected to make zero or negative growth, compared to less than 2% of those who started at the 90th percentile. These expected percentages increase across grade levels and tend to be larger for reading than math.

Table 3.2.1. Percent of Students Expected to Make Zero or Negative Growth on NWEA MAP Growth — Pre-Pandemic Norms									
Subject	Fall 2020 Percentile	Fall 2020 Grade Level							
		K	1st	2nd	3rd	4th	5th	6th	7th
Mathematics	10th	0.0%	1.8%	1.5%	2.4%	5.7%	21.9%	13.8%	19.5%
	30th	0.1%	1.9%	1.8%	2.7%	5.3%	20.0%	13.0%	17.0%
	50th	0.1%	1.9%	2.0%	3.0%	5.0%	18.8%	12.4%	15.4%
	70th	0.3%	1.9%	2.2%	3.3%	4.8%	17.5%	11.9%	13.9%
	90th	0.6%	1.9%	2.6%	3.7%	4.4%	15.9%	11.2%	11.9%
Reading	10th	0.2%	5.0%	5.2%	6.8%	9.5%	17.8%	18.2%	19.4%
	30th	0.4%	3.7%	4.6%	8.1%	11.7%	19.8%	22.2%	23.3%
	50th	0.6%	3.0%	4.2%	9.0%	13.4%	21.4%	25.3%	26.2%
	70th	1.1%	2.4%	3.8%	10.1%	15.2%	23.0%	28.6%	29.3%
	90th	2.1%	1.7%	3.3%	11.8%	18.2%	25.3%	33.7%	34.0%

Notes: The percentages in this table represent the percentile ranks corresponding to zero fall-to-fall growth on NWEA's conditional growth distribution (Thum & Kuhfeld, 2020) for students with initial scores at the 10th, 30th, 50th, 70th, or 90th percentiles.

Results

Although, as we noted earlier in this section, the "average" student demonstrated less-than-expected growth from fall 2020 to fall 2021, we find that this was not a universal experience across Michigan students. On average across grades, subjects, and assessment providers, we find that about 40% of Michigan students met or exceeded their growth targets, and about a quarter did not demonstrate *any* growth from fall 2020 to fall 2021 (i.e., their scale scores decreased or did not change). The proportion that met or exceeded their targets is smaller than would be expected in a typical pre-pandemic school year, and the proportion that made zero or negative growth is much larger than would have been expected.

We also find that, as expected, students with lower baseline achievement were more likely to reach their growth targets, while those with the highest baseline scores were more likely to make zero or negative growth. However, this pattern is driven to some extent by grade level, pointing to the "at-home advantage" for early elementary students in fall 2020. For middle school students, in many cases we find evidence of bifurcated growth, whereby students were more likely to make zero or negative growth or to meet or exceed their growth targets than they were to make less than

expected growth. These patterns underscore how important it is to examine not only *average* achievement outcomes, but to also understand the extent to which these outcomes varied across Michigan students.

Figure 3.2.1 shows that, across all grade levels and subjects, fewer students met or exceeded their growth targets and more students made zero or negative growth than would have been expected before the pandemic. Figure 3.2.1 also highlights clear differences across grades in the percentages of students who made zero or negative, partial, and full expected growth between fall 2020 and fall 2021 on the MAP Growth mathematics and reading assessments. In particular, mathematics and reading outcomes for kindergarten students look considerably different compared to those for other grade levels. Across both subjects, kindergarteners were by far the most likely to make less than expected growth (66% and 58% in mathematics and reading, respectively). Rates of partial expected growth were also high for other elementary grades, at around 40%. Conversely, kindergarteners were the least likely to meet or exceed expected growth (24% and 29%). This is likely due to the "at-home advantage" that NWEA identified for some early grade-level students. If remote testing in fall 2020 led to inflated scale scores for that administration period, maintaining a similar level of achievement would be difficult once in-person testing became more prominent in the fall 2021 semester.

Fewer older students made less than expected growth between fall 2020 and fall 2021, decreasing in the middle school years such that 29% and 16% of 7th-grade students made less than expected growth in mathematics and reading, respectively, between the fall of 2020 and the fall of 2021. However, relatively high proportions of middle school students made zero or negative growth, with 32% and 41% of 7th graders making zero or negative growth between fall 2020 and fall 2021 in math and reading, respectively. This is markedly higher than what we might expect in a typical pre-pandemic school year; Table 3.2.1 shows that about 15% and 25% of 7th graders were expected to make zero or negative growth in math and reading, respectively, prior to the pandemic. This finding is notable and deserves reiterating: over a quarter and nearly a third of middle school students made no or negative growth on the MAP Growth mathematics assessment and a third or more middle school students made no or negative growth on the reading assessment between fall of 2020 and fall of 2021.

The proportion of students meeting or exceeding expectations increased over grade levels such that, in mathematics, over 45% of 3rd and 4th graders and approximately 40% of 5th through 7th graders met or exceeded expectations, and just over 40% of students across 2nd through 7th grade met or exceeded expectations in reading.

Students with lower initial mathematics and reading scores were the most likely to meet or exceed expected growth while students with higher initial scores were the most likely to make zero or negative growth. These trends are as expected, but the magnitude of students making zero or negative growth at each initial score percentile is substantially greater than what is predicted in a typical pre-pandemic year. Figure 3.2.2 shows that approximately half of students with the lowest initial fall 2020 mathematics and reading scores (45% and 54%) met or exceeded expected growth over the sample period. Conversely, those with fall 2020 mathematics and reading scores in the 81st to 99th percentiles were considerably more likely to make zero or negative growth between fall 2020 and fall 2021 (31% and 41%). This is not likely due to a "ceiling effect" (which is when students start the year at such a high level that there is, in essence, nowhere to grow), as each growth target represents the median change in scale scores for students with the same initial percentiles in the pre-pandemic norming sample. Thus, by definition, 50% of students in the norming sample *with the same initial percentile rank* reached these targets. Moreover, although expected percentages of students making zero or negative growth from before the pandemic (shown in Table 3.2.1) increase across initial percentile ranks for some grades in some subjects, these expected differences are far smaller than the differences we observe in the growth outcomes shown in Figure 3.2.2.

To better understand what's driving these patterns, Figure 3.2.3 and Figure 3.2.4 provide more detail by grade *and* initial performance quintile. Although for the most part, growth patterns across initial performance quintiles for each grade level follow the same average trend shown in Figure 3.2.2, these patterns are particularly stark for students in early elementary grades. For example, among students in the 4th-grade cohort, those with initial math scores in the highest quintile were nearly twice as likely as those in the second-highest quintile to make zero or negative growth (about 24% and 13%, respectively). For students in the kindergarten cohort, on the other hand, those in the highest quintile were nearly *five times* as likely to make zero or negative growth as those in the second-highest quintile (about 30% and 6%, respectively). Across subjects and assessment providers, we find that students in early elementary grade levels are unlikely to make zero or negative growth *unless* their initial scores were in the top quintile. This could indicate that the fall 2020 scores of many of the K-2 students who completed their tests remotely *overstate* their true achievement at that time due to the "at-home advantage;" this would mean that these students' fall 2021 scores are the same as or lower than their fall 2020 scores because these scores better reflect their true skill level, and not because the students did not experience any growth over the year. We find similar patterns across subjects and assessment providers.

Figure 3.2.5 through Figure 3.2.8 repeat a portion of these analyses for students in i-Ready and Star 360 districts and the analyses examining differences by grade and

initial performance level are presented in Appendix Figures A2.1 through A2.4. We see many of the same trends as we did for students in districts that administer the MAP Growth assessment. Here we highlight the few differences unique to students who completed these two assessments.

Figure 3.2.5 shows that outcomes for kindergarten and 1st-grade students in i-Ready districts look more like their peers in higher grade levels than did these early learners in districts administering the MAP Growth assessments. In particular, in math, the most common outcome for students across all elementary grades was to make less than expected growth, with approximately one-quarter to one-third of students making zero or negative growth between fall semesters. By middle school, approximately equal proportions (just over a third) of students were meeting or exceeding expected growth and making negative or zero growth. There was a similar, though more accentuated, bifurcation of outcomes in reading, this time for upper elementary and middle school grades. Approximately half of students in the 3rd- through 7th-grade fall 2020 cohorts met or exceeded expected growth, and over a third of the 6th- and 7th-grade cohorts made zero or negative growth.

When considering the proportions of middle school students who made less than expected, zero, or negative growth, i-Ready districts perform similarly to MAP Growth districts in math; approximately two-thirds of middle school students made less than expected, zero, or negative growth in math between fall 2020 and fall 2021. In reading, on the other hand, about 50% of middle school students in i-Ready districts made less than expected, zero, or negative growth in reading, suggesting that these students were about as likely to reach their reading growth targets as students with similar initial scores before the pandemic.

As was the case for the MAP Growth assessments, students with the highest initial i-Ready math and reading scores were substantially more likely than their peers to make zero or negative growth. As Figure 3.2.6 shows, across the lower four initial performance levels, 23% to 28% of students made zero or negative growth in math, compared to 45% of students in the highest initial performance level. Similarly, 23% to 29% of students in the bottom four initial performance levels made zero or negative growth in reading, compared to 43% of students in the highest prior performance level.

Unlike in the MAP Growth districts, students in i-Ready districts with lower initial mathematics and reading scores were *not* the most likely to meet or exceed expected growth in mathematics or reading. Figure 3.2.6 shows that approximately one-third of students across all five proficiency levels met or exceeded expected growth in mathematics, while slightly less than half of students across all performance levels

met or exceeded expected growth in reading. However, it is important to note that the fall placement levels from Curriculum Associates are defined in terms of grade-level placement relative to the student's assessed grade level. Thus, Curriculum Associates does not define scale score thresholds for the lowest two performance levels ("3+ Grades Below" and "2 Grades Below" grade level) for some early grade levels. Given the many differences we've noted about results for students in early elementary grade levels compared to upper elementary and middle school students, the fact that not all early elementary grade levels are represented in the lowest two performance categories may explain why we don't observe all the same patterns that we did for NWEA. In fact, the percentages of students who reached their growth targets decrease consistently across the three initial performance levels ("1 Grade Below," "On Grade [Early]," and "On Grade [Mid] or Above") for which all grade levels are represented.

The final two figures show the percentages of students who made zero or negative, partial, and full expected growth between fall 2020 and fall 2021 on Renaissance Learning's Star Math, Reading, and Early Literacy assessments by grade level (Figure 3.2.7) and percentile range of fall 2020 performance (Figure 3.2.8). The trends shown in Figure 3.2.7 generally resemble those previously discussed for students in MAP Growth districts, with one major exception. Across most grade levels and both subjects, students were the most likely to meet or exceed expected growth compared to other growth outcomes; well over one-third of students met or exceeded expected mathematics and reading growth over the time period under study. In particular, over half of the fall 2020 2nd- and 4th-grade cohorts met or exceeded expected growth in math and over half of the 3rd-grade cohort did so in reading. At the same time, relatively few students made less than zero or negative growth in either subject across grades; only in the 6th- and 7th-grade cohorts did more than a quarter of students make zero or negative growth.

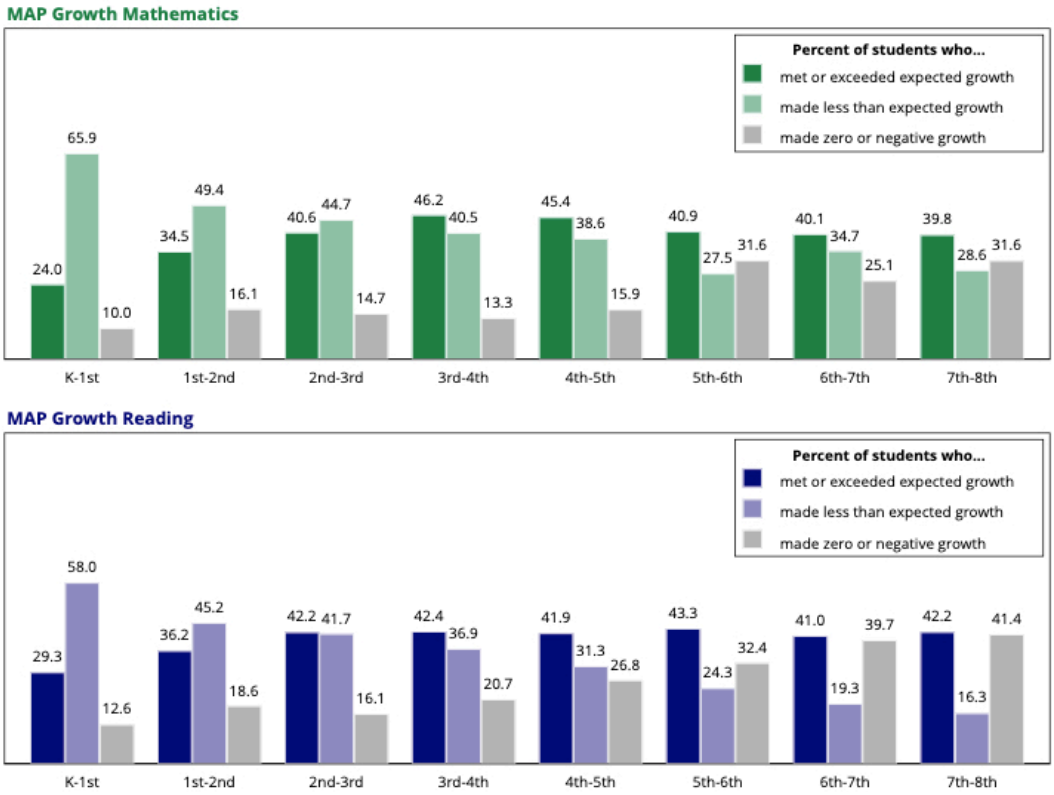
Figure 3.2.8 shows that, similar to NWEA districts, students in the lowest quintile on the Star 360 assessments were far more likely to meet or exceed expected growth, and by contrast, students in the highest quintile were most likely to make zero or negative growth. This is the case across both subjects. As with NWEA, we find that this pattern is strongest among students in early elementary grade levels, which may indicate that these students' fall 2020 quintiles do not accurately reflect their true skill levels at that time, but rather that their scores were inflated due to an "at-home advantage" (see Appendix Figures A2.3 and A2.4).

Appendix Figures A2.1 through A2.4 examine differences by grade and initial performance level for students in i-Ready and Star 360 districts. They show the mathematics and reading outcome trends for students in i-Ready districts reflect the same average trends across grades and initial performance levels as we observed for

MAP Growth districts. Mathematics and reading outcome trends in Star 360 districts were also generally similar, however, 7th- and 8th-grade students in the highest initial achievement quintile were considerably more likely to make zero or negative growth compared to students in MAP Growth districts.

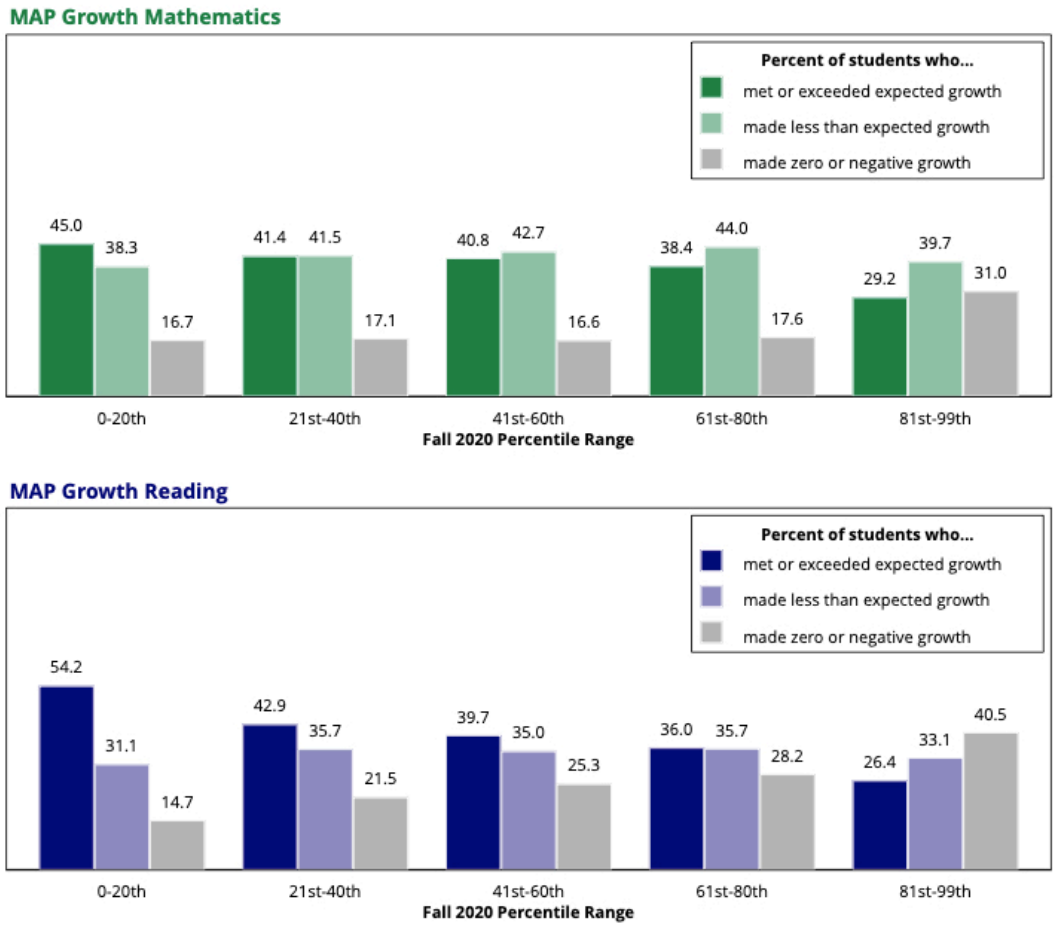
Overall, we find that students in lower grade levels typically were most likely to make less than expected growth in mathematics and reading between fall 2020 and fall 2021. Students in middle school grade levels were the most likely to make zero or negative growth in both subjects. One contributing factor is that scores tend to change less from year to year for students in higher grade levels (this is evident in the slopes of and relative distances between the norm lines for each grade level in Figure 3.1.1 through Figure 3.1.3), so growth targets tend to be smaller and fewer students are expected to experience increases in scale scores, compared to younger students (as is evident in the increasing percentages across grade levels in Table 3.2.1). For initial achievement levels, those scoring at the bottom of the distribution in fall 2020 were most likely to meet or exceed expectations in mathematics and reading by spring 2021, while initially high-performing students were the most likely to make zero or negative growth. Early elementary students in the top performance level are far more likely than any other students to make zero or negative growth. This is likely because many of the students who scored in the highest level in fall 2020 did so as a result of some "at-home advantage," and this was not a true reflection of their fall 2020 achievement levels.

Figure 3.2.1. Percent of Students Making Expected Fall-to-Fall Growth on NWEA's MAP Growth Mathematics and Reading Assessments by Grade



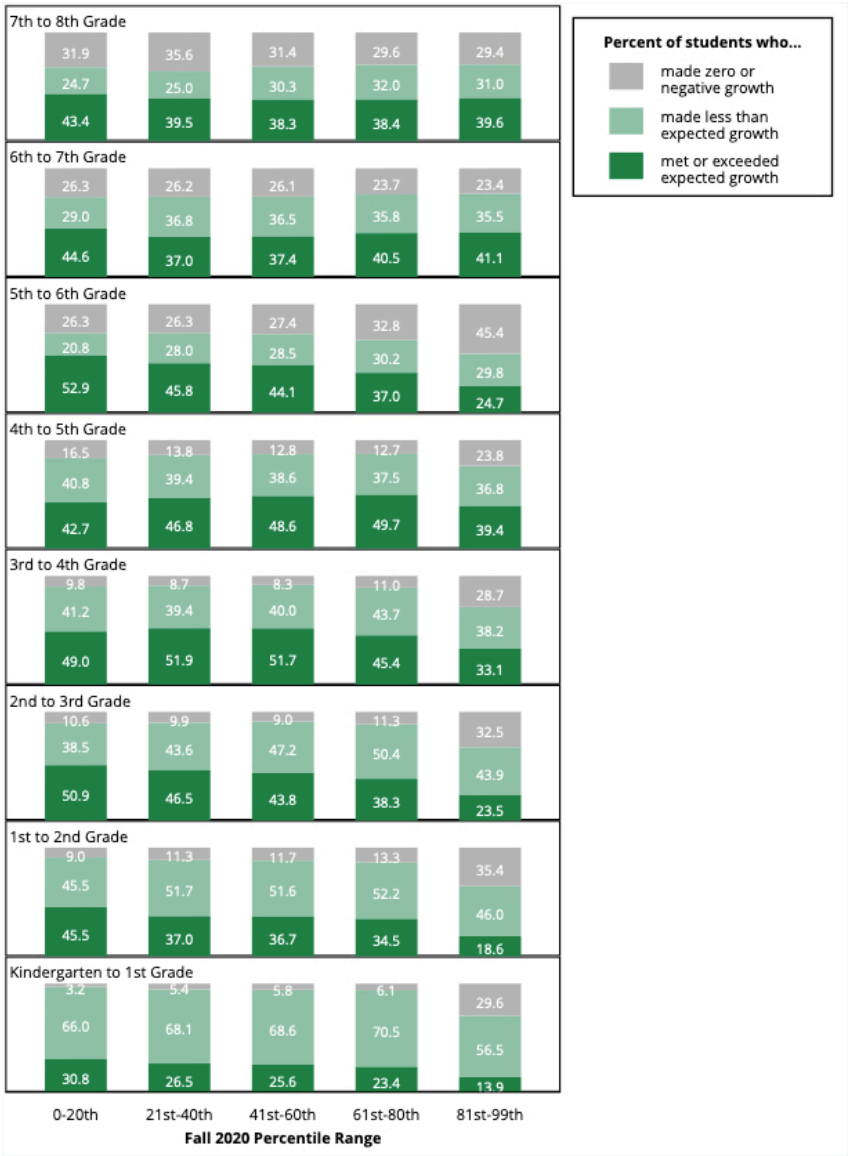
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For students who completed NWEA's MAP Growth assessments, we use as a growth norm the 50th percentile of the fall-to-fall conditional growth distribution for students with the same initial (fall 2020) percentile rank (Thum & Kuhfeld, 2020). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.2. Percent of Students Making Expected Fall-to-Fall Growth on NWEA's MAP Growth Mathematics and Reading Assessments by Fall 2020 Performance Level



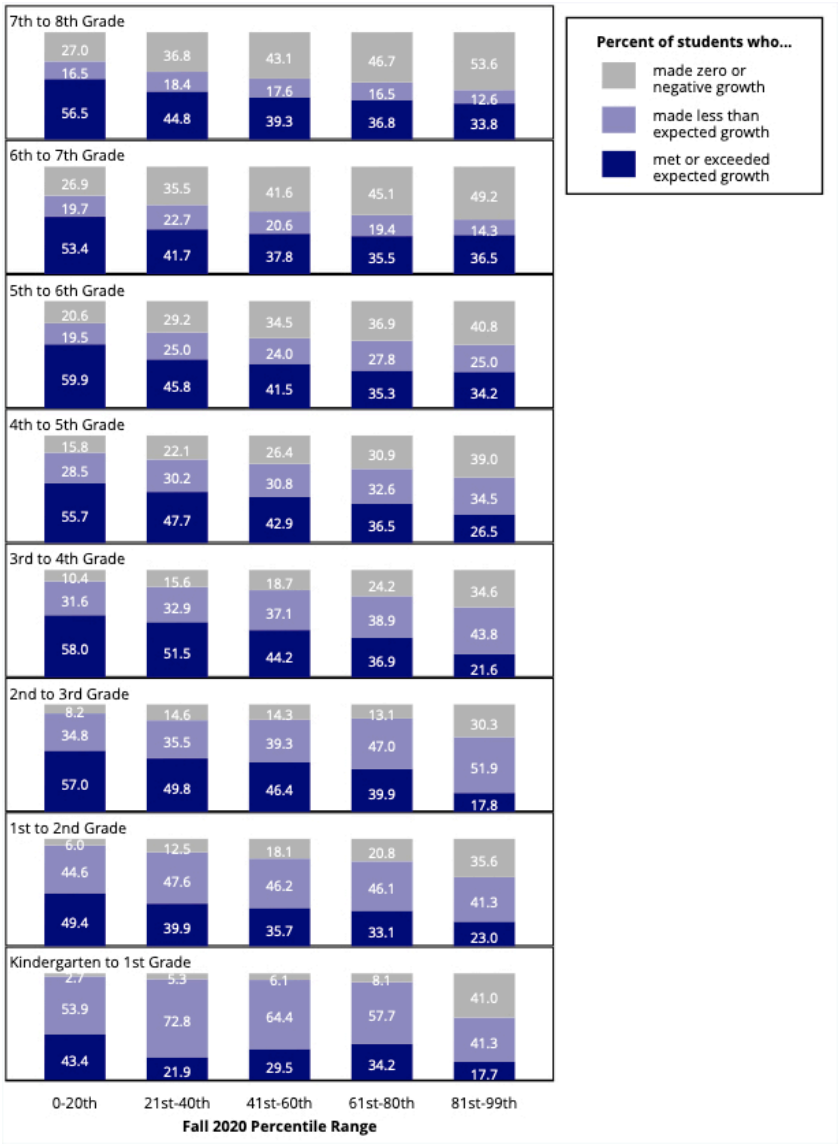
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For students who completed NWEA's MAP Growth assessments, we use as a growth norm the 50th percentile of the fall-to-fall conditional growth distribution for students with the same initial (fall 2020) percentile rank (Thum & Kuhfeld, 2020). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.3. Percent of Students Making Expected Fall-to-Fall Growth on NWEA's MAP Growth Mathematics Assessment by Grade and Fall 2020 Performance Level



Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For students who completed NWEA's MAP Growth assessments, we use as a growth norm the 50th percentile of the fall-to-fall conditional growth distribution for students with the same initial (fall 2020) percentile rank (Thum & Kuhfeld, 2020). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

