WestEd
WestEd.org

Early Literacy Performance in Massachusetts

Results of Ongoing Analysis of Literacy Screening Assessments

Mariann Lemke, Dan Murphy, Aaron Soo Ping Chow, and Angela Acuña

Fall 2024

WestEd is a nonpartisan, nonprofit agency that conducts and applies research, develops evidence-based solutions, and provides services and resources in the realms of education, human development, and related fields, with the end goal of improving outcomes and ensuring equity for individuals from infancy through adulthood. 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.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Executive Summary

Study Background

Beginning with the 2020/21 school year, the Massachusetts Department of Elementary and Secondary Education (DESE) began an ongoing effort to collect and analyze literacy screening assessment data from schools and districts participating in certain state grants to inform improvement efforts. Grantee schools and districts that provide literacy screener data to DESE select their screening assessments from a list of state-approved, commercially available literacy screener products. Each assessment is typically administered to studhispents three times per year (most commonly in the fall/beginning of year [BOY], winter/middle of year [MOY], and spring/end of year [EOY]).

Analysis of 2021/22 and earlier data provided a first look at early literacy performance and progress of students in Massachusetts,[1] offering a number of key insights. This report follows up on that analysis using data from the 2022/23 school year. Though direct comparisons of results across reports cannot be made due to changes in methods (see box at right), patterns of student performance and progress in 2022/23 are consistent with earlier analysis. This year's reporting also offers some new information related to performance of students with different background characteristics, performance of students in Massachusetts relative to the nation, and comparisons of screening assessment benchmarks.

Changes in Data and Methods in 2022/23 Analysis

- Available data in 2022/23 doubled from 2021/22, including nearly 200,000 scores and 67,000 students, or about 26 percent of the state's K–3 student population, with scores from 308 schools, 88 districts, and 12 screening assessments.
- More students in 2022/23 had scores from all three time periods, improving the ability to analyze student progress.
- 2022/23 reporting focuses on students "significantly below benchmark" instead of "below benchmark" to align with new state guidance.

Key Findings

- **As in prior reporting, more students were identified as at risk (in this case, significantly below benchmark) at the beginning of the school year than at the middle or end of the year.** The percentage of students significantly below benchmark decreased from 36 percent at BOY to 29 percent at EOY. Across all grades and all assessment time periods combined, 45 percent of students were *ever* classified as significantly below benchmark. Additionally, 31 percent were significantly below benchmark multiple

[1] Initial reports with 2020/21 and 2021/22 data available at the DESE First Look at Early Reading Performance in Massachusetts page.

times in the school year. See the student performance section on identification of students by time period (page 56).

- **Kindergarten students again showed the greatest change from beginning to end of year compared to students in grades 1 through 3.** The percentage of kindergarten students identified as significantly below benchmark decreased by 21 percentage points between BOY and EOY, compared to 3 percentage points for grades 1 and 2 and no change for grade 3 students. By EOY, a smaller percentage of kindergarten students were significantly below benchmark than the percentage for all other grade levels. Kindergarten students were also least often significantly below benchmark more than once during the year. See the student performance section on identification of students by grade level (page 57).
- **As in the previous analysis, most students who start the year below benchmark or significantly below benchmark still perform at those levels at the end of the year, though there are differences by grade.** Overall, about 60 percent of students who started the year significantly below benchmark were still significantly below benchmark at EOY. About 30 percent of students who were significantly below benchmark at BOY met benchmark by EOY, and another 16 percent improved to the next performance level, though they did not meet benchmark. Students who improved to the meeting benchmark level MOY were significantly more likely to meet benchmark at EOY than those who were still classified as at risk at MOY. See the student progress section on changes in performance across the school year (page 71).
- **As was true in past reporting, most students who were at risk in one grade level were also at risk in the next grade level, but students at earlier grade levels were more likely to get on track.** More students in later grades remained below benchmark or significantly below benchmark across years than did students in earlier grades. For example, 77 percent of grade 2 students who were significantly below benchmark were still significantly below benchmark at the end of grade 3 (compared to 64 percent of kindergarten students who moved to grade 1). See the student progress section on changes in performance across grade levels (page 82).
- **Differences in performance and progress between student groups are again evident.** Outcome data suggest that the current educational system often does not provide adequate support for students from historically marginalized groups, such as those learning English or students with disabilities. Low income students, English learner students, students receiving special education services, Black students, and Hispanic students were more likely than their peers not in those groups to be classified as significantly below benchmark at each time period and more than once during the school year. Asian and White students were less likely than their peers to be classified as such. See the student performance section on identification of students by student group (page 58).

- **New analysis examining students' intersecting social and economic background characteristics in combination shows that the probability of students needing additional support increases as their association with historically undersupported groups overlap, but not always in the same ways across groups.** See student performance section on identification of students by overlapping characteristics (page 65).
    - Students who belong to a historically underserved student group, such as low income students, are more likely to be identified as at risk of reading difficulty, and the more of these groups a student belongs to, the greater the chance of being identified as at risk. For example, students who receive special education or English learner support services *and* come from a low income background have a greater likelihood of being identified as at risk than do students with only one of these characteristics. And students belonging to all three groups have an even higher likelihood of being identified as at risk.
    - Although the pattern of being identified as at risk increasing in likelihood as background characteristics intersect is the same for all students, the increase varies by gender and other background factors. Female students who are English learners or who receive special education services are more likely to be identified as at risk than males across ethnoracial groups. Asian students show the smallest increases in likelihood of being identified as at risk across ethnoracial groups.
    - School characteristics affect students' chances of being identified as at risk. Students in schools with above-average percentages of low income students, below-average teacher retention rates, and below-average student attendance have a higher likelihood of being identified as at significant risk. In Massachusetts, more Black and Hispanic students attend schools with these characteristics.
- **Extending prior analysis that used the Massachusetts Comprehensive Assessment System (MCAS) to compare risk level benchmarks to additional grade levels shows that some benchmarks shift from higher to lower (or vice versa) on the MCAS scale within and across grades.** These shifts mean that students with the same skills might be classified differently at different time periods, and changes in the percentages of students identified as significantly below benchmark over time may be due in part to changes in the benchmarks themselves. See the comparison of screening assessment benchmarks (page 35).
- **New analysis also shows that despite differences between student groups, students in Massachusetts overall perform above the national average, based on norms provided by assessment publishers.** At EOY, the median national percentile of performance for Massachusetts students was 58 (with 50 the median national percentile). Black students in Massachusetts perform about the same as all other students around the country, White students perform slightly better, Asian students perform significantly better, and Hispanic students perform less well than the national

sample. Growth of students using publisher-provided growth norms shows 66 percent of students growing at average or above average rates compared to about 60 percent nationally. See the section that compares Massachusetts students' performance with overall national performance (page 87).

- **Based on data from 2022/23, early childhood (EC) learning experience (and more specifically, EC experience in a formal environment) reduces the likelihood of students in kindergarten and beyond being identified as at significant risk of reading difficulty.** New analysis shows that more kindergarten students *without* EC experience were identified as significantly below benchmark within any time period and multiple times during the school year than were students with any EC experience. For all students in grades K–3, students with formal EC experience were less often significantly below benchmark than were students with informal EC experience. Effects of formal EC experience were most pronounced among English learner students and among Black English learner students in particular. See the section on identification of students by EC program (page 95).

The following sections provide more detail on these findings.

Introduction

State-level efforts to improve student literacy skills continued around the country in 2023, including new policies and practices related to curriculum, instruction, educator preparation, and systems of support for struggling readers. Following on strategic planning and legislative efforts aimed at literacy that began in Massachusetts in 2018, an amendment to state regulation was introduced in September 2022 requiring 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 (see box below).

Massachusetts Early Literacy Screening Regulation

Effective July 1, 2023, each school district shall at least twice per year assess each student's reading ability and progress in literacy skills, from kindergarten through at least third grade, using a valid, developmentally appropriate screening instrument approved by the Department. Consistent with section 2 of chapter 71B of the general laws and the Department's dyslexia and literacy guidelines, if such screenings determine that a student is significantly below relevant benchmarks for age-typical development in specific literacy skills, the school shall determine which actions within the general education program will meet the student's needs, including differentiated or supplementary evidence-based reading instruction and ongoing monitoring of progress. Within 30 school days of a screening result that is significantly below the relevant benchmarks, the school shall inform the student's parent or guardian of the screening results and the school's response and shall offer them the opportunity for a follow-up discussion.

Source: Early Literacy Screening, Regulation 603 CMR 28.03(1)(f) (Code of Massachusetts Regulations Title 603)

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 "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 to implement 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.

Beginning with the 2020/21 school year, DESE began collecting literacy screening assessment data from schools and districts participating in certain state grants. Since then, the number of grantees and program participants that report data has increased, allowing for more robust analysis of student performance in the early grades. As all schools began to implement the new regulation in the 2023/24 school year, additional data may become available.

This report follows up on the initial analysis based on 2020/21 and 2021/22 screening assessment data (see WestEd's 2023 report *A First Look at Early Literacy Performance in Massachusetts: Results of Initial Analysis Based on State Grantee Literacy Screening Assessment*).

In collaboration with DESE and other interest holders in the state, WestEd, a nonpartisan education research and service organization, developed a set of research questions aimed at informing educators and policymakers about trends in student performance and about areas in which additional information and research is needed and at identifying potential levers for policy and practice. Research questions will be reviewed and updated each year. Table 1 shows current research questions (i.e., those addressed in this report).

Table 1. Research Questions

Topic	Research question
Comparing screening assessment benchmarks	• How do screening assessment benchmarks identifying students as significantly below benchmark compare in Massachusetts Comprehensive Assessment System (MCAS) and national percentile terms?
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? How many benchmark scores are available by grade and student group? How many benchmark scores are available by time period? • To what extent does the sample of students with available benchmark scores represent the overall K–3 student population in the state?
Student performance	• How many (and what percentage of) 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 intersecting and school background factors interact and affect student likelihood of being identified as significantly below benchmark more than once?

Topic	Research question
Student progress	• How does student performance change from beginning of year (BOY) to middle of year (MOY) to end of year (EOY)? Do students identified as significantly below benchmark remain significantly below benchmark? • How does student progress from BOY to MOY to EOY vary by benchmark level, grade, and student background characteristics? • How does student performance change across grade levels? Do students at risk remain at risk across years? • What is the relationship between grade 2 (BOY, EOY) and grade 3 (BOY, EOY) literacy screening assessment scores and MCAS performance?
Comparing student performance to the nation	• How does achievement and growth of Massachusetts students compare to national samples and growth rates?
Kindergarten student performance and early childhood (EC) experience	• How many kindergarten students attended any type of EC program, by type? • What percentage of kindergarten students were significantly below benchmark at BOY, MOY, and EOY overall and by EC program type? How many students overall and by program type were significantly below benchmark multiple times? • How do student background characteristics interact and intersect with EC program enrollment and affect the likelihood of students being identified as significantly below benchmark?

This report provides the results of analysis of the approved early literacy universal screening assessment data collected from state grantees in the 2022/23 school year, following on initial analysis carried out with 2020/21 and 2021/22 data. The report is organized as follows:

- Available Data
- Analysis and Findings
- Discussion and Next Steps

Available Data

This report draws on 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.[2] The student-level data include:

- early literacy universal screening assessment data for K–3 students in districts receiving specific state grants,
- the state's Student Information Management System (SIMS) data,
- Massachusetts Comprehensive Assessment System data, and
- Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs, or ACCESS) data.

Early Literacy Universal Screening Assessment Data

Beginning in the 2020/21 school year, recipients of certain state grants were required to provide their students' early literacy universal screening assessment data to DESE. Over time, the number of included grants and programs has expanded. Current examples include the Early Grades Literacy, Early Literacy Screening Assessment and Professional Development, Growing Literacy Equity Across Massachusetts, Accelerating Literacy, and High Quality Instructional Materials Implementation grants.[3]

As part of early efforts to encourage screening of students for potential reading difficulties, 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 for certain grantees. In 2022, the list of approved assessments was updated to better reflect recent Massachusetts Dyslexia Guidelines, and assessments may continue to be added to the approved list on a rolling basis.[4] A summary of the 2022 universal screening assessment criteria and descriptions of approved assessments can be found in appendix A.[5]

Currently, nine early literacy screening assessments are approved for use in elementary schools in Massachusetts, with four that were rated during the review process as "Meet Expectations" (Amira, Dynamic Indicators of Basic Early Literacy Skills [DIBELS] 8th Edition, EarlyBird, and mCLASS) and five rated as "Partially Meet Expectations" (Acadience Reading, FastBridge's suite

[2] School- and district-level data can be found at the DESE School and District Profiles page.
[3] Additional information about DESE's grant funding opportunities can be found at DESE's Grants and Other Financial Assistance Programs page.
[4] The Massachusetts Dyslexia Guidelines can be viewed on DESE's Special Education page.
[5] The currently approved assessment list and the state criteria used to review the assessments can also be found on DESE's Early Literacy Universal Screening Assessments page.

of assessments, i-Ready, MAP Reading Fluency, and Star's suite of assessments). See appendix A for DESE's brief description of each of these approved assessments.

For this report, data from the following 12 screening assessments—including from current and previously approved lists—were included in some or all analyses:

- Acadience Reading
- aimswebPlus
- DIBELS 8th Edition
- mCLASS
- EarlyBird
- FastBridge (aReading, CBMreading, earlyReading)
- i-Ready (Diagnostic)
- Istation's Indicators of Progress Early Reading (ISIP ER)
- Lexia RAPID
- MAP Growth
- MAP Reading Fluency
- Star (Curriculum-Based Methods [CBM], Early Literacy English and Spanish, Reading)

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. Files included data such as student ID (i.e., state and/or local ID), school year, school and/or district name, assessment period (e.g., fall/beginning of year [BOY], winter/middle of year [MOY], spring/end of year [EOY]), test administration date, and test name, along with screening assessment data such as composite scores, benchmark levels, national percentile ranks, and/or reading risk flags.

Although each of these assessments is commonly used for early literacy screening, they can vary in significant ways, including the content assessed, technical characteristics of the assessments, mode of administration, type of scores provided (e.g., composite scores, reading risk flags), benchmark and risk definitions, and cut score calculations. These differences are important to keep in mind when comparing students' scores across these assessments.

The state's goal in approving these assessments was to help schools and districts choose technically sound tools to identify students at risk of reading difficulties (including dyslexia) so that support services can be provided to them.[6] However, most of the approved screening

[6] DESE guidance notes that "early literacy universal screening assessments do not diagnose dyslexia" but may be used to identify risk of dyslexia.

assessments do not simply identify risk using a single cut score. Instead, most provide several performance benchmarks or risk levels (e.g., low risk, some risk, high risk, below benchmark, at benchmark, above benchmark; table 2).

These levels differ in how they were determined and in what they represent. For example, the cut scores for each level for a screening assessment may have been determined based on data about students' later performance on that screening assessment or other reading assessments later in the year or in later grades. Alternatively, benchmarks may have been set based on normative data—selecting a percentile, such as the 40th percentile, to identify the lowest-performing group of students.[7] 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 all the specific reading subtests administered at each grade level. These differences mean there is no truly common definition of risk across screening assessments.

For the main analyses in this report, we focus on benchmarks DESE identifies in its June 2023 Early Literacy Screening Guidance and Dyslexia Guidelines documents as best aligned with state regulation to identify students at risk.[8] The state's regulation requires that schools or districts take action, including informing parents or guardians of results, if a screening assessment determines that a student is "significantly below relevant benchmarks," but the regulation does not specifically define "significantly below benchmark." The state's Dyslexia Guidelines, however, recommend using the national 25th percentile or below as a metric to define the category of "significantly below relevant benchmarks." In its Early Literacy Screening Guidance, DESE provides the relevant publisher-provided reporting categories for each approved screening assessment that they recommend schools and districts use to determine whether or not students are performing "significantly below relevant benchmarks." For example, for DIBELS 8th Edition, DESE recommends using the "at risk" or "well below benchmark" performance level to identify students significantly below benchmark. Table 2 describes the levels used in analysis for each of the early literacy screening assessments and provides additional information on the definition of risk according to each assessment. We report using both the publisher-provided benchmark categories and the 25th percentile or below metric, noting where differences occur (though national percentile data were not available for a few assessments). Using both types of reporting metrics can provide complementary information—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 may provide a more consistent measure across assessments, though norming procedures and samples can also vary. Additionally, analysis suggests that there can be variation in the rigor of benchmarks within assessments at different time periods (BOY/MOY/EOY), which makes interpreting measures of growth within and across

[7] The populations on which norms were based may also differ across assessments.

[8] Available at https://www.doe.mass.edu/instruction/screening-guide.pdf and https://www.doe.mass.edu/sped/dyslexia-guidelines.pdf.

assessments challenging. Using the 25th percentile or below metric allows for more consistent measurement of change despite potential differences in norming procedures across assessments. The "Comparing Screening Assessment Benchmarks" section provides more detail on analysis and findings related to benchmarks.

Note that initial reporting on results of literacy screening assessments based on 2020/21 and 2021/22 data (carried out prior to release of DESE guidance on early literacy screening) focused on students identified as having at least some risk of reading difficulty, which generally maps to "below benchmark" performance rather than "significantly below benchmark." The change in focus from "below benchmark" or "at least some risk" to "significantly below benchmark" (or "significant risk" as described in earlier reporting) means that direct comparisons of results across reports cannot be made. However, for analyses in this report that describe data over time, we consistently apply the 2023 DESE guidance categories and/or 25th percentile or below metric. Initial reports with 2020/21 and 2021/22 data available in the DESE web page First Look at Early Reading in Massachussetts.

One of the main goals of the analysis was to provide Massachusetts with an estimated percentage of students across the state who are at risk of reading difficulties, according to the screening assessment data. 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 (this situation occurred in about 2 percent of records). The rules used to calculate benchmark levels are detailed in appendix C, alongside other business rules regarding the processing and merging of the screening assessment and other student-level data.

Table 2. Description of Risk Levels for Early Literacy Screening Assessments Included in Analysis Aligned to DESE Early Literacy Screening Guidance

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
Acadience Reading Acadience provides four levels to describe student performance for a reading composite score and for subtest scores. The levels indicate the overall likelihood of achieving subsequent proficiency goals (without targeted instructional support) and the overall level of need for students in these benchmark categories.	2020/21, 2022/23	• Above Benchmark: Likelihood of achieving subsequent early literacy goals is 90%–99% (Core Support needed) • At Benchmark: Likelihood of achieving subsequent early literacy goals is 70%–85% (Core Support needed; students near the benchmark cut score may require monitoring and/or Strategic Support on specific skills)	• Below Benchmark: Likelihood of achieving subsequent early literacy goals is 40%–60% (Strategic Support needed)	• Well Below Benchmark: Likelihood of achieving subsequent early literacy goals is 10%–20% (Intensive Support needed; students in this benchmark category are at risk of not achieving reading goals unless Intensive Support is provided)	No
aimswebPlus aimswebPlus provides three levels to describe student performance for a reading composite score. The levels indicate the overall likelihood of meeting the spring performance target.	2022/23	• Tier 1/low risk: Continue general instruction; 75%–90% of students expected to meet EOY target	• Tier 2/moderate risk: Require intervention; 25%–65% of students expected to not meet EOY target	• Tier 3/high risk: Require intensive, individual instructional intervention; 50%–90% of students expected to not meet EOY target	Yes

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
DIBELS 8th Edition DIBELS 8th Edition provides four levels to describe student performance for a reading composite score and for subtest scores. Scores represent the overall level of need for students and their risk of not achieving proficiency goals.	2020/21, 2021/22, 2022/23	• Above Benchmark: Core Support; Negligible Risk; nearly all students in this category score at or above the 40th percentile on criterion measure • At Benchmark: Core Support; Minimal Risk; 80% of students who score at or above the 40th percentile on criterion measure fall in this category	• Below Benchmark: Strategic Support; Some Risk; 80% of students who score below the 40th percentile on criterion measure fall in this category	• Well Below Benchmark: Intensive Support; At Risk; classifies students who are at risk of reading difficulties, including dyslexia; 80% of students who score below the 20th percentile on criterion measure fall in this category	Yes
mCLASS See description for DIBELS 8th Edition (mCLASS assessments are based on DIBELS 8th Edition, and reported performance levels are the same).	2020/21, 2021/22, 2022/23	See description for DIBELS 8th Edition	See description for DIBELS 8th Edition	See description for DIBELS 8th Edition	Yes

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
EarlyBird EarlyBird provides a dyslexia risk flag that indicates the likelihood that a student will be at risk of severe word reading struggles at the end of the school year (provided the student doesn't receive appropriate remediation). According to the publisher, severe word reading struggles are defined as "performing at or below the 20th percentile on the SAT-10 (Stanford Achievement Test Series, Tenth Edition, 2018, Pearson Education, Inc.)." This flag is used to indicate students significantly below benchmark. EarlyBird provides different metrics at each time period to identify students at any risk of reading difficulties. In BOY, no metric is available, other than the significant risk indicator described above (i.e., dyslexia risk flag). In MOY, students receive a Potential for Word Reading (PWR) likelihood percentage, which is the probability that a student will reach grade-level expectations in word reading by EOY without remediation. According to the publisher, "Reaching expectations, for the purposes of this analysis, is defined as performing above the 40th percentile on the SAT-10: a reasonable standard for measuring grade-level expectation word reading" (Gaab & Petscher, 2022, p. 10). In EOY, EarlyBird refers to the Word Reading subtest score, which is available only to kindergarten students at EOY. Percentile ranks are used to describe a student's performance on each subtest.	2021/22, 2022/23	• BOY: Not available • MOY: Meets Expectations (based on PWR) • EOY: At/Above Benchmark (based on Word Reading subtest): 41st–99th percentile	• BOY: Not available • MOY: Below Expectations (based on PWR) • EOY: Below Benchmark (based on Word Reading subtest): 21st–40th percentile • EOY: Well below benchmark (based on Word Reading subtest): Below the 21st percentile	• BOY, MOY, EOY: Flagged (based on dyslexia risk flag)	No

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
FastBridge aReading FastBridge aReading provides four levels to describe student performance for the composite scaled scores. These benchmarks "were established for FastBridge aReading to help teachers accurately identify students who are at risk for not meeting the current grade level expectations as measured by future performance on important tests such as the state assessment" (FastBridge Learning, 2019c, p. 56). The FastBridge benchmarks are based on its national norms and correspond to percentile ranges.	2021/22, 2022/23	• Advanced/College Pathway: 71st–99th percentile • Low risk: 40th–70th percentile	• Some risk: 15th–39th percentile	• High risk: Below the 15th percentile	Yes
FastBridge CBMreading FastBridge CBMreading provides four levels to describe student performance for the words read correctly per minute score. Benchmark levels are not available for kindergarten students. Benchmarks "were set by examining data from students who completed both the FastBridge CBMreading assessment and another 'high stakes' assessment such as a state test. … Results indicate that FastBridge™ CBMreading is highly predictive of student's scores on other reading assessments" (FastBridge Learning, 2019c, p. 17). The FastBridge benchmarks are based on its national norms and correspond to the percentile ranges.	2020/21, 2021/22, 2022/23	• Advanced/College Pathway: 71st–99th percentile • Low risk: 40th–70th percentile	• Some risk: 15th–39th percentile	• High risk: Below the 15th percentile	Yes

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
FastBridge earlyReading FastBridge earlyReading provides three levels to describe student performance for composite and subtest scores. Benchmarks are not available for grade 2 and grade 3 students. Benchmarks "were developed from a criterion study examining FastBridge™ earlyReading assessment scores in relation to scores on the Group Reading Assessment and Classification Evaluation" (FastBridge Learning, 2019c, p. 36). The benchmarks are based on the national norms and correspond to percentile ranges.	2020/21, 2021/22, 2022/23	• Low risk: 40th–99th percentile	• Some risk: 15th–39th percentile	• High risk: Below the 15th percentile	Yes
i-Ready Diagnostic The i-Ready Diagnostic test provides three benchmarks for composite and subtest scale scores. These benchmarks are criterion-referenced (i.e., based on judgments about performance relative to expectations set by the Common Core State Standards, not based on normative data about student performance). Benchmarks can be used to determine whether students are meeting grade-level expectations.	2020/21, 2021/22, 2022/23	• No observed risk	• Some risk	• At risk	Yes

Early literacy screening assessment	Years with data	Description of levels(s) corresponding to "at/above benchmark"	Description of level(s) corresponding to "below benchmark"	Description of level(s) corresponidng to "significantly below benchmark"	National percentiles available in data
Lexia RAPID Lexia RAPID reports three performance levels based on its Reading Success Probability (RSP) score. In grades K–2, the RSP is calculated by a combination of a student's performance in the Phonlogical Awareness, Letter Sounds, Word Reading, Vocabulary Pairs, Spelling, and Following Directions tasks, with the combination differing by grade level (Lexia Learning, 2020). In grade 3, The RSP "is calculated by a combination of a student's performance in the Word Recognition, Vocabulary Knowledge, Syntactic Knowledge, and Reading Comprehension tasks. This formula is based on the student's grade level, since the factors that are most predictive of reading comprehension success change as a student grows older" (Lexia Learning, 2022).	2020/21, 2021/22, 2022/23	• High likelihood of EOY grade-level success: An RSP of 70% or higher means that a student has a high likelihood of reaching EOY grade-level success. A student with an RSP in this range will continue to benefit from universal instruction.	• Moderate likelihood of EOY grade-level success: An RSP between 69% and 31% means that a student has a moderate likelihood of reaching EOY grade-level success. A student with an RSP in this range may need additional instruction to target skill weaknesses.	• Low likelihood of EOY grade-level success: An RSP of 30% or lower means that a student has a lower likelihood of reaching EOY grade-level success. A student with an RSP in this range may need more intensive instruction to target skill weaknesses.	No
Star Early Literacy and Star Reading Star Early Literacy and Star Reading provide four levels based on the composite scaled score, which are established based on normative data. The default benchmark is the 40th percentile ("based on a review of proficiency cut scores from several state assessments and guidance from RTI [response to intervention] experts"), which identifies students who "require some form of intervention to accelerate their growth and bring them into benchmark range" (Renaissance Learning, 2022a, p.1).	2021/22, 2022/23	• At/Above Benchmark: Students meeting or exceeding the benchmark score (at or above the 40th percentile)	• On Watch: Students slightly below the benchmark score (automatically calculated range between at/above benchmark level and intervention level)	• Intervention: Students below the benchmark score (below the 25th percentile) • Urgent intervention: Students far below the benchmark score (below the 10th percentile)	Yes

Source: Authors' compilation based on assessment documentation and/or communication with publishers (see references for list of technical reports and other documentation reviewed)

Note: i-Ready Diagnostic, combined with the i-Ready Literacy Tasks, is currently approved to be administered in Massachusetts. The information presented in the table and in the report only pertain to i-Ready Diagnostic scores, as Literacy Task data were not available. EarlyBird was a kindergarten-only assessment; 2021/22 EarlyBird data were not included in prior analysis. The performance levels considered as significantly below benchmark for Star Early Literacy and Star Reading include the "Intervention" and "Urgent Intervention" levels. Prior analysis did not include "Intervention" as a significantly below benchmark level.

Student-Level State Education Data

In addition to the K–3 early literacy screening assessment data, other student-level data were used for analysis, including Student Information Management System data, Massachusetts Comprehensive Assessment System data, and Assessing Comprehension and Communication in English State-to-State for English Language Learners data.[9] 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 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, Massachusetts Comprehensive Assessment System (MCAS), and ACCESS data collections. A brief description of each dataset follows.

Student Information Management System

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 dataset, students can be identified using their unique state-assigned statewide 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 datasets, then they were included only in analyses not requiring demographic data. Only 189 observations (0.09 percent of cases and 87 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 so cannot be matched to SIMS data.

Only the variables that were relevant to K–3 students were used when creating the student-level dataset for the analyses. 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, and Version 20.9 for the 2022/23 school year.[10]

Massachusetts Comprehensive Assessment System

The MCAS data for the 2022/23 school year provide student assessment scores in mathematics and English language arts (ELA) for the grade 3 students in the sample. These data provide a standardized measure of ELA achievement for most grade 3 students, allowing for an analysis of

[9] The Office of Planning and Research at DESE provides a guide for researchers with more specific information about the student-level data provided by DESE.

[10] The current SIMS Data Handbook can be viewed on DESE's SIMS page.

grade 3 screening assessment cut scores by linking them to the MCAS cut scores that describe student performance levels (Not Meeting Expectations, Partially Meeting Expectations, Meeting Expectations, or Exceeding Expectations). English learner students in their first year in the United States are exempt from taking the MCAS ELA assessment and are therefore excluded from MCAS-related analysis.

Assessing Comprehension and Communication in English State-to-State for English Language Learners

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 and February of each school year.[11]

The 2022/23 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. The proficiency levels are on a scale of 1 through 6 and can be used to describe a student's performance in terms of the 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 consider their performance on each of the four language domains and 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 2022/23 were retrieved from DESE's school and district profiles website to provide contextual data about the sample of students used in

[11] Information regarding the ACCESS for ELLs assessment (and its associated performance levels and interpretations) can be viewed on DESE's MCAS page.

the analysis. Overall, the data pertain to the following four main categories of 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 low income or economically disadvantaged students, attrition rate, retention rate), and (d) financial (e.g., per-pupil expenditure amounts).

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

All student-, school-, and district-level data for the 2020/21, 2021/22, and 2022/23 school years were combined into one primary longitudinal analytic file. This file was provided to DESE, along with an accompanying codebook with a description of each variable and its associated values/codes. Data are organized as a single longitudinal dataset with one observation per student, per test period (i.e., BOY, MOY, and EOY), per screening assessment, per year. Some students have multiple screening assessment scores per test period, as they were delivered multiple early literacy screening assessments during the school year. Appendix C details the process of merging different data sources and discusses the data issues that arose during the data cleaning process and the decisions that were made to resolve these issues.

This report follows DESE's standard procedure for suppression of student demographic and assessment data. DESE uses a minimum sample size of 6 students for reporting any student demographic information and a minimum sample size of 10 students for reporting student assessment outcomes. Two dashes (--) represent suppressed data in this report.

Analysis and Findings

As described in the introduction, this report provides follow-up analysis on the literacy performance of students in grades K–3 in Massachusetts based on available data. In this section, we first provide additional information on screening assessment benchmarks and then describe available data and discuss how well the data represent students in grades K–3 in Massachusetts as a whole. We also provide updated information about student performance and progress.

Comparing Screening Assessment Benchmarks

One of the challenges of the screening assessment data being collected is that it includes data from assessments that were designed differently and that use different approaches to determine whether students are at risk or on track.

However, there is one assessment that all students in grade 3 in the state do take: MCAS. The MCAS ELA assessment therefore provides a mechanism to look across screening assessments using a common metric. Overall, we can link scores from over 3,500 students who took grade 2 screening assessments in 2021/22 to their MCAS scores from 2022/23 and about 14,000 students who took grade 3 assessments in 2022/23 to their MCAS scores in 2022/23. In this section, we describe results of an equipercentile linking analysis similar to that carried out previously using 2021/22 data for grade 3 EOY only. 2022/23 analysis links benchmark cut scores from grade 2 and grade 3 literacy screening assessments at multiple time periods to MCAS Grade 3 ELA scale scores.

