WestEd
WestEd.org

Early Literacy Performance in Massachusetts

Results of Ongoing Analysis of Literacy Screening Assessments (2023/24 Data)

Mariann Lemke, Molly Faulkner-Bond, Brandon LeBeau, Aaron Soo Ping Chow, Sarah Quesen

Fall 2025

Suggested citation: Lemke, M., Faulkner-Bond, M., LeBeau, B, Soo Ping Chow, A., & Quesen, S. (2025). *Early literacy performance in Massachusetts: Results of ongoing analysis of literacy screening assessments (2023/24 data)*. WestEd.

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, increasing opportunity, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit WestEd.org. For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at WestEd.org/subscribe.

Contents

Executive Summary

Since the 2020/21 school year, the Massachusetts Department of Elementary and Secondary Education (DESE) has collected and analyzed early literacy screening assessment data from schools and districts that participated in targeted state grants. In 2023/24, the 1st year that screening was required for all kindergarten through grade 3 (K–3) students statewide, data were collected for nearly half the state's K–3 population.

Using 2023/24 and longitudinal data, this report examines patterns of early literacy performance and progress across student groups, grades, and schools. The findings reaffirm patterns seen in prior years—such as the persistence of risk and disparities across student populations—while also offering new information on benchmark comparability, multilingual learners, and school-level variation in outcomes.

Data Used in 2023/24 Analysis

- The data available in 2023/24 doubled from 2022/23.
- The data include 388,000 scores from 137,000 students, about 53 percent of the state's K–3 student population.
- Scores are from 612 schools and 144 districts, about half of schools serving K–3 students.
- The data are from 13 screening assessments.
- Six screening assessments account for 92 percent of the data.

Key Findings

- **Student benchmark performance is consistent over time.** About one quarter to one third of students were identified as significantly below benchmark at any given time period (beginning, middle, or end of year), and between 16 and 19 percent were significantly below benchmark in all three time periods. These rates have been consistent over 3 years of analysis.
- **Inequities in literacy performance persist.** Students from low income backgrounds; English learners; students receiving special education services; and Black, Hispanic, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander students were more likely to be identified as significantly below benchmark than were their peers not in those groups. For example, students from low income backgrounds were 2.5 times as likely to be repeatedly classified as at significant risk than were students without low income backgrounds. Students from backgrounds that include membership in multiple historically underserved groups were even more frequently

identified as at risk, although there is variation across groups. For example, students from low income backgrounds in all racial and ethnic groups were more often classified as significantly below benchmark, but Hispanic White students were about 1.8 times as likely to be at significant risk than were non–Hispanic White students. Prior results show these same patterns of performance.

- **Early literacy performance tends to predict later performance.** Students who started the school year significantly below benchmark tended to remain so throughout the year and across grade levels, and students who met reading benchmarks also tended to stay at those levels. More than half the students identified as at significant risk at the beginning of the year were still at that level at the end of the year. However, students who improved their performance beyond the significant risk threshold typically maintained higher scores on screening assessments and the grade 3 Massachusetts Comprehensive Assessment System (MCAS) test. Also, students in earlier grades consistently showed a lower likelihood of remaining at significant risk from year to year compared with those in higher grades. These findings also mirror past results.
- **Meeting benchmarks early is associated with better longer term achievement.** This year's analysis followed students identified as at significant risk at the beginning of grade 1 through grade 3. Students who met reading benchmarks by the end of grade 1 were at least 1.3 times as likely to meet expectations on the MCAS English language arts (ELA) assessment in grade 3 as were students who met benchmarks later, and they were at least 5 times as likely to meet expectations as students who never met benchmarks. These and similar previous findings underscore the need for early, targeted interventions to disrupt trajectories of reading difficulty.

  At the same time, analysis has repeatedly shown that meeting screening assessment benchmarks is not always sufficient for meeting expectations on MCAS. In many cases, even students who met screening assessment benchmarks did not meet expectations on MCAS, pointing to the need for comprehensive literacy instruction that addresses higher-level ELA and writing skills not typically measured by screening assessments.
- **Students just above and below significant risk threshold perform similarly on MCAS.** New exploratory analysis this year focused on students just above and below risk thresholds to examine differences in later performance. Students just below risk thresholds were identified as needing intervention and support, and students just above risk thresholds were not identified as needing such supports. There were no observed differences between these groups in later MCAS performance, although this finding could be due to differences between screening assessments and MCAS (as mentioned in the finding above). Or it could simply suggest that any supports being provided are helping students keep up but not gain ground compared with their peers.
- **School factors influence literacy outcomes.** Previous analyses showed that school environments are related to rates of risk identification, and this is shown in the current

results also. School-level factors like higher concentrations of students from low income backgrounds correlated with the increased likelihood of repeated risk identification. Zooming out to the district level this year—through an analysis of district-level school segregation measures—showed that historically underserved students in more racially segregated contexts were slightly more likely to be identified as significantly below benchmark multiple times. However, some schools do outperform the risk rates predicted by statistical models that take into account student demographics and resources, suggesting avenues for potential further study.

- **Screening results for the same students from different assessments can vary considerably.** In any time period, about 3 percent to 4 percent of students took more than one screening assessment. Some students took assessments in English and Spanish (mostly Star assessments). Some took assessments that targeted different grade or skill levels or that varied in length or mode of administration. Rates of agreement between screening assessments—that is, whether students would be considered at risk of reading difficulty—varied from 47 percent to 85 percent depending on screening assessment combinations.

  Among English learners taking screening assessments in English and Spanish, patterns of performance varied by language. In English, about 21 percent of students who performed above the significant risk threshold at the beginning of the year were still there by the end of the year, and another 19 percent improved to that level. In Spanish, fewer students (11%) improved beyond the significant risk level. This difference may be related to the fact that most students are being taught primarily in English, and Spanish language development outside of school may not align with the types of skills included on screening assessments. Taken together, these findings suggest that schools and districts carefully consider whether administering multiple screening assessments provides new information that supports student learning.

Some caution is needed in generalizing from these results given differences between screening assessments and between screening assessment data samples over time. Addressing these challenges could improve DESE's ability to develop a comparable system for reporting and analysis. Nonetheless, although exact numbers may vary depending on the benchmarks used in analysis or characteristics of the available data, patterns of performance over 3 years of analysis have been remarkably consistent. Although universal screening provides valuable data that can be used to identify students who need support, current interventions are not yet addressing these challenges fully. Continued efforts to investigate and support effective school practices are essential for advancing literacy outcomes statewide.

Introduction

Over the past few years, the Commonwealth of Massachusetts, like many other states, has implemented a series of policies and programs aimed at improving the literacy skills of young children. A change to state regulation in 2022 required elementary schools to assess each student's reading abilities and early literacy skills at least twice per year from kindergarten through at least grade 3 (K–3), beginning in the 2023/24 school year.

The early literacy screening regulation also requires that schools determine how to meet the needs of students whose screening results are "significantly below relevant benchmarks for age-typical development" based on use of a "valid, developmentally appropriate, screening instrument approved by the Department."

To assist schools and districts in implementing the regulation, the Massachusetts Department of Elementary and Secondary Education (DESE) published guidance in June 2023 that includes information on selecting and administering screening assessments and interpreting their results in the context of the regulation. The guidance also provides information on considerations for screening students with disabilities and multilingual learners, as well as information on the relationship between screening and identification of dyslexia. DESE's goal is to support schools in implementing an effective screening and data-based decision-making process that will identify students at risk of reading difficulty early and provide evidence-based and effective supports. Even before the amendment to the regulation, DESE had begun supporting early literacy screening by providing a list of vetted literacy screening assessments to schools and districts and offering grant funding to purchase and implement them.

Beginning with the 2020/21 school year, DESE began collecting literacy screening assessment data from schools and districts participating in certain state grants. DESE contracted with WestEd, a national nonpartisan research and service organization, to analyze these data. Following on earlier analysis, this report uses data from 2023/24 and earlier years to address key questions about the extent to which students are reaching critical early reading benchmarks and how performance and progress differ across student groups and schools.[1] The specific research questions addressed in this year's report are in the next section. Several associated issue briefs provide additional information on related topics.

[1] Earlier reports and issue briefs are available on the DESE website Early Reading Performance in Massachusetts.

Research Questions

Screening Assessment Data Overview

- How many benchmark scores are available overall and for each assessment?
- How many students, schools, and districts are represented by the data in 2023/24?
- How many benchmark scores are available by grade and student group? To what extent does the sample of students with available benchmark scores represent the overall grade K–3 student population in the state?
- How many benchmark scores are available by time period in 2023/24?

Student Performance

- How many students were identified as significantly below benchmark by time period, grade, and student group?
- How many students were identified as significantly below benchmark two or three times overall and by grade and student group?
- How do student and school background factors interact and relate to the likelihood of students being identified as significantly below benchmark more than once?
- Which students are tested "off grade level"? How does their performance compare with that of other students?

Student Progress

- How does student performance change as the school year progresses? Do students identified as significantly below benchmark remain significantly below benchmark?
- How does student progress vary by grade and student background characteristics?
- How does student performance change across grade levels? Do students at risk remain at risk across years?
- How does student performance change over the summer?
- What is the relationship between screening assessment scores and MCAS performance in grade 3?
- How do students identified as at risk in grades 1, 2, and 3 perform on MCAS in later grades (e.g., 3, 4, and 5)? Among students identified as at risk early (e.g., beginning of grade 1), how does MCAS performance in grade 3 differ between those who remain at risk over time and those who reach benchmark?

School Characteristics and Student Performance

- How do school-level characteristics relate to the risk of being identified as at significant reading difficulty multiple times?
- To what extent does district-level school segregation (as measured by indices of concentration of poverty and race/ethnicity) moderate the relationship of school-level factors to student performance?
- How does MCAS performance differ for those just below compared with those just above screener benchmarks identifying students as at risk?
- Are there schools where students are "outperforming" expectations in terms of achievement and/or growth within the school year, based on the demographics of students and/or schools or other factors?

English Learner Student Performance

- How many students participate in Spanish-language screening assessments?
- What Spanish-language screening assessments are used in the state?
- What are the characteristics of students who participate in Spanish-language screening assessments?
- To what extent are students who participate in Spanish-language screening assessments also being taught in Spanish?
- To what extent are through-year score patterns on Star Spanish-language assessments similar to patterns on English-language assessments?
- To what extent do Spanish-language screening assessment outcomes agree with English-language versions? When the results of each screener diverge, are there any patterns to the disagreement?

This report draws mainly on early literacy screening assessment and other administrative data. More details on the specific screening assessments used and details on analytic approaches are provided in the section on data and methods, followed by the results of analysis for each of the research questions and a summary discussion of limitations and findings.

Data and Methods

This report uses data from multiple sources, including extant student-level data provided by DESE and publicly available school- and district-level data obtained from DESE's school and district profiles website and other sources. The following sections provide an overview of data included in analysis and how it is used:

- early literacy universal screening assessment data for K–3 students in districts that receive specific state grants and programs
- the state's Student Information Management System (SIMS) data
- MCAS data
- Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs, or ACCESS) data
- school- and district-level data from DESE and other sources

More detail on data processing is provided in Appendix A.

Early Literacy Universal Screening Assessment Data

As part of early efforts to encourage screening of students for potential reading difficulties and early intervention, DESE approved a set of early literacy screening assessments developed by various publishers for use in the state, although their use was not required except by certain grantees. Beginning in the 2020/21 school year, recipients of certain state grants, including those related to early literacy screening, were required to provide their students' early literacy universal screening assessment data to DESE. Over time, the number of grants and programs included has expanded.

For the 2023/24 school year, screening became required for all schools in the state. Currently, nine early literacy screening products (some of which include multiple assessment versions that target younger or older learners or are available in English and Spanish) are approved for use in elementary schools in Massachusetts.

Eleven English-language screening assessments are currently approved or partially approved for use in the state. Six Spanish-language assessments offered by publishers of the English-language assessments are also available for district or school use, although they are not formally approved.[2] Approved English-language products and related Spanish-language assessments include Acadience Reading, Amira and Amira Spanish, Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 8th Edition, EarlyBird, FastBridge's suite of assessments (aReading, earlyReading, earlyReading Spanish), i-Ready, MAP Reading Fluency and MAP Reading Fluency

[2] Note that data from Spanish-language assessments is included in the analysis if available.

Spanish, mCLASS and mCLASS Lectura, and Star's suite of assessments (Star Early Literacy, Star Early Literacy Spanish, Star Reading, and Star Reading Spanish). Note that mCLASS and DIBELS 8th Edition are the same assessment—DIBELS 8th Edition users assess students with paper tasks, and mCLASS users deliver the same tasks online. Assessments may continue to be added to the approved list on a rolling basis.

For this report, data from the following 13 specific screening assessments for grades K–3 were included in some or all analyses:

- Acadience Reading
- DIBELS 8th Edition
- EarlyBird
- FastBridge (aReading, earlyReading)[3]
- i-Ready (Diagnostic)
- MAP Reading Fluency
- mCLASS (mCLASS, mCLASS Lectura)
- Star Early Literacy (English and Spanish)
- Star Reading (English and Spanish)

DESE collected the K–3 screening assessment data from individual schools and districts and from assessment publishers (with whom districts signed data-sharing agreements to report data to DESE on their behalf) and provided it to the WestEd research team.

Comparability of Screening Assessments and Reporting

Approved early literacy screening assessments vary in many ways, including the content they assess, their technical characteristics, the mode of administration (e.g., paper, digital), the grade levels and languages targeted, costs, types of scores provided, and other aspects. Descriptions and details for each assessment can be found on DESE's early literacy screening assessment site and in Appendix B.

Perhaps most importantly, however, screening assessments differ in how they identify students who may be at risk for reading difficulty. Because the goal of screening is to identify these students so that additional support can be provided to them, all the screening assessments offer some indicator of risk. Indeed, most provide several performance benchmarks or risk levels (e.g., low risk, some risk, high risk or below benchmark, at benchmark, above benchmark). It is the lowest performance categories that are typically used to identify students who are "at risk" or "at significant risk."

[3] FastBridge earlyReading Spanish did not include sufficient data for reporting.

In its June 2023 *Early Literacy Screening Guidance*, DESE recommends that schools and districts use publisher-provided reporting categories for each approved screening assessment in order to determine whether or not students are performing "significantly below relevant benchmarks" as required by state regulation (Table 1). For example, for DIBELS 8th Edition, DESE recommends using the at risk or well below benchmark performance level to identify students significantly below benchmark. Using these benchmarks in analysis describes the numbers of students who would be identified as at risk locally according to DESE's guidance and for whom additional supports should be provided and information reported to parents.[4]

Table 1. Publisher-Provided Language for Benchmarks That Indicate Risk Based on DESE Guidance

Early literacy universal screening assessments	At/above benchmark	Below benchmark	Significantly below benchmark
DIBELS 8th Edition *(paper version)*	Minimal/negligible risk	Some risk	At risk
EarlyBird	Minimal/negligible risk	Some risk	Flag icon
mCLASS and DIBELS 8th Edition *(digital version)*	At/above benchmark	Below benchmark/ strategic support	Well below benchmark/intensive intervention
Acadience	At/above benchmark	Below benchmark	Well below benchmark
FastBridge	Low/no risk	Some risk	High risk
i-Ready	No observed risk	At some risk	At risk
MAP Reading Fluency	At grade level	Approaching grade level expectations	Flag and/or below grade level expectations
STAR	At/above benchmark	Below benchmark/ on watch	Intervention and urgent intervention

Source: DESE. (2023). *Early literacy screening guidance*. https://www.doe.mass.edu/instruction/screening-guide.pdf

[4] Note that we use "at significant risk" and scoring "significantly below benchmark more than once" and scoring "significantly below benchmark" interchangeably in the sections that follow.

The levels that identify students as significantly below benchmark differ in how they were determined by assessment developers and in what they represent. The most common approach across screening assessments is a predictive approach that establishes a risk cut score by using an external criterion—generally another standardized test of reading skills—that is thought to be a good indicator of success in reading. A cut score is then established by identifying, through data analysis, students who have a low likelihood of being successful on that measure. One screening assessment used a judgment-based standard setting process in which panels of educators reviewed the assessment and decided how much of the content and skills measured on the test students needed to demonstrate to be "on grade level." The risk cut score identifies students farthest away from being on grade level according to that test.

Finally, benchmarks can be set based on normative data—selecting a percentile, such as the 25th percentile, to identify the lowest-performing group of students. Assessment publishers may also use both types of information to set their cut scores. Benchmark levels are generally, although not always, based on composite scores derived from specific reading subtests administered at each grade level. These differences in approach to setting benchmarks mean there is no truly common definition of risk across screening assessments. Table 2 provides an overview of how the risk levels for each screening assessment included in this report were set. Appendix B provides additional details on the methods and samples used in this report to set cut scores.

Table 2. Approaches to Setting Risk-Level Benchmarks

Early literacy screening assessment	Performance level indicating significant risk in reporting	Method for setting risk cut scores	Additional details on risk cut score
Acadience Reading (grades K–3)	Well below benchmark	Predictive analysis	Low likelihood (10% to 20% chance) of performance above 40th percentile on GRADE assessment
DIBELS 8th Edition (grades K–3)	Well below benchmark	Predictive analysis	Low likelihood of performance above 20th percentile on DIBELS Next (grade K) or Iowa Assessments (grades 1–3)
MAP Reading Fluency (grades K–3)	Universal screener flag	Predictive analysis	Low likelihood of performing above 10th percentile on MAP Growth Reading or below 25th percentile on words correct per minute (WCPM) measure

Early literacy screening assessment	Performance level indicating significant risk in reporting	Method for setting risk cut scores	Additional details on risk cut score
mCLASS (grades K–3)	Well below benchmark	Predictive analysis	Low likelihood of performance above 20th percentile on DIBELS Next (grade K) or Iowa Assessments (grades 1–3)
mCLASS Lectura (grades K–3)	Well below benchmark	Predictive analysis	Low likelihood of performance above 20th percentile on Woodcock-Muñoz Análisis de Palabras (grade K) or Star Early Literacy Spanish (grades 1–3)
EarlyBird (grade K)	Dyslexia risk flag	Predictive analysis	Low likelihood of performing above the 16th percentile on the Kaufman Test of Educational Achievement, 3rd Edition (KTEA-3), Phonological Processing subtest
FastBridge aReading (grades 2 and 3)	High risk	Predictive analysis	Performance below 15th percentile (selected based on low likelihood of performing above 20th percentile on Gates-MacGinitie Reading Tests or MAP Growth Reading)
FastBridge earlyReading (grades K and 1)	High risk	Predictive analysis	Performance below 15th percentile (selected based on low likelihood of performing above 15th percentile on Group Reading Assessment & Diagnostic Evaluation [GRADE] assessment)
i-Ready Diagnostic (grades K–3)	At risk	Judgment-based standard setting	Performance below grade-level expectations (typically one or more grade level below current grade); grade-level expectations established based on college and career readiness standards
Star Early Literacy (English and Spanish; grades K and 1)	Intervention and urgent intervention	Normative	Performance below 25th national percentile

Early literacy screening assessment	Performance level indicating significant risk in reporting	Method for setting risk cut scores	Additional details on risk cut score
Star Reading (English and Spanish; grades 2 and 3)	Intervention and urgent intervention	Normative	Performance below 25th national percentile

Source: WestEd's compilation based on technical documentation and/or communication with publishers. (See the references for list of technical reports and related documentation.)

The lack of a common definition of "risk" can create challenges at the school, district, and state levels. First, students who move across schools or districts that use different screening assessments may not be consistently identified as in need of additional support or services. In the screening assessment data, agreement rates between results for students who were administered different screening assessments at the same time periods ranged between 62 percent and 85 percent (see Table B2 in Appendix B). Since schools are required to notify parents if students are identified as at risk, differences in how screening assessments identify students as at risk may also cause confusion for parents and families who could receive conflicting information about student performance and needs. It also complicates the ability of schools, districts, and the state to examine performance trends across schools and districts that use different screening assessments because patterns of performance may vary across assessments. A key goal of this analysis is to provide Massachusetts with a summary of how many students across the state are at risk of reading difficulties and how risk varies by student group, by school, and over time; therefore, selecting indicators of risk to use in reporting is an important analytic decision.

However, as noted, publisher-provided benchmarks differ in how they were established, and prior analysis[5] shows that screening assessments identify different numbers of students as at risk based on their benchmarks. Further, some screening assessment benchmarks are designed differently at different times of the year—for example, a beginning of year benchmark may be easier to achieve than an end-of-year benchmark on some screening assessments.

An alternative approach is to use publisher-provided national percentile scores. DESE's Dyslexia Guidelines recommend using the national 25th percentile or below as a metric to define the category of "significantly below relevant benchmarks." Thus, using these scores could provide a more consistent way to identify how many students in Massachusetts are among the lowest

[5] See earlier reports on Massachusetts data available at the DESE Early Reading Performance in Massachusetts site.

performers relative to the national population. However, two approved English-language screening assessments—EarlyBird and MAP Reading Fluency—do not report national percentile scores for reading risk overall, nor does mCLASS Lectura. Additionally, there may be some variability in the samples included and in the norming procedures used for other assessments, which may mean that national percentiles are not fully comparable either. Further, norms may change over time. Publishers often update their norms in response to broader educational trends. For example, several publishers recently updated their prepandemic norms for the 2024/25 school year, meaning that future analysis using normative metrics like the 25th percentile would need to acknowledge and/or take these changes into account.[6]

Table 2 also shows which grade levels for each assessment were included in this analysis. Note that for 2023/24 analysis, we excluded data from overall reporting for students who take assessments outside of the recommended grade levels (though we include additional analysis focused specifically on these students where possible). This change also applies to historical data (i.e., for analyses that compare data from prior years to 2023/24 data, we apply the same exclusions to prior year data).

As part of the analysis for this year's report, WestEd tested several different approaches to developing comparable metrics for reporting reading risk. This work demonstrated several potential methods DESE could use to create and apply consistent benchmarks indicating risk across assessments. However, given available data, no method works for all currently approved screening assessments. Each approach has other benefits and limitations as well. Details of this testing and its outcomes can be found in Appendix B.

For this reason, and for consistency with 2022/23 and earlier reporting, we report on 2023/24 data using both the publisher-provided benchmark categories and a 25th percentile or below metric, noting where differences in findings occur (though, as noted, national percentile data were not available for a few assessments). Using both types of reporting metrics provides complementary information that can be used to triangulate findings. Benchmark categories may be most commonly used in schools for identifying students and can therefore provide a picture of performance of students for whom schools are providing extra support, whereas national percentiles provide a more consistent measure across assessments and over time.

Student-Level State Education Data

In addition to the K–3 early literacy screening assessment data, other student-level data used in analysis included SIMS data, MCAS data, and ACCESS for ELLs data. These data provided additional information (demographic and assessment) about the K–3 students in the sample and were used to determine how representative the sample is of the state's K–3 student

[6] Star Early Literacy and Star Reading and i-Ready Diagnostic norms were updated. In both cases, results show that some students whose scale scores would have been identified as below the 25th percentile previously would not be so identified under updated norms.

population. DESE's guide to researchers using the statewide educational data in Massachusetts provides further details about which students are included/excluded in the SIMS, MCAS, and ACCESS data collections. A brief description of each data set follows.

Student Information Management System (SIMS)

SIMS collects data pertaining to various student characteristics (e.g., gender, race/ethnicity, English learner status, immigrant status, native language, attendance) for the majority of students across the state. In the SIMS data set, students can be identified using their unique state-assigned student identifier (SASID) or their locally assigned student identifier (LASID), which is unique at the district level. SIMS data are submitted three times per school year (October, March, and June) by districts across Massachusetts. For this analysis, data from the June collection were used to provide background characteristics for students in the screening assessment data unless it was missing. In those cases, October data were used. If a student was missing from both data sets, then they were included only in analyses not requiring demographic data. Only 4,057 observations (1% of cases and 1,668 students) were missing from both the October and the June SIMS data or do not have state student IDs available in the screening assessment data and cannot be matched to SIMS data.

Only the variables relevant to K–3 students were used when creating the student-level data set for analysis. These variables (and their associated codes) were defined using Version 20.3 of the SIMS Data Handbook for the 2020/21 school year, Version 20.7 for the 2021/22 school year, Version 20.9 for the 2022/23 school year, and Version 30.0 for the 2023/24 school year.

Massachusetts Comprehensive Assessment System (MCAS)

The MCAS data for the 2023/24 school year provide student assessment scores in mathematics and ELA for students in grade 3 and beyond who also have screening assessment scores in 2023/24 or earlier. MCAS data provide a statewide measure of ELA achievement for students, allowing for an analysis of how well screening assessment performance relates to later outcomes. These data include scaled scores and achievement levels describing performance in reading, writing, and language relative to grade-level expectations (Not Meeting Expectations, Partially Meeting Expectations, Meeting Expectations, Exceeding Expectations). Note that English learner students in their 1st year in the United States are exempt from taking the MCAS ELA assessment and are therefore excluded from MCAS-related analysis in those years.

Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs)

English learner students in grades K–12 in Massachusetts are tested annually using the ACCESS assessment to satisfy federal and state laws that require measuring the English proficiency of these students each year. ACCESS for ELLs is used to measure student proficiency in reading, writing, listening, and speaking, typically in January or February of each school year. ACCESS assessment data provide scale scores for each of the four language domains (i.e., listening,

speaking, reading, and writing) and an associated proficiency level. Proficiency levels describe a student's performance in terms of six English Language Proficiency Levels (i.e., Entering [Level 1], Emerging [Level 2], Developing [Level 3], Expanding [Level 4], Bridging [Level 5], and Reaching [Level 6]).

Students also receive four composite scores and proficiency levels that are derived from a weighted combination of domain scale scores:

- Overall: listening (15%), speaking (15%), reading (35%), and writing (35%)
- Oral Language: listening (50%) and speaking (50%)
- Comprehension: listening (30%) and reading (70%)
- Literacy: reading (50%) and writing (50%)

In addition to the scaled scores and proficiency levels, other assessment data are provided in the ACCESS files that can be used to evaluate the performance of English learner students, including progress toward proficiency and attainment of English proficiency resulting in students exiting English learner status.

Publicly Available School- and District-Level Data

Publicly available school- and district-level data for 2023/24 for Massachusetts were retrieved from DESE's school and district profiles website[7] to provide contextual data about the sample of students used in the analysis. Overall, the data pertain to the following four main categories of aggregate information: (a) educator characteristics (e.g., teacher race/ethnicity, teacher retention rate, percentage of experienced teachers), (b) student performance (e.g., MCAS scores, MCAS achievement levels), (c) student enrollment and demographic characteristics (e.g., grade-level enrollment, race/ethnicity, gender, percentage of students from low income backgrounds or who are economically disadvantaged, attrition rate, retention rate), and (d) financial (e.g., per-pupil expenditure amounts).

Additionally, for 2023/24 analysis, WestEd utilized publicly available data from the Segregation Tracking Project,[8] which provides several different metrics representing differences in the composition of schools and districts for students of different racial/ethnic and economic backgrounds. These data were used in analysis of the relationship between school segregation and early literacy performance.

[7] This website can be found at DESE's School and District Profiles website.

[8] This resource can be found at The Educational Opportunity Project's Segregation Tracking Project website.

Merging and Reporting of Student-, School-, and District-Level Data

All student-, school-, and district-level data for the 2020/21, 2021/22, 2022/23, and 2023/24 school years were combined into one primary longitudinal analytic file. Data are organized as a single longitudinal data set with one observation per student, per test period (i.e., beginning of year [BOY], middle of year [MOY], and end of year [EOY]), per screening assessment, per year. Some students had multiple screening assessment scores per test period because they participated in multiple early literacy screening assessments during the school year. Appendix A details the process of merging different data sources and describes the data cleaning process and associated business rules.

This report follows DESE's standard procedure for suppression of student demographic and assessment data: A minimum sample size of 6 students is required for reporting any student demographic information, and a minimum sample size of 10 students is required for reporting student assessment outcomes. "Sup data" represents suppressed data in this report.

Analysis and Findings

In this section, we describe how well the data represent students in grades K–3 in Massachusetts and provide updated information about student performance and progress.

Screening Assessment Data Overview

This subsection provides an overview of the literacy screening assessment data available for analysis. Specific questions to be addressed include the following:

- How many benchmark scores are available overall and for each assessment?
- How many students, schools, and districts are represented by the data in 2023/24?
- How many benchmark scores are available by grade and student group? To what extent does the sample of students with available benchmark scores represent the overall grade K–3 student population in the state?
- How many benchmark scores are available by time period in 2023/24?

How many benchmark scores are available overall and for each assessment?

Data Summary

- The data available in 2023/24 doubled from 2022/23.
- The data include about 388,000 scores and 137,000 students, about 53 percent of the state's K–3 student population.
- Scores from 612 schools and 144 districts, about half of schools serving K–3 students.
- The data are from 13 screening assessments.
- Six screening assessments account for 92 percent of the data.

Early literacy screening assessment data for the 2023/24 school year includes a total of 388,204 records with benchmark scores (i.e., levels that identify whether students are at risk of reading difficulty given their grade level and time of testing, as described in Table 2).[9] This number includes all screening assessments, grades K–3, and all time periods, because most assessments are administered several times per year.[10] The number of records available in 2023/24 doubled from 2022/23.

The most commonly used assessments were mCLASS (37%) and DIBELS 8th Edition (21%), which together account for about 58 percent of the sample. Star assessments (Star Early Literacy and Star Reading) represented 15 percent of the sample, i-Ready 12 percent, and MAP Reading Fluency 8 percent (see Table 3 for a full list).

Compared with 2022/23, the sample includes more mCLASS and MAP Reading Fluency records; fewer Star Early Literacy, Star Reading, i-Ready records; and somewhat fewer DIBELS 8th Edition records.

Analyses based on the 25th percentile include a smaller number of scores than those using publisher-provided benchmarks because not all screening assessments report national percentiles.

[9] About 65.7 thousand (65,657) records were removed from the data due to missing benchmark scores or other data issues. (Most of the observations removed came from DIBELS 8th Edition and Star CBM data that had only subtest scores.) Almost 14 thousand (13,949) observations from the FastBridge and Star suite of assessments were not used for the main analyses due to "off grade level" testing, but they were used for sub-analyses.

[10] Almost 6,000 (5,958) students had multiple scores from different screening assessments in the same time period in the 2023/24 school year (e.g., Star Early Literacy, Star Early Literacy Spanish).

Table 3. Number and Percentage of Screening Assessment Benchmark Scores by Year and Assessment for Grades K–3

Early literacy screening assessment	2021/22 (*N*)	2021/22 (%)	2022/23 (*N*)	2022/23 (%)	2023/24 (*N*)	2023/24 (%)	2023/24 (25th percentile metric) (*N*)	2023/24 (25th percentile metric) (%)
Acadience Reading	0	0%	4,615	3%	8,838	2%	8,838	3%
DIBELS 8th Edition	21,686	32%	46,938	26%	79,914	21%	79,914	23%
EarlyBird	1,360	2%	2,689	1%	4,214	1%	0	0%
FastBridge aReading	1,291	2%	628	<1%	4,053	1%	4,053	1%
FastBridge earlyReading	390	<1%	1,168	<1%	5,452	1%	5,452	2%
MAP Reading Fluency	314	<1%	0	0%	32,414	8%	0	0%
Star Early Literacy	12,530	19%	26,418	15%	29,337	8%	29,337	8%
Star Early Literacy Spanish	1,534	2%	0	0%	3,564	<1%	3,564	1%
Star Reading	2,799	4%	25,977	14%	28,680	7%	28,680	8%
Star Reading Spanish	0	0%	0	0%	1,728	<1%	1,728	<1%
i-Ready	16,426	24%	34,784	19%	45,568	12%	45,568	13%
mCLASS	9,383	14%	36,131	20%	143,949	37%	143,942	41%
mCLASS Lectura	0	0%	0	0%	493	<1%	0	0%

Early literacy screening assessment	2021/22 (*N*)	2021/22 (%)	2022/23 (*N*)	2022/23 (%)	2023/24 (*N*)	2023/24 (%)	2023/24 (25th percentile metric) (*N*)	2023/24 (25th percentile metric) (%)
Total	67,713	100%	179,348	100%	388,204	100%	351,076	100%

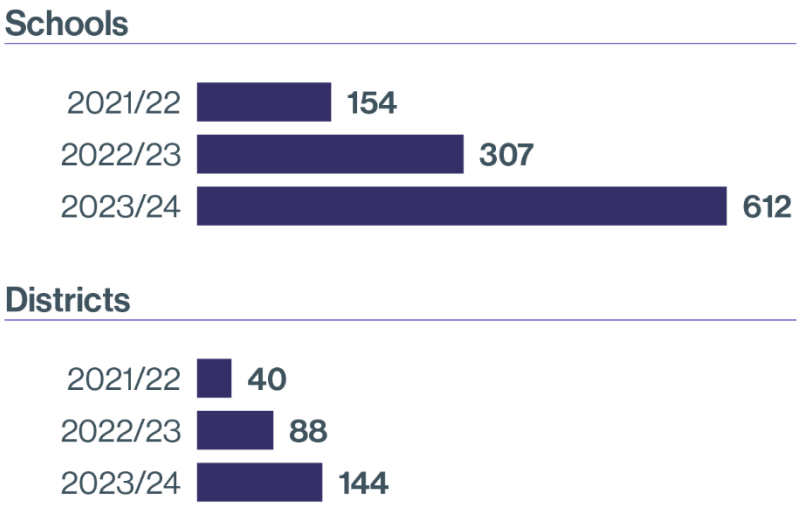
Note. Excludes screening assessments with no data or insufficient data for reporting in 2023/24. EarlyBird includes only data for grade K. FastBridge earlyReading and Star Early Literacy (English and Spanish) include only grades K and grade 1. FastBridge aReading and Star Reading (English and Spanish) include only grades 2 and 3. *Source*: Early literacy screening assessment data.

How many students, schools, and districts are represented by the data in 2023/24?

Most screening assessments are administered at least three times per school year (i.e., BOY, MOY, and EOY), although the start and end dates for these time periods vary across schools and districts. As a result, most students have multiple scores within a school year, and some students have more than one score per time period if they took multiple different screening assessments during the school year.

The 388,204 scores shown in Table 3 represent 137,159 unique students across 144 districts and 612 schools in 2023/24, about double the numbers of students, schools, and districts included in 2021/22 (Figure 1). About half of elementary schools and one third of school districts in the state are represented in the data.

Figure 1. Between School Years 2022/23 and 2023/24, Numbers of Schools and Districts Included in Analysis Approximately Doubled



Note. Five hundred and six schools across 136 districts are represented in data using the 25th percentile or below metric *Source*: Early literacy screening assessment data.

The 137,159 students represent about 53 percent of the total grade K–3 student population in the state in 2023/24, up from 26 percent in 2022/23.[11] About 30 percent of students (40,473 students) across the sample have at least one benchmark score in 2022/23 and 2023/24, and about 5 percent of students have scores for 2023/24 and 2 prior years.

[11] Enrollment data were retrieved from DESE's School and District Profiles page.

Table C1 in Appendix C provides the number of students with data from each literacy screening assessment. Some students took multiple screening assessments within the school year, resulting in larger totals than the actual number of unique students.

How many benchmark scores are available by grade and student group? To what extent does the sample of students with available benchmark scores represent the overall grade K–3 student population in the state?

Key Findings

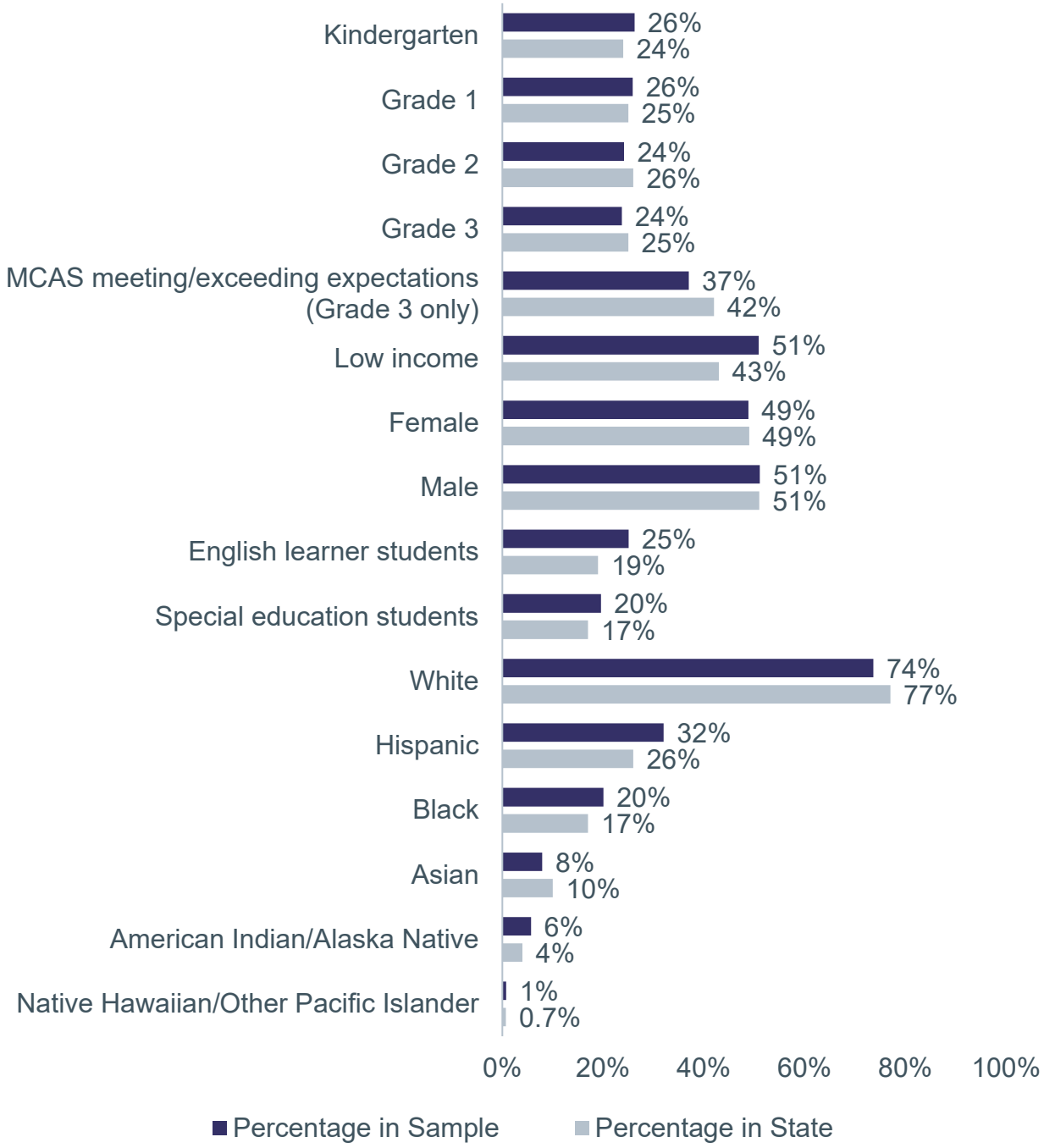
- There are more Hispanic students, students from low income backgrounds, and English learner students in the screening data sample than there are in the state.
- Student background characteristics differ across assessments.
- The screening assessment sample may be somewhat lower performing than that of the state overall.

The results in this report are based on a sample of students. Because the sample slightly overrepresents students who may be at higher risk of being identified as significantly below benchmark, the findings may slightly underestimate overall student performance across the state.

Compared with the state, the background characteristics of students in the screening assessment sample include between 6 and 8 percent more students who are Hispanic, from low income backgrounds, and are English learners (Figure 2). About 51 percent of the sample was classified as from low income backgrounds compared with 43 percent in the state, about one quarter were identified as English learner students compared with 19 percent in the state, and 32 percent were Hispanic compared with 26 percent in the state.

The early literacy screening assessment data have 2 to 3 percent more Black, American Indian or Alaska Native, and kindergarten students and students receiving special education services than the state and about 2 to 3 percent fewer Asian, White, and grade 2 students. Thirty-seven percent of 3rd grade students met expectations on MCAS in the literacy screening assessment sample compared with 42 percent in the state overall.

Figure 2. Screening Assessment Data Includes More Hispanic, English learner, and Students From Low Income Backgrounds Than the State Population At Grades K–3



Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). For figure data, see Figure 2 in Appendix F. *Source*: Early literacy screening assessment data, October and June SIMS collection data, and state-provided MCAS data.

The students in the screening data sample and the state are clustered in schools, which may affect the resources and support available to them. Compared with schools in the state, the schools included in early literacy screening data had higher concentrations of Black, Hispanic, and English learner students and lower concentrations of White students than most schools in the state. They also had higher mobility and chronic absenteeism rates and lower MCAS proficiency rates than most schools in the state. A later section of this report enumerates differences in student performance by school characteristics.

Student data are also clustered by screening assessment, and student background varies by assessment, showing differences in which assessments schools with different student populations choose. As noted in the previous section, six assessments account for 92 percent of the sample. Among these assessments, MAP Reading Fluency had higher proportions of students from low income backgrounds and English learner, Hispanic, and Black students than other assessments (Table 4). Fewer DIBELS 8th Edition and i-Ready test takers were from low income backgrounds and more were White compared with other assessments.

Table 4. Screening Assessment Data by Student Group

Demographic	DIBELS 8th Edition	MAP Reading Fluency	Star Early Literacy	Star Reading	i-Ready	mCLASS	State
Low income	39%	69%	53%	49%	46%	55%	43%
English learner	17%	41%	30%	25%	20%	26%	19%
Students receiving special education services	20%	21%	16%	20%	21%	20%	17%
White	86%	53%	71%	77%	86%	68%	77%
Hispanic/Latino	26%	44%	36%	31%	30%	32%	26%
Black	10%	42%	17%	15%	13%	26%	17%
Asian	8%	9%	6%	5%	5%	10%	10%
American Indian or Alaska Native	5%	3%	14%	10%	3%	3%	4%
Native Hawaiian or Other Pacific Islander	0%	1%	1%	1%	0%	1%	1%

Demographic	DIBELS 8th Edition	MAP Reading Fluency	Star Early Literacy	Star Reading	i-Ready	mCLASS	State
Not MCAS meeting/exceeding	58%	68%	N/A	65%	64%	63%	58%
MCAS meeting/exceeding	42%	32%	N/A	35%	36%	37%	42%

Note. No MCAS performance is listed for Star Early Literacy because only K–1 students were analyzed for Star Early Literacy. Some students may appear multiple times per row if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories).
Source: Early literacy screening assessment data, October and June SIMS collection data, and state-provided MCAS data.

To the extent that school- and student-level characteristics related to early literacy performance and progress differ between the early literacy screening assessment sample and the state as a whole, the description of results based on the sample may not be fully generalizable to the state's K–3 population. In this case, screening assessment data were obtained from schools and districts participating in certain grants for which they had to compete, which could suggest a higher level of capacity or willingness to take on improvement initiatives than other districts in the state. Further, prior year analysis shows that within the sample of screening assessment data, students from low income backgrounds; students receiving special education services; and Hispanic, Black, and English learner students are more likely to be identified as at risk than their peers not in those groups (as are students who attend schools with more students in those groups). Because these groups are somewhat overrepresented in the screening sample, the sample may be somewhat lower performing than the state overall.

An additional challenge is that, in this case, student performance measures themselves can vary (and are also related to student and school background characteristics). As described in the introduction, this year's analysis included additional work to examine approaches to improve comparability of performance measures; future work could also incorporate other approaches to adjust for sample differences and improve generalizability.

How many benchmark scores were available by time period in 2023/24?

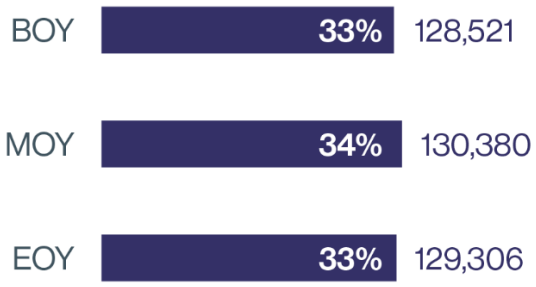
Key Findings

- Scores were evenly distributed by time period.
- Seventy-six percent of students have scores in all three time periods.
- More White students and fewer English learners, Black and Hispanic students, students from low income backgrounds, and students receiving special education services had three scores compared with students with only one or two scores.

Of the 388,204 records with benchmark scores in 2023/24, about 33 percent were from the BOY time period, 34 percent were from the MOY time period, and 33 percent were from the EOY time period (Figure 3).[12]

Most screening assessments had about the same number of scores in each time period, but about 40 percent of EarlyBird scores were from BOY compared with only 23 percent at EOY. mCLASS Lectura also had more scores at BOY (37%) compared with EOY (27%), while Star Early Literacy Spanish and Star Reading Spanish had more scores at the end of the year. Table C5 in Appendix C provides details. Students with different background characteristics also had about the same number of scores in each time period (see Table C6 in Appendix C for data by time period and student group).

Figure 3. Data Included About the Same Number of Scores in Each Time Period

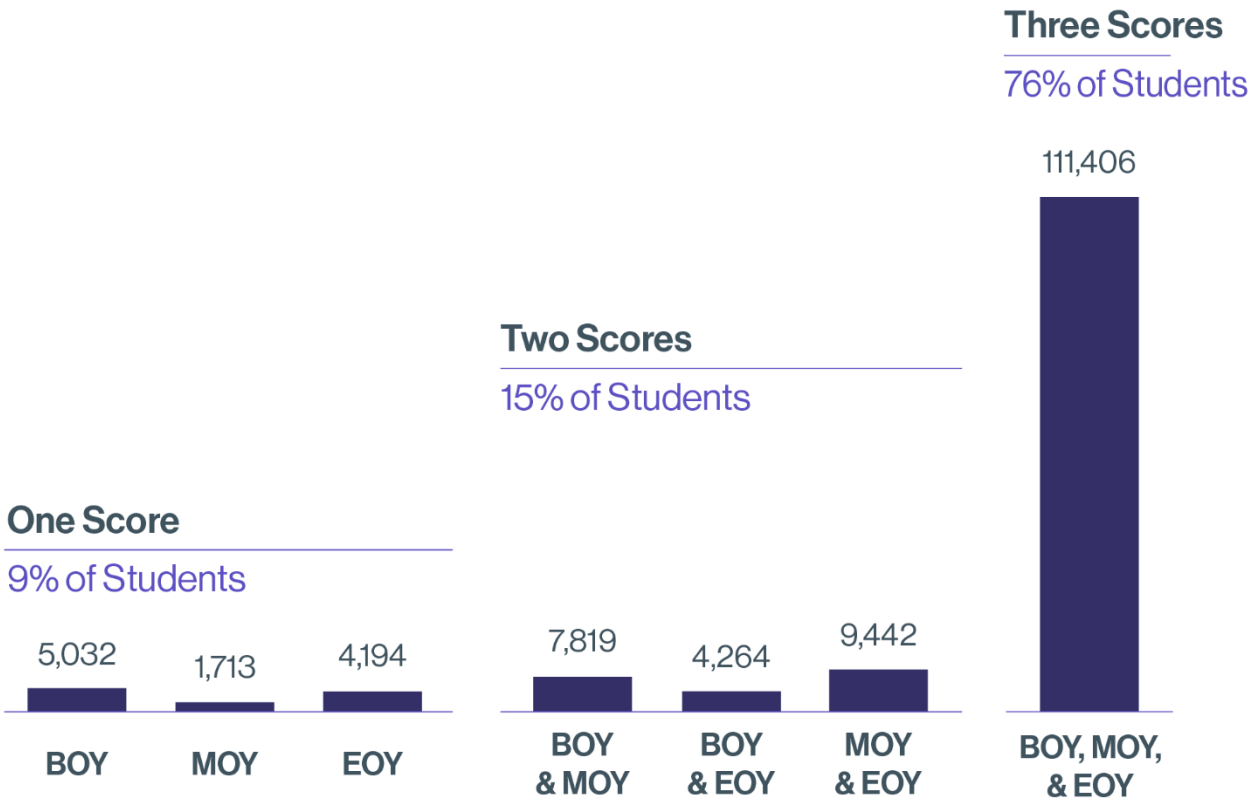


Note. Some students are included more than once in a time period if they took multiple screening assessments in that time period. *Source*: Early literacy screening assessment data.

[12] Under the 25th percentile metric, 35 percent of scores are from EOY, 33 percent from MOY, and 32 percent from BOY.

Just over three quarters of the K–3 students in the screening assessment data (76%) had benchmark scores from the same screening assessment in all three time periods (Figure 4), about the same as in 2022/23. About 15 percent had benchmark scores only from two time periods, and 9 percent had benchmarks only from a single time period. Of the students with two benchmark scores, most had MOY and EOY scores (Figure 4). Most students with only one benchmark had that score at BOY.

Figure 4. Seventy-Six Percent of Students Had Benchmark Scores in All Three Time Periods



Note. Students with multiple screening assessments per time period are counted for each screening assessment. Using the 25th percentile or below metric, about 6 percent of students had one score, 13 percent had two scores, and 81 percent had scores in all three time periods. (Although the total number of scores was smaller than the publisher-provided benchmark scores because several screening assessments do not report national percentiles.) *Source*: Early literacy screening assessment data.

Students with all three scores included a higher percentage of White students and lower percentages of students from low income backgrounds, English learner, Black, and Hispanic students and students receiving special education services compared with students with only one or two scores (see Table C7 in Appendix C). For example, 22 percent of students with three scores were ELs compared with 34 percent of students with two scores and 43 percent of

students with one score. Kindergarten students were about twice as likely as students in grades 1 to 3 to have only two scores available than three scores. Most had MOY and EOY scores, which may reflect a preference to begin screening kindergarteners later in the year after they first enter school.

Student Performance

There are multiple approaches to describing how many students may be at risk of reading difficulty based on the literacy screening assessment data. In addition to differences between screener benchmarks, different interest holders may be interested in answering different questions for different purposes—for example, at what time of year are most students at risk? How many students are at risk? How many students are consistently at risk? In the following sections, we provide information that addresses several questions of student performance:

- How many students were identified as significantly below benchmark by time period, grade, and student group?
- How many students were identified as significantly below benchmark two or three times overall and by grade and student group?
- How do student and school background factors interact and relate to the likelihood of students being identified as significantly below benchmark more than once?
- Which students are tested "off grade level"? How does their performance compare with that of other students?