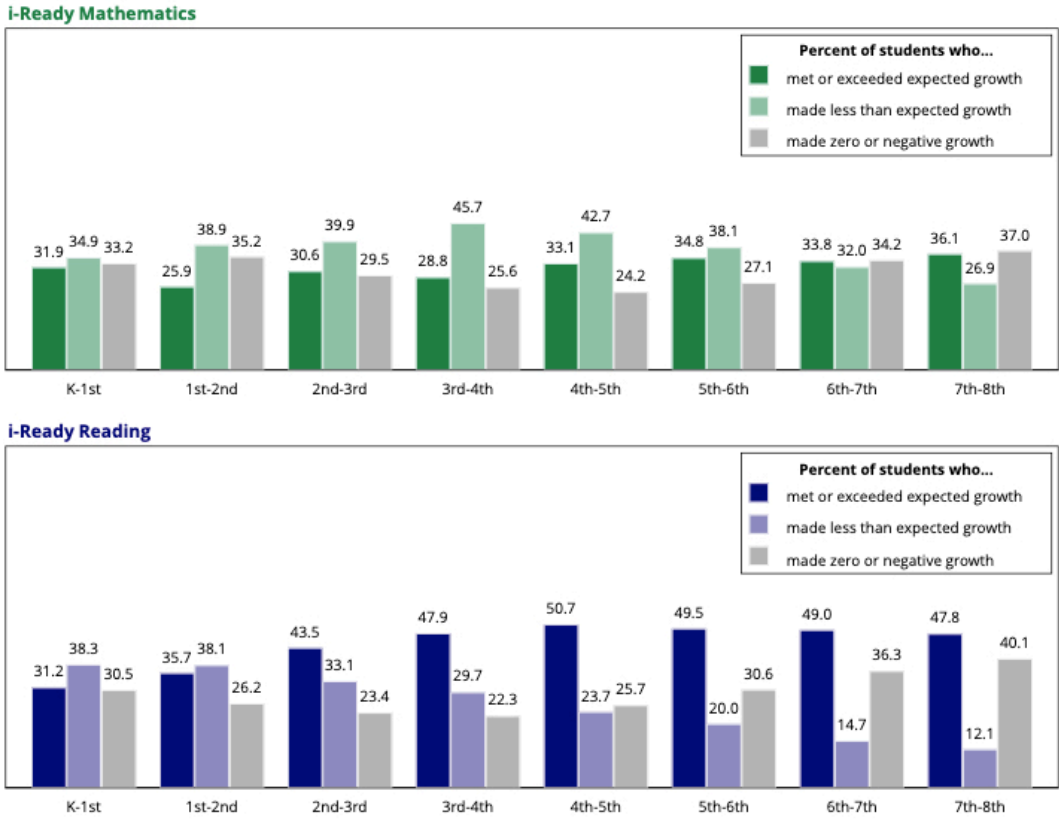
Figure 3.2.4. Percent of Students Making Expected Fall-to-Fall Growth on NWEA's MAP Growth Reading Assessment by Grade and Fall 2020 Performance Level



Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For students who completed NWEA's MAP Growth assessments, we use as a growth norm the 50th percentile of the fall-to-fall conditional growth distribution for students with the same initial (fall 2020) percentile rank (Thum & Kuhfeld, 2020). Source:

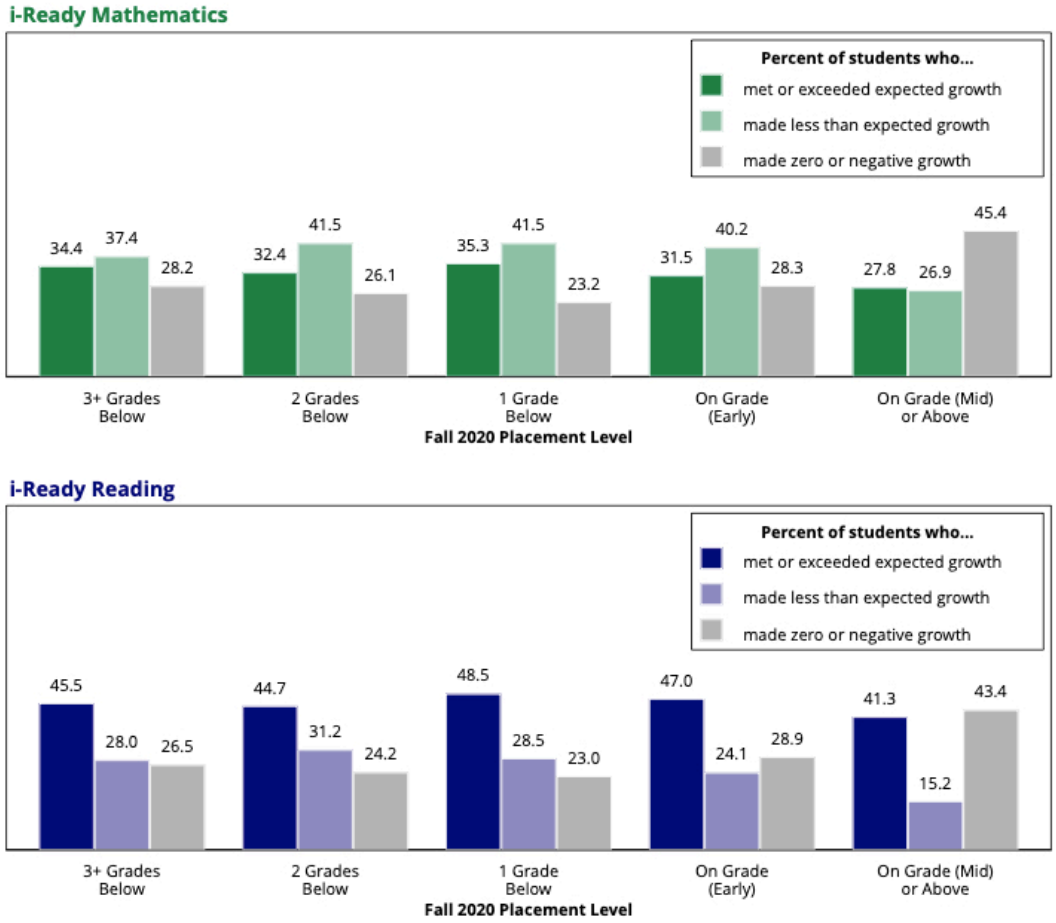
School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.5. Percent of Students Making Expected Fall-to-Fall Growth on Curriculum Associates' i-Ready Mathematics and Reading Assessments by Grade



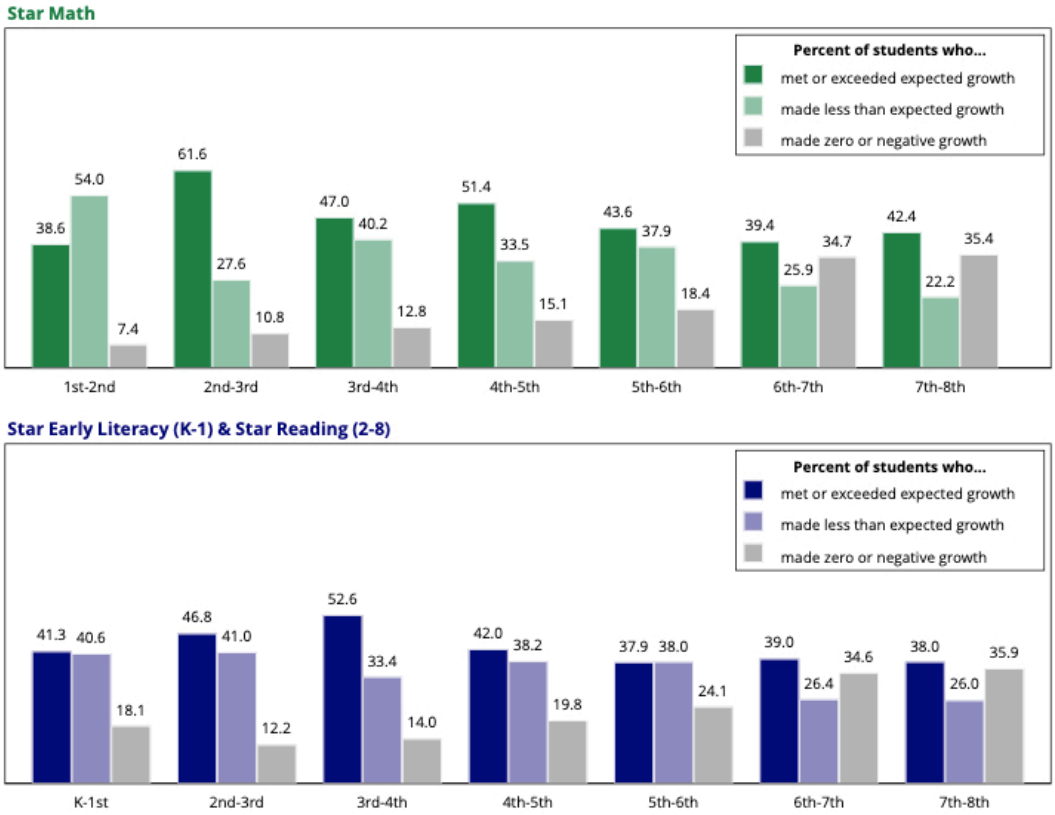
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associate's i-Ready assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). However, we note that these targets are designed to measure growth from fall to spring and may not perfectly represent growth norms from fall to fall. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.6. Percent of Students Making Expected Fall-to-Fall Growth on Curriculum Associates' i-Ready Mathematics and Reading Assessments by Fall 2020 Placement Level



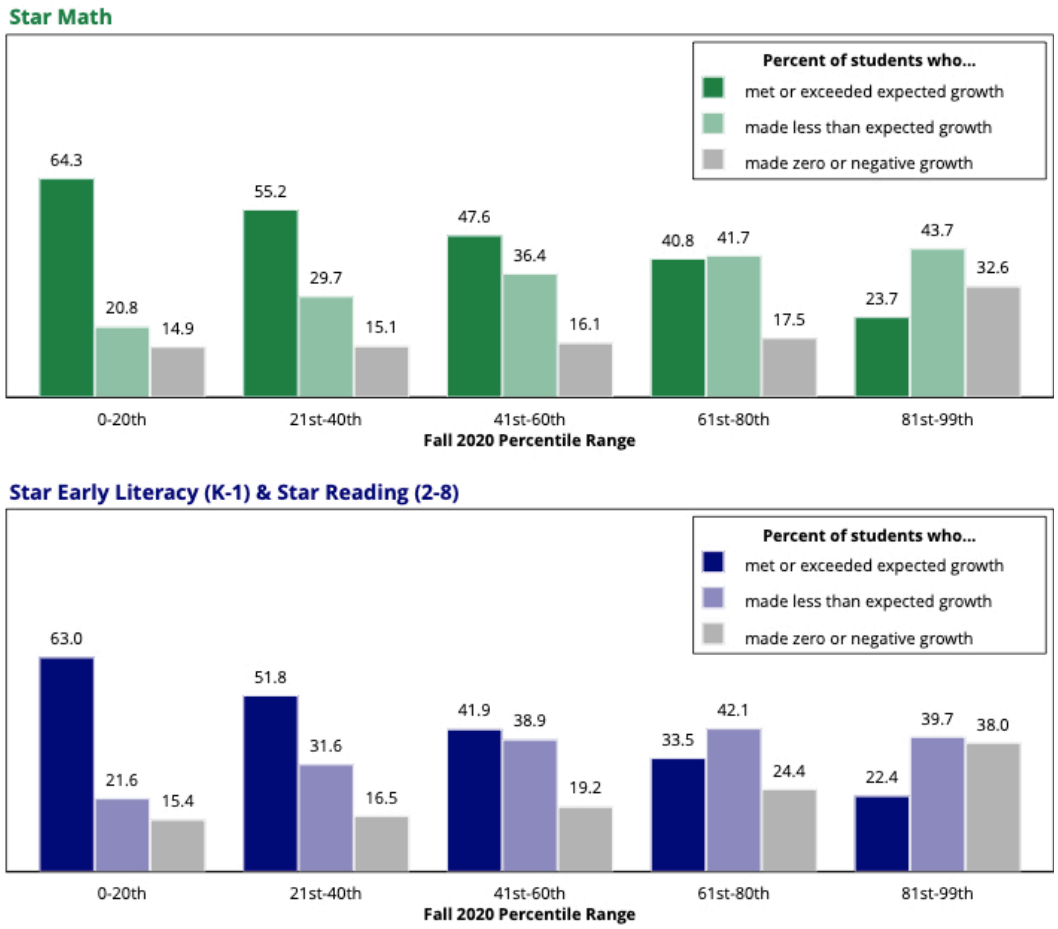
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associate's i-Ready assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). However, we note that these targets are designed to measure growth from fall to spring and may not perfectly represent growth norms from fall to fall. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.7. Percent of Students Making Expected Fall-to-Fall Growth on Renaissance Learning's Star Math, Reading, and Early Literacy Assessments by Grade



Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Math, Reading, or Early Literacy assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For the Star 360 we use pre-pandemic scale score distributions to identify "expected growth" as the change in scale scores necessary for a student to maintain the same percentile rank as they did in the previous year (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure 3.2.8. Percent of Students Making Expected Fall-to-Fall Growth on Renaissance Learning's Star Math, Reading, and Early Literacy Assessments by Fall 2020 Performance Level



Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Math, Reading, or Early Literacy assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject, and students' initial achievement level. For the Star 360 we use pre-pandemic scale score distributions to identify "expected growth" as the change in scale scores necessary for a student to maintain the same percentile rank as they did in the previous year (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

STANDARDIZED ACHIEVEMENT GAPS

Figures in the previous two sub-sections showed changes in average scale scores between fall 2020 and fall 2021 and summarized how these trajectories translate into "expected growth" over the same period. In this sub-section, we examine achievement gaps between subgroups of students. We first compare these to historical gaps for Michigan assessments before the pandemic, and then compare gaps for benchmark assessments across the fall 2020, spring 2021, and fall 2021 testing periods.

To help us understand whether achievement gaps in the benchmark assessment data differ from pre-pandemic achievement gaps for Michigan students on other assessments, we show standardized gaps from Michigan's 2018-19 summative assessments (M-STEP and PSAT 8/9) in Table 3.3.1. For instance, the top row in the table shows that Black students scored between 0.90 and 0.99 standard deviations below their White peers, on average, on the 2018-19 end-of-year summative math exams.

Table 3.3.1. 2018-19 M-STEP and PSAT Standardized Achievement Gaps

Achievement Gap	Subject	Grade Level					
		3rd	4th	5th	6th	7th	8th
Black – White	Math	-0.91	-0.95	-0.95	-0.99	-0.94	-0.90
	ELA	-0.87	-0.84	-0.81	-0.78	-0.75	-0.77
Latino/a/x – White	Math	-0.45	-0.48	-0.48	-0.50	-0.48	-0.50
	ELA	-0.43	-0.43	-0.42	-0.42	-0.39	-0.43
Asian – White	Math	0.55	0.57	0.67	0.70	0.68	0.83
	ELA	0.32	0.37	0.46	0.49	0.50	0.57
Economically Disadvantaged– Not Economically Disadvantaged	Math	-0.83	-0.87	-0.88	-0.89	-0.87	-0.80
	ELA	-0.80	-0.81	-0.80	-0.78	-0.78	-0.72
Male – Female	Math	0.11	0.11	0.09	0.02	0.00	-0.04
	ELA	-0.15	-0.16	-0.20	-0.23	-0.28	-0.27
Special Education – General Education	Math	-0.86	-0.91	-0.98	-1.16	-1.13	-1.02
	ELA	-0.69	-0.80	-0.90	-0.95	-1.00	-0.90

Source: This table was created using publicly available subgroup means and standard deviations obtained from: https://www.mischooldata.org/grades-3-8-state-testing-includes-psat-data-performance/

We use these as comparison points to help interpret gaps in the benchmark assessment data within the context of Michigan's longstanding achievement gaps that pre-date the pandemic. However, we note that these historical gaps are based on the statewide populations of students who took each of these assessments. Discrepancies between these gaps and the ones we observe in the benchmark assessment data may reflect differences between the populations of students who participated in benchmark assessments from each provider and the populations of Michigan

students who participated in the 2018-19 summative assessments, as well as differences between the assessments themselves.

Fall 2021 Achievement Gaps

Figure 3.3.1 and Figure 3.3.2 show standardized math and reading achievement gaps in fall 2021 benchmark assessment scores for students of Black, Latino/a/x, Asian, and "other" races/ethnicities, relative to White students. In these figures, we convert the average scores for each subgroup to percentile ranks relative to the distribution of scores for the largest race/ethnicity subgroup (White students). A percentile rank of 50 indicates that the average score for a subgroup is equal to the average score for the reference group, while percentile ranks above 50 indicate that students in a given subgroup scored higher, on average, than students in the reference group, and percentile ranks below 50 indicate that students in the subgroup scored lower than students in the reference group. Figures 3.3.3 and 3.3.4 show achievement gaps for economically disadvantaged students relative to students who are not economically disadvantaged, for special education students relative to general education students, and for male students relative to female students.

We calculated these gaps based on the full fall 2021 sample. In other words, the results in these figures represent all students, districts, and assessment providers that contributed fall 2021 benchmark assessment data through the MDH. Later in this sub-section, we delve deeper into each of these achievement gaps and examine how they've changed over time across a constant cohort of students, but we first examine general patterns in the fall 2021 gaps and compare these patterns across benchmark assessments. As pre-pandemic comparison points, we also show overall gaps between the same subgroups of students on Michigan's 2018-19 summative assessments.

The green, navy, bright blue, and purple bars in Figures 3.3.1 and 3.3.2 show the percentile ranks corresponding to the average scores for Black students, Latino/a/x students, Asian students, and other non-White students, respectively, averaged across early elementary (K-2), upper elementary (3-5), and middle school (6-8) grade ranges. The grey bars represent the average score for White students; because this is the reference group, these bars are always exactly equal to 50. In Figure 3.3.3 and Figure 3.3.4, the darker shades of green, navy, and bright blue represent each of the focal subgroups (economically disadvantaged students, special education students, and male students, respectively), and the lighter shades of each color represent the reference group for each comparison (students who are not economically disadvantaged, general education students, and female students). Once again, percentile ranks for each of the reference groups are always equal to 50. We note that there are several student race/ethnicity subgroups missing from the figure for the K-

2s and ICAs, as there were too few students in these subgroups to calculate achievement gaps or relative percentiles. We also note that there are no M-STEP/PSAT bars for K-2 students, as students begin taking these assessments in grade 3.

Across all assessments, subjects, and grade ranges, relative percentiles for Black, Latino/a/x, economically disadvantaged, and special education students are all lower than 50, indicating that, on average, students in these subgroups consistently score lower than students in their respective reference groups. The relative percentiles for Asian students, on the other hand, are always above 50, indicating that Asian students consistently score higher than White students, on average. Racial and socioeconomic achievement gaps are consistently large across all grade ranges, while gaps between special education and general education students are more pronounced for older students. In general, mathematics achievement gaps and reading achievement gaps followed the same patterns, but reading gaps tended to be slightly smaller in magnitude. The one exception to this is that male students tend to score slightly higher than female students in math, but slightly lower than female students in reading.

Notably, across all gaps and in both subjects, patterns in relative percentiles across subgroups on Michigan's 2018-19 summative assessments are generally very consistent with those for the fall 2021 benchmark assessments. This indicates that, on average, most of the achievement gaps for upper elementary and middle school students on the fall 2021 benchmark assessments are very similar in size to gaps from Michigan's 2018-19 summative assessments.

In the remainder of this sub-section, we discuss standardized gaps in benchmark assessment scores among students in the fall-spring-fall sample. We focus on this sample so we can examine how gaps among the exact same groups of students have changed across the three assessment periods. This sample does not include the cohort of students who began kindergarten in fall 2021, districts that did not administer an approved benchmark assessment in 2020-21, students who participated in MDE's K-2 assessments or the Smarter Balanced ICAs, or any other students who do not have scores from comparable assessments for all three testing periods. The results that follow examine grade-specific standardized achievement gaps between districts that offered different modes of instruction throughout the 2020-21 school year. We also more closely examine gaps by race/ethnicity, economically disadvantaged status, and special education status to help us understand how these gaps vary across individual grade levels and how they have changed over time. Similar to the analysis of average scale score trends, we calculate and present standardized achievement gaps separately for each assessment provider, subject, and grade level over time (i.e., fall 2020 gaps are shown in the lightest green and blue bars, spring 2021 gaps in the middle shades, and fall 2021 gaps are

represented by the darkest green and blue bars). Although not discussed in this section, we also present mathematics and reading gaps between male and female students in Appendix Figures A3.11 and A3.12.

Instructional Modality

Figure 3.3.5 through 3.3.8 show standardized MAP Growth scale score gaps between districts that offered different instructional modalities in the 2020-21 school year by semester, grade, and subject. We compare districts that offered only in-person, only remote, or only hybrid instruction during both the fall and spring 2020-21 benchmark administration periods (i.e., "in-person all year," "remote all year," and "hybrid all year," respectively). We also provide results for districts that offered different instructional modalities in fall 2020 than in spring 2021: Districts that offered in-person instruction during one administration period and hybrid or remote instruction during the other period (i.e., "in-person part-year"), and districts that offered hybrid instruction during one administration period and remote instruction during the other period (i.e., "hybrid part-year").

We present results separately for each instructional modality, and students in "in-person all year" districts are the reference category for each gap. This means that negative gaps would indicate that students in either "remote all year," "in-person part-year," or "hybrid part-year" performed worse than their peers in districts offering in-person instruction during both benchmark administration periods throughout the year. We focus only on results for the MAP Growth assessments, and do not discuss gaps across instructional modalities for any of the other assessments, as there were fewer than 10 districts in each modality category (except for "in-person all year") that used the i-Ready and Star 360 assessments. We provide equivalent figures for the i-Ready and Star 360 assessments in the appendix (see Appendix Figures A3.1 through and A3.8), however, we urge readers to interpret them with caution, as very few districts are represented in each modality category.

For the instructional modality results discussed below, there are considerable differences in the number of students and districts included in achievement gap calculations across instructional modalities. In total, 451 districts and 293,496 students were included in the mathematics or reading gap calculations for MAP Growth districts. Fifty-seven percent of these districts offered fully in-person instruction in both benchmark assessment periods (48% of students), approximately a fifth offered fully in-person instruction during only one assessment period (21% of students), and no more than 10% of MAP Growth districts were classified into each of the remaining three instructional modality classifications (13%, 4%, and 10% of students for hybrid, remote, and "hybrid part-year," respectively).

The mathematics and reading gaps between "remote all year" and "in-person all year" districts provide strong evidence of an "at-home advantage" for students in early grade levels (see Figure 3.3.5). Specifically, for students in "remote all year" districts, kindergarten and 1st-grade gaps were between 0.65 and 1.12 standard deviations in fall 2020; this means that on average, students in "remote all year" districts scored between the 74th and 87th percentiles for "in-person all year" districts. By fall 2021, kindergarten and first grade cohort gaps were less than 25% of a standard deviation and negative; this means that early elementary students in "remote all year" districts, on average, scored between the 40th and 50th percentiles for "in-person all year" districts.

For students in 2nd through 7th grade, some spring 2021 and fall 2021 mathematics and reading gaps are considerably larger compared to the same gaps for kindergarten and 1st-grade students. For instance, 2nd- through 5th-grade mathematics are at least half a standard deviation in spring 2021. Additionally, across nearly all grade levels, mathematics and reading gaps increased dramatically between fall 2020 and spring 2021, then decreased over the summer months.

Gaps for "in-person part-year" districts (shown in Figure 3.3.6) varied by assessment provider, where mathematics and reading gaps were typically small. The gaps also tended to increase across all three semesters between fall 2020 and fall 2021. This suggests that students who received less in-person instruction in 2020-21 fell further behind over time. For "hybrid part-year" districts (see Figure 3.3.7), mathematics and reading gaps were generally larger than the gaps for "in-person part-year" districts; this suggests that students who did not receive fully in-person instruction at all during 2020-21 faced greater effects than those who spent part of the year fully in-person. Across nearly all grade levels, these gaps increased between fall 2020 and spring 2021, reflecting the continued effect of missed in-person instruction. However, the gaps decreased between spring 2021 and fall 2021, as the return to in-person instruction for most students began to alleviate the effects of disrupted instruction throughout the previous year. Figure 3.3.8 shows standardized mathematics and reading achievement gaps between students in "hybrid all year" districts and those in "in-person all year" districts. These gaps tended to be small and negative.

Overall, these results suggest that students who had access to in-person instruction during both 2020-21 testing periods performed better on the MAP Growth benchmark assessments. Additionally, mathematics and reading gaps for students in districts that were fully in-person for part of the year or hybrid for part of the year were typically smaller in magnitude compared to gaps for students in "remote all year" districts, implying that access to a limited amount of in-person instruction was beneficial for student outcomes compared to fully remote instruction. Many of these gaps increased during the 2020-21 school year, suggesting that the effects of disrupted instruction

grew over time. However, decreases in gaps from spring 2021 to fall 2021 may indicate that students who received less in-person instruction in 2020-21 are starting to catch up to their peers now that most of their districts have resumed in-person instruction.

Race/Ethnicity

Figures 3.3.9 through 3.3.14 show standardized scale score gaps between students of different races/ethnicities by assessment provider, subject, grade, and time period. White students are the reference category for all outcome gaps (i.e., we estimate gaps between Black and White students, Latino/a/x and White students, and Asian and White students). We also present gaps between White students and students of ethnicities other than White, Black, Latino/a/x, or Asian (i.e., students reported as American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, or more than one race/ethnicity) in Appendix Figures A3.9 and A3.10. Overall, the following results show relatively stable or slightly increasing Black-White and Latino/a/x standardized mathematics and reading gaps in upper elementary and middle school grade levels, and larger increases in these gaps for lower elementary grade levels. Asian-White gaps, on the other hand, generally either remained stable or decreased slightly over time.

Table 3.3.1 shows that, before the pandemic, there was approximately a full standard deviation Black-White gap in math M-STEP/PSAT scores and over three-quarters of a standard deviation gap in ELA scores across all tested grade levels. However, Figures 3.3.9 and 3.3.10 show that, across all elementary grade levels, Black-White gaps in benchmark assessment scores are substantially smaller in fall 2020 than historical gaps. These gaps generally returned to pre-pandemic magnitudes by spring and fall 2021.

For example, the 3rd-grade Black-White mathematics gaps in MAP Growth districts increased from -0.61 in fall 2020 (indicating that Black students scored at about the 27th percentile for White students) to -1.05 and -1.02 standard deviations in spring 2021 and fall 2021 (approximately the 15th percentile in both semesters), respectively. These gaps are slightly larger than, but much closer to, the Black-White gap on the 2018-19 M-STEP for 3rd-grade mathematics (-0.91 standard deviations, meaning that Black students scored at about the 18th percentile for White students). We find similar results for districts that administered i-Ready assessments for both math and reading. These patterns may indicate that remote testing irregularities and the "at-home advantage" inflated scores for Black students to a greater extent than for White students, as districts that were fully remote in fall 2020 tended to serve larger shares of Black students (Hopkins, Kilbride, & Strunk, 2021), therefore resulting in unrealistically small achievement gaps. These increases in the Black-White gaps may at least in part reflect a return to at-school testing and more reliable data rather than worsening inequities.

Figures 3.3.11 and 3.3.12 show similar patterns for Latino/a/x-White achievement gaps, however, the relative size of mathematics and reading benchmark gaps compared to historical gaps differ across assessment providers. As seen in Table 3.3.1, Latino/a/x-White math and ELA gaps on 2018-19 summative assessments were approximately half of a standard deviation across grade levels. Fall 2021 mathematics and reading benchmark gaps in MAP Growth and Star 360 districts were equal to or less than historical gaps, while benchmark gaps in i-Ready districts tended to be larger. This may be due to differences in the demographic composition of i-Ready districts, compared to districts that used MAP Growth or Star 360 and to the state as a whole.