Key Findings

- Linking estimates show most benchmarks identifying students as significantly below benchmark map to the MCAS Partially Meeting Expectations performance level, though there is variation between assessments within that level.
- Screening assessment benchmarks indicating significant risk do not always map to the same MCAS and national percentile scores at each time period or across grade levels, which means that students with the same skills might be classified differently at different time periods and that changes in the percentages of students identified as significantly below benchmark over time may be due in part to changes in the benchmarks themselves.
- Screening assessments accurately predict that most students at significant risk will not meet MCAS standards, but are somewhat less accurate at predicting MCAS proficiency for students classified as at little or no risk.

We also carried out predictive linking and examined the accuracy of grade 3 screener benchmarks in predicting MCAS performance. Grade 3 EOY linking results are similar to those from 2021/22, but examining benchmarks within and across grade levels for screening assessments provides new information about how benchmarks can vary in relation to MCAS, which has implications for how many and which students are identified as significantly below benchmark over time.

All analyses use a single-group design, in which students included in the sample took both a screening assessment and the MCAS. This approach helps control for differences in proficiency between students (Brennan, 2006). Also, the amount of time between tests can affect the accuracy of linking. We focus here on links between grade 3 screening assessments and MCAS, given the larger sample size available for analysis at grade 3. Grade 2 results are discussed later in this section when comparing the literacy screening assessment benchmark scores at each

time period. Additional grade 2 results can be found in appendix D, along with details about the methods and the sample included in analysis.

How Many Scores Are Available for Linking Analysis? To What Extent Are the Samples Representative of the State as a Whole?

The 2022/23 grade 3 literacy screening assessment concordance samples used for equipercentile and predictive linking include 14,309 students from 211 schools within 69 districts and test scores from nine early literacy screening assessments (table 3). Additional data from FastBridge CBMreading (73 records) were omitted from the study because the number and distribution of valid test scores within the samples were considered unrepresentative of the population and unlikely to produce reliable results. Samples from the screening assessments listed in table 3 were considered sufficiently large and representative to produce reasonable preliminary estimates. However, there are some limitations to the representativeness of the samples.

Table 3. Number of Schools, Districts, and Grade 3 Students with 2022/23 EOY Literacy Screening Assessment Data Linked to MCAS

Early literacy screening assessment	Number of districts	Number of schools	Number of grade 3 students
Acadience Reading	2	3	137
aimswebPlus	3	3	169
DIBELS 8th Edition	18	42	3,046
mCLASS	19	46	2,834
FastBridge aReading	2	3	105
i-Ready Diagnostic	11	30	2,555
Lexia RAPID	1	4	468
Star Early Literacy	12	49	411
Star Reading	16	82	4,584
Total	84	262	14,309

Source: District-provided screening assessment data
Note: Districts and schools can be counted for more than one screening assessment.

The equipercentile method of linking depends entirely on the data in the sample; therefore, the accuracy of the linkages depends on how well the sample test score distribution approximates the population test score distribution. The screening assessment data were collected by convenience (i.e., screening assessments were chosen by districts that participated in a variety of state grants as opposed to being randomly assigned for research purposes). Different relationships between observed test score distributions and cut scores of interest across samples, for example, can lead to different linking results. Sample sizes are expected to continue to increase over time (as they have between 2020/21 and 2022/23) and linking estimates to improve.

For this study, the sampled literacy screening assessment scale score distributions are assumed to be representative of the scale score distribution from the population of Massachusetts grade 3 public school students. Comparing student demographic characteristics of the entire concordance sample with demographic characteristics of all Massachusetts public school students enrolled in grade 3 during the 2022/23 school year shows that Hispanic students, low income students, and English learner students are slightly overrepresented in the concordance sample, and Black and Asian students are slightly underrepresented (see table D.2 in appendix D). However, there is also variation between assessments (see tables D.3–D.11).

Discrepancies between the concordance samples and statewide population could affect the equipercentile linking results if they affect the test score distributions used to calculate linking estimates. As a further check on the reasonableness of the sample distributions, we examined the distribution of MCAS scores for the students taking each of the screening assessments and compared those test score distribution statistics with the known MCAS population statistics (see table D.12). Results suggest that linking estimates for benchmark cut scores should be reliable across the range of scores for the largest concordance samples. Estimates for benchmark cut scores near the middle of the MCAS scale are likely to be more reliable than linking estimates toward extremes of the scale for the smaller concordance samples. These estimates are discussed next. Results of predictive modeling used as a further check on equipercentile linking estimates that provided similar values can be found in appendix D.

How Do Literacy Screening Assessment Benchmark Scores From Different Screening Assessments Compare to Each Other in MCAS Terms?

This section describes results of the equipercentile method used to link literacy screening assessment scale scores to MCAS ELA scale scores and the associated findings. The equipercentile linking procedure (Kolen & Brennan, 2004) is a statistical method that assumes two test scores from the same group of examinees can be considered equivalent when the scores on each test have the same percentile rank. A description of the equipercentile linking procedure can be found in appendix D, along with the equations used to produce the results included in table 4 and figure 1 and results of predictive linking.

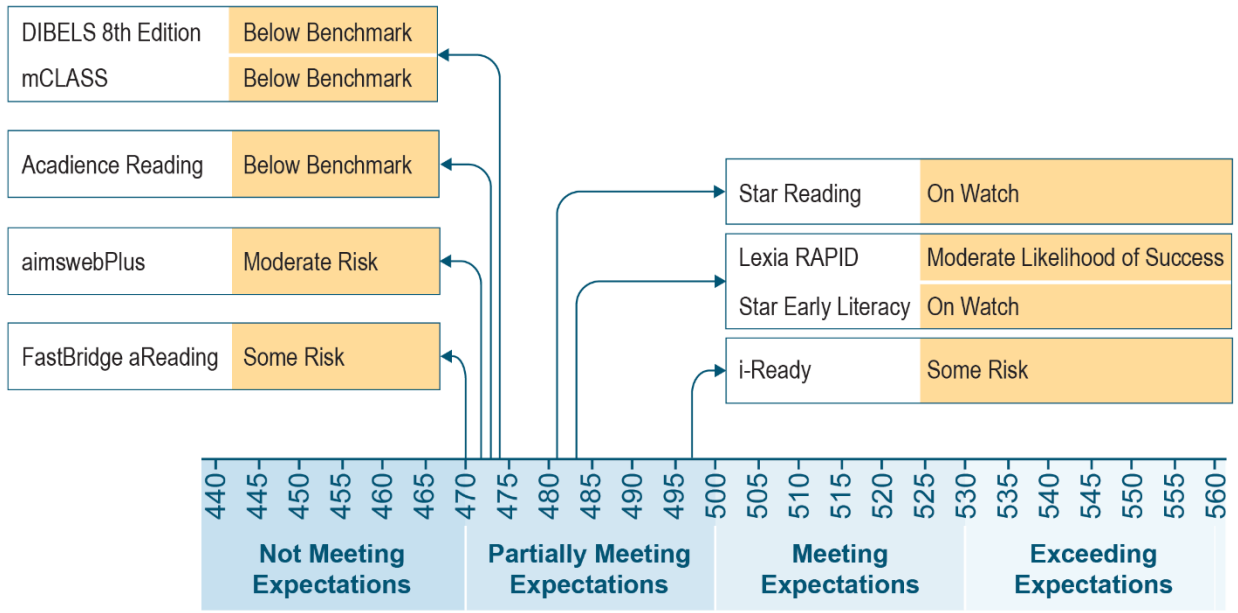
The linking procedure shows that publisher-provided benchmark categories representing performance significantly below benchmark at BOY and EOY largely link to MCAS scores in the Partially Meeting Expectations performance level, which ranges from 470 to 500 on the MCAS scale (table 4). One assessment benchmark falls into the Not Meeting Expectations level at BOY. Figure 1 provides a visual representation of the grade 3 EOY "significantly below" screening assessment benchmarks, showing their relationship to MCAS and to one another in MCAS terms. Their mapping to the Partially Meeting Expectations MCAS performance level is not surprising given that the benchmarks aim to identify students in need of additional support.

Table 4. Literacy 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	479	473	Partially meeting	Partially meeting
aimswebPlus	328	362	Moderate risk	476	472	Partially meeting	Partially meeting
DIBELS 8th Edition	314	424	Below benchmark	477	474	Partially meeting	Partially meeting
mCLASS	314	424	Below benchmark	474	474	Partially meeting	Partially meeting
FastBridge aReading	468	483	Some risk	469	470	Not meeting	Partially meeting
i-Ready	474	545	Some risk	483	497	Partially meeting	Partially meeting
Lexia RAPID	30	30	Moderate likelihood of success	494	483	Partially meeting	Partially meeting
Star Early Literacy	865	901	Intervention	478	475	Partially meeting	Partially meeting
Star Reading	865	901	Intervention	474	470	Partially meeting	Partially meeting

Source: District-provided screening assessment data and state-provided MCAS data

Figure 1. Most Grade 3 Literacy Screening Assessment EOY Benchmark Cut Scores Indicating Significant Risk Map to the MCAS Partially Meeting Expectations Level, Though There is Some Variation



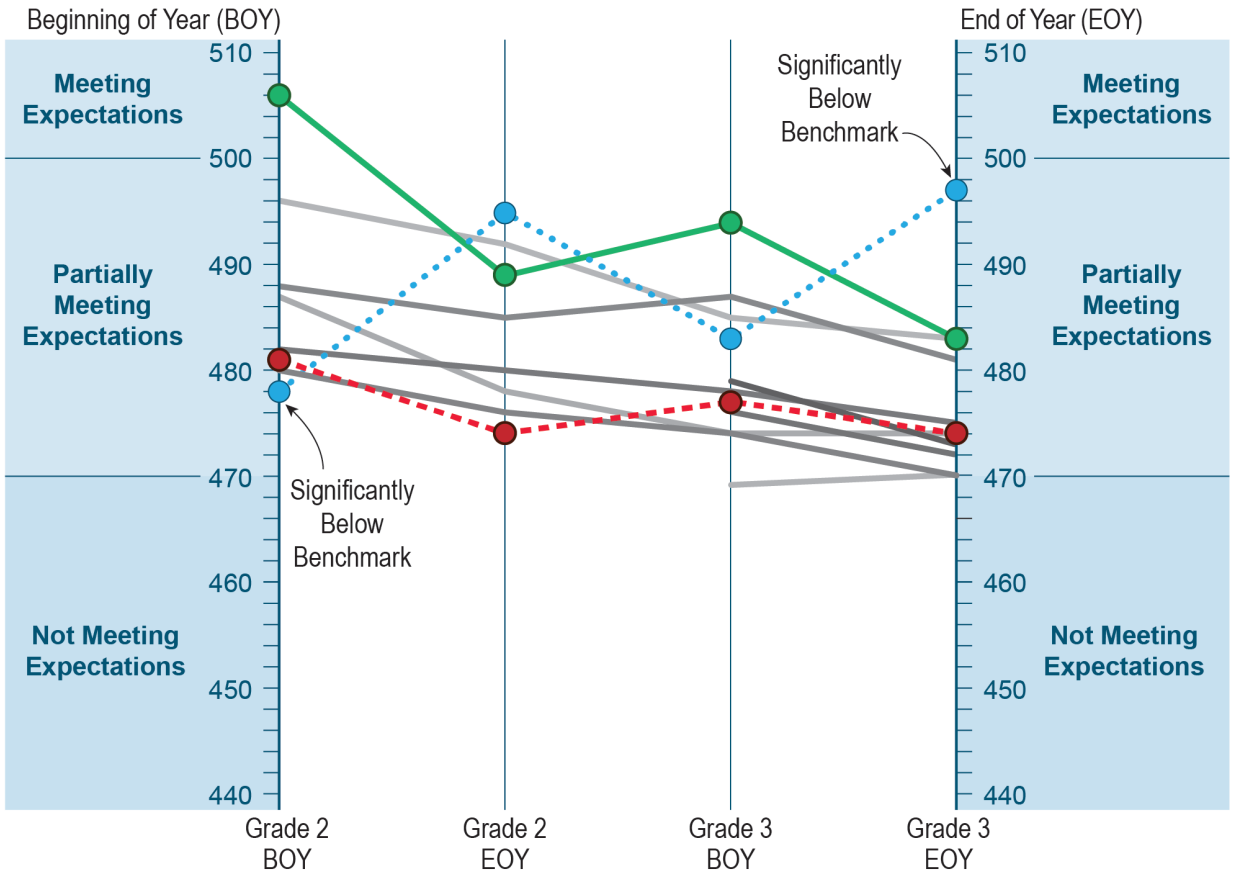
Source: District-provided screening assessment data and state-provided MCAS data
Note: Benchmarks indicate levels that will result in a student being classified as significantly below benchmark.

Prior analysis using 2021/22 data examined grade 3 EOY screening assessment benchmarks only. Using 2022/23 data, we also linked BOY benchmarks to MCAS (see table 4). To ensure maximum comparability when examining BOY and EOY benchmarks, we match students so that the equipercentile estimates represent the same students in each time period (i.e., linking estimates are based on a sample that includes only students who have BOY, MOY, EOY, and MCAS scores).

If the interpretation of scores across time periods for a given screening assessment is intended to be the same, we would expect each time period's score to map to the same MCAS score. However, analysis shows some variation within assessments where benchmarks link to MCAS between BOY and EOY, ranging from a decrease of 11 points from BOY to EOY to an increase of 14 points; other benchmarks stay relatively constant over time. Additionally, there can be changes in how screening assessment benchmarks map to MCAS across grades. Figure 2 presents these results: Each line represents a screening assessment mapped to MCAS at four different time points (grade 2 BOY and EOY and grade 3 BOY and EOY). The lines are highlighted to illustrate different patterns of change (e.g., increasing over time, decreasing over time, staying relatively constant over time).

These results mean that, depending on the benchmark patterns, students with the same skills at BOY and EOY might be classified differently. At one time period, they might be identified as significantly below benchmark and at another they might not.

Figure 2. Screening Assessment Benchmarks Indicating Significant Risk Vary Within and Across Grade Levels on the MCAS Scale, Meaning it May Be Harder or Easier to Be Identified as Significantly Below Benchmark at Different Times



Source: District-provided screening assessment data and state-provided MCAS data
Note: The figure includes all students with scores in BOY and EOY and MCAS scores in 2022/23. Each line represents a screening assessment mapped to MCAS at four different time points (grade 2 BOY and EOY and grade 3 BOY and EOY). Lines are highlighted to illustrate different patterns of change (e.g., increasing over time, decreasing over time, staying relatively constant over time).

Some change may be due to imprecision in linking estimates, but larger differences demonstrate variation in how benchmarks for different times of year were set. This variation may reflect differences in intended test purpose or use. Some assessments may prioritize measuring growth; others may prioritize growth toward a particular standard. As noted earlier, screening assessments vary in many ways, from content to administration to benchmark-setting procedures.

These differences do not indicate that one assessment is better than another but rather that users must be aware of how their assessments were designed, because those differences can result in differences in the numbers of students identified as significantly below benchmark in each time period, which may not reflect differences in student knowledge and skills.

Benchmark-setting differences can also affect how student growth is understood. For example, students in schools that use an assessment with a benchmark that shifts from lower to higher (i.e., easier to harder) on the MCAS scale between BOY and EOY may improve their performance relative to the BOY benchmark, but appear not to show progress relative to the EOY benchmark. For example, if an assessment's benchmark indicating significant risk maps to an MCAS score of 478 at BOY and a student scores below this level, then improves their skills by EOY, but the benchmark now maps to an MCAS score of 495 (like the blue dotted line in figure 2), they may still be classified as significantly below benchmark. Again, it is possible that changes like this are intentional, but nonetheless they do affect how users should interpret performance changes relative to benchmarks.

Conversely, students in schools that use an assessment with a benchmark that shifts from higher to lower (i.e., harder to easier) on the MCAS scale (like the red or green lines in figure 2 between BOY and EOY) may appear to grow out of the significantly below benchmark category by EOY while still performing at a skill level similar to the BOY benchmark. Across grades, some students may appear to have lost ground over the summer if benchmarks shift upward from the end of one grade to the beginning of another, whereas other students may appear to progress over the summer if benchmarks shift downward. Understanding how benchmarks within assessments compare from time period to time period is important for schools to take into account when reflecting on student performance and is especially important for analysis of growth within and across assessments.

Using national percentiles provides another way to examine screening assessment benchmarks over time and across assessments, which tells a similar story. As described in the "Available Data" section, benchmarks in DESE guidance for schools and districts focus on publisher-provided reporting categories (such as "well below benchmark") likely to be found in screening assessment reporting systems and used for identifying students in need of support. These screening benchmarks can vary from a national percentile of 13 (meaning about 13 percent of students nationally would be identified as significantly below benchmark and in need of Intensive Support) to a national percentile of 58, which would likely identify a much larger proportion of students (see figure 3 in report and table D.13 in appendix D).

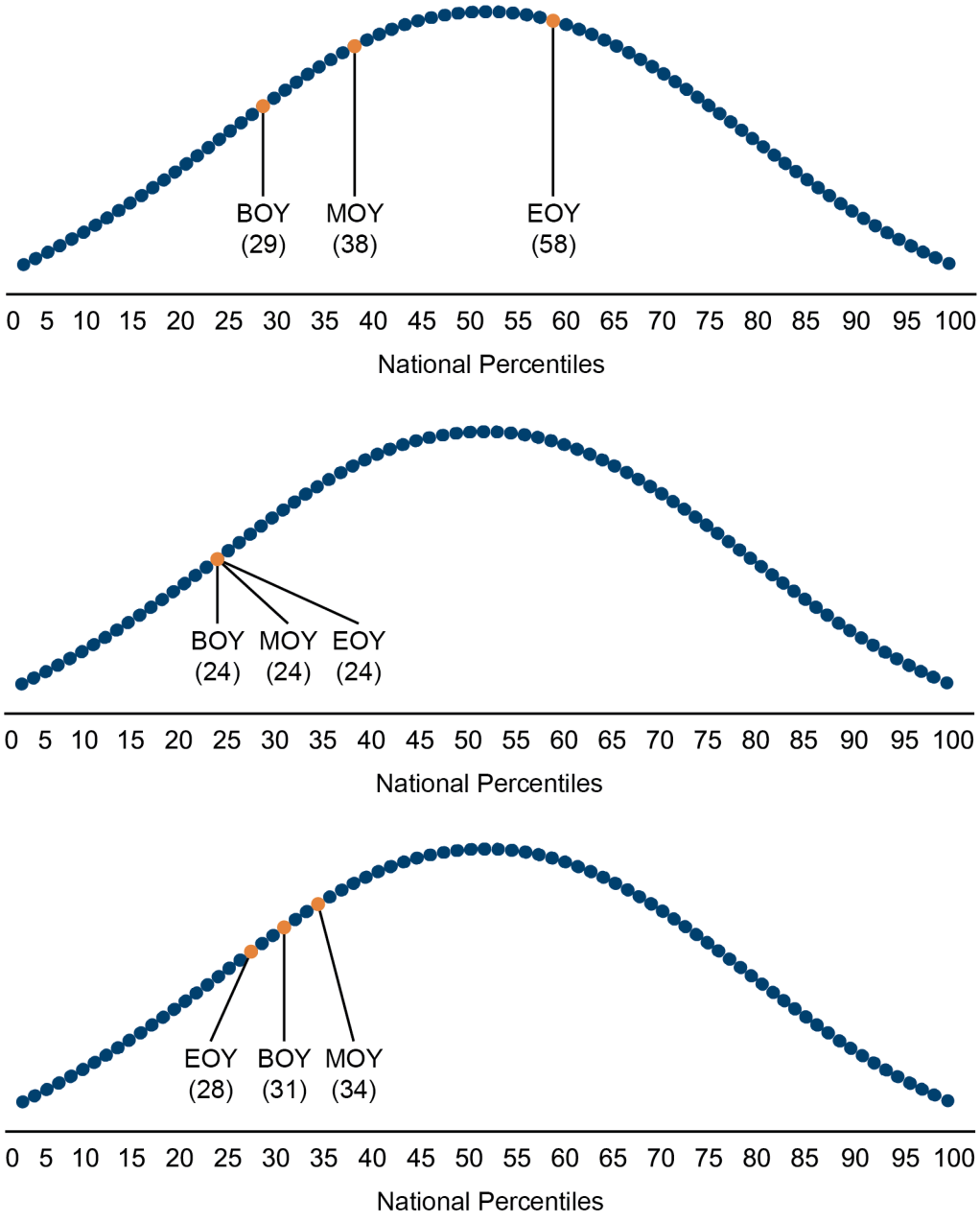
The percentile scores associated with being identified as at significant risk for some screening assessments shift between BOY, MOY, and EOY. Figure 3 shows BOY, MOY, and EOY benchmarks from several different assessments to illustrate how they can vary across assessments and across time periods within assessments. For example, in the first panel in figure 3, about 29 percent of students would be identified as significantly below benchmark at BOY, but 58 percent at EOY. In the middle panel, the benchmark indicating significant risk is the

same at each time period and would be expected to identify about 24 percent of students each time. Finally, the third panel shows benchmarks that would identify more students at MOY (34%) than BOY or EOY.

As noted earlier for MCAS, because some screening assessment benchmarks indicating significant risk do not map consistently to national percentile scores at each time period or across grade levels, students with the same skills might be classified differently at different time periods, and changes in the percentages of students identified as significantly below benchmark over time may be due in part to changes in the benchmarks themselves.

The percentages of students identified as at significant risk using publisher-provided (DESE-identified) benchmarks and the 25th percentile or below metric differ by time period and overall (see table E.6 in appendix E). For example, for one assessment, the percentage of students at significant risk decreases by 13 percentage points from BOY to EOY using publisher-provided benchmarks aligned to DESE guidance, but it does not change at all using the 25th percentile or below metric.

Figure 3. Grade 3 Screening Assessment Benchmarks Representing Significantly Below Benchmark Performance at BOY, MOY, and EOY With Corresponding National Percentiles



Source: District-provided screening assessment data
Note: Colored dots and lines indicate BOY, MOY, and EOY benchmarks for three different early literacy screening assessments (included in the analysis) to illustrate how they differ in the percentages of students identified as significantly below benchmark.

How Well Do Literacy Screening Assessment Benchmark Scores From Different Screening Assessments Discriminate Between Students Meeting and Not Meeting MCAS Proficiency Standards?

Prior and current analyses show that screening assessment scores in general discriminate reasonably well between students who will meet and not meet MCAS proficiency standards, with classification accuracy ranging from ~75 percent to ~89 percent, which can be considered in the good range (70%–90%). Full results of the ordinary least squares, or OLS, regression carried out in 2022/23 can be found in table D.16 in appendix D.

However, given the focus on students identified as performing significantly below benchmark, additional analysis was carried out to examine the diagnostic accuracy of those benchmark cut scores—that is, how well they discriminate between students who did and did not meet MCAS ELA proficiency standards in 2022/23. We used logistic regression to predict the probability of scoring at or above MCAS Meeting Expectations or Exceeding Expectations performance levels (see Equation 4 in appendix D for more details) and receiver operating characteristic (ROC) analysis to assess accuracy. Table 5 reports classification accuracy and area under the curve (AUC) for each assessment.[12]

Table 5. ROC Analysis Of Accuracy With Which Grade 3 BOY and EOY Literacy Screening Assessments Discriminate Between Students Who Meet and Do Not Meet Proficiency Standards on MCAS Grade 3 ELA by Assessment and Cut Score

Early literacy screening assessment	Benchmark	Classification accuracy (percent)		TP		FN		TN		FP		AUC
		BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	
Acadience Reading	Below Benchmark	55	53	0.97	1.00	0.03	0.00	0.24	0.18	0.76	0.82	0.88
aimswebPlus	Moderate Risk	60	56	0.97	0.99	0.03	0.01	0.29	0.18	0.71	0.82	0.89
DIBELS 8th Edition	Below Benchmark	64	64	0.97	0.98	0.03	0.02	0.4	0.37	0.60	0.65	0.83

[12] The AUC is equivalent to the probability that the literary screening assessment will rank a randomly chosen student who met proficiency standards higher than a randomly chosen student who did not (Carrington et al., 2021).

Early literacy screening assessment	Benchmark	Classification accuracy (percent)		TP		FN		TN		FP		AUC
		BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	
mCLASS	Below Benchmark	60	64	0.97	0.98	0.03	0.02	0.36	0.35	0.64	0.65	0.84
FastBridge aReading	Some Risk	69	68	0.98	1	0.02	0	0.21	0.15	0.79	0.85	0.90
i-Ready	Some Risk	74	86	0.98	0.83	0.02	0.17	0.57	0.87	0.43	0.13	0.93
Lexia RAPID	Moderate Likelihood of Success	79	68	0.83	0.95	0.17	0.05	0.78	0.56	0.22	0.44	0.90
Star Early Literacy	Intervention	73	70	0.99	0.95	0.01	0.05	0.69	0.64	0.31	0.36	0.93
Star Reading	Intervention	61	55	0.99	1	0.01	0	0.38	0.3	0.62	0.70	0.92

Source: District-provided screening assessment data and state-provided MCAS data
Notes: TP = True Positive (students at or above benchmark who met or exceeded expectations), FN = False Negative (students below benchmark who met or exceeded expectations), TN = True Negative (students below benchmark who did not meet expectations), FP = False Positive (students at or above benchmark who did not meet expectations), AUC = area under the curve

An AUC of 0.80–0.90 can generally be considered good, and 0.90–1.00 can be considered excellent. The AUC estimates for all early literacy screening assessments are in the good to excellent range (see table 5).

The additional columns in table 5 represent true positive (TP), false negative (FN), true negative (TN), and false positive (FP) predictions. TP rates (also known as sensitivity) describe the proportion of students accurately predicted to meet expectations on MCAS, and TN rates (also known as specificity) describe the proportion of students accurately predicted to not meet proficiency standards.

The FN and FP rates describe classification errors that accompany imperfect models. FN errors include students who meet proficiency standards despite being predicted not to meet them. FP errors include students who do not meet proficiency standards when they were predicted to meet them. On the one hand, from an educational policy perspective, FN errors may be preferable to FP errors because students who meet proficiency standards despite being predicted not to meet them can include cases of effective intervention. On the other hand, FP errors could be the costliest, as students need support but are not identified as needing it.

Focusing on publisher-provided benchmarks indicating significantly below benchmark performance shows, not surprisingly, that students identified as significantly below benchmark are rarely proficient on MCAS (see table 5). For example, 3 percent of students identified as significantly below benchmark in grade 3 at BOY using the DIBELS 8th Edition assessment were proficient on MCAS. Larger proportions of students who were not identified as significantly below benchmark (between 22 percent and 89 percent of students) were still not proficient on MCAS, which is also not surprising given that those benchmarks aim to identify students most in need of support, and, as shown in equipercentile and predictive linking, scores above those benchmarks often still fall in the MCAS Partially Meeting Expectations performance level. Table D.17 shows results for all benchmarks.

As seen in prior analysis, benchmarks that indicate little to no risk (typically "above benchmark" or similar) may still not accurately predict MCAS proficiency, likely due at least in part to differences between the skills included on MCAS and screeners. Among the most commonly used assessments, about 16 percent of students above those benchmarks did not meet expectations on MCAS. The National Center on Intensive Intervention's (n.d.) tools charts rate a screening tool highest when it has a TP rate of 0.70 or higher and a TN rate of at least 0.80, which implies an FN rate of less than 0.30 and an FP rate of less than 0.20—these metrics are commonly used by review committees to evaluate screening tools. Several assessments' benchmarks indicating low or no risk meet these standards at BOY (e.g., Acadience's Above Benchmark level, Fastbridge aReading's College Pathways level, i-Ready's No Risk level, Star Early Literacy's At/Above Benchmark performance level at BOY and EOY). Star Early Literacy's On Watch performance level also met this standard at BOY. Nearly all assessments accurately guarded against false negative outcomes (students who outperformed expectations on MCAS based on screener), but they were less accurate in preventing false positives (students who perform above benchmark but fail to meet expectations on MCAS), which is likely due to their intended use as screeners of reading risk rather than MCAS predictors.

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

How Many Benchmark Scores Are Available Overall and for Each Assessment?

Early literacy screening assessment data for the 2022/23 school year includes a total of 199,122 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 6).[13] This number includes all assessments, grades K–3, and all time periods, since most assessments are administered several times per year.[14] The number of records available in 2022/23 more than doubled from 2021/22, when over 81,000 records were available. Only about 10,000 records were available in 2020/21 (and thus initial reporting combined data from 2021/22 and 2020/21).

Key Findings

- Available data in 2022/23 doubled from 2021/22 and includes nearly 200,000 scores and 67,000 students, or about 26 percent of the state's K–3 student population, with scores from 308 schools and 88 districts.
- The most commonly used screening assessments in the sample were Star Early Literacy and Star Reading (32%), DIBELS 8th Edition (24%), mCLASS (18%), and i-Ready (17%).

The most commonly used assessments were Star Early Literacy and Star Reading (32%), DIBELS 8th Edition (24%), mCLASS (18%), and i-Ready (17%; see table 6). These screening assessments together represent about 91 percent of all observations. DIBELS 8th Edition and mCLASS, both of which are based on the same assessment tasks, together made up approximately 42 percent of all scores. Several assessments with data available in 2020/21 and/or 2021/22 had no records for 2022/23 (MAP Growth, ISIP ER, MAP Reading Fluency, and Star Early Literacy Spanish) and were therefore not included in 2022/23 calculations. Conversely, though no longer an approved screening assessment, aimswebPlus scores for 2022/23 were available and included in analysis. Additionally, about 12,600 Star CBM observations were excluded as the assessment only provides scores for each specific subtest/skill (e.g., letter naming)—that is, there is no overall description of student reading risk.

Analyses using the 25th percentile or below metric include a smaller number of scores than those using publisher-provided benchmarks (174,714 compared to 199,122) because not all files collected from publishers or districts contained national percentile scores.

[13] 23,698 records were removed from the data due to missing benchmark scores or other data issues.

[14] 7,519 students have multiple scores from different screening assessments in the same time period in the 2022/23 school year.