This section first describes students "significantly below benchmark" following DESE guidance[13] (refer to Table 3 for details on reporting of benchmark categories for each assessment) and then students at or below the 25th percentile.

In the analysis that follows, we provide information for each time period during which students typically take tests (BOY, MOY, and EOY) separately and summarize information across time periods, describing how often students are classified as significantly below or below benchmark multiple times during the year.

[13] DESE's early literacy screening guidance can be found in the Early Literacy Screening Guidance document.

How many students were identified as significantly below benchmark by time period, grade, and student group?

Key Findings

- About one third of students were identified as significantly below benchmark at each time point according to publisher-provided benchmarks, and about one quarter were identified with the 25th percentile as the risk indicator.
- Risk rates were higher for students from low income backgrounds; students receiving special education services; and English learner, Black, Hispanic, AIAN, and NHPI students.
- Risk rates varied from BOY to EOY by screening assessment and benchmark type.
- Fewer former English learner students were at significant risk than current or never English learner students.

As in prior reporting, using publisher-provided benchmarks shows that about one third of students were classified as significantly below benchmark at BOY (33%), MOY (33%) and EOY (30%) (Figure 5). Using the 25th percentile as a risk metric (which excludes MAP Reading Fluency, EarlyBird, and mCLASS Lectura) identifies about 26 percent of students as significantly below benchmark in any time period.[14]

Figure 5. About One Third of Students Were Identified as At Significant Risk At Each Time Period

Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Includes students with any number of scores. Percentages identified as significantly below benchmark for students with scores in all three time periods

[14] Removing Spanish-language assessments reduces the percentages of students identified as at risk by less than 1 percent.

are 33 percent at BOY, 32 percent at MOY, and 27 percent at EOY. *Source*: Early literacy screening assessment data.

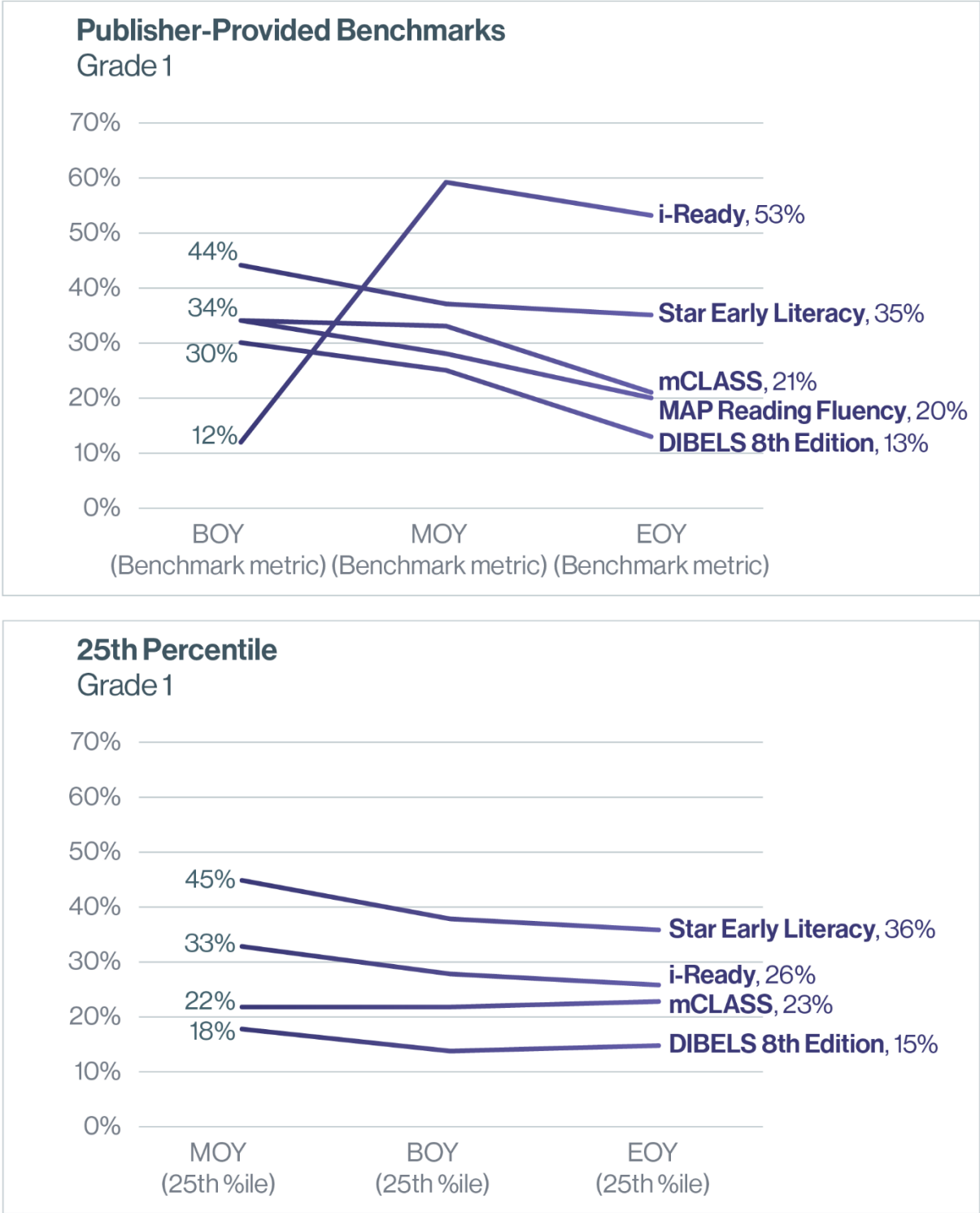
Figure 5 represents students with scores in any number of time periods. Among students who had scores in all three time periods, the percentages of students identified as significantly below benchmark varied by grade level and test over time. Figure 6 shows how percentages changed for the most commonly used assessments (DIBELS 8th Edition, MAP Reading Fluency, Star Early Literacy, i-Ready, mCLASS) for 3 as an example. Table C8 in Appendix C provides data for all grade levels and assessments. **Note that these figures include different students at each grade level.** For information on how individual students progress across years, see the section on student progress.

The percentage of students identified as at significant risk using publisher-provided benchmarks generally decreased between BOY and MOY for DIBELS 8th Edition, Star Early Literacy, and mCLASS and generally increased for i-Ready. That is, fewer students are identified as at significant risk over time, except for i-Ready, for which those percentages increase likely due to the manner in which i-Ready sets its EOY risk cut scores to move closer to grade-level performance.[15] Percentages decrease for MAP Reading Fluency at grades K and 1 and increase at grades 2 and 3. Prior reporting showed greater decreases in the percentages of students identified as at risk at earlier grade levels, which is also true using publisher-provided benchmarks in 2023/24 (again, except for i-Ready).

The percentage of students identified as at risk—based on scoring below the 25th percentile—either decreases or remains the same across grade levels and assessments except for mCLASS at grades K and 1. DIBELS 8th Edition and mCLASS show greater decreases with the publisher-provided benchmarks than with the 25th percentile. Unlike with the publisher-provided benchmarks, the percentages of students identified as at risk on i-Ready decrease using the 25th percentile. As noted, these differences between results using publisher-provided benchmarks and the 25th percentile are likely due to the manner in which different publishers set their risk benchmarks. (See Comparability of Screening Assessments and Reporting in the data and methods section for more information.) Because Star's publisher-provided benchmark is based on the 25th percentile, percentages of students identified as at risk do not change using that metric.

[15] i-Ready documentation recommends analyzing student performance over time based on changes in scaled scores compared with typical changes. See, for example, description of i-Ready student growth metrics.

Figure 6. Percentages of Students Identified as At Significant Risk Decrease Between BOY and EOY for Most Assessments At Grade 1



Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

As in prior reporting, in each time period, students from low income backgrounds; English learner students; students receiving special education services; and Black, Hispanic, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander students were more likely to be identified as at risk than their peers not in those groups (Table 5). For example, 47 percent of students from low income backgrounds were significantly below benchmark at BOY compared with 20 percent of students not from low income backgrounds, meaning students from low income backgrounds were more than twice as likely to be identified as significantly below benchmark as students not from low income backgrounds. Conversely, Asian students were about half as likely as non-Asian students to be classified as significantly below benchmark.

Table 5. Percentage of Students Identified as Significantly Below Benchmark and At Relative Risk of Being Significantly Below Benchmark At BOY, MOY, and EOY by Student Group

Demographic	Significantly below benchmark at BOY	Significantly below benchmark at MOY	Significantly below benchmark at EOY	Change from BOY to EOY	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Low income	47%	46%	39%	–7%	2.3	2.4	2.4
Non–low income	20%	19%	16%	–4%	N/A	N/A	N/A
Female	32%	31%	26%	–6%	1.0	1.0	0.9
Male	33%	33%	28%	–5%	N/A	N/A	N/A
Nonbinary	Sup data	Sup data	Sup data	Sup data	NC	NC	NC
English learner	60%	58%	49%	–11%	2.4	2.3	2.3
Non-English learner	25%	25%	21%	–4%	N/A	N/A	N/A
Students receiving special education services	56%	58%	53%	–4%	2.1	2.2	2.5
Students not receiving special education services	27%	26%	21%	–6%	N/A	N/A	N/A
White	31%	30%	25%	–5%	0.8	0.8	0.8
Non-White	39%	39%	33%	–6%	N/A	N/A	N/A

Demographic	Significantly below benchmark at BOY	Significantly below benchmark at MOY	Significantly below benchmark at EOY	Change from BOY to EOY	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Hispanic/Latino	50%	49%	43%	–7%	1.9	2.0	2.0
Non-Hispanic/Latino	26%	25%	21%	–5%	N/A	N/A	N/A
Black	41%	41%	34%	–7%	1.3	1.4	1.3
Non-Black	31%	30%	26%	–5%	N/A	N/A	N/A
Asian	18%	18%	14%	–4%	0.5	0.5	0.5
Non-Asian	34%	33%	28%	–6%	N/A	N/A	N/A
American Indian or Alaska Native	59%	54%	50%	–9%	1.9	1.8	1.9
Non-American Indian or Alaska Native	31%	31%	26%	–5%	N/A	N/A	N/A
Native Hawaiian or Other Pacific Islander	45%	44%	37%	–8%	NC	NC	NC
Non-Native Hawaiian or Other Pacific Islander	33%	32%	27%	–5%	N/A	N/A	N/A
Total	**33%**	**32%**	**27%**	**–6%**	**N/A**	**N/A**	**N/A**

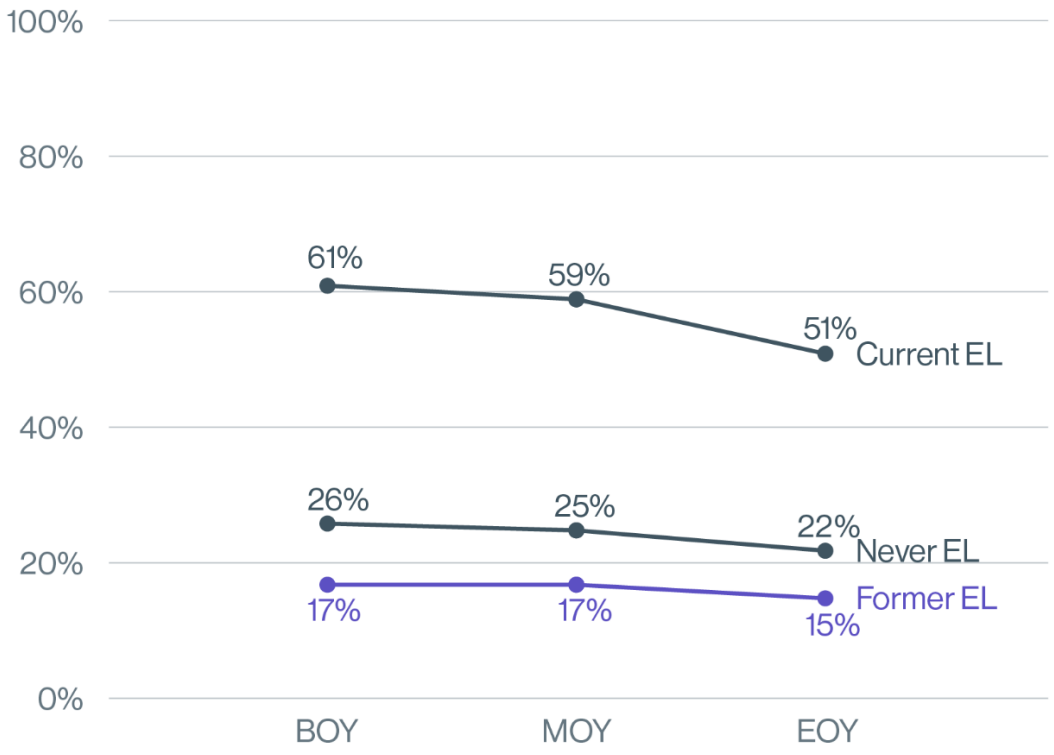
Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (i.e., Hispanic/not Hispanic and multiple race/ethnicity

categories). The relative risk ratio indicates the likelihood of a student group being classified as significantly below benchmark compared with students not in that group. For example, students from low income backgrounds were 2.3 times more likely than students who are not from low income backgrounds to be classified as significantly below benchmark at BOY. Includes only students with three scores. "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. NC means that the percentage was not computed because the group made up 5 percent or less of the sample.
Source: Early literacy screening assessment data and October and June SIMS collection data.

Analyzing data with the 25th percentile as the indicator of significant risk shows the same patterns by student group (see Table C13 in Appendix C).

In 2023/24, we also explored in more depth the performance of English learner students—specifically, achievement differences between students who exited from English learner status (former English learners) compared with those still receiving service or who never received services. In general, research shows that students who exit from English learner status in early elementary grades tend to be high achievers overall and often are students who entered school with relatively higher levels of English language proficiency compared with their peers (see, e.g., Greenberg Motamedi, Singh & Thompson, 2016; Kieffer & Parker, 2016; Slama et al., 2017; Thompson, 2017). As Figure 7 shows, this trend held true among screener data as well: the percentage of former English learners who were deemed at risk (by any screener in any language) was lower at all time points than for current or never English learners, by comparison.

Figure 7. The Percentage of Students Significantly Below Benchmark Is Lower Among Former English Learners Than Current or Never English Learners At All Time Points



Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

How many students were identified as significantly below benchmark two or three times overall and by grade and student group?

Key Findings

- About 25 to 30 percent of students were significantly below benchmark more than once during the school year.
- Between 16 and 19 percent of students were significantly below benchmark three times.
- Fewer kindergarten students were significantly below benchmark multiple times than were students in other grades.
- Repeated risk identification rates were higher for students from low income backgrounds; students receiving special education services; and English learner, Black, Hispanic, AIAN, and NHPI students.
- Likelihood of being identified more than once varies by screening assessment.

As described in earlier sections, because students are typically assessed more than once during the school year, there are multiple approaches to summarizing performance. One relevant metric is the number of students who are repeatedly identified as significantly below benchmark, because these students may be particularly in need of additional support.

In the previous section, we summarized performance by grade and student groups for each time period. In the following sections, we focus on the performance of students with scores in at least two time periods who were identified as significantly below benchmark in more than one time period. Such analysis necessarily excludes students with only one score. About 76 percent of students had scores in all three time periods, and about 91 percent had scores in at least two time periods. Focusing on students identified as significantly below benchmark multiple times therefore excludes about 9 percent of students because those students only had one score (Figure 4).

In 2023/24, among students with at least two scores, 30 percent of students were identified as significantly below benchmark two or more times (Table 6). That percentage was 25 percent using the 25th percentile to identify students as significantly below benchmark. Despite the fact that this year's screening assessment sample differs from prior year samples, these percentages are nearly identical to those from 2022/23.

Table 6. Percentage of Students Identified as Significantly Below Benchmark Multiple Times by Number of Available Scores

Benchmark scores available	Significantly below benchmark multiple times	25th percentile or below multiple times
Two benchmarks available	30% (*n* = 6,481), 2 times	30% (*n* = 4,822), 2 times
Three benchmarks available	30% (*n* = 33,639), 2 or 3 times 12% (*n* = 13,011), 2 times 19% (*n* = 20,628), 3 times	24% (*n* = 24,621), 2 or 3 times 8% (*n* = 8,000), 2 times 16% (*n* = 16,621), 3 times
Total	**30% (*n* = 40,120)**	**25% (*n* = 29,443)**

Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. The total for each column includes the number of students significantly below benchmark two or more times among students with two or three benchmarks available. *Source*: Early literacy screening assessment data.

Similarly, among students with scores in all three time periods, fewer students overall were identified as at significant risk more than once during the school year using the 25th percentile as an indicator of risk compared with publisher-provided benchmarks (24% compared with 30%). Similar to prior reporting, between 16 and 19 percent of students with all three scores were identified as at significant risk in all three time periods.

Of the students who had at least two scores, fewer kindergarten students were identified as significantly below benchmark multiple times than were students in other grades. As was the case in prior reporting, students from low income backgrounds; students receiving special education services; English Learner students; and Black, Hispanic, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander students were more likely than their peers not in those groups to score significantly below benchmark multiple times during the school year (Table 7).

Table 7. Relative Risk of Being Identified as At Significant Risk Multiple Times by Student Group

Demographic	Significantly below benchmark two or more times	Relative risk of significantly below benchmark two or more times
Low income	43%	2.5
Non–low income	17%	N/A
Female	29%	0.9
Male	31%	N/A
English learner	56%	2.5
Non-English learner	22%	N/A
Students receiving special education services	56%	2.3
Students not receiving special education services	24%	N/A
White	28%	0.8
Non-White	37%	N/A
Hispanic/Latino	47%	2.0
Non-Hispanic/Latino	23%	N/A
Black	38%	1.4
Non-Black	28%	N/A
Asian	16%	0.5
Non-Asian	32%	N/A
American Indian or Alaska Native	54%	1.9
Non–American Indian or Alaska Native	29%	N/A
Native Hawaiian or Other Pacific Islander	42%	NC
Non–Native Hawaiian or Other Pacific Islander	30%	N/A
Total	**30%**	**N/A**

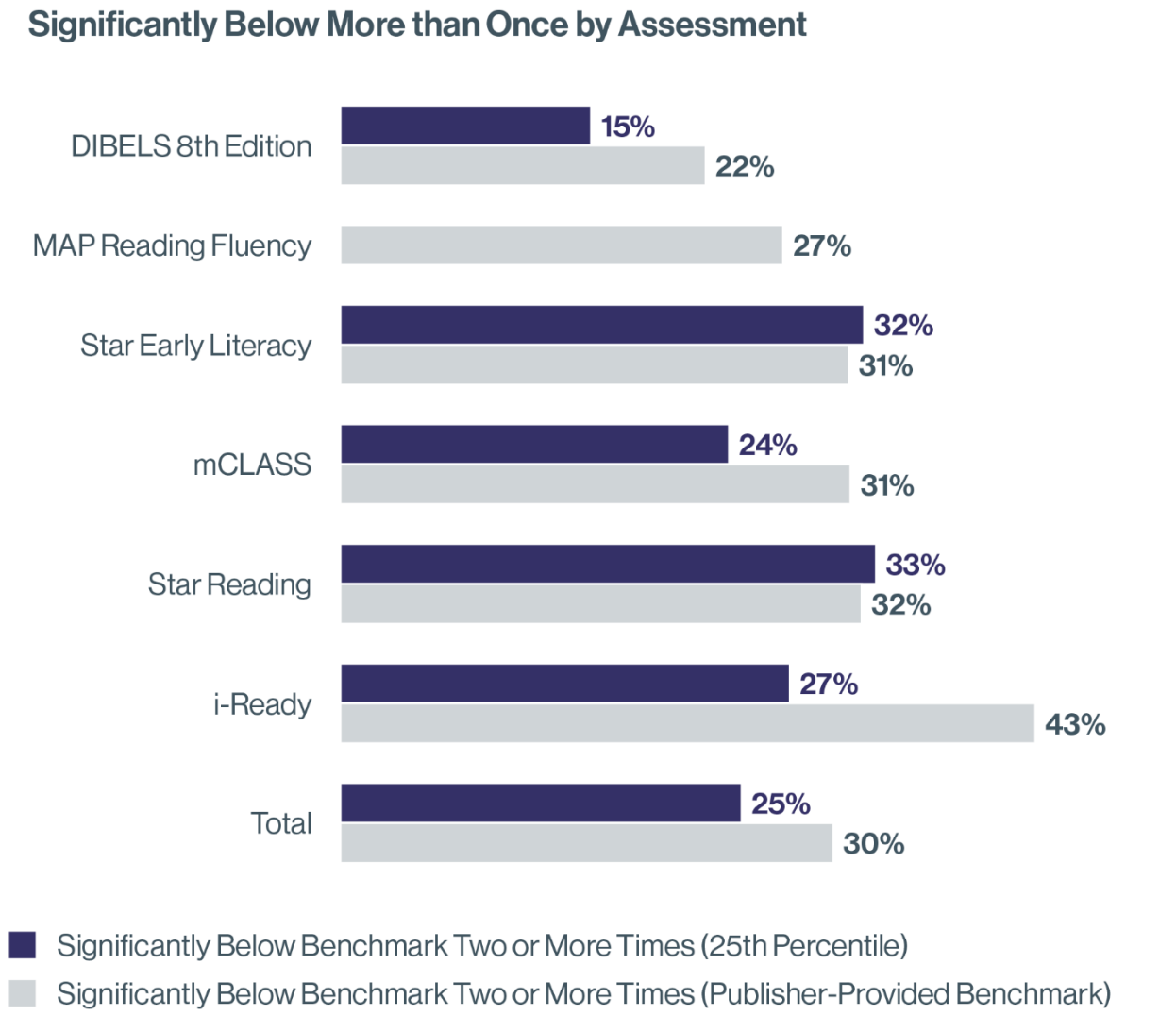
Note. Some students may appear multiple times per row if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (i.e., Hispanic/not Hispanic and multiple race/ethnicity categories). The relative risk ratio indicates the likelihood of a student group being classified as significantly below benchmark compared with students not in that group. For example, students from low income backgrounds were 2.5 times more likely than students who are not from low income backgrounds to be classified as significantly below benchmark multiple times. NC means that the percentage was not computed because the group made up 5 percent or less of the sample. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

For example, 43 percent of students from low income backgrounds were significantly below benchmark at least twice during the school year compared with 17 percent of students not from low income backgrounds. The relative risk of students from low income backgrounds being classified as significantly below benchmark multiple times during the school year was 2.5 times that of students not from low income backgrounds. Asian students and White students were less likely than their peers to be significantly below benchmark more than once during the school year; about the same proportions of female and male students were classified as significantly below benchmark. Table C14 in Appendix C provides the percentages of students identified as significantly below benchmark using the 25th percentile and below metric. These values follow the same patterns as shown for the publisher-provided benchmarks.

Across the most commonly used screening assessments (DIBELS 8th Edition, i-Ready, mCLASS, Star Early Literacy, MAP Reading Fluency, and Star Reading), the percentages of students classified as significantly below benchmark multiple times during the school year vary from 22 percent to 43 percent using the publisher-provided benchmarks (Figure 8 and Table C15 in Appendix C).[16] Using the 25th percentile as the indicator of significant risk results in smaller overall percentages of students identified as at risk, and smaller differences between the percentages identified by each assessment. These disparities are likely due to differences in the ways each screening assessment's benchmark scores are calculated and to differences in the background characteristics of students participating in each assessment.

[16] Results of a multilevel model that examines student and school-level factors associated with being identified as at risk multiple times also showed differences in the likelihood of identification for some screening assessments. See Appendix E for details.

Figure 8. Percentage of Students Identified as At Risk More Than Once Varies by Screening Assessment



Note. Some students may appear multiple times if they were administered multiple screening assessments. MAP Reading Fluency does not report national percentiles. For figure data, see Figure 8 in Appendix F. *Source*: Early literacy screening assessment data.

How do student and school background factors interact and relate to the likelihood of students being identified as significantly below benchmark more than once?

Key Findings

- Students who experienced intersecting educational inequities related to factors like race/ethnicity or socioeconomic status were more likely to be repeatedly identified as at risk.
- Patterns of repeated risk differed by gender and race/ethnicity.

Early literacy screening assessments aim to identify students who are not on track to become successful readers and who require additional support to be so. Analysis in the prior section and prior reporting show disparities in performance between students with different background characteristics, as are often found in analysis of educational and other data, suggesting that current systems are not providing adequate support for all students. For example, screening assessment data show that students from low income backgrounds, English learner students, and students receiving special education services were more likely than their peers not in those groups to score significantly below benchmark multiple times during the school year, which means these students are in need of additional support.

However, taking an intersectional approach to analysis offers an opportunity to more closely examine outcomes. Such an approach may inform development of policies and supports that better address the complexity of students' experiences and recognize that membership in multiple historically marginalized groups is likely associated with disparities in achievement. McCall (2005) identified several different approaches to intersectional analysis. In prior reporting, WestEd took an "intercategorical" approach, analyzing data across student background categories such as low income status and race/ethnicity rather than only analyzing categories separately.[17] This analysis used a multilevel regression modeling approach with interactions to examine the relationships between multiple characteristics and risk of reading difficulty. Results showed variation in the relationship between performance and factors like low income status, race/ethnicity, and gender. For example, results showed that all students from low income backgrounds were more likely to need additional support than students not from low income backgrounds, but the likelihood increased by 5 percentage points for Asian students and 10 percentage points for Hispanic students.

[17] See the brief Lemke, M., Murphy, D., Soo Ping Chow, A., & Acuña, A. (2024). *Intersections: Student Background and Early Literacy Performance*, WestEd. https://www.doe.mass.edu/instruction/ela/research/intersections.pdf

Using data from 2023/24, we repeated this analysis using a similar statistical model to validate previous findings and found very similar results. A detailed description of the statistical model and full model results can be found in Appendix E.

As in prior reporting, although a pattern of increasing likelihood of being identified as at significant risk as background characteristics intersect is the same across student groups, the increases vary by gender, race/ethnicity, and other background characteristics. For example, female students receiving special education services were again shown to be more likely to be identified as at significant risk more than once during the school year across all ethnoracial groups, while there were no differences between males and females among non-special education participants (Figure 9). Other findings from prior reporting, such as the relevance of early childhood participation in reducing risk differences across English learner students by gender and race/ethnicity were also observed again. See the issue brief *Intersections: Student Background and Early Literacy Performance* (Lemke et al., 2024) for details.

Figure 9. Likelihood of Repeated Risk Identification Was Higher for Female Students Receiving Special Education Services Than It Was for Male Students Across Ethnoracial Groups

Note. Model probabilities indicate the likelihood of being identified as significantly below benchmark multiple times during the school year. For example, White female students receiving special education services had about a 50 percent probability of being identified as significantly below benchmark multiple times. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

In last year's and current analyses, ethnoracial categories were defined as mutually exclusive. It was not possible to estimate a model with non-exclusive ethnoracial categories given the small numbers of students in each group when combined with other background data. However, with a larger sample in 2023/24, it is possible to descriptively analyze data with overlapping racial and ethnic categories, though some caution is needed in interpreting results.

Student racial and ethnic background is typically collected by asking parents of students or students themselves to identify first if they are Hispanic/Latino and then to select from other racial categories (e.g., White, Black, Asian). In Massachusetts, common intersections of student race and ethnicity are Hispanic/White, Hispanic/Black, Not Hispanic/White, Not Hispanic/Black, Not Hispanic/Asian, and Not Hispanic/Multiple. These intersecting categories also intersect with other background characteristics in variable ways, including coming from a low income background, being an English learner, and receiving special education services (Table 8). For example, 83 percent of Hispanic/Black students are classified as being from a low income background compared with 76 percent of Hispanic/White students and 73 percent of Not Hispanic/Black students. About 40 percent of Hispanic/Black and Not Hispanic/Asian students were classified as English learners, but Hispanic/Black English learner students were about twice as likely to also be classified as being from a low income background and about four times as likely to receive special education services as Not Hispanic/Asian students.

Table 8. Percentage of Students From Low Income Backgrounds, English Learner Students, and Students Receiving Special Education Services by Race/Ethnicity

Race/ethnicity	Number	% of total	LI	EL student	Receiving special education services	LI and EL student	LI student and receiving special education services
Hispanic/White	26,783	19%	76%	47%	21%	40%	17%
Hispanic/Black	5,164	4%	83%	40%	21%	36%	18%
Hispanic/Asian	280	<1%	55%	28%	12%	19%	9%
Hispanic/Multiple	3,657	3%	77%	43%	21%	37%	18%
Hispanic/AIAN	4,524	3%	85%	83%	12%	71%	10%
Hispanic/NHPI	338	<1%	85%	51%	19%	44%	17%
Not Hispanic/White	62,123	44%	28%	5%	20%	4%	8%

Race/ethnicity	Number	% of total	LI	EL student	Receiving special education services	LI and EL student	LI student and receiving special education services
Not Hispanic/Black	14,562	10%	73%	28%	22%	23%	16%
Not Hispanic/Asian	7,652	5%	34%	38%	12%	19%	5%
Not Hispanic/Multiple	6,075	4%	45%	6%	20%	4%	12%
Not Hispanic/AIAN	326	<1%	68%	44%	17%	37%	13%
Not Hispanic/NHPI	130	<1%	62%	38%	25%	30%	20%

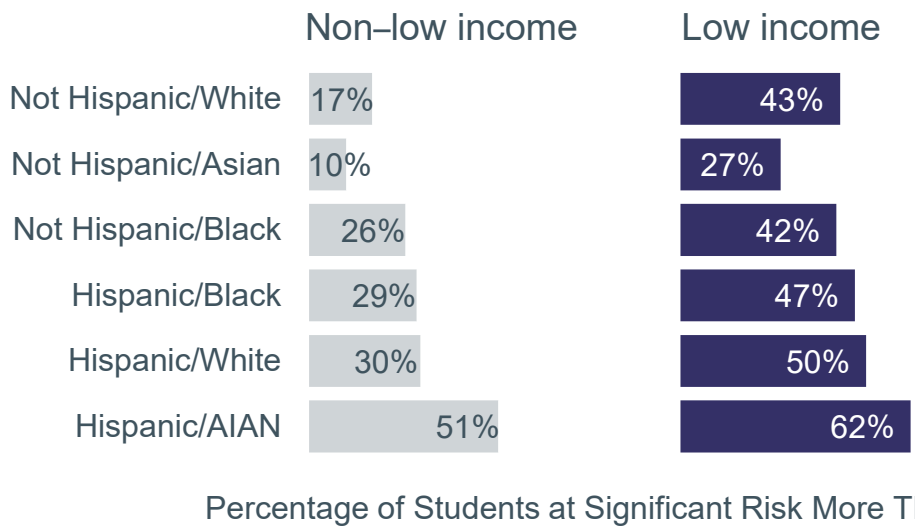
Note. Some students may appear multiple times per row if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Following up on previous analyses, we again explore how identifying or being identified as a member of multiple student groups relates to consistent risk of reading difficulties—that is, being identified as significantly below benchmark more than once within a school year. For 2023/24, we focus on common overlapping background characteristics, including Hispanic/White, Hispanic/Black, Hispanic/AIAN, non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian students who are also classified as coming from low income backgrounds, being English learners, and/or receiving special education services.

As seen in this year's and previous analyses, students who receive special education services, are English learners, or come from a low income background are more often identified as significantly below benchmark more than once than are students without those backgrounds, and the chance of being identified as in need of support increases as these background characteristics intersect.

For example, across ethnoracial groups, more students from low income backgrounds were identified as at significant risk than were students not from low income backgrounds (Figure 10). However, within students not from low income backgrounds, Hispanic White students were about 1.8 times as likely to be at significant risk than were non-Hispanic White students and about 2.9 times as likely as non-Hispanic Asian students.

Figure 10. Across Ethnoracial Student Groups, More Students From Low Income Backgrounds Were Identified as At Significant Risk Than Were Students Not From Low Income Backgrounds

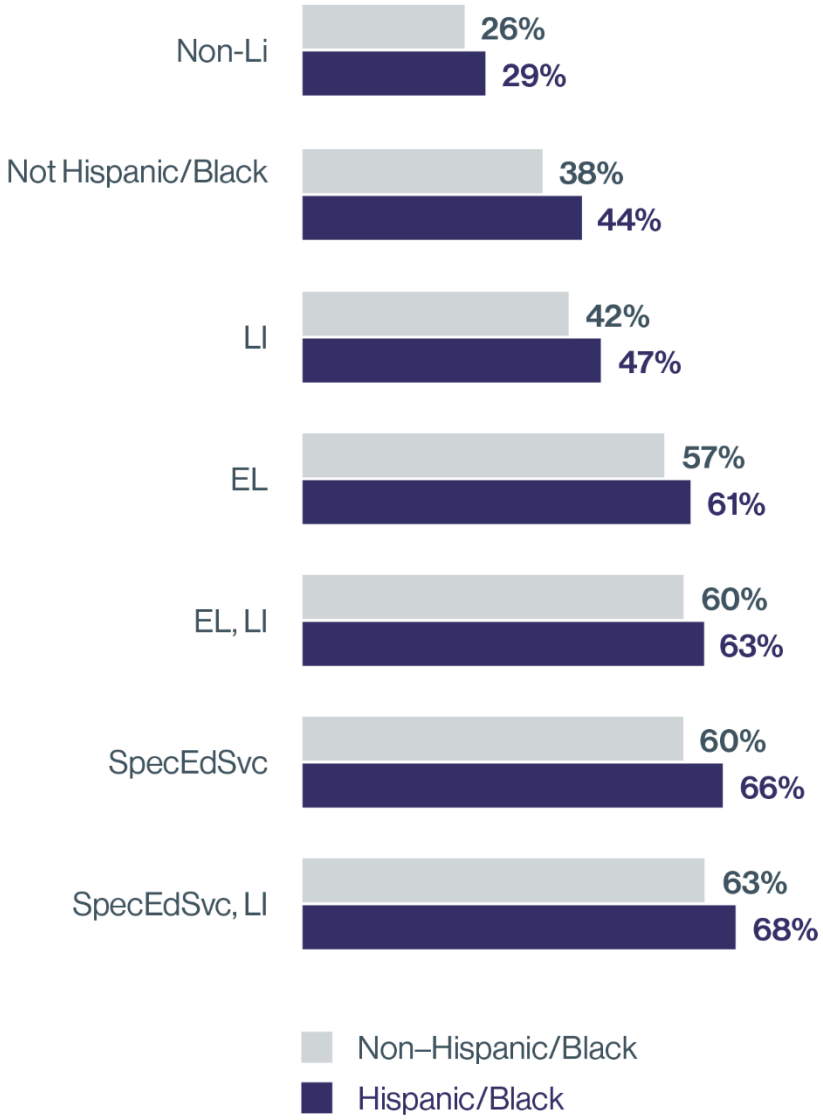


Note. Some students may appear multiple times per group if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Additionally, across ethnoracial groups, students who receive special education services and are also classified as from low income backgrounds are most often identified as significantly below benchmark, followed by students receiving special education services (particularly female students receiving special education services), students classified as English learners and from low income backgrounds, and English learner students not from low income backgrounds (see Appendix C for details).

However, the pattern is the same, there are differences between students based on race and ethnicity. For example, as shown in Figure 11, 44 percent of Hispanic Black students were identified as at significant risk compared with 38 percent of non-Hispanic Black students. Data also show other differences within ethnoracial groups (e.g., between non-Hispanic White and Hispanic White students). The percentages of students identified as at risk was always smallest for non-Hispanic Asian students and highest for Hispanic AIAN students.

Figure 11. More Hispanic Black Students Were At Significant Risk More Than Once Than Were Non-Hispanic Black Students



Note. Some students may appear multiple times per group if they were administered multiple screening assessments. LI means low income, "SpecEdSvc" means students receiving special education services. For figure data, see Figure 11 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Which students are tested "off grade level"? How does their performance compare with that of other students?

Key Findings

- Most students taking "off grade" assessments were English learners or students who received special education services.
- Students tested "off grade level" in later grades were more likely to be repeatedly identified as at risk.
- Most students taking both grade-level and "off grade level" assessments were consistently identified as at risk.

Several of the approved early literacy screening tools include multiple separate assessments designed to be used on their own or together. For example, the Star assessment suite includes Star Early Literacy and Star Reading, separate assignments that target different skills and, as a result, different grade levels. Star also offers assessments in Spanish and a curriculum-based measure (CBM).[18] Other publishers also offer multiple assessments for different purposes or students. As noted in the introduction, to maximize comparability in reporting, we restricted Star and FastBridge reporting to include students from the grade levels for which the publisher or DESE recommends use of each assessment. Recommended grade levels for Star Early Literacy and FastBridge earlyReading are grades K–1 and grades 2 and 3 for Star Reading and FastBridge aReading.[19]

However, publisher and DESE guidance also recommend that schools consider which assessments are most appropriate for students, and in 2023/24 and prior years, some data were reported for students using assessments outside of the recommended grade levels (e.g., Star Early Literacy for a grade 3 student). In this section, we examine the numbers, background characteristics, and performance of students taking Star Early Literacy and Star Reading in English by grade. A later section more closely examines the numbers, background, and performance of students taking assessments in Spanish and English. FastBridge "off grade level" test data include fewer than 10 students and are excluded from this analysis.

Most students taking Star Early Literacy were in grades K and 1 and most students taking Star Reading were in grades 2 and 3 (Table 9). Students taking Star Early Literacy in later grades were more often students receiving special education services, English learners, or both than students taking Star Early Literacy in grades K and 1. For example, 41 percent of students taking

[18] Star CBM is not currently approved for use on its own.

[19] See Renaissance guidance for Star at the K–3 Guidance for Star Early Literacy and Star Reading site and Illuminate Education's FastBridge FAQs site.

Star Early Literacy in grade 3 were receiving special education services compared with 16 percent of Star Early Literacy test takers in grades K and 1.

Table 9. Students Taking "Off Grade Level" Star Assessments by Student Background Characteristics

Grade	Assessment	Number	Receiving special education services	EL student	EL student and receiving special education services	Not EL or receiving special education services
K	Star Early Literacy	5,343	15%	29%	4%	60%
1	Star Early Literacy	5,621	18%	31%	5%	56%
2	Star Early Literacy	3,132	25%	45%	8%	38%
3	Star Early Literacy	1,137	41%	52%	16%	23%
K	Star Reading	59	17%	25%	2%	59%
1	Star Reading	1,435	14%	11%	2%	76%
2	Star Reading	4,697	18%	23%	3%	63%
3	Star Reading	5,575	22%	27%	5%	56%

Source: Early literacy screening assessment data and October and June SIMS collection data.

However, between 23 and 38 percent of students taking Star Early Literacy assessments in grades 2 and 3 were not receiving special education services or were English learners, suggesting that schools may also be administering these assessments to students for reasons unrelated to disability or language proficiency. Among English-proficient students not receiving special education services, students taking Star Early Literacy in grades 2 and 3 were about 1.5 to 3 times as likely to be repeatedly identified as at significant risk than were students in grades K and 1 taking the same assessment (Table 10). Students taking Star Early Literacy in grade 3 were also more likely to have been at significant risk more than once the year before (while in grade 2).

Table 10. Percentage of Students Taking "Off Grade Level" Star Assessments Identified as At Significant Risk Multiple Times

Grade	Assessment	All	Receiving special education services	EL student	EL student and receiving special education services	Not an EL student, not receiving special education services
K	Star Early Literacy	26%	42%	55%	62%	11%
1	Star Early Literacy	37%	60%	66%	79%	16%
2	Star Early Literacy	64%	88%	79%	94%	33%
3	Star Early Literacy	75%	92%	90%	95%	25%
K	Star Reading	Sup data	Sup data	Sup data	Sup data	Sup data
1	Star Reading	6%	21%	13%	27%	4%
2	Star Reading	29%	60%	51%	70%	16%
3	Star Reading	34%	64%	69%	85%	12%

Note. "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Most students taking Star Reading were in grades 2 and 3, but students taking it at earlier grades were somewhat less likely to be receiving special education services than were students in grades 2 and 3, and at grade 1, they were less likely to be English learners. They were also less likely to be identified as at significant risk on Star Reading than were students in higher grade levels, suggesting that these students may be advanced learners.

Many students taking "off grade" assessments also took typical grade-level assessments in the same time period—88 percent of grade 3 students took Star Early Literacy and 69 percent of grade 2 students also took Star Reading. In these cases, about 84 percent of students were identified as significantly below benchmark on both assessments (Table 11). This finding would seem to suggest that in general, no new information about identification is being gained by

giving students both assessments, though in about 15 percent of cases, students would be identified as at significant risk on one assessment and not the other.

Table 11. Percentage of Students Identified as Significantly Below Benchmark on Star Early Literacy and Star Reading

Assessments	BOY (N, % agree)	MOY (N, % agree)	EOY (N, % agree)
Star Early Literacy (English) and Reading (English)	1,870 83%	1,998 84%	2,046 83%

Note. Includes students with any number of scores (i.e., agreement is within time period). Including only students with all three scores decreases sample size to 628 and increases agreement rate to 87 percent at BOY, 90 percent at MOY, 89 percent at EOY. *Source*: Early literacy screening assessment data.

Student Progress

In the previous sections, we reported on student performance at BOY, MOY, and EOY and across time periods and on students who performed significantly below benchmark at multiple time periods. In this section, we examine how individual student performance changes over the course of the year and across years. Specifically, we address the following questions:

- How does student performance change as the school year progresses? Do students identified as significantly below benchmark remain significantly below benchmark?
- How does student progress vary by grade and student background characteristics?
- How does student performance change across grade levels? Do students at risk remain at risk across years?
- How does student performance change over the summer?
- What is the relationship between screening assessment scores and MCAS performance in grade 3?
- How do students identified as at risk in grades 1, 2, and 3 perform on MCAS in later grades (e.g., 3, 4, and 5)? Among students identified as at risk early (e.g., beginning of grade 1), how does MCAS performance in grade 3 differ between those who remain at risk over time and those who reach benchmark?

How does student performance change as the school year progresses? Do students identified as significantly below benchmark remain significantly below benchmark?

Key Findings

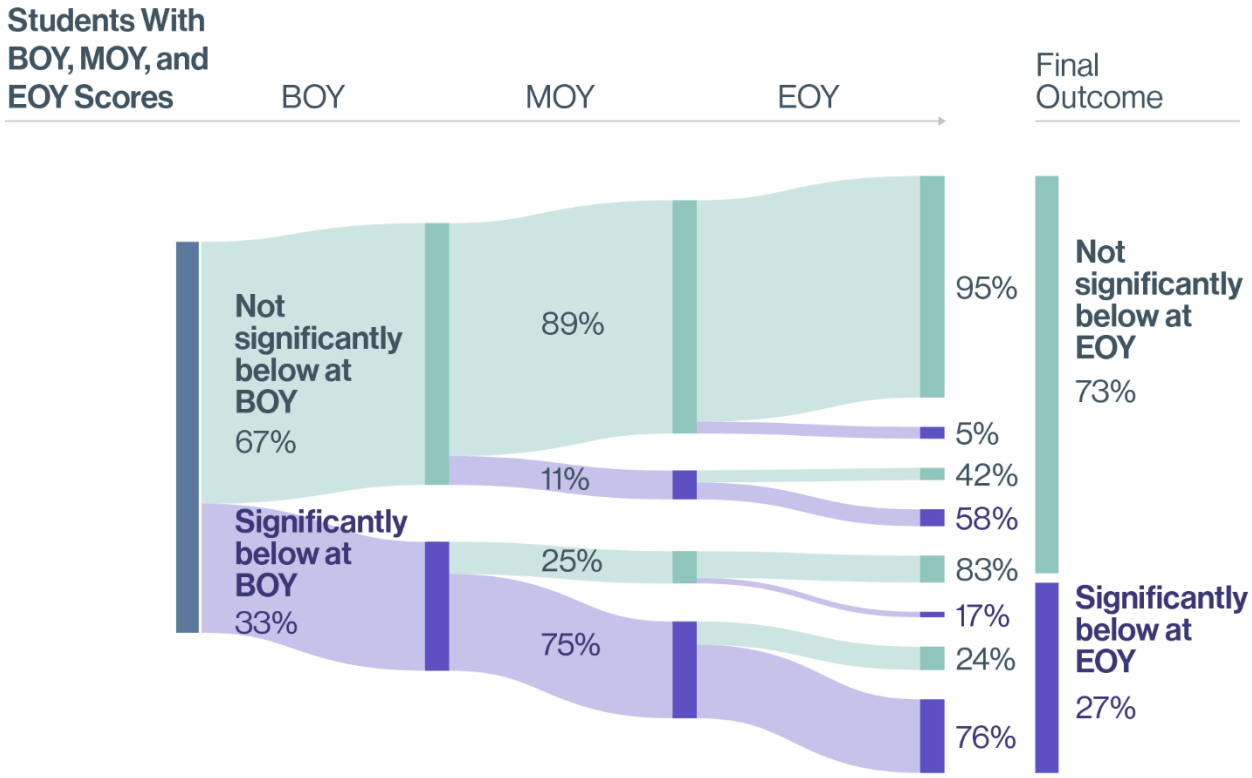
- Students mostly remain either significantly below benchmark or at benchmark across the school year.
- Most students who get on track and meet benchmark by MOY stay on track, whereas most who fall off track stay off track through EOY.

As in prior reporting, most students who were significantly below benchmark in one time period were also significantly below benchmark in later time periods. Among students who had a score in all three time periods of the school, 75 percent of students who were identified as significantly below benchmark at BOY were also significantly below benchmark at MOY, and 76 percent of the students who performed significantly below benchmark at BOY and MOY were also at that level at EOY (Figure 12).

Conversely, the vast majority of students who were *not* identified as significantly below benchmark at the beginning of the year were also not identified as such at the end of the year. Also, most students who improved their performance between BOY and MOY maintained that performance at EOY. About 25 percent of students who were classified as significantly below benchmark at BOY improved their performance at MOY, and of these students, the vast majority (83%) were not significantly below benchmark at EOY.

However, most students who were identified as at significant risk for the first time at MOY remained at significant risk at EOY. Of the 11 percent of students who were **not** significantly below benchmark at BOY but were at MOY, 58 percent continued to be significantly below benchmark at EOY.

Figure 12. Most Students Who Started Significantly Below Benchmark Were Still Significantly Below Benchmark in Later Time Periods



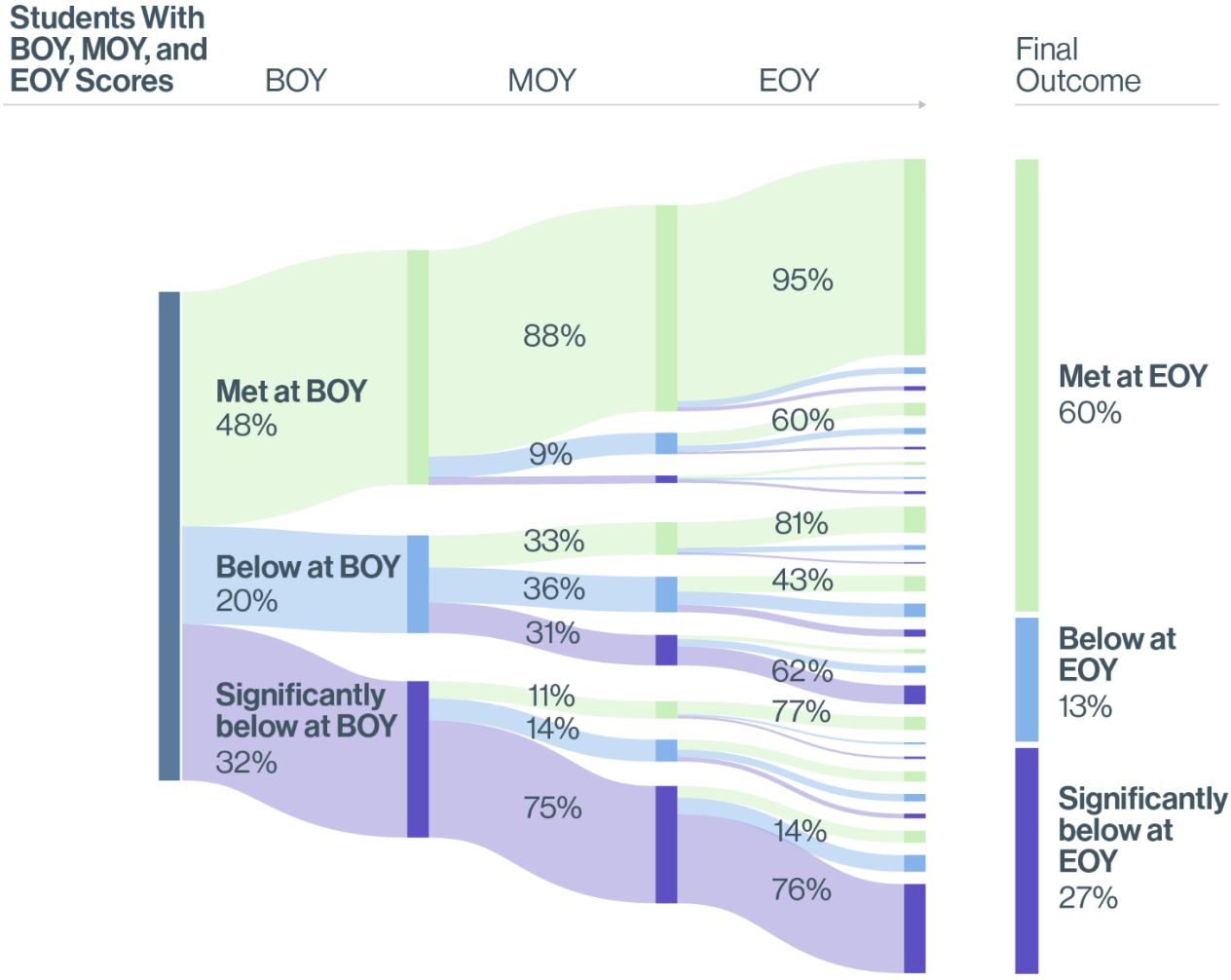
Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (76% of students) were included in the figure. See Figure 12 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

Using the 25th national percentile to identify students as at significant risk shows the same patterns of performance, though fewer students were identified as at significant risk at BOY (Figure D1 in Appendix D).

Figure 12 divides student performance into two categories: significantly below benchmark and not significantly below benchmark. However, because most assessments include multiple risk levels, some students identified as **not** significantly below benchmark might still be at some risk of reading difficulty.