Across all three assessment providers, Asian-White gaps generally decreased slightly or remained stable over time (see Figures 3.3.13 and 3.3.14). As was the case with Asian-White gaps on 2018-19 summative assessments, these gaps are consistently larger in mathematics than in reading across all testing periods. For example, in MAP Growth districts, the 3rd-grade Asian-White gaps in fall 2020 were 0.94 standard deviations in mathematics (83rd percentile for 3rd-grade White students) and 0.55 in reading (71st percentile). By fall 2021, these same gaps decreased to 0.80 and 0.53 standard deviations in mathematics and reading, respectively (83rd and 70th percentiles).

For each of the Black-White, Latino/a/x-White, and Asian-White gaps discussed above, we are wary of placing too much weight on the fall 2020 gap analyses. Given the noise inherent in remote testing outcomes and the differential testing patterns described earlier and in EPIC's second report in this series, changes in achievement gaps between fall 2020 and fall 2021 may be at least partially driven by changes in students' testing environments as they transitioned back to in-person instruction. For example, Black and Latino/a/x students were consistently more likely to be educated using fully remote instruction throughout the 2020-21 school year compared to White students (see Hopkins, Kilbride, & Strunk, 2021). If these students were able to benefit from remote testing in fall 2020, initial benchmark assessment scores would likely be inflated, on average, artificially reducing achievement gaps between White and Black or Latino/a/x students. Thus, these data make clear that, by fall 2021, the persistent achievement gaps between White and Black, Latino/a/x, or Asian students were generally similar in size to statewide gaps on summative assessments prior to the pandemic.

Economically Disadvantaged Status

Figures 3.3.15 and 3.3.16 provide similar analyses of standardized gaps between students who are and are not economically disadvantaged. Students who are not economically disadvantaged are the reference category for these gaps, such that negative gaps indicate that economically disadvantaged students have lower average scores than their wealthier peers. Overall, mathematics and reading gaps between economically disadvantaged students and their more advantaged peers are negative

for every grade level and assessment provider. These gaps grew larger between fall 2020 and spring 2021/fall 2021 for students in the earliest grades. Since these trends are so consistent across subjects and assessment providers, it is important to note that economically disadvantaged students were more likely to be educated through remote instruction in fall 2020 (Hopkins, Kilbride, & Strunk, 2021). Therefore, these students may have had increased access to parental help or other aids in a remote setting, potentially biasing fall 2020 achievement upwards and understating fall 2020 achievement gaps in elementary grade levels.

By late elementary and middle school, the gaps were only slightly larger in spring and fall 2021 than they were in fall 2020, suggesting consistent—and still negative—gaps over the course of pandemic schooling. The magnitude of these gaps is also generally consistent with those shown in pre-pandemic M-STEP scores. Mathematics and reading gaps for students in MAP Growth and Star 360 districts were slightly smaller than similar gaps from the 2018-19 summative assessments, while the same gaps for students in i-Ready districts were typically larger than the historical gaps. For example, 5th-grade 2018-19 M-STEP math gaps between students who are and are not economically disadvantaged were 0.87 standard deviations. On the benchmark assessments, the fall 2021 mathematics gap for 5th-grade students in MAP Growth districts was -0.88 standard deviations (indicating that economically disadvantaged students scored, on average, at the 19th percentile for students who are not economically disadvantaged) and -1.05 standard deviations for similar students in i-Ready districts (15th percentile). Once again, this may be due to differences in the characteristics of the student population in i-Ready districts compared to the rest of the state.

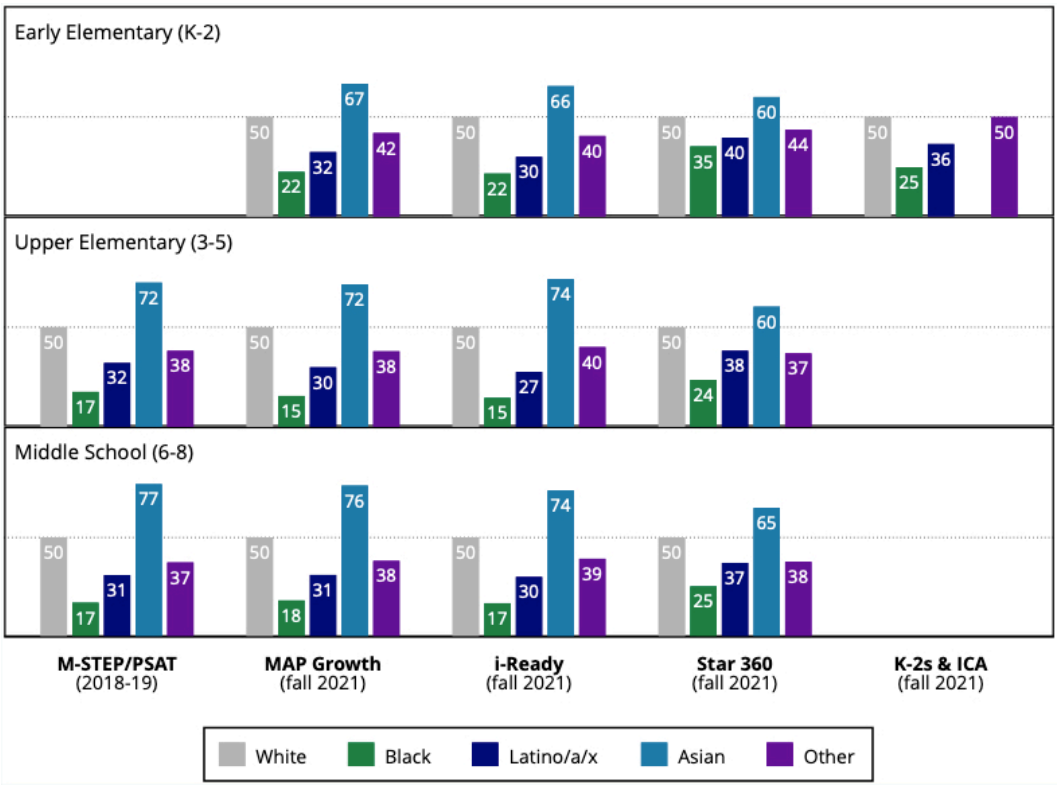
Across all three assessment providers, fall 2021 mathematics and reading gaps were larger for younger students and increased throughout the elementary grades, which is consistent with pre-pandemic gaps on state summative assessments. For example, fall 2021 mathematics and reading gaps for kindergarten students in MAP Growth districts were each about two-thirds of a standard deviation, meaning that economically disadvantaged students in kindergarten scored in the 25th percentile relative to students who are not economically disadvantaged. Similar gaps for 4th- and 5th-grade students were approximately -0.90 and -0.80 standard deviations in mathematics and reading, respectively, translating to average scores between the 18th and 22nd percentiles for wealthier students.

Special Education Status

Finally, Figures 3.3.17 and 3.3.18 show standardized gaps in average scale scores for special education students relative to general education students. These gaps are negative across all grade levels, subjects, and assessment providers, indicating that special education students consistently score lower than their general education peers. K-2 mathematics and reading gaps were smallest in fall 2020 and generally widened across all three assessment periods. In upper elementary and middle school grades, these gaps were generally more stable over time, but tended to decrease slightly from fall 2020 to spring 2021, then increase slightly in fall 2021.

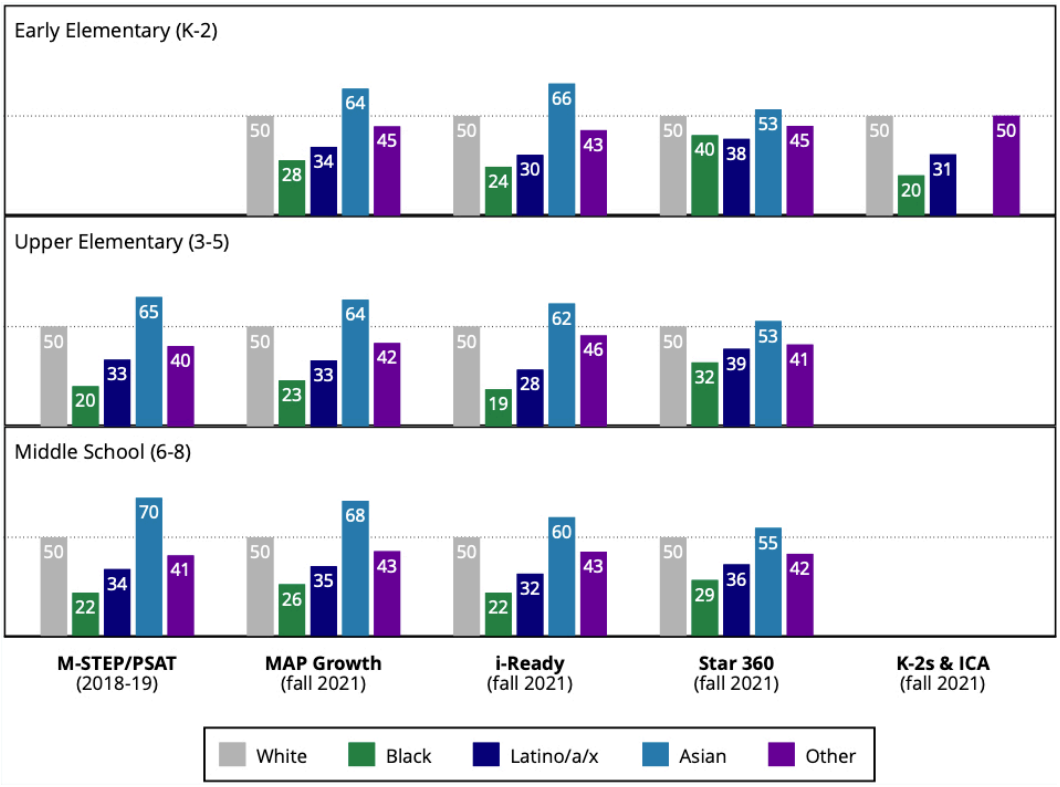
In general, across all assessment providers and both subjects, standardized gaps between special education and general education students were smaller in early grades and increased with each grade level. This pattern is consistent with the gaps between special education and general education students on the 2018-19 state summative assessments, however, the benchmark gaps for older students seen in Figures 3.3.17 and 3.3.18 are considerably larger than historical gaps. For example, the 7th-grade special education-general education mathematics score gap on the 2018-19 M-STEP was about one standard deviation, while fall 2021 mathematics gaps for students in the same grade level ranged from -1.20 to -1.50 standard deviations across assessment providers (between the 7th and 12th percentiles of 7th-grade general education students). This may be due to differences in the populations of special education students who take the M-STEP (as opposed to MI-Access, an alternative summative assessment designed for students with significant cognitive impairments) and those who participate in benchmark assessments. However, to the extent that this pattern is reflective truly of widening achievement gaps between special education and general education student during the pandemic, it is plausible that special education students received fewer services during the pandemic due to logistical challenges stemming from remote learning and necessary safety precautions taken even when students were learning in the school building.

Figure 3.3.1. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 Summative Assessments by Race/Ethnicity (Math)



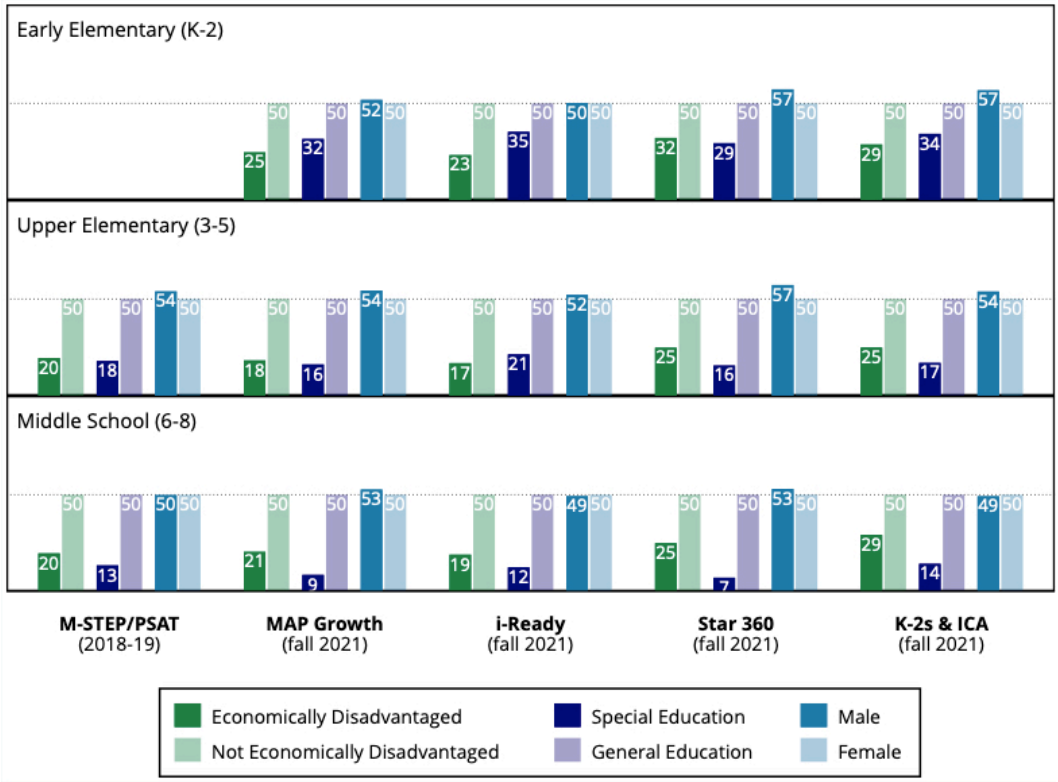
Notes: The numbers in this figure represent the percentile rank of the average score for each subgroup, relative to the distribution of scores for the largest subgroup (White students). We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. There were too few Asian students who participated in the K-2s, and too few students in any non-White subgroup who participated in the Smarter Balanced ICAs, to calculate relative percentiles. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Figure 3.3.2. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 Summative Assessments by Race/Ethnicity (Reading)



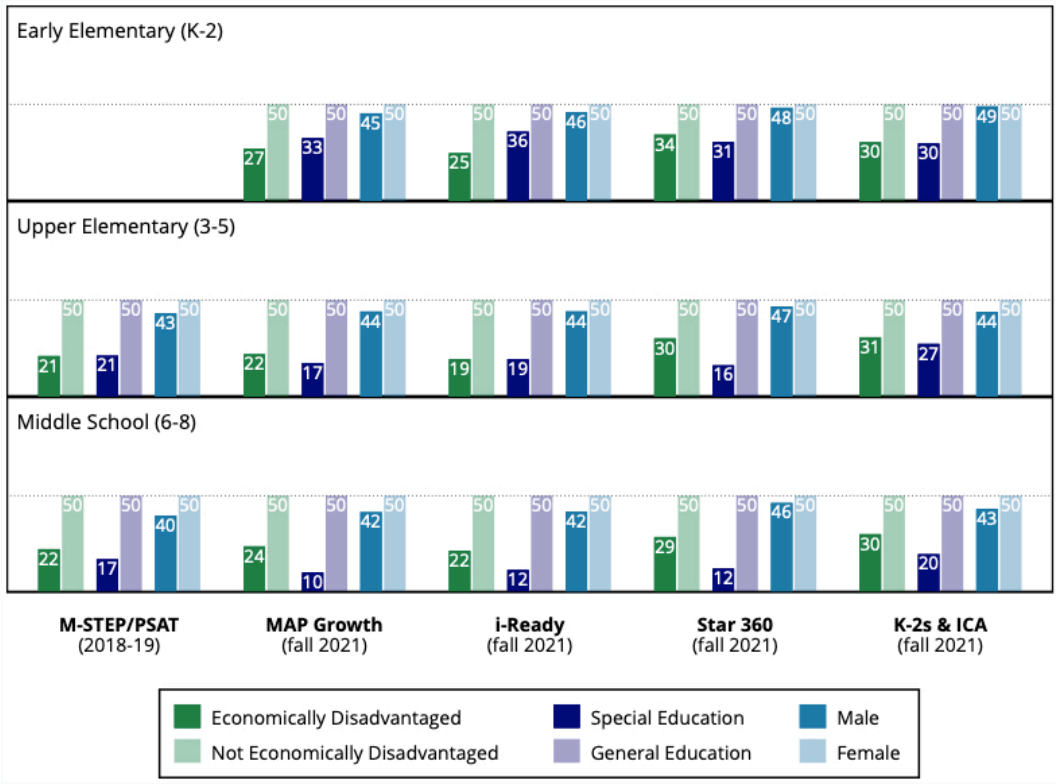
Notes: The numbers in this figure represent the percentile rank of the average score for each subgroup, relative to the distribution of scores for the largest subgroup (White students). We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. There were too few Asian students who participated in the K-2s, and too few students in any non-White subgroup who participated in the Smarter Balanced ICAs, to calculate relative percentiles. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Figure 3.3.3. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 State Summative Assessments by Subgroup (Math)



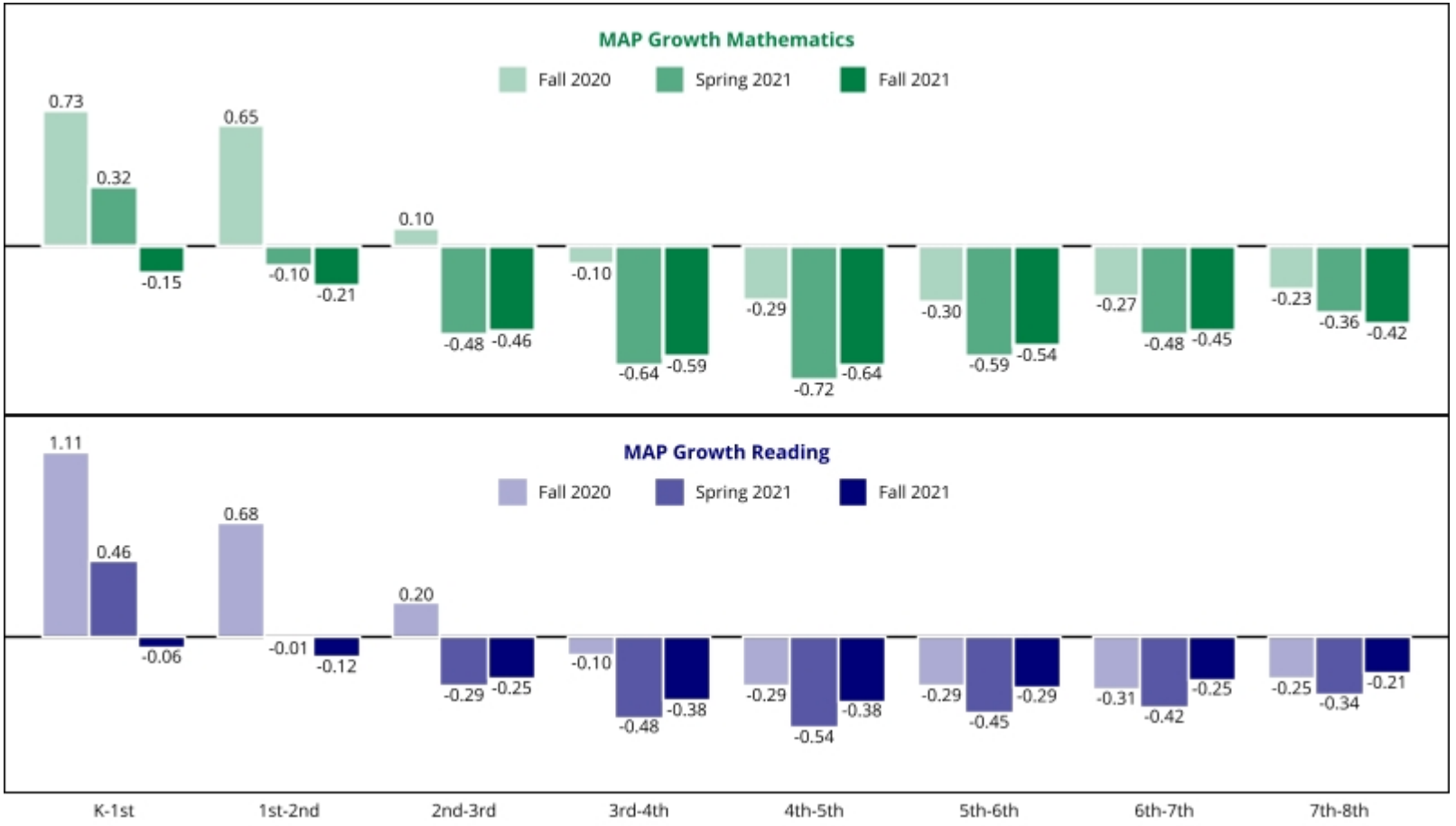
Notes: The numbers in this figure represent the percentile rank of the average score for a subgroup (economically disadvantaged students, special education students, or male students), relative to the distribution of scores for students who are not part of the subgroup. We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Figure 3.3.4. Relative Percentiles on Fall 2021 Benchmark Assessments and 2018-19 State Summative Assessments by Subgroup (Reading)



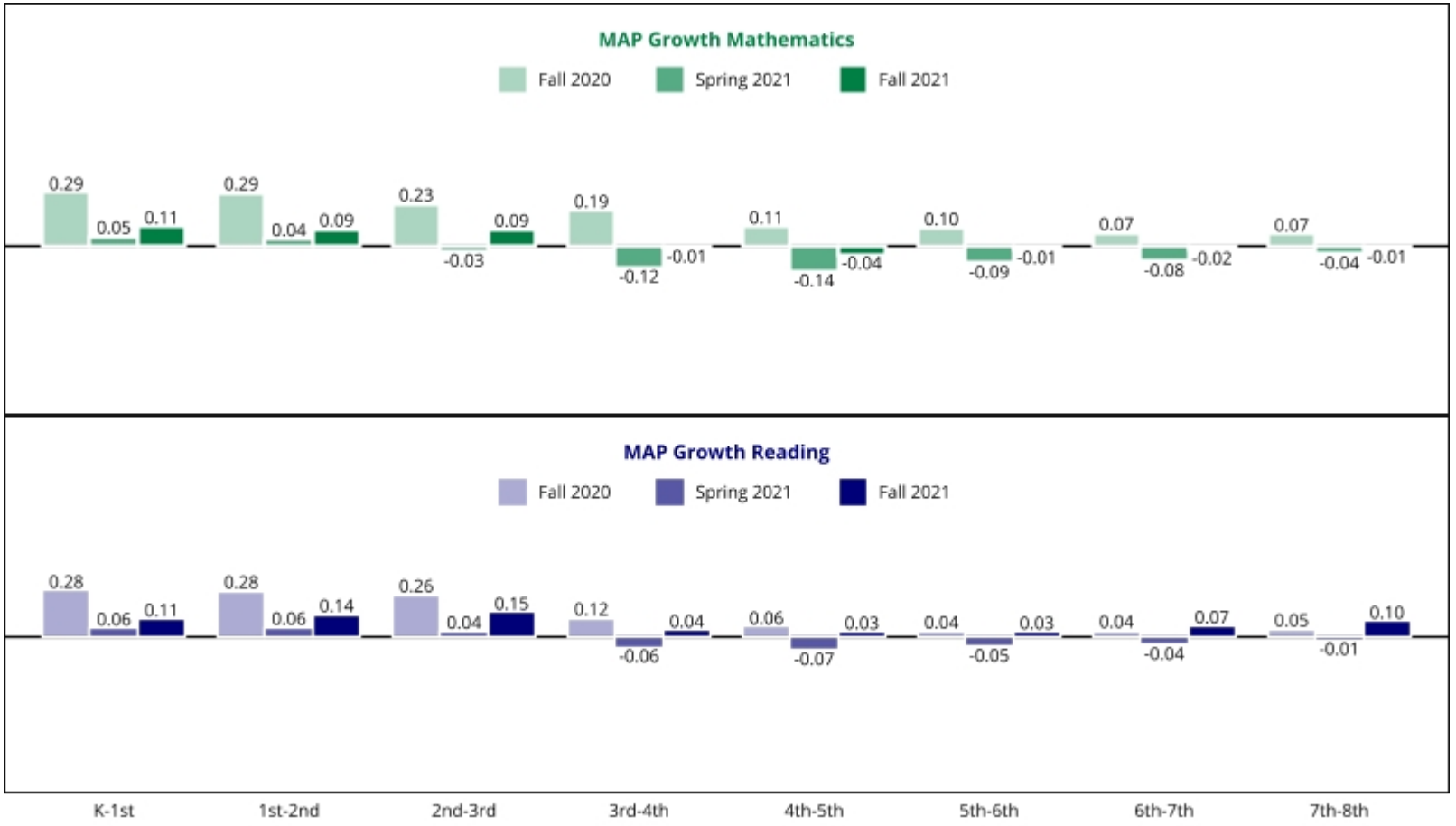
Notes: The numbers in this figure represent the percentile rank of the average score for a subgroup (economically disadvantaged students, special education students, or male students), relative to the distribution of scores for students who are not part of the subgroup. We cannot calculate early elementary percentile ranks for the M-STEP/PSAT because only students in grades 3-8 participate in these assessments. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE. We use publicly available subgroup averages and standard deviations from CEPI's "Grades 3-8 State Testing" report to calculate gaps for the 2018-19 M-STEP and PSAT assessments.

Figure 3.3.5. "Remote All Year" vs. "In-Person All Year" Districts Standardized Achievement Gaps; MAP Growth



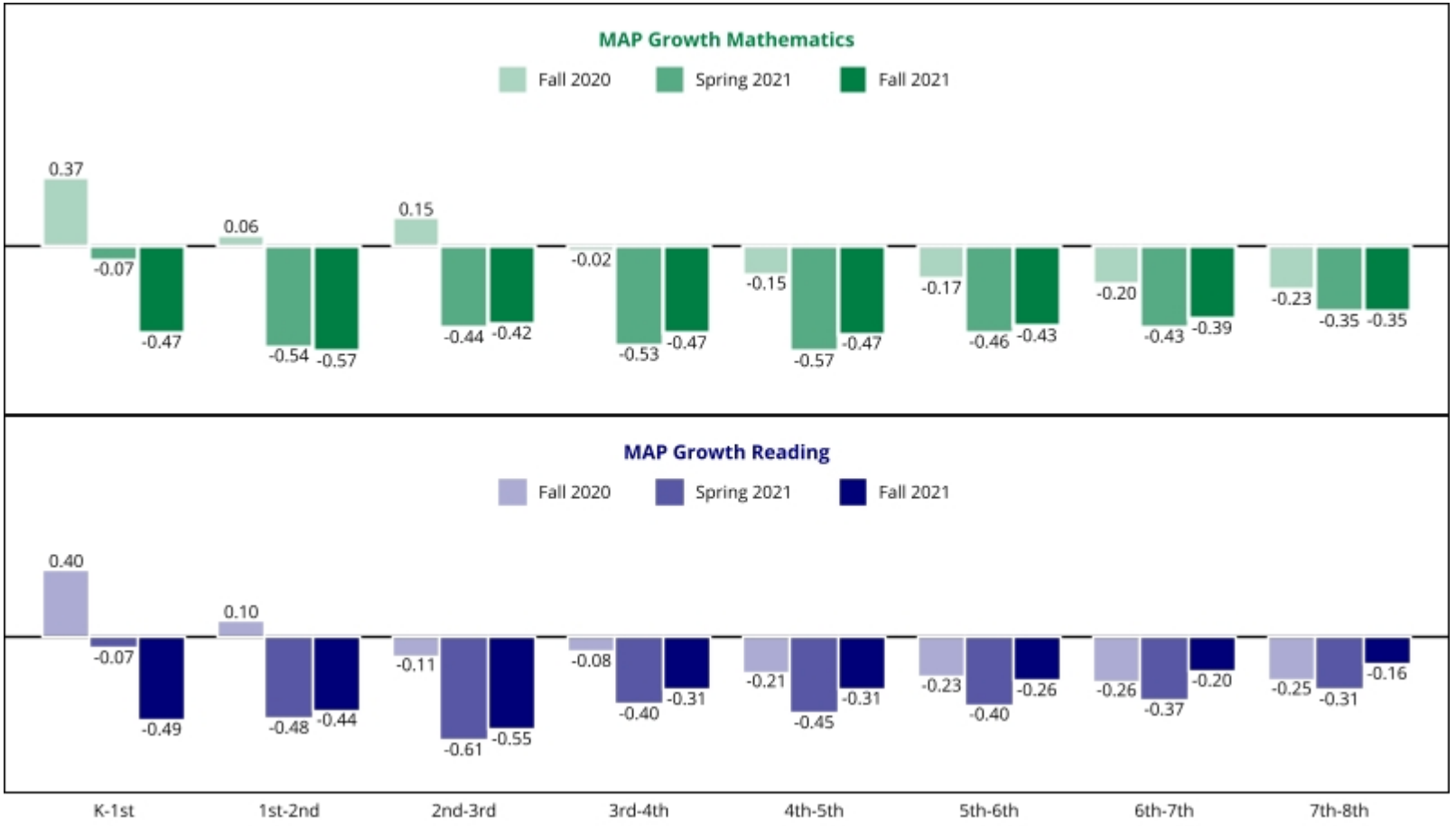
Notes: For more information about this figure, please see Section Three Note 1.