Table 6. Number of Literacy Screening Assessment Benchmark Scores by Year and Assessment for Grades K–3

Early literacy screening assessment	Number in 2021/22	Percent in 2021/22	Number in 2022/23	Percent in 2022/23	Number in 2022/23 (25th percentile metric)	Percent in 2022/23 (25th percentile metric)
Acadience Reading	0	0%	4,615	2%	0	0%
aimswebPlus	0	0%	1,951	1%	1,951	1%
DIBELS 8th Edition	21,686	26%	46,938	24%	35,058	20%
mCLASS	9,383	11%	36,131	18%	36,129	21%
EarlyBird	N/A	N/A	2,689	1%	0	0%
FastBridge aReading	1,342	2%	756	<1%	756	<1%
FastBridge CBMreading	1,029	1%	778	<1%	778	<1%
FastBridge earlyReading	418	1%	1,168	1%	1,168	1%
i-Ready	16,426	20%	34,784	17%	34,784	20%
ISIP ER	1,902	2%	0	0%	0	0%
Lexia RAPID	6,203	8%	5,222	3%	0	0%
MAP Growth	1,908	2%	0	0%	0	0%
MAP Reading Fluency	314	<1%	0	0%	0	0%
Star Early Literacy	15,869	19%	34,602	17%	34,602	20%
Star Early Literacy Spanish	2,247	3%	0	0%	0	0%
Star Reading	3,126	4%	29,488	15%	29,488	17%
Total	81,853	100%	199,122	100%	174,714	100%

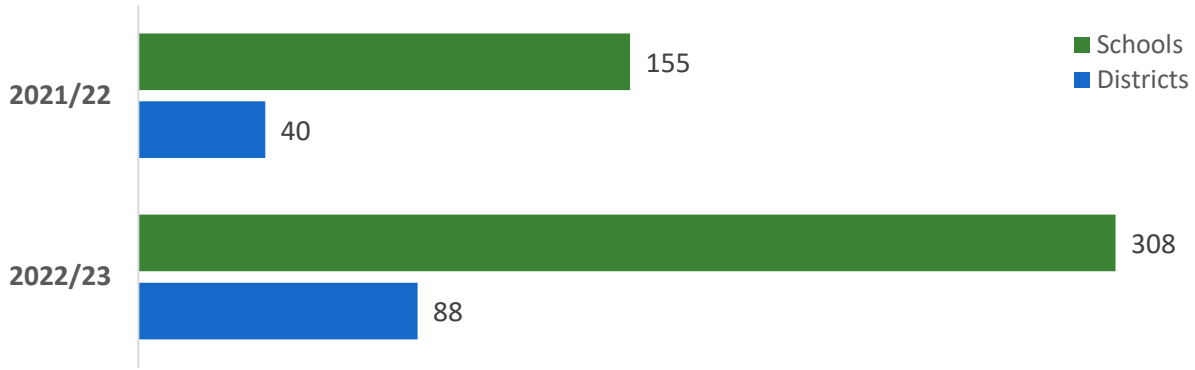
Source: District-provided screening assessment data
Note: EarlyBird data were not included in 2021/22 analysis and are therefore not shown in this table.

How Many Students, Schools, and Districts Are Represented by the Data?

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 199,122 observations shown in table 6 represent 67,090 students across 88 districts and 308 schools in 2022/23, about double the numbers of students, schools, and districts included in 2021/22 (figure 4). The median number of students per district was 546, with 10 districts comprising about 40 percent of the data overall (see figure 4, which shows numbers of schools and districts in the sample in 2022/23 and 2021/22).

Figure 4. Between School Years 2021/22 and 2022/23, the Numbers of Schools and Districts Included in Analysis Approximately Doubled



Source: District-provided screening assessment data
Note: 272 schools across 78 districts are represented in data using the 25th percentile or below metric.

The 67,090 students represent about 26 percent of the total grade K–3 student population in the state in 2022/23, up from 12 percent in 2021/22.[15] About 23 percent of students (15,588 students) across the sample have at least one benchmark score in 2021/22 and 2022/23, and 0.9 percent of students have scores from 2020/21, 2021/22, and 2022/23.

Table 7 provides the number of students with data from each literacy screening assessment in either 2021/22 or 2022/23. Some students took multiple screening assessments within the school year, resulting in larger totals than the actual number of unique students; most of these instances were due to the students taking multiple Star assessments (Early Literacy and Reading) or DIBELS 8th Edition and i-Ready.

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

Table 7. Number of Students by Year and Assessment

Early literacy screening assessment	Number in 2021/22	Percent in 2021/22	Number in 2022/23	Percent in 2022/23	Number in 2022/23 (25th percentile metric)	Percent in 2022/23 (25th percentile metric)
Acadience Reading	0	0%	1,591	2%	0	0%
aimswebPlus	0	0%	656	1%	656	1%
DIBELS 8th Edition	8,362	24%	17,533	23%	12,676	19%
mCLASS	3,359	10%	13,882	18%	13,882	21%
EarlyBird	N/A	N/A	1,279	2%	0	0%
FastBridge aReading	629	2%	269	<1%	269	<1%
FastBridge CBMreading	619	2%	364	<1%	364	1%
FastBridge earlyReading	285	<1%	406	1%	406	1%
i-Ready	6,627	19%	12,622	17%	12,622	19%
ISIP ER	675	2%	0	0%	0	0%
Lexia RAPID	2,171	6%	1,838	2%	0	0%
MAP Growth	992	3%	0	0%	0	0%
MAP Reading Fluency	164	<1%	0	0%	0	0%
Star Early Literacy	7,671	22%	13,373	18%	13,373	20%
Star Early Literacy Spanish	1,352	4%	0	0%	0	0%
Star Reading	1,738	5%	11,428	15%	11,428	17%
Total	34,644	100%	75,241	100%	65,676	100%

Source: District-provided screening assessment data
Notes: Includes students with at least one benchmark score/level within each year. Students may take more than one assessment in a school year, and the total includes those duplicated students. EarlyBird data were not included in 2021/22 analysis and are therefore not shown in this table.

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?

Compared to the state as a whole, the background characteristics of students in the screening assessment sample are relatively similar, with the proportions of students with various background characteristics differing at most by 4 percentage points.

Key Findings

- The demographic composition of the early literacy screening assessment data sample is similar to the state, suggesting that results are broadly generalizable.
- Background characteristics of students differ by screening assessment.

The early literacy screening assessment data have a somewhat smaller percentage of grade 3 students than the state overall (21 percent compared to 25 percent; see figure 5), and somewhat larger percentages of Hispanic and English learner students (22 percent compared to 18 percent for English learner students and 28 percent compared to 25 percent for Hispanic students). The sample used for analyses using the 25th percentile or below metric is similar in terms of its demographic composition to the sample used for the benchmark level analyses. Compared to the benchmark level sample, the sample for the 25th percentile or below metric has slightly fewer kindergarten students (24 percent compared to 22 percent) and slightly more grade 1–3 students. All differences aside from kindergarten are only 1 percent.

About 45 percent of the screening assessment sample is classified as low income (compared to 44 percent in the state) and 18 percent as receiving special education services (compared to 16 percent in the state).[16] Additionally, 79 percent are White, 28 percent are Hispanic, 14 percent are Black, 9 percent are Asian, 5 percent are American Indian or Alaskan Native, and less than 1 percent are Native Hawaiian or Pacific Islander. Overall, the racial/ethnic distribution of the sample is generally similar to that of the state, with slightly larger percentages of American Indian or Alaskan Native and Hispanic students and a slightly smaller percentage of Black students.

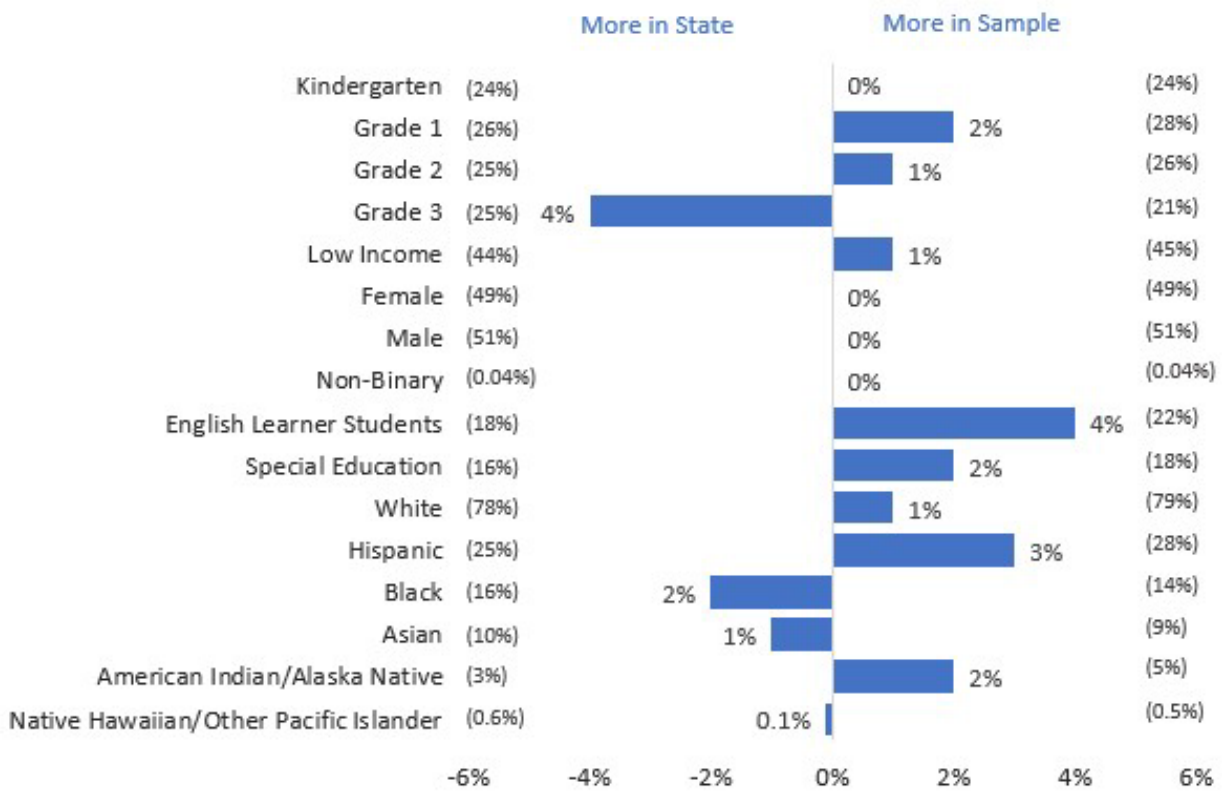
Appendix E provides information on the background characteristics of students by screening assessment, which suggests differences in the schools and districts choosing different screening assessments (see tables E.1.1 and E.1.2). For example, Star Early Literacy has about twice as many English learner students as the state overall, whereas other screening assessments (e.g., EarlyBird) include fewer. Appendix E also provides a comparison of the MCAS scores of grade 3 students across screening assessments, showing that grade 3 students in the sample have

[16] A student was classified as a special education student if they were identified as receiving special education services at the time of the SIMS collection or were previously identified during the same school year.

slightly smaller percentages of students meeting MCAS proficiency standards than the state average (see table E.2).

Overall, although some caution should be used, the relatively small magnitude of observed differences between students included in the literacy screening assessment data and the state (4 percent or less) suggest that the results are broadly generalizable to the state population of K–3 students.

Figure 5. Screening Assessment Data Include Somewhat Fewer Grade 3 Students, More English Learner Students, and More Hispanic Students Compared to K–3 Students in the State Overall



Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Percentages in parentheses represent the overall percentage in the state or screening assessment data sample and percentages next to the bars represent the differences between the state and the sample (e.g., 26 percent of students in the state are in grade 1 compared to 28 percent in the sample, or there are 2 percent more grade 1 students in the sample than in the state overall). The percentage of students classified as special education students includes those classified as special education at the time of SIMS reporting and those who were previously classified during the current school year. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

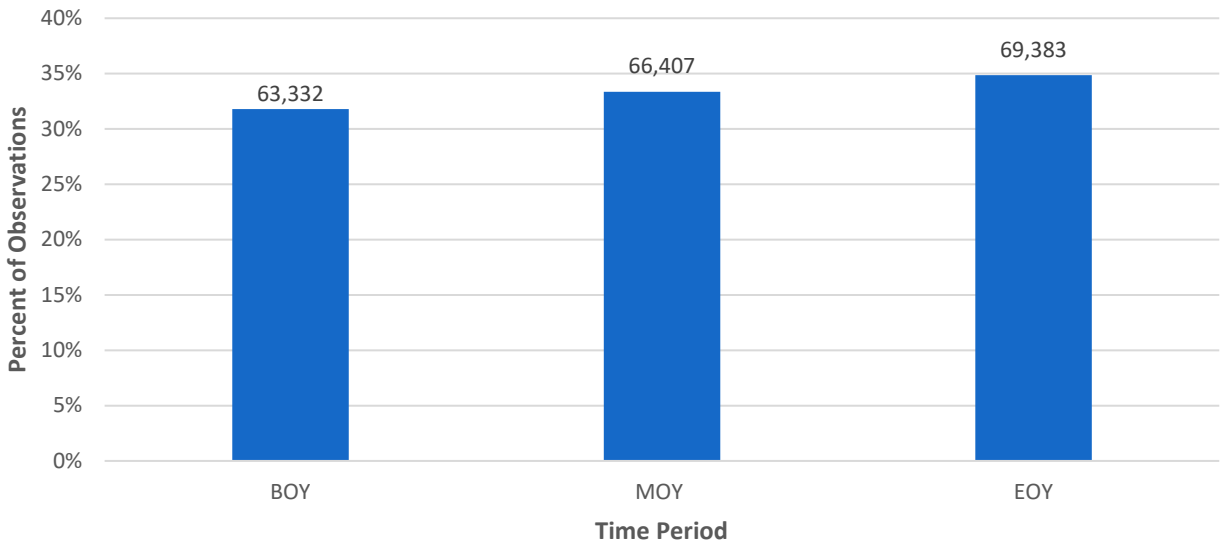
How Many Benchmark Scores Are Available by Time Period?

Of the 199,122 records with benchmark scores in 2022/23, about 32 percent were from the BOY time period, 33 percent were from the MOY time period, and 35 percent were from the EOY time period (figure 6). Each of the screening assessments, except EarlyBird and FastBridge CBMreading, had a relatively similar number of data points across time periods (see table E.3 in appendix E). For Early Bird and FastBridge CBMreading, only 10 percent and 8 percent of their observations, respectively, were from the BOY time period; the MOY and EOY time periods had roughly the same number of observations.

Key Findings

- The number of records with benchmark scores is somewhat greater at EOY than MOY or BOY (35% in EOY, 33 percent in EOY, and 32% in BOY).
- 75 percent of students have a benchmark score in all three time periods for the same assessment; 14 percent have a score in two of the three time periods; 11 percent have a score in a single time period.

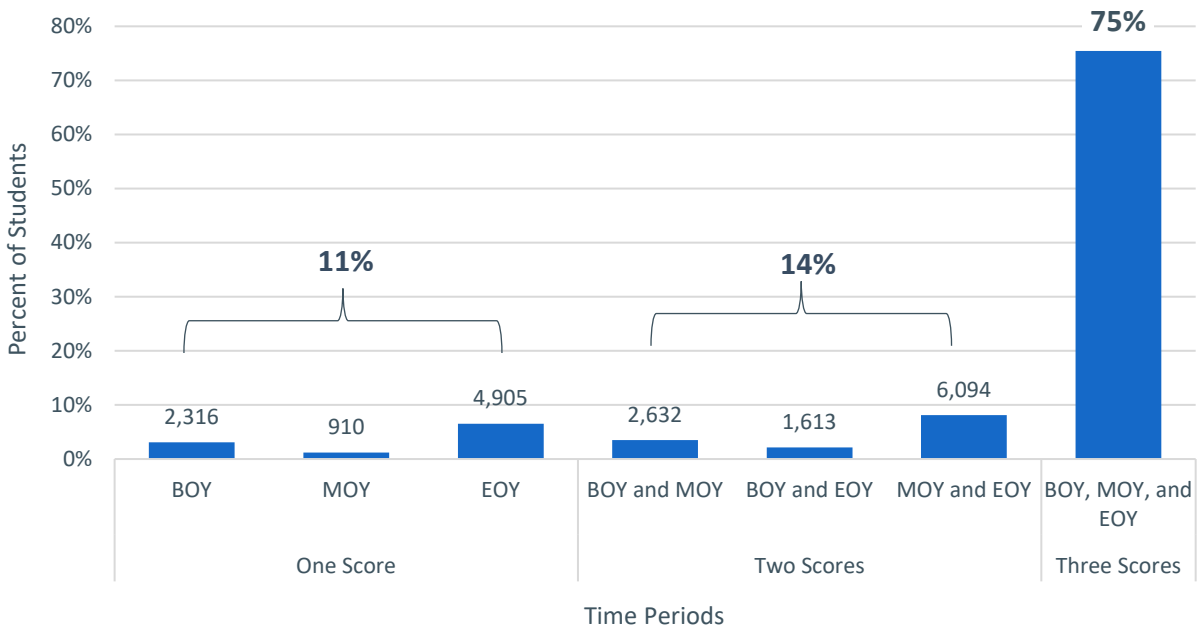
Figure 6. BOY Time Period Has Smallest Number of Scores



Source: District-provided screening assessment data
Notes: The distribution of scores across time periods in the analytic sample for the 25th percentile or below metric matches that of the benchmark analytic sample (32 percent in BOY, 33 percent in MOY, and 35 percent in EOY). Some students may appear multiple times per time period if they were administered multiple screening assessments.

Three quarters of the K–3 students in the screening assessment data (75%) had benchmark scores from the same screening assessment in all three time periods (figure 7), an increase from 2020/21 and 2021/22 when only about half of students had scores in all three time periods. About 14 percent had benchmark scores only from two time periods, and 11 percent had benchmarks only from a single time period. Of the students with two benchmark scores, most had MOY and EOY scores (see figure 7). Most students with only one benchmark had that score at EOY.

Figure 7. Seventy-Five Percent of Students Have Benchmark Scores for All Three Time Periods



Source: District-provided screening assessment data
Note: Students with multiple screening assessments per time period are counted for each screening assessment. The distribution of scores across time periods in the analytic sample for the 25th percentile or below metric roughly matches that of the benchmark analytic sample (77 percent of students had scores in all three time periods, 13 percent had scores in only two time periods, and 11 percent had scores in only one time period).

Students with all three scores included a higher percentage of White students and lower percentages of low income, English learner, Black, Hispanic, Asian, and American Indian/Alaskan Native students and students receiving special education services compared to students with only one or two scores (see table E.5 in appendix E). For example, 20 percent of students with three scores were English learners compared to 26 percent of students with two scores and 30 percent of students with one score.

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 stakeholders 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 questions of student performance:

- How many students were identified as significantly below benchmark by time period, grade, and student group?
- How many sutdents were idenitfied as significantly below benchmark two or three times overall and by grade and student group?
- How do intersecting student and school background factors interact and relate to the likelihood of students being identified as significantly below more than once?

This section first describes students "significantly below benchmark," following DESE guidance[17] (refer to table 2 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 that 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.

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

How Many Students Were Identified As Significantly Below Benchmark by Time Period, Grade, and Student Group?

Key Findings

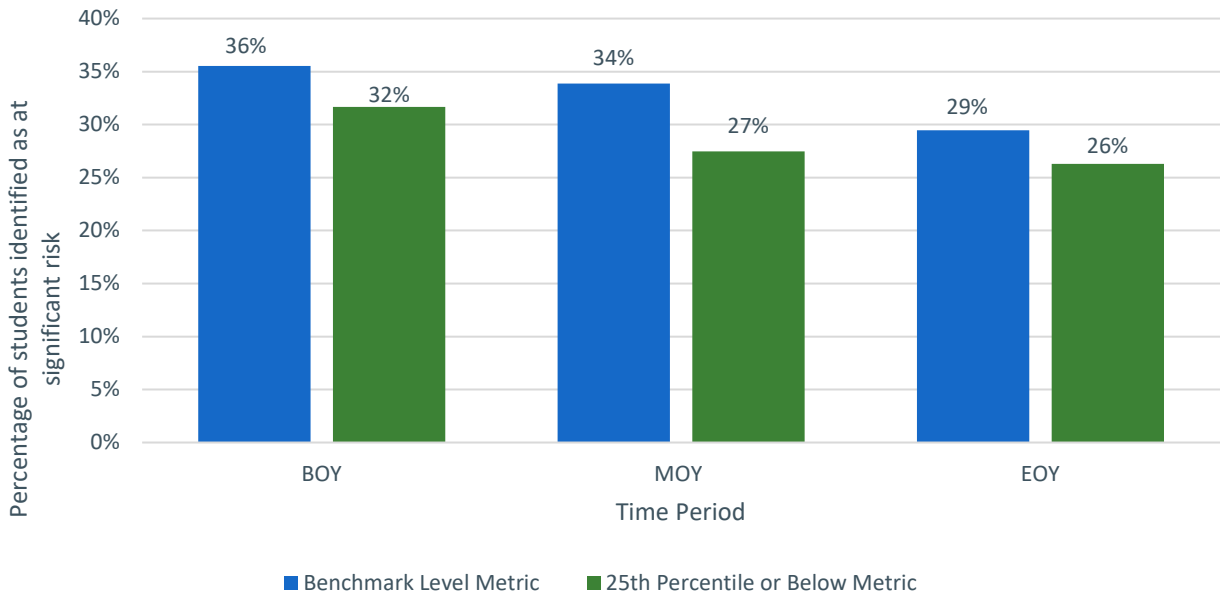
- As described in the introduction, 2022/23 analysis focuses on performance of students identified as "significantly below benchmark" rather than "below benchmark" as in prior analysis from 2021/22 and earlier. Despite this change, patterns of performance and progress from prior reporting remain the same.
- Across time periods and grade levels, 45 percent of students were ever classified as significantly below benchmark level.
- Using the 25th percentile or below provides a similar picture of performance to publisher-provided benchmarks, but specific percentages are generally smaller.
- Depending on the time period, between 29 percent and 36 percent fell significantly below benchmark. As in prior reporting, more students were classified as significantly below benchmark at BOY than at MOY or EOY.
- Kindergarten students showed greater change in the percentages of students identified as at significant risk between BOY and EOY and across years than students in grades 1 through 3; grade 3 students showed no change from BOY to EOY. This pattern was also observed in prior analysis.

As in prior reporting, in each time period, students from low income backgrounds; English learner students; students receiving special education services; and Black, Hispanic, and American Indian/Alaskan Native students were more likely to be identified as at risk than their peers not in those groups.

Using publisher-provided benchmarks aligned to DESE guidance, the percentage of students identified as significantly below benchmark ranged from 29 percent to 36 percent of students, depending on the time period (figure 8). As shown in prior reporting using 2020/21 and 2021/22 data, more students were classified as significantly below benchmark at BOY (36%) than MOY (34%) or EOY (29%). Across time periods, 45 percent of students were classified as significantly below benchmark levels of performance. Using the 25th percentile or below metric provides a similar picture of performance, though the specific percentages vary. With the 25th percentile or below metric, 28 percent of students were identified as at significant risk across time periods, with 32 percent at BOY and 26 percent at EOY (see also table E.6 in appendix E, which compares the percentage of students at significant risk according to the benchmark level

and the 25th percentile or below metrics). Figure 8 represents students with scores in any number of time periods.

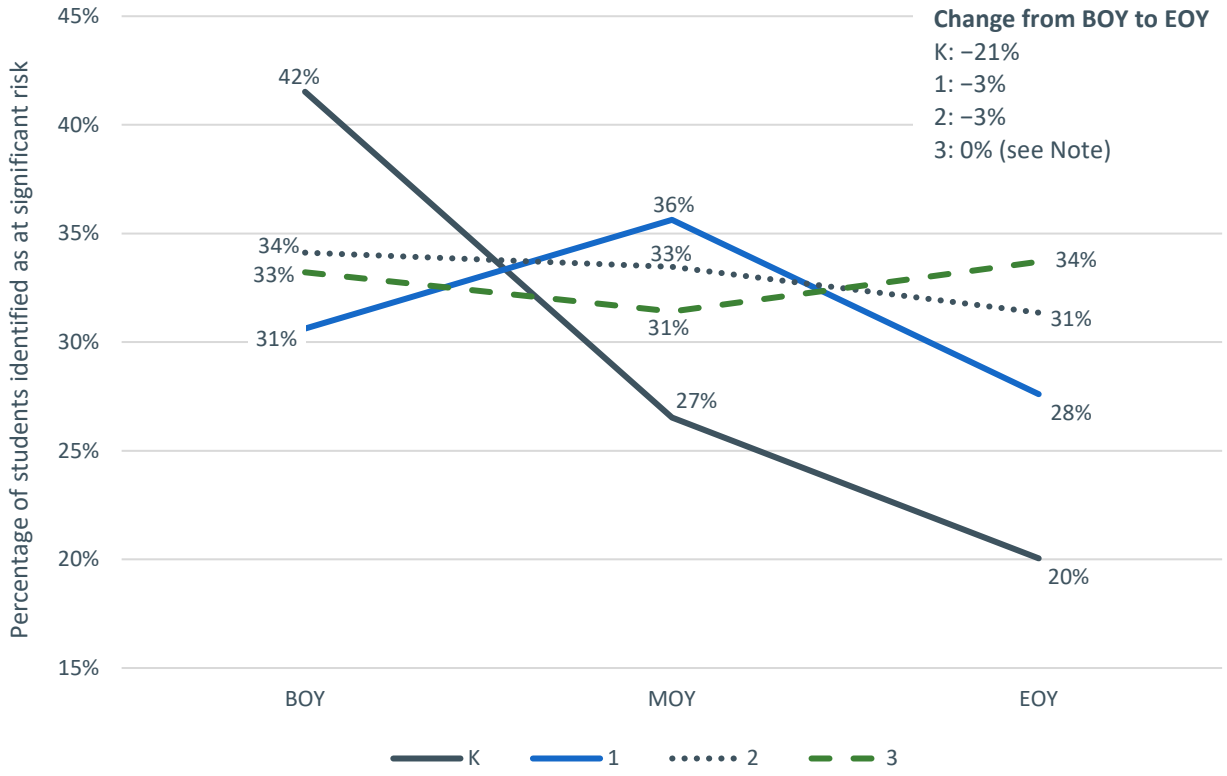
Figure 8. Fewer students were identified as at Significant Risk at EOY Compared to Earlier Time Periods



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Includes students with any number of scores. Percentages for students with three scores are BOY: 35 percent, MOY: 32 percent, EOY: 28 percent.

Among students who have scores in all three time periods, the percentages of students in grades K–2 identified as significantly below benchmark decreases from BOY to EOY, although there was greater change among students in kindergarten than in other grade levels (figure 9). This pattern was also observed in prior reporting. Although 42 percent of kindergarten students were classified as significantly below benchmark at BOY, that percentage was 20 percent at EOY, a 21 percentage point decrease. The percentage of students classified as significantly below benchmark decreased by 3 percentage points for grades 1 and 2, and there was no change in the percentage for grade 3 students. By EOY, the percentage of students in grades 1 through 3 identified as significantly below benchmark was smallest for grade 1 and greatest for grade 3.

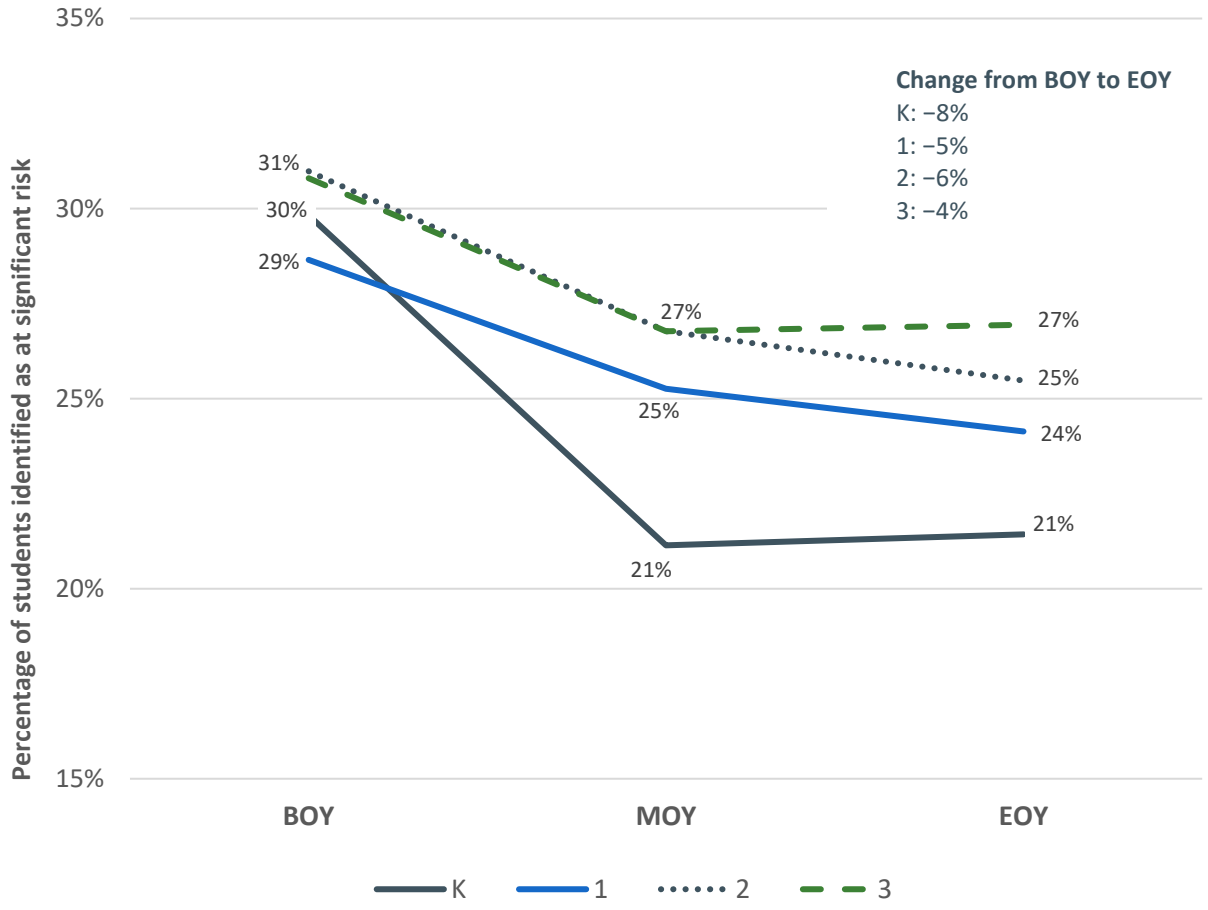
Figure 9. The Percentage of Students Identified as Significantly Below Benchmark Decreased From BOY to EOY for Students in Grades K–2, With the Largest Change Among Kindergarten Students



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Includes only students with three scores (i.e., the same students are included in each time period). The change in the percentage of students significantly below benchmark in grade 3 from BOY to EOY rounds to zero.

Using the 25th percentile or below metric, kindergarten students still show the greatest decrease in the percentage of students classified as at significant risk from BOY to EOY, but the change is much smaller than when using publisher-provided benchmarks (an 8 percentage point decrease compared to 21%). The percentages of students in grades 1 through 3 classified as at significant risk all decreased from 4 percent to 6 percent between BOY and EOY (figure 10). These differences in percentages of students identified as at significant risk between BOY and EOY using publisher-provided benchmarks and the 25th national percentile may be due at least in part to differences in how publishers set benchmarks within and across grade levels. (See the "Comparing Screening Assessment Benchmarks" section for more information.)