Figure 13 displays progress using these more detailed categories of risk. As in the earlier analysis, students who met benchmark or who were significantly below benchmark at BOY typically remained at those levels at MOY and EOY, but students at a lower level of risk ("below benchmark" but not "significantly below benchmark") have more mixed results.

Figure 13. Most Students Who Improved Performance Between BOY and MOY Maintained That Improvement



Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (76%) were included in the figure. See Figure 13 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

Student performance followed several patterns. Students who were below benchmark at the beginning of the year (BOY)—represented by the yellow segment in Figure 13—showed mixed trajectories. By the middle of the year (MOY), their outcomes were fairly evenly split: About one third had improved to meet benchmark, one third remained below benchmark, and one third declined further, falling into the significantly below benchmark category. By EOY, approximately half of students reached benchmark, while the remainder were evenly divided between those still below benchmark and those significantly below.

Students who fell off track at MOY were unlikely to recover by EOY. For example, among the 12 percent of students who were on track at BOY but dropped below benchmark at MOY, about half continued to perform below or significantly below benchmark by the end of the year.

Conversely, students who improved between BOY and MOY were more likely to succeed by EOY.

- Students who were below benchmark at BOY but reached benchmark at MOY were nearly twice as likely to meet benchmark at EOY compared with those who remained below benchmark at both BOY and MOY.
- Students who began the year significantly below benchmark but met benchmark by MOY were
  - **1.6 times more likely** to meet benchmark at EOY than students who only improved to just below benchmark at MOY and
  - **7.7 times more likely** to meet benchmark than students who remained significantly below benchmark throughout the year.

How does student progress vary by grade and student background characteristics?

Key Findings

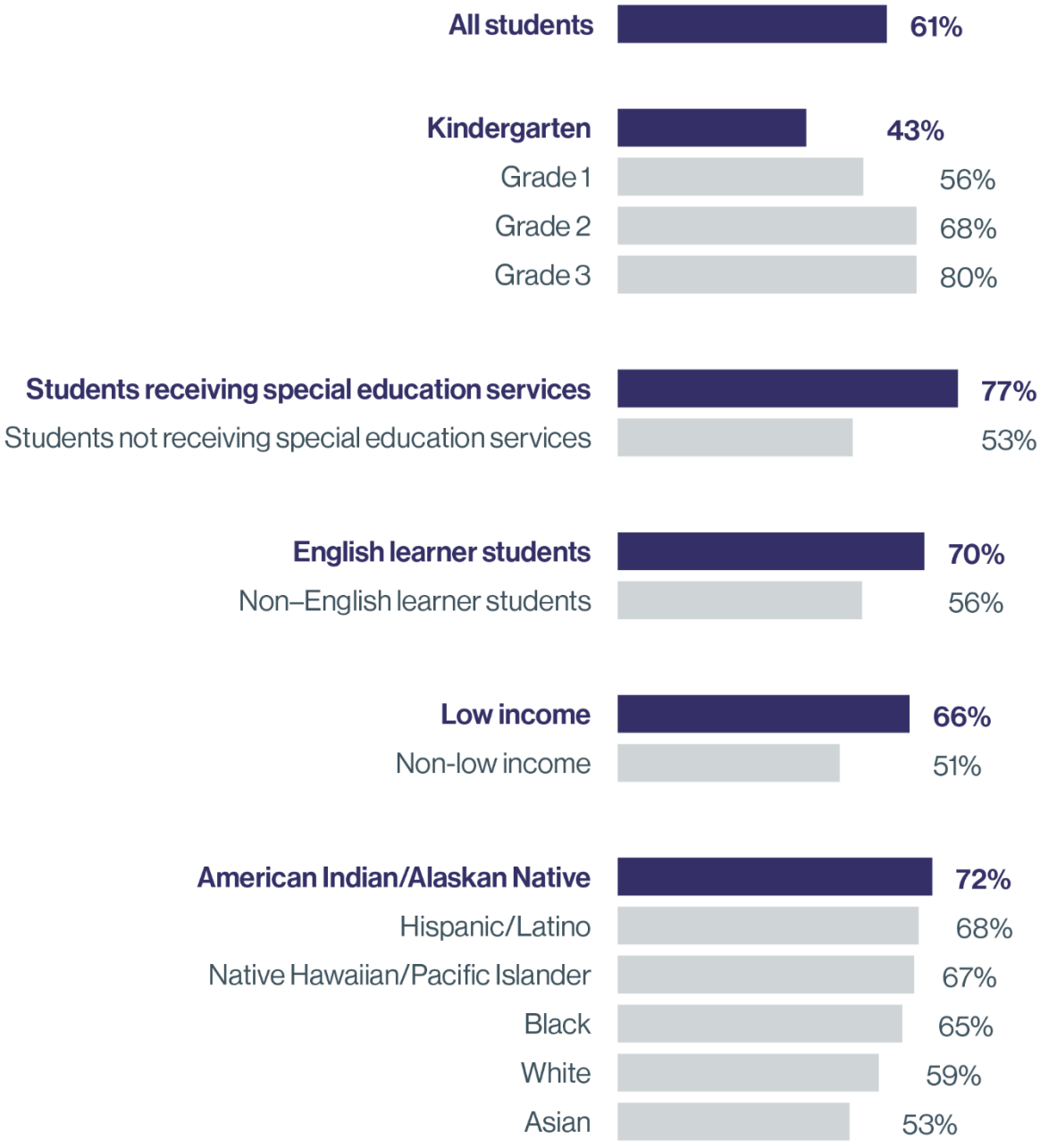
- Fewer students at later grade levels get on track after being identified as significantly below benchmark than do students in earlier grade levels.
- More English learner students, students from low income backgrounds, students receiving special education services, and non-White and non-Asian students were significantly below benchmark at EOY after being identified at BOY or MOY than were their peers not in those groups.

Among students with scores at all three time points, the percentage who remained significantly below benchmark after being identified as such at BOY increased across grade levels, consistent with prior findings. For instance, 43 percent of kindergarten students identified as significantly below benchmark at BOY remained in that category at EOY compared with 80 percent of grade 3 students (Figure 14). Grade 3 students were approximately 1.9 times more likely than kindergarten students to remain significantly below benchmark at EOY if they began the year in that category; grade 2 students were 1.6 times more likely, and grade 1 students were 1.3 times more likely. Notably, students in the earlier grades were more likely to meet benchmark by EOY, even if they started the year significantly below benchmark.

Additionally, as shown in prior analyses, more English learner students, students from low income backgrounds, students receiving special education services, and non-White and non-

Asian students were significantly below benchmark at EOY after being identified as such at BOY than were their peers not in those groups (Figure 14). For example, 66 percent of students from low income backgrounds who were significantly below benchmark at BOY were still significantly below benchmark at EOY compared with 51 percent of students not from low income backgrounds. Similarly, among students who started the year below benchmark (but not significantly below), fewer students from low income backgrounds, students receiving special education services, and English learner students improved and met benchmark at EOY than did their peers not in those groups.

Figure 14. English Learner Students, Students From Low Income Backgrounds, and Students Receiving Special Education Services Were More Likely to Remain Significantly Below Benchmark From BOY to EOY Compared With Peers Not in Those Groups



Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). *Source*: Early literacy screening assessment data and October and June SIMS collection data.

How does student performance change across grade levels? Do students at risk remain at risk across years?

Key Findings

- Most students who were significantly below benchmark at EOY in grades K and 1 were still significantly below benchmark 2 years later.
- More students in grades 2 and 3 remained significantly below benchmark across years than did students in grades K and 1.
- Students who were English learners, from low income backgrounds, Hispanic, and receiving special education services were more likely to remain significantly below benchmark across years.

In this section, we examine changes in performance for students with scores from multiple years. Prior analysis focused on changes in performance over a single year—between EOY 2021/22 and EOY 2022/23. In 2023/24, we can track student performance for some students over 2 years. Additionally, in 2023/24 we examine performance from EOY of 1 year to BOY of the next to explore how student performance may change over the summer.

In total, 29,826 (about 23% of the scores in 2023/24) had EOY scores in 2023/24 and 1 year earlier (Tables 12 and 13). There were 4,272 students who had scores at EOY in 2021/22, 2022/23, and 2023/24 on the same assessment.

Table 12. Number of Students With EOY Scores Over 1 Year

Students with scores from EOY 2022/23 and EOY 2023/24 (EOY 1 year later)	Number
Grade K to grade 1	11,305
Grade 1 to grade 2	7,776
Grade 2 to grade 3	10,745
Total	**29,826**

Note. Screening assessment data are available only for students in grades K–3, so we cannot track scores 1 year later for students in grade 3. *Source*: District-provided early literacy screening assessment data.

Table 13. Number of Students With EOY Scores Over 2 Years

Students with scores from EOY 2021/22, 2022/23 and 2023/24 (EOY 2 years later)	Number
Grade K to grade 2	2,263
Grade 1 to grade 3	2,009
Total	**4,272**

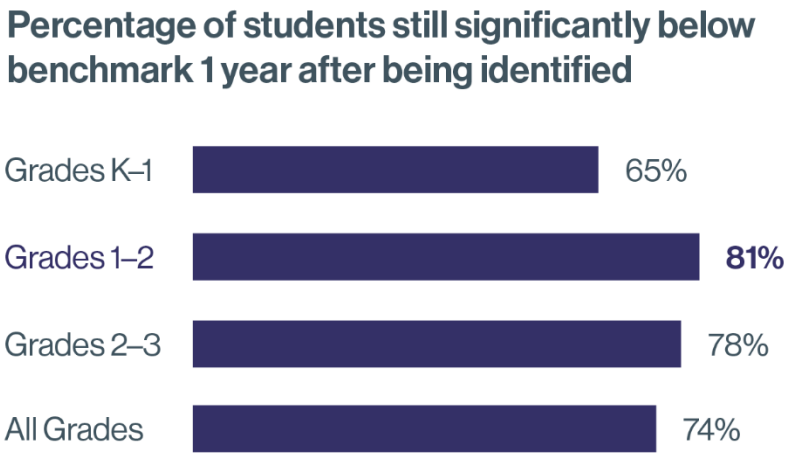
Note. Screening assessment data are available only for students in grades K–3, so we cannot track scores 2 years later for students who were screened in grades 2 or 3. *Source*: District-provided early literacy screening assessment data.

Of the students who had scores in 2022/23 and 2023/24, about one quarter were identified as at significant risk at EOY of 2022/23. By EOY of 2023/24, most of these students (74%) were still at significant risk (Figure 15).

As in previous analyses, more students in later grades remained significantly below benchmark across years than did students in earlier grades. For example, 65 percent of students who ended kindergarten significantly below benchmark were still significantly below benchmark at the end of grade 1 compared with 81 percent of grade 1 students who moved to grade 2 and 78 percent of grade 2 students who moved to grade 3.[20]

[20] Using the 25th percentile and below to identify students as at significant risk shows a similar pattern (62% of grade K students remained significantly below benchmark at the next grade level compared to 70% of grade 1 students and 73% of grade 2 students).

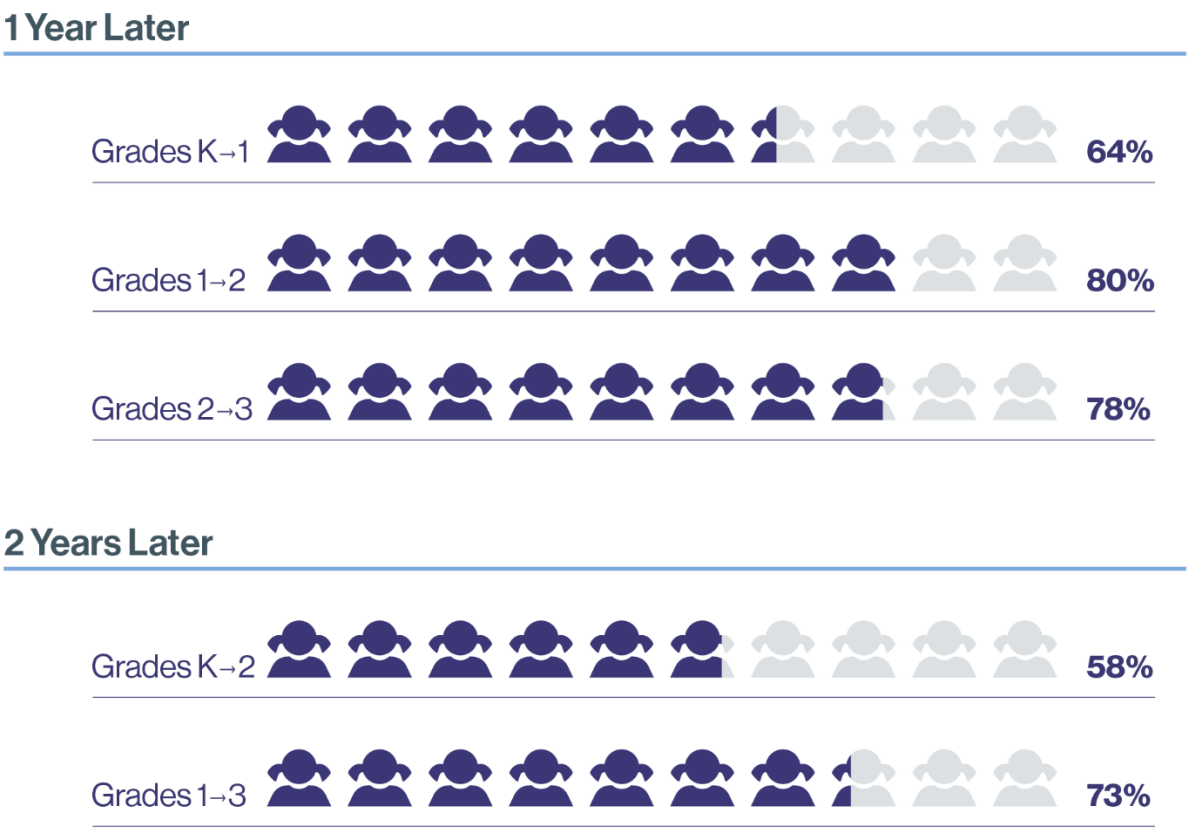
Figure 15. More Students At Higher Grade Levels Stayed At Risk from Grade to Grade Than Students At Lower Grade Levels



Note. Graph shows EOY to EOY performance across grades. For example, 65 percent of students significantly below benchmark at EOY in kindergarten were still significantly below benchmark at EOY of grade 1. *Source*: Early literacy screening assessment data.

Figure 15 shows students who had scores in 2022/23 and 2023/24. Of students who had EOY scores in kindergarten or grade 1 that were significantly below benchmark and EOY scores both 1 and 2 years later, most were also still classified as significantly below benchmark (Figure 16). Again, however, fewer students who were identified earlier were at significant risk than students identified later.

Figure 16. Most Students Identified as At Significant Risk in Grades K or 1 Were Still Significantly Below Benchmark 1 and 2 Years Later



Note. Graph shows EOY to EOY performance across grades. For example, 64 percent of students significantly below benchmark at EOY in kindergarten were still significantly below benchmark at EOY of grade 1, and 58 percent were still significantly below benchmark at EOY of grade 2.
Source: District-provided early literacy screening assessment data.

More English learner students, students from low income backgrounds, students receiving special education services, and Hispanic students remained significantly below benchmark across grade levels and years than did their peers not in those groups (Figures D5–D14 in Appendix D). For example, 72 percent of Hispanic students who ended kindergarten significantly below benchmark remained at that level at the end of grade 1 compared with 58 percent of non-Hispanic students.

These results indicate that students who finish a school year below benchmark are likely to continue below benchmark in subsequent years. This risk increases as students move from kindergarten through grade 3, highlighting the need for early, targeted interventions to disrupt these trajectories. They also indicate that the current supports, interventions, and opportunities provided to students performing below benchmark have not yet fully achieved desired improvements in student outcomes and progress.

How does student performance change over the summer?

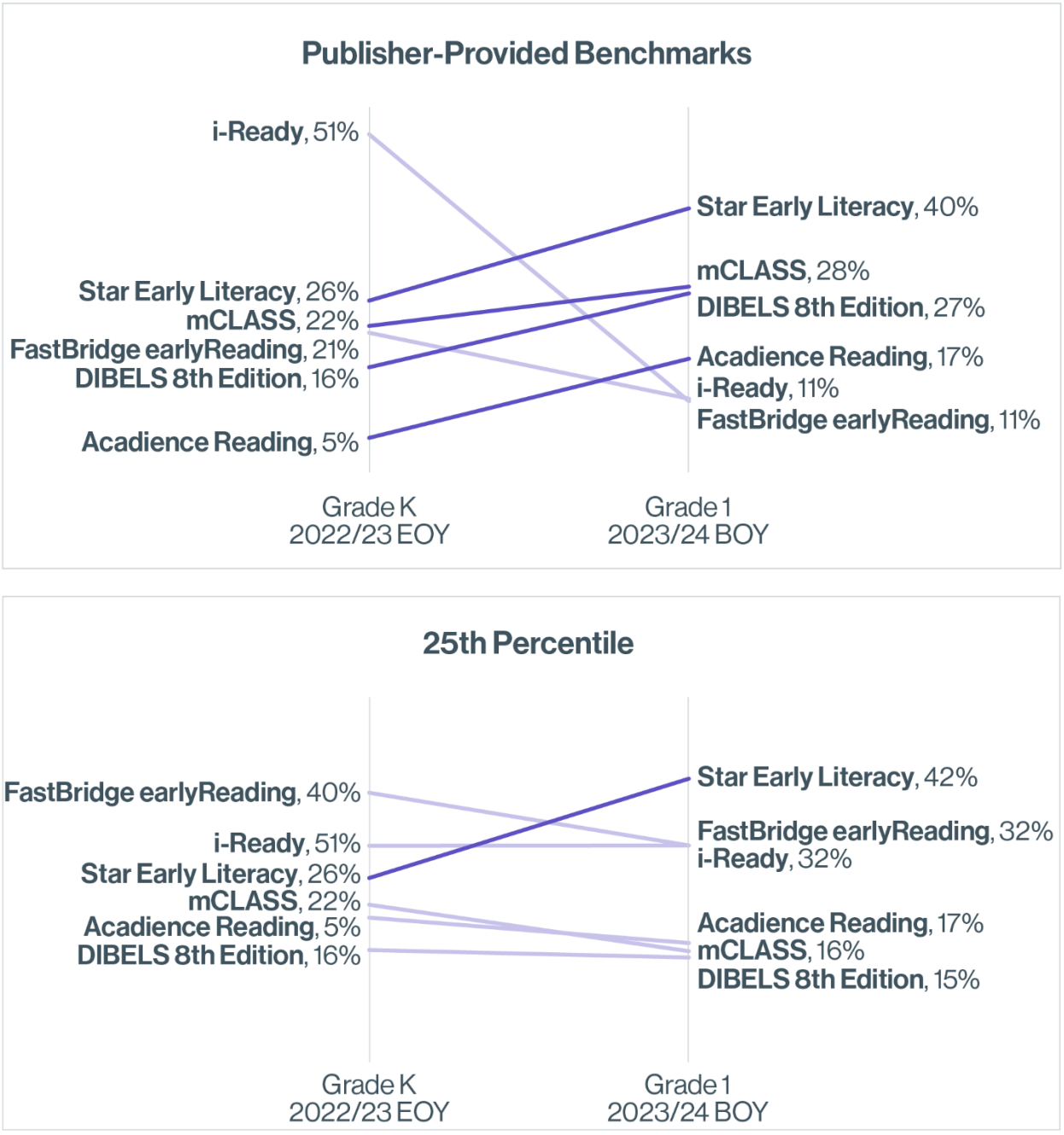
Key Findings

- Measuring summer changes in student risk is challenging due to differences in assessments and shifting benchmarks.
- Results vary across assessments. Publisher benchmarks generally show summer declines, whereas percentile benchmarks indicate stable or improved performance, highlighting the need for further research.

Focusing on a shorter time period—between EOY of 1 year and BOY of the next—addresses a common question about how student performance may change over the summer. However, changes in assessments, score scales across grade levels, and shifts in risk benchmarks over time complicate the interpretation of results. Figure 17 (next page) shows data for the most commonly used screening assessments for students with scores at EOY and the following BOY between grades K and 1.

In general, the publisher-provided benchmarks appear to suggest that student performance declines over the summer, such that the percentages of students identified as at significant risk at the beginning of the school year are larger than they were at the end of the year before. A few screening assessments show a different pattern (i-Ready, FastBridge earlyReading), in which the percentages of students identified as at risk decreases between EOY and BOY of the following year, which may be due at least in part to how those publishers set their benchmarks. However, examining the same data with the 25th percentile as the benchmark to identify students as at significant risk, however, shows the percentages staying the same or decreasing between EOY and the following BOY except for Star Early Literacy. Both FastBridge and Star Early Literacy (which use normative metrics as their publisher-provided benchmark) continue to show the same patterns of student performance, with FastBridge percentages of students at significant risk decreasing from EOY to the following BOY and Star Early Literacy increasing. Taken together, the inconsistent patterns suggest that additional research may be needed to determine the most accurate approaches to analyzing changes in the percentages of students identified as significantly below benchmark, particularly when analyzing transitions between grade levels.

Figure 17. Differences in Benchmarks Across Assessments Make Assessing Changes From EOY to the Following BOY Difficult to Measure



Note. Some students may appear multiple times per time period if they were administered multiple assessments. *Source*: Early literacy screening assessment data.

What is the relationship between screening assessment scores and MCAS performance in grade 3?

Key Findings

- The 2023/24 analysis confirms prior findings that show that most screening assessment benchmarks align with the MCAS level of partially meeting performance.
- Most students identified as significantly below benchmark on screening assessments were not proficient on MCAS.
- About one third of students who met publisher benchmarks were not proficient on MCAS.

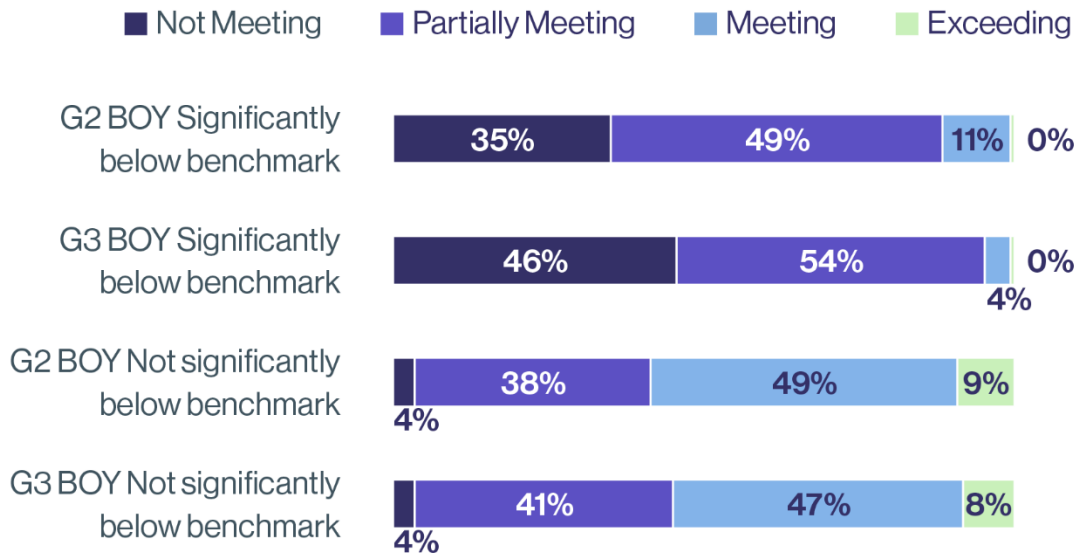
Prior analyses examined the relationship between student performance on screening assessments and later performance on MCAS in several different ways: examining the strength and accuracy of the predictive relationship between screener scores and MCAS performance through statistical analysis, linking screener benchmarks directly to the MCAS scale, and cross-tabulating the percentages of students at each MCAS performance level by screening assessment performance.

This analysis showed that despite differences between screening assessments and MCAS, scores correlate between 0.64 to 0.84 depending on the screening assessment, time of year, and grade level when administered. Linking to the MCAS scale shows that most screening assessment benchmarks identifying students as at significant risk of reading difficulty map to the MCAS Partially Meeting Expectations performance level. Additionally, screening assessment scores do discriminate between students who will meet and not meet MCAS proficiency standards, with classification accuracy ranging from about 75 to 89 percent. Unsurprisingly, given these findings, most students identified as significantly below benchmark on screening assessments did not meet expectations on MCAS.

The 2023/24 data show these same patterns. Table B74 in Appendix B provides MCAS-linked values. Figure 18 shows the MCAS performance of students who were identified as significantly below benchmark at BOY of grade 2 or grade 3 compared with peers who were not. About 11 percent of students in grade 2 and about 4 percent of students in grade 3 who were identified as significantly below benchmark at BOY met or exceeded standards on MCAS, with about one half in the partially meeting category and the remainder in the not meeting category. More students who were identified as significantly below benchmark in grade 2 met or partially met MCAS standards than did students who were identified as significantly below benchmark in grade 3. As described in the Comparability of Screening Assessments and Reporting section, most benchmarks indicating reading risk map to the MCAS partially meeting performance level, so these results are not surprising. However, as also previously shown, benchmarks that indicate little to no risk do not necessarily mean that all students above them will meet MCAS

standards. In 2023/24, about 65 percent of students who met screening assessment benchmarks also met MCAS standards (meaning about one third did not).

Figure 18. Relatively Few Students Identified as At Significant Risk on Grade 2 and Grade 3 Screening Assessments Met or Exceeded Expectations on Grade 3 MCAS ELA



Note. Some students may appear multiple times per time period if they were administered multiple assessments. For figure data, see Figure 18 in Appendix F. *Source*: Early literacy screening assessment data and state-provided MCAS data.

How do students identified as at risk in grades 1, 2, and 3 perform on MCAS in later grades (e.g., 3, 4, and 5)? Among students identified as at risk early (e.g., beginning of grade 1), how does MCAS performance in grade 3 differ between those who remain at risk over time and those who reach benchmark?

Key Findings

- Among students identified as at risk at EOY in 2021/22, few were proficient on MCAS 2 years later.
- More students who were identified as significantly below benchmark in earlier grades met MCAS expectations than those identified at later grades.

Prior analysis focused on grade 3 MCAS performance for students in grades 2 and 3 given the smaller numbers of students with data beyond grade 3. In 2023/24, additional data are

available to track student MCAS performance in later grades and also to examine the performance of multiple cohorts of students into grade 3 (Cohorts 2 and 3). Specifically, we can now examine grade 4 MCAS performance for two groups of students (Cohorts 3 and 4), examine Grade 5 performance for one group (Cohort 4), and examine MCAS performance both 1 and 2 years after students took screening assessments (Table 14).

Table 14. Available Data in 2023/24

Year	K	1	2	3	4	5
2020/21	Limited data	Limited data	Limited data	Limited data	Limited data	Limited data
2021/22	Cohort 1	Cohort 2	Cohort 3	Cohort 4	No screening data	No screening data
2022/23	Cohort 5	Cohort 1	Cohort 2	Cohort 3	Cohort 4	No screening data
2023/24	Cohort 6	Cohort 5	Cohort 1	Cohort 2	Cohort 3	Cohort 4

Source: Early literacy screening assessment data and state-provided MCAS data.

Among students identified as at risk at EOY in 2021/22, few met expectations on MCAS 2 years later, but as in prior analysis, more students who were identified as significantly below benchmark in earlier grades met expectations than did those who were identified later. Eleven percent of students identified in grade 1 met expectations by grade 3 compared with 10 percent of those identified in grade 2 and tested in grade 4, and only 6 percent of those identified in grade 3 tested in grade 5 (Figure 19). Similarly, more students who were identified as significantly below benchmark at grade 2 in either 2021/22 or 2022/23 met expectations as 3rd graders than did students identified at grade 3 in 2021/22 or 2022/23 as 4th graders (about 10% compared with 5%).

Figure 19. Rates of Meeting or Exceeding on MCAS Higher for Students Identified as At Risk in Earlier Grades

Note. Using the 25th percentile as an indicator of risk, percentages meeting or exceeding expectations on MCAS are 7 percent (grades 1 to 3), 6 percent (grades 2 to 4), and 2 percent (grades 3 to 5). *Source*: Early literacy screening assessment data and state-provided MCAS data.

To further examine patterns of screening assessment and MCAS performance, we analyzed data for a group of students who were identified as at risk early (BOY of grade 1), creating four categories: students who never got on track (i.e., never met benchmark) prior to grade 3, students who first got on track (i.e., met benchmark) at EOY of grade 1, students who first got on track at BOY of grade 2, and students who first got on track at EOY of grade 2. Not all students who met benchmark at these time periods stayed at benchmark; some students met benchmark but then were identified as at risk or at significant risk in a subsequent time period. Nonetheless, the students did get on track at least temporarily prior to MCAS and did so at different times.

Overall, using publisher-provided benchmarks, between 41 and 59 percent of students who met benchmark at some point after being identified as at risk at grade 1 BOY met expectations on MCAS compared with only 15 percent of students who never met benchmark in grades 1 or 2 (Tables 15 and 16). Using the 25th percentile and the 40th percentile as indicators of significant risk and some risk, the percentages of students who met expectations on MCAS ranged from 18 to 34 percent, with students who got on track earlier more likely to meet expectations on MCAS (Figure 20). Differences between publisher benchmarks within and across grade levels likely explain differences in the patterns of performance.

Table 15. Pathways for Students Identified as At Risk At BOY Grade 1 to Grade 3 MCAS, Publisher Benchmarks

Time period and grade when first on track	Number	% of students	% who met or exceeded expectations on MCAS
On track EOY G2	209	9%	41%
On track BOY G2	143	6%	59%
On track EOY G1	768	33%	50%
Never on track	1,178	51%	15%
Total	**2,298**	**100%**	**32%**

Source: Early literacy screening assessment data.

Table 16. Pathways for Students Identified as At Risk At BOY Grade 1 to Grade 3 MCAS, National Percentiles

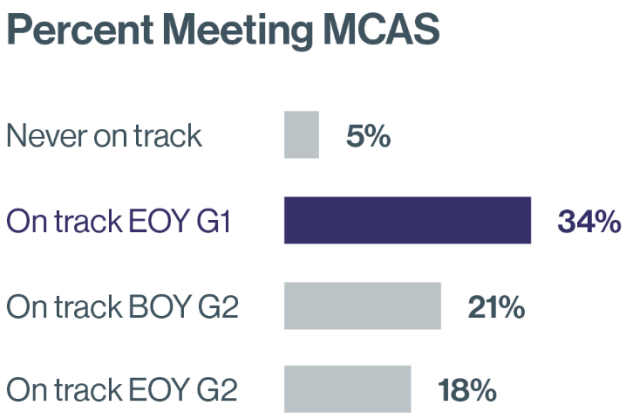
Time period and grade when first on track	Number	% of students	% who met or exceeded expectations on MCAS
On track EOY G2	80	8%	18%
On track BOY G2	103	10%	21%
On track EOY G1	422	42%	34%
Never on track	396	40%	5%
Total	**1,001**	**100%**	**20%**

Note. National percentile thresholds used in analysis were 25th and below for significant risk/significantly below benchmark and 40th percentile for at risk/below benchmark.
Source: Early literacy screening assessment data.

The most common pathways for students identified as below benchmark at BOY in grade 1 were to remain at risk through the remainder of grade 1 and grade 2 ("never on track") or to get on track by end of grade 1 ("on track EOY G1"). In total, over 80 percent of students followed these patterns. Only a small number of students got on track anytime after grade 1.

Among students who met benchmark, those who remained at benchmark were more likely to meet expectations on MCAS.

Figure 20. Students Who Get On Track and Who Get On Track Earlier Based on Screening Assessment Data Have Much Higher MCAS Meeting Expectation Rates Than Do Students Who Never Meet Screener Benchmarks



Note. Not all students who met benchmark at each time period stayed at benchmark. *Source*: Early literacy screening assessment data and state-provided MCAS data.

Along with other data, these findings once again point to the importance and benefit of helping students meet benchmark as early as possible, though it is important to note that differences in screening assessment benchmarks mean that performance pathways to MCAS may not be perfectly comparable across screening assessments.

School Characteristics and Student Performance

Key Findings

- School-level enrollment of students from low income backgrounds was significantly associated with higher rates of repeated reading risk identification. Schools with higher populations of English learners showed a marginal association with increased risk.
- District-level segregation measures were highly correlated with school characteristics.
- One racial segregation index (White–Minority isolation) was modestly but significantly associated with increased student risk after adjusting for other factors.
- Students just above and students just below benchmark cut scores showed similar MCAS outcomes, suggesting interventions might help maintain—but not significantly

improve—performance among at-risk students or that measurement limitations prevent detection of effects of any interventions.

- Some schools notably outperformed similar schools based on student demographics and resources, demonstrating lower risk rates than predicted by statistical models.

Earlier sections of this report focus on student-level performance and progress. However, research has long shown that school-level factors are associated with individual student performance. The characteristics of the communities in which schools are located, the backgrounds of the families of students in different schools, and the actions of and resources available to schools themselves can all influence student opportunities and learning.

Prior descriptive analysis of screening assessment data (see the Opportunity Gaps issue brief [Lemke et al., 2023]) showed that students enrolled in schools with the highest mobility rates, lowest attendance rates, highest discipline rates, fewest experienced teachers, lowest teacher retention rates, and highest percentages of historically marginalized student groups were more likely to be below benchmark and to stay there from BOY through EOY than were their peers in schools without those characteristics. Additional multilevel modeling analysis also showed that students in schools with above-average percentages of students from low income backgrounds and below-average teacher retention and student attendance rates were more likely to be identified as significantly below benchmark than were students in average schools.

In 2023/24, new analyses aimed to provide additional information about how school-level and district-level factors may influence student performance. Questions to be addressed include the following:

- How do school-level characteristics relate to the risk of being identified as at significant reading difficulty multiple times?
- To what extent does district-level school segregation (as measured by indices of concentration of poverty and race/ethnicity) moderate the relationship of school-level factors to student performance?
- How does MCAS performance differ for those just below compared with those just above screener benchmarks identifying students as at risk?
- Are there schools where students are “outperforming” expectations in terms of achievement and/or growth within the school year, based on the demographics of students and/or schools or other factors?

How do school-level characteristics relate to the risk of being identified as at significant reading difficulty multiple times?

The multilevel statistical model described in the Student Performance section and Appendix E also examined how school-level factors relate to students' likelihood of being identified as significantly below benchmark more than once during the school year.

The model included four school-level variables:

- Percentage of students from low income backgrounds
- Student mobility rate (measured as schoolwide student stability)
- Average student attendance rate
- Percentage of English learner students

Among these factors, the percentage of students from low income backgrounds in a school showed a strong association with repeated risk identification. Specifically, for every one standard deviation increase in the percentage of students from low income backgrounds (about 25 percentage points), the odds of a student being identified as at risk multiple times increased by 37 percent (Odds ratio = 1.37, $p < .001$).

Other school-level variables (student mobility, attendance rate, and percentage of English learner students) did not reach conventional levels of statistical significance, though some showed marginal associations. For example, higher attendance rates were modestly associated with a lower likelihood of repeated identification ($p = .06$), while schools with higher percentages of English learner students showed a trend toward increased risk ($p = .06$). These patterns may merit further investigation but did not represent statistically significant effects in this model.

To what extent does district-level school segregation (as measured by indices of concentration of poverty and race/ethnicity) moderate the relationship of school-level factors to student performance?

Schools and their districts help shape the conditions for student learning. Segregation by race/ethnicity and income can influence students' access to opportunities, including experienced teachers, rigorous coursework, and adequate resources, all factors known to impact academic achievement and long-term outcomes (e.g., Owens, 2018; Reardon, 2016).

To explore these dynamics, we incorporated measures of district-level school segregation into our analysis of factors predicting students' repeated identification as significantly below benchmark. Using 2023/24 data and segregation indices from the Education Opportunity Project at Stanford University (Reardon et al., 2025), we examined the extent to which patterns of racial and income-based school segregation at the district level correspond to differences in school composition and risk, after adjusting for student and school-level attributes.

These segregation indices quantify the extent to which students from different groups attend school together within a given district. Values range from 0 to 1, with higher values indicating greater isolation. For example, a Black-White isolation index of 0.70 indicates that Black students attend schools where, on average, 70 percent of the student body (considering only Black and White students) is Black.

Table 17 shows how segregation indices correlate with key school-level characteristics. Racial segregation (e.g., White–Black, White–Hispanic) is strongly negatively correlated with indicators such as the percentages of students from low income backgrounds (e.g., –0.81), English learner students (e.g., –0.69), and high-need students (e.g., –0.84). These patterns suggest that decreasing concentrations of White students at the district level are associated with higher school-level percentages of historically underserved student groups. District-level school racial segregation also shows moderate to strong positive associations with school conditions typically associated with higher achievement, such as higher attendance rates and teacher retention, particularly for White–Black and White–Hispanic indices.

In contrast, income-based segregation (e.g., comparing free/reduced-price lunch students to those not eligible) is positively correlated with school-level socioeconomic and linguistic diversity. For example, the FRL–non-FRL segregation index correlates at 0.90 with percentage of students from low income backgrounds and 0.67 with percentage of English learners. These findings suggest that income segregation in schools within a district often coincides with concentrations of students in schools who often face systemic barriers to educational opportunity.

Because the segregation indices are correlated with one another, we selected one racial segregation measure (White–Minority isolation) and one income-based measure (FRL–non-FRL isolation) for inclusion in a multilevel logistic regression model predicting repeated risk identification.[21] These variables were added to the model described earlier that already included student and school-level predictors.

In this model, White–Minority segregation was modestly but significantly associated with increased risk (estimate = 0.12, t = 2.26, p = .02), indicating that students in more racially segregated contexts were slightly more likely to be identified as significantly below benchmark multiple times. The FRL–non-FRL segregation measure, however, was not statistically significant (p = .41).

These results suggest that within district segregation may influence student outcomes primarily through its effects on school composition, many examples of which (e.g., concentration of students from low income backgrounds, percentage of English learner students) were already included in the model. While district-level school segregation measures added some

[21] According to Education Opportunity project documentation, “Minority” includes Black, Hispanic, and Native American students (Documentation for Segregation Estimates, Segregation Explorer Version 1.0).

explanatory value, especially for racial isolation, their incremental effect was relatively small when controlling for school-level characteristics.

Table 17. Correlations Between School Attributes and District-Level School Segregation Measures

Segregation index measure	% churn	% LI students	% EL students	% avg attendance rate	% students disciplined	% teacher retention	% high needs	% stability	Avg absences
White—Black	–0.56	–0.68	–0.51	0.47	–0.23	0.34	–0.68	0.54	–0.45
White—Hispanic	–0.65	–0.81	–0.69	0.59	–0.34	0.41	–0.82	0.61	–0.57
White—Asian	–0.31	–0.19	–0.46	0.08	–0.09	0.19	–0.30	0.31	–0.06
Seg Wht—NAM	–0.19	–0.38	–0.32	0.22	–0.08	0.22	–0.38	0.20	–0.22
Seg Wht—Min	–0.66	–0.82	–0.69	0.58	–0.34	0.41	–0.84	0.62	–0.55
Seg WAS—Min	–0.64	–0.83	–0.67	0.58	–0.34	0.41	–0.83	0.59	–0.56
White—Non-White	–0.65	–0.76	–0.70	0.52	–0.31	0.40	–0.80	0.62	–0.50
FRL—non-FRL	0.70	0.90	0.67	–0.63	0.33	–0.40	0.90	–0.65	0.61

Note. The "% churn" is the percentage of students who transfer in and out of the district in a school year. The "% LI students" is the percentage of students in the school classified as being from a low income background. The "% EL students" is the percentage of students classified as English learners. The "% attendance rate" is the percentage of days attended in the school year. The "% students disciplined" is the percentage of students disciplined. The "% teacher retention" is the percentage of teachers working in the same position across 2 school years. The "% high needs" is the percentage of students classified as at least one of the following: being from a low income background, being an English learner, or having an IEP. The "% stability" is the percentage of students remaining in the same school throughout the school year. The "Avg absences" is the average number of days students are absent. White–Black isolation measures how concentrated White students are in their typical school environment within a district, considering only White and Black students. White—Hispanic isolation measures how concentrated White students are in their typical school environment within a district, considering only White and Hispanic students. White—Asian isolation measures how concentrated White students are in their typical school environment within a district, considering only White and Asian students. Seg WH—NAM isolation measures how

concentrated White students are in their typical school environment within a district, considering only White and Native American students. Seg WH—Min isolation measures how concentrated White students are in their typical school environment within a district, considering only White and Minority students. Seg WAS—Min isolation measures how concentrated White and Asian students are in their typical school environment within a district, considering only White, Asian, and minority students. White–Non-White isolation measures how concentrated White students are in their typical school environment within a district, considering only White and Non-White students. FRL–non-FRL isolation measures how concentrated FRL-eligible students are in their typical school environment within a district. *Source*: Massachusetts public school profile data and education opportunity segregation index data.

How does MCAS performance differ for those just below compared with those just above screener benchmarks identifying students as at risk?

The goal of early literacy screening is to identify students who may benefit from additional support in getting on track as readers. Under a Multi-Tiered System of Supports (MTSS), students identified as significantly below benchmark are expected to receive intensive intervention, while their peers just above the benchmark may not. Although we lack data on whether or how services were provided, we conducted an exploratory analysis comparing outcomes for grade 3 students just below versus those just above each benchmark.

We used a quasi-experimental method, regression discontinuity design (RDD), to examine differences in grade 3 spring MCAS scores for students near the cut scores on BOY screening assessments. Two analyses were conducted: one focused on students near the significantly below benchmark cut score (within ±0.5 SD, n = 8,933) and another focused on students near the below benchmark cut score (within ±0.9 SD, n = 15,926). Because students took MCAS at the end of the year, this design allows time for any interventions to take effect. The analysis was conducted separately by screening assessment given the known differences in risk benchmarks.

Overall, students just above and students just below the benchmark cut scores performed similarly on MCAS. One exception was in i-Ready (below benchmark), in which students above the cut score scored 4.2 scale score points higher on average than those below (t = 2.89, p < .01). We also found that for DIBELS 8th Edition, the relationship between screening scores and MCAS outcomes was significantly stronger for students below the cut score than for those above, with a slope difference of approximately 9 points (t = –5.03, p < .001).

Some screening assessments, including i-Ready and Star Reading, also showed stronger screening–MCAS relationships for students below the cut score compared with those above. However, most other screener-specific effects were not statistically significant.

In the analysis using the significantly below benchmark cut score, we found no statistically significant differences in MCAS outcomes for students just above versus those just below the cut score across screening assessments. While there were descriptive differences in slope across some screening assessments—for example, FastBridge aReading appeared to have a negative relationship below the cut and a positive one above—these differences were not statistically significant.

Importantly, the absence of a detectable MCAS difference for students significantly below benchmark does not imply that early literacy interventions are ineffective. First, MCAS scores may not detect gains among the lowest performing students well. A student could move from very low to moderately low reading ability without showing a measurable MCAS gain. Second, interventions may be preventive, helping students maintain their skills rather than fall further behind. Such benefits would not be captured in our comparison unless a true control group

(with no services) were available. Therefore, these results should be interpreted with caution and not as evidence against the value of early intervention.

Are there schools where students are "outperforming" expectations in terms of achievement and/or growth within the school year, based on the demographics of students and/or schools or other factors?

Educational researchers have long sought to identify schools that "beat the odds"—that is, schools that achieve better outcomes than would be expected based on the demographics and characteristics of the students they serve. In 2023/24, we used a multilevel logistic regression model to identify such schools, statistically adjusting for differences in student background, school context, and district-level school segregation. This approach enables comparisons across schools that are more valid by identifying those that perform better than predicted, not simply those with the lowest rates of risk.

Using the school-level random effects from our model, we identified seven schools with the lowest adjusted likelihood of students being flagged as significantly below benchmark more than once during the year (Table 18). The analysis was conducted both with and without segregation measures, and the same schools were identified in both cases, suggesting a robust pattern. These schools represent the strongest outliers in terms of performance relative to their contexts rather than absolute performance alone. Several additional schools had low rates of risk identification, but either they served relatively advantaged populations or we could not identify more than two similar schools for comparison within the screener data sample.

The range in risk identification among these outperforming schools is substantial, spanning from 14 to 28 percent.

Most schools on the list serve moderate- to high-need student populations. School 1 and School 2 both serve student bodies in which more than 70 percent of students are from low income backgrounds, and yet they report risk rates of only 14.9 and 16.5 percent, respectively. These figures are lower than predicted based on their school contexts, suggesting that effective practices may be helping to buffer students from academic risk.

Several schools on the list serve some of the highest need student populations in the state. For example, School 7 and School 6 serve student populations with about 80 percent of students from low income backgrounds and English learner rates exceeding 30 percent. Despite serving a population of students with relatively high need, their risk rates were significantly lower than expected after adjusting for all student, school, and district variables. Their presence on this list illustrates that having high populations of students from historically marginalized groups does not preclude strong performance.

Across schools, certain features appear consistently. Every school maintained an attendance rate above 90 percent, with most exceeding 93 percent. Discipline incidents were rare, with the vast majority of schools reporting that fewer than 2 percent of students had disciplinary events during the year. Notably, many of these schools also serve high percentages of students

classified as high-need—including students who are from low income backgrounds, are English learners, or receive special education services. The fact that these schools emerged as outperformers after adjusting for those characteristics provides compelling evidence that strong outcomes are possible even in schools serving complex student populations.

Table 18. Schools Identified as Outperforming

School identifier	% significantly below benchmark multiple times	% LI students	Average school attendance rate	Number of students disciplined	% EL students	% high needs students	Number of students
1	14.9	71.2	94.5	0.2	26.9	82.3	483
2	16.5	75	94.2	0.0	14.8	81.1	176
3	19.6	61.8	92.7	1.6	35.5	82.2	176
4	24.4	64.4	92	1.9	46.5	83.6	331
5	25.2	66.1	93.8	0.5	49.2	85.3	203
6	25.5	78.7	92.8	1.4	39.3	88.5	359
7	28.0	83.3	92.6	0.0	31.8	91.4	139

Source: Early literacy screening assessment data and October and June SIMS collection data.

To further illustrate how these outperforming schools differ from their peers, we identified comparison schools with similar demographic and contextual characteristics. For each outperforming school, matches were selected based on a set of proximity criteria: comparison schools had to fall within 15 percent of the focal school's values on key variables—percentage of students from low income backgrounds, attendance rate, percentage of English learners, and percentage of high needs students—and within 30 percent of enrollment size. For instance, a school with 50 percent students from low income backgrounds would be compared with schools with rates between 42.5 and 57.5 percent.

Table 19 summarizes these comparisons. Of the seven outperforming schools, all had at least four comparable schools.

The results show consistent and often substantial differences in the percentage of students identified as significantly below benchmark multiple times ranging from about 11 to 27.5 percentage points in favor of the outperforming schools.

The size of these gaps is particularly notable among schools serving high-need populations. For example, School 2 had 16.5 percent of students identified as multiply at risk compared with 34.6 percent in its matched peers—a gap of 18.1 percentage points. Similarly, School 1 reported a rate of 14.9 percent versus 42.3 percent in comparison schools, a difference of 27.4 percentage points. Other high-need schools, such as Schools 6 and 4, also showed sizable differences of 23.0 and 19.9 percentage points, respectively.

These comparisons reinforce that the identified schools are not simply succeeding because of advantageous demographics. Instead, they represent clear exceptions to prevailing performance patterns. The results suggest that substantial improvements in early literacy outcomes are achievable—even in high-need contexts—when schools implement effective instructional practices and foster strong, supportive learning environments.

Table 19. Comparison School Percentages of Students Significantly Below Benchmark Multiple Times

School identifier	% significantly below benchmark multiple times	Comparison % significantly below benchmark multiple times (average)	Number of comparison schools	Difference in % of significantly below benchmark
1	14.9	42.3	4.0	–27.5
2	16.5	34.6	6.0	–18.2
3	19.6	30.3	8.0	–10.7
4	24.4	44.3	8.0	–19.9
5	25.2	41.2	7.0	–15.9
6	25.5	48.6	8.0	–23.0
7	28.0	45.4	7.0	–17.4

Source: Early literacy screening assessment data and October and June SIMS collection data.

English Learner Student Performance

Key Findings

- The vast majority of English learners in the available data are screened only in English (about 10 percent of English learner students took Spanish language screening assessments).
- Star screening assessments were the most commonly administered Spanish-language assessments in the screening assessment data.
- Most students who took Star Spanish-language assessments were English learners taught primarily in English with a home language of Spanish.
- Risk identification rates for English learners remained relatively stable throughout the year, particularly for students assessed in Spanish, suggesting limited change in Spanish literacy skills over the school year.
- Screening in both languages generally yielded consistent results, though differences emerged based on assessment level and language.
- Students screened with mismatched assessment levels (different levels in English versus Spanish) showed distinct risk patterns, emphasizing the need for careful interpretation of such data at the school level.

Massachusetts's K–3 population includes several thousand students who speak Spanish as a home language, participate in Spanish-English dual language bilingual education (DLBE) programs, take Spanish-language literacy screening assessments, or engage in some combination of these three activities. For a variety of reasons, these students' screener results may differ from students who speak, test, and learn only in English. For example, students participating in DLBE programs have been shown to follow different language development and achievement trajectories over time (Umansky & Reardon, 2014) compared with multilingual students educated only in English. Based on these differential trajectories, literacy norms based on monolingual students receiving monolingual English instruction may be inappropriate or inaccurate for predicting whether DLBE students are at risk or are developing appropriately. Relatedly, research on multilingualism generally confirms that individuals have different strengths, knowledge, and profiles in each language (as opposed to knowing and being able to do exactly the same things in each language) such that measuring only one language may paint a distorted or incomplete picture of students' language abilities (Faulkner-Bond et al., 2020; National Academies of Sciences, Engineering and Medicine, 2017).

DESE's *Early Literacy Screening Guidance* (DESE, 2023) and *Massachusetts Dyslexia Guidelines* (DESE, 2020) both include several recommendations for screening multilingual learners. These recommendations include screening students in both their home language and English when

appropriate tools are available, supplementing assessment-based information with observation-based protocols, and comparing these students' results to a set of more closely matched "true peers" rather than to the full population of primarily monolingual English speaking students. A general theme across these recommendations is to collect additional information about the literacy development of multilingual students when possible and to interpret their results with more nuance and context than may be needed for monolingual English students.