Figure 3.3.6. “In-Person Part-Year” and “In-Person All Year” Standardized Achievement Gaps; MAP Growth



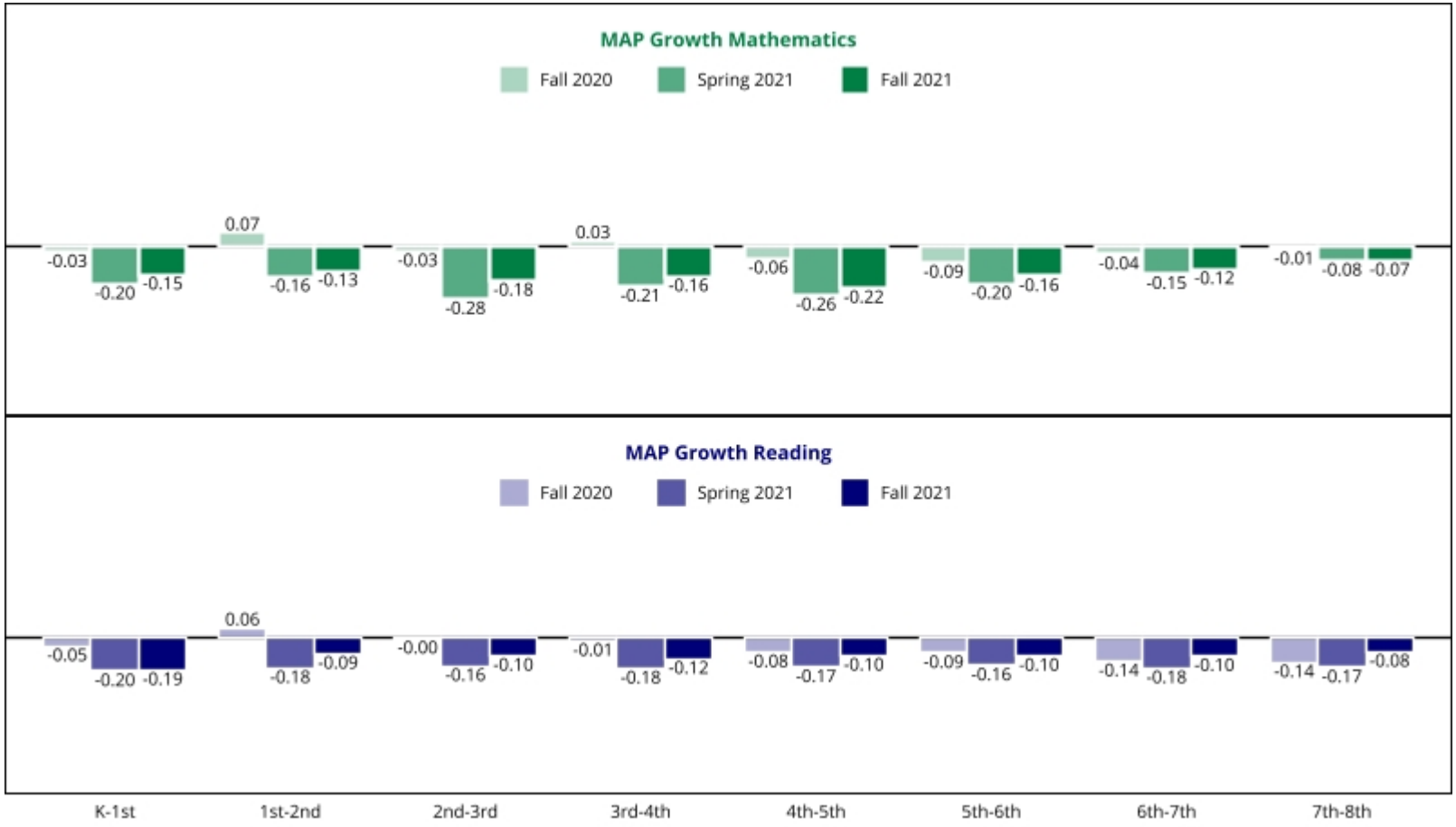
Notes: For more information about this figure, please see Section Three Note 1.

Figure 3.3.7. "Hybrid Part-Year" vs. "In-Person All Year" Standardized Achievement Gaps; MAP Growth



Notes: For more information about this figure, please see Section Three Note 1.

Figure 3.3.8. “Hybrid All Year” vs. “In-Person All Year” Standardized Achievement Gaps; MAP Growth



Notes: For more information about this figure, please see Section Three Note 1.

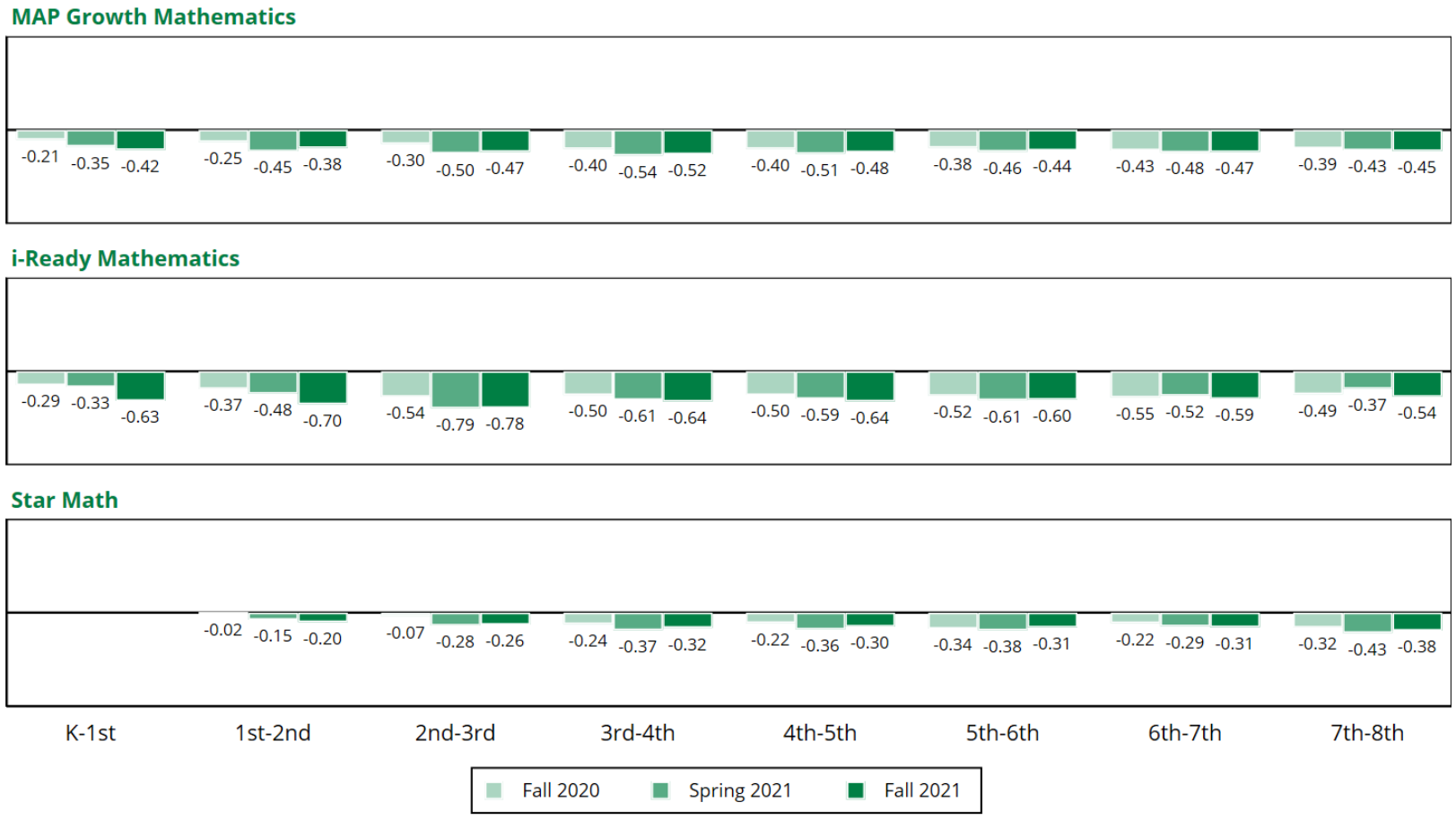
Figure 3.3.9. Black-White Standardized Mathematics Achievement Gaps, All Assessment Providers

Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.10. Black-White Standardized Reading Achievement Gaps, All Assessment Providers

Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.11. Latino/a/x-White Standardized Mathematics Achievement Gaps, All Assessment Providers

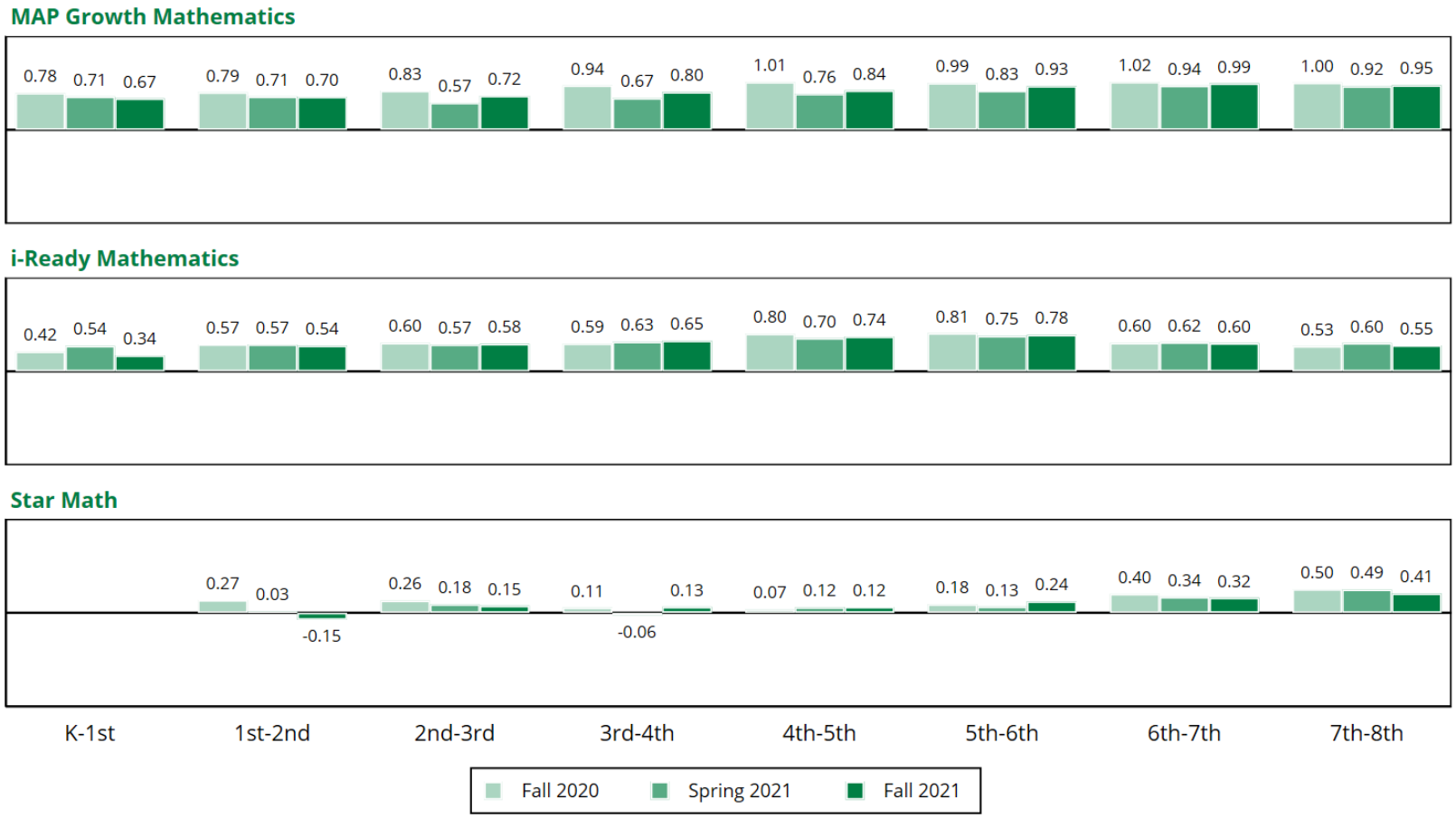


Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.12. Latino/a/x-White Standardized Reading Achievement Gaps, All Assessment Providers

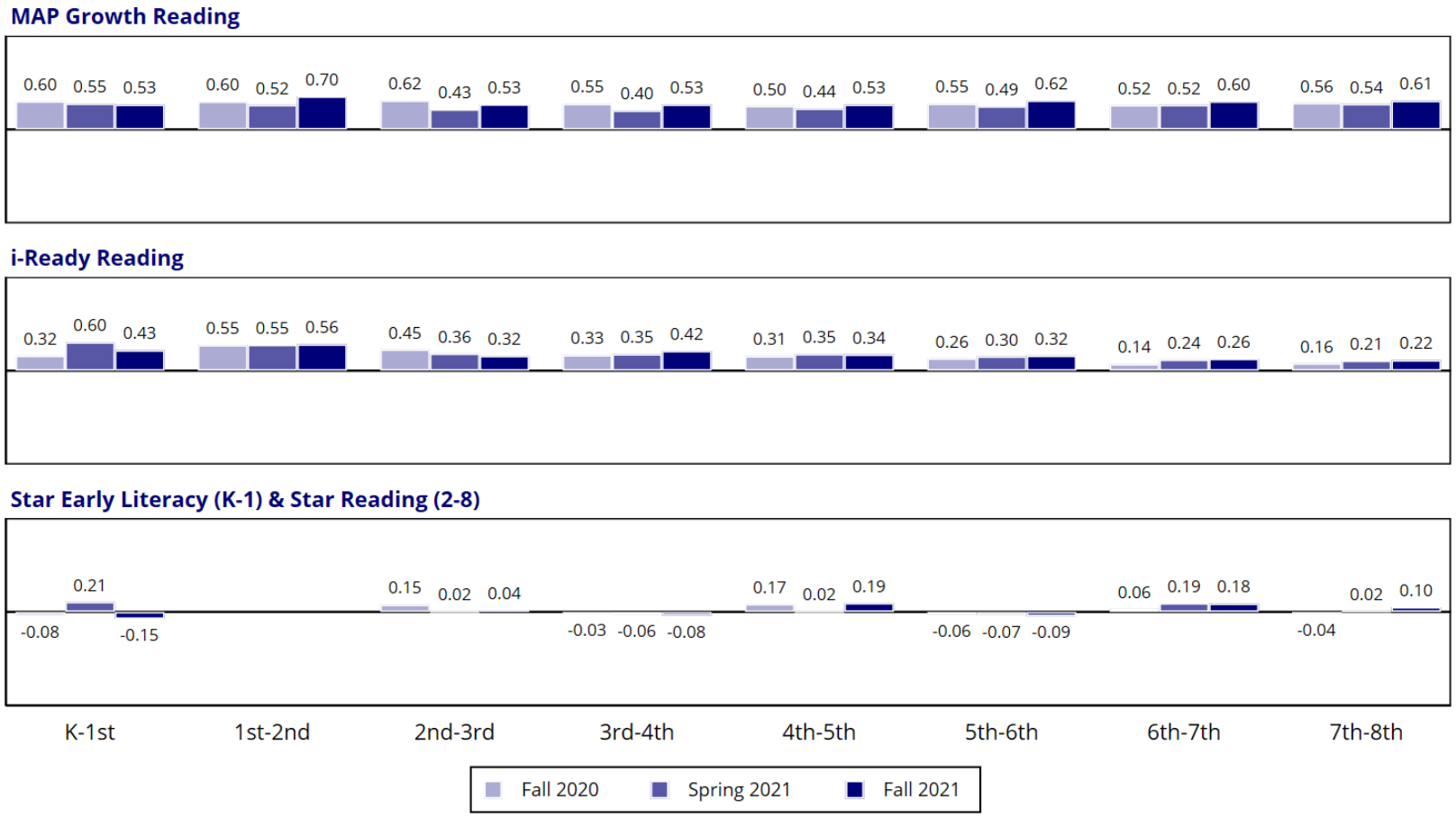
Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.13. Asian-White Standardized Mathematics Achievement Gaps, All Assessment Providers



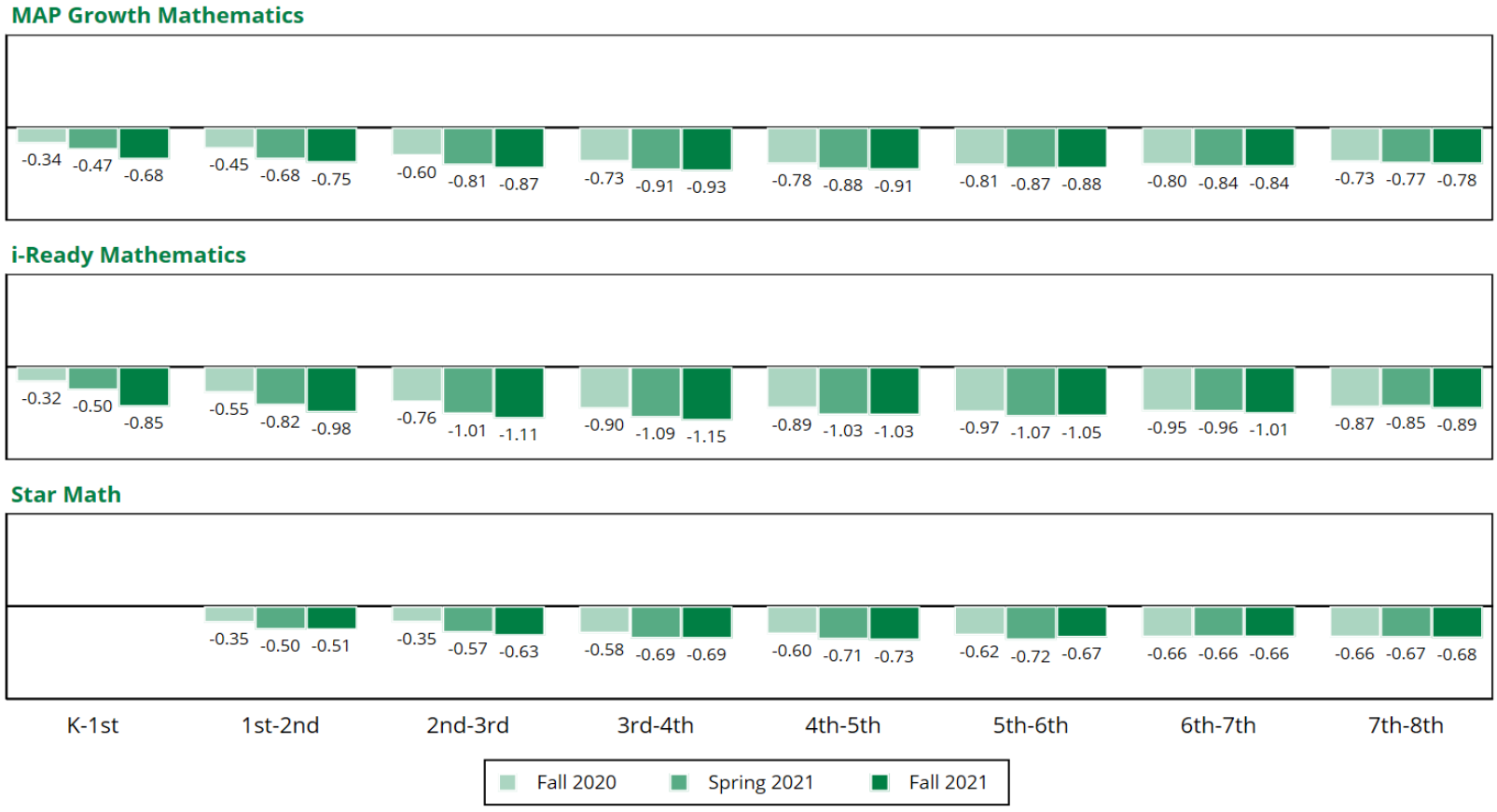
Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.14. Asian-White Standardized Reading Achievement Gaps, All Assessment Providers



Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.15. Economically Disadvantaged-Not Economically Disadvantaged Standardized Mathematics Achievement Gaps, All Assessment Providers



Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.16. Economically Disadvantaged-Not Economically Disadvantaged Standardized Reading Achievement Gaps, All Assessment Providers

Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.17. Special Education-General Education Standardized Mathematics Achievement Gaps Special Education, All Assessment Providers

Notes: For more information about this figure, please see Section Three Note 2.

Figure 3.3.18. Special Education-General Education Standardized Reading Achievement Gaps Special Education, All Assessment Providers

Notes: For more information about this figure, please see Section Three Note 2.

SUMMARY

These results show that K-8 students in Michigan, on average, demonstrated growth on benchmark assessments in reading and mathematics between fall 2020 and fall 2021, but are still falling below pre-pandemic norms for their grade levels. While *on average,* students made progress toward, but did not reach, the appropriate growth targets, the *majority* of students either met their growth targets or did not demonstrate any growth at all. On average across grade levels, subjects, and assessment providers, about 40% of students met or exceeded their growth targets (compared to 50% in a typical year), while about one-third grew by an amount less than their growth target, and a quarter did not demonstrate any growth from fall 2020 to fall 2021.

In addition, we find that longstanding achievement gaps have persisted throughout the 2020-21 school year and into 2021-22, and that these gaps tend to be about the same size as pre-pandemic gaps on state summative assessments. Students with access to in-person instruction outperformed their peers who did not, and these disparities grew over the course of the 2020-21 school year. The gaps were largest for students whose districts were fully remote all year, suggesting that even a limited amount of in-person instruction was beneficial for student achievement.

We also find repeated and consistent evidence that fall 2020 scores overstate students' true ability levels, especially in early elementary grade levels. As a result of the "at-home advantage" for students who completed their fall 2020 tests remotely, we see unrealistically high average scores in lower grade levels, often significantly higher than pre-pandemic norms. Since students' growth targets are based on their initial (fall 2020) scores, many students with inflated baseline scores could not demonstrate growth from fall 2020 to fall 2021 because their growth targets were not realistic for their true achievement levels. Finally, abnormal testing conditions in fall 2020 may explain why many of the standardized achievement gaps for this time period are much smaller than historical gaps on summative assessments. For instance, districts that were fully remote in fall 2020 tended to serve larger proportions of economically disadvantaged, Black, and Latino/a/x students (Hopkins, Kilbride, & Strunk, 2021). Thus, if average scores for these groups were disproportionately inflated due to the "at-home advantage," this likely resulted in unrealistically small achievement gaps that did not accurately reflect inequities between student groups at that time.

SECTION 3 NOTES

1. This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth assessment in fall 2020, spring 2021, and fall 2021. "Remote all year" districts offered only fully remote instruction in both the fall 2020 and spring 2021 testing periods. "In-person part-year" districts offered in-person instruction during either fall 2020 or spring 2021, and hybrid or remote instruction in the other assessment period. "Hybrid part-year" districts offered hybrid instruction in either fall 2020 or spring 2021, and fully remote instruction in the other testing period. "Hybrid all year" districts offered hybrid instruction in both fall 2020 and spring 2021. To calculate achievement gaps, we divide the difference in fall 2021 and fall 2020 average scale scores for the focal group by the standard deviation of scale scores for the reference group. Students in districts that offered in-person instruction during both the fall 2020 and spring 2021 assessment periods are the reference group. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

2. This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, or Renaissance Learning's Star 360 assessments in fall 2020, spring 2021, and fall 2021. To calculate achievement gaps, we divide the difference in average scale scores for the focal group by the standard deviation of scale scores for the reference group. White students, students who are not economically disadvantaged, general education students, and female students are the reference group in each respective analysis (the gender analyses are included in the appendix). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Section Four: Takeaways and Implications

This report expands on the results presented in our previous reports by incorporating another semester of benchmark assessment data and providing a more refined estimate of academic growth. In particular, we explore achievement trends for Michigan students compared to pre-pandemic national or state norms, student progress toward appropriate growth targets, and differences in achievement trends across student subgroups.

While this report helps to deepen our understanding of how Michigan public school students progressed and learned between fall 2020 and fall 2021, there are several limitations of the data that must be stressed. Most importantly, the analyses presented in Section Three are based on imperfect and incomplete data. Our analytic samples include only a subset of students in the state, and prior research has shown that the pandemic has had a greater negative effect on achievement and achievement growth for the specific student populations who are underrepresented in these analyses. Additionally, at-home testing continues to make it difficult to assess student performance, especially for students in younger grades, in fall 2020 and to understand achievement growth using fall 2020 as a baseline, especially for students in younger grades. Further, the sample of students who began kindergarten in fall 2020 is especially small due to pandemic-related declines in kindergarten enrollment.

Nonetheless, the analyses and results discussed in this report enable policymakers, educators, and stakeholders to reflect on the—often differential—effect of the pandemic on student achievement growth throughout the 2020-21 school year and into the fall of 2021. Many of our key findings mirror trends in recent research from other districts and states across the U.S., and also offer new insight about student learning during the pandemic. These data should be used to guide local, state, and national education agencies and policymakers as they continue to work to address the challenges wrought by the pandemic.