Figure 10. The Percentage of Students at the 25th Percentile or Below Decreased From BOY to EOY for all Grade Levels, With the Largest Change Among Kindergarten Students



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Includes only students with three scores (i.e., the same students are included in each time period).

In each time period, students from low income backgrounds; English learner students; students receiving special education services; and Black, Hispanic, and American Indian/Alaskan Native students were more likely to be identified as at risk than their peers not in those groups (table 8). For example, 51 percent of low income students were significantly below benchmark compared to 22 percent of non-low income students, meaning the relative risk or likelihood of low income students being identified as significantly below benchmark was more than twice that of non–low income students.

Low income students and English learner students showed greater decreases in percentages of students identified as significantly below benchmark from BOY to EOY but were still about twice as likely to be identified as significantly below benchmark compared to peers not in those groups.

Students receiving special education services showed less improvement over time than their peers. Hispanic, American Indian/Alaskan Native, and Native Hawaiian/Pacific Islander students showed greater change from BOY to EOY than White, Asian, or Black students but were still more frequently identified as significantly below benchmark than their peers not in those groups.

Table 8. Percentage of Students Identified as Significantly Below Benchmark and Relative Risk of Being Significantly Below Benchmark at BOY, MOY, and EOY by Student Group

Demographic	Significantly below benchmark at BOY (percent)	Significantly below benchmark at MOY (percent)	Significantly below benchmark at EOY (percent)	Change BOY to EOY (percent)	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Low income	51%	46%	41%	−10%	2.31*	2.14*	2.19*
Non–low income	22%	21%	19%	−3%	N/A	N/A	N/A
Female	34%	31%	27%	−7%	0.95*	0.95*	0.93*
Male	35%	33%	29%	−6%	N/A	N/A	N/A
Nonbinary	--	--	--	--	†	†	†
English learner	65%	60%	54%	−11%	2.45*	2.41*	2.49*
Non–English learner	27%	25%	22%	−5%	N/A	N/A	N/A
Students receiving special education services	58%	59%	54%	−4%	1.94*	2.21*	2.33*
Students not receiving special education services	30%	27%	23%	−7%	N/A	N/A	N/A
White	33%	31%	27%	−6%	0.81*	0.83*	0.80*
Hispanic/Latino	58%	52%	47%	−11%	2.22*	2.09*	2.17*
Black	42%	39%	35%	−6%	1.24*	1.27*	1.29*

Demographic	Significantly below benchmark at BOY (percent)	Significantly below benchmark at MOY (percent)	Significantly below benchmark at EOY (percent)	Change BOY to EOY (percent)	Relative risk at BOY	Relative risk at MOY	Relative risk at EOY
Asian	19%	19%	16%	−3%	0.53*	0.58*	0.55*
American Indian/Alaskan Native	67%	56%	55%	−12%	2.04*	1.83*	2.03*
Native Hawaiian/ Pacific Islander	47%	42%	35%	−12%	†	†	†
Total	**35%**	**32%**	**28%**	**−7%**	**N/A**	**N/A**	**N/A**

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). The relative risk ratio indicates the likelihood of a student group classified as significantly below benchmark compared to students not in that group (e.g., students from low income families were 2.31 times more likely than students who are not from low income families to be classified as significantly below benchmark at BOY). Includes only students with three scores. Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).
† Not computed because groups are 5 percent or less of the sample.
*** Indicates statistical significance at an alpha level less than .05.**

The same trends by student group are observed when using the 25th percentile or below metric (see table E.7 in appendix E). Low income students; English learner students; students receiving special education services; and Hispanic, Black, American Indian/Alaskan Native, and Native Hawaiian/Pacific Islander students are more likely than their peers to be identified as at significant risk in any time period. Low income and English learner students showed greater decreases in the percentage of students at significant risk over time, and students receiving special education services showed less improvement from BOY to EOY than their peers. However, across each student group the percentage of students at significant risk is typically lower in each time period when using the 25th percentile or below metric than when using the publisher-provided benchmarks.

How Many Students Were Identified as Significantly Below Benchmark Two or Three Times Overall and by Grade and Student Group?

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

In the previous section, we summarize performance by grade and student groups for each time period. In the following sections, we focus on 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. Students may be missing scores in given time periods for many reasons, such as moving, being absent during test administration, or not all scores being provided for analysis. About 75 percent of students had scores in all three time periods, and 89 percent had scores in at least two time periods, so focusing on students identified as significantly below benchmark excludes about 11 percent of students (because those students only had one score; see figure 7).

Key Findings

- Among students with scores in at least two time periods, about 31 percent were significantly below benchmark multiple times during the school year.
- Low income students, English learner students, students receiving special education services, Black students, and Hispanic students were again more likely than their peers not in those groups to be classified as significantly below benchmark multiple times during the school year. Asian students were less likely than their peers to be classified as such.

In 2022/23, among students with at least two scores, about 31 percent of students were identified as significantly below benchmark two or more times (table 9).

Table 9. 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* = 3,127)—2 times	31% (*n* = 2,657)—2 times
Three benchmarks available	31% (*n* = 17,551)—2 or 3 times 12% (*n* = 6,985)—2 times 19% (*n* = 10,566)—3 times	25% (*n* = 12,767)—2 or 3 times 8% (*n* = 4,152)—2 times 17% (*n* = 8,615)—3 times
Total	**31% (*n* = 20,678)**	**26% (*n* = 15,424)**

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Total for each column includes the number of students significantly below benchmark two times (regardless of their number of benchmarks available) and the number of students significantly below benchmark three times.

Fewer students overall were identified as at significant risk two or more times during the school year when using the 25th percentile or below metric than when using publisher-provided benchmarks (26% compared to 31%). Among students with scores in two benchmark periods, the two metrics identified similar percentages of students; however, among students with scores in all three time periods, the 25th percentile or below metric identified fewer students (8% compared to 12% were below two times and 17% compared to 19% were below all three times).

Of the students who had at least two scores, fewer kindergarten and grade 1 students were significantly below benchmark multiple times than were students in grades 2 and 3. As was the case in prior reporting, low income students; English learner students; students receiving special education services; and Black, Hispanic, and American Indian/Alaskan Native students were more likely than their peers not in those groups to score significantly below benchmark multiple times during the school year (table 10). For example, 44 percent of low income students were significantly below benchmark at least twice during the school year compared to 19 percent of non–low income students. The relative risk of low income students being classified as significantly below benchmark multiple times during the school year was 2.4 times that of non–low income students. 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. Across the five most commonly used screening assessments (DIBELS 8th Edition, i-Ready, mCLASS, Star Early Literacy, and Star Reading), the percentages of students classified

as significantly below benchmark multiple times during the school year vary from 24 percent to 41 percent (see table E.8 in appendix E); these differences may be due in part 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.

Using the 25th percentile or below metric shows the same patterns for student groups, though with fewer students identified as at significant risk multiple times throughout the school year.

Some additional differences exist, however, when examining the results by assessment. Both metrics identify about the same number of students for the Star Early Literacy and Star Reading assessments, but fewer students are identified with DIBELS 8th Edition, mCLASS, and i-Ready assessments when using the 25th percentile or below metric, and more are identified for the aimswebPlus and FastBridge assessments (see table E.8).

Table 10. Percentage of Students at Significant Risk Multiple Times by Student Group

Demographic	Significantly below benchmark two or more times (percent)	Relative risk of significantly below benchmark two or more times	25th percentile or below two or more times (percent)	Relative risk of being 25th percentile or below two or more times
Kindergarten	25%	N/A	22%	N/A
Grade 1	31%	N/A	25%	N/A
Grade 2	34%	N/A	29%	N/A
Grade 3	32%	N/A	29%	N/A
Low income	45%	2.35*	40%	2.83*
Non–low income	19%	N/A	14%	N/A
Female	30%	0.93*	25%	0.91*
Male	32%	N/A	28%	N/A
Nonbinary	--	†	--	†
English learner	59%	2.56*	55%	3.08*
Non–English learner	23%	N/A	18%	N/A

Demographic	Significantly below benchmark two or more times (percent)	Relative risk of significantly below benchmark two or more times	25th percentile or below two or more times (percent)	Relative risk of being 25th percentile or below two or more times
Students receiving special education	57%	2.26*	51%	2.46*
Students not receiving special education	25%	N/A	21%	N/A
White	29%	0.81*	24%	0.73*
Hispanic/Latino	51%	2.22*	47%	2.61*
Black	38%	1.27*	33%	1.32*
Asian	17%	0.54*	13%	0.48*
American Indian/Alaskan Native	58%	1.98*	58%	2.37*
Native Hawaiian/ Pacific Islander	41%	†	36%	†
Total	**31%**	**N/A**	**26%**	**N/A**

Source: District-provided screening assessment data
Notes: 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 categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).
† Not computed because groups are 5 percent or less of the sample.
* Indicates statistical significance at an alpha level less than .05.

How Do Intersecting Student and School Background Factors Interact and Relate to the Likelihood of Students Being Identified as Significantly Below Benchmark More Than Once?

Outcome data suggest that the current educational system often does not provide adequate support for students from historically marginalized groups, such as those learning English or students with disabilities. Early literacy screening assessments aim to identify students who are not on track to become successful readers and who require additional support to do so.

The prior section provides descriptive information about the percentages of students with different background characteristics who were identified as significantly below benchmark at each time period and multiple times during the school year. This analysis shows what decades of research has shown—that there are differences in student performance associated with students' backgrounds. For example, screening assessment data show that low income students, English learner students, and students receiving special education services were more likely than their peers not in those groups to score significantly below benchmark more than once during the school year, which means these students are at significant risk for reading difficulties.[18]

However, that analysis considers each background characteristic separately, and students often belong to multiple overlapping groups. For example, a student may be simultaneously classified as low income, Hispanic, and an English learner, and

Key Findings

- Students receiving special education or English learner services *and* coming from a low income background have a greater likelihood of being identified as at risk than students with any one of these characteristics across ethnoracial groups. Students belonging to all three groups have an even greater likelihood of being identified as at risk.
- Although the pattern of the increasing likelihood of being identified as at risk as background characteristics intersect is the same for all students, the increases vary by gender and other background factors. Female students who are English learners or who receive special education services are more likely to be identified as at risk than males across ethnoracial groups. Asian students show the smallest increases in likelihood of being identified as at risk across ethnoracial groups.
- Students in schools with above-average percentages of low income students, below-average teacher retention rates, and below-average student attendance have a greater likelihood of being identified as at significant risk. In Massachusetts, more Black and Hispanic students attend schools with these characteristics.

[18] 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.

these characteristics may independently and together influence a student's literacy performance. In Massachusetts, common intersections of student background characteristics include low income status and Black or Hispanic race/ethnicity, low income status and English learner status, English learner status and Hispanic, Black, or Asian race/ethnicity, and low income and English learner status and Hispanic ethnicity (table 11). For example, 76 percent of Hispanic students are low income and 56 percent are English learner students, and they are more than three times as likely to be classified as low income and English learners than are White students, more than twice as likely as Black students, and almost four times as likely as Asian students. Black and Hispanic students are also more commonly low income and more commonly receiving special education services than White or Asian students.[19]

Table 11. Percentage of low income and English learner students and students receiving special education services by race/ethnicity

Race/Ethnicity	Number	Low income (percent)	English learner (percent)	Receiving special education services (percent)	Low income and English learner (percent)	Low income and receiving special education services (percent)
White	34,497	27%	5%	18%	3%	7%
Hispanic/Latino	16,809	76%	56%	17%	47%	13%
Black	4,305	67%	23%	18%	17%	13%
Asian	4,691	24%	33%	10%	12%	3%
Other	3,320	31%	6%	17%	13%	9%
Total	**63,622**	**43%**	**22%**	**17%**	**16%**	**9%**

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per row if they were administered multiple screening assessments. In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model. The "Other" racial group includes American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, and multiracial students.

[19] In this section of the report, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

To explore how identifying or being identified as a member of multiple student groups (e.g., Hispanic and English learner) relates to risk of reading difficulties, we estimated a multilevel statistical model that examines how these characteristics and school-level factors interact with the outcome of being classified as significantly below benchmark multiple times. A detailed description of the model can be found in appendix E.

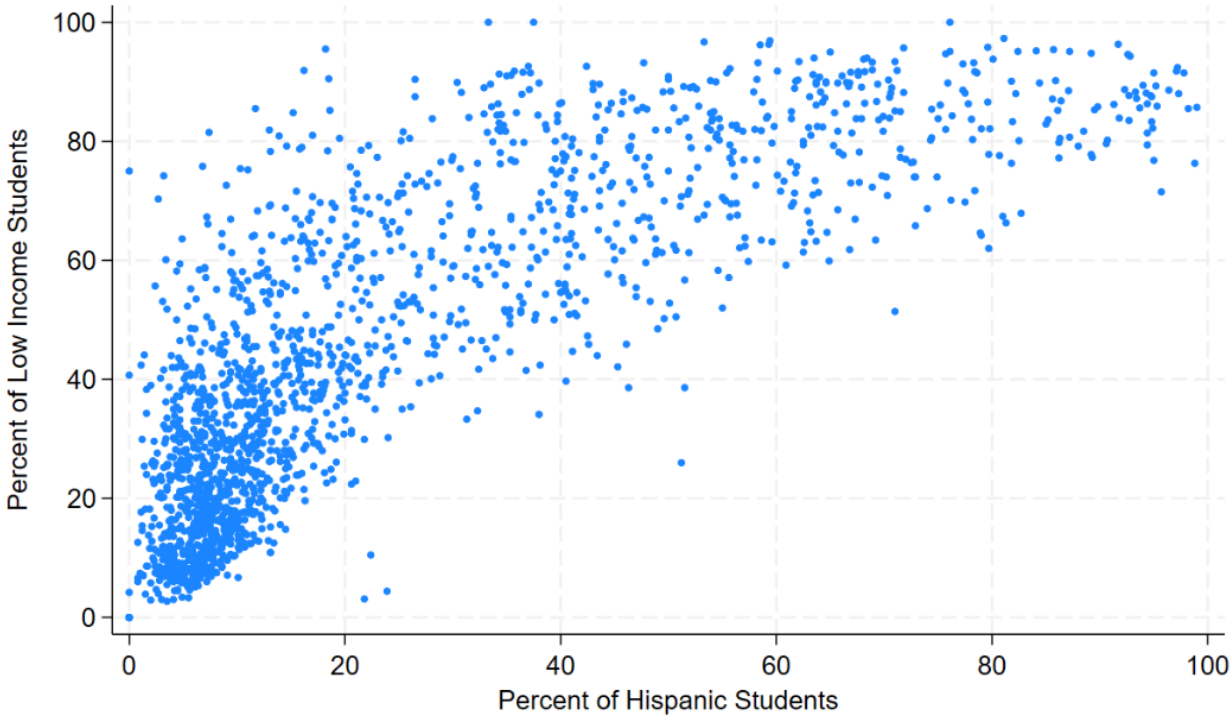
Model results show that the likelihood of students being identified as needing additional support increases as their association with groups historically under-supported in the general education system coincide, but that the increases vary by student groups and school characteristics. The probability of being identified as significantly below benchmark increases as students have more background characteristics individually associated with that outcome, but the increases are not the same for all groups. For example, being low income does not increase the probability of students being identified as significantly below benchmark equally across other student groups.

Student-level performance also varies by school characteristics. Prior descriptive analysis of screening assessment data (see the Opportunity Gaps issue brief [Lemke et al., 2003]) 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 among grantees were more likely to be below benchmark and to stay there from BOY through EOY than were their peers in schools without those characteristics. To account for these differences, several school-level factors found to be relevant during exploratory analysis were included in the statistical model (percent of low income students and teacher retention and student attendance rates).

Current analysis again demonstrates the relevance of schools to performance and shows how differences between them affect student performance. The model results shown here report the likelihood of being identified as at significant risk at an *average* school (based on the school-level factors included in the model). This means that students in schools with above-average percentages of low income students and below-average teacher retention and student attendance rates are more likely to be identified as significantly below benchmark than students in average schools.

The Massachusetts screening assessment data show that more Black and Hispanic students attend schools with these characteristics. For example, although the average percentage of low income students in schools in the early literacy screening assessment data is about 39 percent, many Hispanic students attend schools with far greater percentages of low income students (figure 11). Hispanic students, on average, attend schools where the median percentage of low income students is 70 percent, meaning the likelihood of being identified as significantly below benchmark is even higher for many Hispanic students than for those shown in the following sections, due to school characteristics. Conversely, students in schools with below average percentages of low income students and above average teacher retention and student attendance are less likely to be identified as significantly below benchmark.

Figure 11. Schools With Greater Percentages of Low Income Students Also Have Greater Percentages of Hispanic Students

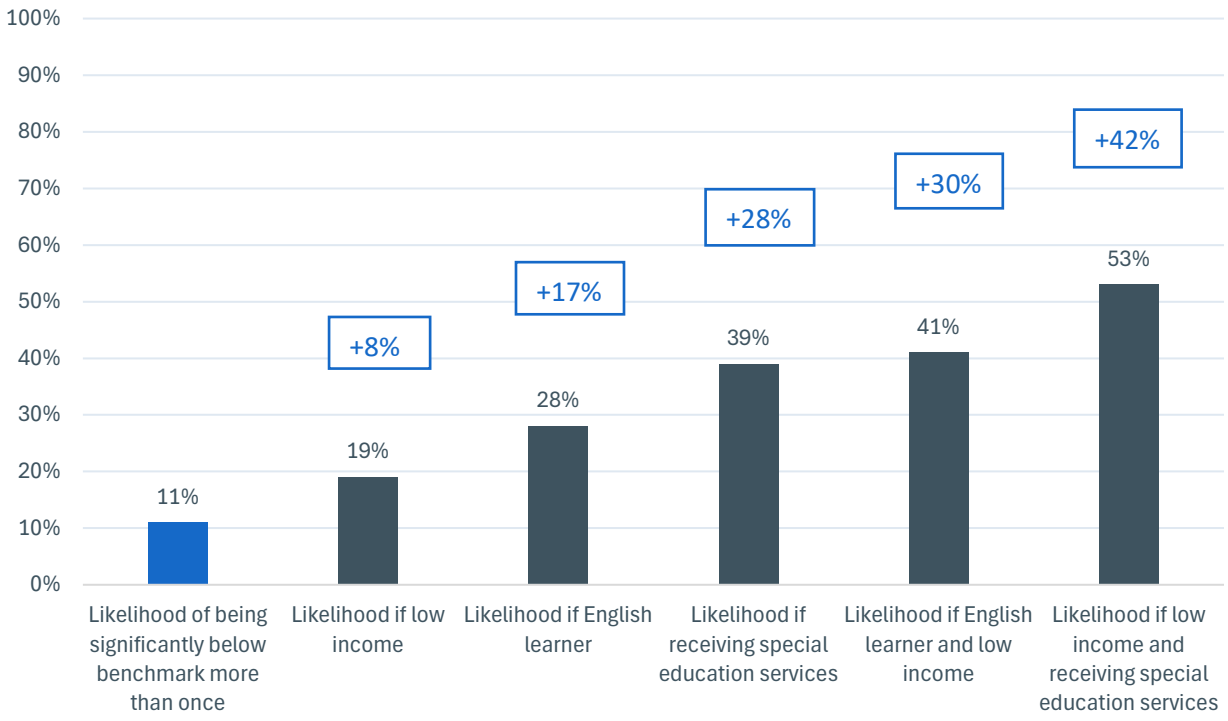


Source: District-provided screening assessment data

Students who receive special education services, are English learners, or come from a low income background are more likely to be identified as at risk of reading difficulty than students without those backgrounds, and the chance of being identified as in need of support increases as these background characteristics intersect.

On average, students have about an 11 percent chance of being identified as significantly below benchmark more than once. That likelihood increases by about 8 percentage points if students come from a low income background, 17 percentage points if students are English learners, and about 28 percentage points if students receive special education services (figure 12). The likelihood increases by 30 percentage points for students who come from a low income background **and** are English learners and about 42 percentage points for students who come from a low income background **and** receive special education services. These students have a 53 percent chance overall of being identified as in need of additional support. A small number of students who come from low income backgrounds and receive both special education and English learner services have an even higher likelihood of being identified as at significant risk.

Figure 12. The Likelihood of Being Identified as in Need of Additional Support Increases as Student Background Characteristics Intersect



Source: District-provided screening assessment data and October and June SIMS collection data
Notes: In this analysis, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). The likelihood of being significantly below benchmark more than once (11%) is defined by the multiple student and school-level characteristics included in the statistical model. See appendix E for details.

Although the pattern of the increasing likelihood of being identified as at risk as background characteristics intersect is the same for all students, the increases vary by gender and other background factors.

For example, across all ethnoracial groups, students from a low income background are more likely to need additional support, but the likelihood increases by 5 percentage points for Asian students and 10 percentage points for Hispanic students (table 12). Across intersecting categories, Asian students show the smallest increases in the likelihood of being identified as at risk.

Additionally, intersecting categories differentially affect males and females, with females generally having greater likelihood of risk. For example, Hispanic females who are English learners have a 47 percent chance of being identified as significantly below benchmark more than once compared to 43 percent for Hispanic English learner students who are male. Female English learner students across ethnoracial groups are 3 to 4 percentage points more likely to be identified as at risk than males.

Similarly, across ethnoracial groups, female students who receive special education services are more likely than male students to be identified as at risk. Female students who receive special education services and who come from low income backgrounds have the highest likelihood of being identified as at risk when examining background characteristics together. Black females who receive special education services and come from a low income background have a 70 percent chance of being identified as requiring additional support to be successful readers, the highest among students included in this analysis, suggesting this group is particularly underserved by the current system. Interestingly, however, there is a 9 percentage point difference in the probability of being identified as at risk between low income Black males and females receiving special education services, but a greater difference between Asian males and females (16 percentage points—low income Asian females receiving special education services have a 47 percent chance of being identified as at risk compared to 31 percent for males).

Table 12. Increased Likelihood of Needing Additional Support for English Learner Students and Students Receiving Special Education Services Who Also Come From Low Income Backgrounds Varies by Gender and Race/Ethnicity

	Asian Females	Asian Males	White Females	White Males	Black Females	Black Males	Hispanic /Latino Females	Hispanic/ Latino Males
Likelihood of being significantly below benchmark more than once	7%	7%	14%	14%	16%	16%	18%	18%
Likelihood if low income	12%	12%	22%	22%	25%	25%	28%	28%
Likelihood if English learner	24%	21%	40%	36%	44%	40%	47%	43%
Likelihood if receiving special education services	40%	26%	56%	46%	57%	48%	56%	44%
Likelihood if English learner and low income	35%	31%	53%	49%	58%	54%	61%	57%
Likelihood if receiving special education services and low income	54%	38%	68%	60%	70%	61%	68%	58%

Source: District-provided screening assessment data

Student Progress

In the previous sections, we report on student performance at BOY, MOY, and EOY and across time periods and on students who perform 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 from BOY to MOY to EOY vary by grade, benchmark level, and student background characteristics?
- How does student performance change across grade levels? Do students at risk remain at risk across years? What is the relationship between screening assessment scores and MCAS performance?

How Does Student Performance Change as the School Year Progresses? Do Students Identified as Significantly Below Benchmark Remain Significantly Below Benchmark?

Of students with at least two benchmark scores, most students who were significantly below benchmark in one time period were also significantly below benchmark in a later time period, but between 27 percent and 40 percent of students did improve their performance after being identified as significantly below benchmark in an earlier time period (table 13). Similar patterns of progress were shown in prior reporting using 2020/21 and 2021/22 data.

Sixty percent of students who were significantly below benchmark at BOY were also significantly below benchmark at EOY, whereas 40 percent of students who started off the year significantly below benchmark were **not** significantly

Key Findings

- Sixty percent of students who begin the year significantly below benchmark end the year at that same level; conversely, most students meeting benchmark in one time period also met benchmark in later time periods, as in prior reporting.
- Some students who started the year identified as significantly below benchmark did improve, either moving from significantly below benchmark to below benchmark (16%) or from significantly below benchmark to meeting benchmark (24%) by the end of the year.
- As shown in previous analyses, improving performance between BOY and MOY significantly improves the likelihood of success at EOY. Students who were significantly below benchmark at BOY but met benchmark at MOY were about 1.6 times more likely to meet benchmark at EOY than those who were significantly below at BOY but were just below at MOY, and 8.4 times more likely than those were significantly below at both BOY and MOY.

below benchmark by EOY. 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 (87%). The same patterns were observed when analyzing 2020/21 and 2021/22 data.

Similar trends were observed when analyzing the percentages of students at the 25th percentile or below, though they were more pronounced (table 14). Students identified as at significant risk by this metric were even more likely to still be at significant risk in a later time period, and students not identified as at significant risk were also more likely not to be at significant risk during a later time period.

Table 13. Later Performance of Students Classified as at Significant Risk at BOY and/or MOY

If significantly below benchmark	Then significantly below at BOY	Then significantly below at MOY	Then significantly below at EOY
BOY	N/A	70% (*n* = 14,672)	60% (*n* = 11,925)
MOY	N/A	N/A	73% (*n* = 15,213)
If not significantly below benchmark	**Then not significantly below at BOY**	**Then not significantly below at MOY**	**Then Not Significantly Below at EOY**
BOY	N/A	88% (*n* = 33,792)	87% (*n* = 33,357)
MOY	N/A	N/A	93% (*n* = 39,025)

Table 14. Later Performance of Students at or Below 25th Percentile at BOY and/or MOY

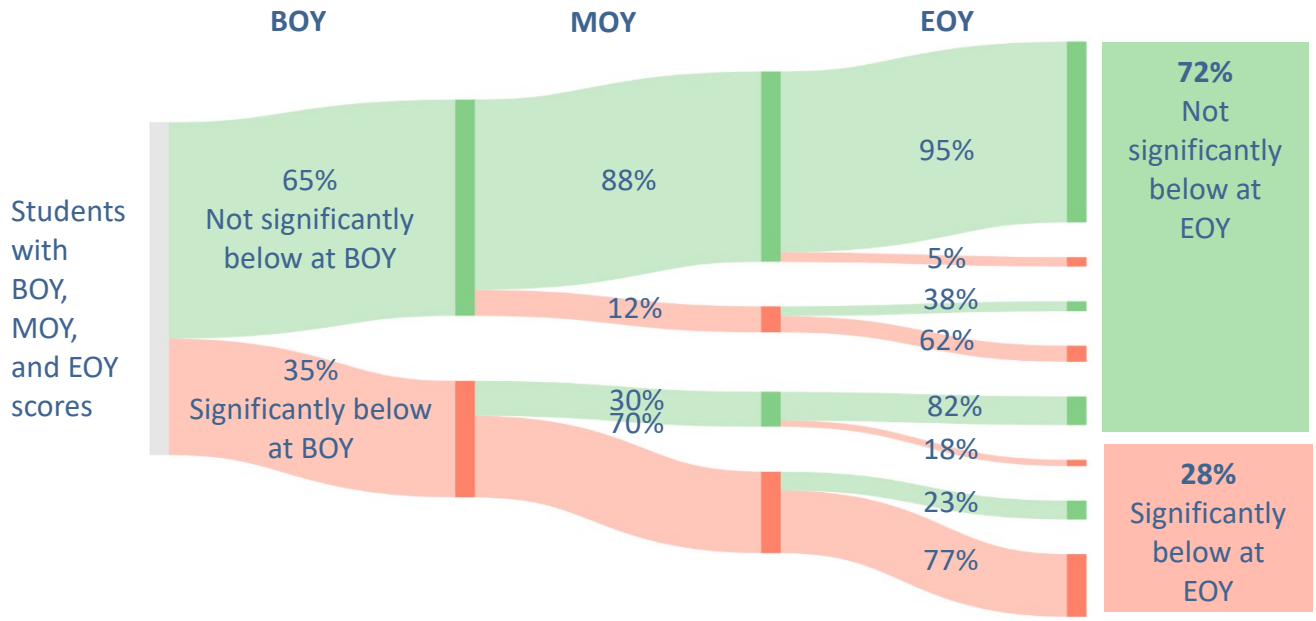
If 25th percentile or below	Then 25th percentile or below at BOY	Then 25th percentile or below at MOY	Then 25th percentile or below at EOY
BOY	N/A	70% (*n* = 11,248)	65% (*n* = 10,176)
MOY	N/A	N/A	77% (*n* = 11,230)

If not 25th percentile or below	Then not 25th percentile or below at BOY	Then not 25th percentile or below at MOY	Then not 25th percentile or below at EOY
BOY	N/A	94% (*n* = 33,976)	92% (*n* = 33,328)
MOY	N/A	N/A	93% (*n* = 37,691)

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Tables 13 and 14 show how students with at least two scores (about 89 percent of students in the sample) progress. Students who had a score during each time period of the school year (75 percent of the sample) had similar patterns of progress (figure 13).

Figure 13. Most Students Who Started Significantly Below Benchmark Were Still Significantly Below at EOY, but Improving Performance Between BOY and MOY Reduces Risk at EOY



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (75%) were included in the figure.

Of the students who were identified as significantly below benchmark at BOY, most continued to perform at that level for the remainder of the school year (as was the case with prior years' data); 70 percent of these students performed significantly below benchmark at MOY, and

77 percent of the students who performed significantly below benchmark in both time periods were also at that level at EOY. However, about 30 percent of students who were classified as significantly below benchmark at BOY improved their performance at MOY, and of these students, the vast majority (82%), as in prior reporting, were also not significantly below benchmark at EOY.

Among the 65 percent of students who were not significantly below benchmark at BOY, the vast majority were also not significantly below benchmark at MOY and EOY (88% and 95%). However, of the 12 percent of students who were not significantly below benchmark at BOY but were at MOY, 62 percent continued to be significantly below benchmark at EOY.

Using the 25th percentile or below metrics shows the same trends (see figure F.1 in appendix F); however, fewer students were identified as at significant risk at BOY. Additionally, when using this metric, students who were not at significant risk at BOY were more likely to also not be at significant risk at EOY, and students at significant risk at BOY were more likely to still be at significant risk at EOY, regardless of their performance at MOY.

Because most assessments include multiple risk levels, some students identified as significantly below benchmark may make progress during the school year but still be identified as at some risk (table 15).