Following up on preliminary analyses conducted with 2020/21 and 2021/22 data with the larger data set available in 2023/24, we further explore the characteristics, achievement profiles, and trajectories of students who participated in Spanish-language screening in grades K–3. The goals and research questions of these analyses include the following:

1. Understand which students participate in these assessments.
    a. How many students participate in Spanish-language screening assessments?
    b. What Spanish-language screening assessments are used in the state?
    c. What are the characteristics of students who participate in Spanish-language screening assessments?
    d. To what extent are students who participate in Spanish-language screening assessments also being taught in Spanish?
2. Describe achievement profiles in each language and relative to the profiles of students who are screened in English, and explore whether students' language of instruction or time in English learner status is associated with different achievement profiles, trajectories, or predictive relationships.
    a. To what extent are through-year score patterns on Spanish-language assessments similar to patterns on English-language assessments?
    b. To what extent do Spanish-language screening assessment outcomes agree with English-language versions? When the results of each screener diverge, are there any patterns to the disagreement?

How many students participate in Spanish-language screening assessments?

Available data suggest that screening students in languages other than English is still a rare practice in the state. In the screening data sample, only 2.4 percent of all students and approximately 9 percent of all English Learners took a Spanish language screener at any point

during the 2023/24 school year.[22] These percentages suggest that the most common experience for all students, regardless of what languages they use in school or at home, is still to be screened in English only.

What Spanish-language screening assessments are used in the state?

Many of the approved screening assessment publishers provide Spanish language assessments that are designed to function similarly to their English-language counterparts. In the 2023/24 school year data, data for four of these Spanish-language screening assessments were available: FastBridge earlyReading Spanish, mCLASS Lectura, Star Early Literacy Spanish, and Star Reading Spanish. A total of 3,561 unique students (2.4% of all students in our sample) participated in one of these four Spanish-language screening assessments during the 2023/24 school year (Table 20).[23]

Of these, the two Star assessments were by far the most widely used in the screening assessment data sample, representing 95 percent of all students who participated in Spanish-language screening assessments (3,372 students). The mCLASS Lectura and FastBridge earlyReading Spanish accounted for <1 percent and 6 percent of students tested in Spanish, respectively. Based on the dominance of the Star assessments and the small sample sizes available for the other two Spanish-language screening assessments, in this section we focus only on students taking the Star assessments.

Table 20. Number and Percentage of Students Who Participated in Spanish-Language Screening Assessments in 2023/24

Early literacy screening assessment	Number of participants	Percentage of all participants	Percentage of Spanish-language participants
FastBridge earlyReading Spanish	3	<0.1%	<0.1%
Star Early Literacy Spanish	2,533	1.7%	71.1%
Star Reading Spanish	839	0.56%	23.6%
mCLASS Lectura	186	0.12%	5.2%
Total	**3,561**	**2.4%**	**100%**

[22] The screening data sample includes about half of the schools serving K–3 students in the state, but it is possible that some schools reported only English-language screening assessment data.

[23] Note that these numbers include some students who took "off grade level" Spanish-language screening assessments such as Star Early Literacy in grades 2 or 3.

Note. Students may appear in multiple rows if they took more than one screening assessment.
Source: Early literacy screening assessment data.

What are the characteristics of students who participate in Spanish-language screening assessments?

The 3,372 students who participated in Star Spanish-language screening assessments were largely homogeneous on several key characteristics (Figure 21). First, and unsurprisingly, the vast majority of these students (3,171; 94%) spoke Spanish as their home language. The second most reported home language in this group was English—a common finding in English learner data (e.g., see NCES, 2024) that is generally interpreted as representing students who live in multilingual households or students who were raised speaking another language but currently live in households where English is spoken. Fewer than 3 percent of students reported speaking a language other than Spanish or English at home.

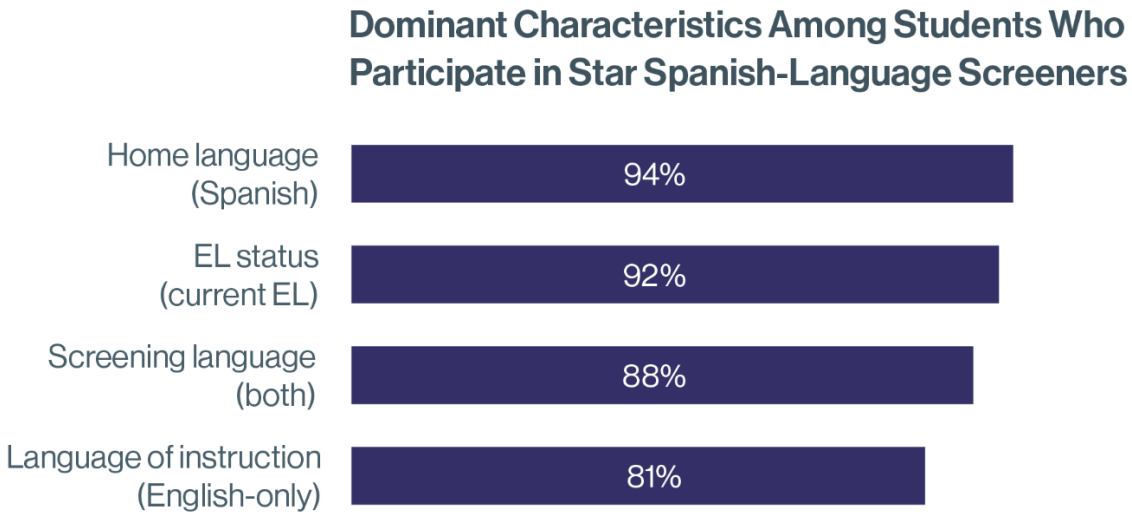
Second, and also not surprisingly, the vast majority of students participating in Star Spanish-language (3,100 students; 92%) were English learners at the time they were screened. Approximately 7 percent of participants (235 students) had never been English learners, and just over 1 percent of participants (37 students) were former English learners.

Third, the vast majority of students participating in Star Spanish-language (3,216; 95%) participated in assessments in both English and Spanish. These data suggest that most districts that screen students in Spanish are adhering to DESE's recommended practice of screening students in all of their languages to get a picture of student skills and capabilities across their full linguistic repertoire.

Finally, as we discuss more in the next section, the majority of students who took these screening assessments (2,757 individuals; 82%) were enrolled in English-only instructional programs (structured English immersion [SEI] programs).

In other respects, the population of students participating in Spanish-language screening assessments largely resembled the population of students taking English-language screening assessments: 50 percent (1,688) were female, 13 percent (433) received special education services, and 85 percent (2,864) were classified as coming from low income backgrounds.

Figure 21. Most Students Participating in Star Spanish-Language Screening Assessments Were Current English Learners, Were Screened in English and Spanish, Had a Home Language of Spanish, and Were Taught in English



Note. The first bar shows the percentage of EL students whose native language is Spanish. The second bar shows the percentage of EL students who are current EL students (with the remainder former or never English learners). The third bar shows the percentage of students taking Spanish and English. The fourth bar shows the percentage of students in different EL programs who are in SEI programs (other programs include dual language, not enrolled, parent opted out). *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Among students who took Star assessments in Spanish, one other notable characteristic within the sample is the prevalence of "off grade" testing, particularly in Grade 2 (Table 21). As noted elsewhere in this report, Star Early Literacy is generally recommended for students in grades K–1, while Star Reading is recommended for most students in grades 2–3. Table 21 shows that these recommendations are generally being followed. Seventy-two percent of all students (2,131 individuals) took either the Early Literacy assessments in grades K–1 (1,423 students) or the Reading assessments in grades 2–3 (708 students). However, the next largest group of students was 1,678 students (57% of all students tested in both languages) in grades 2–3 who took the Early Literacy screener in one (755 students) or both (923 students) languages, despite this screener being recommended for earlier grades only. It is possible that this trend reflects a practice of screening students considered at risk in English using the more foundational assessments as a check, though this could not be confirmed in the data. As Table 21 shows, this group did not stand out from others based on their disability status, nor did they appear to participate in DLBE programs at notably higher rates than other groups. This group (students in grades 2–3 who were also screened using the Early Literacy assessment in English) did have a slightly higher proportion of students classified as new to U.S. schools, suggesting

that schools may also be using this practice to assess foundational skills for students who are enrolling in the U.S. school system for the first time after the earliest grades.

Table 21. Percentage of Students Taking "Off Grade Level" Star Assessments in Spanish and/or English

Student grade levels	Assessment	All (*N*)	All (%)	Receiving special education services	SEI	DLBE	1st year in U.S.
K–1	On grade in both languages	1,423	37%	10%	81%	13%	56%
2–3	On grade in both languages	708	18%	14%	87%	4%	8%
K–1	"Off-grade" (above) in Spanish	14	<1%	0%	43%	29%	36%
K–1	"Off grade" (above) in English	55	1%	5%	84%	4%	27%
K–1	"Off grade" (above) both languages	Sup data	Sup data	Sup data	Sup data	Sup data	Sup data
2–3	"Off grade" (below) in Spanish	432	11%	14%	94%	0%	9%
2–3	"Off grade" (below) in English	323	8%	17%	89%	3%	16%
2–3	"Off grade" (below) both languages	923	24%	17%	87%	6%	14%

Note. Students may appear in multiple rows if they took more than two screening assessments. "Sup data" means that data for student groups with fewer than 10 students are not shown in

order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

To what extent are students who participated in Spanish-language screening assessments also being taught in Spanish?

Spanish language screening may be particularly appropriate for students who are participating in DLBE programs, in which anywhere from 50 to 90 percent of instruction (including instruction of academic content) is provided in a language other than English (LOTE) or a partner language. Since students in DLBE programs are used to learning and expressing their ideas in a LOTE, they may be more likely to benefit from an assessment that invites their response in the language they are accustomed to using in school. By contrast, Spanish-language assessments may be less helpful for students—including multilingual students—who are instructed only in English because these students may be unaccustomed to using their home language in the context of academic activities and expression. For example, multilingual students who have been taught "heart words" or "sight words" in English only may not be able to easily identify these same words if presented with them Spanish.

DLBE Programs and Related Research

Depending on the model it uses, a DLBE program may enroll all or exclusively English learners (a one-way immersion program), all or exclusively English-only students (a one-way immersion world language program), or a mix of English learners and English-only students (a two-way immersion program). Research (Bilber, 2022; Morales, 2024; Palacios et al, 2024; Steele et al, 2017; Umansky & Reardon, 2014; Valentino & Reardon, 2015) has consistently found evidence that DLBE programs produce comparable or superior academic and linguistic outcomes in English for English learner students over time, with the additional advantages of social–emotional benefits and full bilingualism and biliteracy in the partner language.

Starting in 2018 with the passage of the Language Opportunities for Our Kids (LOOK) Act, DESE has sought to increase the number of DLBE programs in the state and thereby expand students', and particularly English learners', access to such instructional opportunities. As these expansion efforts continue, a majority of all students in Massachusetts, including English learners, continue to receive instruction in monolingual English learning environments.

In the 2023/24 school year, DESE reported 2,790 K–3 English learners (5.6%) enrolled in DLBE programs compared with 46,483 (93.1%) in SEI programs. Given this context, it is perhaps unsurprising that, among students who took Spanish-language screening assessments, the vast majority of students (2,528 students; 89%) were enrolled in SEI programs to help them learn English (Table 22). Just over 10 percent of participants (324 individuals; all but one student

currently classified as English learners) participated in DLBE programs. This finding is important for contextualizing the results that are reported in the remainder of this section because a student's language of instruction is likely to interact with and affect their literacy development in that language and the potential validity and utility of their scores on screening assessments in that language.

Table 22. Number of Students Who Took Star Spanish-Language Screening Assessments by English Learner Status and English Language Education (ELE) Program

ELE program	Current EL	Never EL	Former EL
No program	0	210 (100%)	33 (97%)
Structured English immersion	2,528 (89%)	0	0
Dual language education	323 (11%)	0	Sup data
Guardian opted out	Sup data	0	0
Total	**2,851**	**210**	**334**

Note. Dual language education includes DLBE, other bilingual programs, and transitional bilingual education. Never EL is defined as a student who was never classified as an English Learner. Former EL is defined as a student who is not currently an English learner but who has been in the past. "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

To what extent are through-year score patterns on Star Spanish-language assessments similar to patterns on English-language assessments?

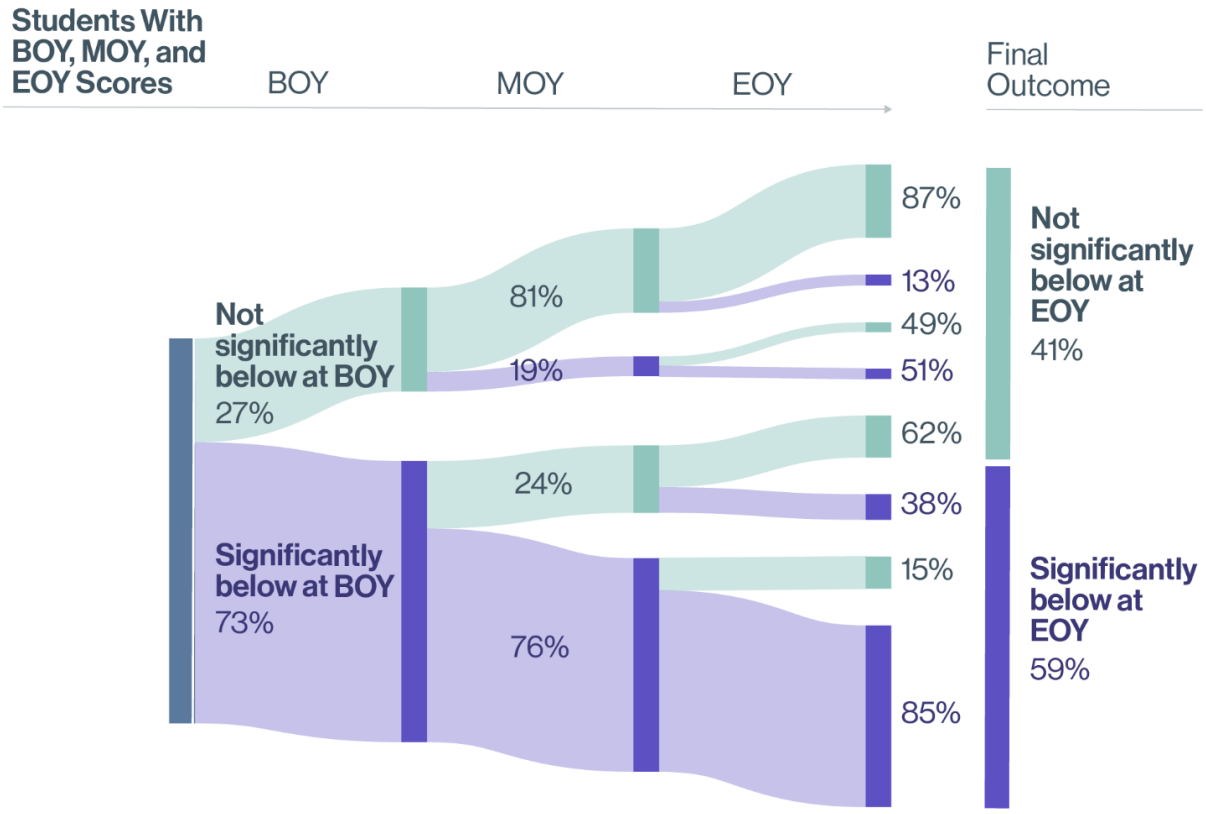
Current and prior reporting on early literacy screening assessment data has found that, in general, students tend to stay in the same risk category over the course of the year. In the 2023/24 school year, for example, 61 percent of students who began the year significantly below benchmark were still at that level at the end of the year, while 89 percent of students who began the year not significantly below ended the year at the same status (not significantly below). English learner students who take the English-language Star assessments are even more likely to remain significantly below benchmark if they begin the year that way (73% of students end the year in this status compared with only 61% among all students).

Among the subset of English learner students who had a score on an English-language Star screener during each time period of the school year (5,189 students; 65% of the sample), the proportion not significantly below benchmark increased from 27 percent at BOY to 41 percent at EOY, while the proportion of students significantly below benchmark decreased from

73 percent at BOY to 59 percent at EOY (Figure 22). This decrease is largely driven by students who improved at MOY and then remained not significantly below benchmark at EOY. In general, students who started significantly below benchmark remained significantly below benchmark across the year.

Although the patterns observed for this subset of English learner students are similar to the pattern in the overall population (Figure 22), these students showed more movement in their risk statuses over the course of the year compared with the overall population (Figure 12). They were also more likely to go from not being at significant risk to being at significant risk in the following time period than the overall population was. For example, in the overall student population, among students who were not significantly below benchmark at MOY and at BOY, 95 percent were not significantly below benchmark at EOY (Figure 12); this drops to 87 percent among the English-screened English learner students. So, although it appears there are improvements in performance across the school year for English learner students screened in English, positive performance is less consistent throughout the school year than among the overall sample of students.

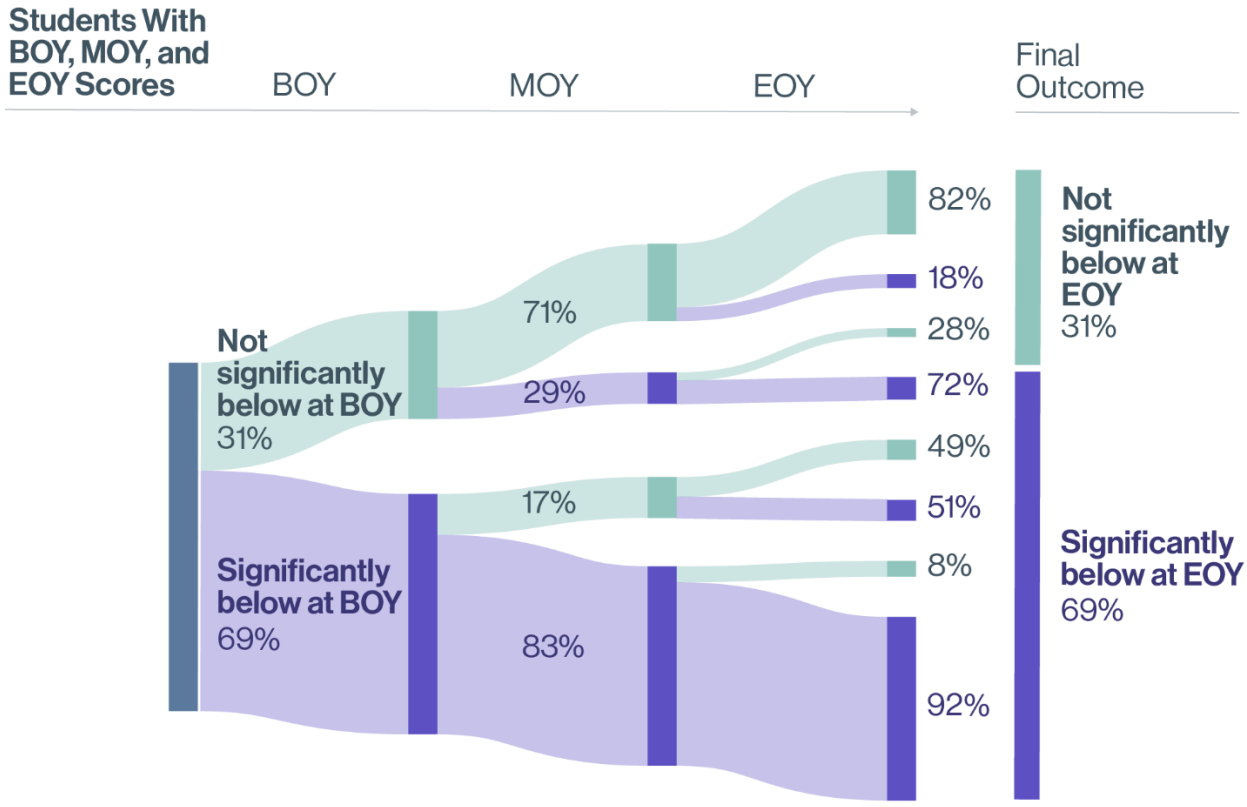
Figure 22. Most English Learner Students Who Were Screened in English Maintained the Same Risk Level From BOY to EOY



Note. See Figure 22 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

Interestingly, English learner students who took Star Spanish-language screening assessments were also likely to end the year at the same risk level—but with more downward movement compared with those who took Star English-language screening assessments. For example, 87 percent of students who started the year not significantly below benchmark on an English-language screener remained there at EOY compared with 82 percent of students who took Spanish-language screening assessments. Additionally, the percentage of English learner students significantly below benchmark dropped from 73 percent in BOY to 59 percent in EOY for students taking an English-language screener but stayed virtually the same for English learner students taking a Spanish-language screener (Figure 23).

Figure 23. The Proportion of Students At Risk Based on Spanish-Language Screening Assessments Was Largely Stable From BOY to EOY



Note. See Figure 23 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

This lack of change from BOY to EOY in the proportion of students at significant risk is explained by student shifts in performance over time. Among English learner students who started not significantly below benchmark in Spanish at BOY, nearly a third (29%) moved to a lower performance level at MOY (significantly below), and a majority of these (72%) remained in that status at EOY. Similarly, although 17 percent of students who started significantly below

benchmark in Spanish at BOY moved to a higher performance level at MOY, roughly half of these students (51%) moved back down to significantly below benchmark at EOY. These patterns may stem from the fact that the vast majority of students being screened in Spanish are not being instructed in Spanish. While students' development of their home language presumably continues outside of school, their knowledge, and ways of using Spanish may not align with the skills assessed on a literacy screener.

To what extent do Spanish-language screening assessment outcomes agree with English-language versions? When the results of each screener diverge, are there any patterns to the disagreement?

As noted in earlier sections, research suggests that multilingual learners may have distinct strengths and weaknesses in their different languages. Such differences may stem from a variety of factors, including the language(s) in which they are instructed at school and the ways they use their home language outside of school (e.g., the extent to which they read or are read to in their home language; the extent to which they speak or are spoken to in their home language; or the extent to which they consume media like television, movies, games, or podcasts in their home language) (NASEM, 2017). This reality makes it possible—even likely—that students may receive different scores or risk categorizations from the different screening assessments they take. Particularly in cases of students who are deemed to be at risk by the English-language screener, divergent results can be informative and may add important context for determinations of which students are genuinely at risk of reading problems or disabilities. For example, if students are deemed at risk in English but not in Spanish, this may suggest a need for more information or screening to differentiate between typical second language development (which may not be used to justify a disability determination) and a potential learning disability (U.S. Department of Justice & U.S. Department of Education, Office of Civil Rights, 2015). Such students may benefit from opportunities to more directly and explicitly connect and leverage their home language to support their English development, or from changes to how they are being taught English as a language.

Overall, the data for students who take Spanish and English-language assessments show that the screening assessments agree with one another more often than not. As seen in Table 23, when students are screened in both English and Spanish, their risk levels align across the two assessments a majority of the time (between 63% and 73%, depending on the time period and the combination of tests). Notably, this agreement held even when students took different screening assessments in Spanish versus English (e.g., if they took the Early Literacy screener in one language and the Reading screener in the other; we address these instances of mismatched concurrent screening in more depth later). The data in this section consist primarily of 18 schools; there are a few additional schools that have fewer than 10 students each.

Table 23. Risk Identification Agreement Rates for Students Who Took English and Spanish-Language Star Screening Assessments

Risk levels match	BOY	MOY	EOY	Any time period
Early Literacy English—Early Literacy Spanish (matched assessments)	71%	73%	72%	87%
Reading English—Reading Spanish (matched assessments)	71%	65%	67%	76%
Early Literacy English—Reading Spanish (mismatched; Spanish targets more advanced reading skills)	73%	64%	66%	70%
Reading English—Early Literacy Spanish (mismatched; English targets more advanced reading skills)	64%	63%	64%	67%

Note. Includes EL and non-EL students. *Source:* Early literacy screening assessment data.

When risk levels did not align between the two assessments, several patterns emerged. The largest group of students with divergent risk levels (n = 455) were those who took Star Early Literacy Assessment in both languages in grades K or 1 (Figure 24). This group included a roughly even split between English learner students who had attended schools in the United States for less than 1 year (n = 199) and those who had attended schools in the United States for more than 1 year (n = 230).[24] It also included a relatively higher rate of participation in DLBE programs (between 2% and 10% of students at any time period) compared with other groups we describe.

Figure 24 shows the proportion of students identified as at significant risk in each language (English and Spanish) for all students in this group, with results disaggregated for students who attended U.S. schools for less than 1 year (panel 2) or had already spent 1 or more years in

[24] Note that while some might consider some of these students "newcomers" based on their attendance in U.S. schools for less than 1 year, the majority of English learner students start school in kindergarten, like all students, and are thus not necessarily at any disadvantage in terms of prior schooling. Additionally, while most English learner students who have attended U.S. schools for more than 1 year are presumably in 1st grade and beyond, this group might include some kindergarteners if they have participated in a state-sponsored preschool program.

U.S. schools (panel 3).[25] The first panel (for all students in this group) shows that, on the Early Literacy assessment, the proportion of students deemed to be at risk by the English screener was higher than the proportion on the Spanish screener at all points during the year, though the discrepancy was highest at BOY (a difference of 40 percentage points). While the risk rates in the two languages did converge over the year, this convergence appeared to stem both from a decrease in the proportion of students deemed to be at risk by the English-language screener and from an increase in the proportion of students deemed to be at risk in Spanish.

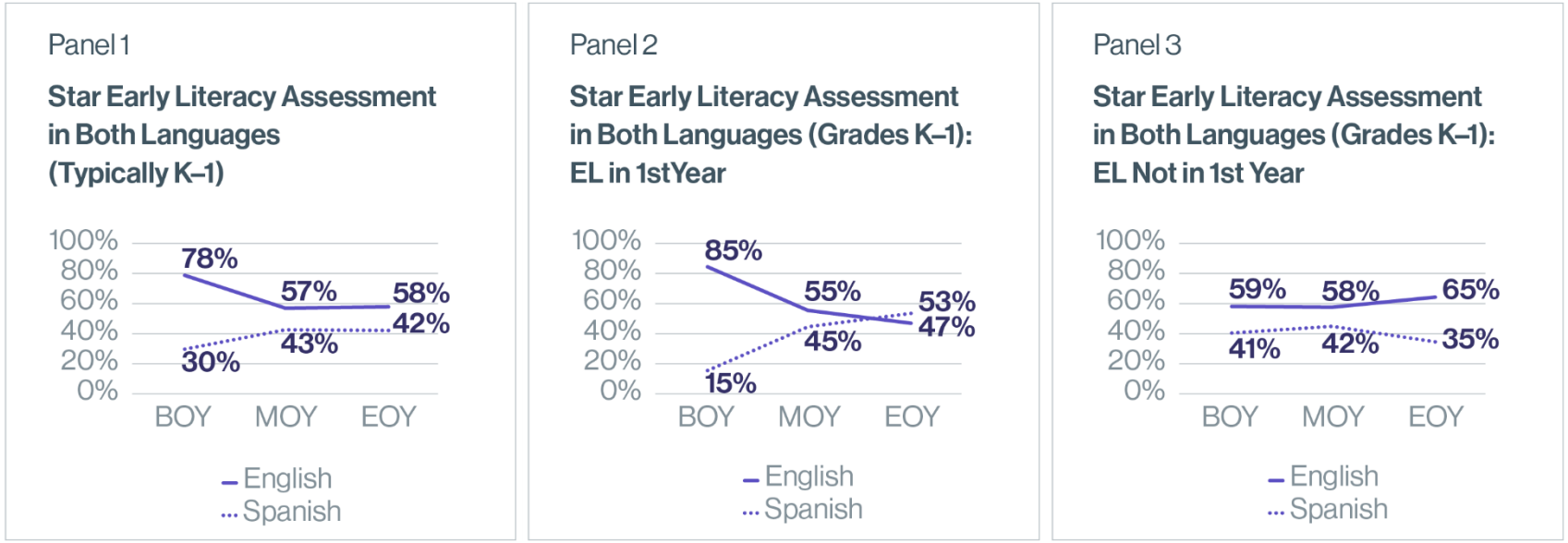
Panels 2 and 3 suggest that English learner students in their 1st year in U.S. schools largely drove this trend. These students (shown in panel 2) showed sharp decreases (30 percentage points) in English risk rates from BOY to MOY and a comparable increase (of 30 percentage points) in Spanish risk rates over the same period. This group stayed on the same trajectory from MOY to EOY, but with less dramatic changes (8 percentage points for each group). These trends suggest that as these students grew more accustomed to their instructional settings (which, as noted, are overwhelmingly English-only), their English screening performance improved and their Spanish screening performance decreased. While both screening assessments eventually settled at similar risk rates (about 50% in each language by EOY), the EOY agreement was 71 percent (similar to Table 23), indicating that nearly one third of students were deemed to be at risk in one language but not in the other.

English learner students attending schools in the United States for 1 or more years (panel 3), by contrast, showed essentially no change from BOY to MOY, and slight divergences from MOY to EOY (English risk rates increased by 7 percentage points, while Spanish risk rates decreased by the same amount). This group also included the most students in DLBE programs (between 2% and 10% depending on the time of year).

Taken together, the findings for this group of students suggest that risk rates and agreement across languages are affected by the language of instruction. English learner students in their 1st year in U.S. schools (panel 2) showed considerable decreases in their risk rates in English over the course of the year and showed corresponding increases in their Spanish risk rates because they spent more time receiving instruction in English. In panel 3, meanwhile, an increase in the proportion of students participating in DLBE over the year, where students are instructed in both languages, corresponded with a decrease in risk rates on the Spanish language screener.

[25] Note that per DESE guidance, all English learner students enrolled in their 1st year of kindergarten are considered 1st year in the United States.

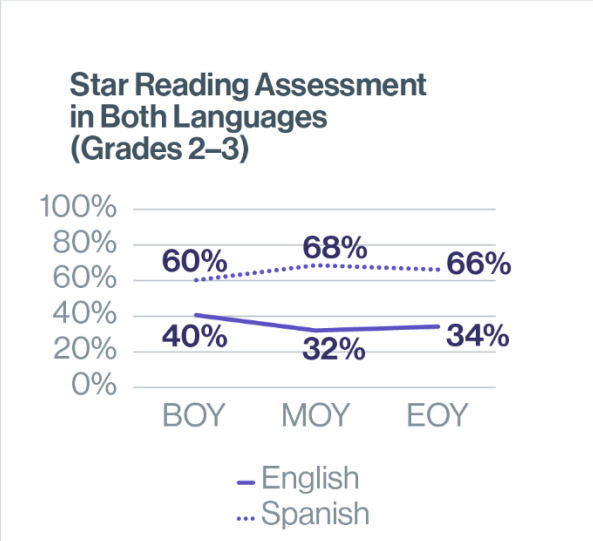
Figure 24. Percentages of Students Identified as At Risk on Star Early Literacy English and Spanish Converged Over the Year, Particularly for English Learners in Their 1st Year in U.S. Schools



Note. In panel 1, *N* = 455 at BOY, *N* = 440 at MOY, and *N* = 450 at EOY. In panel 2, *N* = 199 at BOY, *N* = 180 at MOY, and *N* = 181 at EOY. In panel 3, *N* = 230 at BOY, *N* = 243 at MOY, and *N* = 237 at EOY. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure 25 shows that risk rates for students who took Star Reading in both languages and had divergent risk categorizations were consistently higher on the Spanish screening assessment than the English-language screening assessment—the opposite pattern to that observed for students taking Star Early Literacy. Since Star Reading is generally recommended for students in grades 2 and 3, and since the vast majority of students screened in both languages (between 86% and 93%, depending on the time of year) are instructed solely in English, it would follow that screening in Spanish might be most challenging or irrelevant for this group because in some cases they will be completing their 4th year of English-only education. In alignment with this interpretation, more than 80 percent of all students in this group had spent at least 1 year in U.S. schools. For this group of students, it might be appropriate to seek other data sources (e.g., academic achievement scores, teacher or family input) to determine the likelihood of a disability or genuine reading risk rather than also administer Spanish-language screening assessments.

Figure 25. Percentages of Students Identified as At Risk Were Relatively Consistent From BOY to EOY in Spanish and English for Star Reading



Note. *N* = 111 at BOY, *N* = 158 at MOY, and *N* = 156 at EOY. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Among students with divergent risk categorizations in English and Spanish, the majority at each time point (between 83% and 87%) took the same screener in both languages. However, in the remaining 13 to 17 percent of cases, students took Star Early Literacy (generally recommended

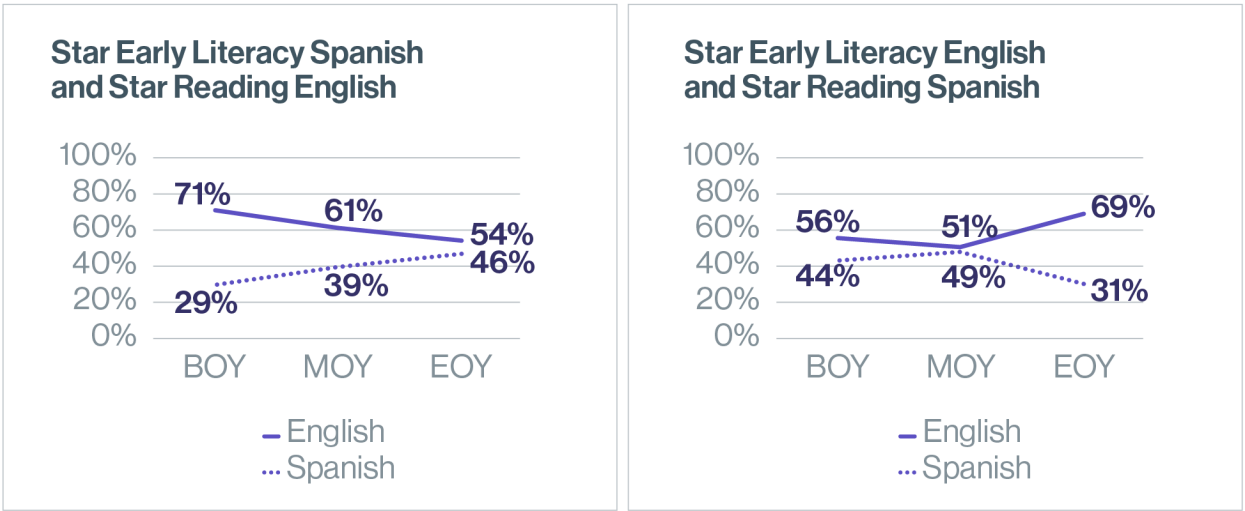
for grades K or 1) in one language and Star Reading (generally recommended for grades 2 or 3) in the other.[26]

Risk rates on mismatched screening assessments are shown in Figure 26. The first panel shows students who took the more advanced of the two assessments (Star Reading) in English. As one might expect, risk rates on the English assessment are higher at all time points for this group, though they also decrease over time and generally converge with the risk rates in Spanish. This assessment scenario might occur if a school seeks to use the Spanish Early Literacy screener as a check on basic skills or linguistic knowledge or a follow-up to an English result that is ambiguous or does not seem to align with teacher or family observations. The through-year trends suggest that students in this scenario generally improve on their English literacy skills over the course of the year, but they do not improve similarly in their Spanish literacy skills. As noted elsewhere, these divergent trends may stem from the fact that nearly all students in the sample are being instructed in English. To the extent that students are continuing to develop skills in Spanish, those skills may not be well-captured by a school-based assessment of literacy skills.

The second panel shows results for students with divergent outcomes who took the more advanced assessment (the Reading screener) in Spanish. This scenario was least likely to be observed in the sample (no more than 100 students at any time point during the year). The data show that, for students screened with this combination of assessments, risk rates were consistently higher on the English screener and actually increased over the course of the year, while Spanish risk rates decreased. Given these divergent trends and the decrease in risk classifications on the more challenging Spanish Reading assessment (relative to the English Early Literacy assessment), this group of students might be one that districts should look into more carefully or with other data sources before making decisions based on English results alone.

[26] Notably, very few students in this group (12 or fewer at any time point) were English learners in U.S. schools for less than 1 year. As such, it is unlikely that any trends in Figure 26 are driven by gaps in formal education or students being new to U.S. schools.

Figure 26. Patterns of Performance Vary Based on Screener Language and Difficulty



Note. In the first panel, *N* = 59 at BOY, *N* = 75 at MOY, and *N* = 95 at EOY. In the second panel, *N* = 32 at BOY, *N* = 35 at MOY, and *N* = 29 at EOY. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

This analysis offers a first look at screening patterns and performance across the year for Massachusetts students taking English- and Spanish-language screening assessments, though it is limited by the numbers of schools and students with data in both languages. It offers a window into practices and results for a small number of cases, which may prove useful in identifying areas for further investigation. One such area may be the relationship between language of instruction and screening results, including common development trajectories over time for students being taught solely in English versus in a DLBE program.

Given the small sample sizes observed here, ongoing analyses of screening results in all languages may help DESE to identify meaningful patterns and best practices for district guidance. Districts, in turn, may benefit from receiving guidance on when to screen in both languages, how to interpret divergent results, and when (under what circumstances) other sources of evidence may be useful to understand student development instead of or in addition to Spanish-language screening assessments. Such guidance could also benefit students with a home language other than Spanish for whom home language screening assessments are not yet available.

Discussion and Next Steps

This report provides additional analysis of early literacy screening data using data that represents about half of the K–3 student population in Massachusetts. Despite challenges in

interpreting results due to differences in definitions of risk used across screening assessments, 3 years of analysis have consistently shown several patterns.

First, students identified as at risk tend to stay at risk. This pattern is true both within and across years and is even more pronounced for historically marginalized students, including Black, Hispanic, and English Learner students; students from low income backgrounds; and students receiving special education services. At the other extreme, students who begin a school year at or above benchmark also tend to stay there within and across school years. This year's analysis allows us to follow students for 2 years rather than just year to year, and these same patterns persist.

At the same time, there is evidence that being identified as significantly below benchmark (or above benchmark) is not a permanent designation. In each of the past 3 years of analysis, data show that students who move above the significant risk threshold typically maintain that improved performance over time. For example, students who improve by midyear tend to stay out of the significant risk category through the end of the year. Additionally, students at earlier grade levels are consistently less likely to remain at significant risk than are students in higher grade levels. Moving students by midyear seems to matter for screening assessment performance at EOY or in the following school years and so does moving students in kindergarten and grade 1.

Analysis also shows the importance of schools to individual student performance. In this year's data and in prior reporting, results show that repeated risk identification is associated with the environments in which students learn. This year's analysis also examines district-level factors such as school segregation by race/ethnicity and socioeconomic status. Interestingly, the results show that we cannot statistically identify independent effects of these factors over and above related student- and school-level characteristics—their impact is likely already embedded at these levels. Similarly, analysis first carried out last year and repeated again this year shows that students from backgrounds that include membership in multiple historically underserved groups are more frequently identified as at risk. Further, analyses of English and Spanish screening practices highlight the complex relationships between learners, home languages, and languages of instruction and assessment. Data suggest that information in two languages may tell different stories about student strengths and progress, though the results of Spanish-language screening assessments tend to show higher risk and less growth over time—likely due to the fact that almost all students who are screened in multiple languages are instructed only in English.

And yet, results from last year and this year do point to some schools who appear to be outperforming expectations. Continuing to track these schools over time may offer an opportunity to identify successful practices.

New analysis this year attempts to check for effects of being identified as at significant risk on later academic performance—specifically on MCAS. We did not identify any such effects, which may be due to the fact that interventions meant to be provided to students at risk are not yet

having their full intended effect or they are keeping students from falling further behind their peers. It may also relate to another consistent finding about the relationship between screening assessment performance and MCAS. In each of the past 3 years, data show that students who are significantly below benchmark are very unlikely to be successful on MCAS, and even students performing at benchmark on many screening assessments will not meet expectations on MCAS. Without strong and sustained intervention *and* core instructional programs aligned with challenging standards, changes in grade 3 assessment scores are improbable.

Results of this and prior analysis are descriptive in nature—they do not "prove" anything in the same way that a rigorous study of a particular program or intervention does. They simply describe what we see in the data, accounting as best we can for factors such as differences between screening assessments and other data limitations.

Addressing data limitations and screening assessment comparability will strengthen DESE's ability to identify important and recurrent trends in performance and progress, but universal screening to identify students at risk is just the beginning.

References

Amplify. (n.d.). *mCLASS Lectura technical manual*. https://amplify.com/pdf/uploads/2024/04/mCLASS_Lectura_Technical_Manual_final_draft.pdf

Biancarosa, G., Kennedy, P. C, Park, S., Otterstedt, J., Gearin, B, Ives, C, & Yoon, H. J. (2018–2020). *8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Technical manual*. University of Oregon. https://dibels.uoregon.edu/sites/default/files/DIBELS8-TechnicalManual_04152020.pdf

Bibler, A. J. (2022). Language immersion and student achievement. *Education Economics*, *30*(5), 451–464. https://doi.org/10.1080/09645292.2021.2001788

Curriculum Associates. (2019). i-Ready® Assessments *technical manual*.

Curriculum Associates. (2025). *Universal early literacy screening using* i-Ready: *Guidance to identify students who are significantly below benchmark*.

EarlyBird Education. (2024). *EarlyBird Kindergarten Dyslexia and Early Literacy Assessment: Technical manual*. https://knowledgebase.earlybirdeducation.com/knowledge/earlybird-tech-manuals

Faulkner-Bond, M., Spycher, P., Olsen, L., & Gándara, P. (2020). The power and promise of California's multilingual learners. In California Department of Education, *Improving education for multilingual and English learner students: Research to practice* (pp. 27–62). https://www.cde.ca.gov/sp/el/er/documents/mleleducation.pdf

Gray, J. S., Warnock, A. N., Good, R. H., III, & Powell-Smith, K. A. (2021). *Acadience Reading K–6 national norms* (Technical Report No. 32). Acadience Learning. www.acadiencelearning.org

Greenberg Motamedi, J., Singh, M., & Thompson, K. D. (2016). *English learner student characteristics and time to reclassification: An example from Washington state* (REL 2016 128). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. http://ies.ed.gov/ncee/edlabs

Illuminate Education. (2021). *Psychometric evidence of FastBridge Universal Screening & Progress Monitoring System*. https://www.illuminateed.com/wp-content/uploads/2021/07/Psychometric-Evidence-of-FastBridge-Universal-Screening-Progress-Monitoring-System-2021.pdf

Kieffer, M. J., & Parker, C. E. (2016). *Patterns of English learner student reclassification in New York City public schools* (REL 2017–200). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. http://ies.ed.gov/ncee/edlabs

Lemke, M., Murphy, D., Soo Ping Chow, A., & Acuña, A. (2024). *Intersections: Student background and early literacy performance*. WestEd. https://www.wested.org/resource/intersections-student-background-and-early-literacy-performance/

Massachusetts Department of Elementary and Secondary Education. (2021). *Massachusetts dyslexia guidelines*. https://www.doe.mass.edu/specialeducation/families/links/dyslexia-guidelines.pdf

Massachusetts Department of Elementary and Secondary Education. (2023). *Early literacy screening guidance*. https://www.doe.mass.edu/instruction/screening-guide.pdf

McCall, L. (2005). The complexity of intersectionality. *Signs*, *30*(3), 1771–1800. http://www.jstor.org/stable/10.1086/426800

Morales, C. (2024). Dual language immersion programs and student achievement in early elementary grades. *Educational Evaluation and Policy Analysis*, *47*(2), 616–623. https://doi.org/10.3102/01623737241228829

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. The National Academies Press. https://doi.org/10.17226/24677

National Center for Education Statistics. (2024). English Learners in public schools. *Condition of Education.* U.S. Department of Education, Institute of Education Sciences. Retrieved May 30, 2024, from https://nces.ed.gov/programs/coe/indicator/cgf

Olsen, L. (with Martinez, M., Herrera, C. B., & Skibbins, H.). (2020). Multilingual programs and pedagogy: What teachers and administrators need to know and do. In California Department of Education, *Improving education for multilingual and English students: Research to practice* (pp. 115–188). https://www.cde.ca.gov/sp/el/er/documents/mleleducation.pdf

Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. *Sociology of Education*, *91*(1), 1–27. https://www.asanet.org/wp-content/uploads/attach/journals/jan18soefeature.pdf?ref=postcommon.com

Palacios, N., Bohlmann, N. L., Bell, B.A., Oh, M. H., & Yue, Y. (2024). Does early elementary dual language instruction deliver as promised? *AERA Open, 10*. https://doi.org/10.1177/23328584241264513

Peng, F., Xue, K., Hall, C., & Newburn, J. (2024). *English MAP reading fluency technical report: Based on assessments administered during the 2020–2022 school year*. NWEA. https://www.nwea.org/research/publication/english-map-reading-fluency-technical-report/

Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 34–57. https://doi.org/10.7758/RSF.2016.2.5.03

Reardon, S. F., Owens, A., & Kalogrides, D. (2025). *Documentation for segregation estimates, Segregation Explorer (Version 1.0)* [Data set]. http://edopportunity.org/segregation

Renaissance Learning. (n.d.). *Defining benchmarks in Star Computer Adaptive tests*. https://star-help.renaissance. com/hc/en-us/articles/24424240092827-Defining-Benchmarks-in-Star-Assessments Renaissance Learning

Renaissance Learning (2024). *FastBridge benchmarks and norms interpretation guide*. https://fastbridge.illuminateed.com/hc/en-us/articles/1260802348670-FastBridge-Benchmarks-and-Norms-Interpretation-Guide

Renaissance Learning (2025a). *Star Assessments™ for Early Literacy technical manual*. https://star-help. renaissance.com/hc/en-us/articles/12483321397019-Star-Assessments-for-Early-Literacy-Technical-Manual Renaissance Learning

Renaissance Learning (2025b). *Star Assessments™ for Reading technical manual*. https://star-help. renaissance.com/hc/en-us/articles/12542471051803-Star-Assessments-for-Reading-Technical-Manual

Slama, R., Molefe, A., Gerdeman, R. D, Herrera, A., Brodziak de los Reyes, I., August, D., & Cavazos, L. (2017). *Time to proficiency for Hispanic English learner students in Texas* (REL 2018–280). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. http://ies.ed.gov/ncee/edlabs

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal*, *54*(1_suppl), 282S-306S. https://doi.org/10.3102/0002831216634463

Thompson, K. D. (2017). English learners' time to reclassification: An analysis. *Educational Policy*, *31*(3), 330–363. https://doi.org/10.1177/0895904815598394

Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, *51*(5), 879–912. https://doi.org/10.3102/0002831214545110

U.S. Department of Justice, Civil Rights Division, & U.S. Department of Education, Office for Civil Rights. (2015, January 7). *Dear colleague letter: English learner students and limited English proficient parents*. https://www2.ed.gov/about/offices/list/ocr/letters/colleague-el-201501.pdf

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*, *37*(4), 612–637. https://doi.org/10.3102/0162373715573310

Appendix A. Business Rules and Data Processing Specifications

This report draws on data from multiple sources, including extant student-level data provided by Massachusetts Department of Elementary and Secondary Education (DESE) and publicly available school- and district-level data obtained from DESE's school and district profiles website and Stanford University's Segregation Tracking Project. The data includes

- early literacy universal screening assessment data for K–3 students in districts receiving certain state grants (e.g., the Early Grades Literacy grant, the Early Literacy Screening Assessment and Professional Development grant, the Growing Literacy Equity Across Massachusetts grant, Accelerating Literacy, and the High Quality Instructional Materials Implementation grant);
- the state's Student Information Management System (SIMS) data;
- Massachusetts Comprehensive Assessment System (MCAS) data;
- Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) data; and
- publicly available school- and district-level data pertaining to educator characteristics, student performance, student enrollment and demographic characteristics, racial/ethnic segregation, and finances/expenditures.

These data sources were each cleaned separately using R and Stata and were merged into one primary longitudinal analytical file that was used for the analysis. In the following sections, we describe the data cleaning and merging progress, data issues that arose, and the decisions that were made to resolve these issues.

Cleaning Early Literacy Universal Screening Assessment Data

Cleaning of the 2023/24 early literacy screening assessment data primarily consisted of dropping student identifiers and assessment-specific variables that would not be needed for the analyses (e.g., publisher-assigned ID, race/ethnicity, other measures); renaming variables to create a standardized format across assessments; creating variables to contain school and district codes for the merging process; creating variables containing the composite (and subtest) benchmark levels and reading risk flag status (as defined by the publisher); creating a time/test period variable, when needed, that describes when the screening assessment was administered (i.e., beginning of year [BOY], middle of year [MOY], and end of year [EOY]); and selecting one score per student per time period per grade level per assessment.

The analytic team used publisher-defined cut scores (obtained through the assessment technical manuals or communication with the publishers) to create composite and subtest benchmark variables and a reading risk flag variable. Although these variables were typically available in DESE- or publisher-provided files, for some assessments, schools and districts could establish their own local benchmarks and/or could calculate them manually. To ensure as much comparability as possible in the data, benchmark levels were recalculated according to the technical documentation provided by the assessment publishers. Where this was not possible due to missing information or other reasons, we used the school- or district-provided benchmark scores. The MAP Reading Fluency Universal Screener flag, the EarlyBird dyslexia risk flag, and the aimswebPlus composite benchmark level were not calculated by the analytic team because publisher-defined cut scores were not available. The MAP Reading Fluency flag is generated by NWEA using a multivariate predictive model, the EarlyBird dyslexia risk flag is generated by EarlyBird using subtests determined to be most predictive of dyslexia, and the aimswebPlus composite benchmark is generated by comparing the student's composite score with the seasonal cut scores set after the teacher or school selects the spring performance target. Additionally, for the 25th percentile or below analyses, we used publisher-provided documentation that identifies the score at the 25th percentile if percentile data were not included in the assessment file.

In some instances, differences between the benchmark level provided in the screening assessment file and the benchmark level generated based on the publisher-defined cut scores were likely due to the administration of an "off grade level" test; this was observed in 214 observations. In these instances, a note was included in the database and the observations were excluded from the analyses because their scores were not generated from the expected grade-level assessment.

The time period corresponding to each score was typically determined by using a variable within the file or was indicated in the file name. For cases with a missing time period, the assessment administration date provided within the file and the default testing periods provided by publishers were used to determine the expected time period. The Istation and Star assessments had additional ways to determine the appropriate time period. Because Istation Indicators of Progress (ISIP) Early Reading was typically delivered each month during the school year, the analytic team used the September scores as the BOY scores, the January scores as the MOY scores, and the May scores as the EOY scores. If September scores were missing, October scores were used as the BOY scores; if January scores were missing, February scores were used as the MOY scores; and if May scores were missing, June scores were used as the EOY scores. The Star assessments also include a variable that identifies the period that the student growth percentile for that observation covers. This variable was also used to determine the current time period of an observation.