KEY FINDINGS

- **There is clear evidence that students who completed their fall 2020 tests remotely had an "at-home advantage," where fall 2020 benchmark assessment scores overstate students' true achievement levels, especially in early grade levels.** This makes it especially difficult to measure fall-to-fall growth, as we do not have an accurate measure of baseline achievement for many younger students. Fall 2020 achievement gaps may also be understated; students who are economically disadvantaged, Black, or Latino/a/x were more likely to attend fully remote districts in fall 2020 (Hopkins, Kilbride, & Strunk, 2021), so average scores for these groups were likely disproportionately affected by the "at-home advantage."

- **Across all grade levels and assessment providers, the average Michigan student experienced growth in mathematics and reading achievement between fall 2020 and fall 2021, but students' scores in fall 2021 were still below pre-pandemic norms.** Aside from early elementary grade levels (whose initial scores were far above pre-pandemic norms due to the "at-home advantage"), Michigan students generally performed close to or slightly below national norms in fall 2020. Average scores increased over the course of the school year, but at a slower rate than would have been expected before the pandemic. In spring 2021, average scores for early elementary students were closer to pre-pandemic norms, and those for upper elementary and middle school students fell even further below the norms. By fall 2021, across nearly all grade levels, subjects, and assessment providers, Michigan students consistently scored below pre-pandemic norms.

- **Although *on average,* students made less-than-expected growth on their benchmark assessments, many Michigan students met their growth targets for the year, and still many others made no growth at all.** Student growth targets represent the median growth for students with similar prior achievement scores before the pandemic, so in a typical year, we would expect about 50% of students to reach these targets. Across grades, subjects, and assessment providers, we found that only about 40% of Michigan students met or exceeded their expected growth targets, while about a quarter did not demonstrate any growth from fall 2020 to fall 2021. Students with lower fall 2020 achievement scores were more likely to reach their growth targets, while students with the highest initial achievement were more likely to make zero or negative growth. However, these patterns vary by grade level. In early elementary grades, students with the highest initial scores were substantially

more likely than their peers to make zero or negative growth, presumably because many of the students in this group had inflated baseline scores due to the "at-home advantage," and as a result, could not maintain the same achievement level after returning to in-person testing. For middle school students, in many cases we find evidence of bifurcated growth, whereby students were more likely to make zero or negative growth or to meet or exceed their growth targets than they were to make less than expected growth.

- **Longstanding racial/ethnic and socioeconomic achievement gaps persisted into the 2021-22 school year.** On the MAP Growth assessments, average fall 2021 math scores for Black students were between the 15th and 22nd percentiles of White students' scores, while average math scores for Latino/a/x students were between the 30th and 32nd percentiles, and those for Asian students were between the 67th and 76th percentiles. Average reading scores for Black students were between the 23rd and 28th percentiles of White students' scores, while those of Latino/a/x students were between the 33rd and 35th percentiles, and those of Asian students were between the 64th and 68th percentiles. Fall 2021 gaps between students who are and are not economically disadvantaged were even larger; average MAP Growth math scores for economically disadvantaged students were between the 18th and 25th percentiles for students who are not economically disadvantaged, while their average reading scores were between the 22nd and 27th percentiles.

  We found similarly large racial/ethnic and socioeconomic achievement gaps across all assessment providers and in all grade levels. Each of these achievement gaps on the fall 2021 benchmark assessments are comparable in size to the achievement gaps on Michigan's 2018-19 summative assessments, suggesting that achievement gaps—while large and concerning—may not have substantially grown throughout the 2020-21 school year, at least not among the tested population. We note that these gaps represent overall differences between entire populations of students (not accounting for any additional student characteristics) rather than differences between otherwise-similar students from two different subgroups, and that there are likely several interrelated factors that contribute to these overall gaps.

- **Achievement gaps between students in districts offering fully remote instruction and those that offered in-person instruction grew over the course of the 2020-21 school year but improved slightly over the summer.** Across all upper elementary and middle school grade levels, students learning in-person consistently outperformed students learning remotely on the MAP

Growth assessments, and these gaps grew substantially from fall 2020 to spring 2021. This provides some evidence that not only did students in in-person districts generally begin the 2020-21 school year at higher achievement levels than their peers in districts offering remote or hybrid instruction, but the missed opportunities to learn in person during the 2020-21 school year exacerbated these discrepancies. However, most of these gaps decreased slightly from spring 2021 to fall 2021, particularly in reading, suggesting that students are starting to recover from missed opportunities for face-to-face instruction now that most of them have returned to fully in-person learning modalities. In fall 2020, students in early elementary grade levels in fully remote districts greatly out-performed their peers in fully in-person districts, showing clear evidence of an "at-home advantage." These gaps decreased in spring 2021 and reversed in direction by fall 2021 as more students returned to at-school testing.

- **Students with access to some in-person instruction in 2020-21 were less affected than those whose districts only offered fully remote instruction.** Achievement gaps between students in districts that offered fully in-person or hybrid instruction during *either* fall 2020 or spring 2021 (but not both) and those whose districts offered in-person instruction all year were generally negative, but smaller than the gaps for students whose districts were fully remote all year. Further, gaps were smaller for students whose districts offered fully in-person instruction for part of the year, compared to those that switched between remote and hybrid instruction. These patterns again suggest that students with access to more in-person schooling fared better than those with fewer opportunities for face-to-face instruction.

IMPLICATIONS

Our findings make it clear that interruptions to student learning and instruction during the COVID-19 pandemic negatively affected average student achievement and growth outcomes for Michigan students, and many students made no gains on benchmark assessments during the 2020-21 school year. Moreover, disruptions to learning were not spread out equitably amongst the student population. This research highlights a few important considerations for moving forward:

- **Stakeholders should interpret and use fall 2020 benchmark assessment results with caution.** Without giving due consideration to the "at-home advantage" for early elementary students, stakeholders could come to misleading conclusions about student performance, growth, and achievement gaps based on

the fall 2020 data. We find that most abnormal patterns in the benchmark assessment data are concentrated among K-2 students whose fall 2020 scores were above the 80th percentile. Focusing on students whose fall 2020 scores were at or below the 80th percentile will provide a more reliable (albeit less representative) picture of student performance outcomes. This is true in our data and analyses, as well as in national reporting from assessment vendors.

- **It will be especially important to continue monitoring learning outcomes for economically disadvantaged, Black, and Latino/a/x students.** Research exploring trends in academic achievement over the past two years makes clear that the effects of the COVID-19 pandemic on students varied across student populations and the pandemic has had a greater and more negative effect on economically disadvantaged, Black, and Latino/a/x students. Given that these student groups are underrepresented in our analyses, and those who are represented are more likely to be affected by fall 2020 testing irregularities, we still do not have a comprehensive understanding of learning trajectories for these populations of Michigan students during the pandemic. Any decisions to reduce monitoring student learning progress risk placing these populations of students—those who have been most disadvantaged both prior to and during the pandemic—at greater risk of falling through the cracks.

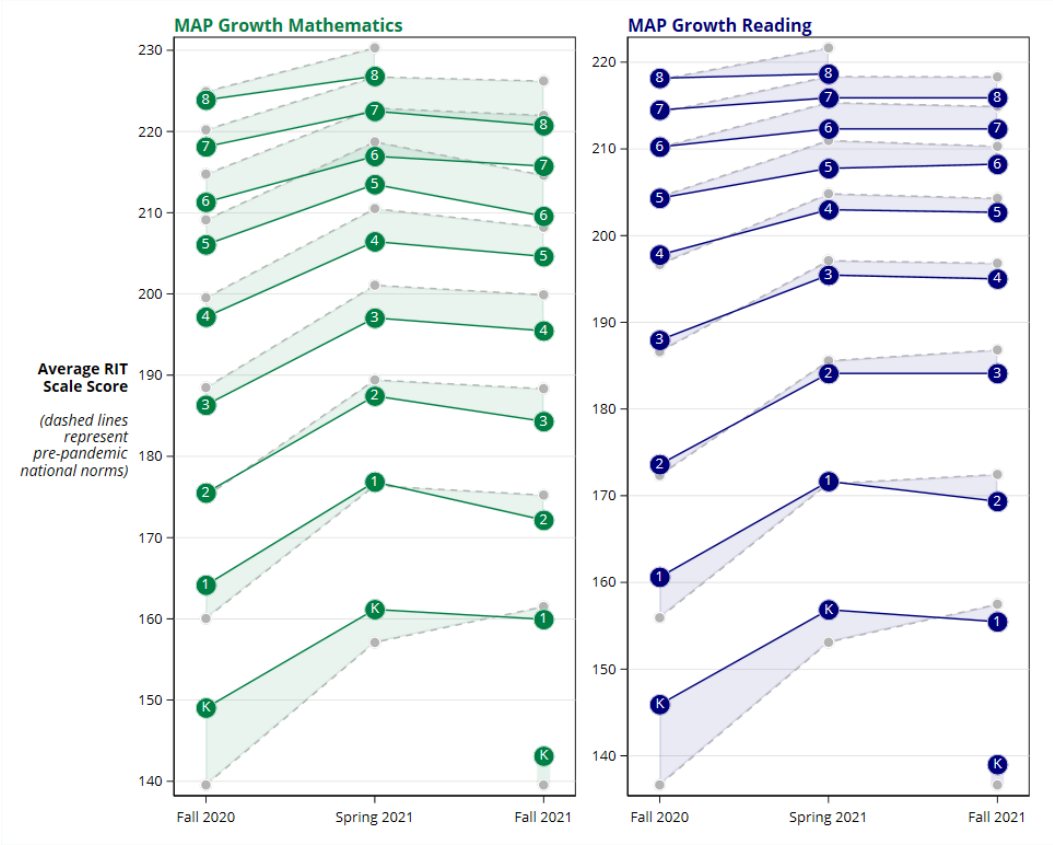
- **Focusing too narrowly on average learning outcomes likely masks the substantial number of students who are deeply struggling.** Based on average growth outcomes alone, it appears that Michigan students grew at a slower rate than would have been expected before the pandemic. However, when we examine the *distribution* of growth outcomes across Michigan students, we find that it was more common for students to reach or exceed their growth targets or to make no growth at all than it was for them to make less than expected growth. In other words, substantial proportions of Michigan students are doing well, and substantial proportions suffered from missed learning opportunities. Thus, state leaders and local educators should target resources and interventions towards the large portion of Michigan students who most need assistance with accelerating their learning—those students who made no or negative achievement growth during the 2020-21 school year (while making adjustments for students who exhibited lower rates of learning given artificially high fall 2020 scores).

- **The magnitude of the discrepancies between expected and actual performance and achievement growth that occurred as a result of the**

pandemic will not be addressed quickly, or without a substantial and sustained influx of resources to support education in Michigan. As noted above, both average achievement and achievement growth was lower during the pandemic than would have been the case in a more typical period, and substantial proportions of students made no achievement growth whatsoever during the 2020-21 school year and into the fall of 2021. This is not unique to Michigan, but rather echoes recent findings from across the U.S. that consistently show lower average achievement (e.g., Jack et al., 2021; Kuhfeld et al., 2022; North Carolina State Board of Education, 2022) and less than typical achievement growth (e.g., Regional Educational Laboratory, 2022) than would have been expected for similar students before the pandemic. It will be critical for local, state, and the federal governments to continue to prioritize both short- and longer-term investments into public education, in Michigan and elsewhere, as educators and students work to recover.

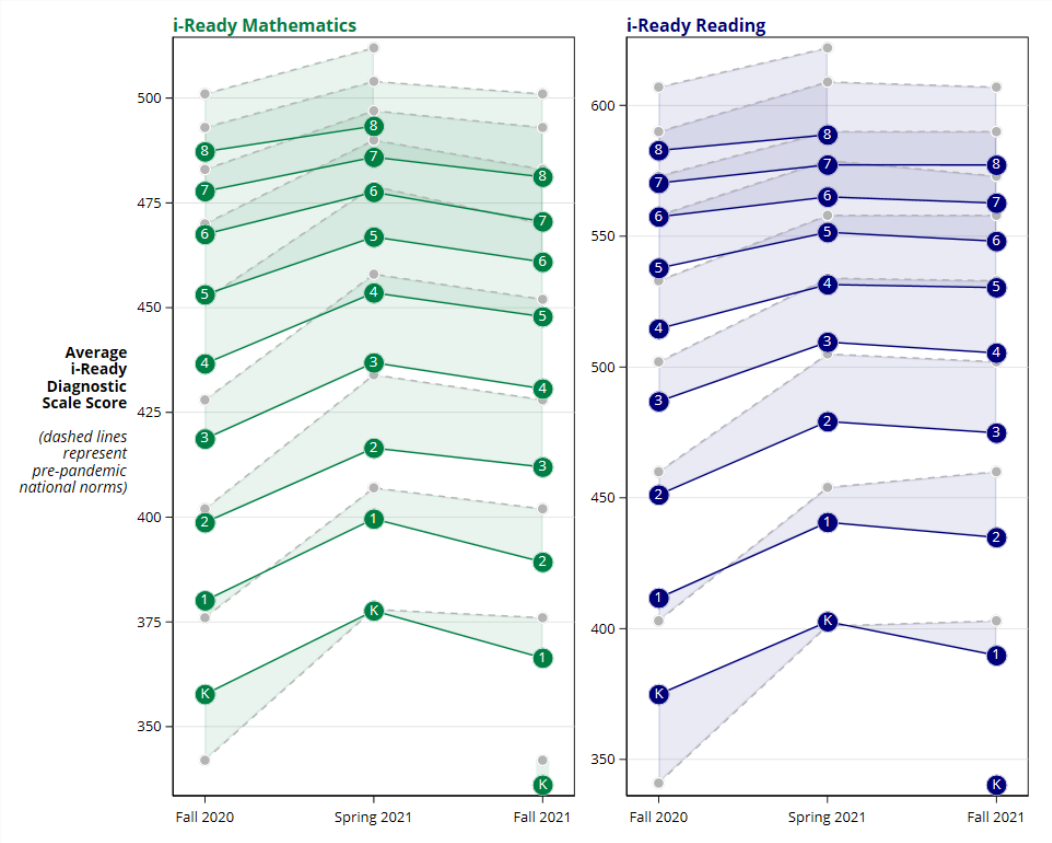
Appendix

Figure A1.1 Trends in Average Scale Scores on NWEA's MAP Growth Mathematics and Reading Assessments



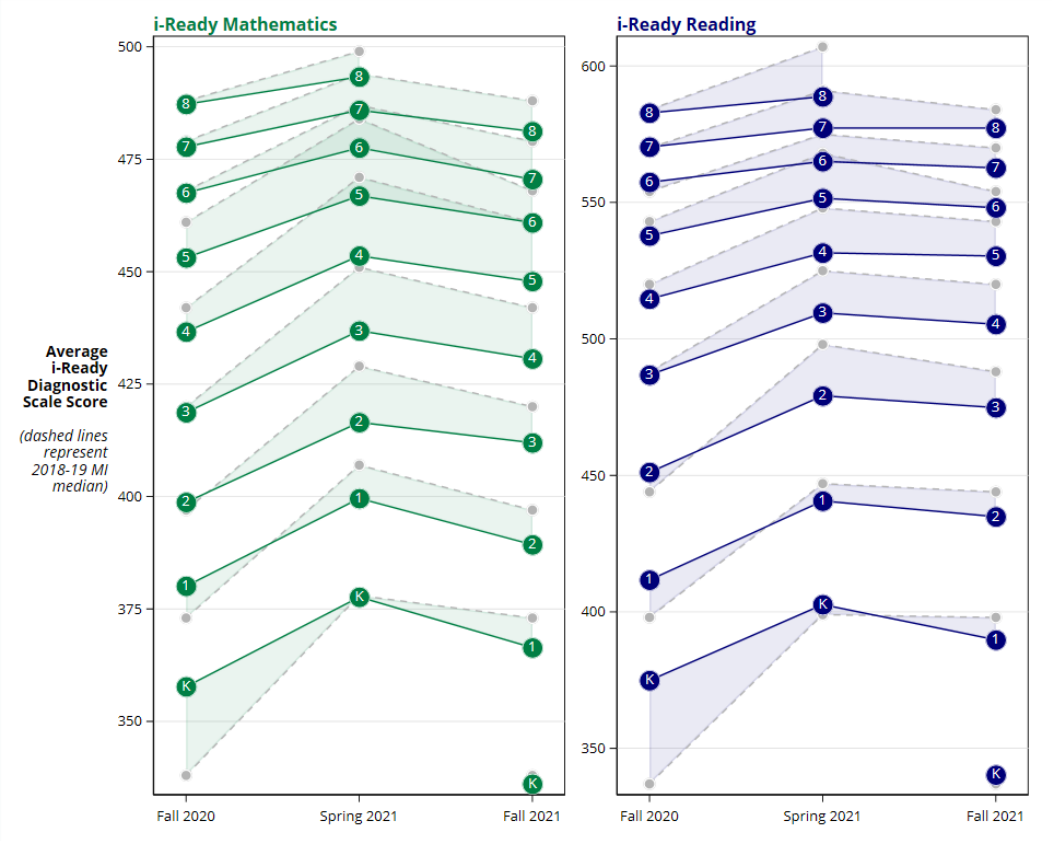
Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth Mathematics or Reading assessments in fall 2021. The comparison points in the figure represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. These data can be found here: https://teach.mapnwea.org/impl/normsResearchStudy.pdf. NWEA's MAP Growth assessments are scored on a vertical scale which ranges from 100 to 350, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A1.2. Trends in Average Scale Scores on Curriculum Associates i-Ready Mathematics and Reading Assessments



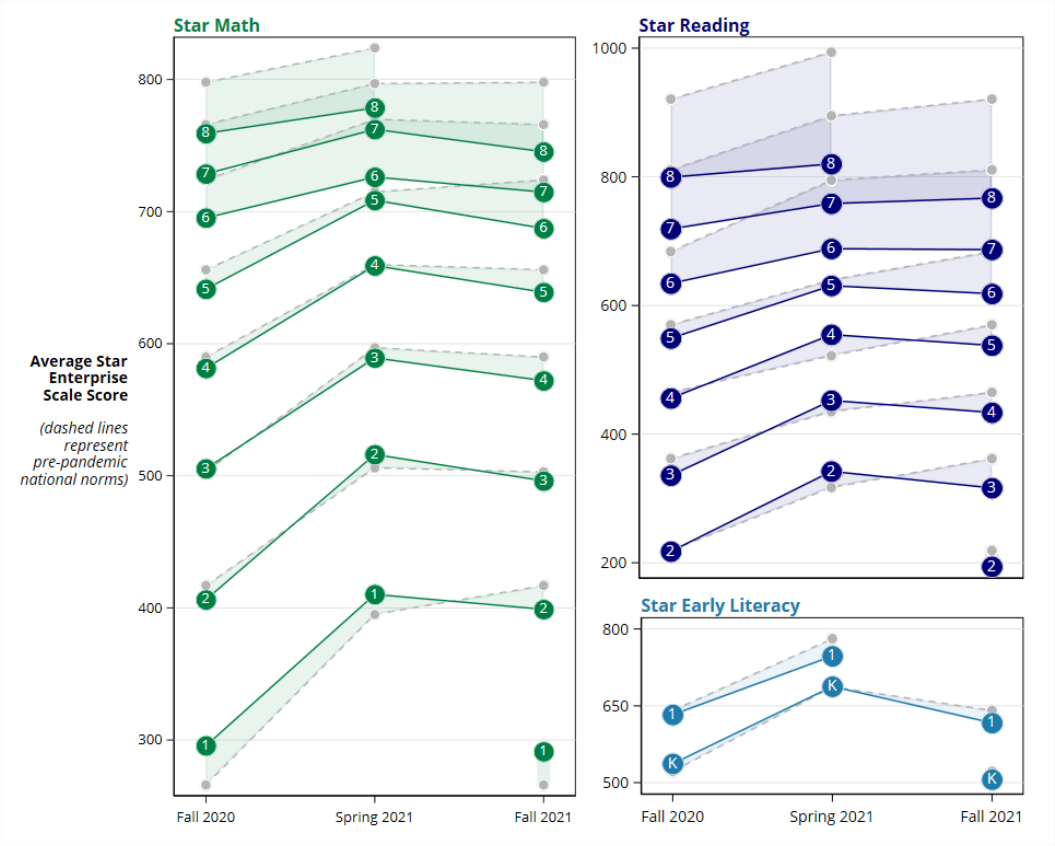
Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2021. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A1.3. Trends in Average Scale Scores on Curriculum Associates i-Ready Mathematics and Reading Assessments



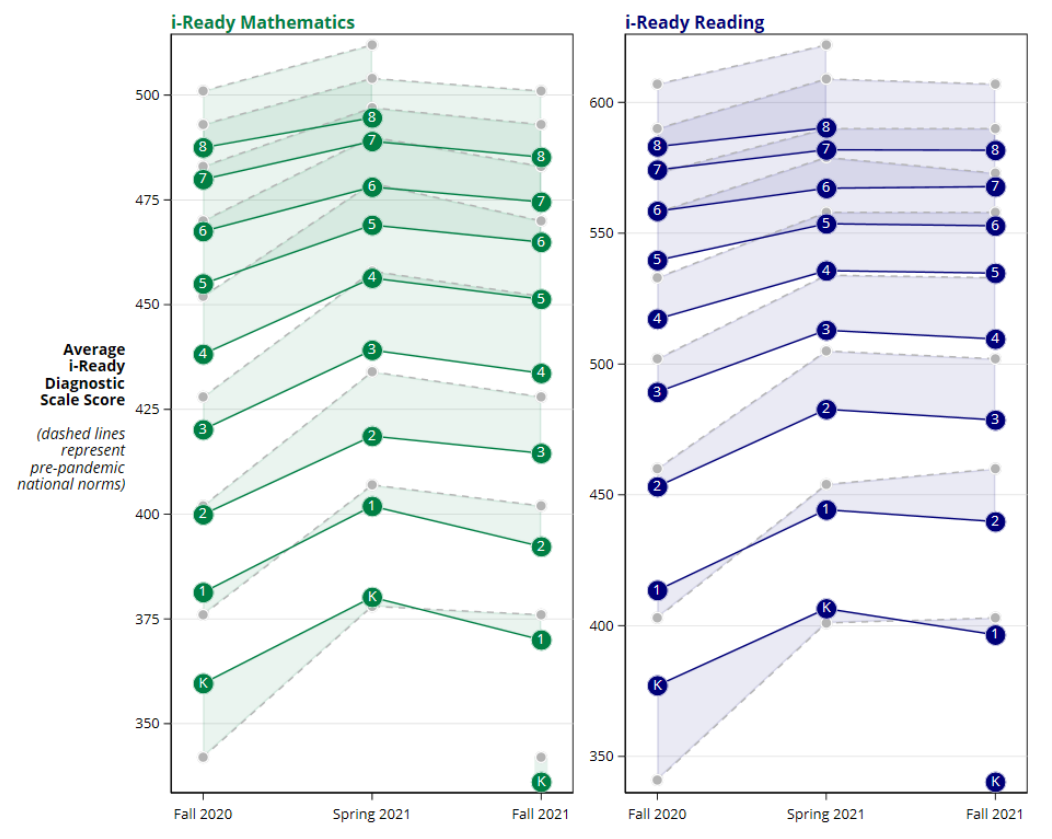
Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2021. The Curriculum Associates comparison points represent median scale scores derived from the fall and spring distributions of scale scores unique to the group of Michigan students who completed an i-Ready benchmark assessment during the 2018-19 school year. Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A1.4. Trends in Average Scale Scores on Renaissance Learning's Star Math, Reading, and Early Literacy Assessments



Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Mathematics, Reading, or Early Literacy assessments in fall 2021. We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Renaissance Learning's Star Math and Reading assessments are scored on a vertical scale which ranges from 0 to 1400, and the Early Literacy assessment is scored from 300 to 900. The y-axis scales in this figure differ across the math, reading, and literacy assessments to reflect differences in the scales and grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A1.5. Trends in Average Scale Scores on Curriculum Associates i-Ready Mathematics and Reading Assessments



Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2020, spring 2021, and fall 2021. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.1. Trends in Average Scale Scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, and Renaissance Learning's Star Mathematics Assessments; Fall-Spring-Fall Sample and National Norms

Grade 2020	Semester	MAP Growth		i-Ready		Star 360	
		Average	Norm	Average	Norm	Average	Norm
K ('21)	F21	143.11	139.6	336.1	342		
K	F20	149.38	139.6	359.6	342		
	S21	162.55	157.1	380.2	378		
	F21	160.86	161.5	370.0	376		
1st	F20	163.56	160.1	381.3	376	293.8	266
	S21	177.51	176.4	401.9	407	415.6	395
	F21	173.12	175.2	392.3	402	407.8	417
2nd	F20	175.51	175.0	399.9	402	405.9	417
	S21	188.30	189.4	418.6	434	522.0	506
	F21	185.56	188.3	414.5	428	506.4	503
3rd	F20	186.74	188.5	420.2	428	504.6	503
	S21	198.17	201.1	439.2	458	592.7	597
	F21	196.88	199.9	433.6	452	582.0	590
4th	F20	197.76	199.6	438.2	452	582.9	590
	S21	207.59	210.5	456.3	479	664.0	660
	F21	206.13	208.2	451.3	470	653.1	656
5th	F20	206.94	209.1	455.0	470	646.4	656
	S21	215.04	218.7	469.0	490	717.3	715
	F21	210.88	214.6	464.9	483	700.5	724
6th	F20	212.33	214.8	467.5	483	701.4	724
	S21	218.59	222.9	478.1	497	735.1	770
	F21	217.54	222.0	474.5	493	724.7	766
7th	F20	218.71	220.2	479.9	493	730.1	766
	S21	223.64	226.7	489.0	504	764.4	797
	F21	222.64	226.2	485.2	501	755.4	798
8th	F20	224.17	224.9	487.5	501	762.6	798
	S21	227.53	230.3	494.6	512	780.5	824

Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates i-Ready, or Renaissance Learning Star Mathematics or Reading assessments in fall 2020, spring, 2021, and fall 2021. The comparison points for NWEA represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.2. Trends in Average Scale Scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, and Renaissance Learning's Star Reading Assessments; Fall-Spring-Fall Sample and National Norms

Grade 2020	Semester	MAP Growth		i-Ready		Star 360	
		Average	Norm	Average	Norm	Average	Norm
K ('21)	F21	139.01	136.7	340.3	341	506.3	522*
K	F20	145.72	136.7	377.1	341	537.6	522*
	S21	157.96	153.1	406.5	401	692.4	687*
	F21	156.57	157.5	396.5	403	624.5	641*
1st	F20	159.85	155.9	413.5	403	624.1	641*
	S21	172.16	171.4	444.3	454	750.6	781*
	F21	170.26	172.5	439.8	460	194.1	219
2nd	F20	173.65	172.4	453.1	460	219.9	219
	S21	185.15	185.6	482.7	505	353.5	317
	F21	184.99	186.8	478.6	502	334.6	362
3rd	F20	188.41	186.6	489.3	502	335.5	362
	S21	196.49	197.1	512.9	534	458.8	435
	F21	196.15	196.8	509.6	533	450.8	465
4th	F20	198.42	196.7	517.2	533	457.6	465
	S21	204.02	204.8	535.7	558	560.0	522
	F21	203.79	204.3	534.7	558	552.1	570
5th	F20	205.14	204.5	539.6	558	550.3	570
	S21	208.89	211.0	553.6	579	638.0	640
	F21	209.16	210.3	552.8	573	637.0	684
6th	F20	211.23	210.2	558.4	573	642.0	684
	S21	213.60	215.4	567.2	590	704.0	795
	F21	213.51	214.9	567.8	590	706.0	811
7th	F20	215.40	214.2	574.2	590	728.0	811
	S21	217.21	218.4	581.9	609	770.1	895
	F21	217.28	218.3	581.7	607	791.2	921
8th	F20	218.46	218.0	583.1	607	804.8	921
	S21	219.23	221.7	590.3	622	827.7	994

Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates i-Ready, or Renaissance Learning Star Mathematics or Reading assessments in fall 2020, spring, 2021, and fall 2021. The comparison points for NWEA represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.3. Trends in Average Scale Scores on Curriculum Associates' i-Ready Mathematics and Reading Assessments; Fall-Spring-Fall Sample and Michigan Norms					
Grade 2020	**Semester**	**Mathematics**		**Reading**	
		Average	**Norm**	**Average**	**Norm**
K ('21)	F21	336.1	338	340.3	337
K	F20	359.6	338	377.1	337
	S21	380.2	378	406.5	399
	F21	370.0	373	396.5	398
1st	F20	381.3	373	413.5	398
	S21	401.9	407	444.3	447
	F21	392.3	397	439.8	444
2nd	F20	399.9	397	453.1	444
	S21	418.6	429	482.7	498
	F21	414.5	420	478.6	488
3rd	F20	420.2	420	489.3	488
	S21	439.2	451	512.9	525
	F21	433.6	442	509.6	520
4th	F20	438.2	442	517.2	520
	S21	456.3	471	535.7	548
	F21	451.3	461	534.7	543
5th	F20	455.0	461	539.6	543
	S21	469.0	484	553.6	568
	F21	464.9	468	552.8	554
6th	F20	467.5	468	558.4	554
	S21	478.1	487	567.2	575
	F21	474.5	479	567.8	570
7th	F20	479.9	479	574.2	570
	S21	489.0	494	581.9	591
	F21	485.2	488	581.7	584
8th	F20	487.5	488	583.1	584
	S21	494.6	499	590.3	607

Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2020, spring, 2021, and fall 2021. The Curriculum Associates comparison points represent median scale scores derived from the fall and spring distributions of scale scores unique to the group of Michigan students who completed an i-Ready benchmark assessment during the 2018-19 school year. Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.4. Trends in Average Scale Scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, and Renaissance Learning's Star Mathematics Assessments; Fall 2021 Sample and National Norms

Grade 2020	Semester	MAP Growth		i-Ready		Star 360	
		Average	Norm	Average	Norm	Average	Norm
K ('21)	F21	143.1	139.6	336.1	342		
K	F20	149.0	139.6	357.7	342		
	S21	161.2	157.1	377.7	378		
	F21	159.9	161.5	366.4	376		
1st	F20	164.2	160.1	380.1	376	295.7	266
	S21	176.9	176.4	399.6	407	410.3	395
	F21	172.2	175.2	389.3	402	398.9	417
2nd	F20	175.5	175.0	398.7	402	406.2	417
	S21	187.4	189.4	416.5	434	516.2	506
	F21	184.3	188.3	412.0	428	496.1	503
3rd	F20	186.3	188.5	418.7	428	505.4	503
	S21	197.0	201.1	436.9	458	589.3	597
	F21	195.5	199.9	430.6	452	572.0	590
4th	F20	197.2	199.6	436.6	452	581.4	590
	S21	206.5	210.5	453.5	479	659.1	660
	F21	204.6	208.2	447.9	470	639.0	656
5th	F20	206.1	209.1	453.1	470	641.4	656
	S21	213.5	218.7	466.9	490	708.6	715
	F21	209.6	214.6	460.9	483	687.3	724
6th	F20	211.3	214.8	467.5	483	695.2	724
	S21	217.0	222.9	477.6	497	726.2	770
	F21	215.7	222.0	470.5	493	714.8	766
7th	F20	218.1	220.2	477.8	493	728.5	766
	S21	222.5	226.7	485.9	504	762.3	797
	F21	220.7	226.2	481.2	501	745.3	798
8th	F20	223.9	224.9	487.2	501	759.3	798
	S21	226.8	230.3	493.3	512	778.6	824

Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates i-Ready, or Renaissance Learning Star Mathematics or Reading assessments in fall 2021. The comparison points for NWEA represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.5. Trends in Average Scale Scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, and Renaissance Learning's Star Reading Assessments; Fall 2021 Sample and National Norms

Grade 2020	Semester	MAP Growth		i-Ready		Star 360	
		Average	Norm	Average	Norm	Average	Norm
K ('21)	F21	139.0	136.7	340.3	341	506.3	522*
K	F20	145.9	136.7	374.9	341	537.1	522*
	S21	156.9	153.1	402.7	401	687.7	687*
	F21	155.5	157.5	389.8	403	616.7	641*
1st	F20	160.6	155.9	411.7	403	632.7	641*
	S21	171.7	171.4	440.6	454	747.1	781*
	F21	169.4	172.5	434.9	460	194.1	219
2nd	F20	173.6	172.4	451.2	460	217.6	219
	S21	184.1	185.6	479.2	505	342.2	317
	F21	184.1	186.8	474.8	502	316.3	362
3rd	F20	188.0	186.6	486.8	502	335.7	362
	S21	195.5	197.1	509.5	534	452.4	435
	F21	195.0	196.8	505.4	533	433.5	465
4th	F20	197.8	196.7	514.6	533	455.8	465
	S21	203.0	204.8	531.5	558	554.8	522
	F21	202.7	204.3	530.4	558	537.8	570
5th	F20	204.3	204.5	537.8	558	549.0	570
	S21	207.8	211.0	551.5	579	630.7	640
	F21	208.2	210.3	548.1	573	618.3	684
6th	F20	210.2	210.2	557.4	573	634.2	684
	S21	212.3	215.4	565.0	590	688.3	795
	F21	212.3	214.9	562.7	590	686.8	811
7th	F20	214.5	214.2	570.3	590	719.0	811
	S21	215.9	218.4	577.3	609	758.5	895
	F21	215.9	218.3	577.2	607	767.0	921
8th	F20	218.2	218.0	582.8	607	799.5	921
	S21	218.7	221.7	588.8	622	820.1	994

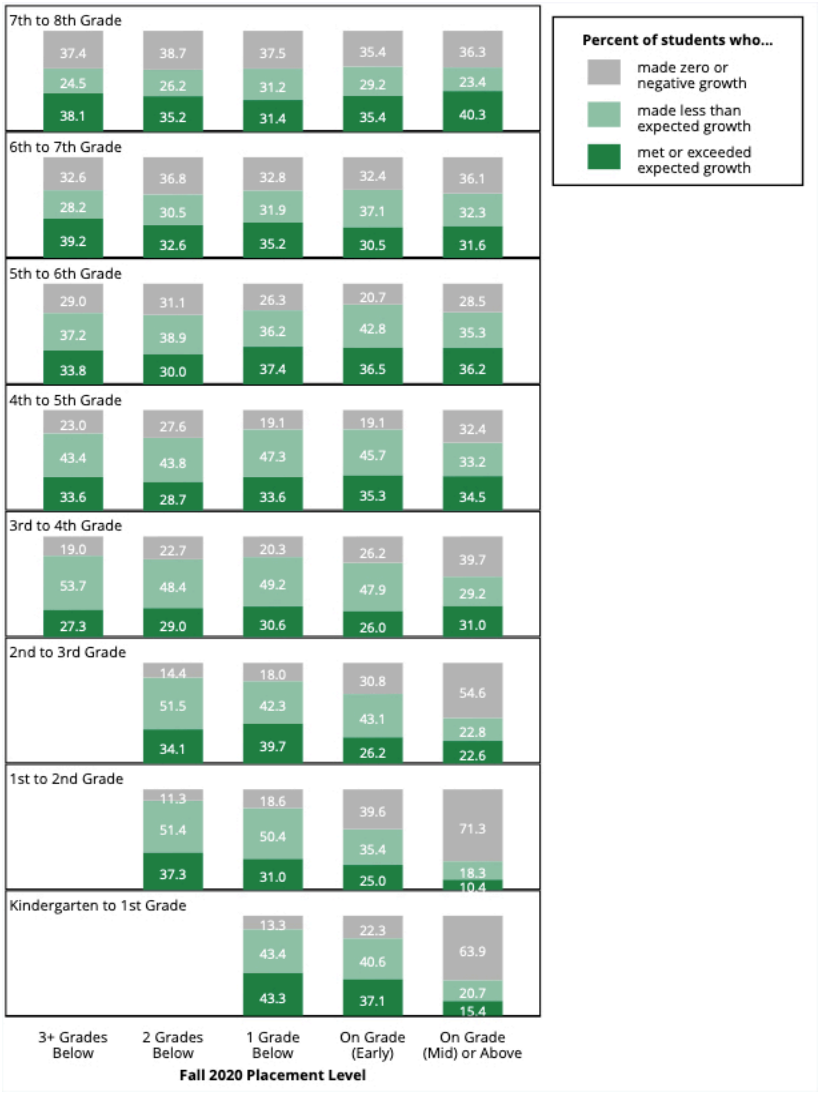
Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates i-Ready, or Renaissance Learning Star Mathematics or Reading assessments in fall 2021. The comparison points for NWEA represent the 50th percentile of the fall-to-fall conditional growth from NWEA's growth norms for students with the same fall 2020 percentile rank. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). We use the fall and spring distributions of scale scores for the Renaissance Learning norming sample to identify fall and spring medians as pre-pandemic comparison points (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Table A1.6. Trends in Average Scale Scores on Curriculum Associates' i-Ready Mathematics and Reading Assessments; Fall 2021 Sample and Michigan Norms

Grade 2020	Semester	Mathematics		Reading	
		Average	Norm	Average	Norm
K ('21)	F21	336.1	338	340.3	337
K	F20	357.7	338	374.9	337
	S21	377.7	378	402.7	399
	F21	366.4	373	389.8	398
1st	F20	380.1	373	411.7	398
	S21	399.6	407	440.6	447
	F21	389.3	397	434.9	444
2nd	F20	398.7	397	451.2	444
	S21	416.5	429	479.2	498
	F21	412.0	420	474.8	488
3rd	F20	418.7	420	486.8	488
	S21	436.9	451	509.5	525
	F21	430.6	442	505.4	520
4th	F20	436.6	442	514.6	520
	S21	453.5	471	531.5	548
	F21	447.9	461	530.4	543
5th	F20	453.1	461	537.8	543
	S21	466.9	484	551.5	568
	F21	460.9	468	548.1	554
6th	F20	467.5	468	557.4	554
	S21	477.6	487	565.0	575
	F21	470.5	479	562.7	570
7th	F20	477.8	479	570.3	570
	S21	485.9	494	577.3	591
	F21	481.2	488	577.2	584
8th	F20	487.2	488	582.8	584
	S21	493.3	499	588.8	607

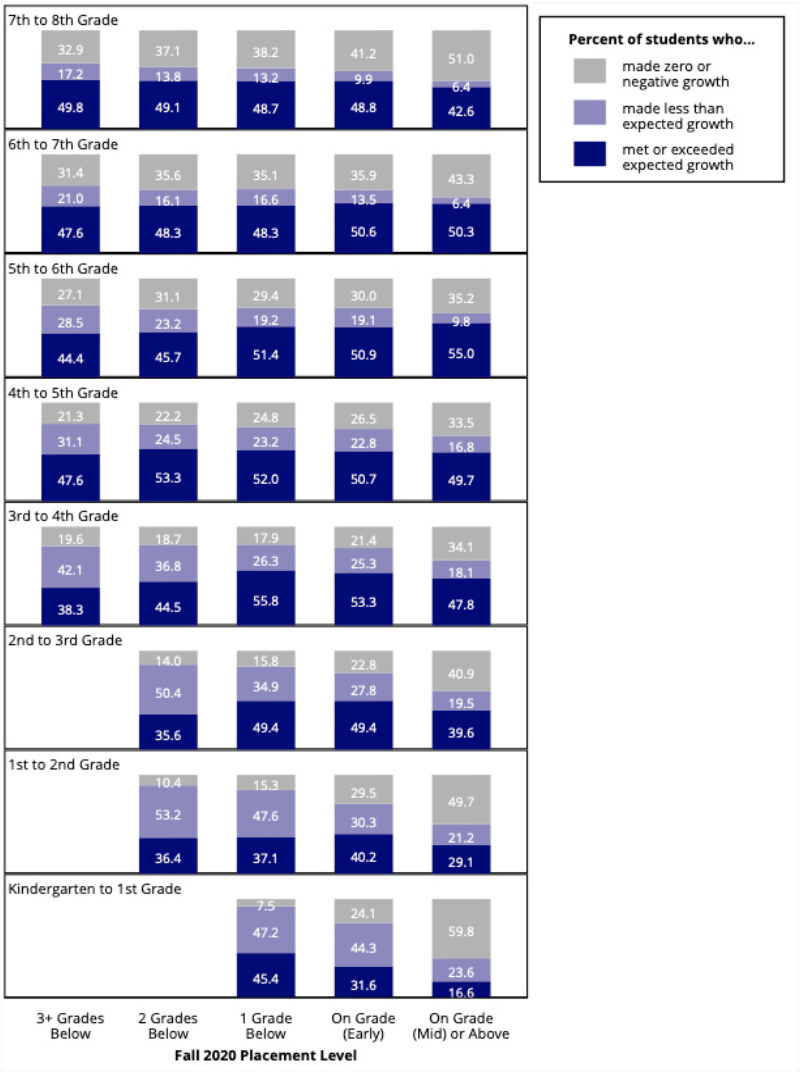
Notes: This figure uses the fall 2021 sample, which includes all Michigan students who have valid test scores on Curriculum Associates' i-Ready Mathematics or Reading assessments in fall 2021. The Curriculum Associates comparison points represent median scale scores derived from the fall and spring distributions of scale scores unique to the group of Michigan students who completed an i-Ready benchmark assessment during the 2018-19 school year. Curriculum Associates' i-Ready assessments are scored on a vertical scale which ranges from 0 to 800, however, the y-axis scales in this figure differ slightly across the math and reading assessments to reflect differences in the grade-level norms for each subject. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A2.1. Percent of Students Making Expected Fall-to-Fall Growth on Curriculum Associates' i-Ready Mathematics Assessment by Grade and Fall 2020 Placement Level



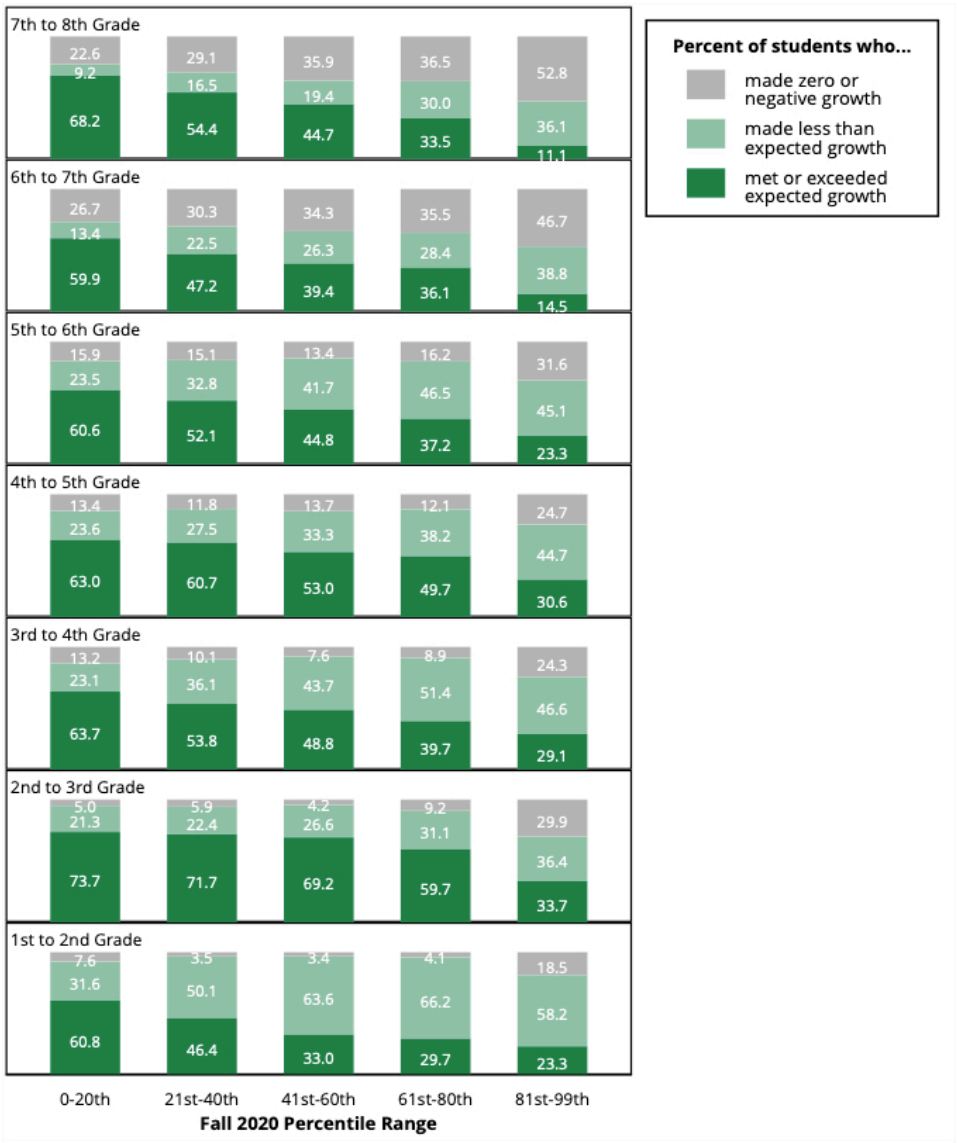
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associate's i-Ready assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject and students' initial achievement level. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). However, we note that these targets are designed to measure growth from fall to spring and may not perfectly represent growth norms from fall to fall. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A2.2. Percent of Students Making Expected Fall-to-Fall Growth on Curriculum Associates' i-Ready Reading Assessment by Grade and Fall 2020 Placement Level



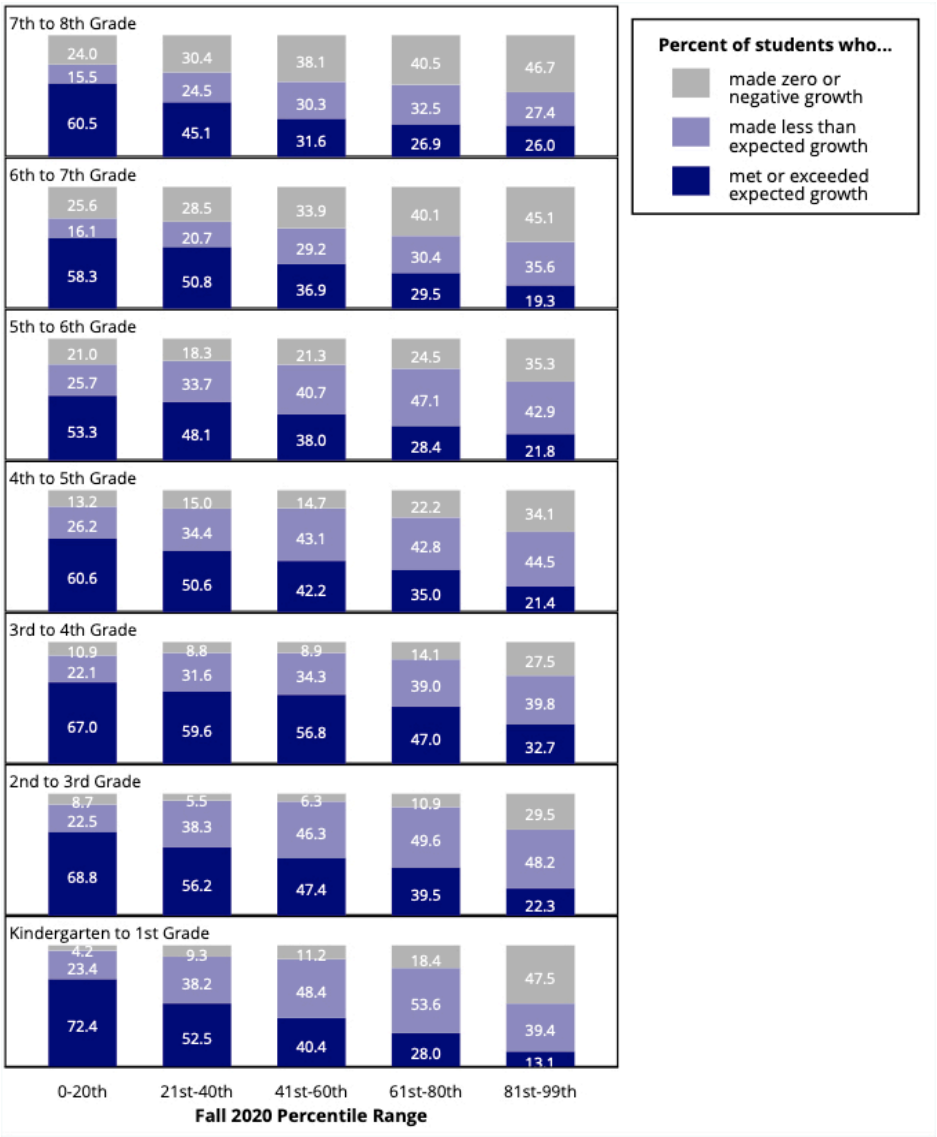
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associate's i-Ready assessments in fall 2020 and fall 2021. This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Curriculum Associate's i-Ready assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject and students' initial achievement level. For Curriculum Associates, we use typical growth targets from the i-Ready assessment growth models which indicate the median growth of students in the same grade level with the same initial placement levels (nationwide before the pandemic; Curriculum Associates, 2020). However, we note that these targets are designed to measure growth from fall to spring and may not perfectly represent growth norms from fall to fall. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A2.3. Percent of Students Making Expected Fall-to-Fall Growth on Renaissance Learning's Star Math Assessment by Grade and Fall 2020 Performance Level



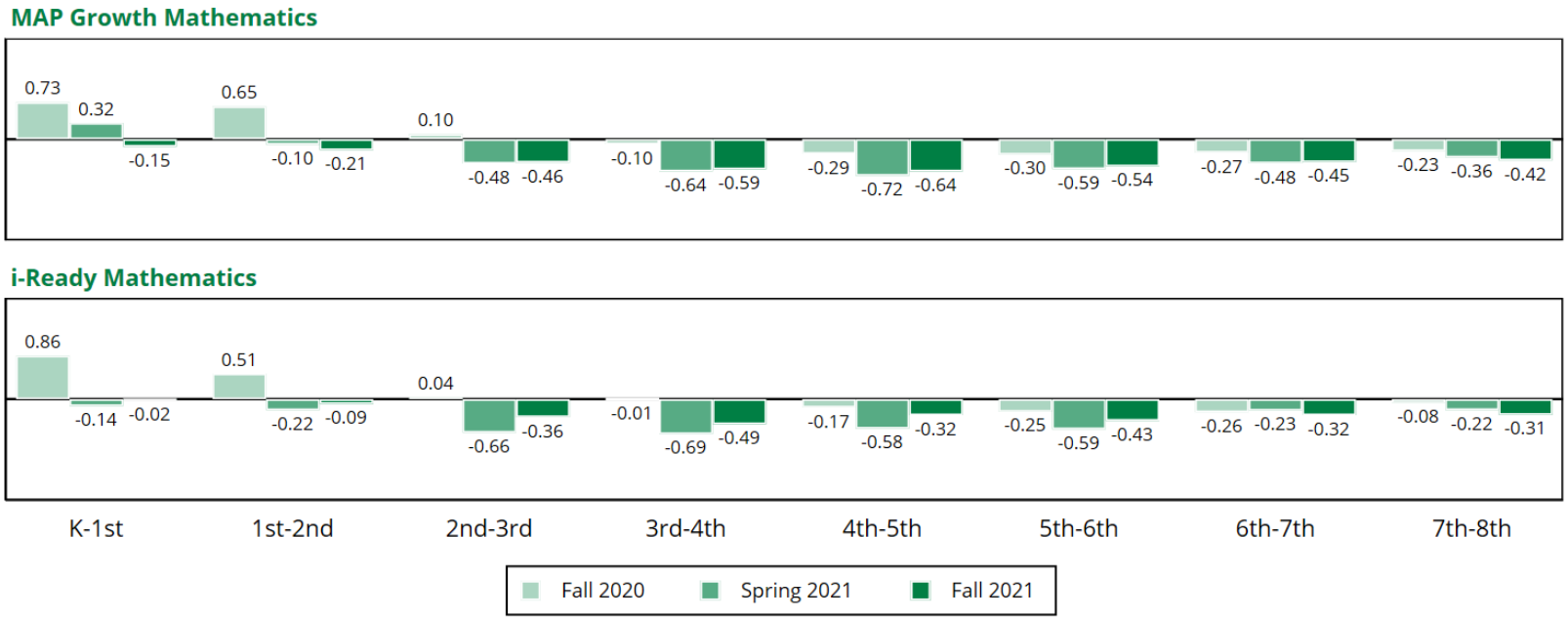
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Math, Reading, or Early Literacy assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject and students' initial achievement level. For the Star 360 we use pre-pandemic scale score distributions to identify "expected growth" as the change in scale scores necessary for a student to maintain the same percentile rank as they did in the previous year (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A2.4. Percent of Students Making Expected Fall-to-Fall Growth on Renaissance Learning's Star Reading and Early Literacy Assessments by Grade and Fall 2020 Performance Level