Table 15. Later Performance of Students Classified as at Significantly Below Benchmark in BOY and/or MOY

If significantly below benchmark	Then significantly below at BOY	Then significantly below at MOY	Then significantly below at EOY
BOY	N/A	70% (*n* = 14,615)	60% (*n* = 11,903)
MOY	N/A	N/A	73% (*n* = 15,213)
If significantly below benchmark	**Then below but not significantly below at BOY**	**Then below but not significantly below at MOY**	**Then below but not significantly below at EOY**
BOY	N/A	16% (*n* = 3,291)	16% (*n* = 3,238)
MOY	N/A	N/A	16% (*n* = 3,305)

If significantly below benchmark	Then met at BOY	Then met at MOY	Then met at EOY
BOY	N/A	14% (n = 2,896)	24% (n = 4,794)
MOY	N/A	N/A	11% (n = 2,312)

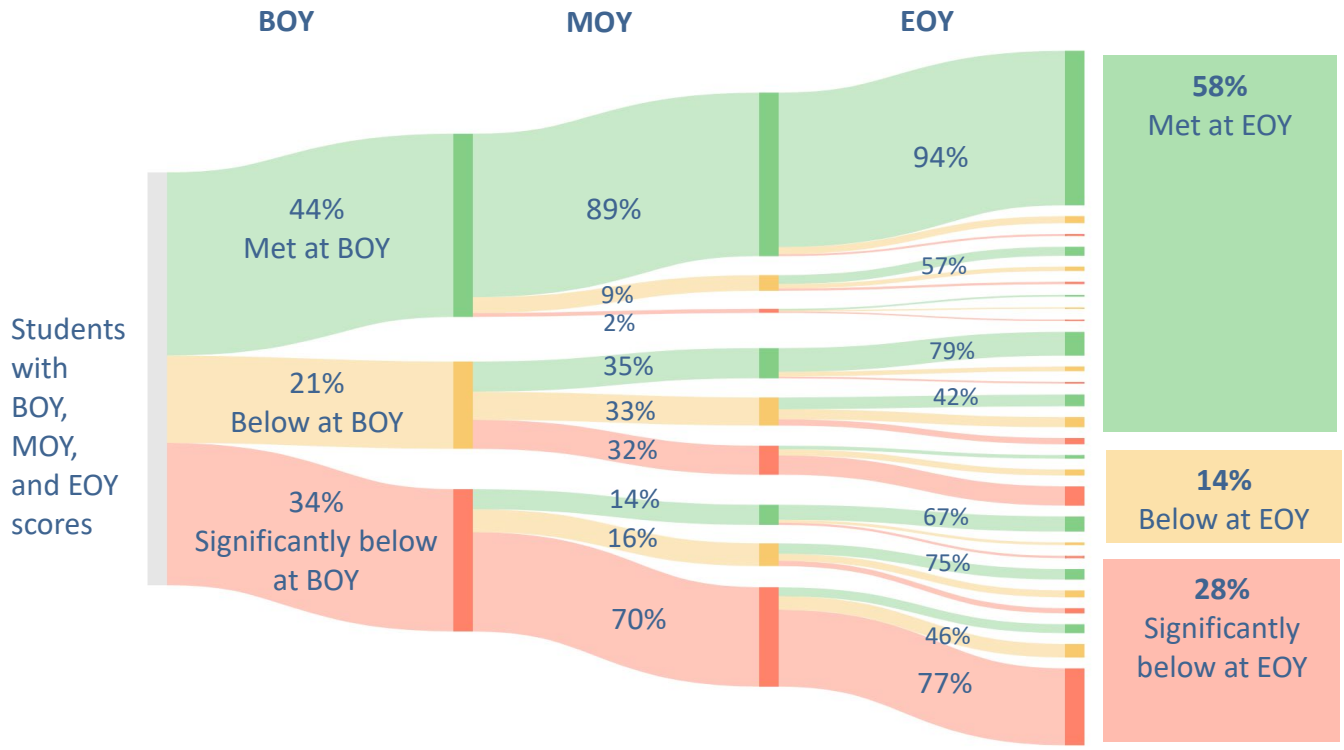
Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Of the students with at least two scores who were significantly below benchmark at BOY, 60 percent were still significantly below benchmark at EOY, and 40 percent had improved performance on their screening assessment; 16 percent were below benchmark (but not significantly below benchmark), and 24 percent met benchmark. In total, 76 percent of students who were identified as at significant risk at BOY were still at some level of risk by EOY. Similar trends are observed when comparing progress from BOY to MOY and from MOY to EOY—that is, most students significantly below benchmark at an earlier time period remained significantly below benchmark.

More improvement occurred between BOY and EOY than between BOY and MOY or MOY and EOY (i.e., the longer the period between testing, the more likely it was that students would improve). For example, of the students who were significantly below benchmark at BOY or MOY (but not at EOY), 19 percent more of the BOY students met benchmark at EOY than did the MOY students (60% compared to 41%).

Similar patterns of improvement are observed when examining students with scores in all three time periods (figure 14).

Figure 14. Student Performance Remained Relatively Consistent Throughout the School Year, but Most Students Who Did Improve Between BOY and MOY Maintained That Improvement at EOY



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (75%) were included in the figure.

At the beginning of the year, 44 percent of students met benchmark, 21 percent were below benchmark but not significantly below, and 34 percent were significantly below benchmark. Students who met benchmark or were significantly below benchmark at BOY typically remained at these levels at MOY and EOY. Of the students who met benchmark at BOY, 89 percent also met benchmark at MOY, and of those, 94 percent met benchmark at EOY.

Of the students who were significantly below benchmark at BOY, 70 percent were also significantly below at MOY, and of those, 77 percent were still significantly below at EOY. However, although students who met benchmark at BOY were very likely to meet benchmark at MOY and EOY, those who fell off track by MO Y struggled to get back on track at EOY. Of the 11 percent of students who met benchmark at BOY but not at MOY, 45 percent remained at risk at EOY.

Students who were below benchmark at MOY but not significantly below benchmark experienced more mixed results—about a third of these students met benchmark at MOY, about a third remained below benchmark, and about a third moved to significantly below

benchmark. As a result, more students who were below benchmark at BOY met benchmark at EOY than students who were significantly below benchmark at BOY.

Additionally, improving performance between BOY and MOY significantly improves the likelihood of success at EOY. Students who met benchmark at MOY after being below benchmark (but not significantly below) at BOY were almost twice as likely to meet benchmark at EOY than those who were below benchmark in both previous time periods. Similarly, students who were significantly below benchmark at BOY but met benchmark at MOY were about 1.6 times more likely to meet benchmark at EOY than those who were significantly below at BOY but were just below at MOY and 8.4 times more likely to meet benchmark than those who were significantly below at both BOY and MOY.

How Does Student Progress Vary by Grade, Benchmark Level, and Student Background Characteristics?

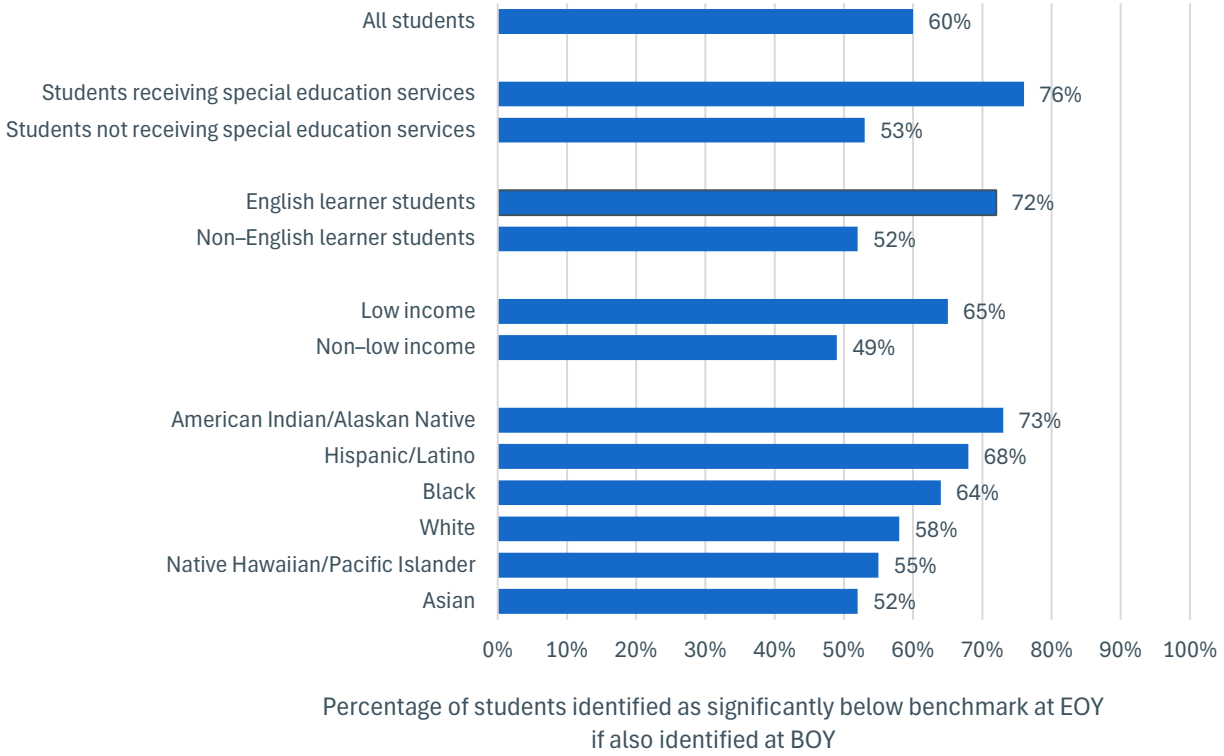
Among students with at least two scores, the percentage of those who remained significantly below benchmark if identified as such at BOY increases across grade levels. The same trend was observed in prior reporting. For example, 37 percent of kindergarten students identified as significantly below benchmark at BOY are still classified as such at EOY, compared to 78 percent of grade 3 students (see tables F.1–F.4 in appendix F). Grade 3 students are more than twice as likely as kindergarten students to be identified as significantly below benchmark at EOY if identified as such at BOY, grade 2 students are 1.9 times as likely, and grade 1 students are 1.5 times as likely.

Key Findings

- Fewer students at higher grade levels get on track after being identified as significantly below benchmark than students in lower grade levels (similar to patterns seen in prior reporting).
- More English learner students, low income students, students receiving special education services, and Black and Hispanic students were signficantly below benchmark at EOY after being identified as such at BOY or MOY than their peers not in those groups.
- Fewer students significantly below benchmark at BOY achieved benchmark at EOY than students at lower levels of risk.

Additionally, as shown in prior analyses of 2020/21 and 2021/22 data, more English learner students, low income students, students receiving special education services, and Black, Hispanic, and American Indian/Alaskan Native students were significantly below benchmark at EOY after being identified as such at BOY or MOY than their peers not in those student groups (see figure 15 and table 16 below, and tables F.5–F.24 in appendix F). For example, 65 percent of low income students who were significantly below benchmark at BOY were still significantly below benchmark at EOY, compared to 49 percent of non–low income students.

Figure 15. More English Learner Students, Low Income Students, Students Receiving Special Education Services, and Black, Hispanic, and American Indian/Alaskan Native Students Were Still Significantly Below Benchmark at EOY After Being Identified as Such at BOY Than Their Peers Not in Those Groups



Source: District-provided screening assessment data

Similarly, English learner students, low income students, students receiving special education services, Black students, Hispanic students, and students in later grade levels were also less likely than their peers not in those groups to meet benchmark at EOY if they were just below benchmark at BOY (see table 16). They were also more likely not to show any improvement in their benchmark performance than their peers.

Table 16. Performance at EOY for Students Significantly Below and Below Benchmark at BOY by Student Group

	Significantly below at EOY (percent)	Below at EOY (but not significantly below) (percent)	Met at EOY (percent)
Significantly below benchmark at BOY			
Kindergarten	37%	18%	45%
Grade 1	56%	19%	26%
Grade 2	70%	16%	15%
Grade 3	78%	12%	10%
Low income	65%	14%	20%
Non–low income	49%	20%	31%
Female	58%	18%	24%
Male	61%	15%	24%
Nonbinary	--	--	--
English learner students	72%	12%	16%
Non–English learner students	52%	19%	29%
Students receiving special education services	76%	11%	12%
Students not receiving special education services	53%	18%	29%
White	58%	17%	25%

	Significantly below at EOY (percent)	Below at EOY (but not significantly below) (percent)	Met at EOY (percent)
Hispanic/Latino	68%	13%	19%
Black	64%	15%	21%
Asian	52%	19%	29%
American Indian/Alaskan Native	73%	11%	16%
Native Hawaiian/Pacific Islander	55%	16%	29%
Significantly below benchmark at BOY: Total	**60%**	**16%**	**24%**
Below benchmark (but not significantly below) at BOY			
Kindergarten	29%	19%	53%
Grade 1	34%	22%	44%
Grade 2	35%	28%	37%
Grade 3	34%	29%	38%
Low income	38%	22%	40%
Non–low income	30%	25%	45%
Female	32%	25%	43%
Male	34%	23%	43%
Nonbinary	--	--	--

	Significantly below at EOY (percent)	Below at EOY (but not significantly below) (percent)	Met at EOY (percent)
English learner students	44%	20%	36%
Non–English learner students	31%	24%	45%
Students receiving special education services	48%	22%	31%
Students not receiving special education services	30%	24%	46%
White	33%	24%	43%
Hispanic/Latino	40%	21%	39%
Black	37%	22%	41%
Asian	30%	24%	46%
American Indian/Alaskan Native	35%	19%	46%
Native Hawaiian/Pacific Islander	50%	18%	32%
Below benchmark (but not significantly below) at BOY: Total	**33%**	**24%**	**43%**

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

How Does Student Performance Change Across Grade Levels? Do Students at Risk Remain at Risk Across Years? What is the Relationship Between Screening Assessment Scores and MCAS Performance?

In this section, we examine changes in performance for students with scores from 2021/22 and 2022/23. In total, 11,859 students (about 16% of students in 2022/23) had a score from EOY in one grade level to EOY in the next on the same assessment. A much smaller number of students (395 students, or 0.5%) had two prior years with EOY scores, largely because the 2020/21 dataset had many fewer students than the 2021/22 file. Of those with an EOY score in 2021/22 and 2022/23, about a third were classified as significantly below benchmark in 2021/22 (table 17).

Key Findings

- Seventy-one percent of students who were significantly below benchmark at EOY at one grade level were still significantly below benchmark at EOY in the following grade level.
- Students in earlier grades were again more likely to get on track across years than students in later grades; 64 percent of students who ended kindergarten significantly below benchmark were still significantly below benchmark at the end of grade 1, compared to 73 percent of grade 1 students who moved to grade 2, and 77 percent of grade 2 students who moved to grade 3.

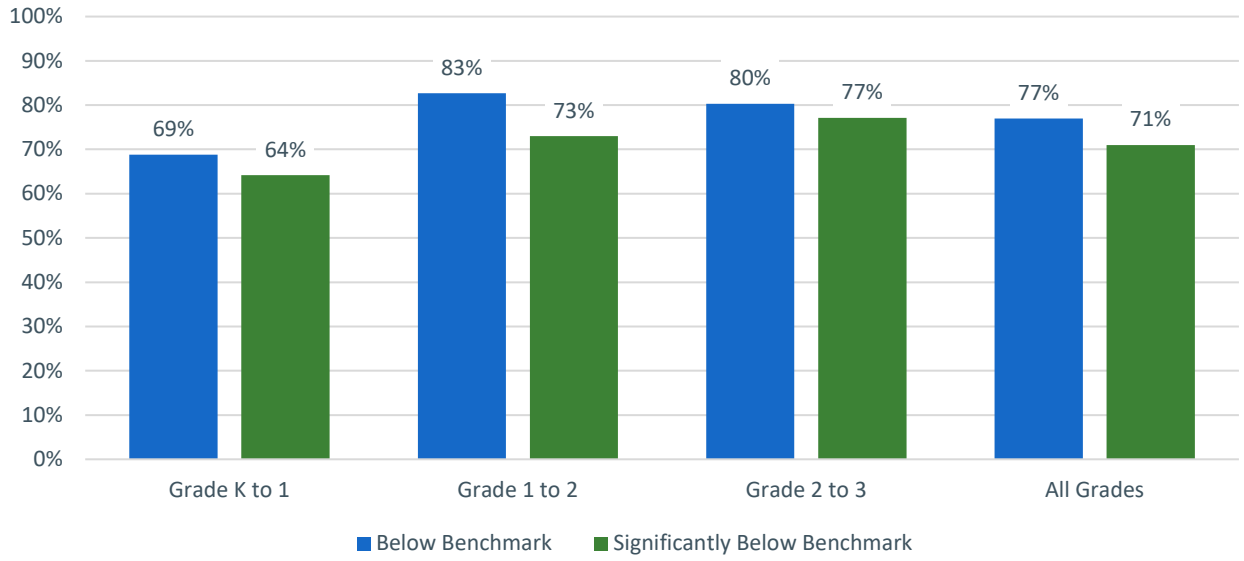
Table 17. Percentage of Students With EOY Scores in 2021/22 and EOY Scores in 2022/23 by Grade Level

Students with scores from EOY 2021/22 to EOY 2022/23	Number	Below benchmark at EOY of 2021/22 (percent)	Significantly below benchmark at EOY of 2021/22 (percent)	25th percentile or below at EOY of 2021/22 (percent)
Grade K to grade 1	4,340	49%	31%	29%
Grade 1 to grade 2	4,473	48%	36%	27%
Grade 2 to grade 3	3,046	52%	34%	28%
Total	**11,859**	**50%**	**33%**	**28%**

Source: District-provided screening assessment data
Note: Only 8,175 observations were included in the 25th percentile or below column, as some assessments or students did not have a recorded national percentile.

Within a school year, about 60 percent of students who started the year significantly below benchmark were still classified as such at EOY, and about 69 percent who began below benchmark were still below benchmark at EOY. Across years, the percentages increase to over 70 percent for students transitioning from grade 1 to grade 2 and grade 2 to grade 3, but there are substantial differences across grades (figure 16).

Figure 16. More Students at Higher Grade Levels Stayed at Risk From Grade to Grade Than Students at Lower Grade Levels



Source: District-provided screening assessment data
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.

As in previous analyses, more students in later grades remained below benchmark or significantly below benchmark across years than did students in earlier grades. For example, 64 percent of students who ended kindergarten significantly below benchmark were still significantly below benchmark at the end of grade 1, compared to 73 percent of grade 1 students who moved to grade 2 and 77 percent of grade 2 students who moved to grade 3. Similar patterns in progress appear for students identified as below benchmark at the end of the year.

The 25th percentile or below metric produces similar results (see figure F.2) with fewer students remaining at significant risk from one year to the next. The metric also showed that students who were at significant risk at EOY in kindergarten in 2021/22 less frequently remained at significant risk at EOY in 2022/23 than students who were in grade 1 or grade 2 in 2021/22. Sixty percent of kindergarten students who ended the year at significant risk were still at significant risk at the end of grade 1, compared to 72 percent of grade 1 students who

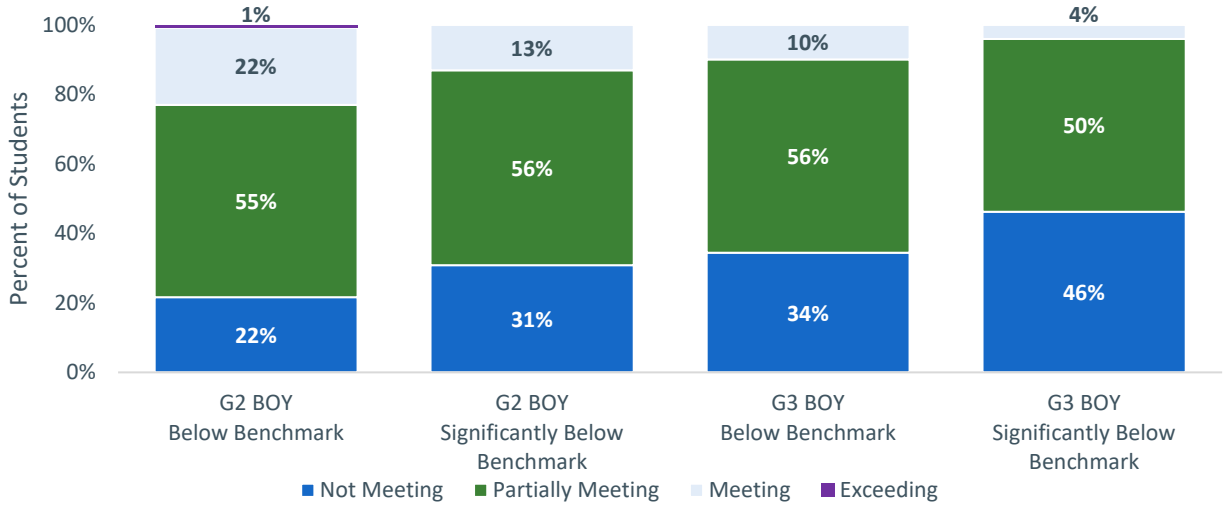
remained at significant risk at the end of grade 2, and 71 percent of grade 2 students who remained at significant risk at the end of grade 3.

English learner students, low income students, students receiving special education services, and students in all racial/ethnic groups show the same pattern of having more students in later grades remain significantly below or below benchmark across grade levels and years, though the percentages were larger for English learner students, low income students, students receiving special education services, and Hispanic and Black students than for their peers not in those groups (see figures F.3–F.11 in appendix F). For example, 73 percent of Hispanic students who ended kindergarten significantly below benchmark were still significantly below benchmark at the end of grade 1, compared to 57 percent of non-Hispanic students.

These results suggest that most students who end the school year below benchmark or significantly below benchmark will still be performing at those levels at the end of the year in subsequent grades, and the chances are even greater as students move from kindergarten through grade 3.

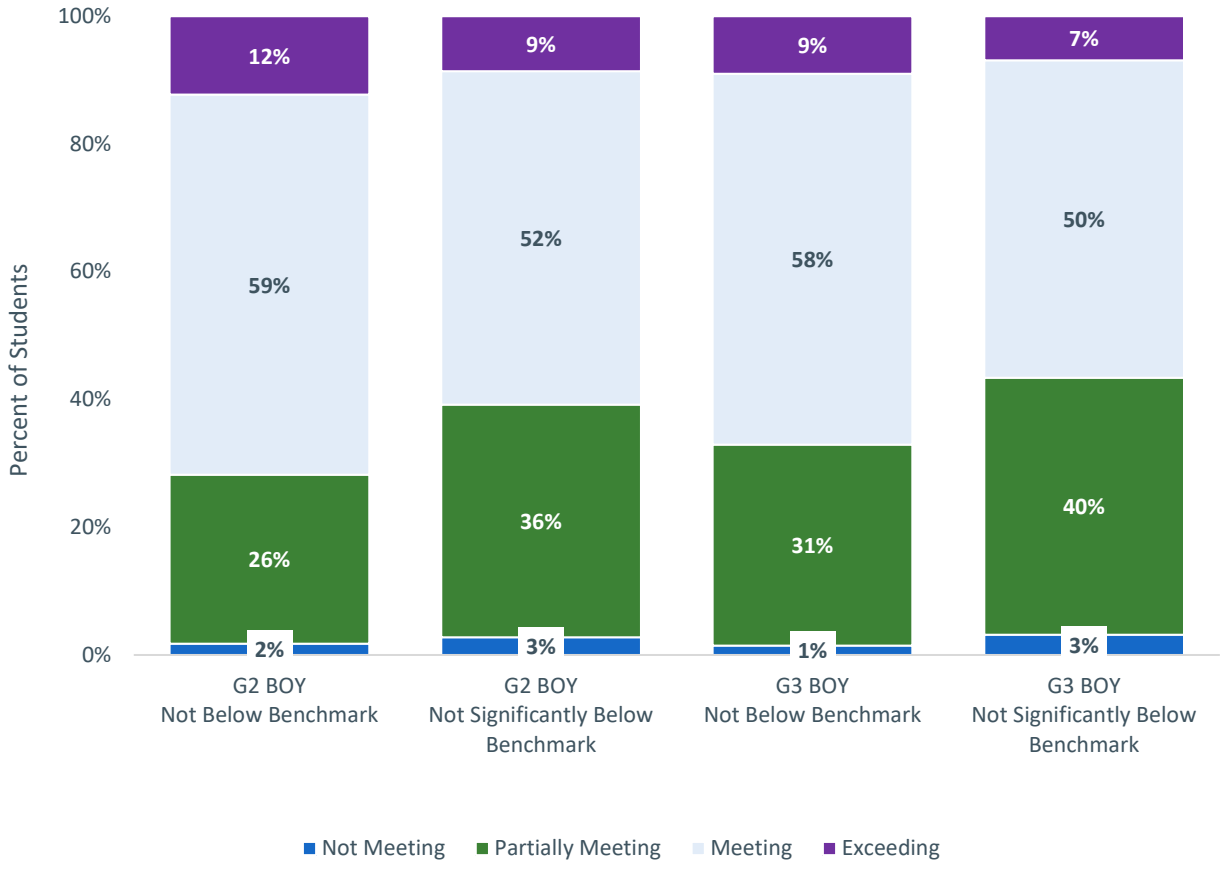
Similarly, student performance at BOY in grades 2 and 3 also corresponds to MCAS performance at the end of grade 3 (figures 17 and 18). That is, most students who were below benchmark or significantly below benchmark on screening assessments in grades 2 or 3 did not meet MCAS proficiency standards, though more students at grade 2 students than grade 3 did meet proficiency standards. For example, 13 percent of students who were significantly below benchmark at BOY in grade 2 were proficient on MCAS at the end of grade 3, compared to 4 percent of grade 3 students.

Figure 17. Relatively Few Students Identified as at Risk on Grade 2 and Grade 3 Literacy Screening Assessments Meet or Exceed Expectations on Grade 3 MCAS ELA



Source: District-provided screening assessment data and state-provided MCAS data

Figure 18. Most Students not Identified as at Risk or Significant Risk Meet or Exceed MCAS Standards, Though About a Third of Students Do Not



Source: District-provided screening assessment data and state-provided MCAS data

As described in the "Comparing Screening Assessment Benchmarks" section, most benchmarks indicating reading risk map to the MCAS "Partially Meeting" performance level, so these results are not surprising. However, as that section also points out, benchmarks that indicate little to no risk do not necessarily mean that all students above them will meet MCAS standards (figure 18), which may be due in part to differences in the assessments. (See also table D.15.) Between 57 percent and 72 percent of students who met screening assessment benchmarks also met MCAS standards.

Comparing Student Performance to the Nation

Screening assessment publishers typically provide normative reporting information that enables users to compare performance of their students to a nationally representative sample of students. Additionally, policymakers may find it useful to compare data across schools and districts within the state to try to identify places where students may be outperforming their peers and whose literacy instruction or intervention practices may provide insights for others. In this section, we examine how well Massachusetts students compare to their peers nationally, using the norms provided by screening assessment publishers, and whether or not any schools within the state stand out based on their performance.

How Does Achievement and Growth of Massachusetts Students Compare to National Samples and Growth Rates?

Massachusetts has long been considered a high-performing state. Massachusetts students' performances on the National Assessment of Educational Progress and on international comparative assessments have often placed them at the top of the rankings compared to other U.S. states. The normative data provided by literacy screening assessment publishers provides a way for the state to compare literacy performance and progress to students in grades K–3 around the country as well. In this section, we use national percentiles and publisher-provided growth norms to examine how students in the state compare to students in other states. For purposes of this analysis, we limit the Massachusetts sample to students participating in assessments with comparable norms (for 2022/23, this included only DIBELS 8th Edition and mClass, which use the same norms). These two assessments are also fully approved for use in the state and represent over a third of the screening assessment data sample.

Key Findings

- Despite differences between student groups in the state, students in Massachusetts overall perform above the national average based on norms provided by assessment publishers.
- Black students in Massachusetts perform about the same as all other students around the country, while White students perform slightly better, Asian students perform significantly better, and Hispanic students perform less well than the national sample.
- Growth of students using publisher-provided growth norms shows more than half of students (66%) of students growing at average or above average rates compared to about 60 percent nationally.

National percentile values range from 1 to 99, with a value of 50 representing the point at which half the students performed above that value and half below. Overall, students in Massachusetts perform slightly above the 50th national percentile (median percentiles of 55, 56, and 58 at BOY, MOY, and EOY), meaning

Massachusetts students are performing above the national average (table 18). EOY performance is generally higher than MOY and BOY, and grade 2 and grade 3 students perform better than their peers nationally compared to kindergarten and grade 1 students in the state. More students perform above the 75th percentile than the 25 percent that would be expected if the Massachusetts sample performed exactly like the national sample, which likely accounts for the higher median national percentile in the state. As with the publisher-provided significantly below benchmark levels, smaller proportions of students were identified as below the 25th percentile at EOY than BOY.

Table 18. National Percentiles at BOY, MOY, and EOY

Grade	Early literacy screening assessment	N	Median national percentile			Percent above 75th percentile			Percent below 25th percentile		
			BOY	MOY	EOY	BOY	MOY	EOY	BOY	MOY	EOY
K	DIBELS 8th Edition and mCLASS	6,859	53	54	57	25%	27%	29%	23%	20%	20%
1	DIBELS 8th Edition and mCLASS	7,085	52	55	56	26%	26%	29%	21%	18%	18%
2	DIBELS 8th Edition and mCLASS	5,790	57	58	59	31%	32%	31%	19%	19%	18%
3	DIBELS 8th Edition and mCLASS	5,311	60	59	59	33%	32%	32%	20%	20%	18%
Total	**DIBELS 8th Edition and mCLASS**	**25,045**	**55**	**56**	**58**	**28%**	**29%**	**30%**	**21%**	**19%**	**19%**

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Includes students with one, two, or three scores.

Although students overall in the state perform slightly better than their peers nationally, there is variation between student groups (table 19). For example, Hispanic/Latino students in Massachusetts perform less well on average than students around the country. Black students perform about the same as students in the national sample, and White and Asian students perform better.[20]

Table 19. National Percentiles at BOY, MOY, and EOY by Student Group

Race/Ethnicity	N	Median National Percentile			Percent Above 75th Percentile			Percent Below 25th Percentile		
		BOY	MOY	EOY	BOY	MOY	EOY	BOY	MOY	EOY
White	19,191	54	55	56	26%	27%	28%	21%	19%	19%
Hispanic/Latino	5,449	34	38	40	15%	16%	17%	40%	36%	34%
Black	3,302	49	50	49	25%	25%	23%	24%	24%	24%
Asian	3,634	75	73	74	49%	46%	48%	11%	10%	9%
American Indian/Alaskan Native	470	40	45	47	18%	19%	21%	31%	27%	25%
Native Hawaiian/Pacific Islander	116	54	52	55	29%	29%	23%	26%	23%	22%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Includes students with one, two, and three scores.

DIBELS 8th Edition and mClass also provide nationally normed growth metrics, classifying the change in student performance using the composite score within the year into five categories, or Zones of Growth. These categories represent the extent to which students have grown—some more, some less—by EOY than their peers with similar BOY skill levels (see University of Oregon [2022] for details on the metrics). The categories are as follows:

- Well below average (below 20th percentile for growth)
- Below average (between 20th and below 40th percentile for growth)

[20] Note that the national sample includes students from all racial/ethnic groups—specific norms for different student groups were not available.

- Average (between 40th and below 60th percentile for growth)
- Above average (between 60th and below 80th percentile for growth)
- Well Above Average (80th percentile and above)

Across grades, 66 percent of students in Massachusetts grew at an average rate or above, with 35 percent below or well below average (table 20). Compared to national growth percentiles, which would predict 20 percent of students in each category, fewer students in the state grew at a below or well below average rate compared to a national sample and more students grew at an average or above average rate.