Some students had multiple scores within the same grade level and time period and on the same assessment. To select one score per grade level per time period per assessment, the analytic team used the following rules:

- The observation(s) with a composite score was selected if additional observations did not have one.
- Observations with earlier administration dates were selected (because later scores in the same time window were potentially scores being used for progress monitoring rather than screening).
- If multiple observations had the same date or the administration dates were not available, the lowest score within the time period was selected because the higher score was assumed to be a progress monitoring measure after instruction.
- If the observations had no composite and the assessments were administered on the same date (or the date was unavailable), the observation with more subtest data was selected.

The Star, i-Ready, and NWEA assessments included additional metrics that were used to select the score for analysis.

- Star Curriculum-Based Methods files contained a test purpose variable that indicated whether the administration purpose was "screening," "progress monitoring," or "other." Some scores did not have an associated test purpose. Observations that were used for progress monitoring were dropped during the cleaning process. Observations with no stated purpose or with an "Other" purpose were kept if that was the only observation for the student for that subtest in the specific time period. The remaining duplicates were removed using the previously mentioned rules.
- i-Ready files contained a variable (i.e., Rush Flag) that indicates whether a student may have "rushed" through the diagnostic assessment. Students received a red Rush Flag if they answered questions in less than 11 seconds on average per item and a yellow Rush Flag if they spent between 12 and 15 seconds on average per item. In determining which observations to keep, if there were multiple observations per time period, scores without rush flags were kept regardless of administration date. The remaining duplicates were removed using the previously mentioned rules.
- The MAP Reading Fluency assessment included four types of tests: (a) Foundational Skills, (b) Foundational Skills: Beginner, (c) Adaptive Oral Reading, and (d) Adaptive Oral Reading: Passages Only. If students had observations from multiple assessment types, the observation for the assessment type most in line with what is typically administered to students in that grade level was used. The remaining duplicates were removed using the previously mentioned rules.

Finally, during the screening assessment cleaning process for the 2023/24 data, the composite score for 2,909 Acadience Reading observations and 3,867 DIBELS 8th Edition observations were generated by the analytic team using the composite score formulas in the technical manuals. These composite scores were only generated if the student either had all necessary

subtest data or met the discontinue rules (i.e., the student struggled significantly and did not take increasingly difficult subtests) or the gating rules (i.e., the student was high performing and skipped easier subtests [difficulty level determined by the publisher]). These cases typically occurred in district-provided files where the district did not calculate or provide the composite score. Further, 69 Acadience Reading scores, 455 DIBELS 8th Edition scores, and 182 mCLASS Lectura scores were replaced with composite scores generated by the analytic team using the composite score formulas in the technical manual. This recalculation was typically done for files in which the data appeared to be manually entered and was done primarily for two main reasons: First, some of the observations had incorrectly calculated composite scores based on the subtest data provided in the file (specifically, missing scores were treated as zeros). Second, some of the observations had composite scores even though the student did not complete the necessary grade-level subtests and did not meet the discontinue/gating rules.

Cleaning Student-Level State Education Data

In addition to the K–3 early literacy screening assessment data, other student-level data from the 2020/21, 2021/22, 2022/23, and 2023/24 school years were used for the analysis, including October and June SIMS data, MCAS data, and ACCESS for ELLs data.

The state education data required minimal cleaning. The cleaning process was conducted in Stata and generally consisted of renaming variables to meet the standardized format used for the early literacy screening files and dropping variables that were unnecessary for the analysis or not applicable for K–3 students (e.g., the High School Completer Plan). Additionally, some variables were used/manipulated to create indicator variables for the analysis (e.g., DESE-provided race/ethnicity variables were used to create a separate variable for each racial/ethnic group).

Merging Student-Level Early Literacy Screening Data and State Education Data

Following the cleaning of the student-level screening assessment data and state education data, a student-level file was created by merging the two data sources. The following steps were taken:

1. The combined screening assessment file was merged with the June SIMS data where possible. The analytic team first attempted to match student screening assessment scores with their June SIMS data using the state assigned student identifier (SASID), grade level, school code, and district code. Because students may appear multiple times within the SIMS data if they transferred to a different school and district within the school year, we attempted to connect a student's screening assessment data with the SIMS data that corresponded to the same school and district.

2. If a match did not occur between the assessment data and June SIMS data using these identifiers, we then attempted to match the assessment data, using these same identifiers, with the October SIMS data.
3. If a match did not occur between the assessment data and October SIMS data using these identifiers, we then attempted to match the assessment data with their appropriate June SIMS observation using the student's SASID, grade level, and district code (i.e., without the school code as a matching variable). The process was repeated with the October SIMS data for remaining observations, followed by a merge using SASIDs and grade level alone, then a merge using SASIDs and school codes, then a merge using SASIDs and district codes, and then a merge based solely on SASIDs.
4. Some files did not contain the SASID for a student; rather, they only contained the student's locally assigned identifier (LASID), which is unique at the district level. Because the identifier is not unique at the state level, all attempted merges used the district code as a matching variable.[27]

Some screening assessment observations merged with multiple SIMS observations when merging on fewer variables than the student identifier, grade level, school code, and district code. In these instances, we used the *Researcher's Guide to Massachusetts State Education Data*[28] to determine the appropriate assessment-SIMS match to keep. Typically, we used the enrollment status variable, which describes the enrollment status of a student within the school (e.g., enrolled students, students who dropped out, students who transferred into the state), and the days of membership variable to select the appropriate match. In these instances, matches that were higher in the enrollment hierarchy were used; and if two observations had the same enrollment status, the observation with the larger number of days of membership was selected.

Because the analysis includes student characteristics, it was important for each observation to have the same key demographics, regardless of the SIMS observation that they matched with. For the following demographic characteristics, we used the June SIMS data to ensure consistency across each observation: gender, race/ethnicity, low income status, English learner status, English learner program status, indicator for English learner students attending schools in the United States for less than 1 year, Section 504 status, special education status, early childhood experience, and native language.

[27] Observations with no SASIDs were merged using LASIDs, grade levels, school codes, and district codes first, followed by a merge using LASIDs, grade levels, and district codes for remaining observations, followed by a merge using LASIDs, school codes, and district codes, and finally followed by a merge using LASIDs and district codes.

[28] This resource can be downloaded from *Researchers' Guide to Massachusetts State Education Data*.

Out of the 694,858 screening assessment observations used for the analysis for the 2020/21, 2021/22, 2022/23, and 2023/24 school years, 690,355 (99.4%) matched with a corresponding observation in the SIMS data. Of the observations that matched, 688,262 (99.7%) matched with an observation in the June SIMS data and 2,093 (0.3%) matched with an observation in the October SIMS data. Of the screening assessment observations, 32,247 did not match exactly with the SIMS data (i.e., did not match using the grade level, school code, and district code); 17,457 had a different grade level than their corresponding SIMS observation, 239 had different district codes, and 14,832 had different school codes. Almost all grade level differences occurred in the 2023–24 school year (99.9%); these differences occurred primarily because some screening assessment data were exported by the publisher during the 2024/25 school year and thus had the incorrect grade level. Additionally, 93 percent of the school differences (13,782 observations) occurred because the screening assessment file did not identify which school the student attended. In conducting the analyses, the school and district codes from the screening assessment data set were used and the grade levels from the SIMS data set were used. Any student with a grade level outside of K–3, according to the SIMS database, was not included in the analysis.

After the screening assessment data were merged with the SIMS collection data, the combined file was merged with the MCAS and ACCESS data sets using SASID. The MCAS and ACCESS files did not have any duplicate observations; therefore, only the student's state identifier was used to merge the data sets with the combined screening assessment and SIMS file.

Cleaning and Merging Publicly Available School- and District-Level Data

Publicly available school- and district-level data for 2020/21, 2021/22, 2022/23, and 2023/24 were retrieved from DESE's school and district profiles website to provide contextual data about the sample of the students used in analysis. School-level racial/ethnic segregation data were retrieved from the Educational Opportunity Project at Stanford University. Overall, the data pertain to the following four main categories of information: (a) educator characteristics, (b) student performance, (c) student enrollment and demographics, and (d) financial information. These data were merged with the student-level longitudinal file using the school and district codes from the screening assessment data set. For the observations missing the school code from the screening assessment data set, the SIMS school code was used to merge with the publicly available school-level data.

This finalized file with student-level screening assessment data, state education data, and publicly available school- and district-level data was provided to DESE. Data are organized because a single longitudinal data set with one observation per student, per grade, per time period, per screening assessment, per year. Some students have multiple screening assessment scores per time period as they took multiple early literacy screening assessments during the school year.

Appendix B. Comparability Analysis Details

Screening Assessment Benchmark Details

Table B1. Descriptions of Significantly Below Benchmark Score Levels and Availability of National Percentile Scores

Early literacy screening assessment	Description of levels corresponding to significantly below benchmark	National percentiles available
Acadience Reading	At EOY, a score that indicates performance significantly below benchmark ("well below benchmark" in Acadience terms) means a student is estimated to have a 10 percent o 20 percent chance of being above the 40th percentile on the Group Reading Assessment and Diagnostic Evaluation (GRADE) assessment. MOY risk cut points were established such that students were predicted to have a 10 percent to 20 percent chance of achieving the EOY Acadience Reading composite score needed to be above the 40th percentile on the GRADE assessment. The BOY score was set such that students had the same odds of achieving the MOY Acadience Reading composite score. These cut points were established based on a study in 13 schools in five school districts in 2009/10 with 3,816 participants total (1,306 who took the GRADE).	Yes

Early literacy screening assessment	Description of levels corresponding to significantly below benchmark	National percentiles available
DIBELS 8th Edition	At BOY, MOY, and EOY, a score that indicates performance significantly below benchmark ("well below benchmark" or "at risk" in DIBELS and mCLASS terms) identifies most students who would be expected to score at or below the 20th percentile on an EOY assessment. For kindergarten, the EOY assessment used in analysis was DIBELS Next, and it was the Iowa Assessments (total reading score) for grades 1–3. The Iowa Assessments test is described as "a published, group-administered, multiple-choice, norm-referenced measure of reading achievement," and technical documentation notes, "Whereas DIBELS Next includes letter naming and phonemic awareness component skills in the composite score, the Iowa total reading score does not assess these same component skills, making it a more distal criterion measure." Based on studies carried out between 2017 and 2019 with about 7,000 K–3 students, the well below benchmark cut score will accurately identify 80 percent of students who would perform at the 20th percentile or below at EOY. In other words, being well below benchmark identifies students whose reading skills are still likely to be less well developed than those of most of their peers by EOY if they do not receive intensive intervention.	Yes
MAP Reading Fluency	MAP Reading Fluency does not provide a composite score based on its subtests. However, it provides a binary "Universal Screener outcome flag" that suggests possible risk of reading difficulty. MAP Reading Fluency benchmark assessments may include different test content for different students depending on their grade and skill level, and the Universal Screener outcome flag is set differently depending on which subtests students take. For students who take a foundational skills test form (which can be specifically assigned or will be automatically assigned based on performance), the Universal Screener outcome is set based on results of predictive modeling that identifies students as at risk if their performance would likely be below the 10th percentile on MAP Growth in reading at EOY in that grade level.	No

Early literacy screening assessment	Description of levels corresponding to significantly below benchmark	National percentiles available
mCLASS	See the description for DIBELS 8th Edition. (mCLASS assessments use DIBELS 8th Edition tasks, and reported performance levels are the same.)	Yes
mCLASS Lectura	The cut scores were calculated using receiver operating characteristic (ROC) curve analyses, which describe the extent to which mCLASS Lectura scores accurately predicted performance above or below the 20th percentile on external Spanish language criterion measures (i.e., Woodcock-Muñoz Análisis de Palabras for kindergarten and Star Early Literacy Spanish for grades 1–3). This is based on data collected in 2021/2022 from approximately 1,300 students over about 18 schools in 10 districts; it varies by grade level.	No
EarlyBird (grade K)	EarlyBird provides different metrics at each time period to identify students at risk of reading difficulties. At BOY, MOY, and EOY, EarlyBird provides a dyslexia risk flag that is used in analysis to identify students at significant risk. According to the publisher, dyslexia risk is defined as "performing at or below the 16th percentile on the Kaufman Test of Educational Achievement, 3rd Edition (KTEA-3) Phonological Processing subtest. Any child flagged for dyslexia risk is at high risk for low phonological processing skills and therefore subsequent low reading proficiency and needs intensive instruction targeted to the student's skill weaknesses." The dyslexia risk flag was set based on analysis of data collected in 2019/20 from 219 students in eight states. The original sample included 19 schools; it is unclear how many schools were represented in the final sample.	No

Early literacy screening assessment	Description of levels corresponding to significantly below benchmark	National percentiles available
FastBridge aReading (grades 2 and 3)	At BOY, MOY, and EOY, the FastBridge aReading score that indicates performance significantly below benchmark ("high risk" in FastBridge terms) means that students are performing below the 15th national percentile. The 15th percentile was selected through studies that used predictive analysis (ROC analysis) to assess the classification accuracy of those cut points relative to performing below the 20th percentile on the Gates-MacGinitie Reading Tests, 4th Edition (GMRT-4th; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) or the measures of academic progress (MAP) in 2010/11 in two schools (777 students).	Yes
FastBridge earlyReading (grades K and 1)	At BOY, MOY, and EOY, the FastBridge earlyReading score that indicates performance significantly below benchmark ("high risk" in FastBridge terms) means that students are performing below the 15th national percentile. The 15th percentile was selected through studies that used predictive analysis (ROC analysis) to assess the classification accuracy of those cut points relative to performing below the 15th percentile on the GRADE assessment aimed to maximize specificity and sensitivity. Data were collected from two school districts (numbers of students and schools not provided).	Yes
i-Ready Diagnostic	At BOY, MOY, and EOY, scores that indicate performance significantly below benchmark ("at risk" on i-Ready Diagnostic) describe the grade level associated with a student's performance in the context of college and career readiness standards. That is, "significantly below benchmark" generally means that students are performing one or more grade levels below their assigned grade. For example, at BOY, grade 3 students classified as "at risk" based on i-Ready Diagnostic are performing at or below grade 1 standards; at MOY, grade 3 students classified as at risk are performing at grade 2 standards or below; and at EOY, grade 3 students classified as at risk are performing at or below a level that indicates partially meeting grade 3 standards.	Yes

Early literacy screening assessment	Description of levels corresponding to significantly below benchmark	National percentiles available
Star Early Literacy (English and Spanish; grades K and 1)	At BOY, MOY, and EOY, a score that indicates performance significantly below benchmark ("intervention" or "urgent intervention" in Star terms) means that students are performing below the 25th national percentile based on a 2014/15 study including more than 500,000 unique students who took Star assessments. At least 75 percent of students would be expected to perform better than students scoring at the intervention or urgent intervention level.	Yes
Star Reading (English and Spanish; grades 2 and 3)	See the description of Star Early Literacy benchmarks in the previous row.	Yes

Source: WestEd's compilation based on technical documentation and/or communication with publishers. (See the references for a list of technical reports and related documentation.)

Screening Assessment Agreement Rates

Table B2. Percentage of Times Screening Assessment Results Match for Same Students At BOY, MOY, EOY

Early literacy screening assessments	BOY (N, % agree)	MOY (N, % agree)	EOY (N, % agree)
DIBELS 8th Edition and i-Ready	997 74%	1,110 73%	1,100 62%
Star Reading and mCLASS	995 84%	1,103 85%	1,115 85%
EarlyBird and mCLASS	355 47%	624 63%	492 73%
i-Ready and mCLASS	185 84%	176 81%	189 76%
EarlyBird and i-Ready	182 81%	120 77%	310 72%

Early literacy screening assessments	BOY (N, % agree)	MOY (N, % agree)	EOY (N, % agree)
Star Early Literacy and mCLASS	178 68%	172 76%	165 81%

Note. Includes students with any number of scores. *Source*: Early literacy screening assessment data.

Comparability Testing

As part of the analysis for this year's report, WestEd tested several different approaches to developing comparable metrics for reporting reading risk.

We tested three approaches:

1. Equipercentile linking of each screening assessment to MCAS
2. Equipercentile linking of each screening assessment to DIBELS/mCLASS
3. Equipercentile linking of each screening assessment to DIBELS/mCLASS, then pooling data and calculating the 25th percentile scores using only Massachusetts student data

We briefly describe each approach here. The descriptions are followed by information to help evaluate the quality of the results. We also provide tables that show the numbers of students who would be identified as at significant risk at each grade level and time period using these three approaches, plus the existing publisher-provided benchmarks and the 25th national percentile for comparison purposes. A detailed description of the equipercentile linking procedure and additional tables showing grade 3 screening benchmarks linked to the MCAS scale follow.

1. Equipercentile Linking of Each Screening Assessment to MCAS

In this approach, we take students' screening assessment scores where we have complete data for each time point (beginning, middle, and end of year) and grade level and link to 2024 MCAS tests. Using a same-persons (common students) design, we link to the 3rd grade 2024 MCAS ELA assessment:

- Grade 3 screening assessment data are from the 2023/24 school year.
- Grade 2 screening assessment data are from the 2022/23 school year.
- Grade 1 screening assessment data are from the 2021/22 school year.

We only used data from students with both scores. Due to data limitations, sample sizes decreased with greater time intervals between screening assessment administration and spring 2024 MCAS. Note also that we linked Star Reading at grades 2 and 3 and Star Early Literacy at grades K and 1 because these are the generally recommended grade levels for students to take

these assessments. Though there are students who take these assessments at other grade levels, those students represent different populations than the typical students at those grades.

After linking, for purposes of illustrating an approach for the state to select a benchmark, we chose an MCAS score of 470 (the cut between Does Not Meet and Partially Meets) as a sample cut score indicating risk of reading difficulty.

2. Equipercentile Linking of Each Screening Assessment to DIBELS/mCLASS

In this approach, two screening assessment scores were linked within a grade level and time of year. For this design, common students were not available across different screening assessments. Also, the characteristics of the population of students taking different screening assessments may vary. To account for these differences, we used coarsened exact matching, a statistical technique to group similar students based on demographic characteristics and their school MCAS performance. We used matched sample data from the 2023/24 school year where we had at least 250 observations at the beginning, middle, and end of year for the linking study. The following screening assessments met the sample size requirement to link to DIBELS/mCLASS:

- Acadience Reading
- FastBridge aReading and earlyReading
- Star Early Literacy
- Star Reading
- i-Ready

After linking, we used linked scores to illustrate two different approaches for the state to select a benchmark. First, we simply applied the DIBELS/mCLASS Below and Well Below Benchmarks to all other screening assessments. This method shows how the state could capitalize on an existing assessment's risk benchmarks while still using other screening assessments. The second approach uses linked scores to compute a Massachusetts-specific percentile score (in this case, the 25th percentile) to show how the state could create a normative benchmark using Massachusetts student screening data from multiple assessments.

3. Equipercentile Linking of Each Screening Assessment to DIBELS/mCLASS, Then Pooling Data and Calculating the 25th Percentile Scores Using Only Massachusetts Student Data

For this analysis, we followed a similar procedure to that described earlier, except we then combined data by grade and time period and found the 25th percentile value using the linked data and used this as an example benchmark.

Evaluation Metrics

Where possible we evaluated the following metrics to understand the quality of the linking:

- linked scores' correlation with the observed scores
- standard errors of the linking estimates
- similarity of the students across linked and observed scores

In the first section (Tables B3–B29), we provide MCAS linking results. Tables B30–B53 show results for DIBELS/mCLASS linking.

Not surprisingly, overall the MCAS linking is strongest for assessments with larger samples and at later grade levels. Correlations between observed and linked MCAS scores range from about 0.55 to 0.84 depending on the screening assessment and time of year (Tables B3–B5). Tables B6–B8 show descriptive statistics comparing linked scores and observed scores. These show that on average the linked and observed scores are similar, but there is significant variation when comparing individual scores. Tables B9–B11 show similar statistics by MCAS performance level, showing more accurate links in the middle of the scale (for students in the Partially Meeting Expectations and Meeting Expectations levels). Tables B12–B14 provide estimates of the linking error at the specific MCAS cut score of 470, which separates the Not Meeting Expectations from Partially Meeting Expectations. Average error tends to be 1–3 MCAS scale score points, but it is larger for Acadience. Tables B15–B29 show classification accuracy using our illustrative cut point of 470. Applying this as a sample risk indicator to screening assessment data and then comparing it to observed MCAS scores, we see that between 71 and 94 percent of students would be accurately classified into Does Not Meet or above Does Not Meet categories. There is some variation across assessments and grades, but the accuracy would be considered in the "good" range. Tables B18–B29 show the classification data by student subgroup to assess how well they apply for students with different background characteristics. We included results for which we had at least 100 students in each group.

This disaggregated analysis of the proportion of students with the same Not Meeting Expectations (MCAS) scores shows that this linking approach (MCAS) generally maintained good agreement rates across student populations. However, agreement rates were higher for students without IEPs and for students not from low income backgrounds compared with those with IEPs or who were classified as being from low income backgrounds. These variations may reflect differences in score distributions rather than linking quality. Gender differences were minimal across all screening assessments, with agreement rates varying by less than 2 percentage points between female and male students. While some differences were observed across racial/ethnic groups, most screening assessments maintained acceptable agreement levels (0.70 or higher) across demographic categories. These findings suggest that the linking methods provide reasonably consistent results across different student groups, though educators may benefit from considering these patterns when interpreting linked scores, particularly for students with IEPs.

Tables B30–B53 provide similar information for the DIBELS/mCLASS linking. Table B31 shows that the average error was about 1–3 DIBELS/mCLASS scale score points at the Below Benchmark cut-score location. Tables B34–B53 show how often classifications using DIBELS/mCLASS benchmarks agreed with classifications using each assessment's original benchmarks (i.e., whether or not students were classified as at risk using each). We focused our analysis on those identified in the lowest benchmarks (on DIBELS and mCLASS, the performance level Well Below Benchmark). Tables B34–B37 show overall classification accuracy, and Tables B38–B53 break these down by student subgroups. Like the MCAS analysis, we reported findings when sample sizes were at least 100. In most instances, classification accuracy between the observed and linked benchmarks was greater than 70 percent and in many instances was above 80 percent. The differences by demographic characteristics were smaller compared with the MCAS analysis. For example, the rates for those with an IEP and those without an IEP were similar (see Tables B42–45). This disaggregated analysis provides evidence that the linking approach was relatively accurate overall and for each subgroup.

MCAS Linking Results

Table B3. Correlations of Linked MCAS Scores to Observed MCAS Scores by Grade, Time of Year, and Screening Assessment, Including Sample Size, Grade 1

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	NSD	NSD	NSD
DIBELS 8th Edition	0.553 (1,350)	0.604 (1,423)	0.626 (1,467)
FastBridge aReading	N/A	N/A	N/A
Star Early Literacy	0.611 (1,453)	0.657 (1,838)	0.656 (1,861)
Star Reading	N/A	N/A	N/A
i-Ready	0.641 (666)	0.679 (669)	0.698 (409)
mCLASS	0.609 (546)	0.641 (568)	0.632 (575)

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B4. Correlations of Linked MCAS Scores to Observed MCAS Scores by Grade, Time of Year, and Screening Assessment, Including Sample Size, Grade 2

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	0.640 (443)	0.687 (401)	0.705 (403)
DIBELS 8th Edition	0.670 (2,352)	0.674 (2,550)	0.670 (2,605)
FastBridge aReading	NSD	NSD	NSD
Star Early Literacy	N/A	N/A	N/A
Star Reading	0.683 (3,513)	0.750 (3,760)	0.761 (3,931)
i-Ready	0.763 (2,199)	0.789 (2,223)	0.800 (2,224)
mCLASS	0.690 (2,169)	0.685 (2,182)	0.695 (2,570)

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B5. Correlations of Linked MCAS Scores to Observed MCAS Scores by Grade, Time of Year, and Screening Assessment, Including Sample Size, Grade 3

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	0.697 (465)	0.700 (467)	0.714 (466)
DIBELS 8th Edition	0.657 (5,724)	0.662 (5,982)	0.645 (5,612)
FastBridge aReading	0.778 (573)	0.802 (557)	0.824 (583)
Star Early Literacy	N/A	N/A	N/A
Star Reading	0.798 (5,158)	0.824 (5,143)	0.832 (5,207)
i-Ready	0.809 (3,847)	0.833 (3,873)	0.841 (3,858)
mCLASS	0.703 (11,208)	0.694 (11,190)	0.682 (10,981)

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B6. Descriptive Statistics of Difference in MCAS Scale Score Estimates by Screening Assessment and Grade Level (in MCAS Scale Score Points), Grade 1

Early literacy screening assessment	Avg	SD	Min	Max
DIBELS 8th Edition	NSD	19.77	–105	81
Star Early Literacy	–0.55	18.49	–69	93
Star Reading	N/A	N/A	N/A	N/A
i-Ready	–0.21	16.31	–65	55
mCLASS	–0.28	18.95	–68	66
Acadience Reading	NSD	NSD	NSD	NSD
FastBridge aReading	N/A	N/A	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B7. Descriptive Statistics of Difference in MCAS Scale Score Estimates by Screening Assessment and Grade Level (in MCAS Scale Score Points), Grade 2

Early literacy screening assessment	Avg	SD	Min	Max
DIBELS 8th Edition	–0.18	18.08	–99	71
Star Early Literacy	N/A	N/A	N/A	N/A
Star Reading	–0.79	15.86	–83	74
i-Ready	–0.28	15.16	–76	58
mCLASS	–0.53	17.78	–72	69
Acadience Reading	–0.18	16.51	–57	47

Early literacy screening assessment	Avg	SD	Min	Max
FastBridge aReading	NSD	NSD	NSD	NSD

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B8. Descriptive Statistics of Difference in MCAS Scale Score Estimates by Screening Assessment and Grade Level (in MCAS Scale Score Points), Grade 3

Early literacy screening assessment	Avg	SD	Min	Max
DIBELS 8th Edition	–0.22	18.54	–93	87
Star Early Literacy	N/A	N/A	N/A	N/A
Star Reading	–0.35	13.45	–69	83
i-Ready	–0.22	12.97	–66	63
mCLASS	–0.19	18.05	–79	96
Acadience Reading	–0.47	15.78	–59	49
FastBridge aReading	–0.08	13.54	–53	53

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B9. Descriptive Statistics of the Difference in Linked and Observed MCAS Scale Scores by Screening Assessment, MCAS Benchmark, Grade Level, and Performance Levels, Grade 1

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Acadience Reading	Not meeting expectations	NSD	NSD	NSD	NSD

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Acadience Reading	Partially meeting expectations	NSD	NSD	NSD	NSD
Acadience Reading	Meeting expectations	NSD	NSD	NSD	NSD
Acadience Reading	Exceeding expectations	NSD	NSD	NSD	NSD
DIBELS 8th Edition	Not meeting expectations	14.71	19.18	–27	81
DIBELS 8th Edition	Partially meeting expectations	2.24	17.37	–52	68
DIBELS 8th Edition	Meeting expectations	–5.49	17.88	–75	54
DIBELS 8th Edition	Exceeding expectations	–19.03	20.14	–105	24
FastBridge aReading	Not meeting expectations	N/A	N/A	N/A	N/A
FastBridge aReading	Partially meeting expectations	N/A	N/A	N/A	N/A
FastBridge aReading	Meeting expectations	N/A	N/A	N/A	N/A
FastBridge aReading	Exceeding expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Not meeting expectations	9.30	16.44	–27	93
Star Early Literacy	Partially meeting expectations	–1.42	17.08	–54	64
Star Early Literacy	Meeting expectations	–7.65	18.03	–69	47
Star Early Literacy	Exceeding expectations	–19.84	17.41	–62	16

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Star Reading	Not meeting expectations	N/A	N/A	N/A	N/A
Star Reading	Partially meeting expectations	N/A	N/A	N/A	N/A
Star Reading	Meeting expectations	N/A	N/A	N/A	N/A
Star Reading	Exceeding expectations	N/A	N/A	N/A	N/A
i-Ready	Not meeting expectations	13.21	15.87	–23	55
i-Ready	Partially meeting expectations	0.74	15.40	–45	46
i-Ready	Meeting expectations	–4.34	14.59	–59	31
i-Ready	Exceeding expectations	–12.94	17.23	–65	23
mCLASS	Not meeting expectations	14.81	17.28	–26	66
mCLASS	Partially meeting expectations	2.45	16.51	–54	54
mCLASS	Meeting expectations	–4.08	18.22	–68	40
mCLASS	Exceeding expectations	–18.52	15.98	–59	24

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B10. Descriptive Statistics of the Difference in Linked and Observed MCAS Scale Scores by Screening Assessment, MCAS Benchmark, Grade Level, and Performance Levels, Grade 2

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Acadience Reading	Not meeting expectations	14.33	17.01	–25	47
Acadience Reading	Partially meeting expectations	3.80	14.49	–48	40
Acadience Reading	Meeting expectations	–0.62	15.28	–37	35
Acadience Reading	Exceeding expectations	–15.12	16.37	–57	19
DIBELS 8th Edition	Not meeting expectations	10.69	15.91	–25	64
DIBELS 8th Edition	Partially meeting expectations	3.41	16.83	–51	71
DIBELS 8th Edition	Meeting expectations	–4.24	16.90	–77	59
DIBELS 8th Edition	Exceeding expectations	–17.38	16.06	–99	24
FastBridge aReading	Not meeting expectations	NSD	NSD	NSD	NSD
FastBridge aReading	Partially meeting expectations	NSD	NSD	NSD	NSD
FastBridge aReading	Meeting expectations	NSD	NSD	NSD	NSD
FastBridge aReading	Exceeding expectations	NSD	NSD	NSD	NSD
Star Early Literacy	Not meeting expectations	N/A	N/A	N/A	N/A

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Star Early Literacy	Partially meeting expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Meeting expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Exceeding expectations	N/A	N/A	N/A	N/A
Star Reading	Not meeting expectations	7.55	15.64	–27	74
Star Reading	Partially meeting expectations	0.67	15.07	–57	61
Star Reading	Meeting expectations	–4.03	15.20	–81	47
Star Reading	Exceeding expectations	–13.05	14.26	–83	24
i-Ready	Not meeting expectations	6.75	15.10	–26	58
i-Ready	Partially meeting expectations	0.57	14.53	–52	52
i-Ready	Meeting expectations	–3.95	13.78	–63	39
i-Ready	Exceeding expectations	–11.50	16.69	–76	26
mCLASS	Not meeting expectations	9.67	15.94	–27	69
mCLASS	Partially meeting expectations	2.03	16.21	–57	69
mCLASS	Meeting expectations	–4.23	16.86	–70	49
mCLASS	Exceeding expectations	–19.00	15.99	–72	21

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B11. Descriptive Statistics of the Difference in Linked and Observed MCAS Scale Scores by Screening Assessment, MCAS Benchmark, Grade Level, and Performance Levels, Grade 3

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
Acadience Reading	Not meeting expectations	13.79	15.65	–22	43
Acadience Reading	Partially meeting expectations	2.96	13.91	–44	41
Acadience Reading	Meeting expectations	–0.54	14.54	–41	49
Acadience Reading	Exceeding expectations	–14.60	16.14	–59	23
DIBELS 8th Edition	Not meeting expectations	10.74	17.73	–26	87
DIBELS 8th Edition	Partially meeting expectations	3.13	17.15	–54	66
DIBELS 8th Edition	Meeting expectations	–5.38	16.79	–78	57
DIBELS 8th Edition	Exceeding expectations	–18.97	16.11	–93	25
FastBridge aReading	Not meeting expectations	9.12	14.13	–28	53
FastBridge aReading	Partially meeting expectations	0.58	13.89	–45	33
FastBridge aReading	Meeting expectations	–1.97	11.24	–48	28

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
FastBridge aReading	Exceeding expectations	–9.31	14.08	–53	22
Star Early Literacy	Not meeting expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Partially meeting expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Meeting expectations	N/A	N/A	N/A	N/A
Star Early Literacy	Exceeding expectations	N/A	N/A	N/A	N/A
Star Reading	Not meeting expectations	4.90	13.48	–27	83
Star Reading	Partially meeting expectations	0.39	12.69	–57	54
Star Reading	Meeting expectations	–2.89	12.88	–69	46
Star Reading	Exceeding expectations	–12.13	13.20	–53	28
i-Ready	Not meeting expectations	5.98	13.69	–27	63
i-Ready	Partially meeting expectations	–0.17	12.12	–49	50
i-Ready	Meeting expectations	–2.67	11.95	–66	48
i-Ready	Exceeding expectations	–11.44	13.32	–51	28
mCLASS	Not meeting expectations	8.71	16.41	–28	96
mCLASS	Partially meeting expectations	2.04	17.32	–55	83

Early literacy screening assessment	MCAS performance level	Avg	SD	Min	Max
mCLASS	Meeting expectations	–5.20	16.41	–79	54
mCLASS	Exceeding expectations	–19.25	14.92	–69	29

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B12. Average, Minimum, and Maximum Linking Error, in Terms of MCAS Scale Scores, At the Partially Meeting Expectations MCAS Performance Level Cut Score by Screening Assessment and Grade, Aggregated Across Time of Year, Grade 1

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	NSD	NSD	NSD
DIBELS 8th Edition	1.173	0.703	1.694
FastBridge aReading	N/A	N/A	N/A
i-Ready	3.464	3.053	4.100
mCLASS	2.281	0.959	3.938
Star Early Literacy	4.163	3.445	5.045
Star Reading	N/A	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data

Table B13. Average, Minimum, and Maximum Linking Error, in Terms of MCAS Scale Scores, At the Partially Meeting Expectations MCAS Performance Level Cut Score by Screening Assessment and Grade, Aggregated Across Time of Year, Grade 2

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	12.262	7.881	20.171
DIBELS 8th Edition	1.502	1.084	1.856
FastBridge aReading	NSD	NSD	NSD
i-Ready	2.992	2.618	3.189
mCLASS	1.508	0.981	1.980
Star Early Literacy	N/A	N/A	N/A
Star Reading	3.207	2.504	3.801

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data

Table B14. Average, Minimum, and Maximum Linking Error, in Terms of MCAS Scale Scores, At the Partially Meeting Expectations MCAS Performance Level Cut Score by Screening Assessment and Grade, Aggregated Across Time of Year, Grade 3

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	28.284	24.520	30.971
DIBELS 8th Edition	1.118	0.976	1.288
FastBridge aReading	4.066	3.050	5.059
i-Ready	2.492	2.360	2.664
mCLASS	0.880	0.650	1.085
Star Early Literacy	N/A	N/A	N/A

Early literacy screening assessment	Avg	Min	Max
Star Reading	3.528	2.916	4.167

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data

Table B15. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores, Grade 1

Early literacy screening assessment	Accuracy	N
Acadience Reading	NSD	NSD
DIBELS 8th Edition	0.844	4,240
FastBridge aReading	N/A	N/A
i-Ready	0.866	1,744
mCLASS	0.864	1,689
Star Early Literacy	0.765	5,152
Star Reading	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B16. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores, Grade 2

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.935	1,247
DIBELS 8th Edition	0.883	7,507

Early literacy screening assessment	Accuracy	N
FastBridge aReading	NSD	NSD
i-Ready	0.868	6,646
mCLASS	0.872	6,921
Star Early Literacy	N/A	N/A
Star Reading	0.872	11,204

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B17. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores, Grade 3

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.941	1,398
DIBELS 8th Edition	0.870	17,318
FastBridge aReading	0.883	1,713
i-Ready	0.876	11,578
mCLASS	0.851	33,379
Star Early Literacy	N/A	N/A
Star Reading	0.880	15,508

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B18. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and IEP Status, Grade 1

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	NSD	NSD
Acadience Reading	Yes	NSD	NSD
DIBELS 8th Edition	No	0.870	3,623
DIBELS 8th Edition	Yes	0.689	617
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
i-Ready	No	0.896	1,449
i-Ready	Yes	0.715	295
mCLASS	No	0.907	1,402
mCLASS	Yes	0.655	287
Star Early Literacy	No	0.787	4,249
Star Early Literacy	Yes	0.659	903
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B19. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and IEP Status, Grade 2

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.967	993
Acadience Reading	Yes	0.811	254
DIBELS 8th Edition	No	0.918	6,171
DIBELS 8th Edition	Yes	0.719	1,328
FastBridge aReading	No	NSD	NSD
FastBridge aReading	Yes	NSD	NSD
i-Ready	No	0.904	5,285
i-Ready	Yes	0.731	1,353
mCLASS	No	0.905	5,730
mCLASS	Yes	0.710	1,191
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.897	9,381
Star Reading	Yes	0.738	1,823

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B20. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and IEP Status, Grade 3

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.979	1,097
Acadience Reading	Yes	0.804	301
DIBELS 8th Edition	No	0.925	13,390
DIBELS 8th Edition	Yes	0.682	3,921
FastBridge aReading	No	0.941	1,350
FastBridge aReading	Yes	0.669	363
i-Ready	No	0.924	8,785
i-Ready	Yes	0.722	2,783
mCLASS	No	0.900	26,023
mCLASS	Yes	0.674	7,353
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.913	12,071
Star Reading	Yes	0.762	3,434

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students.
Source: Early literacy screening assessment data.

Table B21. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Low Income Status, Grade 1

Early literacy screening assessment	LI	Accuracy	*N*
Acadience Reading	No	NSD	NSD
Acadience Reading	Yes	NSD	NSD
DIBELS 8th Edition	No	0.924	2,237
DIBELS 8th Edition	Yes	0.754	2,003
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
i-Ready	No	0.886	1,240
i-Ready	Yes	0.815	504
mCLASS	No	0.928	880
mCLASS	Yes	0.795	809
Star Early Literacy	No	0.875	1,519
Star Early Literacy	Yes	0.719	3,633
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B22. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Low Income Status, Grade 2

Early literacy screening assessment	LI	Accuracy	*N*
Acadience Reading	No	0.947	1,080
Acadience Reading	Yes	0.856	167
DIBELS 8th Edition	No	0.934	4,582
DIBELS 8th Edition	Yes	0.804	2,917
FastBridge aReading	No	NSD	NSD
FastBridge aReading	Yes	NSD	NSD
i-Ready	No	0.911	3,929
i-Ready	Yes	0.806	2,709
mCLASS	No	0.919	3,922
mCLASS	Yes	0.809	2,999
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.917	5,890
Star Reading	Yes	0.822	5,314

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B23. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Low Income Status, Grade 3

Early literacy screening assessment	LI	Accuracy	*N*
Acadience Reading	No	0.957	1,220
Acadience Reading	Yes	0.831	178
DIBELS 8th Edition	No	0.908	11,020
DIBELS 8th Edition	Yes	0.804	6,291
FastBridge aReading	No	0.920	1,013
FastBridge aReading	Yes	0.830	700
i-Ready	No	0.911	6,764
i-Ready	Yes	0.825	4,804
mCLASS	No	0.909	16,686
mCLASS	Yes	0.792	16,690
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.921	7,904
Star Reading	Yes	0.837	7,601

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students.
Source: Early literacy screening assessment data.

Table B24. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Race/Ethnicity, Grade 1

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
Acadience Reading	White	NSD	NSD
DIBELS 8th Edition	Asian	0.910	288
DIBELS 8th Edition	Black	0.729	166
DIBELS 8th Edition	Hispanic	0.757	1,350
DIBELS 8th Edition	Other	0.851	181
DIBELS 8th Edition	White	0.895	2255
FastBridge aReading	Hispanic	N/A	N/A
FastBridge aReading	Other	N/A	N/A
FastBridge aReading	White	N/A	N/A
i-Ready	Asian	NSD	NSD
i-Ready	Black	0.791	129
i-Ready	Hispanic	0.876	121
i-Ready	Other	NSD	NSD
i-Ready	White	0.868	1,376
mCLASS	Asian	0.900	642
mCLASS	Black	0.743	105
mCLASS	Hispanic	0.834	151
mCLASS	Other	NSD	NSD
mCLASS	White	0.876	709
Star Early Literacy	Asian	0.823	237
Star Early Literacy	Black	0.778	595
Star Early Literacy	Hispanic	0.707	2,228

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
Star Early Literacy	Other	0.771	236
Star Early Literacy	White	0.822	1,856
Star Reading	Asian	N/A	N/A
Star Reading	Black	N/A	N/A
Star Reading	Hispanic	N/A	N/A
Star Reading	Other	N/A	N/A
Star Reading	White	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B25. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Race/Ethnicity, Grade 2

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
Acadience Reading	White	0.938	1,121
DIBELS 8th Edition	Asian	0.884	481
DIBELS 8th Edition	Black	0.809	341
DIBELS 8th Edition	Hispanic	0.804	1,675
DIBELS 8th Edition	Other	0.917	349
DIBELS 8th Edition	White	0.914	4,661
FastBridge aReading	Hispanic	NSD	NSD
FastBridge aReading	Other	NSD	NSD
FastBridge aReading	White	NSD	NSD

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
i-Ready	Asian	0.919	248
i-Ready	Black	0.800	330
i-Ready	Hispanic	0.804	1,838
i-Ready	Other	0.920	311
i-Ready	White	0.897	3,919
mCLASS	Asian	0.939	1,290
mCLASS	Black	0.816	629
mCLASS	Hispanic	0.796	1,519
mCLASS	Other	0.843	383
mCLASS	White	0.896	3,100
Star Early Literacy	Asian	N/A	N/A
Star Early Literacy	Black	N/A	N/A
Star Early Literacy	Hispanic	N/A	N/A
Star Early Literacy	Other	N/A	N/A
Star Early Literacy	White	N/A	N/A
Star Reading	Asian	0.938	291
Star Reading	Black	0.832	762
Star Reading	Hispanic	0.816	2,935
Star Reading	Other	0.865	569
Star Reading	White	0.898	6,647

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B26. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Race/Ethnicity, Grade 3

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
Acadience Reading	White	0.939	1,257
DIBELS 8th Edition	Asian	0.902	961
DIBELS 8th Edition	Black	0.805	734
DIBELS 8th Edition	Hispanic	0.807	4,060
DIBELS 8th Edition	Other	0.880	752
DIBELS 8th Edition	White	0.895	10,811
FastBridge aReading	Hispanic	0.876	177
FastBridge aReading	Other	0.896	134
FastBridge aReading	White	0.886	1,238
i-Ready	Asian	0.922	334
i-Ready	Black	0.868	736
i-Ready	Hispanic	0.808	3,151
i-Ready	Other	0.921	483
i-Ready	White	0.902	6,874
mCLASS	Asian	0.933	2,931
mCLASS	Black	0.807	5,089
mCLASS	Hispanic	0.789	9,520
mCLASS	Other	0.850	1,689
mCLASS	White	0.891	14,150
Star Early Literacy	Asian	N/A	N/A
Star Early Literacy	Black	N/A	N/A
Star Early Literacy	Hispanic	N/A	N/A

Early literacy screening assessment	Race/ ethnicity	Accuracy	N
Star Early Literacy	Other	N/A	N/A
Star Early Literacy	White	N/A	N/A
Star Reading	Asian	0.910	466
Star Reading	Black	0.854	1,093
Star Reading	Hispanic	0.827	4,886
Star Reading	Other	0.862	744
Star Reading	White	0.914	8,319

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students.
Source: Early literacy screening assessment data.

Table B27. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Gender, Grade 1

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	NSD	NSD
Acadience Reading	MA	NSD	NSD
DIBELS 8th Edition	FE	0.862	2,124
DIBELS 8th Edition	MA	0.827	2,113
FastBridge aReading	FE	N/A	N/A
FastBridge aReading	MA	N/A	N/A
i-Ready	FE	0.884	865
i-Ready	MA	0.848	879
mCLASS	FE	0.854	793

Early literacy screening assessment	Gender	Accuracy	N
mCLASS	MA	0.874	896
Star Early Literacy	FE	0.778	2,550
Star Early Literacy	MA	0.752	2,602
Star Reading	FE	N/A	N/A
Star Reading	MA	N/A	N/A

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B28. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Gender, Grade 2

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.930	631
Acadience Reading	MA	0.940	616
DIBELS 8th Edition	FE	0.899	3,729
DIBELS 8th Edition	MA	0.868	3,766
FastBridge aReading	FE	NSD	NSD
FastBridge aReading	MA	NSD	NSD
i-Ready	FE	0.891	3,197
i-Ready	MA	0.848	3,437
mCLASS	FE	0.873	3,332
mCLASS	MA	0.870	3,589
Star Early Literacy	FE	N/A	N/A

Early literacy screening assessment	Gender	Accuracy	N
Star Early Literacy	MA	N/A	N/A
Star Reading	FE	0.880	5,652
Star Reading	MA	0.863	5,550

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics. *Source:* Early literacy screening assessment data.

Table B29. Proportion of Students With the Same Not Meeting Expectations MCAS Benchmark Between the Linked and Observed MCAS Scores by Grade Level and Gender, Grade 3

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.944	697
Acadience Reading	MA	0.939	701
DIBELS 8th Edition	FE	0.886	8,524
DIBELS 8th Edition	MA	0.855	8,778
FastBridge aReading	FE	0.904	874
FastBridge aReading	MA	0.862	839
i-Ready	FE	0.887	5,543
i-Ready	MA	0.865	6,019
mCLASS	FE	0.858	16,302
mCLASS	MA	0.843	17,056
Star Early Literacy	FE	N/A	N/A
Star Early Literacy	MA	N/A	N/A
Star Reading	FE	0.892	7,727

Early literacy screening assessment	Gender	Accuracy	N
Star Reading	MA	0.868	7,775

Note. Statistics for Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students.
Source: Early literacy screening assessment data.

DIBELS/mCLASS Linking Results

Table B30. Average, Minimum, and Maximum Linking Error, in Terms of DIBELS/mCLASS Scale Scores for the Below Benchmark Cut Scores by Screening Assessment and Grade for a Matched Sample, Aggregated Across Time of Year, Grade 1

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	1.284	0.825	1.881
FastBridge aReading	N/A	N/A	N/A
FastBridge earlyReading	1.539	1.267	2.022
i-Ready	0.772	0.486	1.080
Star Early Literacy	0.606	0.368	0.861
Star Reading	N/A	N/A	N/A

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B31. Average, Minimum, and Maximum Linking Error, in Terms of DIBELS/mCLASS Scale Scores for the Below Benchmark Cut Scores by Screening Assessment and Grade for a Matched Sample, Aggregated Across Time of Year, Grade 2

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	2.034	1.299	2.702
FastBridge aReading	2.643	1.910	3.350
FastBridge earlyReading	N/A	N/A	N/A
i-Ready	0.975	0.724	1.212
Star Early Literacy	N/A	N/A	N/A
Star Reading	0.908	0.698	1.099

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B32. Average, Minimum, and Maximum Linking Error, in Terms of DIBELS/mCLASS Scale Scores for the Below Benchmark Cut Scores by Screening Assessment and Grade for a Matched Sample, Aggregated Across Time of Year, Grade 3

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	2.752	2.341	3.161
FastBridge aReading	2.633	2.186	2.913
FastBridge earlyReading	N/A	N/A	N/A
i-Ready	1.070	0.925	1.202
Star Early Literacy	N/A	N/A	N/A
Star Reading	0.972	0.843	1.096

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B33. Average, Minimum, and Maximum Linking Error, in Terms of DIBELS/mCLASS Scale Scores for the Below Benchmark Cut Scores by Screening Assessment and Grade for a Matched Sample, Aggregated Across Time of Year, Grade K

Early literacy screening assessment	Avg	Min	Max
Acadience Reading	2.112	1.649	2.504
FastBridge aReading	N/A	N/A	N/A
FastBridge earlyReading	2.662	2.083	3.232
i-Ready	1.503	1.202	1.991
Star Early Literacy	1.035	0.786	1.350
Star Reading	N/A	N/A	N/A

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source*: Early literacy screening assessment data.