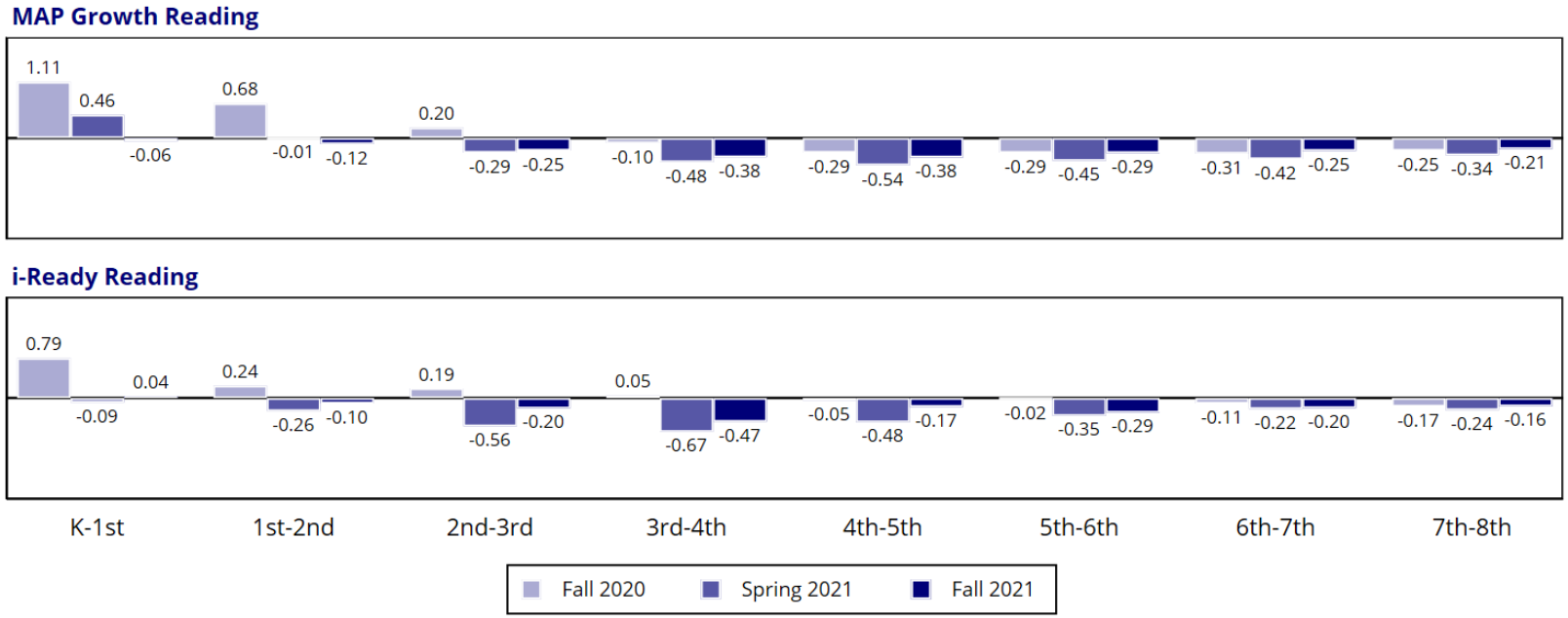
Notes: This figure uses the fall-to-fall sample, which includes all Michigan students who have valid test scores on Renaissance Learning's Star Math, Reading, or Early Literacy assessments in fall 2020 and fall 2021. Growth expectations for both subjects are calculated by dividing the difference in fall 2021 and fall 2020 scale scores by the appropriate expected scale score increase for each student. Growth norms are defined differently for each assessment provider, and they are unique to each grade level, subject and students' initial achievement level. For the Star 360 we use pre-pandemic scale score distributions to identify "expected growth" as the change in scale scores necessary for a student to maintain the same percentile rank as they did in the previous year (Renaissance Learning, 2020b, c, d). Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A3.1. Standardized Mathematics Achievement Gaps Between "Remote All Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



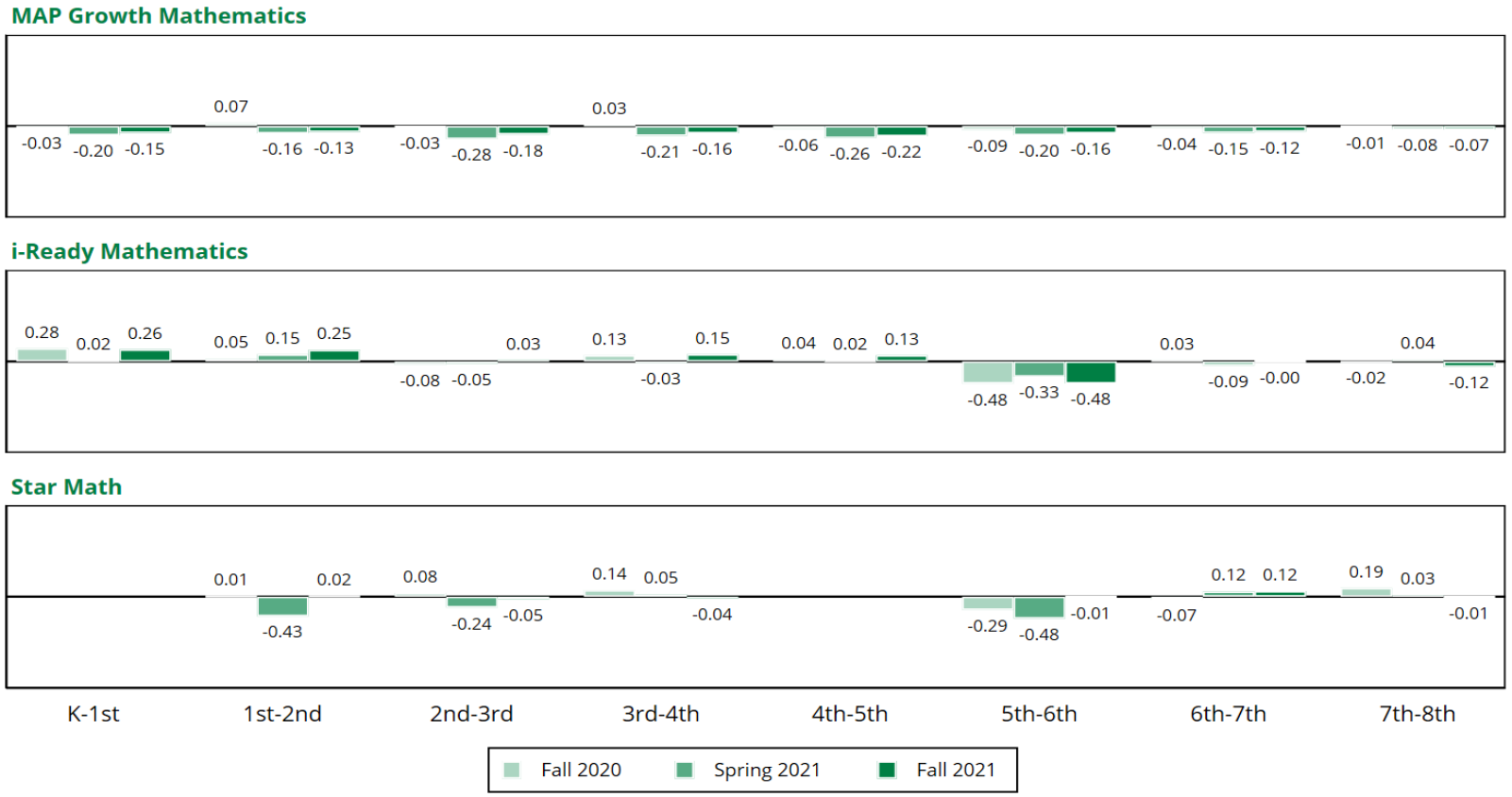
Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, or Renaissance Learning's Star assessment in fall 2020, spring 2021, and fall 2021. "Remote all year" districts offered only fully remote instruction in both the fall 2020 and spring 2021 testing periods. "In-person part-year" districts offered in-person instruction during either fall 2020 or spring 2021, and hybrid or remote instruction in the other assessment period. "Hybrid part-year" districts offered hybrid instruction in either fall 2020 or spring 2021, and fully remote instruction in the other testing period. "Hybrid all year" districts offered hybrid instruction in both fall 2020 and spring 2021. To calculate achievement gaps, we divide the difference in fall 2021 and fall 2020 average scale scores for the focal group by the standard deviation of scale scores for the reference group. Students in districts that offered in-person instruction during both the fall 2020 and spring 2021 assessment periods are the reference group. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A3.2. Standardized Reading Achievement Gaps Between "Remote All Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



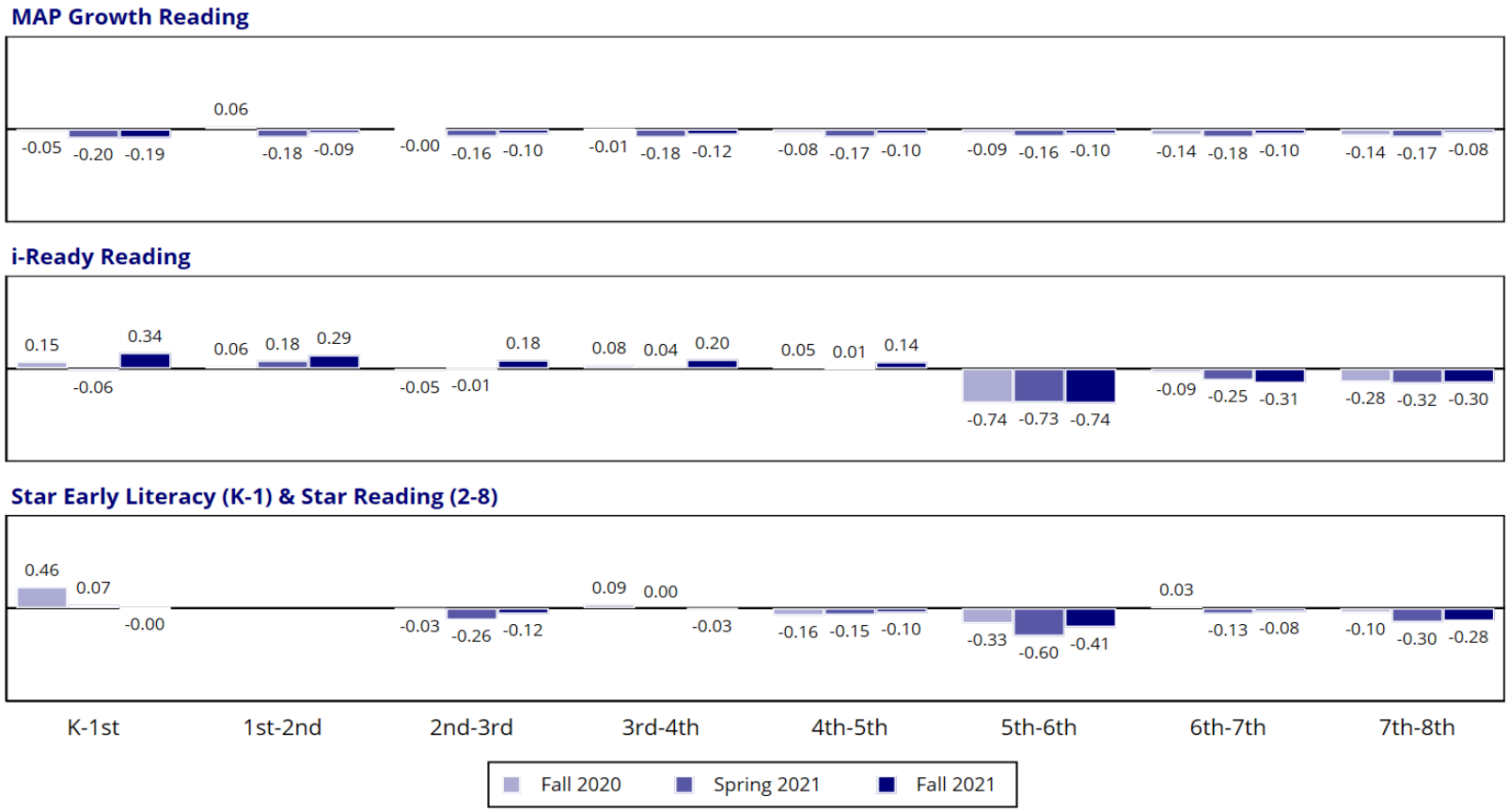
Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, or Renaissance Learning's Star assessment in fall 2020, spring 2021, and fall 2021. "Remote all year" districts offered only fully remote instruction in both the fall 2020 and spring 2021 testing periods. "In-person part-year" districts offered in-person instruction during either fall 2020 or spring 2021, and hybrid or remote instruction in the other assessment period. "Hybrid part-year" districts offered hybrid instruction in either fall 2020 or spring 2021, and fully remote instruction in the other testing period. "Hybrid all year" districts offered hybrid instruction in both fall 2020 and spring 2021. To calculate achievement gaps, we divide the difference in fall 2021 and fall 2020 average scale scores for the focal group by the standard deviation of scale scores for the reference group. Students in districts that offered in-person instruction during both the fall 2020 and spring 2021 assessment periods are the reference group. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A3.3. Standardized Mathematics Achievement Gaps Between "Hybrid All Year" and "In-Person All Year" Districts, All Assessment Providers



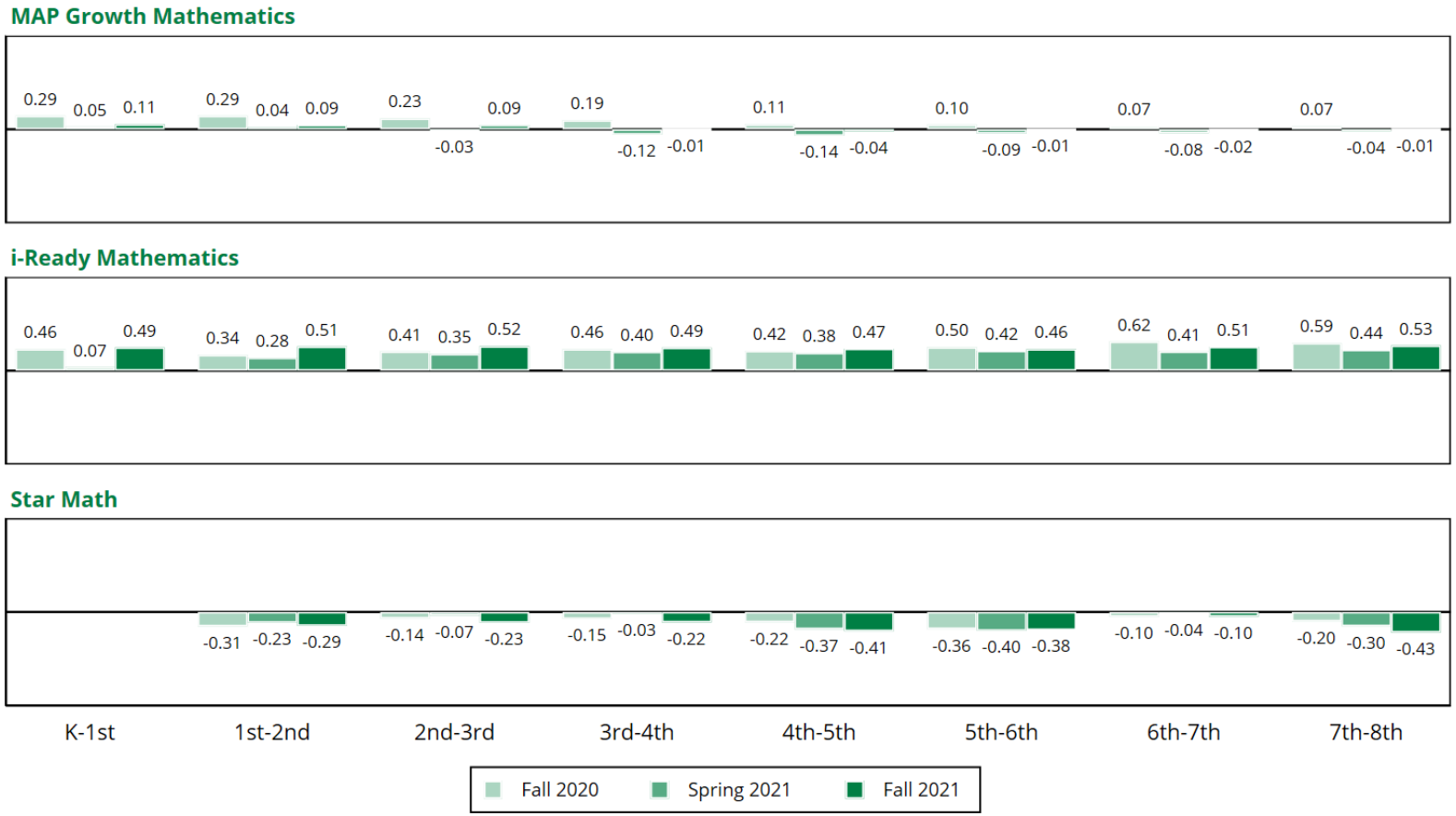
Notes: For more information about this figure, please see the notes for Figure A3.1.

Figure A3.4. Standardized Reading Achievement Gaps Between "Hybrid All Year" and "In-Person All Year" Districts, All Assessment Providers



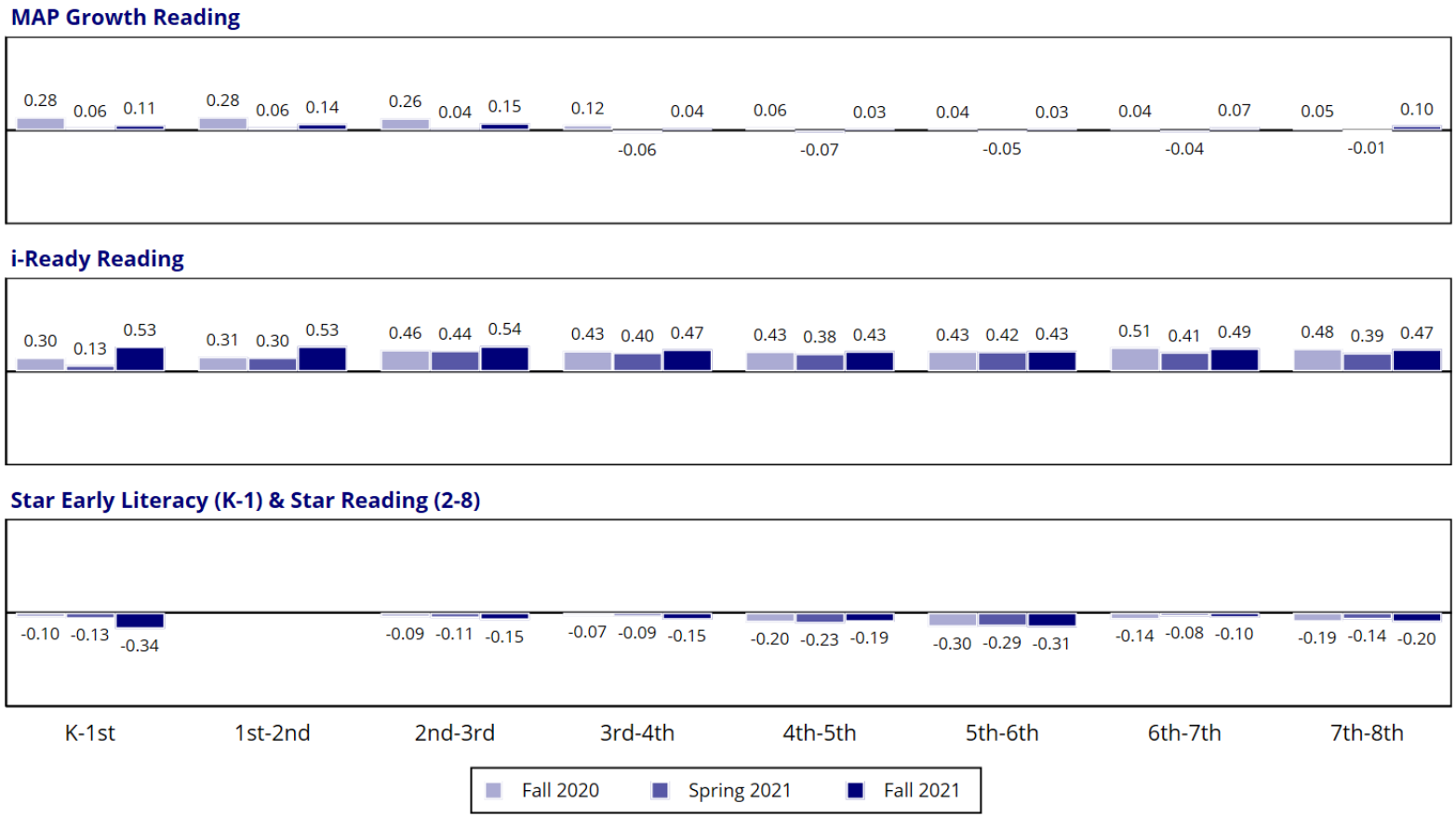
Notes: For more information about this figure, please see the notes for Figure A3.1.

Figure A3.5. Standardized Mathematics Achievement Gaps Between "In-Person Part Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



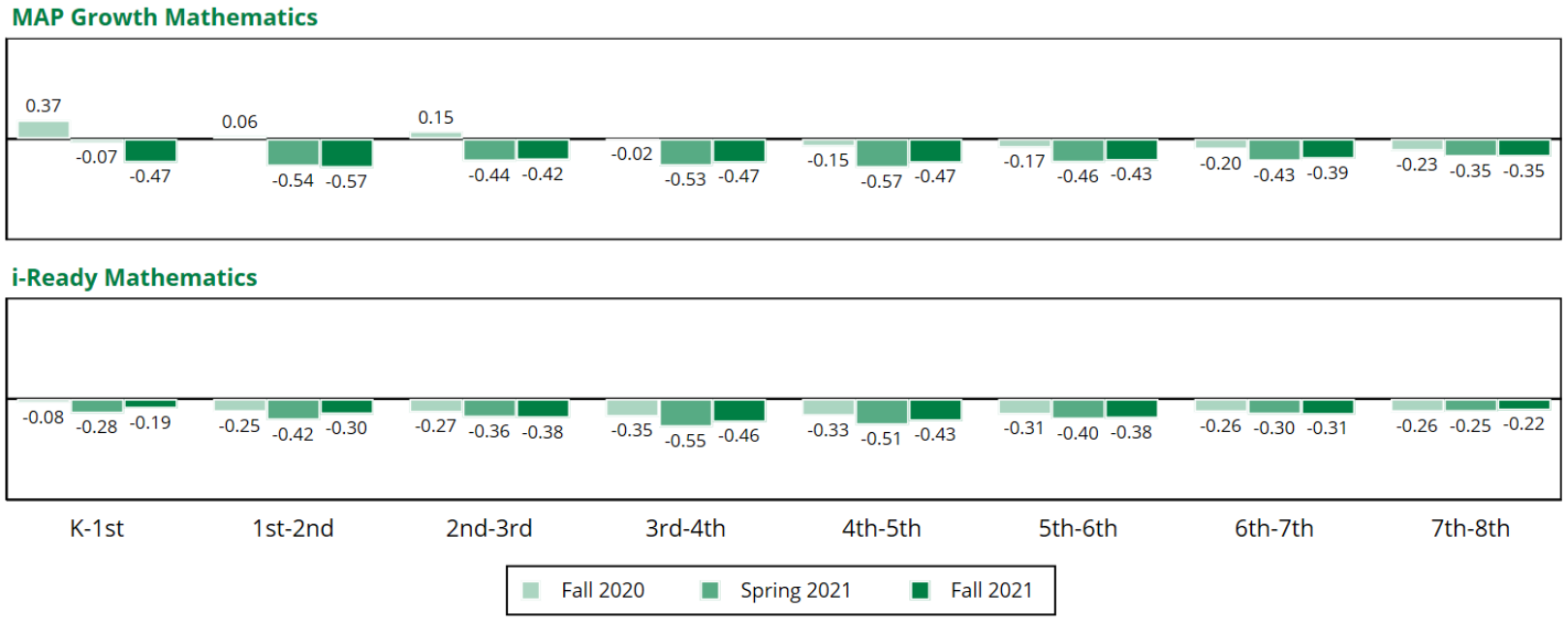
Notes: For more information about this figure, please see the notes for Figure A3.1.

Figure A3.6. Standardized Reading Achievement Gaps Between "In-Person Part Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



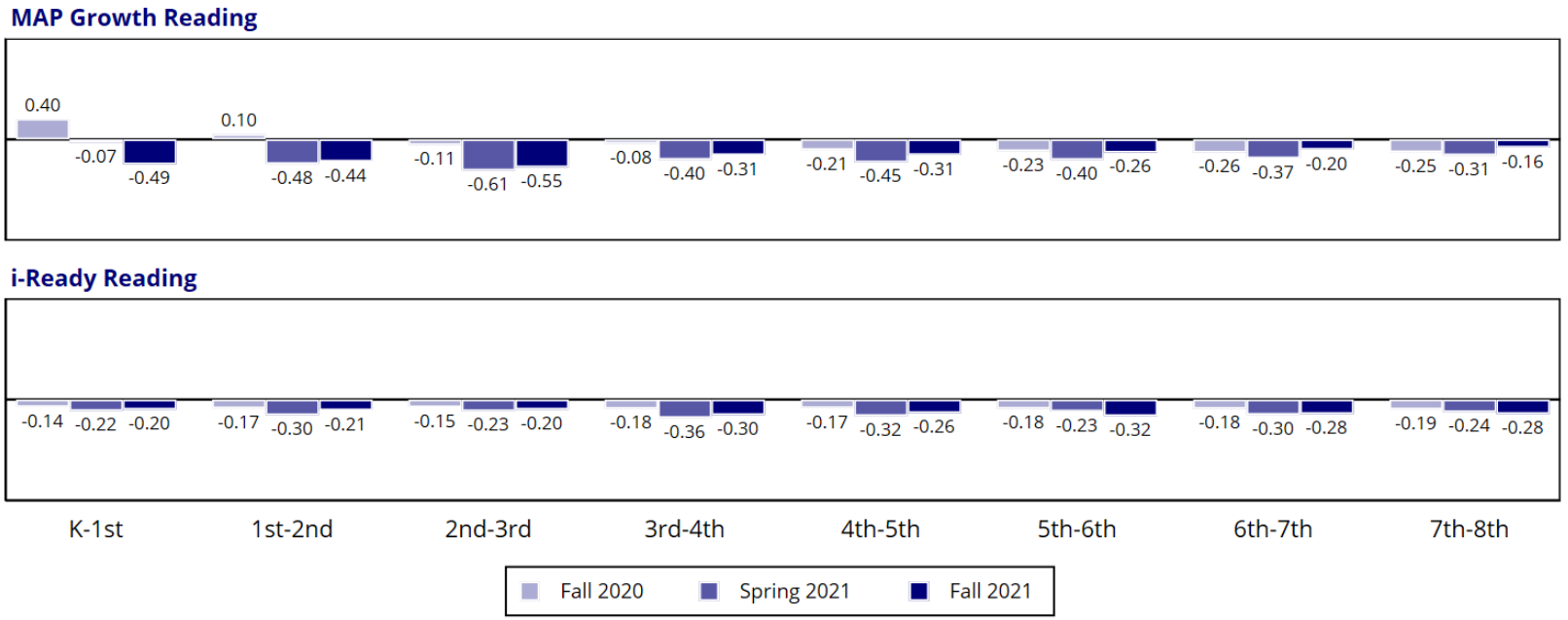
Notes: For more information about this figure, please see the notes for Figure A3.1.