Table 20. Massachusetts Student Performance Using DIBELS 8th Edition and mCLASS Growth Benchmarks Describing Early Literacy Growth From BOY 2023 to EOY 2023 by Grade

Grade	Number	Well below average (percent)	Below average (percent)	Average (percent)	Above average (percent)	Well above average (percent)
K	4,750	15%	19%	23%	22%	22%
1	5,500	15%	19%	26%	21%	19%
2	4,691	17%	18%	26%	21%	19%
3	4,188	17%	19%	28%	19%	16%
Total	**19,129**	**16%**	**19%**	**26%**	**21%**	**19%**

Source: District-provided screening assessment data and October and June SIMS collection data

Early Childhood Program Participation and Kindergarten Student Performance

As seen in earlier sections, students at earlier grade levels more often got on track from BOY to EOY or across years if they were identified as significantly below benchmark. In this section, we examine how student experiences even before kindergarten relate to performance. After examining enrollment in early childhood (EC) programs, we explore how student performance varies by EC program experience, program type, and student background characteristics.

How Many Kindergarten Students Attended Any Type Of Early Childhood Program, by Type? How Does Enrollment Vary by Student Background Characteristics?

In this section, we provide an overview of available data about participation in EC programs and a description of participants. Data are drawn from a survey administered to families of incoming, first-time kindergarten students enrolling in a public school. The survey may be administered during kindergarten enrollment, during screening, or at another time depending on the district. The survey asks parents or guardians if their children participated in formal or informal preschool and how much time per week the children spent in programs (less than 20 hours per week or 20 hours or more per week on average). Districts record this information in their student information systems and report it to the state through the regular state data reporting process. Relevant variables are then added to the state's SIMS datafile for kindergarten students (table 21) and for this analysis were merged to early literacy screening assessment data using the process described earlier in this report.

Key Findings

- Students in kindergarten in the screening assessment sample are similar to the state's overall kindergarten population in terms of EC experience, with over 60 percent of kindergarteners having a record of some type of EC experience.
- Fewer low income, English learner, Hispanic, American Indian/Alaskan Native, and Native Hawaiian/Pacific Islander students had any EC experience compared to students not in those groups; students receiving special education services more often had EC experience than students not receiving those services.

Table 21. Early Childhood Experience Indicators and Definitions

Indicator	Definition
00	Information not provided
01	No formal EC program experience
02	Family Support: Coordinated Family and Community Engagement (CFCE)
03	Family Support: Parent-Child Home Program (PCHP)
04	Family Support: Both CFCE and PCHP
05	Formal: Licensed family child care provider < 20 hours per week
06	Formal: Licensed family child care provider => 20 hours per week
07	Formal: Center-based program < 20 hours per week
08	Formal: Center-based program => 20 hours per week
09	Formal: Both licensed family child care provider and center-based program < 20 hours per week
10	Formal: Both licensed family child care provider and center-based program => 20 hours per week
99	Not applicable

Notes: A formal EC program is a public preschool, licensed community-based preschool/child care, Head Start program, and/or licensed family child care provider. "Licensed family child care" refers to state-licensed child care in a group setting in a home. It may include care in the home of a family member if the provider is both a relative and a licensed child care provider providing care to children from multiple families. "Center-based care" refers to care for children in a group setting, including public and private preschools, Head Start, early education and care centers, and inclusive/integrated public preschools. Informal EC experiences include CFCE services and PCHP participation. CFCE services are locally based programs serving families with children from birth through school age. (e.g., parent/child playgroups, parent-child activities). The PCHP is funded through the Massachusetts Department of Early Education and Care. Not every community has a program. The PCHP is a home-visit model providing low income families with the knowledge, skills, and tools to build school readiness in their homes before their children enter school.

Students in kindergarten in the early literacy screening assessment data sample are very similar to the state's kindergarten population as a whole in terms of EC experience. About 16 percent have no available data regarding EC experience, about 19 percent had no EC experience, and about 63 percent of students participated in a formal EC program, with a very small number participating in informal programs (table 22). Of students with EC experience, then, about 23 percent had no EC experience, 76 percent participated in formal EC programs, and the remaining percentage attended informal EC programs.

Table 22. Numbers and Percentages of Students With Early Childhood Experience in the Early Literacy Screening Assessment Data Sample and the State

EC experience	Percentage sample (kindergarten)	Percentage state (kindergarten)
No data	16% (*n* = 2,985)	16% (*n* = 9,754)
No EC experience	19% (*n* = 3,457)	19% (*n* = 11,945)
Informal (family support) program	1% (*n* = 243)	1% (*n* = 766)
Formal program	63% (*n* = 11,568)	64% (*n* = 39,146)
Total kindergarten students	**18,253**	**61,611**

Source: District-provided literacy screening assessment data and October and June SIMS collection data
Notes: Twenty-eight kindergarten students did not have a recorded EC experience code in SIMS. "No data" consists only of students with a recorded value of 00 (Information not provided) or 99 (Not Applicable).

Fewer low income, English learner, Hispanic, American Indian/Alaskan Native, and Native Hawaiian/Pacific Islander students had any EC experience compared to students not in those groups (table 23). A slightly smaller percentage of Black students (77%) had any EC experience compared to White and Asian students (about 79%). Students receiving special education services had EC experience more often than students not receiving those services, perhaps due to early intervention programs.

Among students who attended any EC program, most attended formal EC programs. Though the overall percentages are small, low income, English learner, Hispanic, and Black students participated in informal programs more often compared to their peers not in those groups. Informal and formal EC programs include several different types of programs; some students attend multiple types (table G.1 in appendix G provides a breakdown by specific program type).

Table 23. Percent of Students With Early Childhood Experience by Student Group

Demographic	No EC experience (percent)	Any EC experience (percent)	Formal (percent)	Informal (percent)
Low income	33%	67%	64%	2%
Non–low income	14%	86%	85%	1%
Female	23%	77%	75%	2%
Male	22%	78%	77%	2%
Nonbinary	--	--	--	--
English learner students	45%	55%	52%	3%
Non–English learner students	16%	84%	83%	1%
Students receiving special education services	15%	85%	83%	2%
Students not receiving special education services	24%	76%	75%	2%
White	20%	80%	78%	1%
Hispanic/Latino	38%	62%	60%	2%
Black	23%	77%	75%	2%
Asian	21%	79%	77%	2%
American Indian/Alaskan	52%	48%	45%	3%
Native Hawaiian/Pacific	30%	70%	69%	--
Total	**23%**	**77%**	**76%**	**2%**

Source: District-provided literacy screening assessment data and October and June SIMS collection data
Notes: The percentage of students attending formal EC programs and the percentage attending informal EC programs do not always add up to the percentage with any EC experience due to rounding. Some students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

What Percentage of Kindergarten Students Were Significantly Below Benchmark at BOY, MOY, and EOY Overall and by EC Program Type? How Many Students Overall and by Program Type Were Significantly Below Benchmark Multiple Times?

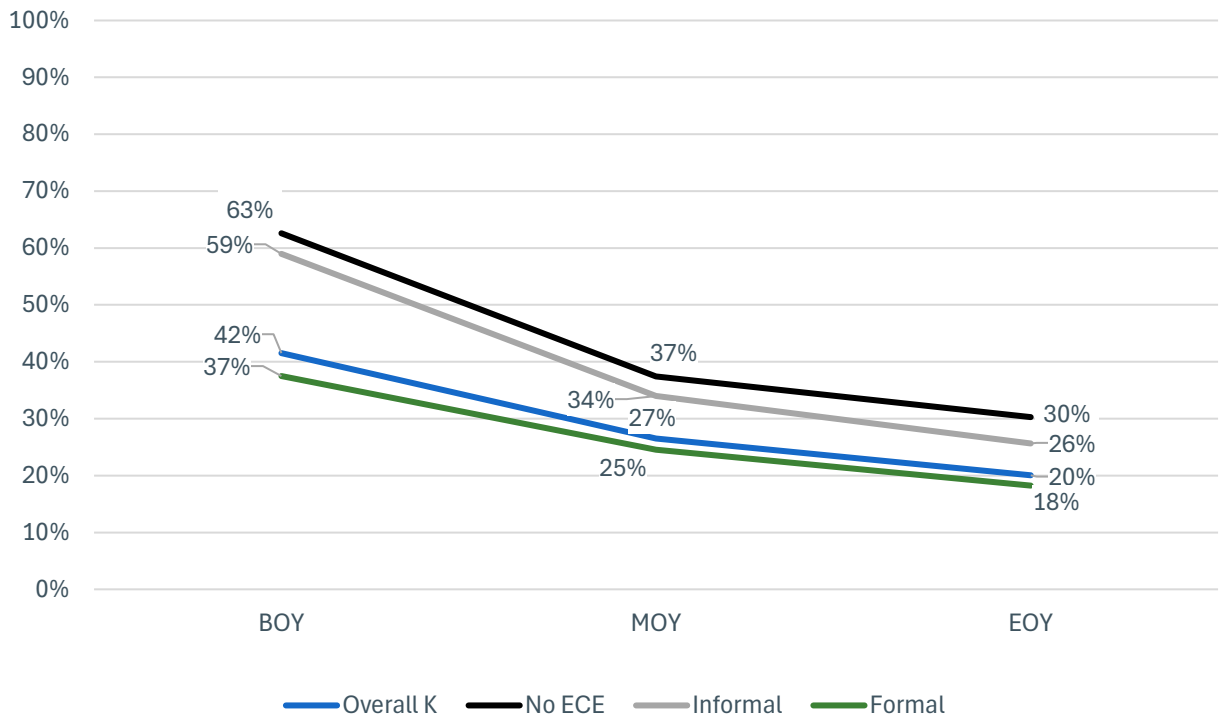
Research dating back decades has shown the impact of preschool on later outcomes for participating children (e.g., the High Scope/Perry Preschool study; Schweinhart et al., 2005), though others note that these programs are not necessarily reflective of current preschool programs and that more research is needed on how and which types of programs are effective (e.g., Stevens & English, 2016). More recent work focuses on identifying specific preschool programs and practices that demonstrate effectiveness in promoting literacy and other skills, (e.g., as summarized by the National Center for Family Literacy [2009], and Herrera et al. [2021]). Evidence from Massachusetts itself shows that preschool can contribute to school readiness. An independent evaluation of the Massachusetts Preschool Expansion Grant (PEG) found that PEG classrooms had a sizeable positive impact on children's early literacy and math skills and a smaller positive impact on vocabulary skills, thereby improving children's readiness for kindergarten (Hofer et al., 2018). Indeed, one reason Massachusetts began systematically collecting data on EC experiences was to enable analysis of how EC experiences relate to later outcomes. In this section, we examine how early literacy performance differs for kindergarten students with and without EC experience.

Key Findings

- Across time periods and student groups, more kindergarten students without EC experience were identified as significantly below benchmark than students with EC experience; students without EC experience were also more frequently identified as significantly below benchmark multiple times during the year.
- Kindergarten students with formal EC experience were less likely to be identified than students with informal experience.

Overall, more kindergarten students without EC experience are identified as significantly below benchmark than students with EC experience (figure 19). This pattern is observed at BOY, MOY, and EOY, though it is important to note that many factors (including experiences even before preschool, preschool quality, and experiences in school after entering kindergarten) can affect student performance. The same pattern is observed using the 25th percentile or below metric, though the percentages of students at significant risk at each time period are smaller.

Figure 19. More Kindergarten Students Without Early Childhood Experience Were Classified as Significantly Below Benchmark at Each Time Period Than Students With Early Childhood Experience



Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Includes students with three scores only. ECE = early childhood experience.

Additionally, kindergarten students without EC experience were more likely to be identified as significantly below benchmark multiple times during the year (table 24). Kindergarten students with formal EC experience were less frequently identified as significantly below benchmark multiple times than were students who participated in informal EC experiences.

Table 24. Percent of Kindergarten Students Significantly Below Benchmark at BOY, MOY, and EOY by Type of Early Childhood Experience

EC experience	BOY (percent significantly below)	MOY (percent significantly below)	EOY (percent significantly below)	Percent significantly below multiple times
No EC program experience	58%	42%	38%	39%

EC experience	BOY (percent significantly below)	MOY (percent significantly below)	EOY (percent significantly below)	Percent significantly below multiple times
Any EC program experience	36%	27%	22%	22%
Informal	59%	41%	32%	36%
Formal	35%	27%	22%	22%
All kindergarten students	39%	30%	24%	25%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Includes students with two or three scores.

Across student groups, kindergarten students with any EC experience were less likely to be identified as significantly below benchmark than were students with no EC experience (table 25). Similarly, students with formal EC experience were generally less likely to be identified than students with informal EC experience. For example, 47 percent of low income students with no EC experience were identified as significantly below benchmark multiple times compared to 35 percent of students with any EC experience.

Table 25. Percent of Kindergarten Students Significantly Below Benchmark Multiple Times by Type of Early Childhood Experience and Student Group

Demographic	No EC experience (percent)	Any EC experience (percent)	Formal (percent)	Informal (percent)
Low income	47%	35%	34%	39%
Non–low income	23%	14%	14%	29%
Female	38%	21%	21%	37%
Male	40%	23%	23%	35%
Nonbinary	--	--	--	--

Demographic	No EC experience (percent)	Any EC experience (percent)	Formal (percent)	Informal (percent)
English learner students	57%	43%	43%	46%
Non–English learner students	25%	18%	18%	30%
Students receiving special education services	50%	42%	42%	55%
Students not receiving special education services	38%	19%	18%	33%
White	37%	21%	21%	30%
Hispanic/Latino	52%	39%	39%	47%
Black	38%	28%	27%	44%
Asian	24%	13%	13%	--
American Indian/Alaskan Native	60%	47%	47%	56%
Native Hawaiian/Pacific Islander	50%	36%	36%	--
Total	**39%**	**22%**	**22%**	**36%**

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Only students with two or three scores were included. Some students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Hispanic students who did not attend EC programs were about 1.9 times more likely to be at risk of being identified as significantly below benchmark multiple times than non-Hispanic students. Black and non-Black students who did not attend EC programs had similar levels of risk.

How Do Student Background Characteristics Interact and Intersect With Early Childhood Program Enrollment and Affect the Likelihood of Students Being Identified as Significantly Below Benchmark in Grades K–3?

EC experience has long been shown to have positive effects on later outcomes for students. However, EC experience may also interact and intersect with other factors affecting student performance in kindergarten and beyond in varying ways.

Key Findings

- EC experience, and more specifically formal EC experience, is associated with reduced likelihood of being identified as significantly below benchmark for students in kindergarten and beyond.
- Effects of formal EC experience were more pronounced for English learner students than for low income students, and for Black English learner students in particular.

In the prior section, we examine the relationship between EC experience and student background factors individually for kindergarten students only. In this section, we explore the effects of student- and school-level factors on the risk of students being identified as significantly below benchmark given EC experience and other characteristics. Results are based on the model described in appendix E, which found formal EC experience to be statistically significant in predicting literacy screener performance during the school year, not only for kindergarten students, but also for all students in grades K–3. This finding suggests that effects of formal EC experience persist over time.

More specifically, results show a decrease in the likelihood of students being identified as significantly below benchmark more than once of 1 to 2 percentage points across gender and racial/ethnic groups for students who attended formal EC programs. Hispanic students were most likely to be identified as significantly below benchmark regardless of EC experience, and Asian students were least likely among racial and ethnic groups to be identified as significantly below benchmark regardless of EC experience. Additionally, results show that formal EC experience decreases the chances of low income students being identified as significantly below benchmark multiple times across racial and ethnic groups (table 26). The likelihood of a low income student being identified as significantly below benchmark multiple times is reduced by about 2 percentage points if they had formal EC experience.

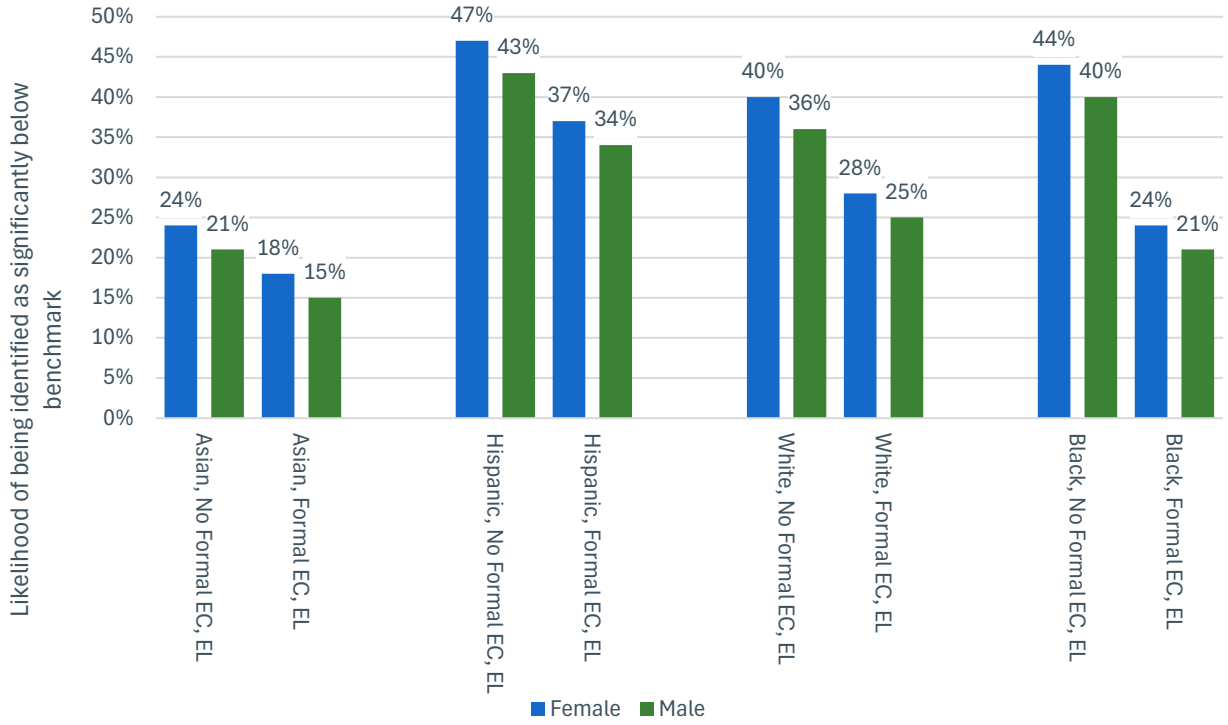
Table 26. Likelihood of Students Being Identified as Significantly Below Benchmark Multiple Times by Formal Early Childhood Experience, Race/Ethnicity, and Low Income Status

Race/ethnicity, formal EC experience, low income status (LI)	Female (percent)	Male (percent)
White, no formal EC	14%	14%
White, formal EC	13%	13%
White, no formal EC, LI	22%	22%
White, formal EC, LI	20%	20%
Hispanic/Latino, no formal EC	18%	18%
Hispanic/Latino, formal EC	17%	17%
Hispanic/Latino, no formal EC, LI	28%	28%
Hispanic/Latino, formal EC, LI	26%	26%
Black, no formal EC	16%	16%
Black, formal EC	15%	15%
Black, no formal EC, LI	25%	25%
Black, formal EC, LI	23%	23%
Asian, no formal EC	7%	7%
Asian, formal EC	6%	6%
Asian, no formal EC, LI	12%	12%
Asian, formal EC, LI	11%	11%

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Effects of formal EC experience were most pronounced among English learner students, and among Black English learner students in particular (figure 20). The probability of Black English learner students being identified as significantly below benchmark is reduced by approximately 20 percentage points (24% compared to 44% for females, 21% compared to 40% for males) when they have attended formal EC programs. Hispanic and White English learner students who attended formal EC programs were also less likely to be identified as significantly below benchmark (a decrease of between 9 and 12 percentage points, depending on race and gender). Asian English learner students who attended formal EC were also less likely to be identified than those who did not attend formal EC, but the difference was smaller than for other groups.

Figure 20. Formal Early Childhood Experience Reduces Risk More for English Learner Students Than it Does for Low Income Students, and Particularly for Black English Learner Students



Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this figure, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model. EL = English learner.

Overall, EC experience, and more specifically, formal EC experience, appears to have benefits for students in kindergarten and beyond and for certain groups of students in particular.

Discussion and Next Steps

This report provides additional analysis of early literacy screening data using a much larger sample of students than previously available. Though this report uses a different approach to reporting on student risk compared to prior analysis, many patterns of student performance are consistent across reporting metrics. For example, though the overall numbers of students identified as at risk are different in previous reporting compared to this year's reporting, most students who are identified as significantly below benchmark (and below benchmark) at the start of the year are still performing below benchmark at the end of the year. As in prior reporting, students at lower grade levels tend to improve more within and across school years than students at higher grade levels, and students' background characteristics are often associated with their performance (e.g., low income students are more often identified as significantly below benchmark than non–low income students).

However, this year's reporting also offers some new information. Additional analysis of screening assessment benchmarks showing how they can vary in difficulty over time provides important information for users at all levels to consider when interpreting student scores. It also reinforces the importance of taking assessment design and intended use into account when using scores. New analysis of the intersections between student background characteristics and performance demonstrates the overlapping, complex, and varying ways in which gender, race, ethnicity, and other factors contribute to risk of reading difficulty. It also highlights student groups that historically have had less access to resources and opportunities, putting them particularly at risk. The analysis also identifies formal early childhood experience as a statistically significant and persistent risk reducer for low income students and English learner students, and especially for Black English learner students. Finally, analysis of national percentiles provided by the publishers shows that Massachusetts students slightly outperform their peers, with about 30 percent of students performing above the 75th percentile at end of year compared to 25 percent nationally. Black students in Massachusetts perform about the same as all other students around the country. White students perform slightly better, Asian students perform significantly better, and Hispanic students perform less well than the national sample that contains all ethnoracial groups. Growth of students using publisher-provided growth norms shows 66 percent of students growing at average or above-average rates compared to about 60 percent nationally.

As more data are collected over time, the extent to which the screening assessment data can provide useful information will likely continue to improve. Future research will be able to track more students across grades and new questions can be explored. Ultimately, however, research and data analysis can only inform literacy policy and practice. It is the design and implementation of policy and practice at the state and local levels that has the potential to truly impact students' learning.

References and Further Reading

Amplify Education, Inc. (2019). *mClass dyslexia screening measures technical manual*. https://amplify.com/pdf/uploads/2019/12/mCLASS_Dyslexia-Screening-Measures-Technical-Manual.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

Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Praeger Publishers.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., McInnes, M., & Holzinger, A. (2021). Deep ROC analysis and AUC as balanced average accuracy to improve model selection, understanding and interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45* (1). https://doi.org/10.1109/tpami.2022.3145392

Cheung, A., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of e3eEducational Research, 82*(4), 351–395. https://doi.org/10.3102/0034654312465472

Condict, D. (2023, April 12). *FastBridge benchmarks and norms overview*. FastBridge. https://fastbridge.illuminateed.com/hc/en-us/articles/1260802348670-FastBridge-Benchmarks-and-Norms-Interpretation-Guide

Cummings, K. D., Smolkowski, K., & Baker, D. L. (2019). Comparison of literacy screener risk selection between English proficient students and English learners. *Learning Disability Quarterly*, *44*(2), 96–109. https://doi.org/10.1177/0731948719864408

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

Dabach, D. B. (2014). "I am not a shelter!": Stigma and social boundaries in teachers' accounts of students' experience in separate "sheltered" English learner classrooms. *Journal of Education for Students Placed at Risk (JESPAR)*, *19*(2), 98–124. https://doi.org/10.1080/10824669.2014.954044

FastBridge Learning. (2019a). *FAST autoreading technical report*. https://support-content.fastbridge.org/KB_Articles/FAST_AUTOreading_Technical_Report.pdf

FastBridge Learning. (2019b). *FastBridge Learning: Benchmarks and norms interpretation and use guidelines (version 4)*. https://viewpointsolution.org/wp-content/uploads/2021/01/FastBridge_Benchmarks_and_Norms_Interpretation_Guide_most-recent.pdf

FastBridge Learning. (2019c). *FastBridge technical manual: Formative Assessment System for Teachers technical manual*. Illuminate Education.

Gaab, N., & Petscher, Y. (2022). *EarlyBird dyslexia and early literacy screener technical manual*. EarlyBird Education.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007-4011). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/PracticeGuide/6

Good, R. H., III, Kaminski, R. A., Cummings, K., Dufour-Martel, C., Peterson, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2011). *DIBELS Next assessment manual.* Dynamic Measurement Group. https://dys-add.com/resources/DiblesNext/DIBELSNext_AssessmentManual.pdf

Good, R. H., III, Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. (2013). *Acadience Reading K–6 technical manual*. Acadience Learning. https://acadiencelearning.org/wp-content/uploads/2020/01/Acadience_Reading_K-6_Technical_Manual.pdf

Gutierrez, K. D., Zepeda, M., & Castro, D. C. (2010). Advancing early literacy learning for all children: Implications of the NELP report for dual-language learners. *Educational Researcher, 39*, 334–339.

Herrera, S., Phillips, B. M., Newton, Y., Dombek, J. L., & Hernandez, J. A. (2021). *Effectiveness of early literacy instruction: Summary of 20 years of research* (REL 2021–084). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Heubeck, E. (2023, February 23). *Universal screening for dyslexia is not enough.* Education Week. https://www.edweek.org/teaching-learning/universal-screening-for-dyslexia-isnt-enough/2023/02

Ho, A. (2012, October). *Off track: Problems with "on track" inferences in empirical and predictive standard setting* [Working paper]. Harvard Graduate School of Education. https://scholar.harvard.edu/files/andrewho/files/off_track_-_andrew_ho_working_paper.pdf

Hofer, K. G., Checkoway, A., Goodson, B., & Nichols, A. (2018). *Massachusetts Preschool Expansion Grant (PEG) Impact Evaluation Report (Research Report).* Abt Associates, Inc.

i-Ready. (2022, August). *Dyslexia screening with* i-Ready Assessment*: Recommendations for using i-Ready to meet Massachusetts dyslexia screening guidelines.* Curriculum Associates Guidance Brief. https://i-readycentral.com/download/?res=64534&view_pdf=1

Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: Evidence from a nine-year longitudinal study. *Journal of Applied Developmental Psychology, 33*(3), 146–157.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*: *Methods and practices,* (2nd ed.). Springer.

Lemke, M., Murphy, D., Soo Ping Chow, A., Spencer, H., & Zhang, A. (2023). *Opportunity gaps: Findings from initial analysis of student performance in early literacy in Massachusetts*. WestEd.

Lexia Learning. (2020). *Lexia RAPID assessment technical manual for RAPID version 5.1: Grades K–2*. https://www.lexialearningresources.com/RAPID/RAPID_K2_Technical_Manual_v5.1.pdf

Lexia Learning. (2022). *Lexia RAPID assessment technical manual for RAPID version 5.1: Grades 3–12*. https://www.lexialearningresources.com/RAPID/RAPID_Technical312.pdf

Mathes, P., Torgesen, J., & Herron, J. (2016). *Istation's Indicators of Progress (ISIP) early reading technical report: Computer adaptive testing system for continuous progress monitoring of reading growth for students pre-K through grade 3*. Istation. https://www.istation.com/hubfs/Content/downloads/studies/er_technical_report.pdf

McClarty, K. L., Murphy, D., Keng, L., Turhan, A., & Tong, Y. (2012, April). *Putting ducks in a row: Methods for empirical alignment of performance standards*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, Canada.

National Center for Family Literacy. (2009). *Developing early literacy: Report of the National Early Literacy Panel*. National Institute for Literacy.

National Center on Intensive Intervention. (n.d). *Classification accuracy*. https://intensiveintervention.org/sites/default/files/Classification_Accuracy_508.pdf

National Research Council. (1998). *Preventing reading difficulties in young children*. National Academies Press. https://doi.org/10.17226/6023

NWEA. (2019). *MAP growth technical report March 2019*. https://www.nwea.org/uploads/2021/11/MAP-Growth-Technical-Report-2019_NWEA.pdf

NWEA. (2020). *Linking study report: Predicting performance on the Massachusetts Comprehensive Assessment System (MCAS) based on NWEA MAP Growth scores*. https://www.nwea.org/uploads/2020/07/MA-MAP-Growth-Linking-Study-Report-2020-07-22.pdf

NWEA. (2021). *Spanish MAP growth reading technical report March 2021*. https://www.nwea.org/uploads/2021/06/Spanish-MAP-Growth-Reading-Technical-Report-2021-03-17.pdf

NWEA. (2021). *English MAP reading fluency technical report August 2021*. https://www.nwea.org/uploads/2021/06/English-MAP-Reading-Fluency-Technical-Report-2022-08-30-1.pdf

Pommerich, M., Hanson, B. A., Harris, D. J., & Sconing, J. A. (2000). *Issues in creating and reporting concordance results based on equipercentile methods* (ACT Research Report No. 2000-1). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2000-1.pdf

Renaissance Learning. (2021). *Star assessments for Spanish: Reading technical documentation*. https://star-help.renaissance.com/hc/en-us/articles/12592399708059-Star-Assessments-for-Spanish-Reading-Technical-Documentation

Renaissance Learning. (2022a). *Defining benchmarks in Star Assessments*. https://star-help.renaissance.com/hc/en-us/articles/24424240092827-Defining-Benchmarks-in-Star-Assessments

Renaissance Learning. (2022b). *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. (2022c). *Star Assessments™ for Reading technical manual.* https://star-help.renaissance.com/hc/en-us/articles/12542471051803-Star-Assessments-for-Reading-Technical-Manual

Sandberg, K. L., & Reschly, A. L. (2011). English learners: Challenges in assessment and the promise of curriculum-based measurement. *Remedial and Special Education, 32*(2), 144–154. https://doi.org/10.1177/0741932510361260

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40 (Monographs of the High/Scope Educational Research Foundation)*. High/Scope Press.

Slavin, R. E., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, *33*(1), 47–58. https://doi.org/10.3102/0162373711398127

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), 282–306. https://doi.org/10.3102/0002831216634463

Stevens, K., & English, E. (2016). *Does pre-K work? The research on ten early childhood programs—and what it tells us*. American Enteprise Institute. https://www.aei.org/wp-content/uploads/2016/04/Does-Pre-K-Work.pdf

Thng, Yi Xe. 2019. *Essays on educational testing in an era of higher ("college-ready") standards* [Doctoral dissertation]. Harvard Graduate School of Education.

Umansky, I., & Reardon, S. (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

Umansky, I. M., Callahan, R. M., & Lee, J. C. (2020). Making the invisible visible: Identifying and interrogating ethnic differences in English learner reclassification. *American Journal of Education, 126*(3), 335–388. https://doi.org/10.1086/708250

University of Oregon. (2022). *Revised zones of growth for DIBELS 8th Edition benchmark assessments (Technical report 2202)*. https://dibels.uoregon.edu

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

Yale Center for Dyslexia & Creativity. (n.d.). *Dyslexia FAQ*. https://dyslexia.yale.edu/dyslexia/dyslexia-faq/

Appendix A. Assessment Review Criteria and Summaries

Table A.1 provides the established criteria that the Massachusetts Department of Elementary and Secondary Education (DESE) used in 2022 to evaluate the early literacy universal screening assessments that were being considered for use in the state. The criteria were developed by a state panel of researchers, teachers, administrators, and other specialists.