Table B34. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared with Observed Screening Assessment Scores, Grade 1

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.934	2,640
FastBridge aReading	N/A	N/A
FastBridge earlyReading	0.951	2,708
Star Early Literacy	0.899	14,708
Star Reading	N/A	N/A

Early literacy screening assessment	Accuracy	N
i-Ready	0.703	11,455

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B35. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared with Observed Screening Assessment Scores, Grade 2

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.969	2,015
FastBridge aReading	0.950	2,114
FastBridge earlyReading	N/A	N/A
Star Early Literacy	N/A	N/A
Star Reading	0.959	12,749
i-Ready	0.808	13,120

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B36. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared with Observed Screening Assessment Scores, Grade 3

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.929	1,420
FastBridge aReading	0.922	1,939

Early literacy screening assessment	Accuracy	N
FastBridge earlyReading	N/A	N/A
Star Early Literacy	N/A	N/A
Star Reading	0.933	15,931
i-Ready	0.813	11,959

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B37. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared with Observed Screening Assessment Scores, Grade K

Early literacy screening assessment	Accuracy	N
Acadience Reading	0.948	2,763
FastBridge aReading	N/A	N/A
FastBridge earlyReading	0.917	2,744
Star Early Literacy	0.953	14,629
Star Reading	N/A	N/A
i-Ready	0.767	9,034

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B38. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Gender, Grade 1

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.920	1,231
Acadience Reading	MA	0.947	1,398
FastBridge aReading	FE	N/A	N/A
FastBridge aReading	MA	N/A	N/A
FastBridge earlyReading	FE	0.943	1,206
FastBridge earlyReading	MA	0.954	1,334
Star Early Literacy	FE	0.899	7,126
Star Early Literacy	MA	0.898	7,561
Star Reading	FE	N/A	N/A
Star Reading	MA	N/A	N/A
i-Ready	FE	0.700	5,590
i-Ready	MA	0.706	5,841

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B39. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Gender, Grade 2

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.972	980

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	MA	0.966	1,032
FastBridge aReading	FE	0.944	857
FastBridge aReading	MA	0.952	973
FastBridge earlyReading	FE	N/A	N/A
FastBridge earlyReading	MA	N/A	N/A
Star Early Literacy	FE	N/A	N/A
Star Early Literacy	MA	N/A	N/A
Star Reading	FE	0.957	6,182
Star Reading	MA	0.960	6,553
i-Ready	FE	0.805	6,296
i-Ready	MA	0.811	6,802

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B40. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Gender, Grade 3

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.944	698
Acadience Reading	MA	0.914	721
FastBridge aReading	FE	0.937	887
FastBridge aReading	MA	0.908	850

Early literacy screening assessment	Gender	Accuracy	N
FastBridge earlyReading	FE	N/A	N/A
FastBridge earlyReading	MA	N/A	N/A
Star Early Literacy	FE	N/A	N/A
Star Early Literacy	MA	N/A	N/A
Star Reading	FE	0.931	7,899
Star Reading	MA	0.935	8,018
i-Ready	FE	0.809	5,685
i-Ready	MA	0.816	6,245

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B41. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Gender, Grade K

Early literacy screening assessment	Gender	Accuracy	N
Acadience Reading	FE	0.950	1,252
Acadience Reading	MA	0.947	1,293
FastBridge aReading	FE	N/A	N/A
FastBridge aReading	MA	N/A	N/A
FastBridge earlyReading	FE	0.902	1,343
FastBridge earlyReading	MA	0.927	1,270
Star Early Literacy	FE	0.951	7,054

Early literacy screening assessment	Gender	Accuracy	N
Star Early Literacy	MA	0.955	7,430
Star Reading	FE	N/A	N/A
Star Reading	MA	N/A	N/A
i-Ready	FE	0.776	4,374
i-Ready	MA	0.759	4,542

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B42. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and IEP Status, Grade 1

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.938	2,198
Acadience Reading	Yes	0.916	431
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
FastBridge earlyReading	No	0.952	2,044
FastBridge earlyReading	Yes	0.938	499
Star Early Literacy	No	0.901	12,067
Star Early Literacy	Yes	0.885	2,622
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A

Early literacy screening assessment	IEP	Accuracy	N
i-Ready	No	0.717	9,319
i-Ready	Yes	0.645	2,115

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B43. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and IEP Status, Grade 2

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.979	1,552
Acadience Reading	Yes	0.935	460
FastBridge aReading	No	0.968	1,425
FastBridge aReading	Yes	0.879	405
FastBridge earlyReading	No	N/A	N/A
FastBridge earlyReading	Yes	N/A	N/A
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.961	10,580
Star Reading	Yes	0.948	2,162
i-Ready	No	0.804	10,236
i-Ready	Yes	0.824	2,868

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B44. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and IEP Status, Grade 3

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.945	1,106
Acadience Reading	Yes	0.872	313
FastBridge aReading	No	0.939	1,364
FastBridge aReading	Yes	0.863	373
FastBridge earlyReading	No	N/A	N/A
FastBridge earlyReading	Yes	N/A	N/A
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.937	12,378
Star Reading	Yes	0.918	3,542
i-Ready	No	0.811	9,035
i-Ready	Yes	0.820	2,901

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B45. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and IEP Status, Grade K

Early literacy screening assessment	IEP	Accuracy	N
Acadience Reading	No	0.954	2,064
Acadience Reading	Yes	0.926	484
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
FastBridge earlyReading	No	0.919	2,193
FastBridge earlyReading	Yes	0.888	420
Star Early Literacy	No	0.955	12,369
Star Early Literacy	Yes	0.944	2,115
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A
i-Ready	No	0.776	7,399
i-Ready	Yes	0.726	1,519

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B46. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Race/Ethnicity, Grade 1

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Black	0.942	120

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Hispanic	0.936	595
Acadience Reading	Other	0.944	107
Acadience Reading	White	0.934	1,790
Acadience Reading	Asian	NSD	NSD
FastBridge aReading	Hispanic	N/A	N/A
FastBridge aReading	Other	N/A	N/A
FastBridge aReading	White	N/A	N/A
FastBridge earlyReading	Black	0.938	176
FastBridge earlyReading	Hispanic	0.932	322
FastBridge earlyReading	Other	0.962	390
FastBridge earlyReading	White	0.952	1,740
Star Early Literacy	Asian	0.898	480
Star Early Literacy	Black	0.896	1,039
Star Early Literacy	Hispanic	0.876	5,471
Star Early Literacy	Other	0.913	793
Star Early Literacy	White	0.915	6,925
Star Reading	Black	N/A	N/A
Star Reading	Hispanic	N/A	N/A
Star Reading	Other	N/A	N/A
Star Reading	White	N/A	N/A
Star Reading	Asian	N/A	N/A
i-Ready	Asian	0.744	395
i-Ready	Black	0.706	660
i-Ready	Hispanic	0.690	3,585

Early literacy screening assessment	Race/ethnicity	Accuracy	N
i-Ready	Other	0.692	520
i-Ready	White	0.709	6,295

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics.
Source: Early literacy screening assessment data.

Table B47. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Race/Ethnicity, Grade 2

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Black	NSD	NSD
Acadience Reading	Hispanic	0.985	328
Acadience Reading	Other	NSD	NSD
Acadience Reading	White	0.967	1,517
Acadience Reading	Asian	NSD	NSD
FastBridge aReading	Hispanic	0.940	151
FastBridge aReading	Other	0.952	398
FastBridge aReading	White	0.947	1,398
FastBridge earlyReading	Black	N/A	N/A
FastBridge earlyReading	Hispanic	N/A	N/A
FastBridge earlyReading	Other	N/A	N/A
FastBridge earlyReading	White	N/A	N/A
Star Early Literacy	Asian	N/A	N/A

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Star Early Literacy	Black	N/A	N/A
Star Early Literacy	Hispanic	N/A	N/A
Star Early Literacy	Other	N/A	N/A
Star Early Literacy	White	N/A	N/A
Star Reading	Black	0.956	904
Star Reading	Hispanic	0.956	3,232
Star Reading	Other	0.968	659
Star Reading	White	0.958	7,598
Star Reading	Asian	0.972	356
i-Ready	Asian	0.852	371
i-Ready	Black	0.795	848
i-Ready	Hispanic	0.815	3,614
i-Ready	Other	0.800	564
i-Ready	White	0.805	7,723

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics.
Source: Early literacy screening assessment data.

Table B48. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Race/Ethnicity, Grade 3

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Black	NSD	NSD

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Hispanic	NSD	NSD
Acadience Reading	Other	NSD	NSD
Acadience Reading	White	0.929	1,274
Acadience Reading	Asian	NSD	NSD
FastBridge aReading	Hispanic	0.962	185
FastBridge aReading	Other	0.886	342
FastBridge aReading	White	0.927	1,244
FastBridge earlyReading	Black	N/A	N/A
FastBridge earlyReading	Hispanic	N/A	N/A
FastBridge earlyReading	Other	N/A	N/A
FastBridge earlyReading	White	N/A	N/A
Star Early Literacy	Asian	N/A	N/A
Star Early Literacy	Black	N/A	N/A
Star Early Literacy	Hispanic	N/A	N/A
Star Early Literacy	Other	N/A	N/A
Star Early Literacy	White	N/A	N/A
Star Reading	Black	0.910	1,140
Star Reading	Hispanic	0.916	5,152
Star Reading	Other	0.939	770
Star Reading	White	0.946	8,393
Star Reading	Asian	0.943	476
i-Ready	Asian	0.799	344
i-Ready	Black	0.788	777
i-Ready	Hispanic	0.822	3,379

Early literacy screening assessment	Race/ethnicity	Accuracy	N
i-Ready	Other	0.795	502
i-Ready	White	0.813	6,957

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. NSD means that there was not sufficient data to report statistics.
Source: Early literacy screening assessment data.

Table B49. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Race/Ethnicity, Grade K

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Acadience Reading	Black	0.932	103
Acadience Reading	Hispanic	0.914	618
Acadience Reading	Other	0.928	333
Acadience Reading	White	0.963	1,583
Acadience Reading	Asian	0.984	126
FastBridge aReading	Hispanic	N/A	N/A
FastBridge aReading	Other	N/A	N/A
FastBridge aReading	White	N/A	N/A
FastBridge earlyReading	Black	0.910	178
FastBridge earlyReading	Hispanic	0.885	331
FastBridge earlyReading	Other	0.921	342
FastBridge earlyReading	White	0.921	1,801
Star Early Literacy	Asian	0.958	721

Early literacy screening assessment	Race/ethnicity	Accuracy	N
Star Early Literacy	Black	0.940	1,090
Star Early Literacy	Hispanic	0.945	5,096
Star Early Literacy	Other	0.957	1,018
Star Early Literacy	White	0.961	6,704
Star Reading	Black	N/A	N/A
Star Reading	Hispanic	N/A	N/A
Star Reading	Other	N/A	N/A
Star Reading	White	N/A	N/A
Star Reading	Asian	N/A	N/A
i-Ready	Asian	0.824	238
i-Ready	Black	0.758	604
i-Ready	Hispanic	0.667	2,783
i-Ready	Other	0.786	457
i-Ready	White	0.819	4,952

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B50. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Low Income Status, Grade 1

Early literacy screening assessment	LI	Accuracy	N
Acadience Reading	No	0.938	1,623

Early literacy screening assessment	LI	Accuracy	N
Acadience Reading	Yes	0.929	1,006
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
FastBridge earlyReading	No	0.970	1,286
FastBridge earlyReading	Yes	0.927	1,257
Star Early Literacy	No	0.912	6,615
Star Early Literacy	Yes	0.887	8,074
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A
i-Ready	No	0.726	6,180
i-Ready	Yes	0.676	5,254

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B51. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Low Income Status, Grade 2

Early literacy screening assessment	LI	Accuracy	N
Acadience Reading	No	0.972	1,384
Acadience Reading	Yes	0.963	628
FastBridge aReading	No	0.969	1,089
FastBridge aReading	Yes	0.918	741

Early literacy screening assessment	LI	Accuracy	N
FastBridge earlyReading	No	N/A	N/A
FastBridge earlyReading	Yes	N/A	N/A
Star Early Literacy	No	N/A	N/A
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.963	6,976
Star Reading	Yes	0.953	5,766
i-Ready	No	0.812	7,301
i-Ready	Yes	0.803	5,803

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B52. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Low Income Status, Grade 3

Early literacy screening assessment	LI	Accuracy	N
Acadience Reading	No	0.935	1,238
Acadience Reading	Yes	0.884	181
FastBridge aReading	No	0.945	1,017
FastBridge aReading	Yes	0.892	720
FastBridge earlyReading	No	N/A	N/A
FastBridge earlyReading	Yes	N/A	N/A
Star Early Literacy	No	N/A	N/A

Early literacy screening assessment	LI	Accuracy	N
Star Early Literacy	Yes	N/A	N/A
Star Reading	No	0.946	7,971
Star Reading	Yes	0.920	7,949
i-Ready	No	0.820	6,837
i-Ready	Yes	0.803	5,099

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Table B53. Proportion of Students Classified At the Well Below Benchmark Using Linked Screening Assessment Scores Compared With Observed Screening Assessment Scores by Grade and Low Income Status, Grade K

Early literacy screening assessment	LI	Accuracy	N
Acadience Reading	No	0.964	1,541
Acadience Reading	Yes	0.925	1,007
FastBridge aReading	No	N/A	N/A
FastBridge aReading	Yes	N/A	N/A
FastBridge earlyReading	No	0.945	1,317
FastBridge earlyReading	Yes	0.883	1,296
Star Early Literacy	No	0.962	7,204
Star Early Literacy	Yes	0.945	7,280
Star Reading	No	N/A	N/A
Star Reading	Yes	N/A	N/A

Early literacy screening assessment	LI	Accuracy	N
i-Ready	No	0.826	4,778
i-Ready	Yes	0.700	4,140

Note. Statistics for FastBridge earlyReading and Star Early Literacy were calculated only for grade 1 students, and statistics for FastBridge aReading and Star Reading were calculated only for grades 2 and 3 students. *Source:* Early literacy screening assessment data.

Tables B54–B72 provide the percentages of students significantly below benchmark using different benchmarks (results of comparability testing).

Table B54. Percentage of Kindergarten Students Significantly Below Benchmark, Publisher-Provided Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	24%	20%	11%
DIBELS 8th Edition	38%	25%	13%
EarlyBird	58%	39%	26%
FastBridge earlyReading	10%	17%	17%
MAP Reading Fluency	32%	26%	17%
Star Early Literacy	42%	24%	22%
Star Early Literacy Spanish	55%	50%	61%
i-Ready	3%	37%	46%
mCLASS	40%	35%	24%
mCLASS Lectura	37%	26%	17%
All screening assessments	**36%**	**30%**	**22%**

Source: Early literacy screening assessment data.

Table B55. Percentage of Kindergarten Students Significantly Below Benchmark, 25th National Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	21%	29%	28%
DIBELS 8th Edition	16%	14%	14%
EarlyBird	N/A	N/A	N/A
FastBridge earlyReading	21%	37%	31%
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	42%	25%	23%
Star Early Literacy Spanish	56%	52%	61%
i-Ready	25%	23%	26%
mCLASS	20%	25%	24%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**24%**	**23%**	**23%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B56. Percentage of Kindergarten Students Significantly Below Benchmark, Massachusetts 25th Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	20%	18%	18%
DIBELS 8th Edition	24%	23%	20%
EarlyBird	N/A	N/A	N/A
FastBridge earlyReading	12%	14%	16%

Early literacy screening assessment	BOY	MOY	EOY
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	27%	26%	27%
Star Early Literacy Spanish	N/A	N/A	N/A
i-Ready	22%	24%	23%
mCLASS	28%	30%	28%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**25%**	**26%**	**25%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B57. Percentage of Kindergarten Students Significantly Below Benchmark, DIBELS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	30%	24%	14%
DIBELS 8th Edition	38%	25%	13%
EarlyBird	N/A	N/A	N/A
FastBridge earlyReading	21%	20%	11%
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	37%	30%	22%
Star Early Literacy Spanish	N/A	N/A	N/A
i-Ready	31%	27%	15%
mCLASS	40%	35%	24%

Early literacy screening assessment	BOY	MOY	EOY
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**37%**	**30%**	**19%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B58. Percentage of Grade 1 Students Significantly Below Benchmark, Publisher-Provided Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	22%	29%	23%
DIBELS 8th Edition	30%	25%	13%
FastBridge earlyReading	19%	24%	20%
MAP Reading Fluency	34%	28%	20%
Star Early Literacy	44%	37%	35%
Star Early Literacy Spanish	76%	79%	79%
i-Ready	12%	59%	53%
mCLASS	34%	33%	21%
mCLASS Lectura	41%	41%	43%
All screening assessments	**32%**	**35%**	**26%**

Source: Early literacy screening assessment data.

Table B59. Percentage of Grade 1 Students Significantly Below Benchmark, 25th National Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	22%	29%	25%
DIBELS 8th Edition	18%	14%	15%
FastBridge earlyReading	37%	37%	32%
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	45%	38%	36%
Star Early Literacy Spanish	77%	79%	81%
i-Ready	33%	28%	26%
mCLASS	22%	22%	23%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**27%**	**25%**	**25%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B60. Percentage of Grade 1 Students Significantly Below Benchmark, Massachusetts 25th Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	17%	17%	16%
DIBELS 8th Edition	21%	20%	19%
FastBridge earlyReading	11%	16%	18%
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	23%	26%	26%

Early literacy screening assessment	BOY	MOY	EOY
Star Early Literacy Spanish	N/A	N/A	N/A
i-Ready	23%	23%	23%
mCLASS	26%	29%	28%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**24%**	**25%**	**24%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B61. Percentage of Grade 1 Students Significantly Below Benchmark, DIBELS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	24%	22%	12%
DIBELS 8th Edition	30%	25%	13%
FastBridge earlyReading	17%	16%	13%
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	34%	32%	19%
Star Early Literacy Spanish	N/A	N/A	N/A
i-Ready	32%	26%	16%
mCLASS	34%	33%	21%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**32%**	**29%**	**18%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B62. Percentage of Grade 1 Students Significantly Below Benchmark, MCAS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	NSD	NSD	NSD
DIBELS 8th Edition	14%	14%	15%
FastBridge earlyReading	N/A	N/A	N/A
MAP Reading Fluency	N/A	N/A	N/A
Star Early Literacy	23%	25%	31%
Star Early Literacy Spanish	N/A	N/A	N/A
i-Ready	10%	9%	7%
mCLASS	8%	10%	10%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**18%**	**18%**	**19%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B63. Percentage of Grade 2 Students Significantly Below Benchmark, Publisher-Provided Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	23%	22%	21%
DIBELS 8th Edition	26%	22%	16%

Early literacy screening assessment	BOY	MOY	EOY
FastBridge aReading	21%	19%	14%
MAP Reading Fluency	27%	29%	30%
Star Reading	38%	26%	23%
Star Reading Spanish	22%	11%	22%
i-Ready	27%	50%	55%
mCLASS	36%	32%	26%
mCLASS Lectura	33%	24%	18%
All screening assessments	**32%**	**31%**	**28%**

Source: Early literacy screening assessment data.

Table B64. Percentage of Grade 2 Students Significantly Below Benchmark, 25th National Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	28%	23%	24%
DIBELS 8th Edition	15%	14%	15%
FastBridge aReading	32%	29%	24%
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	39%	26%	23%
Star Reading Spanish	22%	11%	22%
i-Ready	28%	28%	26%
mCLASS	24%	23%	23%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**25%**	**23%**	**22%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B65. Percentage of Grade 2 Students Significantly Below Benchmark, Massachusetts 25th Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	18%	19%	19%
DIBELS 8th Edition	20%	18%	19%
FastBridge aReading	11%	13%	10%
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	24%	23%	26%
Star Reading Spanish	N/A	N/A	N/A
i-Ready	25%	25%	24%
mCLASS	30%	27%	28%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**25%**	**24%**	**25%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B66. Percentage of Grade 2 Students Significantly Below Benchmark, DIBELS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	24%	16%	17%
DIBELS 8th Edition	26%	22%	16%
FastBridge aReading	15%	17%	8%

Early literacy screening assessment	BOY	MOY	EOY
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	29%	27%	21%
Star Reading Spanish	N/A	N/A	N/A
i-Ready	32%	29%	23%
mCLASS	36%	32%	26%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**31%**	**28%**	**22%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B67. Percentage of Grade 2 Students Significantly Below Benchmark, MCAS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	5%	5%	5%
DIBELS 8th Edition	12%	13%	12%
FastBridge aReading	NSD	NSD	NSD
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	14%	13%	15%
Star Reading Spanish	N/A	N/A	N/A
i-Ready	14%	15%	16%
mCLASS	13%	15%	14%
mCLASS Lectura	N/A	N/A	N/A
All screening assessments	**16%**	**17%**	**17%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. NSD means that there was not sufficient data to report statistics. *Source*: Early literacy screening assessment data.

Table B68. Percentage of Grade 3 Students Significantly Below Benchmark, Publisher-Provided Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	14%	13%	12%
DIBELS 8th Edition	22%	24%	20%
FastBridge aReading	12%	12%	12%
MAP Reading Fluency	39%	42%	45%
Star Reading	39%	30%	32%
Star Reading Spanish	76%	72%	74%
i-Ready	36%	41%	62%
mCLASS	31%	32%	27%
All screening assessments	**31%**	**31%**	**32%**

Source: Early literacy screening assessment data.

Table B69. Percentage of Grade 3 Students Significantly Below Benchmark, 25th National Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	15%	14%	14%
DIBELS 8th Edition	18%	17%	17%
FastBridge aReading	21%	22%	19%
MAP Reading Fluency	N/A	N/A	N/A

Early literacy screening assessment	BOY	MOY	EOY
Star Reading	40%	31%	32%
Star Reading Spanish	77%	74%	75%
i-Ready	31%	29%	29%
mCLASS	26%	24%	24%
All screening assessments	**28%**	**25%**	**26%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B70. Percentage of Grade 3 Students Significantly Below Benchmark, Massachusetts 25th Percentile

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	17%	15%	17%
DIBELS 8th Edition	20%	20%	21%
FastBridge aReading	15%	18%	16%
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	26%	25%	25%
Star Reading Spanish	N/A	N/A	N/A
i-Ready	25%	24%	24%
mCLASS	28%	27%	28%
All screening assessments	**25%**	**24%**	**25%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B71. Percentage of Grade 3 Students Significantly Below Benchmark, DIBELS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	21%	21%	18%
DIBELS 8th Edition	22%	24%	20%
FastBridge aReading	19%	21%	15%
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	28%	28%	24%
Star Reading Spanish	N/A	N/A	N/A
i-Ready	27%	29%	24%
mCLASS	31%	32%	27%
All screening assessments	**27%**	**29%**	**24%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Table B72. Percentage of Grade 3 Students Significantly Below Benchmark, MCAS Benchmark

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	4%	5%	5%
DIBELS 8th Edition	15%	15%	13%
FastBridge aReading	11%	12%	13%
MAP Reading Fluency	N/A	N/A	N/A
Star Reading	18%	19%	19%
Star Reading Spanish	N/A	N/A	N/A

Early literacy screening assessment	BOY	MOY	EOY
i-Ready	19%	17%	18%
mCLASS	19%	18%	18%
All screening assessments	**18%**	**17%**	**17%**

Note. N/A means that the statistic was not calculated for the relevant assessment, test period, and grade level because the risk performance level did not have a national percentile or the assessment was a Spanish-language assessment. *Source*: Early literacy screening assessment data.

Equipercentile Linking Procedure

The equipercentile linking procedure that was used to link literacy screening assessment scale scores to MCAS ELA scale scores involved the following three steps. A similar process was used to link screening assessment scale scores to combined DIBELS/mCLASS data.

Linking Steps:

1. Valid MCAS Grade 3 ELA scale scores were collected for students in the concordance sample.
2. The composite score distribution was obtained for each literacy screening assessment concordance sample, and the cumulative proportion of students who fell at or below each benchmark cut score of interest was estimated. We conducted loglinear pre-smoothing on the distributions before linking.
3. The point was found on the MCAS Grade 3 ELA scale at which the estimated proportion of students equaled the estimated proportion of students who scored at or below the benchmark cut score of interest obtained in the previous step.

To define percentile ranks in the concordance samples, let K_X represent the scale score range on Form X of a test (i.e., the literacy screening assessment). Define $F(x)$ as the continuous density function of the probability that scale score X lies within the range $\mathrm{P}(a \leq \mathrm{X} \leq b)$. Define the cumulative distribution function as the proportion of examinees earning a scale score at or below x; that is, $F(x) = \int_{min}^{x} f(t)dt$, for $Minimum\ Scale\ Score \leq x \leq Maximum\ Scale\ Score$. Define x^* as a scale score that is closest to x such that $x^* - 0.5(x - a) \leq x < x^* + 0.5(b - x)$ where a and b are the scale scores that are immediately below and above x, respectively.

For example, in a case where consecutive scale scores are 98, 100, and 102, if x = 99 then x^* = 98 and if x = 100.99 then x^* = 100. The percentile rank function for Form X can be written as

$$P(x) = 100\left\{F(a) + \left(\frac{x-a}{b-a}\right)[F(x) - F(a)]\right\} \qquad (1)$$

In equipercentile equating, the interest is in finding a score on Form Y (e.g., MCAS Grade 3 ELA) that has the same percentile rank as Form X. To do this, we find the inverse of the percentile rank function for Form Y, Q^{-1}, to find the equipercentile equivalent of score x on Form X. Q^{-1} can be defined as

$$Q^{-1}[P(x)] = \frac{\frac{P(x)}{100} - G(y_{U-1}^*)}{G(y_U^*) - G(y_{U-1}^*)} + \left(y_U^* - \frac{y_U^* - (y_{U-1}^*)}{2}\right) \qquad (2)$$

where y_U^* is the lowest scale score with a cumulative percent that is greater than $P(x)$, and y_{U-1}^* is the scale score that is immediately below y_U^*. Equations 1 and 2 were used to produce the equipercentile linking results shown in Table B73 and earlier results.

MCAS Grade 3 Linked Values

Table B73. Number of Districts, Schools, and Students in Benchmark Linking Analysis

Early literacy screening assessment	Number of districts	Number of schools	Number of grade 3 students
Acadience Reading	3	7	471
DIBELS 8th Edition	38	99	6,435
mCLASS	44	188	11,544
FastBridge aReading	5	15	609
i-Ready Diagnostic	29	56	3,976
Star Early Literacy	24	76	982
Star Reading	24	85	5,378

Source: Early literacy screening assessment data.

Table B74. Screening Assessment Grade 3 BOY and EOY Benchmark Cut Scores Linked to MCAS Grade 3 ELA Scale Scores and Performance Levels Using Equipercentile Linking

Early literacy screening assessment	Screening assessment cut score. BOY	Screening assessment cut score EOY	Benchmark	MCAS scale score BOY	MCAS scale score EOY	MCAS performance level BOY	MCAS performance level EOY
Acadience Reading	180	280	Below benchmark	484	481	Partially meeting	Partially meeting
DIBELS 8th Edition	314	424	Below benchmark	475	476	Partially meeting	Partially meeting
mCLASS	314	424	Below benchmark	479	477	Partially meeting	Partially meeting
FastBridge aReading	468	483	Some risk	472	472	Partially meeting	Partially meeting
i-Ready	474	545	Some risk	484	499	Partially meeting	Partially meeting
Star Early Literacy	909	943	On watch	486	484	Partially meeting	Partially meeting
Star Reading	909	943	On watch	486	478	Partially meeting	Partially meeting

Source: Early literacy screening assessment data.

Appendix C. Screening Assessment Overview and Student Performance

Table C1 provides the number of students by year and assessment.

Table C1. Number of Students by Year and Assessment

Early literacy screening assessment	Number in 2021/22	% in 2021/22	Number in 2022/23	% in 2022/23	Number in 2023/24	% in 2023/24	Number in 2022/23 (25th percentile metric)	% in 2022/23 (25th percentile metric)	Number in 2023/24 (25th percentile metric)	% in 2023/24 (25th percentile metric)
Acadience Reading	0	0%	1,591	2%	3,224	2%	1,591	2%	3,224	3%
DIBELS 8th Edition	8,362	30%	17,533	26%	30,152	21%	17,533	27%	30,152	24%
mCLASS	3,359	12%	13,882	21%	51,006	35%	13,881	21%	51,002	40%
mCLASS Lectura	0	0%	0	0%	186	<1%	0	0%	0	0%
EarlyBird	691	2%	1,279	2%	1,796	1%	0	0%	0	0%
FastBridge aReading	578	2%	217	<1%	1,454	1%	217	<1%	1,454	1%

Early literacy screening assessment	Number in 2021/22	% in 2021/22	Number in 2022/23	% in 2022/23	Number in 2023/24	% in 2023/24	Number in 2022/23 (25th percentile metric)	% in 2022/23 (25th percentile metric)	Number in 2023/24 (25th percentile metric)	% in 2023/24 (25th percentile metric)
FastBridge earlyReading	258	<1%	406	<1%	1,994	1%	406	<1%	1,994	2%
i-Ready	6,627	24%	12,622	19%	16,397	11%	12,622	19%	16,397	13%
MAP Reading Fluency	164	<1%	0	0%	14,021	10%	0	0%	0	0%
Star Early Literacy	5,763	20%	9,638	14%	11,030	8%	9,638	15%	11,030	9%
Star Early Literacy Spanish	906	3%	0	0%	1,502	1%	0	0%	1,502	1%
Star Reading	1,483	5%	9,379	14%	10,282	7%	9,379	14%	10,282	8%
Star Reading Spanish	0	0%	0	0%	823	<1%	0	0%	823	1%
Total	**28,191**	**100%**	**66,547**	**100%**	**143,867**	**100%**	**65,267**	**100%**	**127,860**	**100%**

Note. Excludes screening assessments with no data or insufficient data for reporting for 2023/24. EarlyBird includes only data for grade K. FastBridge earlyReading and Star Early Literacy (English and Spanish) include only grades K and 1. FastBridge aReading and Star Reading (English and Spanish) include only grades 2 and 3. Students may be included in multiple rows for the same school year if they took multiple assessments. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Tables C2 and C3 provide the demographic breakdowns for the early literacy screening assessment sample (for 2023/24) by screening assessment.

Table C2. Comparison of Student Demographics of Early Literacy Screening Assessment Sample to Those of the State (by Early Literacy Screening Assessment)

Grade level/ demographic	Acadience Reading	DIBELS 8th Edition	mCLASS	mCLASS Lectura	EarlyBird	FastBridge aReading	State
Kindergarten	29%	27%	25%	42%	100%	0%	24%
Grade 1	28%	28%	25%	22%	0%	0%	25%
Grade 2	28%	23%	26%	20%	0%	52%	26%
Grade 3	15%	22%	25%	16%	0%	48%	25%
Low income	36%	39%	55%	44%	33%	43%	43%
Female	48%	49%	49%	55%	49%	49%	49%
Male	52%	51%	51%	45%	51%	51%	51%
Nonbinary	Sup data	<1%	<1%	0%	Sup data	0%	<1%
English learner	7%	17%	26%	25%	10%	5%	19%
Students receiving special education services	19%	20%	20%	15%	15%	22%	17%
White	91%	86%	68%	76%	77%	87%	77%
Hispanic/Latino	21%	26%	32%	62%	12%	10%	26%
Black	8%	10%	26%	18%	17%	13%	17%
Asian	4%	8%	10%	8%	15%	6%	10%
American Indian/Alaska Native	<1%	5%	3%	5%	3%	2%	4%
Native Hawaiian/Pacific Islander	<1%	<1%	1%	2%	Sup data	Sup data	<1%

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C3. Comparison of Student Demographics of Early Literacy Screening Assessment Sample to Those of the State (by Early Literacy Screening Assessment)—Continued

Grade level/ demographic	FastBridge early-Reading	i-Ready	MAP Reading Fluency	Star Early Literacy	Star Early Literacy Spanish	Star Reading	Star Reading Spanish	State
Kindergarten	48%	22%	25%	49%	47%	0%	0%	24%
Grade 1	52%	24%	26%	51%	53%	0%	0%	25%
Grade 2	0%	28%	26%	0%	0%	46%	14%	26%
Grade 3	0%	25%	24%	0%	0%	54%	86%	25%
Low income	52%	46%	69%	53%	86%	49%	84%	43%
Female	49%	48%	48%	49%	51%	49%	50%	49%
Male	51%	52%	52%	51%	49%	51%	50%	51%
Nonbinary	Sup data	Sup data	Sup data	Sup data	0%	Sup data	0%	<1%
English learner	7%	20%	41%	30%	92%	25%	90%	19%
Students receiving special education services	18%	21%	21%	16%	10%	20%	13%	17%
White	85%	86%	53%	71%	27%	77%	40%	77%
Hispanic/Latino	14%	30%	44%	36%	96%	31%	97%	26%
Black	18%	13%	42%	17%	15%	15%	13%	17%
Asian	5%	5%	9%	6%	1%	5%	1%	10%

Grade level/ demographic	FastBridge early-Reading	i-Ready	MAP Reading Fluency	Star Early Literacy	Star Early Literacy Spanish	Star Reading	Star Reading Spanish	State
American Indian/Alaska Native	2%	3%	3%	14%	65%	10%	59%	4%
Native Hawaiian/Pacific Islander	<1%	<1%	1%	<1%	<1%	<1%	Sup data	<1%

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C4 compares the sample of grade 3 students with the state's grade 3 population by comparing the percentage of students meeting or exceeding expectations on the Massachusetts Comprehensive Assessment System (MCAS; for each screening assessment) to the state average.

Table C4. Comparison of Percentage of Students Meeting/Exceeding Expectations on MCAS

Early literacy screening assessment	Sample (% meeting/ exceeding expectations)	State (% meeting/ exceeding expectations)
All screening assessments	37%	42%
Acadience Reading	59%	42%
DIBELS 8th Edition	42%	42%
mCLASS	37%	42%
mCLASS Lectura	35%	42%
FastBridge aReading	46%	42%
i-Ready	36%	42%
MAP Reading Fluency	32%	42%
Star Reading	35%	42%

Early literacy screening assessment	Sample (% meeting/ exceeding expectations)	State (% meeting/ exceeding expectations)
Star Reading Spanish	9%	42%

Note. No grade 3 students took EarlyBird. Grade 3 students who took FastBridge earlyReading, Star Early Literacy, or Star Early Literacy Spanish were not included in the analysis. *Source*: Early literacy screening assessment data and state-provided MCAS data.

Table C5 provides the number of observations with available benchmarks by time period—beginning of year (BOY), middle of year (MOY), or end of year (EOY)—for each of the early literacy screening assessments.

Table C5. Number of Observations With Benchmarks by Time Period and Screening Assessment

Early literacy screening assessment	BOY (N, %)	MOY (N, %)	EOY (N, %)	Total
Acadience Reading	3,120 (35%)	2,866 (32%)	2,852 (32%)	**8,838**
DIBELS 8th Edition	26,532 (33%)	27,074 (34%)	26,308 (33%)	**79,914**
mCLASS	47,652 (33%)	48,482 (34%)	47,815 (33%)	**143,949**
mCLASS Lectura	181 (37%)	178 (36%)	134 (27%)	**493**
EarlyBird	1,700 (40%)	1,545 (37%)	969 (23%)	**4,214**
FastBridge aReading	1,357 (33%)	1,340 (33%)	1,356 (33%)	**4,053**
FastBridge earlyReading	1,629 (30%)	1,919 (35%)	1,904 (35%)	**5,452**
i-Ready	14,847 (33%)	14,971 (33%)	15,750 (35%)	**45,568**
MAP Reading Fluency	10,909 (34%)	10,756 (33%)	10,749 (33%)	**32,414**
Star Early Literacy	9,700 (33%)	9,938 (34%)	9,699 (33%)	**29,337**
Star Early Literacy Spanish	1,129 (32%)	1,122 (31%)	1,313 (37%)	**3,564**
Star Reading	9,218 (32%)	9,641 (34%)	9,821 (34%)	**28,680**
Star Reading Spanish	544 (31%)	548 (32%)	636 (37%)	**1,728**
Total	**128,518 (33%)**	**130,380 (34%)**	**129,306 (33%)**	**388,204**

Source: Early literacy screening assessment data.

Table C6 provides the number of benchmark scores by time period, grade level, and demographic characteristic.

Table C6. Number of Benchmark Scores by Time Period, Grade Level, and Demographic Characteristics

Grade level/demographic	BOY (N, %)	MOY (N, %)	EOY (N, %)	Total
Kindergarten	31,448 (32%)	32,532 (33%)	34,551 (35%)	**98,531**
Grade 1	34,328 (34%)	34,165 (34%)	33,474 (33%)	**101,967**
Grade 2	31,482 (33%)	31,990 (34%)	31,466 (33%)	**94,938**
Grade 3	31,263 (34%)	31,693 (34%)	29,815 (32%)	**92,771**
Low income	61,874 (33%)	64,131 (34%)	64,222 (34%)	**190,227**
Female	62,256 (33%)	63,381 (34%)	62,833 (33%)	**188,470**
Male	64,833 (33%)	65,689 (34%)	65,125 (33%)	**195,647**
Nonbinary	42 (37%)	35 (30%)	38 (33%)	**115**
English learner	28,686 (32%)	30,370 (33%)	31,642 (35%)	**90,698**
Students receiving special education services	24,850 (33%)	25,001 (34%)	24,702 (33%)	**74,553**
White	95,009 (33%)	96,078 (34%)	94,940 (33%)	**286,027**
Hispanic/Latino	38,651 (33%)	39,787 (34%)	40,024 (34%)	**118,462**
Black	24,548 (33%)	25,026 (34%)	25,096 (34%)	**74,670**
Asian	10,342 (34%)	10,293 (34%)	9,950 (33%)	**30,585**
American Indian/Alaska Native	6,574 (32%)	7,054 (34%)	7,200 (35%)	**20,828**
Native Hawaiian/Pacific Islander	971 (33%)	982 (34%)	960 (33%)	**2,913**

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). A cell may reference the same student more than once if they took multiple screening assessments within the school year. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C7 provides the demographic breakdown of students with one benchmark available versus students with two benchmarks available versus students with three benchmarks available.

Table C7. Comparison of Demographics of Students With One Available Benchmark Versus Two Available Benchmarks Versus Three Available Benchmarks

Grade level/demographic	One benchmark available	Two benchmarks available	Three benchmarks available	State
Kindergarten	8%	23%	69%	24%
Grade 1	8%	11%	81%	25%
Grade 2	8%	12%	81%	26%
Grade 3	7%	13%	79%	25%
Low income	66%	60%	48%	43%
Female	46%	48%	49%	49%
Male	54%	52%	51%	51%
Nonbinary	Sup data	Sup data	<1%	<1%
English learner	43%	34%	22%	19%
Students receiving special education services	21%	21%	19%	17%
White	64%	69%	76%	77%
Hispanic/Latino	45%	40%	29%	26%
Black	27%	24%	19%	17%
Asian	7%	8%	8%	10%
American Indian/Alaska Native	9%	7%	5%	4%
Native Hawaiian/Pacific Islander	<1%	<1%	<1%	<1%

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). Student records can indicate more than one. "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C8 provides the percentage of students at significant risk by time period and for each assessment, using the benchmark and 25th percentile metrics.

Table C8. Assessment Performance by Time Period and Assessment

Early literacy screening assessment	Benchmark metric BOY	25th percentile metric BOY	Benchmark metric MOY	25th percentile metric MOY	Benchmark metric EOY	25th percentile metric EOY	Benchmark metric total	25th percentile metric total
Acadience Reading	22%	23%	24%	27%	18%	25%	21%	25%
DIBELS 8th Edition	29%	16%	26%	17%	18%	18%	24%	17%
mCLASS	35%	23%	34%	25%	26%	26%	32%	25%
mCLASS Lectura	28%	N/A	23%	N/A	27%	N/A	26%	N/A
EarlyBird	38%	N/A	26%	N/A	23%	N/A	30%	N/A
FastBridge aReading	17%	27%	17%	27%	15%	24%	17%	26%
FastBridge earlyReading	16%	30%	24%	40%	20%	34%	20%	35%
i-Ready	22%	31%	49%	29%	56%	29%	43%	30%

Early literacy screening assessment	Benchmark metric BOY	25th percentile metric BOY	Benchmark metric MOY	25th percentile metric MOY	Benchmark metric EOY	25th percentile metric EOY	Benchmark metric total	25th percentile metric total
MAP Reading Fluency	34%	N/A	32%	N/A	30%	N/A	32%	N/A
Star Early Literacy	42%	43%	32%	33%	31%	32%	35%	36%
Star Early Literacy Spanish	67%	68%	68%	69%	67%	68%	67%	68%
Star Reading	40%	40%	31%	32%	31%	32%	34%	35%
Star Reading Spanish	72%	74%	74%	75%	72%	73%	72%	74%
Total	**33%**	**26%**	**33%**	**26%**	**30%**	**26%**	**32%**	**26%**

Note. The EarlyBird, mCLASS Lectura, and MAP Reading Fluency data did not contain percentiles that could be used for the 25th percentile or below analysis. *Source*: Early literacy screening assessment data.

Tables C9–C12 provide the percentage of students significantly below benchmark by grade level, time period, and assessment.

Table C9. Assessment Performance by Time Period and Assessment for Kindergarten

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	24%	20%	11%
DIBELS 8th Edition	38%	25%	13%
mCLASS	40%	35%	24%
mCLASS Lectura	37%	26%	17%
EarlyBird	58%	39%	26%
FastBridge earlyReading	10%	17%	17%
i-Ready	3%	37%	46%
MAP Reading Fluency	32%	26%	17%
Star Early Literacy	42%	24%	22%
Star Early Literacy Spanish	55%	50%	61%
Total	**36%**	**30%**	**22%**

Source: Early literacy screening assessment data.

Table C10. Assessment Performance by Time Period and Assessment for Grade 1

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	22%	29%	23%
DIBELS 8th Edition	30%	25%	13%
mCLASS	34%	33%	21%
mCLASS Lectura	41%	41%	43%
FastBridge earlyReading	19%	24%	20%

Early literacy screening assessment	BOY	MOY	EOY
i-Ready	12%	59%	53%
MAP Reading Fluency	34%	28%	20%
Star Early Literacy	44%	37%	35%
Star Early Literacy Spanish	76%	79%	79%
Total	**32%**	**35%**	**26%**

Source: Early literacy screening assessment data.

Table C11. Assessment Performance by Time Period and Assessment for Grade 2

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	23%	22%	21%
DIBELS 8th Edition	26%	22%	16%
mCLASS	36%	32%	26%
mCLASS Lectura	33%	24%	18%
FastBridge aReading	21%	19%	14%
i-Ready	27%	50%	55%
MAP Reading Fluency	27%	29%	30%
Star Reading	38%	26%	23%
Star Reading Spanish	22%	11%	22%
Total	**32%**	**31%**	**28%**

Source: Early literacy screening assessment data.

Table C12. Assessment Performance by Time Period and Assessment for Grade 3

Early literacy screening assessment	BOY	MOY	EOY
Acadience Reading	14%	13%	12%
DIBELS 8th Edition	22%	24%	20%
mCLASS	31%	32%	27%
mCLASS Lectura	16%	20%	20%
FastBridge aReading	12%	12%	12%
i-Ready	36%	41%	62%
MAP Reading Fluency	39%	42%	45%
Star Reading	39%	30%	32%
Star Reading Spanish	76%	72%	74%
Total	**31%**	**31%**	**32%**

Source: Early literacy screening assessment data.

Table C13 provides the percentage of students at significant risk (when using the 25th percentile or below metric) and the relative risk of being at significant risk at each time period and by student group.

Table C13. Percentage of Students Significantly At Risk and At Relative Risk of Being At Significant Risk At BOY, MOY, and EOY by Student Group (Using the 25th Percentile or Below Metric)

Demographic	At significant risk at BOY	At significant risk at MOY	At significant risk at EOY	Change from BOY to EOY	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Low income	40%	37%	37%	–3%	2.71	2.96	2.80
Non–low income	15%	13%	13%	–2%			
Female	25%	23%	23%	–2%	0.94	0.94	0.93
Male	27%	25%	25%	–2%			
Nonbinary	Sup data	Sup data	Sup data	Sup data	NC	NC	NC
English learner	54%	51%	48%	–6%	2.87	2.95	2.70
Non–English learner	19%	17%	18%	–1%			
Students receiving special education services	49%	48%	50%	1%	2.36	2.66	2.77
Students not receiving special education services	21%	18%	18%	–3%			
White	24%	22%	22%	–2%	0.74	0.70	0.71

Demographic	At significant risk at BOY	At significant risk at MOY	At significant risk at EOY	Change from BOY to EOY	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Hispanic/Latino	43%	41%	40%	–4%	2.25	2.32	2.27
Black	33%	32%	32%	–1%	1.31	1.41	1.41
Asian	12%	12%	12%	–1%	0.46	0.49	0.47
American Indian/Alaska Native	54%	49%	49%	–5%	2.19	2.15	2.19
Native Hawaiian/ Pacific Islander	36%	34%	35%	–1%	NC	NC	NC
Total	**26%**	**24%**	**24%**	**–2%**	**N/A**	**N/A**	**N/A**

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). Risk ratio indicates the likelihood of a student group ever being classified as at risk compared with students not in that group (e.g., students from LI families were 2.71 times more likely than students who were from NLI families to be classified as at significant risk at BOY). Includes only students with three scores. "Sup data" means that data for student groups with fewer than 10 students are not shown in order to protect student privacy. NC means that the percentage was not computed because the group made up 5 percent or less of the sample. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C14 provides the percentage of students at significant risk two or more times (when using the 25th percentile or below metric), by student group.

Table C14. Relative Risk of Being Identified as At Significant Risk Multiple Times by Student Group

Demographic	25th percentile or below two or more times (%)	Relative risk of 25th percentile or below two or more times (*N*)
Low income	38%	2.9
Non–low income	13%	N/A
Female	24%	1.0
Male	25%	N/A
English learner	51%	3.0
Non-English learner	17%	N/A
Students receiving special education services	49%	2.6
Students not receiving special education services	19%	N/A
White	22%	0.7
Non-White	32%	N/A
Hispanic/Latino	41%	2.3
Non-Hispanic/Latino	18%	N/A
Black	32%	1.4
Non-Black	23%	N/A
Asian	12%	0.5
Non-Asian	26%	N/A
American Indian/Alaska Native	50%	2.2

Demographic	25th percentile or below two or more times (%)	Relative risk of 25th percentile or below two or more times (*N*)
Non-American Indian/Alaska Native	23%	N/A
Native Hawaiian/Pacific Islander	35%	NC
Non-Native Hawaiian/Pacific Islander	25%	N/A
Total	**25%**	**N/A**

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). NC means that the percentage was not computed because the group made up 5 percent or less of the sample. *Source*: Early literacy screening assessment data.

Table C15 provides the percentage of students at significant risk two or more times (when using the benchmark level and 25th percentile or below metrics), by assessment.

Table C15. Benchmark Performance by Screening Assessment

Early literacy screening assessment	Significantly below benchmark two or more times	25th percentile or below two or more times
Acadience Reading	21%	25%
DIBELS 8th Edition	22%	15%
mCLASS	31%	24%
mCLASS Lectura	23%	N/A
EarlyBird	22%	N/A
FastBridge aReading	15%	25%
FastBridge earlyReading	19%	33%
i-Ready	43%	27%
MAP Reading Fluency	27%	N/A

Early literacy screening assessment	Significantly below benchmark two or more times	25th percentile or below two or more times
Star Early Literacy	31%	32%
Star Early Literacy Spanish	65%	67%
Star Reading	32%	33%
Star Reading Spanish	70%	71%
Total	**30%**	**25%**

Note. The EarlyBird, mCLASS Lectura, and MAP Reading Fluency data did not contain percentiles that could be used for the 25th percentile or below analysis. *Source*: Early literacy screening assessment data.

Table C16 shows the number of students across overlapping student groups.

Table C16. Intersectional Student Group Sample Size

Demographic	Number	LI	EL	Receiving special education services	LI and EL	LI and receiving special education services
Hispanic and White	26,783	20,447	12,671	5,533	10,687	4,433
Hispanic and Black	5,164	4,293	2,073	1,090	1,865	953
Hispanic and Asian	280	153	78	34	52	24
Hispanic and American Indian/Alaska Native	4,524	3,823	3,751	543	3,225	450
Hispanic and Native Hawaiian/Pacific Islander	338	286	174	64	148	56

Demographic	Number	LI	EL	Receiving special education services	LI and EL	LI and receiving special education services
Hispanic and Two or More Races	3,657	2,832	1,555	779	1,353	646
Not Hispanic and White	62,123	17,244	3,343	12,154	2,432	4,684
Not Hispanic and Black	14,562	10,639	4,020	3,169	3,343	2,342
Not Hispanic and Asian	7,652	2,581	2,891	887	1,443	352
Not Hispanic and American Indian/Alaska Native	326	222	144	56	120	43
Not Hispanic and Native Hawaiian/Pacific Islander	130	80	49	33	39	26
Not Hispanic and Two or More Races	6,075	2,742	342	1,215	226	728

Note. Each racial/ethnic grouping contains students who belong only to the two identified groups (e.g., "Hispanic and White" refers to students who were identified only as Hispanic and White according to their SIMS collection data; "Not Hispanic and White" refers to students who were identified as White but not Hispanic according to their SIMS data). *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Tables C17 and C20 show the percentages of students in the sample for each student group identified and the corresponding significant risk rates.

Table C17. Intersectional Student Group Sample Size and Significant Risk Rate, Percentage of Sample

Demographic	MA	FE	Total
Low income	25%	24%	**49%**
Non–low income	26%	24%	**50%**
English learner	12%	11%	**23%**
Non-English learner	38%	37%	**76%**
Students receiving special education services	13%	7%	**19%**
Students not receiving special education services	38%	42%	**80%**
Low income and English learner	10%	9%	**19%**
Non-low income and English learner	2%	2%	**5%**
Students receiving special education services and English learner	3%	1%	**4%**
Students not receiving special education services and English learner	9%	10%	**19%**
Low income, English learner, and students receiving special education services	2%	1%	**3%**
Non–low income, non–English learner, and students not receiving special education services	18%	20%	**38%**

Source: Early literacy screening assessment data and October and June SIMS collection data.

Table C18. Intersectional Student Group Sample Size and Significant Risk Rate, Percentage Significantly Below Benchmark Two or More Times

Demographic	MA	FE	Total
Low income	44%	43%	**43%**
Non–low income	18%	16%	**17%**
English learner	56%	57%	**56%**
Non-English learner	23%	21%	**22%**
Students receiving special education services	53%	61%	**56%**
Students not receiving special education services	24%	24%	**24%**
Low income and English learner	60%	61%	**60%**
Non-low income and English learner	39%	41%	**40%**
Students receiving special education services and English learner	66%	72%	**68%**
Students not receiving special education services and English learner	52%	55%	**53%**
Low income, English learner, and students receiving special education services	71%	75%	**72%**
Non–low income, non–English learner, and students not receiving special education services	10%	9%	**9%**

Source: Early literacy screening assessment data and October and June SIMS collection data.

Table C19. Intersectional Student Group Sample Size and Significant Risk Rate (by Race/Ethnicity), Percentage of Sample

Demographic	MA	FE	Total
White	38%	36%	**74%**
White, LI	16%	16%	**32%**
White, NLI	21%	20%	**41%**
White, EL	7%	6%	**13%**
White, spec ed	10%	5%	**15%**
White, EL, LI	6%	5%	**11%**
White, spec ed, LI	5%	3%	**8%**
Hispanic	15%	15%	**31%**
Hispanic, LI	12%	12%	**24%**
Hispanic, NLI	3%	3%	**7%**
Hispanic, EL	8%	7%	**15%**
Hispanic, spec ed	4%	2%	**6%**
Hispanic, EL, LI	7%	6%	**13%**
Hispanic, spec ed, LI	3%	2%	**5%**
Black	10%	10%	**19%**
Black, LI	7%	7%	**14%**
Black, NLI	3%	2%	**5%**
Black, EL	3%	2%	**5%**
Black, spec ed	3%	1%	**4%**
Black, EL, LI	2%	2%	**4%**
Black, spec ed, LI	2%	1%	**3%**
Asian	4%	4%	**8%**
Asian, LI	1%	1%	**3%**

Demographic	MA	FE	Total
Asian, NLI	3%	3%	**5%**
Asian, EL	1%	1%	**2%**
Asian, spec ed	<1%	<1%	**<1%**
Asian, EL, LI	<1%	<1%	**1%**
Asian, spec ed, LI	<1%	<1%	**<1%**

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). LI refers to students from low income backgrounds, NLI refers to students not from low income backgrounds, EL refers to English learner students, and "spec ed" refers to students receiving special education services. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C20. Intersectional Student Group Sample Size and Significant Risk Rate (by Race/Ethnicity), Percentage Significantly Below Benchmark Two or More Times

Demographic	MA	FE	Total
White	29%	27%	**28%**
White, LI	43%	42%	**43%**
White, NLI	18%	16%	**17%**
White, EL	57%	58%	**58%**
White, spec ed	52%	61%	**55%**
White, EL, LI	60%	61%	**60%**
White, spec ed, LI	62%	68%	**64%**
Hispanic	47%	46%	**47%**
Hispanic, LI	52%	50%	**51%**
Hispanic, NLI	31%	30%	**31%**
Hispanic, EL	61%	62%	**61%**
Hispanic, spec ed	65%	71%	**67%**

Demographic	MA	FE	Total
Hispanic, EL, LI	63%	63%	**63%**
Hispanic, spec ed, LI	68%	73%	**70%**
Black	39%	36%	**38%**
Black, LI	44%	41%	**43%**
Black, NLI	26%	22%	**24%**
Black, EL	59%	58%	**58%**
Black, spec ed	59%	66%	**61%**
Black, EL, LI	62%	60%	**61%**
Black, spec ed, LI	63%	68%	**65%**
Asian	16%	15%	**16%**
Asian, LI	28%	28%	**28%**
Asian, NLI	11%	9%	**10%**
Asian, EL	29%	32%	**31%**
Asian, spec ed	38%	47%	**41%**
Asian, EL, LI	37%	40%	**39%**
Asian, spec ed, LI	53%	61%	**56%**

Note. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race/ethnicity categories). LI refers to students from low income backgrounds, NLI refers to students not from low income backgrounds, EL refers to English learner students, and "spec ed" refers to students receiving special education services. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Tables C21 and C22 show the percentages of students in the sample for each student group identified and the corresponding significant risk rates (by Hispanic classification and other race).