Figure A3.7. Standardized Mathematics Achievement Gaps Between "Hybrid Part Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



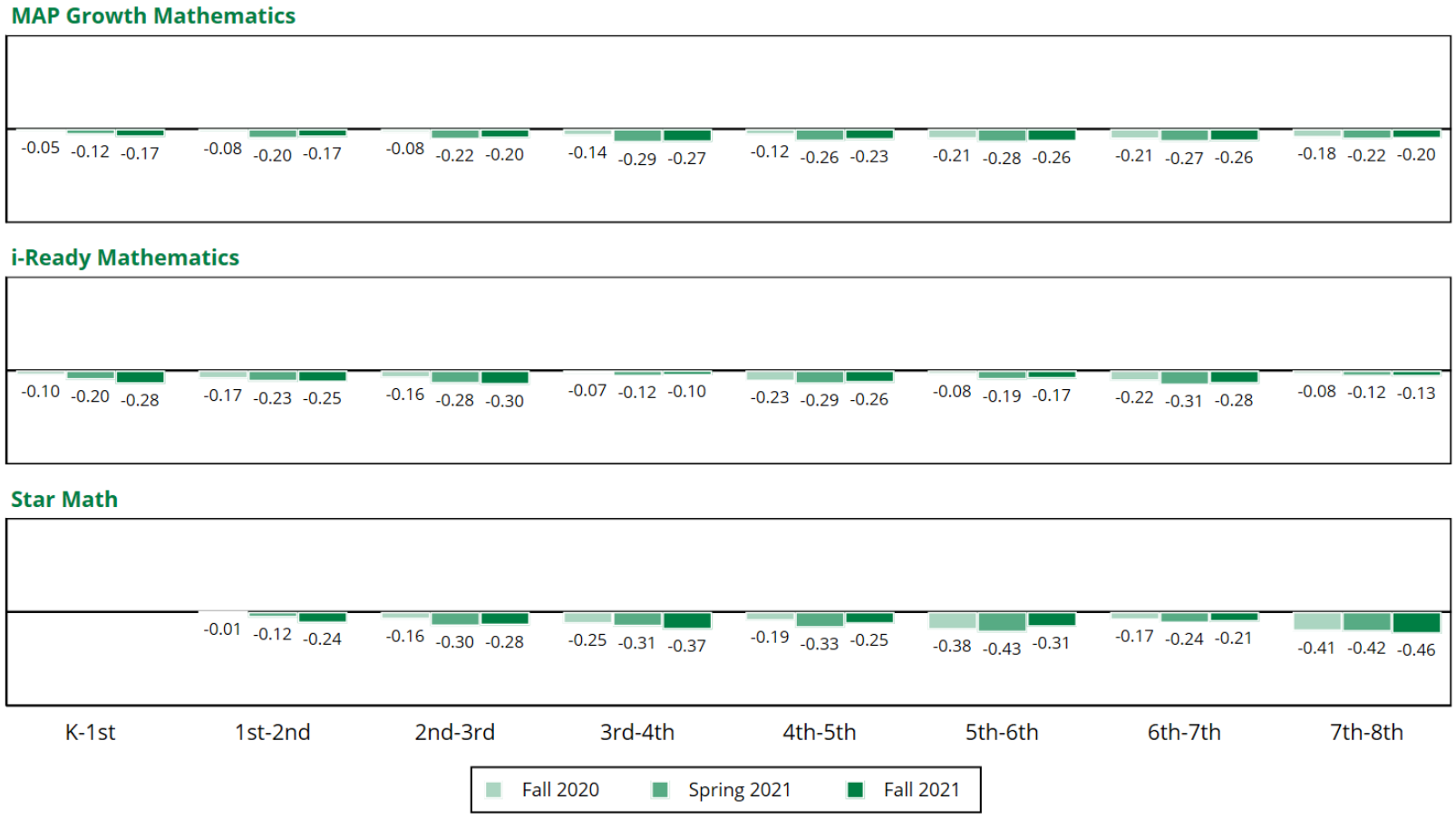
Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, or Renaissance Learning's Star assessment in fall 2020, spring 2021, and fall 2021. "Remote all year" districts offered only fully remote instruction in both the fall 2020 and spring 2021 testing periods. "In-person part-year" districts offered in-person instruction during either fall 2020 or spring 2021, and hybrid or remote instruction in the other assessment period. "Hybrid part-year" districts offered hybrid instruction in either fall 2020 or spring 2021, and fully remote instruction in the other testing period. "Hybrid all year" districts offered hybrid instruction in both fall 2020 and spring 2021. To calculate achievement gaps, we divide the difference in fall 2021 and fall 2020 average scale scores for the focal group by the standard deviation of scale scores for the reference group. Students in districts that offered in-person instruction during both the fall 2020 and spring 2021 assessment periods are the reference group. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A3.8. Standardized Reading Achievement Gaps Between "Hybrid Part Year" and "In-Person All Year" Districts; MAP Growth and i-Ready



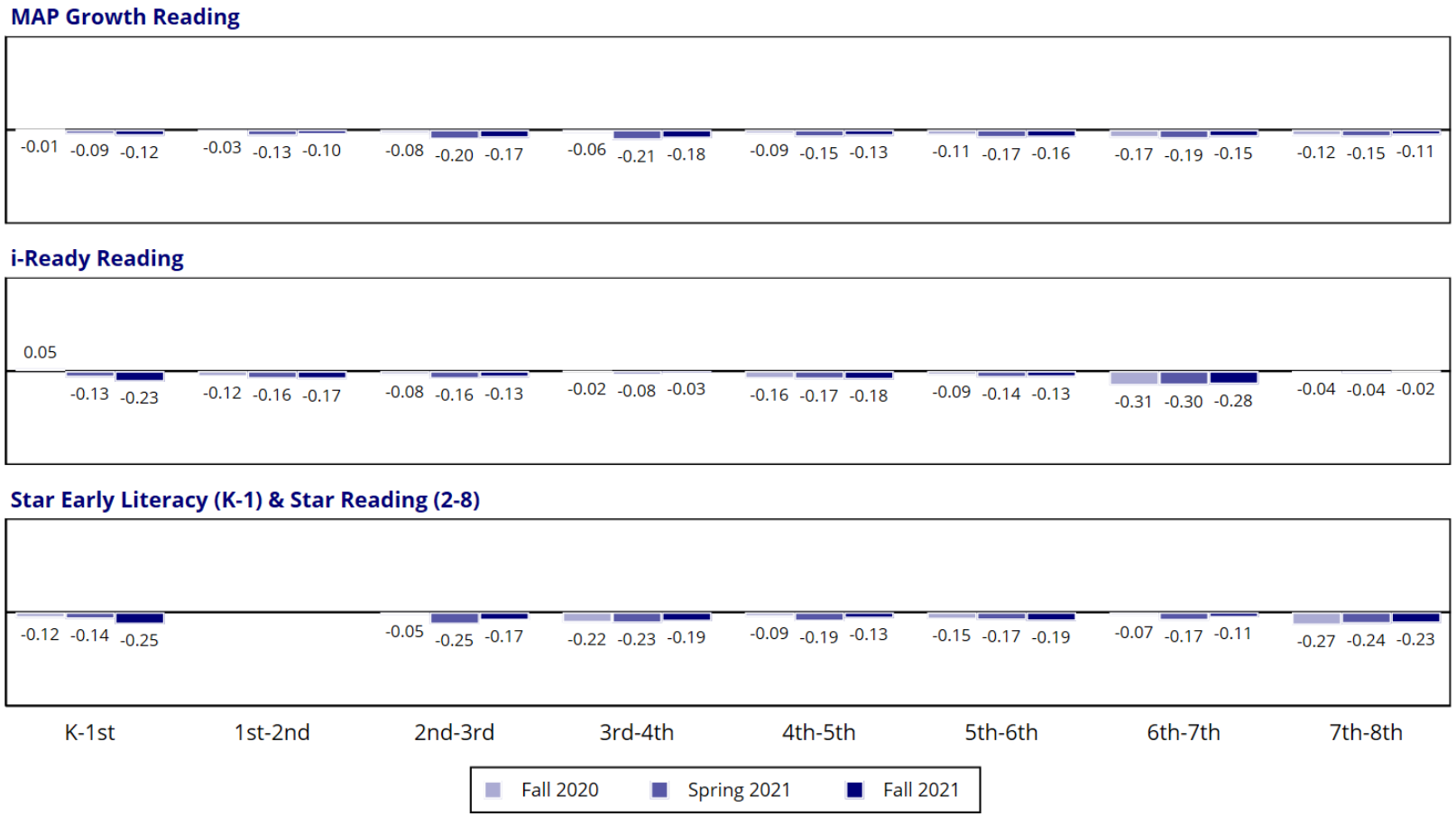
Notes: This figure uses the fall-spring-fall sample, which includes all Michigan students who have valid test scores on NWEA's MAP Growth, Curriculum Associates' i-Ready, or Renaissance Learning's Star assessment in fall 2020, spring 2021, and fall 2021. "Remote all year" districts offered only fully remote instruction in both the fall 2020 and spring 2021 testing periods. "In-person part-year" districts offered in-person instruction during either fall 2020 or spring 2021, and hybrid or remote instruction in the other assessment period. "Hybrid part-year" districts offered hybrid instruction in either fall 2020 or spring 2021, and fully remote instruction in the other testing period. "Hybrid all year" districts offered hybrid instruction in both fall 2020 and spring 2021. To calculate achievement gaps, we divide the difference in fall 2021 and fall 2020 average scale scores for the focal group by the standard deviation of scale scores for the reference group. Students in districts that offered in-person instruction during both the fall 2020 and spring 2021 assessment periods are the reference group. Source: School districts submitted benchmark assessment data directly to MDH, and these data were provided to EPIC through a collaboration between EPIC, MEDC, and MDE.

Figure A3.9. Other-White Standardized Mathematics Achievement Gaps, All Assessment Providers



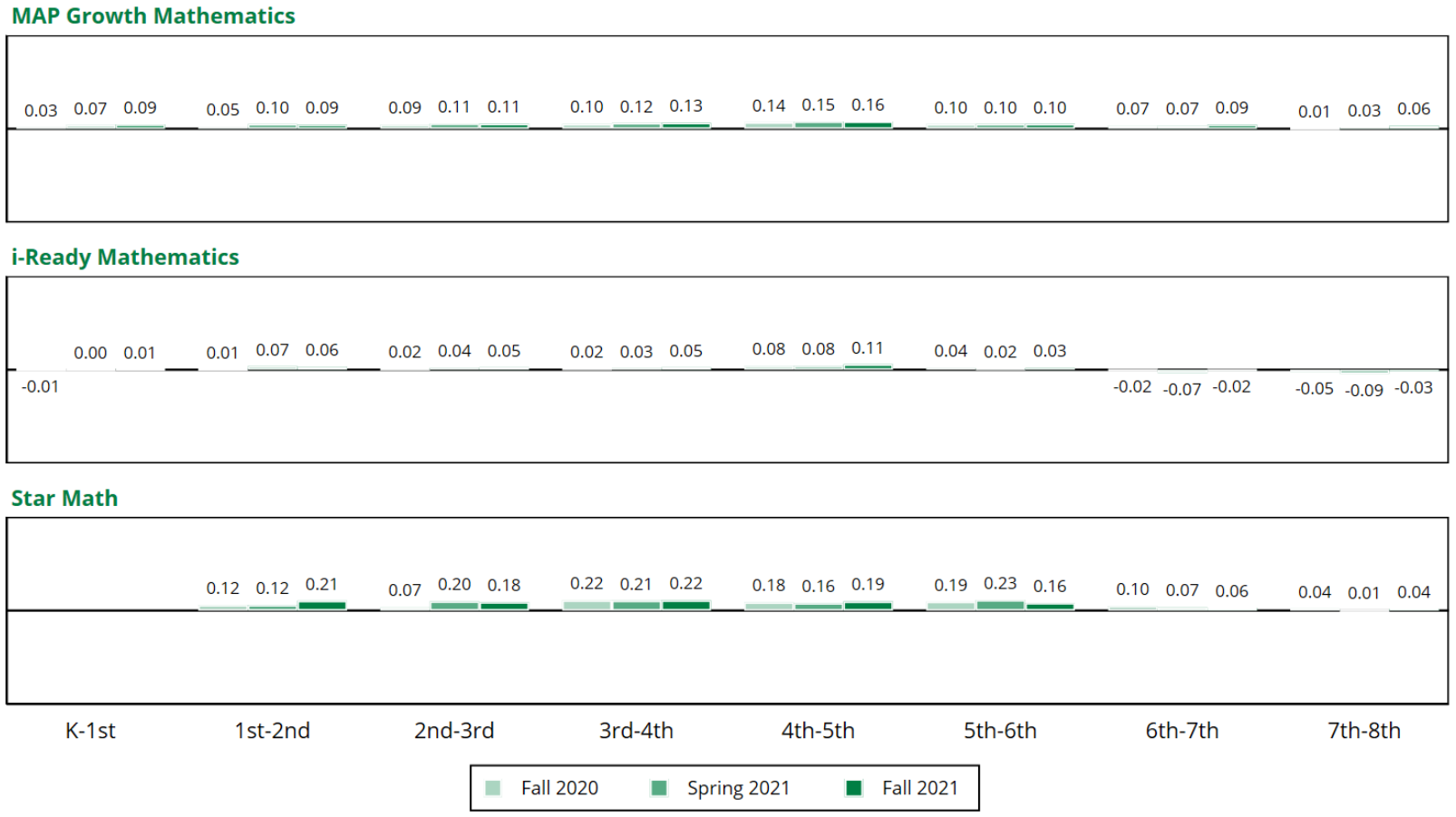
Notes: For more information about this figure, please see Report Note 2.

Figure A3.10. Other-White Standardized Reading Achievement Gaps, All Assessment Providers



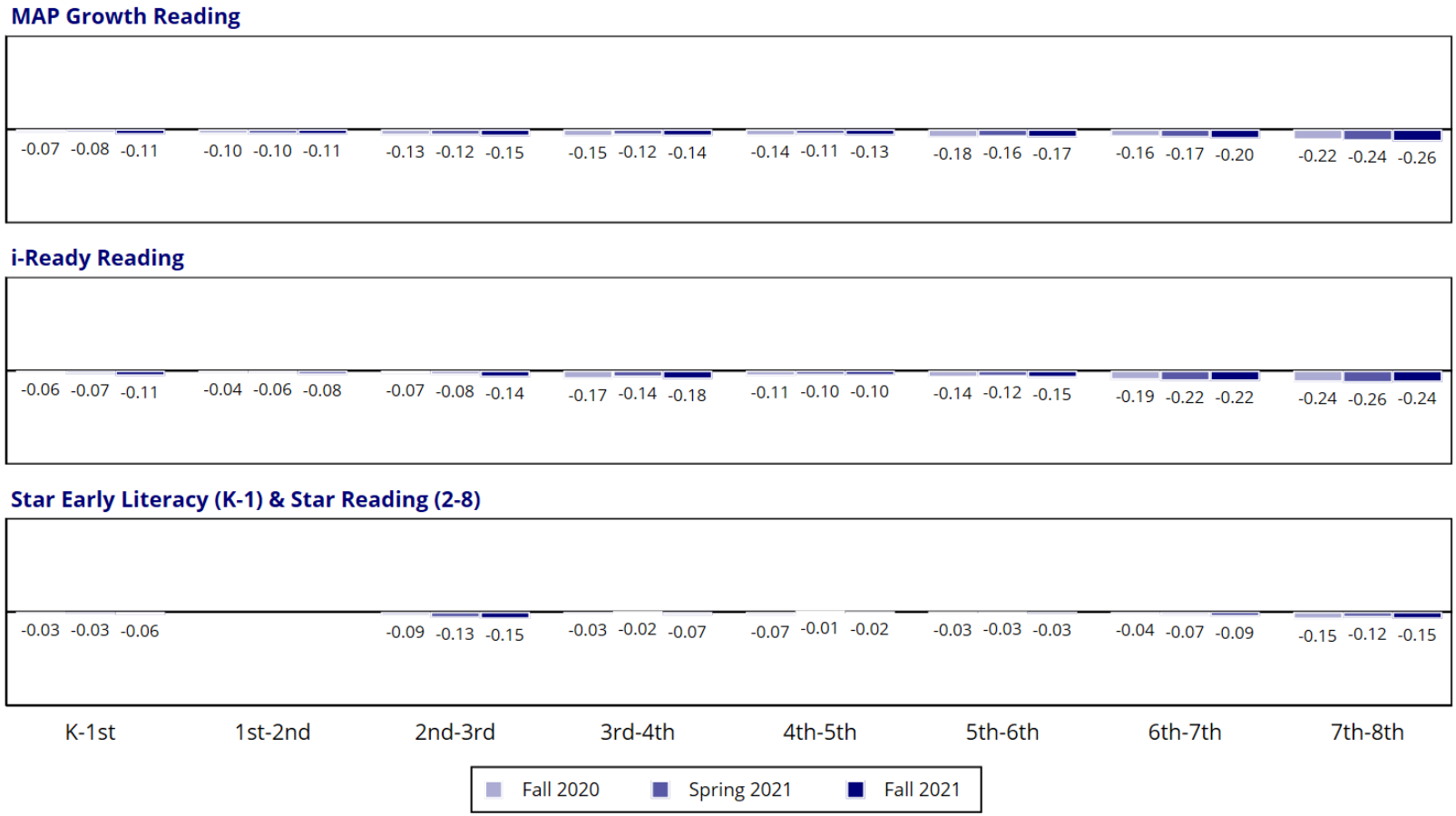
Notes: For more information about this figure, please see Report Note 2.

Figure A3.11. Male-Female Standardized Mathematics Achievement Gaps, All Assessment Providers



Notes: For more information about this figure, please see Report Note 2.

Figure A3.12. Male-Female Standardized Reading Achievement Gaps, All Assessment Providers



Notes: For more information about this figure, please see Report Note 2.

References

Amplify Education. (2021). COVID-19 means more students not learning to read. *Amplify Education Research Brief*. https://amplify.com/wp-content/uploads/2021/02/Amplify-mCLASS_MOY-COVID-Learning-Loss-Research-Brief_022421.pdf.

Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K., & Iqbal, S. A. (2020, June 18). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer*, *36*(1), 1-40. https://www.worldbank.org/en/topic/education/publication/simulating-potential-impacts-of-covid-19-school-closures-learning-outcomes-a-set-of-global-estimates

Baisley, M. E., Benjamin, N. A., Burnham, B., Danielson, K., Draughn, D., Eversman, K. A., Hernandez, R. A., Kamau, N. M., Mancini, V., Miller-Kamara, G., Munger, M., Nelson, S., Nickelsen, J., Ortiz-Crespin, J., Osuji, A., Paro-Strother, D., Radke, K., Scofield, M., & Urevig-Grilz, M. (2021, January). *Disrupted learning, COVID-19, and public education in Minnesota.* Educator Policy Innovation Center. https://educationminnesota.org/wp-content/uploads/2021/08/EPIC-Disrupted-Learning-Report.pdf

Curriculum Associates. (2020, August). i-Ready Diagnostic National Norms Tables for Reading and Mathematics for Grades K-8 2020-21.

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020a, December 8). COVID-19 and learning loss—disparities grow and students need help. *McKinsey & Company*. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-and-learning-loss-disparities-grow-and-students-need-help.

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020b, June 1). COVID-19 and student learning in the United States: The hurt could last a lifetime. *McKinsey & Company*. https://www.mckinsey.com/industries/education/our-insights/covid-19-and-student-learning-in-the-united-states-the-hurt-could-last-a-lifetime.

Gross, B. & Lake, R. (2021, August). How has the pandemic affected students' social-emotional well-being? A review of the evidence to date. *Center on Reinventing*

Public Education. https://www.crpe.org/publications/how-has-pandemic-affected-students-social-emotional-well-being-review-evidence-date.

Hopkins, B., Kilbride, T., & Strunk, K. (2021, May). Instructional delivery under Michigan districts' extended COVID-19 learning plans—May update. *Education Policy Innovation Collaborative.* https://epicedpolicy.org/ecol-reports/.

Huff, K. (2020). National data quantifies impact of COVID learning loss; raises questions about at-home testing. https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/ca-impact-of-covid-learning-loss-fall-2020.pdf.

Jack, R., Halloran, C., Okun, J., & Oster, E. (2021). Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts. https://www.nber.org/system/files/working_papers/w29497/w29497.pdf

Kilbride, T., Hopkins, B., Strunk, K., & Imberman, S. (2021, December) K-8 student achievement and achievement gaps on Michigan's 2020-21 benchmark and summative assessments. *Education Policy Innovation Collaborative.* https://epicedpolicy.org/k-8-student-achievement-and-achievement-gaps-on-michigans-2020-21-benchmark-and-summative-assessments/

Kogan, V., & Lavertu, S. (2021, January 27). *The COVID-19 pandemic and student achievement on Ohio's third-grade English language arts assessment*. John Glenn College of Public Affairs, Ohio State University. http://glenn.osu.edu/educational-governance/reports/reports-attributes/ODE_ThirdGradeELA_KL_1-27-2021.pdf

Kuhfeld, M., Lewis, K, Meyer, P., & Tarasawa, B. (2020, November). Comparability analysis of remote and in-person MAP Growth testing in fall 2020. *NWEA*. https://www.nwea.org/research/publication/comparability-analysis-of-remote-and-in-person-map-growth-testing-in-fall-2020/.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). *Test score patterns across three COVID-19-impacted school years*. Annenberg Institute Working Paper 22-521. https://doi.org/10.26300/ga82-6v47

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020, October). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher*, *49*(8), 549-565. https://doi.org/10.3102/0013189X20965918.

Kuhfeld, M., & Tarasawa, B. (2020, April). The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement. Brief. *NWEA*. https://www.nwea.org/research/publication/the-covid-19-slide-what-summer-learning-loss-can-tell-us-about-the-potential-impact-of-school-closures-on-student-academic-achievement/.

McEachin, A. & Atteberry, A. (2017). The impact of summer learning loss on measures of school performance. *Education Finance and Policy, 12*(4), 468-491. https://direct.mit.edu/edfp/article/12/4/468/10289/The-Impact-of-Summer-Learning-Loss-on-Measures-of

Michigan Public Act 147 of 2020, Mich., MCL § 388.1621f (2020) http://www.legislature.mi.gov/(S(rhqyboiv2ivksh4xj0wbiqpl))/mileg.aspx?page=GetObject&objectname=mcl-388-1621f.

Michigan Public Act 148 of 2020, Mich. MCL § 388.1701 (2020) http://www.legislature.mi.gov/(S(hotihihl5cde5dkkjzjvx4g3))/mileg.aspx?page=getObject&objectName=mcl-388-1701.

Michigan Public Act 149 of 2020, Mich., MCL § 388.1606 (2020) http://legislature.mi.gov/doc.aspx?mcl-388-1606.

Michigan Public Act 48 of 2021, Mich., MCL § 388.1704a (2021) http://legislature.mi.gov/doc.aspx?mcl-388-1704a.

North Carolina State Board of Education. (2022, March). Report to the North Carolina General Assembly: An impact analysis of student learning during the COVID-19 pandemic. https://content.govdelivery.com/attachments/NCSBE/2022/03/02/file_attachments/2091616/JLEOC%20Report%20HB196.%20Impact%20on%20Lost%20Instructional%20Time%20for%20SBE%20March.pdf

Pier, L., Hough, H. J., Christian, M., Bookman, N., Wilkenfeld, B., & Miller, R. (2021, January 25). COVID-19 and the educational equity crisis: Evidence on learning loss from the CORE Data Collaborative. *PACE*. https://edpolicyinca.org/newsroom/covid-19-and-educational-equity-crisis

Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: Updating and extending past research with new national data. *Educational Researcher*, *45*(8), 443-453. https://doi.org/10.3102/0013189X16677965

Regional Education Laboratory. (2021, November). Fact sheet: Test scores, grades, and student engagement during the COVID-19 pandemic. Mathematica. https://ies.ed.gov/ncee/edlabs/regions/midatlantic/app/Docs/Infographics/RELMA_TestScoresGradesStudentEngagementCOVID-19_508c.pdf

Renaissance Learning. (2021a). How kids are performing: Tracking the school-year impact of COVID-19 on reading and mathematics achievement. https://renaissance.widen.net/s/t8trbrt2tc/r63444

Renaissance Learning. (2021b). *Renaissance*: *Star assessments for early literacy technical manual*. https://help2.renaissance.com/US/PDF/SEL/SELRPTechnicalManual.pdf?_gl=1*1svxhgt*_gcl_aw*R0NMLjE2NDk3ODA2MDIuQ2owS0NRand4dFNTQmhEWUFSSXNBRW4wdGhRblFtQnREMnEzUnNBWVNZS25XVzBGMTBRdkdvenQ4aUE1cUJxSnltUlU2aHRkaDU4NUdZZ2FBbl9zRUFMd193Y0I.&_ga=2.169641829.592103100.1649780394-1829603977.1649780394&_gac=1.213792102.1649780602.Cj0KCQjwxtSSBhDYARIsAEn0thQnQmBtD2q3RsAYSYKnWW0F10QvGozt8iA5qBqJymRU6htdh585GYgaAn_sEALw_wcB&utm_source=google&utm_medium=cpc&utm_campaign=Star_Assessment_NonBrand

Renaissance Learning. (2021c). *Renaissance: Star assessments for math technical manual.* https://help2.renaissance.com/US/PDF/SM/SMRPTechnicalManual.pdf?_gl=1*1svxhgt*_gcl_aw*R0NMLjE2NDk3ODA2MDIuQ2owS0NRand4dFNTQmhEWUFSSXNBRW4wdGhRblFtQnREMnEzUnNBWVNZS25XVzBGMTBRdkdvenQ4aUE1cUJxSnltUlU2aHRkaDU4NUdZZ2FBbl9zRUFMd193Y0I.&_ga=2.169641829.592103100.1649780394-1829603977.1649780394&_gac=1.213792102.1649780602.Cj0KCQjwxtSSBhDYARIsAEn0thQnQmBtD2q3RsAYSYKnWW0F10QvGozt8iA5qBqJymRU6htdh585GYgaAn_sEALw_wcB&utm_source=google&utm_medium=cpc&utm_campaign=Star_Assessment_NonBrand

Renaissance Learning. (2021d). *Renaissance: Star assessments for reading technical manual.* https://help2.renaissance.com/US/PDF/SR/SRRPTechnicalManual.pdf?_gl=1*1svxhgt*_gcl_aw*R0NMLjE2NDk3ODA2MDIuQ2owS0NRand4dFNTQmhEWUFSSXNBRW4wdGhRblFtQnREMnEzUnNBWVNZS25XVzBGMTBRdkdvenQ4aUE1cUJxSnltUlU2aHRkaDU4NUdZZ2FBbl9zRUFMd193Y0I.&_ga=2.169641829.592103100.1649780394-1829603977.1649780394&_gac=1.213792102.1649780602.Cj0KCQjwxtSSBhDYARIsAEn0thQnQmBtD2q3RsAYSYKnWW0F10QvGozt8iA5qBqJymRU6htdh58

5GYgaAn_sEALw_wcB&utm_source=google&utm_medium=cpc&utm_campaign=Star_Assessment_NonBrand

Sass, T., & Goldring, T. (2021, May). Student achievement growth during the COVID-19 pandemic: Insights from metro-Atlanta school districts. *Georgia Policy Labs, Andrew Young School of Policy Studies, Georgia State University*. https://gpl.gsu.edu/download/student-achievement-growth-during-the-covid-19-pandemic-report/?ind=1620902643226&filename=Achievement%20Growth%20During%20COVID-19_Report_20210512_FINAL.pdf&wpdmdl=2105&refresh=609e93df82b691621005279

Smarter Balanced Validity Research. (2020, October 14). Percentiles. Retrieved from https://validity.smarterbalanced.org/reports-and-specifications/percentiles/

Thum, Y. M. & Kuhfeld, M. (2020, July). NWEA 2020 MAP growth: Achievement status and growth norms for students and schools. *NWEA*. https://teach.mapnwea.org/impl/normsResearchStudy.pdf

von Hippel, P. T., & Hamrock, C. (2019). Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science*, *6*, 43-80. https://sociologicalscience.com/articles-v6-3-43/