Table A.1. Massachusetts's Early Literacy Universal Screening Assessment Criteria

Criteria category	Assessment details

Constructs measured	• alphabetic knowledge
	  ○ letter identification (Kindergarten [K])
	  ○ letter/sound knowledge (K,1,2)
	  ○ decoding nonsense words (K,1,2)
	• phonemic awareness
	  ○ phoneme segmentation (K,1)
	• word reading
	  ○ word identification (1)
	  ○ passage reading fluency (1,2)
	  ○ reading comprehension (2)
	• vocabulary
	• listening comprehension/oral language comprehension
	• rapid automatized naming
	  ○ included in overall assessment of risk
	  ○ valid
	  ○ features following descriptors:
	  ▪ timed administration
	  ▪ a set of at least 36 familiar objects or letters (stimuli) presented serially and pseudorandomized
	  ▪ five or six different token items (stimuli)
	  ▪ requires student to read continuously from left to right and engage in a "return to sweep" to read on the next line
	  ▪ nationally normed and criterion referenced

Criteria category	Assessment details
Technical adequacy	• classification accuracy ○ reviewed by the National Center on Intensive Intervention (NCII) and rated "Convincing Evidence" or "Partially Convincing Evidence" for Classification Accuracy for a composite/overall score for grades kindergarten, 1, and 2 for fall, winter, and spring. *If the assessment has not, in its current form, been reviewed by NCII, evidence is presented of meeting NCII criteria for Classification Accuracy*. • provides a composite rank and a risk level for each student based on predetermined (external) cut scores • provides a percentile rank for sub-measures (i.e., alphabetic principle) • sample used to set norms is comparable and relevant for Massachusetts • availability of peer-reviewed validation studies
Attention to linguistic diversity	• guidance provided to interpret scores for bi-/multilingual and/or English learner students • normed for English learner students to allow for accurate identification of risk • some or all subtests available in language(s) other than English; subtests are linguistically and culturally authentic rather than directly translated from English • provides specific instructions for whether and how student directions and/or assessment probes should be presented in student's home language ○ bias analysis conducted and reviewed by the NCII to examine the degree to which the tool is or is not biased against certain subgroups (race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners). *If the assessment has not, in its current form, been reviewed by NCII, evidence is presented of meeting NCII criteria for Bias Analysis.*
Administration usability and support	• designed to be administered at least twice per year in kindergarten and three times per year in grades 1 and 2, at a minimum • assessment takes less than 60 minutes to administer as a whole group, regardless of setting; requires less than 15 minutes to administer individually • includes progress monitoring tools; describes how they are used to assess students' rate of improvement or responsiveness to instruction • provides supportive resources to guide school-based administrators and educators in effective assessment administration, data access, and data-based instructional decision-making • provides a student report designed for families • utilized in Massachusetts schools within the last 3 years

Tables A.2–A.10 provide DESE's summaries of the early literacy universal screening assessments that currently meet expectations or partially meet expectations according to their assessment criteria presented above.

Table A.2. MA DESE's Summary of Amira, From Amira Learning

Approval status	Approved: Meets expectations
Grades covered	K–5
Description	Computer adaptive assessment that utilizes Artificial Intelligence and speech recognition software for universal screening and benchmarking. Hand scoring is possible as student recordings are available for playback. **A computer adaptive assessment adjusts to the student's performance.*
Administration time and setting	Up to 15 minutes per student
Paper or digital	Digital
Languages	English and Spanish
Skills assessed in grades K–2	☐ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Objects, letters, numbers ☒ Letter sound correspondence ☒ Vocabulary ☒ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Annual subscription for Amira Assessment is $8 per student. Amira Suite, which includes a tutoring program, is $20 per student.
Initial implementation support available	As part of the subscription for Amira Assessment ($8), a "Getting Started Session," which is a half day of training, is included in the software-as-a-service model. A range of professional development and workshops are available online at no cost. Continuous coaching is available through a subscription at a cost of $4,200.
For more information	https://amiralearning.com Bryan McCorkle Bryan.mccorkle@amiralearning.com 561-521-7692

Table A.3. MA DESE's Summary of DIBELS 8th Edition, From University of Oregon

Approval status	Approved: Meets expectations
Grades covered	K–8
Description	A set of 1-minute fluency measures that can be used for universal screening, benchmarking, and progress monitoring
Administration time and setting	5–8 minutes per student; individually administered
Paper or digital	Paper; digital data services (DDS) available through Amplify
Languages	English and Spanish
Skills assessed in grades K–2	☐ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming (RAN): Uses a letter naming fluency task as an indicator for RAN ☒ Letter sound correspondence: A separate score is included as part of decoding nonsense words task ☐ Vocabulary ☐ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring forms with scoring included
Assessment costs	Materials can be downloaded for free. DIBELS Data System (DDS) for data storage and reporting available through Amplify for a fee.
Initial implementation support available	Introduction: Free training resources Comprehensive training: DIBELS website provides a list of certified trainers to provide in-person or virtual training (not DESE approved or verified).
For more information	https://dibels.uoregon.edu/ University of Oregon DIBELS Data System (DDS) through Amplify: https://dibels.amplify.com

Table A.4. MA DESE's Summary of EarlyBird, From EarlyBird Education

Approval status	Approved: Meets expectations
Grades covered	Kindergarten only
Description	Tablet-based game that assesses the student while they play, with self-administration and auto-scoring.
Administration time and setting	Less than 15 minutes to administer and establish dyslexia risk score; 45 minutes to administer and score entire assessment; small group or individually administered with oversight from an adult
Paper or digital	Digital; computer-adaptive
Languages	English only
Skills assessed in K–2	☒ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☐ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Pictures only ☒ Letter sound correspondence ☒ Vocabulary ☒ Listening comprehension/oral language comprehension
Progress monitoring	Currently, there are no progress monitoring tools.
Assessment costs	Annual per student license $8
Initial implementation support available	Annual platform access and support services $700 per participating cohort (typically organized by elementary school); includes implementation planning, kickoff training, access to data dashboard and Next Steps Resource Library
For More information	https://earlybirdeducation.com/ Steve Sandak steve.sandak@earlybirdeducation.com 617-462-4779

Table A.5. MA DESE's Summary of mCLASS, From Amplify

Approval status	Approved: Meets expectations
Grades covered	K–6
Description	Digital administration of DIBELS 8th Edition, 1-minute fluency measures that can be used for universal screening, benchmarking, and progress monitoring
Administration time and setting	3–6 minutes per student
Paper or digital	Digital
Languages	English and Spanish
Skills assessed in K–2	☐ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming (RAN): Uses a letter naming fluency task as an indicator for RAN ☒ Letter sound correspondence: A separate score is included as part of decoding nonsense words task ☒ Vocabulary ☒ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Annual per-student license $14.90, discounts may apply K–2 class kit (optional) $47 per classroom Add-on charge for Lectura (Spanish version) and dual language reporting
Initial implementation support available	Two half-day remote webinars: $1,500; half-day: $750; 90 minutes: $500 Two days onsite (consecutive days): $4,800; Full-day onsite workshop: $3,200; half-day onsite: $2,500
For more information	https://amplify.com/programs/mclass/ Jesse Paprocki jpaprocki@amplify.com 920-737-2727

Table A.6. MA DESE's Summary of Acadience Reading, From Voyager Sopris

Approval status	Approved: Partially meets expectations
Grades covered	K–6
Description	Previously known as DIBELS Next, a set of 1-minute fluency measures that can be used for universal screening, benchmarking, and progress monitoring
Administration time and setting	3–8 minutes per student
Paper or digital	Paper and digital
Languages	English and Spanish
Skills assessed in K–2	☐ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☐ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Objects, letters, numbers ☒ Letter sound correspondence: A separate score is included as part of decoding nonsense word task ☐ Vocabulary: Task is experimental and untimed ☐ Listening comprehension/oral language comprehension: Task is experimental and untimed
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Paper: Materials sold in bundles of 5, 6 or 25. Contact DESE for cost details Digital: per student $7.95 (discounts may apply) includes digital administration, dashboard, reports using the Acadience Learning Online system
Initial implementation support available	Two full days onsite professional development: $8,000 for up to 65 participants, plus $17 per participant for training manual Virtual Professional development: $129 for 12-hour online course to be completed within 30 days
For more information	https://www.voyagersopris.com/product/assessment/acadience-reading/overview Laurie Carmon Laurie.carmon@voyagersopris.com, 214-932-9404

Table A.7. MA DESE's Summary of FastBridge aReading and earlyReading, From Illuminate Education

Approval status	Approved: Partially meets expectations
Grades covered	K–12
Description	For screening, MA educators will use aReading (a computer adaptive assessment*) for grades 2 and above and use earlyReading (brief, individually administered fluency probes) for kindergarten and grade 1. Other assessments are included in the suite. **Computer adaptive assessment adjusts to the student's performance.*
Administration time and setting	aReading: 30 minutes; adaptive, computer-based; group or individual earlyReading: 5 minutes; student responses recorded electronically; individual
Paper or digital	Paper and digital
Languages	English and Spanish (some subtests available in Spanish)
Skills assessed in K–2	☒ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Uses a letter naming fluency subtest; letters only ☒ Letter sound correspondence ☒ Vocabulary ☒ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Digital: per student $8; plus first year implementation fee based on number of licenses—under 500: $500; 51–1,500: $1,500; 1,501+: $2,000
Initial implementation support available	Onsite professional development: $3,250/day for up to 30 participants per cohort Virtual Professional development: $1,500/day for up to 30 participants per cohort
For more information	https://www.renaissance.com/products/fastbridge/ Jay Anderson Jay.anderson@renaissance.com 612-424-3719

Table A.8. MA DESE's Summary of i-Ready, From Curriculum Associates

Approval status	Approved: Partially meets expectations
Grades covered	K–12
Description	For screening, MA educators will primarily use i-Ready Diagnostic (a computer adaptive diagnostic assessment*) and i-Ready Literacy Tasks (brief, individually administered fluency probes). Other assessments are also included in the suite. **A computer adaptive assessment adjusts to the student's performance.*
Administration time and setting	i-Ready Diagnostic for kindergarten and grade 1: 25–35 minutes active testing time; 40–60 minutes active testing time for grade 2 (computer-based). i-Ready Literacy Tasks: 1–2 minutes each task (administered individually).
Paper or digital	Digital; computer-adaptive
Languages	English and Spanish
Skills assessed in K–2	☒ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Letters, numbers, pictures, colors ☒ Letter sound correspondence ☒ Vocabulary ☐ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Annual per-student license $6.00; discounts available (minimum 150 licenses)
Initial implementation support available	Onsite or virtual support: $2,000 per session for up to 6 hours. Six-hour virtual sessions may be split among several days; onsite sessions are 1 day, one location
For more information	https://www.curriculumassociates.com/products/i-ready/i-ready-assessment Brian O'Mara 978-844-4883 bomara@cainc.com

Table A.9. MA DESE's Summary of MAP Reading Fluency, From NWEA

Approval status	Approved: Partially meets expectations
Grades covered	K–5
Description	A computer-adaptive* assessment, which can be used for universal screening and benchmarking. Hand scoring is possible as student recordings are available for playback. **A computer adaptive assessment adjusts to the student's performance.*
Administration time and setting	20–40 minutes; whole class, small group, or individual student
Paper or digital	Digital; computer-adaptive
Languages	English and Spanish
Skills assessed in K–2	☒ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☒ Word reading/word identification ☒ Letter identification ☐ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ Rapid automatized naming: Pictures only ☒ Letter sound correspondence ☒ Vocabulary ☒ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Annual per-student license $9 (discount if bundled with other NWEA assessments)
Initial implementation support available	Virtual workshop or consulting session: $1,200 Full day onsite workshops: $3,600; half-day workshops: $2,500 Self-directed MAP reading fluency basics: $500
For more information	https://www.nwea.org/map-reading-fluency/ Jackie Cheney Jackie.Cheney@NWEA.org 860-941-1823

Table A.10. MA DESE's Summary of STAR Elementary Bundle (Early Literacy, Reading, Curriculum-Based Methods) From Renaissance

Approval status	Approved: Partially meets expectations
Grades covered	K–3 Early Literacy; 1–12 Reading; K–6 Curriculum-Based Methods (CBM) Reading
Description	For universal screening, MA educators will use two or all three assessments depending on the student's grade level. Star Early Literacy (grades K–3) is a computer adaptive assessment* that measures early literacy skills. Star Reading (grades 1–12) is a computer adaptive reading assessment which measures comprehension and vocabulary of independent readers. Star CBM Reading is administered one-to-one with a focus on foundational skills and fluency. **A computer adaptive assessment adjusts to the student's performance*.
Administration time and setting	Star Early Literacy and Star Reading: 20 minutes each (computer-based) Star CBM Reading: 60–90 second probes (individually administered)
Paper or digital	Digital; computer-adaptive
Languages	English and Spanish
Skills assessed in K–2	☒ Phonological awareness (rhyme, syllable, onset rime) ☒ Phonemic awareness (phoneme isolation, phoneme segmentation) ☐ Word reading/word identification ☒ Letter identification ☒ Decoding nonsense words ☒ Passage reading fluency ☒ Reading comprehension ☒ RAN: Colors, pictures ☒ Letter sound correspondence ☒ Vocabulary ☐ Listening comprehension/oral language comprehension
Progress monitoring	Progress monitoring tools and scoring included
Assessment costs	Annual per-student license $9
Initial implementation support available	Onsite full day: $3,000 Virtual Sessions (Six 60-minute sessions): $1,500
For more information	https://www.renaissance.com/products/star-assessments/ Diane Houle Diane.Houle@renaissance.com 774-413-0061

Appendix B. Screening Assessment Benchmarks

Table B.1 provides the screening assessment risk levels used to identify students as below benchmark or significantly below benchmark for the 2020/21 and 2021/22 analyses.

Table B.1. Description of Risk Levels and Additional Reading Risk or Dyslexia Flags for Early Literacy Screening Assessments Included in 2020/21 and 2021/22 Analysis

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
Acadience Reading	2020/21	Acadience provides four levels to describe student performance for a reading composite score and for subtest scores. The levels indicate the overall likelihood of achieving subsequent proficiency goals (without targeted instructional support) and the overall level of need for students in these benchmark categories. The levels are as follows: • Above Benchmark: Likelihood of achieving subsequent early literacy goals is 90%–99% (core support needed)	Not available

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
		• At Benchmark: Likelihood of achieving subsequent early literacy goals is 70%–85% (core support needed; students near the benchmark cut score may require monitoring and/or Strategic Support on specific skills) • Below Benchmark: Likelihood of achieving subsequent early literacy goals is 40%–60% (strategic support needed)* • Well Below Benchmark: Likelihood of achieving subsequent early literacy goals is 10%–20% (intensive support needed; students in this benchmark category are at risk of not achieving reading goals unless intensive support is provided)*, **	
DIBELS 8th Edition	2020/21, 2021/22	DIBELS 8th Edition provides four levels to describe student performance for a reading composite score and for subtest scores. Scores represent the overall level of need for students and their risk of not achieving proficiency goals, as follows: • Above Benchmark: Core support; negligible risk; nearly all students in this category score at or above the 40th percentile on criterion measure* • At Benchmark: Core support; minimal risk; 80% of students who score at or above the 40th percentile on criterion measure fall in this category* • Below Benchmark: Strategic support; some risk; 80% of students who score below the 40th percentile on criterion measure fall in this category* • Well Below Benchmark: Intensive support; at risk; classifies students who are at risk of reading difficulties, including dyslexia; 80% of students who score below the 20th percentile on criterion measure fall in this category*, **	The publisher suggests that risk on the Letter Naming Fluency and Phonemic Segmentation Fluency subtests in kindergarten and grade 1 and Nonsense Word Fluency in grades 1–3 could be used to understand potential risk of dyslexia. However, these subtests do not aim to provide a dyslexia diagnosis nor are additional results or flags provided based on these subtests.
mCLASS	2020/21, 2021/22	See description for DIBELS 8th Edition (mCLASS assessments are based on DIBELS 8th Edition, and reported performance levels are the same)	mCLASS provides a risk indicator that uses supplemental measures (Vocabulary, Spelling, and Rapid Automatized Naming) to screen for risk related to dyslexia.Indicator scores are as follows:

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
			• Low risk • At risk*** Students are classified as at risk of reading difficulties (including dyslexia) if they are classified as at risk (i.e., Well Below Benchmark) according to their composite score and classified as at risk on the spelling and/or Rapid Automatized Naming subtests.

EarlyBird	2021/22 (available for kindergarten only)	EarlyBird provides a metric at each time period to identify students at risk of reading difficulties. In beginning of year (BOY), a flag identifies students performing "below expectations." In middle of year (MOY), students receive a Potential for Word Reading (PWR) likelihood percentage, which is the probability that a student will reach grade-level expectations in word reading by end of year (EOY) without remediation. According to the publisher, "Reaching expectations, for the purposes of this analysis, is defined as performing above the 40th percentile on the SAT-10: a reasonable standard for measuring grade-level expectation word reading." In EOY, EarlyBird refers to the Word Reading subtest score, which is available only to kindergarten students at EOY. Percentile ranks are used to describe a student's performance on each subtest. The EarlyBird benchmarks are as follows for each time period:
		• BOY
		  ○ At/Above Expectations
		  ○ Below Expectations*, **
		• MOY
		  ○ Not at Risk: At/Above 64th percentile
		  ○ At Risk: Below the 64th percentile*, **
		• EOY
			EarlyBird provides a dyslexia risk flag that indicates the likelihood that a student will be at risk of severe word reading struggles at EOY (provided the student doesn't receive appropriate remediation). According to the publisher, "severe word reading struggles are defined as performing at or below the 20th percentile on the SAT-10 (Stanford Achievement Test Series, 10th Ed., 2018, Pearson Education, Inc.). The calculation involves a selection of our most predictive subtests and an aggregation and weight averaging of that data according to degree of predictability to generate a single output score, which is conveyed as a 'flag.'" Indicator scores are as follows:
			• Not Flagged
			• Flagged***

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
		○ Not at Risk: Above 40th percentile ○ At Risk: 21st–40th percentile* ○ At Significant Risk: Below 21st percentile*, ** Due to missing BOY data and delays in obtaining benchmark information, EarlyBird scores were not used in analyses describing students at risk overall, but were used to describe students at risk of dyslexia (see description of dyslexia risk flag at right).	
FastBridge aReading	2021/22	FastBridge aReading provides four levels to describe student performance for the composite scaled scores. Per FastBridge, these benchmarks "were established for FastBridge aReading to help teachers accurately identify students who are at risk for not meeting the current grade level expectations as measured by future performance on important tests such as the state assessment." The FastBridge benchmarks are based on its national norms and correspond to the following percentile ranges: • Advanced/College Pathway: 71st–99th percentile • Low risk: 40th–70th percentile • Some risk: 15th–39th percentile* • High risk: Below the 15th percentile*, **	Not available
FastBridge CBMreading	2020/21, 2021/22	FastBridge CBMreading provides four levels to describe student performance for the words read correctly per minute score. Benchmark levels are not available for kindergarten students. Per FastBridge, benchmarks "were set by examining data from students who completed both the FastBridge CBMreading assessment and another 'high stakes' assessment such as a state test. … Results indicate that FastBridge™ CBMreading is highly predictive of student's [sic] scores on other reading assessments." The FastBridge benchmarks are based on its national norms and correspond to the following percentile ranges: • Advanced/College Pathway: 71st–99th percentile	Not available

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
		• Low risk: 40th–70th percentile • Some risk: 15th–39th percentile* • High risk: Below the 15th percentile*, **	
FastBridge earlyReading	2020/21, 2021/22	FastBridge earlyReading provides three levels to describe student performance for composite and subtest scores. Benchmarks are not available for grade 2 and grade 3 students. Per FastBridge, benchmarks "were developed from a criterion study examining FastBridge™ earlyReading assessment scores in relation to scores on the Group Reading Assessment and Classification Evaluation." The benchmarks are based on the national norms and correspond to the following percentile ranges: • Low risk: 40th–99th percentile • Some risk: 15th–39th percentile* • High risk: Below the 15th percentile*, **	Not available
i-Ready Diagnostic and Literacy Tasks	2020/21, 2021/22	The i-Ready Diagnostic test provides five benchmarks (referred to as relative placement levels) for composite and subtest scale scores. These benchmarks are criterion referenced (i.e., based on judgments about performance relative to expectations set by the Common Core State Standards, not based on normative data about student performance). Benchmarks can be used to determine whether students are meeting grade-level expectations. The levels are as follows: • Mid or Above Grade Level • Early on Grade Level • 1 Grade Level Below* • 2 Grade Levels Below* • 3 or More Grade Levels Below*	The i-Ready Diagnostic test also provides a specific Reading Difficulty Indicator (iRDI), which is a cut score that identifies students who may be struggling as readers. This indicator is calculated by using below-level cut scores and typical growth measures to determine which scores at each time period and grade may be considered indicators of possible reading difficulty that could require further investigation. Students are either flagged or not flagged based on iRDI cut scores: • No iRDI Flag • iRDI Flag**

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
Istation Indicators of Progress (ISIP) Early Reading (ER)	2021/22	ISIP ER reports three levels based on norms associated with a composite scaled score. Students with an index above the 40th percentile for their grade are placed into Tier 1. Students with an index at or below the 20th percentile are placed into Tier 3. These tiers are used to guide educators in determining the level of instruction for each student, as follows: • Tier 1 students (above the 40th percentile) are on track and performing at grade level. • Tier 2 students (between the 21st and 40th percentile) are at some risk, are performing moderately below grade level, and are in need of intervention.* • Tier 3 students (20th percentile and below) are at risk, are performing seriously below grade level, and are in need of intensive intervention*, **	Not available

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
Lexia RAPID	2020/21, 2021/22	Lexia RAPID reports three performance levels based on its Reading Success Probability score. Per Lexia RAPID, "The Reading Success Probability Score (RSP) is calculated by a combination of a student's performance in the Word Recognition, Vocabulary Knowledge, Syntactic Knowledge and Reading Comprehension tasks. This formula is based on the student's grade level, since the factors that are most predictive of reading comprehension success change as a student grows older." The levels are as follows: • High likelihood of EOY grade-level success: An RSP of 70% or higher means that a student has a high likelihood of reaching EOY grade-level success. A student with an RSP in this range will continue to benefit from universal instruction. • Moderate likelihood of EOY grade-level success: An RSP between 69% and 31% means that a student has a moderate likelihood of reaching EOY grade-level success. A student with an RSP in this range may need additional instruction to target skill weaknesses.* • Low likelihood of EOY grade-level success: An RSP of 30% or lower means that a student has a lower likelihood of reaching EOY grade-level success. A student with an RSP in this range may need more intensive instruction to target skill weaknesses.*, **	Not available

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
MAP Growth	2021/22	MAP Growth reports two levels based on MAP Growth reading scores and associated normative data. Specifically, the publisher reports that "classification accuracy analyses results suggest the benchmarks be set at the 30th percentile in MAP Growth Reading and Mathematics for Grades K–8. … Students who score below those benchmarks are likely at risk for severe learning difficulty and in need of intensive intervention." The levels are as follows: • No intensive intervention • Intensive intervention*, ** Note that MAP Growth also provides other performance levels for grade 3 students (Not Meeting, Partially Meeting, Meeting, and Exceeding Expectations) that are designed to describe which students are on or off track to meet Massachusetts Comprehensive Assessment System proficiency standards.	Not available
MAP Reading Fluency	2021/22	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. Monitoring and/or intervention may be appropriate to improve this student's reading outcomes. A flag on this screener does not indicate a diagnosis of reading disability." Not all students will receive a Universal Screener outcome—receiving a result depends on the test and language they were assigned, skills assessed, and their grade at the time of testing. Indicator scores are as follows: • Not flagged • Flagged*, **	The MAP Reading Fluency Dyslexia Screener provides a binary "Dyslexia Screener outcome flag" that "suggests possible risk factors for dyslexia or other reading difficulties. A flag does not indicate a diagnosis of dyslexia or reading disability." Students are flagged for risk factors of dyslexia or other reading difficulties using a predictive model with multiple measures, including phonological awareness, phonics and word recognition, language comprehension and sentence reading fluency domains, student grade, and time of year. Indicator scores are as follows: • Not flagged • Flagged***

Early literacy screening assessment	Years with data	Benchmark or risk-level descriptions	Dyslexia and/or additional reading risk flag descriptions****
Star CBM	2021/22	Star CBM provides two or three benchmark levels for each of the subtests included in the assessment; however, no composite score or overall benchmark levels to describe reading risk are available. Star CBM was therefore not included in analysis.	Not available
Star Early Literacy	2021/22	Star Early Literacy provides four levels based on the composite scaled score; levels are established based on normative data. Per the publisher, the default benchmark is the 40th percentile "based on review of proficiency cut scores from several state assessments and guidance from RTI [response to intervention] experts," which identifies students who "require some form of intervention to accelerate their growth and bring them into benchmark range." The levels are as follows: • At/Above Benchmark: Students meeting or exceeding the benchmark score (at or above the 40th percentile) • On Watch: Students slightly below the benchmark score (automatically calculated range that is between the at/above benchmark level and the intervention level)* • Intervention: Students below the benchmark score (below the 25th percentile)* • Urgent Intervention: Students far below the benchmark score (below the 10th percentile)*, **	Not available
Star Early Literacy Spanish	2021/22	See description for Star Early Literacy.	Not available
Star Reading	2021/22	See description for Star Early Literacy.	Not available

Source: Authors' compilation based on assessment documentation and/or communication with publishers (see references for list of technical reports and other documentation reviewed).
Note: i-Ready Diagnostic, combined with the i-Ready Literacy Tasks, is currently approved to be administered in Massachusetts. The information presented in the table and in the report only pertain to i-Ready Diagnostic scores as Literacy Task data were not available. MAP Reading Fluency data from 2021/22 do not contain any dyslexia

screening assessment results. EarlyBird is a kindergarten-only assessment. *Single asterisk and red-colored text indicate levels used in reporting on numbers or percentages of students at risk or not meeting benchmarks that indicate likelihood of reading success. For example, students in the "below benchmark" or "well below benchmark" level for DIBELS 8th Edition are reported as "did not meet benchmark" or at risk. **Double asterisks and red-colored text indicate levels used in reporting on numbers or percentages of students at significant risk. For example, students in the "well below benchmark" level for DIBELS 8th Edition are reported as "at significant risk." ***Iasterisks and red-colored text indicate levels used in reporting on numbers or percentages of students at potential risk of dyslexia (based on dyslexia screening assessment indicator). Generally, these indicators are intended to identify students who need additional screening. ****DESE did not review or approve screening assessments specifically for dyslexia flagging.

Appendix C. 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. 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, 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 2022/23 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., vendor-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 vendor); 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 assessment.

The analytic team used vendor-defined cut scores (obtained through the assessment technical manuals or communication with the vendors) to create composite and subtest benchmark variables and a reading risk flag variable. Although these variables were typically available in the DESE or vendor-provided files, it was possible for districts/schools to customize the benchmark cut scores for some screening assessments. Therefore, the team elected to use the vendor-defined benchmark levels and reading risk flags, when possible, for analyses. When it was not possible to generate the vendor-defined benchmark level or reading risk flag status due to missing information, the district- or school-provided levels were used. The MAP Reading Fluency Universal Screener flag and the aimswebPlus composite benchmark level were not calculated by the analytic team because vendor-defined cut scores were not available. The MAP Reading Fluency flag is generated by NWEA using a multivariate predictive model, 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.

In some instances, differences between the benchmark level provided in the screening assessment file and the benchmark level generated based on the vendor-defined cut scores were likely due to the administration of an off-grade-level test. In these instances, the observations were excluded from the analyses as 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 vendors were used to determine the expected time period. Because Istation Indicators of Progress 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.

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

- Observations with earlier administration dates were selected (as later scores in the same time window were potentially scores being used for progress monitoring rather than screening).

- If one observation had more data or if one record had a composite score and another did not, the observation with more data or a composite score was selected.
- If a file did not have administration dates, the lowest score within the time period was selected, as the higher score was assumed to be a progress monitoring measure after instruction.
- 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.

Additionally, during the screening assessment cleaning process for the 2022/23 data, the composite score for 39 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 vendor]). Further, 502 DIBELS 8th Edition composite scores were replaced with composite scores generated by the analytic team using the composite score formulas in the technical manual. This recalculation was done only for files in which the data appeared to be manually entered and was done 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.

Finally, 5 DIBELS 8th Edition observations, 19 mCLASS observations, 35 Star Early Literacy observations, and 18 Star Reading observations were not used for the analyses because students were administered off-grade-level forms (i.e., assessment forms not associated with their grade level at the time of testing).

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, and 2022/23 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., the 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 screening data with the state education data. First, the combined screening assessment file was merged with the SIMS data using the June 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, grade level, school code, and district code. As 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. If a match did not occur between the assessment data and SIMS data using these student, school, and district identifiers, we then attempted to match the assessment data, using these same variables, with the October SIMS data. If a student's screening assessment data did not match with the October SIMS data using these variables, we then attempted to match the assessment data with their appropriate June SIMS observation using the student's State Assigned Student Identifier (SASID), grade level, and district code (i.e., without the school code as a matching variable). The process was repeated with the October data for remaining observations, followed by a merge based on SASID and grade level alone, followed by a merge based solely on SASID. Some files did not contain the SASID; rather, they contained only the student's locally assigned identifier (LASID), which is unique at the district level. As the identifier is not unique at the state level, all attempted merges with LASIDs were conducted using the district code as a merging variable.

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 MA DESE Education Data*[21] to determine the

[21] The *Researchers' Guide to MA DESE Education Data* can be found at https://www.doe.mass.edu/research/researchers-guide.docx

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. If two observations had the same enrollment status, the observation with the larger number of days of membership was selected.

Out of the 325,506 screening assessment observations for the 2020/21, 2021/22, and 2022/23 school years, 325,029 (99.9%) matched with a corresponding observation in the SIMS data. Of the observations that matched, 324,298 (99.8%) matched with an observation in the June SIMS data and 731 (0.2%) matched with an observation in the October SIMS data. Of the screening assessment observations, 8,597 did not match exactly with the SIMS data (i.e., did not match using the grade level, school code, and district code); 45 had a different grade level than their corresponding SIMS observation, 139 had different district codes, and 8,561 had different school codes. Ninety-five percent of the school differences (8,147 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 dataset were used and the grade levels from the SIMS dataset were used.

After the screening assessment data were merged with the SIMS collection data, the combined file was merged with the MCAS and ACCESS datasets 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 datasets 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, and 2022/23 was retrieved from DESE's school and district profiles website to provide contextual data about the sample of the students used in analysis. Overall, the data pertains to the following four main categories of information: (a) educator characteristics, (b) student performance, (c) student enrollment and demographics, and (d) financial information. This data was merged with the student-level longitudinal file using the school and district codes from the screening assessment dataset. For the observations missing the school code from the screening assessment dataset, the SIMS school code was used to merge with the publicly available school-level data.