Table C21. Intersectional Student Group Sample Size and Significant Risk Rate (by Hispanic Classification and Other Race), Percentage of Sample

Demographic	MA	FE	Total
HL and WH	10%	10%	**20%**
HL and WH, LI	8%	8%	**15%**
HL and WH, NLI	2%	2%	**5%**
HL and WH, EL	5%	5%	**10%**
HL and WH, spec ed	3%	1%	**4%**
HL and WH, EL, LI	4%	4%	**8%**
HL and WH, spec ed, LI	2%	1%	**3%**
HL and BAA	2%	2%	**4%**
HL and BAA, LI	2%	2%	**3%**
HL and BAA, NLI	<1%	<1%	**<1%**
HL and BAA, EL	<1%	<1%	**2%**
HL and BAA, spec ed	<1%	<1%	**<1%**
HL and BAA, EL, LI	<1%	<1%	**1%**
HL and BAA, spec ed, LI	<1%	<1%	**<1%**
HL and AIAN	2%	2%	**3%**
HL and AIAN, LI	1%	1%	**3%**
HL and AIAN, NLI	<1%	<1%	**<1%**
HL and AIAN, EL	1%	1%	**3%**
HL and AIAN, spec ed	<1%	<1%	**<1%**
HL and AIAN, EL, LI	1%	1%	**2%**
HL and AIAN, spec ed, LI	<1%	<1%	**<1%**

Demographic	MA	FE	Total
Not HI and BAA	6%	5%	**11%**
Not HI and BAA, LI	4%	4%	**8%**
Not HI and BAA, NLI	2%	1%	**3%**
Not HI and BAA, EL	2%	1%	**3%**
Not HI and BAA, spec ed	2%	<1%	**2%**
Not HI and BAA, EL, LI	1%	1%	**3%**
Not HI and BAA, spec ed, LI	1%	<1%	**2%**
Not HL and AS	3%	3%	**6%**
Not HL and AS, LI	<1%	<1%	**2%**
Not HL and AS, NLI	2%	2%	**4%**
Not HL and AS, EL	1%	1%	**2%**
Not HL and AS, spec ed	<1%	<1%	**<1%**
Not HL and AS, EL, LI	<1%	<1%	**1%**
Not HL and AS, spec ed, LI	<1%	<1%	**<1%**

Note. HL refers to Hispanic or Latino students, WH refers to White students, BAA refers to Black or African American students, AS refers to Asian students, and AIAN refers to American Indian/Alaska Native students. LI refers to students from low income backgrounds, NLI refers to students not from low income backgrounds, EL refers to English learner students, and "spec ed" refers to students receiving special education services. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Table C22. Intersectional Student Group Sample Size and Significant Risk Rate (by Hispanic Classification and Other Race), Percentage Significantly Below Benchmark Two or More Times

Demographic	MA	FE	Total
HL and WH	45%	45%	**45%**
HL and WH, LI	50%	50%	**50%**
HL and WH, NLI	30%	30%	**30%**
HL and WH, EL	59%	61%	**60%**
HL and WH, spec ed	64%	71%	**66%**
HL and WH, EL, LI	62%	63%	**62%**
HL and WH, spec ed, LI	67%	73%	**69%**
HL and BAA	46%	42%	**44%**
HL and BAA, LI	49%	45%	**47%**
HL and BAA, NLI	32%	25%	**29%**
HL and BAA, EL	63%	59%	**61%**
HL and BAA, spec ed	67%	65%	**66%**
HL and BAA, EL, LI	65%	60%	**63%**
HL and BAA, spec ed, LI	68%	67%	**68%**
HL and AIAN	61%	59%	**60%**
HL and AIAN, LI	63%	61%	**62%**
HL and AIAN, NLI	52%	51%	**51%**
HL and AIAN, EL	67%	66%	**67%**
HL and AIAN, spec ed	75%	82%	**77%**
HL and AIAN, EL, LI	68%	68%	**68%**
HL and AIAN, spec ed, LI	77%	83%	**79%**
Not HI and BAA	39%	37%	**38%**

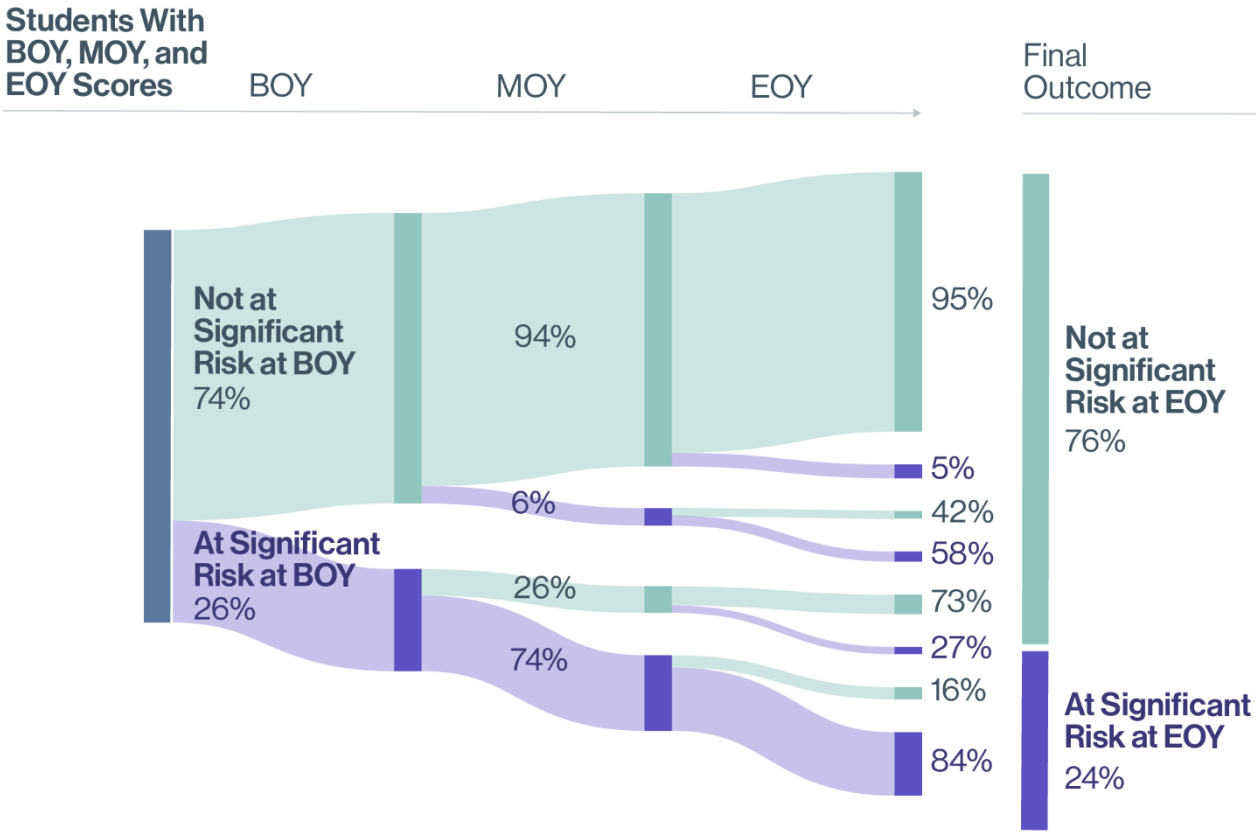
Demographic	MA	FE	Total
Not HI and BAA, LI	43%	41%	**42%**
Not HI and BAA, NLI	27%	24%	**26%**
Not HI and BAA, EL	57%	57%	**57%**
Not HI and BAA, spec ed	57%	64%	**60%**
Not HI and BAA, EL, LI	59%	60%	**60%**
Not HI and BAA, spec ed, LI	61%	66%	**63%**
Not HL and AS	16%	15%	**16%**
Not HL and AS, LI	27%	26%	**27%**
Not HL and AS, NLI	11%	9%	**10%**
Not HL and AS, EL	29%	30%	**30%**
Not HL and AS, spec ed	38%	44%	**39%**
Not HL and AS, EL, LI	36%	38%	**37%**
Not HL and AS, spec ed, LI	53%	57%	**54%**

Note. HL refers to Hispanic or Latino students, WH refers to White students, BAA refers to Black or African American students, AS refers to Asian students, and AIAN refers to American Indian/Alaska Native students. LI refers to students from low income backgrounds, NLI refers to students not from low income backgrounds, EL refers to English learner students, and "spec ed" refers to students receiving special education services. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Appendix D. Student Progress

Figure D1 shows the progression of students at the 25th percentile or below across the school year.

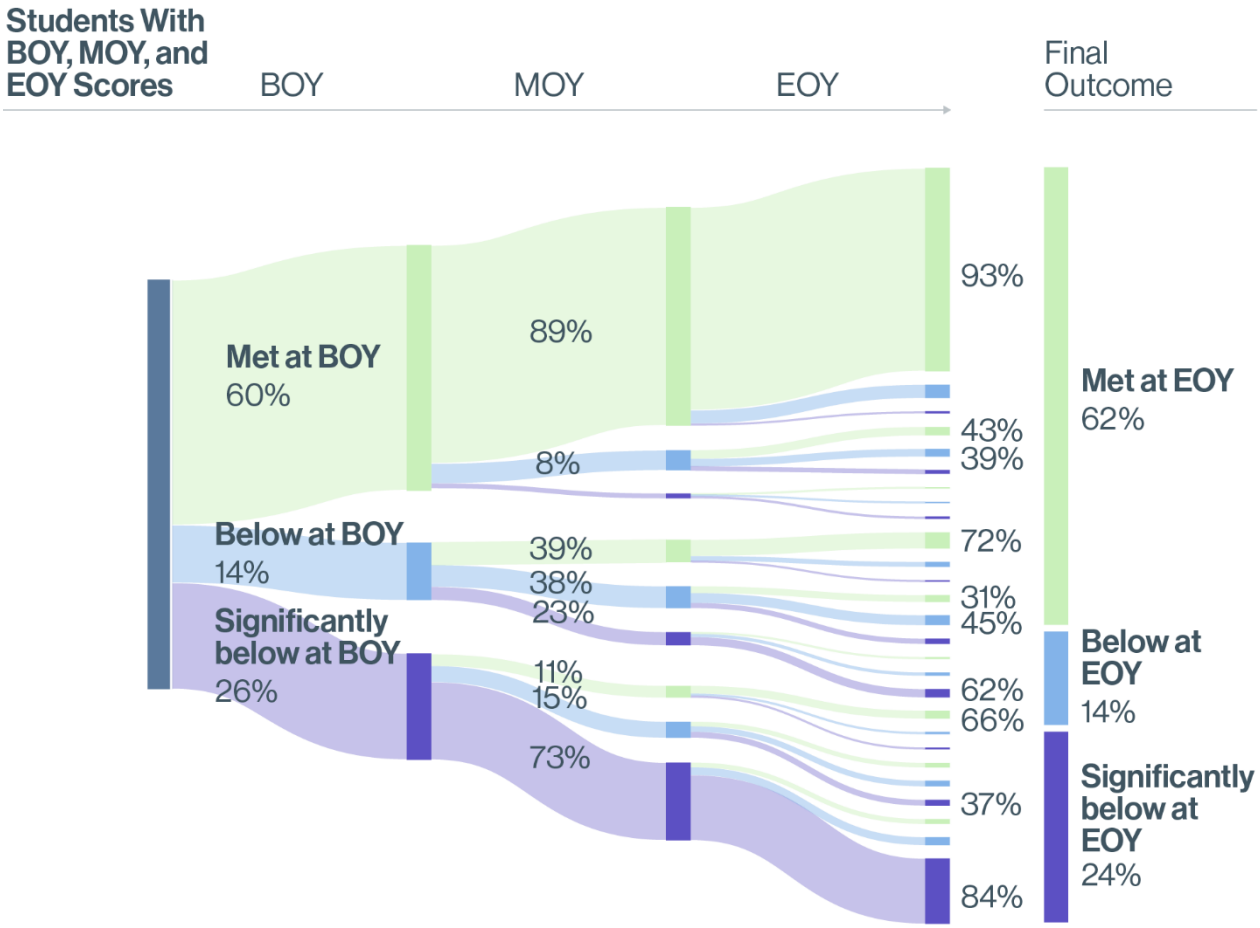
Figure D1. Progression of Students At 25th Percentile or Below Across School Year



Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period on the same assessment were included in the figure. See Figure D1 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

Figure D2 shows the progression of students above the 40th percentile, at or below the 40th percentile, and at or below the 25th percentile across the school year.

Figure D2. Progression of Students Across School Year



Note. Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period on the same assessment were included in the figure. Students met expectations if they scored above the 40th percentile on their early literacy assessment; they were below expectations if they scored at or below the 40th percentile but above the 25th percentile on their assessment; and they were significantly below expectations if they scored at or below the 25th percentile on their assessment. See Figure D2 in Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data.

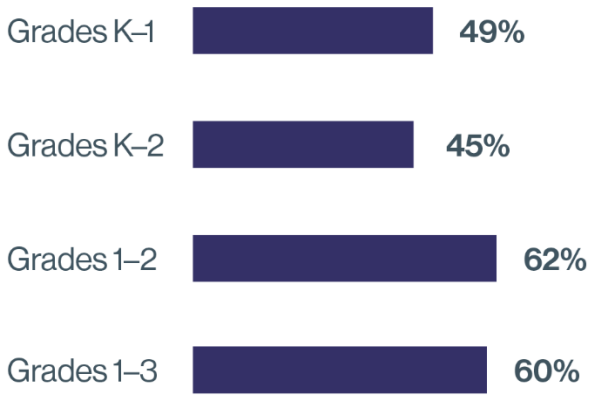
Figure D3 shows the percentage of students who were at significant risk (according to the 25th percentile or below metric) at EOY of one grade level and at EOY of the subsequent grade level.

Figure D3. Progression of Students At 25th Percentile or Below Across School Years

Note. Some students may appear multiple times if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

Figure D4 shows the percentage of students who were at significant risk (according to the 25th percentile or below metric) at EOY of one grade level and at EOY 1 year later and 2 years later.

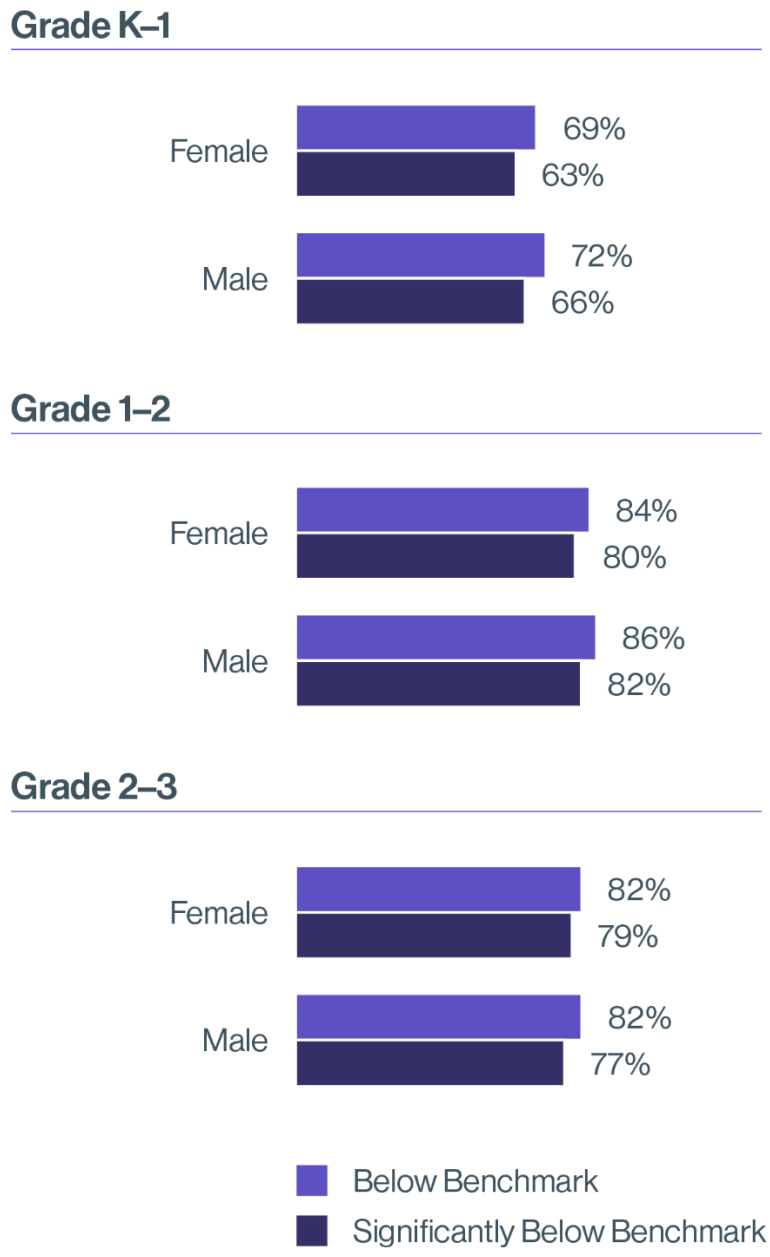
Figure D4. Progression of Students At 25th Percentile or Below Across School Years



Note. Some students may appear multiple times if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

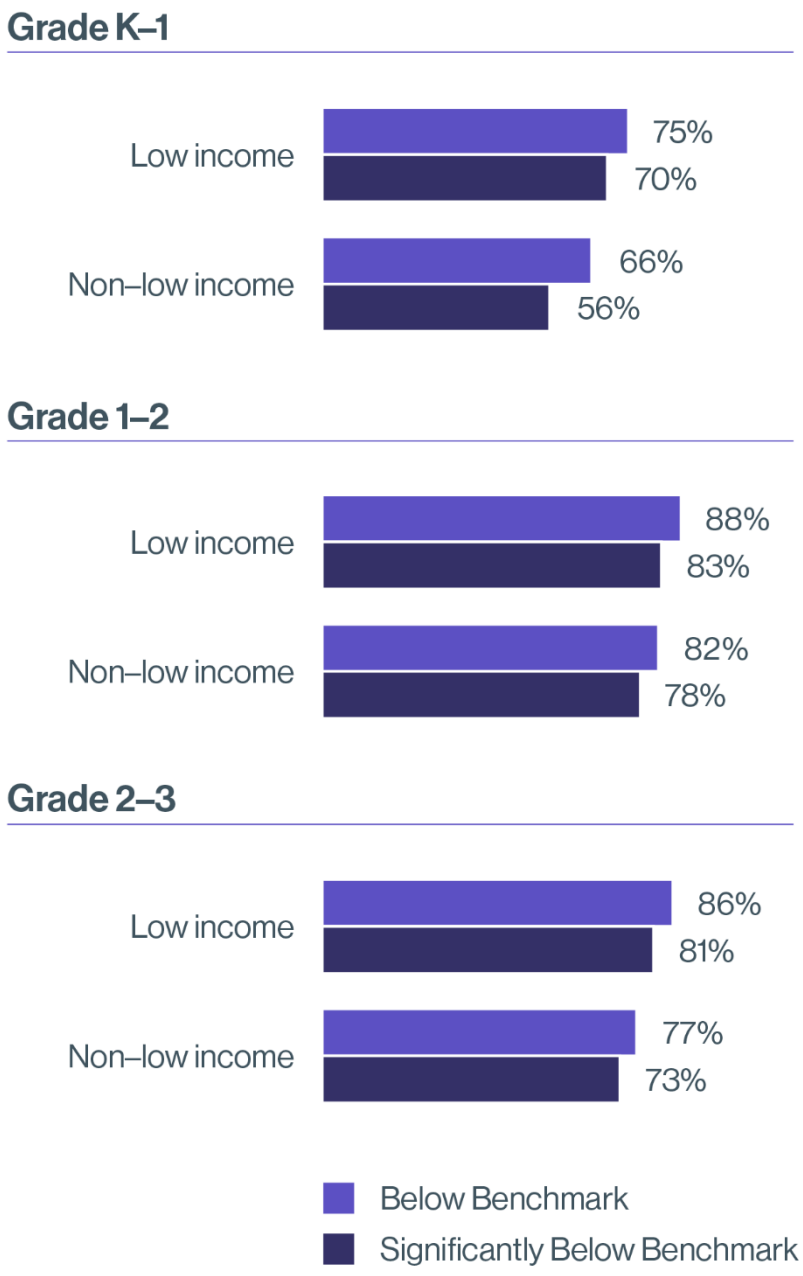
Figures D5–D14 show the percentage of students who were below or significantly below benchmark in one grade level and in the subsequent grade level, by student group.

Figure D5. Progression of Female and Male Students Below or Significantly Below Benchmark Across School Years



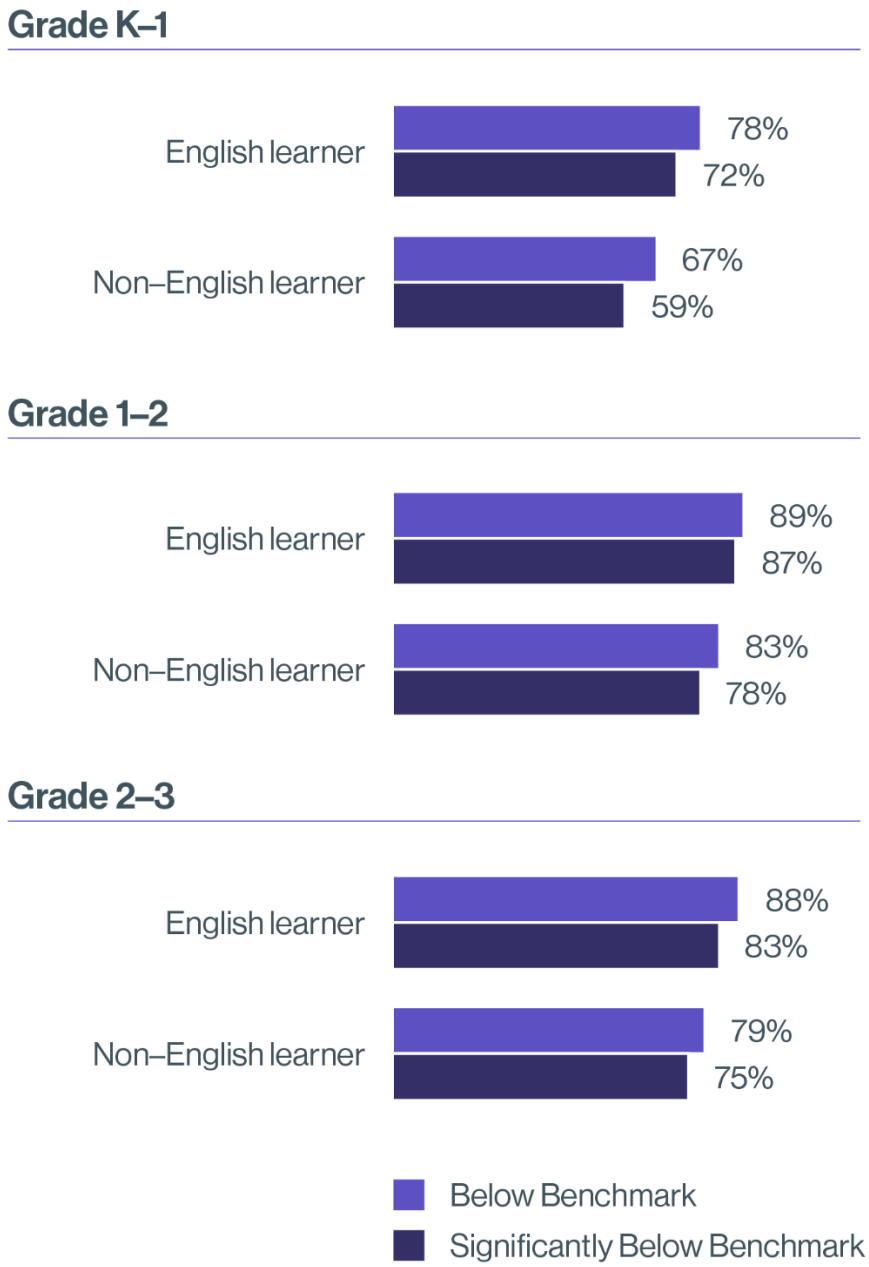
Note. Some students may appear multiple times if they were administered multiple screening assessments. For figure data, see Figure D5 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D6. Progression of Students From Low Income Backgrounds and Students Not From Low Income Backgrounds Below or Significantly Below Benchmark Across School Years



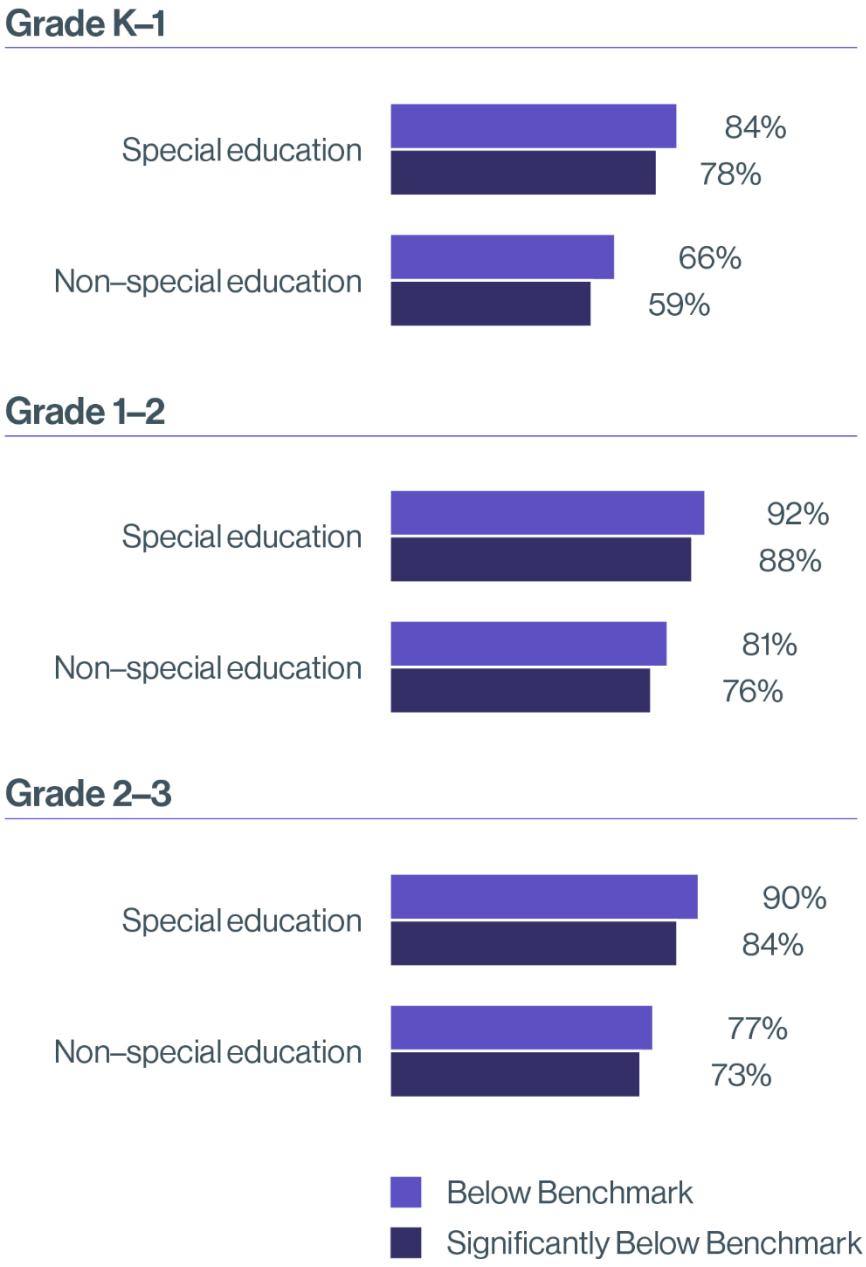
Note. Some students may appear multiple times if they were administered multiple screening assessments. For figure data, see Figure D6 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D7. Progression of English Learner and Non–English Learner Students Below or Significantly Below Benchmark Across School Years



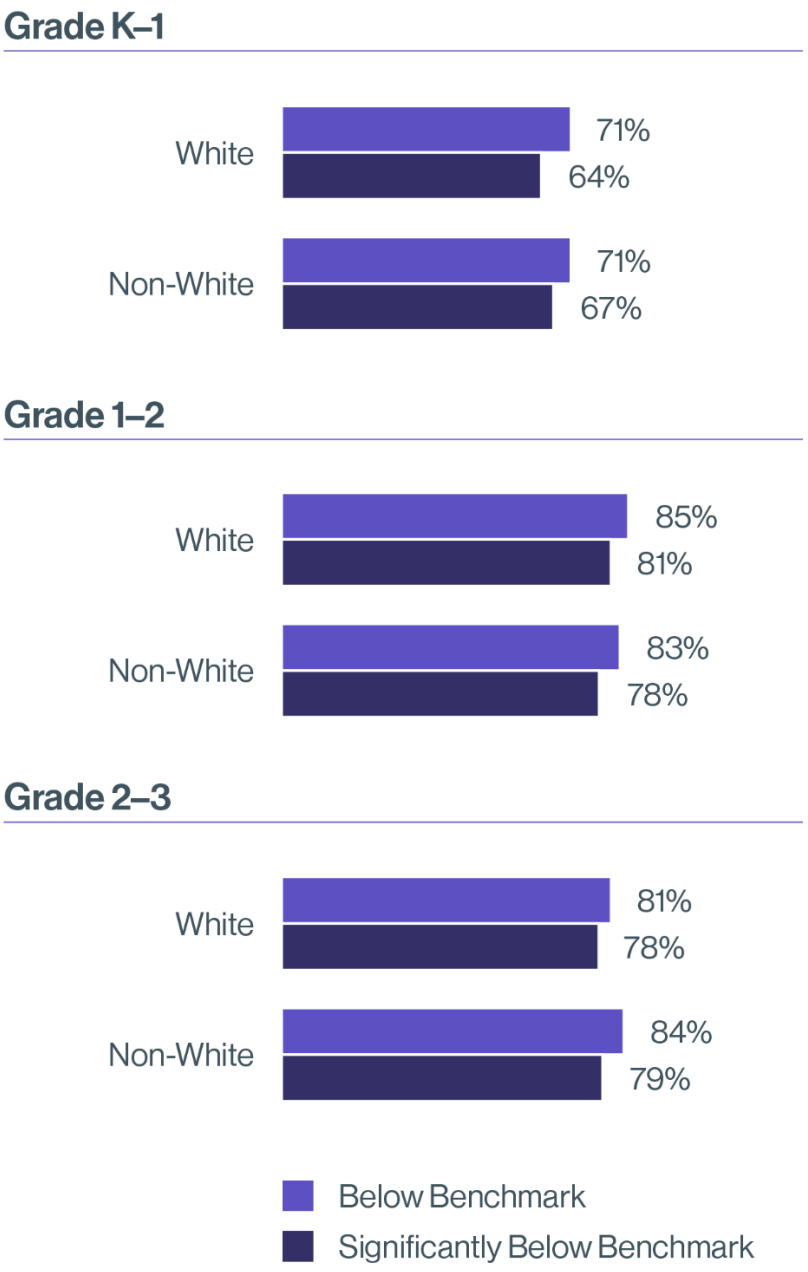
Note. Some students may appear multiple times if they were administered multiple screening assessments. For figure data, see Figure D7 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D8. Progression of Students Receiving and Not Receiving Special Education Services Below or Significantly Below Benchmark Across School Years



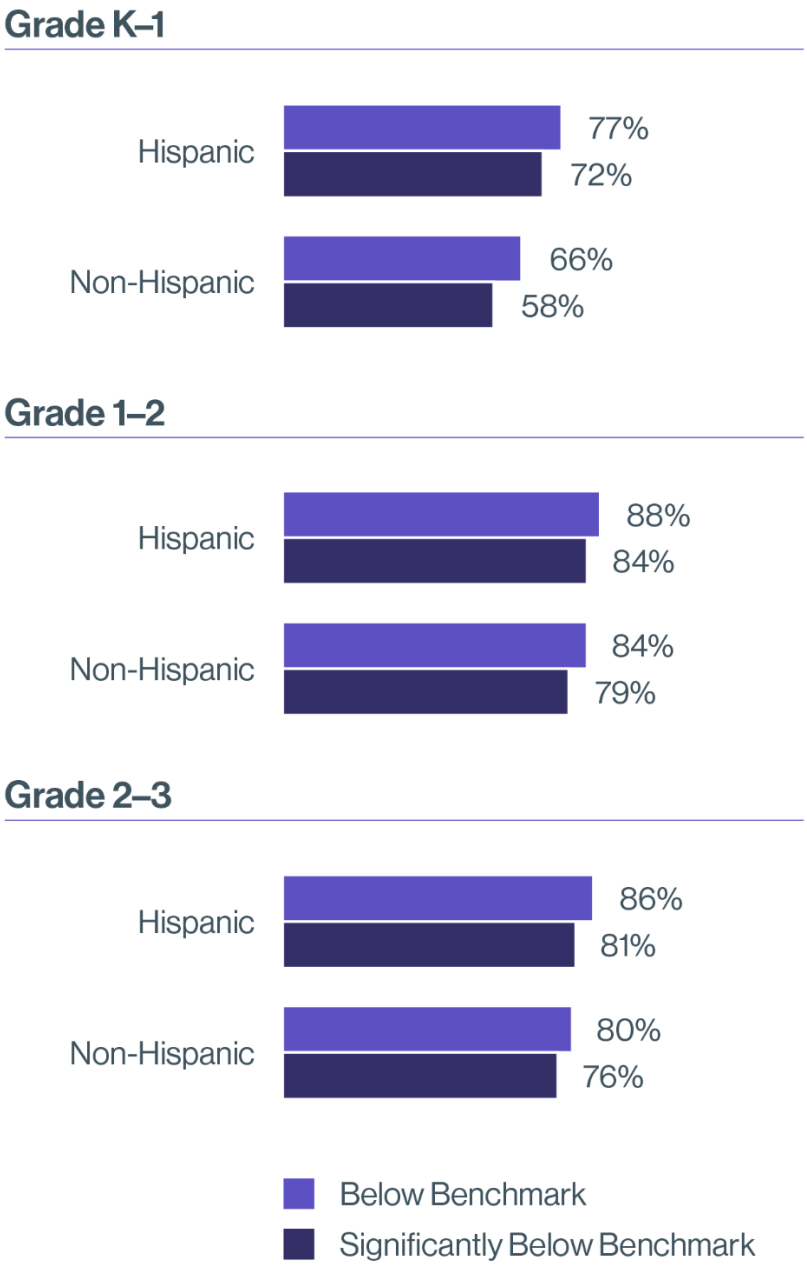
Note. Some students may appear multiple times if they were administered multiple screening assessments. For figure data, see Figure D8 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D9. Progression of White and Non-White Students Below or Significantly Below Benchmark Across School Years



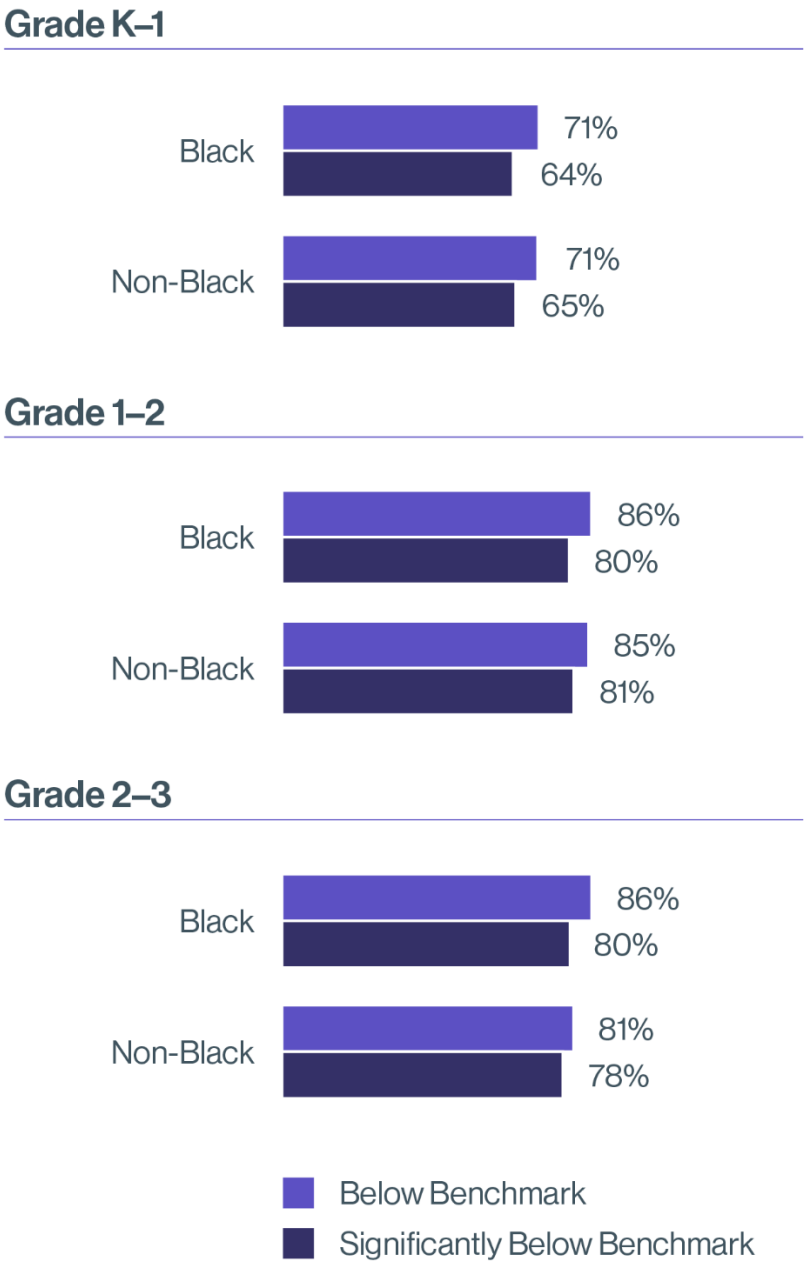
Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D9 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D10. Progression of Hispanic/Latino and Non-Hispanic/Latino Students Below or Significantly Below Benchmark Across School Years



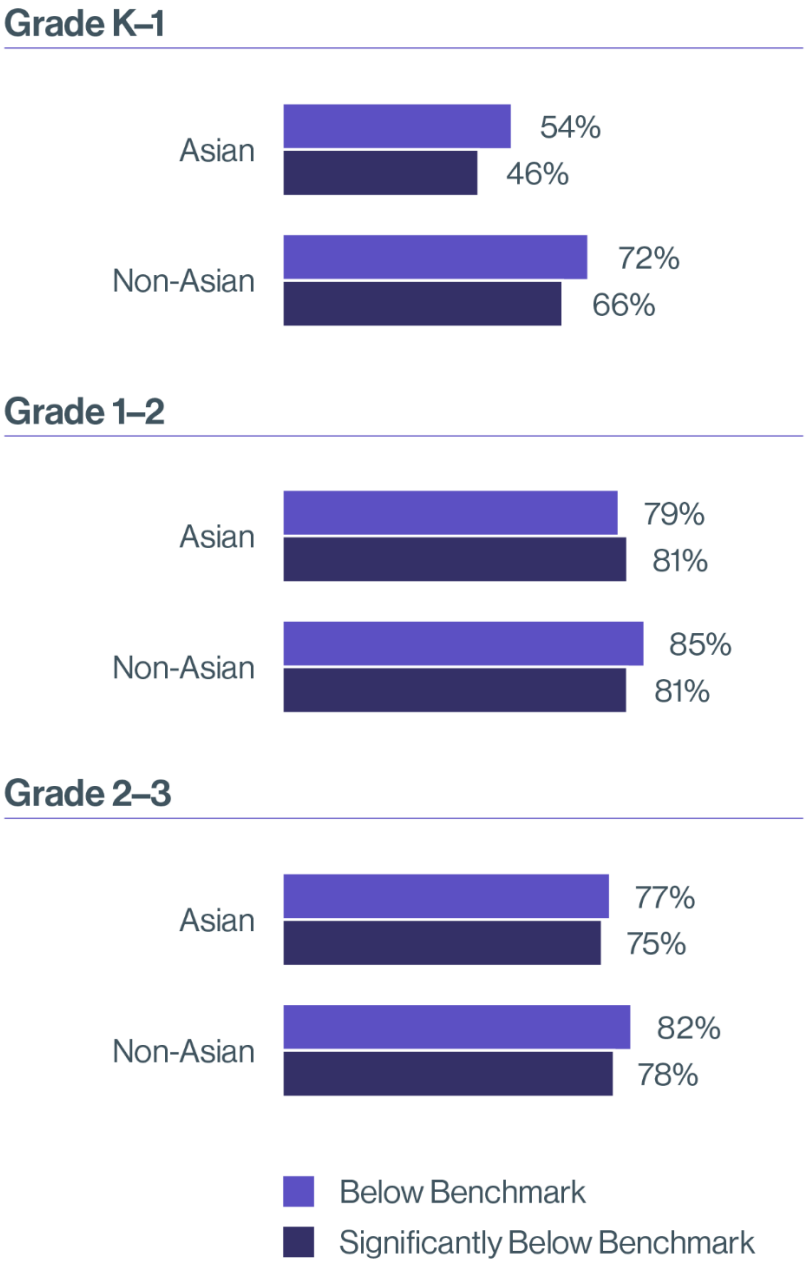
Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D10 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D11. Progression of Black and Non-Black Students Below or Significantly Below Benchmark Across School Years



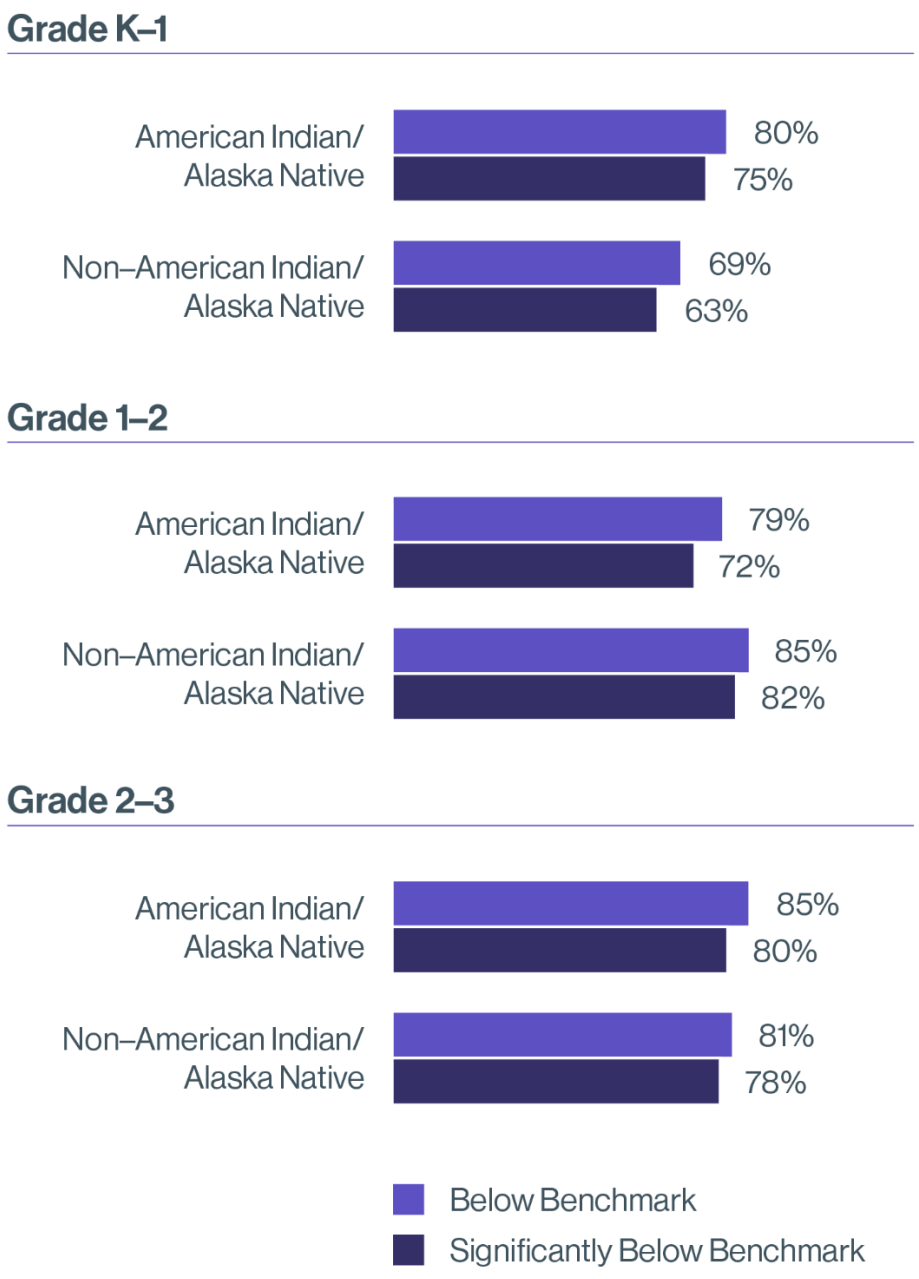
Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D11 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D12. Progression of Asian and Non-Asian Students Below or Significantly Below Benchmark Across School Years



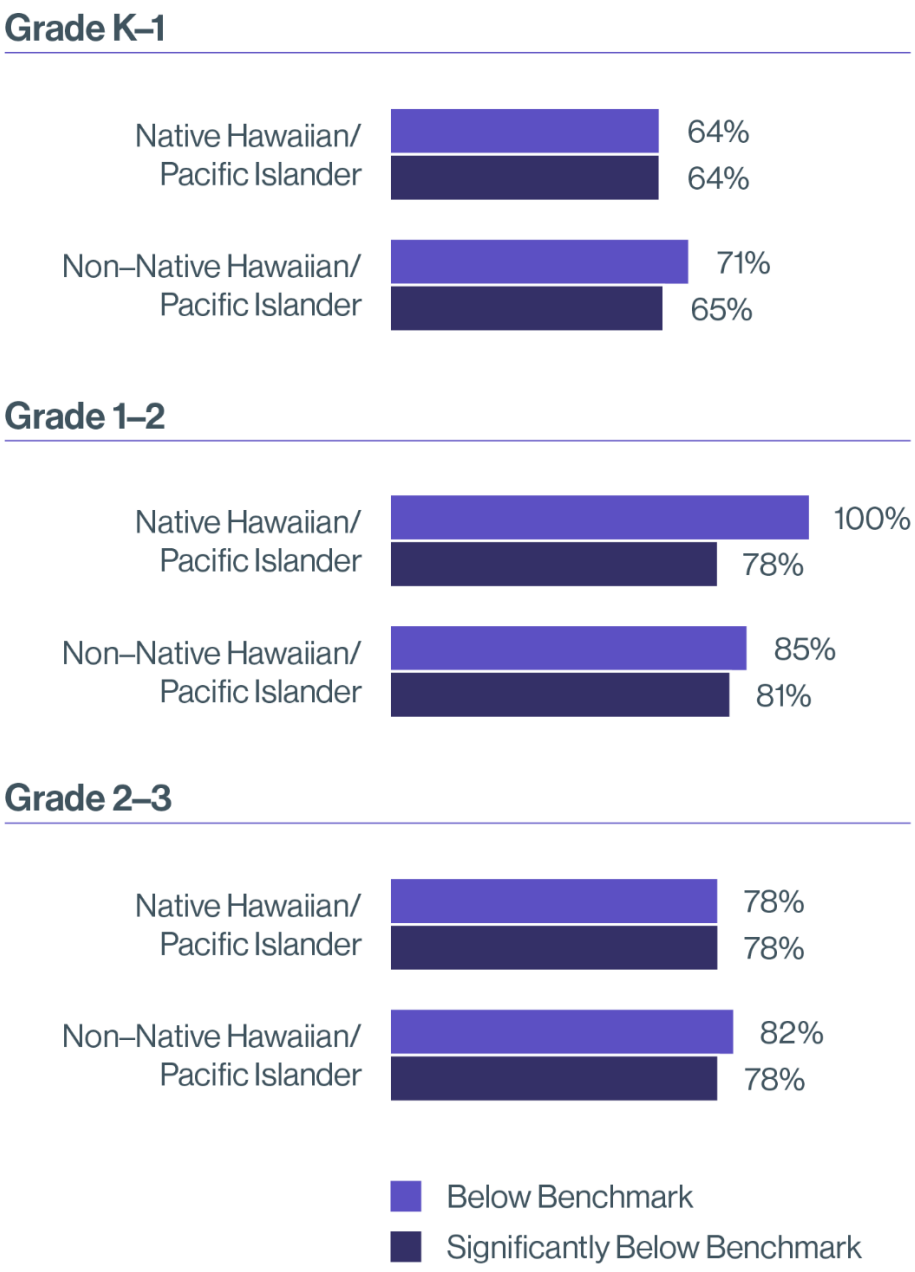
Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D12 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D13. Progression of American Indian/Alaska Native and Non–American Indian/Alaska Native Students Below or Significantly Below Benchmark Across School Years



Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D13 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure D14. Progression of Native Hawaiian/Pacific Islander and Non–Native Hawaiian/Pacific Islander Students Below or Significantly Below Benchmark Across School Years



Note. Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). For figure data, see Figure D14 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Tables D1–D3 shows the progression of students from EOY of one grade level to BOY of the subsequent grade level.

Table D1. Percentage of Students At Significant Risk At End of Kindergarten and Beginning of Grade 1 by Assessment

Early literacy screening assessment	Number	Kindergarten EOY significantly below	Grade 1 BOY significantly below	Kindergarten EOY 25th percentile	Grade 1 BOY 25th percentile
Acadience Reading	454	5%	17%	21%	17%
DIBELS 8th Edition	3,450	16%	27%	16%	15%
mCLASS	2,713	22%	28%	23%	16%
FastBridge earlyReading	102	21%	11%	40%	32%
i-Ready	1,439	51%	11%	32%	32%
Star Early Literacy	3,902	26%	40%	27%	42%
Total	**12,060**	**24%**	**29%**	**24%**	**26%**

Note. Some students may appear multiple times if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

Table D2. Percentage of Students At Significant Risk At End of Grade 1 and Beginning of Grade 2 by Assessment

Early literacy screening assessment	Number	Grade 1 EOY significantly below	Grade 2 BOY significantly below	Grade 1 EOY 25th percentile	Grade 2 BOY 25th percentile
Acadience Reading	459	18%	15%	20%	19%
DIBELS 8th Edition	2,756	14%	25%	16%	14%
mCLASS	2,824	17%	26%	19%	16%
i-Ready	1,718	56%	29%	30%	29%
Total	**7,757**	**25%**	**26%**	**21%**	**18%**

Note. Some students may appear multiple times if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

Table D3. Percentage of Students At Significant Risk At End of Grade 2 and Beginning of Grade 3 by Assessment

Early literacy screening assessment	Number	Grade 2 EOY significantly below	Grade 3 BOY significantly below	Grade 2 EOY 25th percentile	Grade 3 BOY 25th percentile
Acadience Reading	408	14%	15%	19%	15%
DIBELS 8th Edition	2,531	19%	19%	18%	16%
mCLASS	2,656	20%	20%	18%	16%
i-Ready	1,796	58%	35%	28%	32%
Star Reading	3,785	28%	34%	28%	35%
Total	**11,176**	**28%**	**27%**	**23%**	**25%**

Note. Some students may appear multiple times if they were administered multiple screening assessments. *Source*: Early literacy screening assessment data.

Appendix E. School Analysis Details

Multilevel Logistic Regression Model Description

In this section, we describe the results of a statistical model that incorporated several student- and school-level variables to predict students' likelihood of being significantly below benchmark multiple times during the 2023/24 school year considering those variables.

Students are clustered within schools, which may vary in their effects on student performance. To account for this fact, a multilevel logistic regression model, in which students are nested within schools was used to examine the impact of student- and school-level factors on the risk of being identified as significantly below benchmark more than once during the year. The model is shown below.

$$\text{logit}\left(\pi_{ij}\right) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} + \gamma_{0j}$$

This equation represents a two-level multilevel logistic regression model in which

- $\text{logit}\left(\pi_{ij}\right)$is the natural logarithm of the odds of being significantly below the benchmark two or more times for student *i* in school *j*;
- $\beta_0$is the intercept, representing the log odds of being at risk when all predictor variables and random effects are zero;
- β_1, β_2, ..., β_p are the fixed effect coefficients of predictor variables X_{ij1}, X_{ij2}, ..., *Xijp*; and
- γ_{0j} are school-specific effects.

The model controls for student- and school-level predictor variables associated with students' likelihood of being significantly below benchmark more than once during the school year. Note that multiple models were evaluated before the final model was selected. For example, a variable examining the type of English learner program students attended was included in the model and found to not be statistically significant, as were variables examining the amount of student mobility and teacher experience within schools. These variables were therefore removed from the final model. Student-level variables retained in the final model include

- gender,
- race/ethnicity,
- grade-level,
- screening assessment,
- low income status,
- English learner status,

- indicator of whether the student received special education services, and
- early childhood program experience (ECP).

School-level variables include

- percentage of students from low income backgrounds,
- percentage of EL students,
- student mobility rate (i.e., student stability), and
- student attendance rate.

The final model results are shown in Table E1.

Table E1. Multilevel Logistic Regression Results That Predict the Likelihood of Being Significantly Below Benchmark More Than Once During the 2023/24 School Year

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
(Intercept)	–2.08	0.12	–16.81	0.00	0.12	0.10	0.16
Grade 1	–0.01	0.03	–0.31	0.75	0.99	0.93	1.05
Grade 2	0.10	0.03	3.54	0.00	1.11	1.05	1.17
Kindergarten	–0.10	0.03	–3.21	0.00	0.90	0.85	0.96
Male	0.01	0.03	0.33	0.74	1.01	0.95	1.07
Asian	–0.52	0.07	–7.64	0.00	0.60	0.52	0.68
Black	0.32	0.04	7.49	0.00	1.38	1.27	1.50
Hispanic	0.51	0.03	15.36	0.00	1.66	1.56	1.78
Other Race	0.05	0.06	0.84	0.40	1.05	0.93	1.19
Formal ECP	–0.19	0.02	–10.52	0.00	0.82	0.79	0.85
Informal ECP	–0.08	0.08	–1.07	0.29	0.92	0.79	1.07
English learner	1.10	0.05	22.21	0.00	3.00	2.72	3.30
Has IEP	2.25	0.04	61.33	0.00	9.49	8.83	10.20
DIBELS 8th Edition	–0.08	0.13	–0.61	0.54	0.93	0.72	1.19
FastBridge aReading	–0.67	0.20	–3.34	0.00	0.51	0.34	0.76

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
FastBridge earlyReading	–0.31	0.18	–1.68	0.09	0.74	0.51	1.05
i-Ready	0.99	0.14	7.32	0.00	2.70	2.07	3.52
MAP Reading Fluency	–0.98	0.14	–6.88	0.00	0.38	0.29	0.50
mCLASS	–0.04	0.12	–0.30	0.76	0.96	0.76	1.23
Star Early Literacy	0.18	0.13	1.41	0.16	1.20	0.93	1.56
Star Reading	0.03	0.13	0.22	0.83	1.03	0.79	1.33
Low income	0.60	0.02	30.47	0.00	1.82	1.75	1.89
School low income %	0.32	0.04	7.86	0.00	1.37	1.27	1.48
School stability %	0.00	0.03	0.05	0.96	1.00	0.94	1.06
School EL %	0.06	0.03	1.89	0.06	1.06	1.00	1.12
School attendance %	–0.05	0.03	–1.89	0.06	0.95	0.90	1.00
Male:has IEP	–0.35	0.04	–8.78	0.00	0.71	0.65	0.76
Asian:has IEP	–0.34	0.11	–3.25	0.00	0.71	0.58	0.87
Black:has IEP	–0.49	0.06	–8.01	0.00	0.61	0.54	0.69
Hispanic:has IEP	–0.60	0.04	–13.35	0.00	0.55	0.50	0.60
Other Race:has IEP	–0.11	0.09	–1.19	0.24	0.90	0.75	1.07
Grade 1:English learner	0.02	0.05	0.33	0.74	1.02	0.91	1.13
Grade 2:English learner	–0.05	0.06	–0.95	0.34	0.95	0.85	1.06
Kindergarten:English learner	0.23	0.06	4.07	0.00	1.26	1.13	1.41
Male:Asian	–0.05	0.09	–0.53	0.59	0.95	0.80	1.14
Male:Black	–0.03	0.05	–0.60	0.55	0.97	0.87	1.08

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
Male:Hispanic	–0.06	0.04	–1.53	0.13	0.94	0.87	1.02
Male:Other Race	–0.01	0.08	–0.09	0.93	0.99	0.84	1.17
Male:English learner	–0.14	0.04	–3.33	0.00	0.87	0.80	0.95

Note. The estimate column represents the fixed effect coefficients from the linear mixed model. Each estimate shows the expected change in the outcome variable when each term changes by one unit while holding all other variables constant. For continuous terms, the estimate column represents a linear slope; for categorical terms, the estimate represents a mean level change in the outcome compared with the reference group. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

The reference group in the model is grade 3, White, female students who are not English learner (EL) students, are not from low income backgrounds, do not receive special education services (do not have an IEP), do not attend an early childhood program, are from average schools (i.e., school characteristics and random effects equal to 0), and took Acadience. Such students had a 12 percent likelihood of being significantly below benchmark multiple times during the 2023/24 school year.

A handful of student-level predictors showed strong associations with the likelihood of being significantly below benchmark multiple times during the 2023/24 school year. Many student-level attributes were positively associated, such as having Black and Hispanic ethnoracial backgrounds, having English learner status, receiving special education services, or coming from low income backgrounds, increasing the risk by 38 percent, 66 percent, 200 percent, 849 percent, and 82 percent, respectively. Two student-level attributes, Asian students and those involved with formal early childhood programs, were associated with a reduced likelihood of risk identification, reducing the probability by 40 percent and 18 percent, respectively. A handful of interactions were associated with students who had an IEP and were English Learners. For Asian, Black, and Hispanic students who had an IEP, the risk of being identified two or more times decreased by 29 percent, 39 percent, and 45 percent, respectively (Figure E3). For English learners, those in kindergarten were 26 percent more likely to be identified two or more times, whereas males were about 13 percent less likely to be identified two or more times (Figure E2).

There was evidence of a different likelihood of being identified more than once by the screening assessment the student took. Two, FastBridge aReading and Map Reading Fluency, were negatively associated, whereas i-Ready was positively associated when compared with the Acadience Reading reference group. Students who took FastBridge aReading and Map Reading Fluency were 49 percent and 62 percent, respectively, less likely to be identified more than once compared with those who took Acadience Reading. In contrast, students who took

i-Ready were 170 percent more likely to be identified more than once compared with those who took Acadience Reading. No other assessment showed differences compared with the Acadience Reading reference group, and there were no reliable estimates of interactions with the screening assessment that the student took (Figure E1).

Finally, one school-level effect was identified as being associated with being significantly below the benchmark multiple times during the 2023/24 school year (Figure E4). Schools with a larger percentage of students from low income backgrounds were associated with an increased likelihood of students being identified. A one standard deviation increase in the percentage of students from low income backgrounds was associated with an increased likelihood of 37 percent.

Figure E1. Model Fixed Effect Results for Terms Without Interaction Effects Showing the Model-Implied Probabilities of a Student Being Identified Two or More Times During the School Year

Note. For figure data, see Figure E1 in Appendix F. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

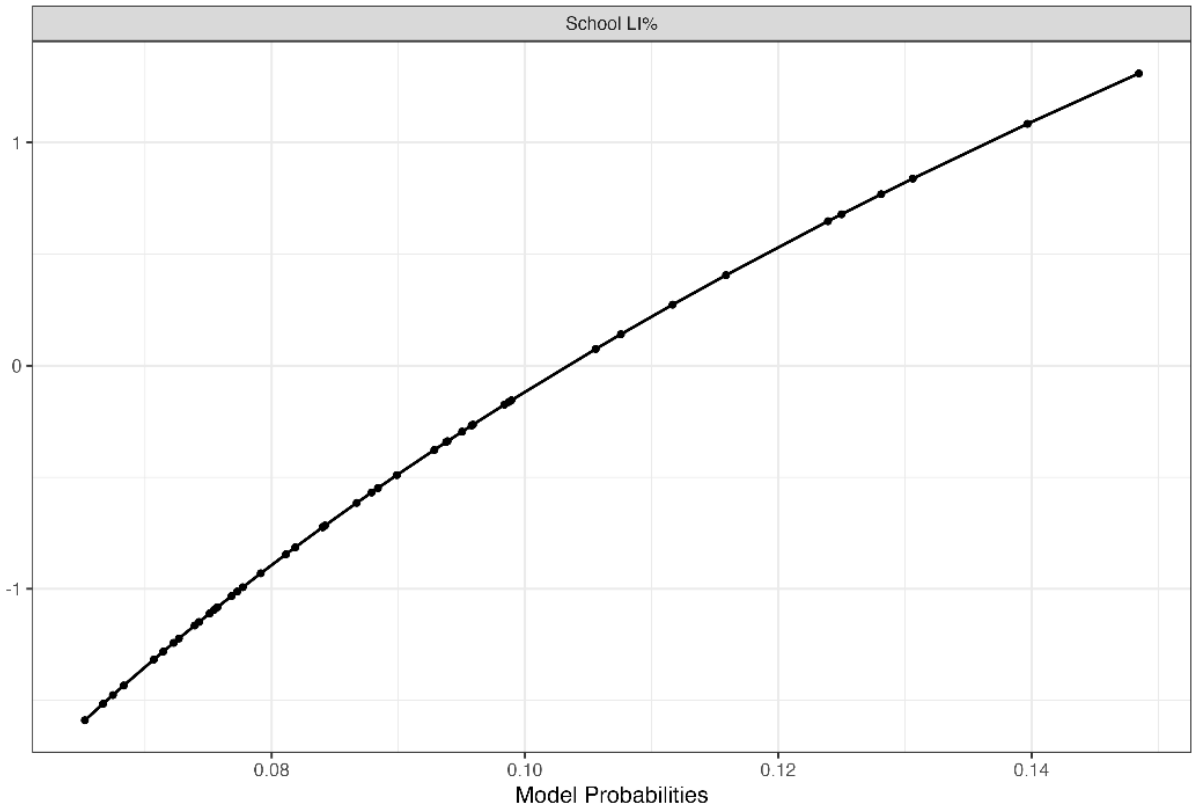
Figure E2. Model-Implied Probabilities of Being Identified Two or More Times During the School Year for the Interaction Between Gender and EL Status

Source: Early literacy screening assessment data and October and June SIMS collection data.

Figure E3. Model Implied Probabilities of Being Identified Two or More Times During the School Year for the Interaction Between Gender, Race/ethnicity, and IEP Status

Source: Early literacy screening assessment data and October and June SIMS collection data.

Figure E4. Model-Implied Probabilities of Being Identified Two or More Times for School Low Income Percentage



Note. *Y*-axis is in standard deviation units. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Neighborhood Segregation Analysis Description

The impact of neighborhood segregation was explored by use of the same modeling approach (Table E2). Additional school-level attributes were added that represented data on neighborhood segregation. The neighborhood segregation attributes were collinear, meaning they had high correlations with one another and represented similar measures. Therefore, the number of characteristics in the model was reduced to estimate the model. No attributes were identified as being important in helping predict the likelihood of a student being identified more than once during the school year. The White to Minority segregation measure showed the largest descriptive association, an estimated odds increase of 13 percent for a one standard deviation increase in the segregation measure.

Table E2. Multilevel Logistic Regression Results That Predict the Likelihood of Being Significantly Below Benchmark More Than Once During the 2023/24 School Year With Neighborhood Segregation

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
(Intercept)	–2.09	0.12	–16.84	0.00	0.12	0.10	0.16
Grade 1	–0.01	0.03	–0.36	0.72	0.99	0.93	1.05
Grade 2	0.10	0.03	3.53	0.00	1.10	1.05	1.17
Kindergarten	–0.10	0.03	–3.23	0.00	0.90	0.85	0.96
Male	0.01	0.03	0.34	0.74	1.01	0.95	1.07
Asian	–0.52	0.07	–7.60	0.00	0.60	0.52	0.68
Black	0.33	0.04	7.62	0.00	1.39	1.27	1.51
Hispanic	0.51	0.03	15.49	0.00	1.67	1.57	1.79
Other	0.05	0.06	0.88	0.38	1.06	0.94	1.19
Formal ECP	–0.19	0.02	–10.50	0.00	0.82	0.80	0.85
Informal ECP	–0.08	0.08	–1.06	0.29	0.92	0.79	1.07
English learner	1.10	0.05	22.20	0.00	3.00	2.72	3.30
Has IEP	2.25	0.04	61.31	0.00	9.49	8.83	10.20
DIBELS 8th Edition	–0.07	0.13	–0.57	0.57	0.93	0.72	1.19
FastBridge aReading	–0.74	0.20	–3.65	0.00	0.47	0.32	0.71
FastBridge earlyReading	–0.38	0.19	–2.05	0.04	0.68	0.47	0.98
i-Ready	0.98	0.14	7.27	0.00	2.67	2.05	3.49
MAP Reading Fluency	–1.00	0.15	–6.82	0.00	0.37	0.28	0.49
mCLASS	–0.02	0.12	–0.13	0.89	0.98	0.77	1.26
Star Early Literacy	0.18	0.13	1.39	0.17	1.20	0.93	1.55

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
Star Reading	0.02	0.13	0.17	0.86	1.02	0.79	1.33
Low income	0.60	0.02	30.42	0.00	1.82	1.75	1.89
School low income %	0.34	0.06	5.65	0.00	1.40	1.25	1.58
School stability %	0.00	0.03	–0.12	0.91	1.00	0.94	1.06
School EL %	0.08	0.03	2.58	0.01	1.09	1.02	1.16
School attendance %	–0.06	0.03	–1.99	0.05	0.95	0.89	1.00
Isolation—White to Minority	0.12	0.05	2.26	0.02	1.13	1.02	1.25
Isolation—FRL to NFRL	0.06	0.07	0.83	0.41	1.06	0.93	1.20
Male:has IEP	–0.35	0.04	–8.78	0.00	0.71	0.65	0.76
Asian:has IEP	–0.34	0.11	–3.24	0.00	0.71	0.58	0.87
Black:has IEP	–0.49	0.06	–8.02	0.00	0.61	0.54	0.69
Hispanic:has IEP	–0.60	0.04	–13.34	0.00	0.55	0.50	0.60
Other Race:has IEP	–0.11	0.09	–1.20	0.23	0.90	0.75	1.07
Grade 1:English learner	0.02	0.05	0.33	0.74	1.02	0.91	1.13
Grade 2:English learner	–0.05	0.06	–0.96	0.34	0.95	0.85	1.06
Kindergarten: English learner	0.23	0.06	4.05	0.00	1.26	1.13	1.41
Male:Asian	–0.05	0.09	–0.54	0.59	0.95	0.80	1.14
Male:Black	–0.03	0.05	–0.60	0.55	0.97	0.87	1.08

Term	Estimate	SE	Statistic	p.value	Odds	odd_ll	odd_ul
Male:Hispanic	–0.06	0.04	–1.53	0.13	0.94	0.86	1.02
Male:Other Race	–0.01	0.08	–0.10	0.92	0.99	0.84	1.17
Male:English learner	–0.14	0.04	–3.33	0.00	0.87	0.80	0.94

Source: Early literacy screening assessment data and October and June SIMS collection data.

"Outperforming" School Selection

We used the multilevel logistic regression results to select the top 10 schools identified by the model. To do these, we identified the schools with the largest negative random effects which would be associated with schools that had the least likelihood of having students identified two or more times during the 2023/24 school year after adjusting for the covariates included in the models.

Regression Discontinuity Design Analysis

Regression discontinuity design (RDD) explores a treatment effect for those just above and those just below a specific cut score. In a strict RDD, those below the cut score receive the treatment, whereas those above the cut score do not. When applied to the screening assessment data, those above the benchmarks are eligible for treatment to help boost reading and English skills. In contrast, those above the cut score typically would not be eligible. School situations are complicated, and the screening assessment score is likely not the only criterion for which a student may become eligible for treatment.

Furthermore, we do not have specific treatment information for these students. Therefore, our setup would be closer to an intent-to-treat framework. We want to show an abundance of caution here that this analysis does not represent a treatment effect for any specific treatment or screening assessment because there is significant school and treatment heterogeneity. Instead, this analysis helps to make groups similar to compare students just above to those just below a screening assessment cut score to help adjust for sample differences across screening assessments.

Within a screening assessment and time of year, the screening assessment scores were standardized by subtracting the specific cut score and dividing by the sample specific standard deviation for each screening assessment. This allows the screening assessments to be analyzed together in a single analysis, with the RDD model using coefficients to adjust for each screening assessment. Centering the screening assessment scores on the specific benchmark allowed us to set up the discontinuity at the benchmark score and identify students below 0 as those

eligible for treatment compared with those who would not be eligible for treatment. The general form of the RDD model was as follows:

$$MCAS_{ij} = \beta_0 + \beta_{1T}Screeners_T + \beta_2 ScreenerScores + \beta_3 Below + \beta_4 ScreenerScores{:}\,Below + \beta_{5T} Screener_T{:}\,Below + \beta_{6T} Screener_T{:}\,ScreenerScores + \beta_{7T} Screener_T{:}\,ScreenerScores{:}\,Below + \gamma_{0j} + \epsilon_{ij}$$

where

- $MCAS_{ij}$ is the grade 3 spring MCAS scores for student *i* and school *j*,
- β_0 is the intercept,
- $\beta_{1T}$is the screening assessment specific effects for screening assessment test, *T*,
- β_2 is the association between the fall screening assessment scores and the grade 3 MCAS scores,
- β_3 is the mean difference at the cut score for those below compared with those above the screening assessment cut score for the reference group,
- β_4 is the interaction between screening assessment composite scores and below the cut score indicator allowing the association between screening assessment scores and MCAS scores to differ above versus below the cut score,
- β_{5T} is the interaction between specific screening assessments and the mean of those below versus above the screening assessment cut score allowing for different effects for different screening assessments,
- β_{6T} is the interaction between screening assessment names and the composite screening assessment scores allowing different associations by screening assessment,
- β_{7T} is the three-way interaction between screening assessment composite scores, specific screening assessments, and below or above the cut score indicators that allows there to be different slopes for different screening assessments below versus above the cut scores.
- γ_{0j} are school-specific random effects to adjust for school effects, and
- ϵ_{ij}, are student-specific residuals.

Two analyses were explored for the 2023/24 school year data, one using the below-benchmark and another using the significantly below-benchmark for each screening assessment to establish the cut-score location. The DIBELS 8th Edition screening assessment served as the reference group for the analysis. The optimal cut score was 0.927 for the below benchmark analysis and 0.544 for the significantly below benchmark analysis. To account for multiple statistical tests being performed and to keep the familywise type I error rate in control, test

statistics larger than 2.58 in absolute value were considered, which would be equivalent to a p-value of 0.01 with a standard normal distribution.

Tables E3 and E4 show the study results for each analysis. For the below benchmark analysis, there was evidence of a positive association between fall screening assessment scores and spring MCAS scores across all screening assessments. For DIBELS 8th Edition, the association was weaker for those below the cut score. In contrast, the association was stronger for i-Ready and Star Reading for those below the cut score. Finally, there was an estimated discontinuity for i-Ready, with those below the cut score having an average spring MCAS score about 4.2 points higher than those who took DIBELS 8th Edition (Figure E5). When looking at the significantly below benchmark analysis, one term was important showing a positive association between fall screening assessment scores and spring MCAS scores across all screening assessments (Figure E6).

Table E3. Regression Discontinuity Analysis Results That Estimate the Difference Between Those Just Above Compared With Those Just Below the Benchmark by Screening Assessment

Term	Estimate	SE	Statistic
(Intercept)	494.415	0.89	553.31
Acadience Reading	3.126	3.76	0.83
FastBridge aReading	–4.471	3.39	–1.32
i-Ready	1.268	1.36	0.93
mCLASS	–1.842	1.09	–1.69
Star Reading	–1.677	1.25	–1.35
Not below cut	–2.227	0.97	–2.29
Screening assessment score	21.678	1.38	15.74
Not below cut:screening assessment score	–9.748	1.94	–5.03
Acadience Reading:not below cut	–4.641	4.01	–1.16
FastBridge aReading:not below cut	–1.005	3.60	–0.28
i-Ready:not below cut	4.223	1.46	2.89
mCLASS:not below cut	1.268	1.21	1.05
Star Reading:not below cut	3.303	1.34	2.47

Term	Estimate	SE	Statistic
Acadience Reading:screening assessment score	–1.516	6.96	–0.22
FastBridge aReading:screening assessment score	–1.301	5.84	–0.22
i-Ready:screening assessment score	–3.686	2.11	–1.75
mCLASS:screening assessment score	–3.841	1.67	–2.30
Star Reading:screening assessment score	–2.757	1.97	–1.40
Acadience Reading:not below cut:screening assessment score	9.620	8.21	1.17
FastBridge aReading:not below cut:screening assessment score	14.426	7.03	2.05
i-Ready:not below cut:screening assessment score	13.861	2.94	4.72
mCLASS:not below cut:screening assessment score	2.349	2.41	0.98
Star Reading:not below cut:screening assessment score	15.836	2.70	5.87

Source: Early literacy screening assessment data and October and June SIMS collection data.

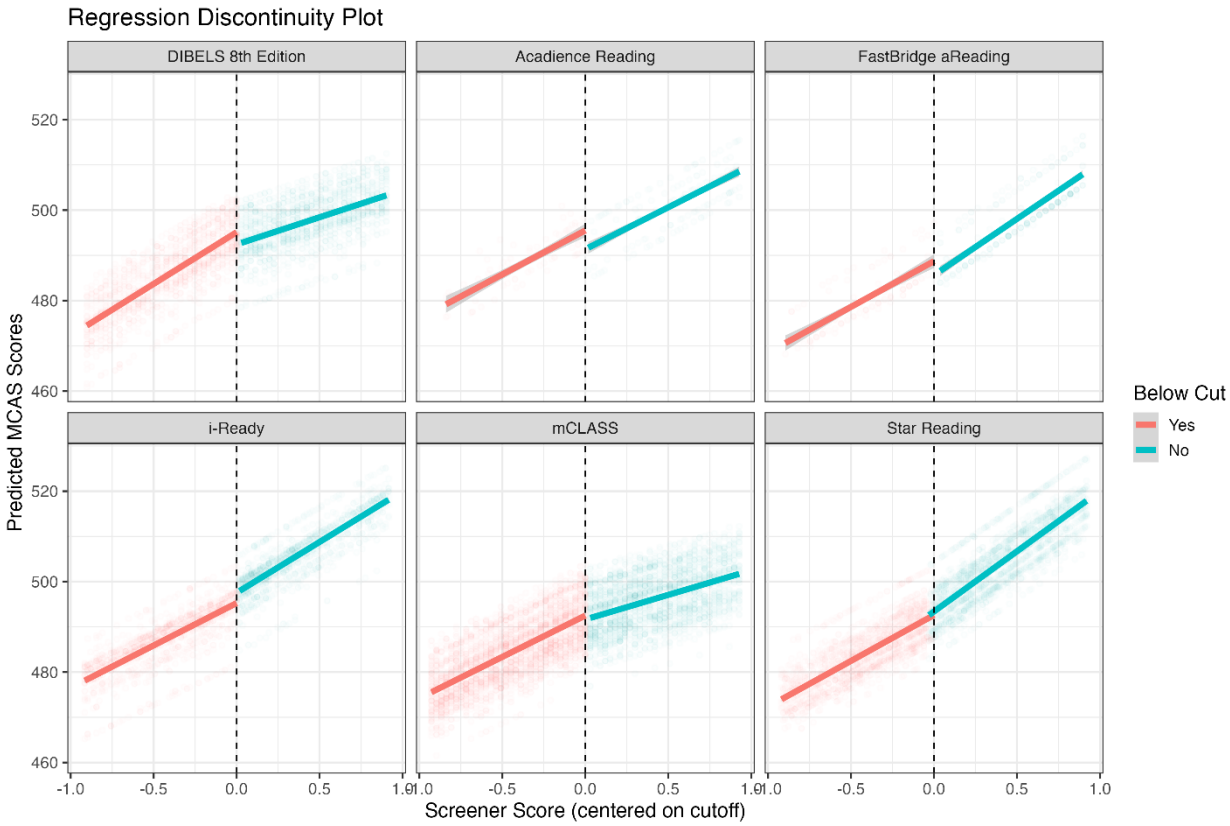
Table E4. Regression Discontinuity Analysis Results That Estimate the Difference Between Those Just Above Compared With Those Significantly Below Benchmark by Screening Assessment

Term	Estimate	SE	Statistic
(Intercept)	482.304	1.20	400.96
Acadience Reading	0.807	5.40	0.15
FastBridge aReading	–11.045	6.13	–1.80
i-Ready	1.117	1.82	0.61
mCLASS	0.474	1.43	0.33
Star Reading	3.018	1.68	1.80
Not below cut	1.217	1.45	0.84

Term	Estimate	SE	Statistic
Screening assessment score	17.689	3.32	5.32
Not below cut:screening assessment score	2.730	4.49	0.61
Acadience Reading:not below cut	12.316	7.14	1.73
FastBridge aReading:not below cut	1.213	7.19	0.17
i-Ready:not below cut	–2.652	2.16	–1.23
mCLASS:not below cut	0.467	1.75	0.27
Star Reading:not below cut	–0.838	1.94	–0.43
Acadience Reading:screening assessment score	–17.249	18.63	–0.93
FastBridge aReading:screening assessment score	–28.985	25.01	–1.16
i-Ready:screening assessment score	–2.416	5.45	–0.44
mCLASS:screening assessment score	1.445	3.97	0.36
Star Reading:screening assessment score	–4.096	4.77	–0.86
Acadience Reading:not below cut:screening assessment score	–12.669	23.55	–0.54
FastBridge aReading:not below cut:screening assessment score	30.959	27.69	1.12
i-Ready:not below cut:screening assessment score	4.570	7.17	0.64
mCLASS:not below cut:screening assessment score	–8.695	5.43	–1.60
Star Reading:not below cut:screening assessment score	7.841	6.16	1.27

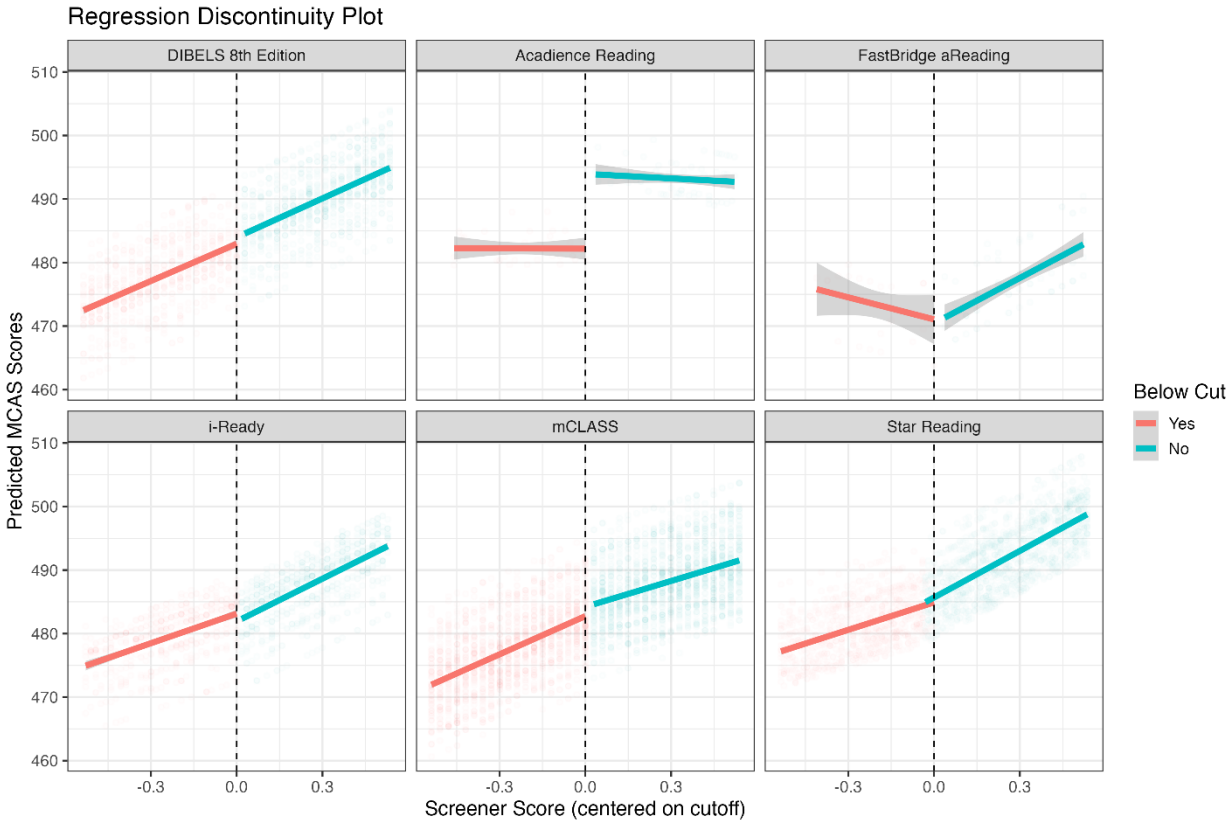
Source: Early literacy screening assessment data and October and June SIMS collection data.

Figure E5. Model-Predicted MCAS Scores for the RDD Analysis, Estimating Gaps Between Those Just Above Versus Just Below the Below Benchmark Cut Score



Note. See Figure E5 Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Figure E6. Model-Predicted MCAS Scores for the RDD Analysis, Estimating Gaps Between Those Just Above Versus Just Below the Significantly Below Benchmark Cut Score



Note. See Figure E6 Appendix F for a full description of this figure. *Source*: Early literacy screening assessment data and October and June SIMS collection data.

Appendix F. Corresponding Data Tables and Figure Descriptions

Figure 2. Corresponding Data Table

Demographic characteristic	Percentage in sample	Percentage in state
Kindergarten	26%	24%
Grade 1	26%	25%
Grade 2	24%	26%
Grade 3	24%	25%
MCAS meeting/exceeding	37%	42%
Low income	51%	43%
Female	49%	49%
Male	51%	51%
English learner students	25%	19%
Special education students	20%	17%
White	74%	77%
Hispanic	32%	26%
Black	20%	17%
Asian	8%	10%
American Indian/Alaska Native	6%	4%
Native Hawaiian/Other Pacific Islander	1%	0.7%

Note. See Figure 2.

Figure 8. Corresponding Data Table

Assessment	Significantly below benchmark two or more times (25th percentile)	Significantly below benchmark two or more times (publisher-provided benchmark)
DIBELS 8th Edition	15%	22%
MAP Reading Fluency	n/a	27%
Star Early Literacy	32%	31%
mCLASS	24%	31%
Star Reading	33%	32%
i-Ready	27%	43%
Total	**25%**	**30%**

Note. See Figure 8.

Figure 11. Corresponding Data Table

Demographic characteristic	Non-Hispanic/Black	Hispanic/Black
Non-LI	26%	29%
Not Hispanic/Black	38%	44%
LI	42%	47%
EL	57%	61%
EL, LI	60%	63%
SpecEdSvc	60%	66%
SpecEdSvc, LI	63%	68%

Note. See Figure 11.

Figure 12. Full Description

This Sankey diagram tracks the academic progression of students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores, showing how

performance levels (either significantly below benchmark or not significantly below benchmark) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Not significantly below at BOY: 67% of students
- Significantly below at BOY: 33% of students

From BOY to MOY

- Of students not significantly below at BOY, 89% maintained that performance level at MOY, while 11% fell to significantly below.
- Of students significantly below at BOY, 75% remained significantly below at MOY, while 25% improved to not significantly below.

From MOY to EOY

- Of students who remained at not significantly below from BOY to MOY, 95% maintained that performance level at EOY, whereas 5% fell to significantly below.
- Of students who fell to significantly below from BOY to MOY, 58% maintained that performance level at EOY and 42% improved to not significantly below.
- Of students who remained at significantly below from BOY to MOY, 76% maintained that performance level at EOY, whereas 24% improved to not significantly below.
- Of students who improved to not significantly below from BOY to MOY, 83% maintained that performance level at EOY and 17% fell to significantly below.

Final Outcomes

- Not significantly below at EOY: 73% of students
- Significantly below at EOY: 27% of students

Note. See Figure 12.

Figure 13. Full Description

This Sankey diagram tracks the academic progression of students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores, showing how performance levels (met benchmark, below benchmark, or significantly below benchmark) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Met at BOY: 48% of students

- Below at BOY: 20% of students
- Significantly below at BOY: 32% of students

From BOY to MOY

- Of students who met at BOY, 88% maintained that performance level at MOY, while 9% fell to below
- Of students who were below at BOY, 33% improved to met, 36% remained below, and 31% fell to significantly below
- Of students who were significantly below at BOY, 11% improved to met, 14% improved to below, and 75% remained significantly below

From MOY to EOY

- Of students who remained at met from BOY to MOY, 95% maintained that performance level at EOY.
- Of students who fell to below from BOY to MOY, 60% improved to met.
- Of students who were below at BOY and improved to met at MOY, 81% maintained that performance level at EOY.
- Of students who remained below from BOY to MOY, 43% improved to met at EOY.
- Of students who were below at BOY and fell to significantly below at MOY, 62% remained at that performance level at EOY.
- Of students who were significantly below at BOY and improved to met at MOY, 77% remained at that performance level at EOY.
- Of students who remained at significantly below from BOY to MOY, 14% improved to below at MOY and 76% remained at that performance level at EOY.

Final Outcomes

- Met at EOY: 60% of students
- Below at EOY: 13% of students
- Significantly below at EOY: 27% of students

Note. See Figure 13.

Figure 18. Corresponding Data Table

Grade and BOY status	Not meeting	Partially meeting	Meeting	Exceeding
G2 BOY significantly below benchmark	35%	49%	11%	0%
G2 BOY not significantly below benchmark	4%	38%	49%	9%
G3 BOY significantly below benchmark	46%	54%	4%	0%
G3 BOY not significantly below benchmark	4%	41%	47%	8%

Note. See Figure 18.

Figure 22. Full Description

This Sankey diagram tracks the academic progression of English learner students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores on the Star Early Literacy or Star Reading assessments in English, showing how performance levels (either significantly below benchmark or not significantly below benchmark) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Not significantly below at BOY: 27% of students
- Significantly below at BOY: 73% of students

From BOY to MOY

- Of students who were not significantly below at BOY, 81% maintained that performance level at MOY, whereas 19% fell to significantly below.
- Of students who were significantly below at BOY, 24% improved to not significantly below, whereas 76% remained significantly below.

From MOY to EOY

- Of students who remained at not significantly below from BOY to MOY, 87% maintained that performance level at EOY, while 13% fell to significantly below.

- Of students who fell to significantly below from BOY to MOY, 49% improved to not significantly below, whereas 51% remained significantly below.
- Of students who improved to not significantly below from BOY to MOY, 62% maintained that performance level at EOY, whereas 38% fell to significantly below.
- Of students who remained at significantly below from BOY to MOY, 15% improved to not significantly below, whereas 85% remained significantly below.

Final Outcomes

- Not significantly below at EOY: 41% of students
- Significantly below at EOY: 59% of students

Note. See Figure 22.

Figure 23. Full Description

This Sankey diagram tracks the academic progression of students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores on Star Early Literacy or Star Reading assessments in Spanish, showing how performance levels (either significantly below benchmark or not significantly below benchmark) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Not significantly below at BOY: 31% of students
- Significantly below at BOY: 69% of students

From BOY to MOY

- Of students who were not significantly below at BOY, 71% maintained that performance level at MOY, whereas 29% fell to significantly below.
- Of students who were significantly below at BOY, 17% improved to not significantly below, whereas 83% remained significantly below.

From MOY to EOY

- Of students who remained at not significantly below from BOY to MOY, 82% maintained that performance level at EOY, whereas 18% fell to significantly below.
- Of students who fell to significantly below from BOY to MOY, 28% improved to not significantly below, whereas 72% remained significantly below.
- Of students who improved to not significantly below from BOY to MOY, 49% maintained that performance level at EOY, whereas 51% fell to significantly below.

- Of students who remained at significantly below from BOY to MOY, 8% improved to not significantly below, whereas 92% remained significantly below.

Final Outcomes

- Not significantly below at EOY: 31% of students
- Significantly below at EOY: 69% of students

Note. See Figure 23.

Figure D1. Full Description

This Sankey diagram tracks the academic progression of students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores, showing how performance levels (where performing at or below the 25th percentile means at significant risk and where performing above the 25th percentile means not at significant risk) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Not at significant risk at BOY: 74% of students
- At significant risk at BOY: 26% of students

From BOY to MOY

- Of students who were not at significant risk at BOY, 94% maintained that performance level at MOY, whereas 6% moved to at significant risk.
- Of students who were at significant risk at BOY, 26% improved to not at significant risk, whereas 74% remained at significant risk.

From MOY to EOY

- Of students who remained not at significant risk from BOY to MOY, 95% maintained that performance level at EOY, whereas 5% moved to at significant risk.
- Of students who moved to at significant risk from BOY to MOY, 42% improved to not at significant risk at EOY, whereas 58% remained at significant risk.
- Of students who improved to not at significant risk from BOY to MOY, 73% maintained that performance level at EOY, whereas 27% moved to at significant risk.
- Of students who remained at significant risk from BOY to MOY, 16% improved to not at significant risk at EOY, whereas 84% remained at significant risk.

Final Outcomes

- Not at Significant Risk at EOY: 76% of students
- At Significant Risk at EOY: 24% of students

Note. See Figure D1.

Figure D2. Full Description

This Sankey diagram tracks the academic progression of students who have beginning-of-year (BOY), middle-of-year (MOY), and end-of-year (EOY) screening assessment scores, showing how performance levels (where above 40th percentile means met benchmark, at or below 40th percentile means below benchmark, and below 25th percentile means significantly below benchmark) change across the academic year and into final outcomes.

Initial Distribution (BOY)

- Met at BOY: 60% of students
- Below at BOY: 14% of students
- Significantly below at BOY: 26% of students

From BOY to MOY

- Of students who met at BOY, 89% maintained that performance level at MOY, whereas 8% fell to below.
- Of students who were below at BOY, 39% improved to met, 38% remained below, and 23% fell to significantly below.
- Of students who were significantly below at BOY, 11% improved to met, 15% improved to below, and 73% remained significantly below.

From MOY to EOY

- Of students who remained at met from BOY to MOY, 93% maintained that performance level at EOY.
- Of students who fell from met at BOY to below at MOY, 43% improved to met and 39% maintained that performance level at EOY.
- Of students who were below at BOY and improved to met at MOY, 72% maintained that performance level at EOY.
- Of students who remained below from BOY to MOY, 31% improved to met and 45% maintained that performance level at EOY.

- Of students who fell from below to significantly below from BOY to MOY, 62% maintained that performance level at EOY.
- Of the students who were significantly below at BOY and improved to met at MOY, 66% maintained that performance level at EOY.
- Of the students who were significantly below at BOY and improved to below at MOY, 37% fell back to significantly below at EOY.
- Of students who remained significantly below from BOY to MOY, 84% maintained that performance level at EOY.

Final Outcomes

- Met at EOY: 62% of students
- Below at EOY: 14% of students
- Significantly below at EOY: 24% of students

Note. See Figure D2.

Figure D5. Corresponding Data Table

Grade level	Gender	Below benchmark	Significantly below benchmark
Grades K–1	Female	69%	63%
Grades K–1	Male	72%	66%
Grades 1–2	Female	84%	80%
Grades 1–2	Male	86%	82%
Grades 2–3	Female	82%	79%
Grades 2–3	Male	82%	77%

Note. See Figure D5.

Figure D6. Corresponding Data Table

Grade level	Income status	Below benchmark	Significantly below benchmark
Grades K–1	Low income	75%	70%
Grades K–1	Non–low income	66%	56%

Grade level	Income status	Below benchmark	Significantly below benchmark
Grades 1–2	Low income	88%	83%
Grades 1–2	Non–low income	82%	78%
Grades 2–3	Low income	86%	81%
Grades 2–3	Non–low income	77%	73%

Note. See Figure D6.

Figure D7. Corresponding Data Table

Grade level	English learner status	Below benchmark	Significantly below benchmark
Grades K–1	English learner	78%	72%
Grades K–1	Non–English learner	67%	59%
Grades 1–2	English learner	89%	87%
Grades 1–2	Non–English learner	83%	78%
Grades 2–3	English learner	88%	83%
Grades 2–3	Non–English learner	79%	75%

Note. See Figure D7.

Figure D8. Corresponding Data Table

Grade level	Special education status	Below benchmark	Significantly below benchmark
Grades K–1	Special education	84%	78%
Grades K–1	Non–special education	66%	59%
Grades 1–2	Special education	92%	88%
Grades 1–2	Non–special education	81%	76%
Grades 2–3	Special education	90%	84%
Grades 2–3	Non–special education	77%	73%

Note. See Figure D8.

Figure D9. Corresponding Data Table

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades K–1	White	71%	64%
Grades K–1	Non-White	71%	67%
Grades 1–2	White	85%	81%
Grades 1–2	Non-White	83%	78%
Grades 2–3	White	81%	78%
Grades 2–3	Non-White	84%	79%

Note. See Figure D9.

Figure D10. Corresponding Data Table

Grade level	Hispanic status	Below benchmark	Significantly below benchmark
Grades K–1	Hispanic	77%	72%
Grades K–1	Non-Hispanic	66%	58%
Grades 1–2	Hispanic	88%	84%
Grades 1–2	Non-Hispanic	84%	79%
Grades 2–3	Hispanic	86%	81%
Grades 2–3	Non-Hispanic	80%	76%

Note. See Figure D10.

Figure D11. Corresponding Data Table

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades K–1	Black	71%	64%
Grades K–1	Non-Black	71%	65%
Grades 1–2	Black	86%	80%

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades 1–2	Non-Black	85%	81%
Grades 2–3	Black	86%	80%
Grades 2–3	Non-Black	81%	78%

Note. See Figure D11.

Figure D12. Corresponding Data Table

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades K–1	Asian	54%	46%
Grades K–1	Non-Asian	72%	66%
Grades 1–2	Asian	79%	81%
Grades 1–2	Non-Asian	85%	81%
Grades 2–3	Asian	77%	75%
Grades 2–3	Non-Asian	82%	78%

Note. See Figure D12.

Figure D13. Corresponding Data Table

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades K–1	American Indian/Alaska Native	80%	75%
Grades K–1	Non–American Indian/Alaska Native	69%	63%
Grades 1–2	American Indian/Alaska Native	79%	72%
Grades 1–2	Non–American Indian/Alaska Native	85%	82%

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades 2–3	American Indian/Alaska Native	85%	80%
Grades 2–3	Non–American Indian/Alaska Native	81%	78%

Note. See Figure D13.

Figure D14. Corresponding Data Table

Grade level	Race/ethnicity	Below benchmark	Significantly below benchmark
Grades K–1	Native Hawaiian/Pacific Islander	64%	64%
Grades K–1	Non–Native Hawaiian/Pacific Islander	71%	65%
Grades 1–2	Native Hawaiian/Pacific Islander	100%	78%
Grades 1–2	Non–Native Hawaiian/Pacific Islander	85%	81%
Grades 2–3	Native Hawaiian/Pacific Islander	78%	78%
Grades 2–3	Non–Native Hawaiian/Pacific Islander	82%	78%

Note. See Figure D14.

Figure E1. Corresponding Data Tables

Screeners

Screening Assessment	Value
Star Reading	Approx. 0.12
Star Early Literacy	Approx. 0.14
mCLASS	Approx. 0.11

Screening Assessment	Value
MAP Reading Fluency	Approx. 0.05
i-Ready	Approx. 0.25
FastBridge aReading	Approx. 0.06
DIBELS 8th Edition	Approx. 0.11
Acadience Reading	Approx. 0.12

Grade

Grade	Value
K	Approx. 0.09
3	Approx. 0.11
2	Approx. 0.12
1	Approx. 0.10

Early Child Programs

Program type	Value
None	Approx. 0.10
Informal	Approx. 0.09
Formal	Approx. 0.08

Low Income

Low income status	Value
Yes	Approx. 0.17
No	Approx. 0.10

Note. See Figure E1.

Figure E5. Full Description

Figure contains six scatterplots with fitted regression lines, each showing screener scores (centered on cutoff below benchmark scores) on the *x*-axis and predicted MCAS scores on the *y*-axis. A vertical dashed line marks the cutoff. For each screener, two regression lines are plotted: one for students scoring below the cutoff and one for students scoring above. In every panel, predicted MCAS scores increase with higher screener scores.

- **DIBELS 8th Edition:** The upward slope is slightly steeper before the cutoff, with a bit of discontinuity at the cutoff, after which the slope flattens slightly.
- **Acadience Reading:** Similar upward slopes are visible on both sides of the cutoff, with a small discontinuity at the cutoff.
- **FastBridge aReading:** Slopes upward at similar rates on both sides, with a bit of discontinuity at the cutoff.
- **i-Ready:** Shows consistent upward slopes on both sides, with a bit of discontinuity at the cutoff.
- **mCLASS:** The upward slope is somewhat steeper before the cutoff, with little to no discontinuity at the cutoff, after which the slope flattens.
- **Star Reading:** Both sides slope upward at similar rates, and the lines meet at nearly the same point at the cutoff, showing little to no discontinuity.

Note. See Figure E5.

Figure E6. Full Description

Figure contains six scatterplots with fitted regression lines, each showing screener scores (centered on cutoff significantly below benchmark scores) on the *x*-axis and predicted MCAS scores on the *y*-axis. A vertical dashed line marks the cutoff. For each screener, two regression lines are plotted: one for students scoring below the cutoff and one for students scoring above.

- **DIBELS 8th Edition:** Predicted MCAS scores increase at similar rates below and above the cutoff, with minimal discontinuity at the cutoff.
- **Acadience Reading:** The regression line below the cutoff is nearly flat, whereas the line above the cutoff is also nearly flat with a slight downward slope. At the cutoff, there is a positive discontinuity of about 10 points, with predicted MCAS scores higher for students above the cutoff.
- **FastBridge aReading:** Below the cutoff, the regression line slopes slightly downward, whereas above the cutoff it slopes upward, with little to no discontinuity at the cutoff.
- **i-Ready:** Predicted MCAS scores increase steadily with screener scores, with minimal discontinuity at the cutoff.

- **mCLASS:** Both lines slope upward with screener scores, with a small discontinuity at the cutoff.
- **Star Reading:** Both lines slope upward with screener scores, and there is little to no discontinuity at the cutoff.

Note. See Figure E6.


WestEd
WestEd.org