This finalized file with student-level screening assessment and state education data, and publicly available school- and district-level data, was provided to DESE. Data are organized as a single longitudinal dataset with one observation per student, 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 D. Linking

Table D.1 provides the correlation of literacy screening assessment scores with Massachusetts Comprehensive Assessment System (MCAS) Grade 3 English language arts (ELA) assessment scores by screening assessment, grade level, and time period.

Table D.1. Correlation of Grade 2 BOY, Grade 2 EOY, Grade 3 BOY, and Grade 3 EOY Literacy Screening Assessment Scores With MCAS Grade 3 ELA Assessment Scale Scores

Early literacy screening assessment	*N*	Composite SS: *M* (*SD*)	MCAS ELA SS: *M* (*SD*)	Correlation
Grade 2 BOY				
DIBELS 8th Edition	1,249	335.22 (33.24)	493.98 (23.16)	0.67
i-Ready	700	465.99 (47.51)	499.03 (20.54)	0.76
Lexia RAPID	489	16.74 (23.24)	488.73 (21.32)	0.64
mCLASS	423	332.75 (25.38)	494.40 (20.43)	0.68
Star Early Literacy	192	811.07 (87.51)	488.76 (21.05)	0.78
Star Reading	56	786.68 (123.08)	485.30 (21.13)	0.71
Grade 2 EOY				
DIBELS 8th Edition	1,285	452.54 (39.74)	493.88 (23.02)	0.7
i-Ready	676	513.67 (45.95)	498.99 (20.51)	0.8
Lexia RAPID	500	39.80 (33.76)	488.27 (21.29)	0.71
mCLASS	442	444.13 (31.80)	491.90 (20.41)	0.72
Star Early Literacy	523	833.88 (90.16)	478.70 (21.22)	0.7
Star Reading	210	895.73 (77.62)	487.52 (18.74)	0.71

Early literacy screening assessment	*N*	Composite SS: *M* (*SD*)	MCAS ELA SS: *M* (*SD*)	Correlation
Grade 3 BOY				
Acadience Reading	137	273.64 (206.04)	498.82 (18.84)	0.68
aimswebPlus	170	387.26 (58.69)	496.69 (21.58)	0.82
DIBELS 8th Edition	2,753	343.29 (37.48)	493.56 (23.16)	0.71
FastBridge aReading	104	501.48 (26.73)	502.28 (18.74)	0.70
i-Ready	2,561	492.27 (60.96)	492.20 (23.88)	0.84
Lexia RAPID	472	33.42 (33.93)	488.71 (22.12)	0.71
mCLASS	2,432	341.24 (32.41)	492.71 (22.54)	0.71
Star Early Literacy	643	840.26 (98.45)	475.75 (20.54)	0.72
Star Reading	4,560	918.00 (94.01)	491.61 (21.80)	0.76
Grade 3 EOY				
Acadience Reading	137	397.36 (112.12)	498.91 (18.79)	0.69
aimswebPlus	169	434.24 (57.43)	497.02 (21.48)	0.83
DIBELS 8th Edition	3,046	454.49 (39.64)	494.38 (23.36)	0.68
FastBridge aReading	105	517.58 (17.81)	502.22 (18.79)	0.78
i-Ready	2,555	524.93 (59.92)	492.41 (23.77)	0.84
Lexia RAPID	468	50.98 (38.42)	488.98 (21.87)	0.77
mCLASS	2,834	450.42 (31.96)	495.66 (23.77)	0.68
Star Early Literacy	411	894.85 (109.30)	476.19 (22.08)	0.73
Star Reading	4,584	968.10 (82.97)	491.38 (21.93)	0.79

Source: District-provided screening assessment data and state-provided MCAS data
Notes: SS: Scaled Score. *M* (*SD*): Mean (Standard Deviation).

Tables D.2–D.11 show student characteristics of the grade 3 EOY assessment concordance sample, overall and by assessment. Students can identify as nonbinary, but samples were not large enough to include in these tables.

Table D.2 Student Characteristics for the State of Massachusetts and Grade 3 EOY Literacy Screening Assessment Concordance Sample

Demographic	Number in state	Percent in state	Number in concordance sample	Percent in concordance sample
Low income	26,765	42.3%	6,356	44.4%
Non–low income	36,453	57.7%	7,953	55.6%
Female	31,016	49.1%	7,098	49.6%
Male	32,168	50.9%	7,209	50.4%
English learner	9,713	15.4%	2,793	19.5%
Non–English learner	53,505	84.6%	11,516	80.5%
Students receiving special education services	13,605	21.5%	2,974	20.8%
Students not receiving special education services	49,613	78.5%	11,335	79.2%
White	34,734	54.9%	7,793	54.5%
Hispanic/Latino	15,057	23.8%	3,952	27.6%
Black	5,502	8.7%	1,003	7.0%
Asian/Native American/ Pacific Islander	4,811	7.6%	896	6.3%
Total	**63,218**	**100%**	**14,309**	**100%**

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students can identify as nonbinary, but samples were not large enough to include in these tables. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Student demographic characteristics for the literacy screening assessment samples vary (see tables E.1.1 and E.1.2 in appendix E); therefore, the demographic differences seen in table D.2 do not necessarily represent the demographic characteristics for any specific literacy screening assessment sample. For example, in contrast with the overall concordance sample, low income students are underrepresented in the FastBridge, Acadience, Lexia Rapid, and i-Ready concordance samples, and are overrepresented in the Star Early Literacy sample (see tables E.1.1 and E.1.2).

Tables D.3 through D.11 show student characteristics of grade 3 EOY assessment concordance sample by early literacy screening assessment. Students can identify as nonbinary, but samples were not large enough to include in these tables.

Table D.3. Acadience Reading (*n* = 137) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	26	19.0%
Non–low income	111	81.0%
Female	76	55.5%
Male	61	45.5%
English learner	--	--
Non–English learner	--	--
Students receiving special education services	29	21.2%
Students not receiving special education services	108	78.8%
White	123	89.8%
Hispanic/Latino	--	--
Black	--	--
Asian/American Indian/Native Hawaiian	--	--

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.4. aimswebPlus (*n* = 169) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	49	29.0%
Non–low income	120	71.0%
Female	81	47.9%
Male	88	52.1%
English learner	--	--
Non–English learner	--	--
Students receiving special education services	36	21.3%
Students not receiving special education services	133	78.7%
White	143	84.6%
Hispanic/Latino	12	7.1%
Black	--	--
Asian/American Indian/Native Hawaiian	--	--

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.5. DIBELS 8th Edition (*n* = 3,046) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	1,247	40.9%
Non–low income	1,799	59.1%
Female	1,495	49.1%
Male	1,551	50.1%
English learner	411	13.5%
Non–English learner	2,635	86.5%

Demographic	Number	Percent in concordance sample
Students receiving special education services	625	20.5%
Students not receiving special education services	2,421	79.5%
White	1,822	59.8%
Hispanic/Latino	698	22.9%
Black	185	6.1%
Asian/American Indian/Native Hawaiian	197	6.5%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.6. FastBridge aReading (n = 105) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	28	26.7%
Non–low income	77	73.3%
Female	48	45.7%
Male	57	54.3%
English learner	--	--
Non–English learner	--	--
Students receiving special education services	20	19.0%
Students not receiving special education services	85	81.0%
White	81	77.7%
Hispanic/Latino	--	--
Black	--	--
Asian/American Indian/Native Hawaiian	--	--

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.7. i-Ready (*n* = 2,555) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	1,041	40.7%
Non–low income	1,514	59.3%
Female	1,268	49.6%
Male	1,286	50.3%
English learner	474	18.6%
Non–English learner	2,081	81.4%
Students receiving special education services	525	18.5%
Students not receiving special education services	2,309	81.5%
White	1,515	59.3%
Hispanic/Latino	690	27.0%
Black	123	4.8%
Asian/American Indian/Native Hawaiian	95	3.7%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.8. Lexia RAPID (*n* = 468) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	255	55.5%
Non–low income	213	45.5%
Female	220	47.0%
Male	248	53.0%
English learner	106	22.6%
Non–English learner	362	77.4%

Demographic	Number	Percent in concordance sample
Students receiving special education services	114	24.4%
Students not receiving special education services	354	75.6%
White	179	38.2%
Hispanic/Latino	244	52.1%
Black	22	4.7%
Asian/American Indian/Native Hawaiian	15	3.2%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.9. mCLASS (*n* = 2,834) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	1,054	37.2%
Non–low income	1,780	62.8%
Female	1,388	49.0%
Male	1,445	51.0%
English learner	430	15.2%
Non–English learner	2,404	84.8%
Students receiving special education services	525	18.5%
Students not receiving special education services	2,309	81.5%
White	1,406	49.6%
Hispanic/Latino	296	10.4%
Black	430	15.2%
Asian/American Indian/Native Hawaiian	416	14.7%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.10. Star Early Literacy (*n* = 411) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	261	63.5%
Non–low income	150	36.5%
Female	191	46.5%
Male	220	53.5%
English learner	185	45.0%
Non–English learner	226	55.0%
Students receiving special education services	162	39.4%
Students not receiving special education services	249	60.6%
White	159	38.7%
Hispanic/Latino	206	50.1%
Black	34	8.3%
Asian/American Indian/Native Hawaiian	--	--

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table D.11. Star Reading (*n* = 4,584) Student Characteristics of Grade 3 EOY Assessment Concordance Sample

Demographic	Number	Percent in concordance sample
Low income	2,395	52.2%
Non–low income	2,189	47.8%
Female	2,331	50.9%
Male	2,253	49.1%
English learner	1,179	25.7%
Non–English learner	3,405	74.3%

Demographic	Number	Percent in concordance sample
Students receiving special education services	887	19.3%
Students not receiving special education services	3,697	80.7%
White	2,365	51.6%
Hispanic/Latino	1,532	33.4%
Black	331	7.2%
Asian/American Indian/Native Hawaiian	164	3.6%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Discrepancies between the concordance samples and statewide population could affect the equipercentile linking results if they affect the test score distributions used to calculate linking estimates. As a further check on the reasonableness of the sample distributions, we examined the distribution of MCAS scores for the students taking each of the screening assessments and compared those test score distribution statistics with the known MCAS population statistics (table D.12).

Table D.12. 2022/23 MCAS Grade 3 ELA Scale Score Mean, Median, 5th, and 95th Percentiles for the State and Grade 3 Literacy Screening Assessment Concordance Samples

Early literacy screening assessment	Number	Mean	*SD*	5th percentile	Median	95th percentile	Minimum	Maximum
Acadience Reading	137	498.91	18.79	471	496	532	440	555
aimswebPlus	169	497.02	21.48	462	498	532	440	560
DIBELS 8th Edition	3,046	494.38	23.36	454	496	532	440	560
mCLASS	2,834	495.66	23.77	454	496	532	440	560
FastBridge aReading	105	502.22	18.79	471	504	526	449	545
i-Ready	2,555	492.41	23.77	454	494	532	440	560

Early literacy screening assessment	Number	Mean	*SD*	5th percentile	Median	95th percentile	Minimum	Maximum
Lexia RAPID	468	488.98	21.87	454	489	526	440	560
Star Early Literacy	411	476.19	22.08	440	473	515	440	538
Star Reading	4,584	491.38	21.93	454	492	526	440	560
State	63,218	495.03	23.56	454	496	532	440	560

Source: District-provided screening assessment data and state-provided MCAS data

As shown in table D.12, when compared to the state in the middle of the distribution

- mean and median MCAS ELA scores are similar for most samples,
- mean and median MCAS ELA scores are slightly higher (<8 points higher) for the aimswebPlus and FastBridge aReading samples (for the Acadience Reading sample, the mean MCAS ELA score is slightly higher, but the median score is the same),
- mean and median MCAS ELA scores are slightly lower (<10 points) for the i-Ready and Star Reading samples, and
- mean and median MCAS ELA scores are more than 10 points lower for the Star Early Literacy sample (which may reflect the fact that students often transition from taking Star Early Literacy to Star Reading while in grade 3, and students taking Star Early Literacy in grade 3 may differ from typical grade 3 MCAS test-takers).

When compared to the state at the tails of the distribution

- sample distributions appear to have good coverage at the extremes of the scale score distribution;
- the Acadience Reading, FastBridge aReading, and Star Early Literacy samples are missing test scores at the top of the scale score range; and
- the FastBridge aReading concordance sample is missing scores at the bottom of the scale score range.

Description of 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 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.
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 $\mathrm{P(a \leq X \leq b)}$. 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 (i.e., 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 4 and figure 1 in the main report.

Table D.13. Screening assessment Benchmarks Representing Significantly Below Benchmark Performance at BOY, MOY, and EOY With Corresponding National Percentiles

Early literacy screening assessment	Grade	Benchmark	BOY national percentile	MOY national percentile	EOY national percentile
aimswebPlus	3	High risk	13	15	13
DIBELS 8th Edition	3	Well Below Benchmark	31	34	28
mCLASS	3	Well Below Benchmark	31	34	28
FastBridge aReading	3	High risk	14	14	14
FastBridge CBMreading	3	High risk	14	14	14
i-Ready	3	At risk	29	38	58
Star Early Literacy	3	Intervention	24	24	24
Star Reading	3	Intervention	24	24	24

Description of Predictive Linking Procedure and Results

To further check on equipercentile linking estimates, we used ordinary least squares (OLS) regression estimates to link literacy screening assessment composite scores to the MCAS scale and predict student performance on MCAS ELA.

When interpreting the results in this section, it is important to note that predictive linking estimates are valid "to the extent that the test can predict the outcome" (Ho, 2012, p. 2). The accuracy of predictive linking estimates is directly related to the correlation between early literacy and MCAS ELA assessments. Table D.14 shows that correlations between literacy and MCAS ELA assessments are in the strong (0.60–0.79) to very strong (0.80–1.00) ranges, although the early literacy screening assessments and MCAS measure different aspects of

literacy development. Though not true in all cases, in general, correlations improve as the time between screening assessment administration and MCAS becomes smaller.

Table D.14. Correlations between grade 3 screening assessment benchmarks and MCAS at BOY, MOY, and EOY

Early literacy screening assessment	BOY correlation	MOY correlation	EOY correlation
Acadience Reading	--	0.68 (*n* = 137)	0.69 (*n* = 137)
aimswebPlus	--	0.82 (*n* = 170)	0.83 (*n* = 169)
DIBELS 8th Edition	0.70 (*n* = 1,285)	0.71 (*n* = 2,753)	0.68 (*n* = 3,046)
mCLASS	0.72 (*n* = 442)	0.71 (*n* = 2,432)	0.68 (*n* = 2,834)
FastBridge aReading	--	0.70 (*n* = 104)	0.78 (*n* = 105)
i-Ready	0.80 (*n* = 676)	0.84 (*n* = 2,561)	0.84 (*n* = 2,555)
Lexia RAPID	0.71 (*n* = 500)	0.71 (*n* = 472)	0.77 (*n* = 468)
Star Early Literacy	0.70 (*n* = 523)	0.72 (*n* = 643)	0.73 (*n* = 411)
Star Reading	0.71 (*n* = 210)	0.76 (*n* = 4,560)	0.79 (*n* = 4,584)

Source: District-provided screening assessment data and state-provided MCAS data
Note: Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Although these correlations are in the strong to very strong ranges, they are not perfect. In the absence of perfect correlation, OLS regression will predict an outcome score that approaches the outcome score mean as the correlation approaches zero. In other words, the weaker the correlation between early literacy and MCAS ELA assessments, the more biased the prediction will be toward the sample MCAS ELA mean. We created a predictive link from literacy screening assessments to the MCAS ELA scale (see Equation 3). Results are shown in table D.15.

Table D.15. Literacy Screening Assessment Grade 3 BOY and EOY Benchmark Cut Scores Linked to MCAS Grade 3 ELA Scale Scores and Performance Levels Using Predictive 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	487	485	Partially Meeting	Partially Meeting
Acadience Reading	220	331	At Benchmark	492	491	Partially Meeting	Partially Meeting
Acadience Reading	289	406	Above Benchmark	501	500	Meeting	Meeting
aimswebPlus	328	362	Moderate risk	479	475	Partially Meeting	Partially Meeting
aimswebPlus	354	386	Low risk	487	482	Partially Meeting	Partially Meeting
DIBELS 8th Edition	314	424	Below Benchmark	481	482	Partially Meeting	Partially Meeting
DIBELS 8th Edition	332	442	At Benchmark	489	489	Partially Meeting	Partially Meeting
DIBELS 8th Edition	365	467	Above Benchmark	503	499	Meeting	Partially Meeting
mCLASS	314	424	Below Benchmark	479	482	Partially Meeting	Partially Meeting
mCLASS	332	442	At Benchmark	488	491	Partially Meeting	Partially Meeting
mCLASS	365	467	Above Benchmark	504	504	Meeting	Meeting
FastBridge aReading	468	483	Some risk	486	483	Partially Meeting	Partially Meeting
FastBridge aReading	490	503	Low risk	497	490	Partially Meeting	Partially Meeting
FastBridge aReading	505	517	College Pathway	504	502	Meeting	Meeting
i-Ready	474	545	Some risk	486	499	Partially Meeting	Partially Meeting

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
i-Ready	511	561	No observed risk	498	507	Partially Meeting	Meeting
Lexia RAPID	30	30	Moderate likelihood of success	487	480	Partially Meeting	Partially Meeting
Lexia RAPID	70	70	High likelihood of success	506	497	Meeting	Partially Meeting
Star Early Literacy	865	901	Intervention	480	477	Partially Meeting	Partially Meeting
Star Early Literacy	909	943	On Watch	486	483	Partially Meeting	Partially Meeting
Star Early Literacy	939	970	At or Above Benchmark	491	487	Partially Meeting	Partially Meeting
Star Reading	865	901	Intervention	482	477	Partially Meeting	Partially Meeting
Star Reading	909	943	On Watch	490	486	Partially Meeting	Partially Meeting
Star Reading	939	970	At or Above Benchmark	495	492	Partially Meeting	Partially Meeting

Source: District-provided screening assessment data and state-provided MCAS data

Comparing tables 4 and D.15 shows that predictive and equipercentile linking estimates are similar and typically predict scores within the same MCAS ELA performance level. When differences occur, lack of perfect correlation between the assessments is a contributing factor. Imperfect correlation between the assessments contributes to the differences seen toward the extremes of the scale by pulling those estimates closer to the sample MCAS ELA mean. Given the effects of imperfect correlation on those estimates, predictive linking estimates near the middle of the scale can be viewed with more confidence than those nearer the extremes, such as those identifying the lowest-performing students. Where differences occur at the extremes of the scale, the equipercentile linking estimate is likely to be more accurate than the predictive linking estimate given a representative score distribution.

The OLS regression model in Equation 3 was used to create the predictive link from literacy screening assessments to the MCAS ELA scale.

$$MCAS\ ELA\ SS_i = \beta_0 + \beta_1 Composite\ SS_i + \varepsilon_i \qquad (3)$$

For student i, β_0 and β_1 are intercept and slope parameters, respectively, and ε_i is a prediction error term. The β_0 and β_1 parameters for each screening assessment were computed, and the scales were linked at each benchmark cut score by rounding the MCAS ELA scale score associated with the corresponding early literacy composite scale score of interest.

Description of Logistic Regression Procedure

Logistic regression predicts the probability p of scoring at or above the criterion score on the outcome test by modeling the intercept and slope of the log-odds of success as a function of a predictor, as shown in Equation 4,

$$ln\left(\frac{P(MCAS\ ELA\ SS \geq 500)}{1-P(MCAS\ ELA\ SS \geq 500)}\right) = \beta_0 + \beta_1 Composite\ SS \qquad (4),$$

where β_0 and β_1 are intercept and slope parameters, respectively.

Table D.16. Accuracy With Which Grade 3 BOY and EOY Literacy Screening Assessments OLS Regression Predictions Discriminate Between Students Who Meet and Do Not Meet Proficiency Standards on MCAS Grade 3 ELA

Early literacy screening assessment	Classification accuracy (percent)		TP		FN		TN		FP	
	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY
Acadience Reading	77	78	0.72	0.81	0.28	0.19	0.81	0.76	0.19	0.24
aimswebPlus	81	82	0.82	0.80	0.18	0.20	0.81	0.82	0.19	0.18
DIBELS 8th Edition	75	74	0.76	0.65	0.24	0.35	0.75	0.81	0.25	0.19
mCLASS	77	75	0.72	0.80	0.28	0.20	0.81	0.72	0.19	0.28
FastBridge aReading	80	83	0.81	0.85	0.19	0.15	0.76	0.80	0.24	0.20

Early literacy screening assessment	Classification accuracy (percent)		TP		FN		TN		FP	
	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY
i-Ready	84	86	0.82	0.81	0.18	0.19	0.85	0.89	0.15	0.11
Lexia RAPID	83	83	0.77	0.83	0.23	0.17	0.85	0.83	0.15	0.17
Star Early Literacy	89	88	0.90	0.50	0.10	0.50	0.89	0.97	0.11	0.03
Star Reading	83	85	0.80	0.75	0.20	0.25	0.84	0.90	0.16	0.10

Source: District-provided screening assessment data and state-provided MCAS data
Notes: TP = True Positive (students at or above benchmark who met or exceeded expectations), FN = False Negative (students below benchmark who met or exceeded expectations), TN = True Negative (students below benchmark who did not meet expectations), FP = False Positive (students at or above benchmark who did not meet expectations)

Table D.17. ROC Analysis of Accuracy With Which Grade 3 BOY and EOY Literacy Screening Assessments Discriminate Between Students Who Meet and Do Not Meet Proficiency Standards on MCAS Grade 3 ELA by Assessment and Cut Score

Early literacy screening assessment	Benchmark	Classification accuracy (percent)		TP		FN		TN		FP		AUC
		BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	
Acadience Reading	Below Benchmark	55	53	0.97	1.00	0.03	0.00	0.24	0.18	0.76	0.82	0.88
Acadience Reading	At Benchmark	61	64	0.95	0.98	0.05	0.02	0.37	0.39	0.63	0.61	0.88
Acadience Reading	Above Benchmark	77	78	0.72	0.83	0.28	0.17	0.81	0.75	0.19	0.25	0.88
aimswebPlus	Moderate risk	60	56	0.97	0.99	0.03	0.01	0.29	0.18	0.71	0.82	0.89
aimswebPlus	Low risk	72	63	0.96	0.97	0.04	0.03	0.52	0.32	0.48	0.68	0.89
DIBELS 8th Edition	Below Benchmark	64	64	0.97	0.98	0.03	0.02	0.4	0.37	0.6	0.65	0.83
DIBELS 8th Edition	At Benchmark	72	71	0.90	0.89	0.10	0.11	0.59	0.58	0.41	0.42	0.83
DIBELS 8th Edition	Above Benchmark	74	75	0.53	0.69	0.47	0.31	0.89	0.79	0.11	0.21	0.83
mCLASS	Below Benchmark	60	64	0.97	0.98	0.03	0.02	0.36	0.35	0.64	0.65	0.84
mCLASS	At Benchmark	71	72	0.91	0.91	0.09	0.09	0.57	0.56	0.43	0.44	0.84
mCLASS	Above Benchmark	77	76	0.64	0.75	0.36	0.25	0.86	0.77	0.14	0.23	0.84

Early literacy screening assessment	Benchmark	Classification accuracy (percent)		TP		FN		TN		FP		AUC
		BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	
FastBridge aReading	Some risk	69	68	0.98	1.00	0.02	0.00	0.21	0.15	0.79	0.85	0.90
FastBridge aReading	Low risk	79	73	0.92	0.95	0.08	0.05	0.56	0.37	0.44	0.63	0.90
FastBridge aReading	College Pathway	77	79	0.74	0.78	0.26	0.22	0.82	0.80	0.18	0.20	0.90
i-Ready	Some Risk	74	86	0.98	0.83	0.02	0.17	0.57	0.87	0.43	0.13	0.93
i-Ready	No Observed Risk	84	83	0.85	0.66	0.15	0.34	0.83	0.95	0.17	0.05	0.93
Lexia RAPID	Moderate likelihood of success	79	68	0.83	0.95	0.17	0.05	0.78	0.56	0.22	0.44	0.90
Lexia RAPID	High likelihood of success	83	81	0.54	0.85	0.46	0.15	0.96	0.79	0.04	0.21	0.90
Star Early Literacy	Intervention	73	70	0.99	0.95	0.01	0.05	0.69	0.64	0.31	0.36	0.93
Star Early Literacy	On Watch	85	81	91	0.89	0.09	0.11	0.84	0.79	0.16	0.21	0.93
Star Early Literacy	At or Above Benchmark	89	88	0.76	0.86	0.24	0.14	0.92	0.88	0.08	0.12	0.93
Star Reading	Intervention	61	55	0.99	1	0.01	0	0.38	0.3	0.62	0.7	0.92

Early literacy screening assessment	Benchmark	Classification accuracy (percent)		TP		FN		TN		FP		AUC
		BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	BOY	EOY	
Star Reading	On Watch	74	69	0.96	0.99	0.04	0.01	0.61	0.52	0.39	0.48	0.92
Star Reading	At or Above Benchmark	82	77	0.88	0.96	0.12	0.04	0.78	0.67	0.22	0.33	0.92

Source: District-provided screening assessment data and state-provided MCAS data
Notes: TP = True Positive (students at or above benchmark who met or exceeded expectations), FN = False Negative (students below benchmark who met or exceeded expectations), TN = True Negative (students below benchmark who did not meet expectations), FP = False Positive (students at or above benchmark who did not meet expectations), AUC = Area under the ROC curve

Appendix E. Screening Assessment Overview and Student Performance

Tables E.1.1 and E.1.2 provide the demographic breakdowns for the early literacy screening assessment sample (for 2022/23) by screening assessment.

Table E.1.1. Comparison of Student Demographics of Early Literacy Screening Assessment Sample to the State (By Early Literacy Screening Assessment)

Grade Level/Demographic	Acadience Reading (percent)	aimswebPlus (percent)	DIBELS 8th Edition (percent)	mCLASS (percent)	EarlyBird (percent)	FastBridge aReading (percent)	State (percent)
Kindergarten	31%	24%	29%	25%	100%	0%	24%
Grade 1	30%	25%	30%	26%	--	19%	26%
Grade 2	30%	25%	22%	26%	0%	41%	25%
Grade 3	9%	26%	19%	23%	0%	40%	25%
Low income	17%	25%	39%	43%	32%	33%	44%
Female	49%	50%	49%	49%	51%	44%	49%
Male	51%	50%	51%	51%	48%	56%	51%

Grade Level/Demographic	Acadience Reading (percent)	aimswebPlus (percent)	DIBELS 8th Edition (percent)	mCLASS (percent)	EarlyBird (percent)	FastBridge aReading (percent)	State (percent)
Nonbinary	--	--	--	--	--	--	<1%
English learner	3%	2%	16%	23%	12%	5%	18%
Students receiving special education services	17%	17%	17%	16%	14%	20%	16%
White	96%	95%	85%	69%	73%	88%	78%
Hispanic/Latino	4%	6%	23%	22%	13%	9%	25%
Black	4%	6%	10%	16%	15%	8%	16%
Asian	3%	3%	9%	20%	20%	7%	10%
American Indian/Alaskan Native	--	--	2%	1%	1%	--	3%
Native Hawaiian/Pacific Islander	--	0%	<1%	<1%	--	0%	<1%

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table E.1.2. Comparison of Student Demographics of Early Literacy Screening Assessment Sample to the state (By Early Literacy Screening Assessment)—Continued

Grade Level/Demographic	FastBridge CBMreading	FastBridge earlyReading	i-Ready	Lexia RAPID	Star Early Literacy	Star Reading	State
Kindergarten	0%	49%	20%	23%	34%	<1%	24%
Grade 1	56%	51%	29%	25%	38%	17%	26%
Grade 2	23%	0%	30%	25%	21%	39%	25%
Grade 3	21%	0%	22%	27%	7%	43%	25%

Grade Level/Demographic	FastBridge CBMreading	FastBridge earlyReading	i-Ready	Lexia RAPID	Star Early Literacy	Star Reading	State
Low income	30%	34%	38%	56%	61%	50%	44%
Female	44%	48%	49%	48%	48%	50%	49%
Male	56%	52%	51%	52%	52%	50%	51%
Nonbinary	--	--	--	--	--	--	<1%
English learner	5%	4%	19%	27%	36%	24%	18%
Students receiving special education services	18%	16%	19%	21%	18%	18%	16%
White	91%	93%	86%	87%	70%	77%	78%
Hispanic/Latino	8%	8%	26%	52%	43%	31%	25%
Black	12%	10%	12%	11%	18%	15%	16%
Asian	6%	5%	9%	5%	5%	5%	10%
American Indian/Alaskan Native	3%	2%	1%	<1%	16%	9%	3%
Native Hawaiian/Pacific Islander	--	--	<1%	--	<1%	<1%	<1%

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Table E.2 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 Massachusetts Comprehensive Assessment System (MCAS; for each screening assessment) to the state average.

Table E.2. Comparison of Percentage of Students Meeting/Exceeding Expectations on MCAS

Early literacy screening assessment	Sample (percent Meeting/Exceeding Expectations)	State (percent Meeting/Exceeding Expectations)
All Screening Assessments	39%	44%
DIBELS 8th Edition	42%	44%
mCLASS	46%	44%
FastBridge aReading	44%	44%
FastBridge CBMreading	44%	44%
i-Ready	62%	44%
Lexia RAPID	58%	44%
MAP Growth	41%	44%
Star Early Literacy	31%	44%
Star Reading	12%	44%

Source: District provided screening assessment data and state-provided MCAS data
Note: No grade 3 students took EarlyBird or FastBridge earlyReading.

Table E.3 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 E.3. Number of Observations With Benchmarks by Time Period and Screening Assessment

Early literacy screening assessment	BOY	MOY	EOY	Total
Acadience Reading	1,557	1,529	1,529	4,615
aimswebPlus	647	653	651	1,951
DIBELS 8th Edition	14,903	15,693	16,342	46,938
mCLASS	11,354	11,680	13,097	36,131
EarlyBird	274	1,240	1,175	2,689
FastBridge aReading	236	259	261	756
FastBridge CBMreading	65	357	356	778
FastBridge earlyReading	393	383	392	1,168
i-Ready	11,251	11,337	12,196	34,784
Lexia RAPID	1,717	1,759	1,746	5,222
Star Early Literacy	11,978	11,638	10,986	34,602
Star Reading	8,957	9,879	10,652	29,488
Total	63,332	66,407	69,383	199,122

Source: District-provided screening assessment data

Table E.4 provides the number of benchmark scores by time period, grade level, and demographic characteristic.

Table E.4. Number of Benchmark Scores by Time Period, Grade Level, and Demographic Characteristics

Grade level or demographic	BOY	MOY	EOY	Total
Kindergarten	13,989	15,874	17,388	47,251
Grade 1	17,671	18,687	19,372	55,730
Grade 2	17,425	17,365	17,832	52,622
Grade 3	14,247	14,481	14,791	43,519
Low income	28,369	29,666	30,182	88,217
Female	31,098	32,584	34,074	97,756
Male	32,150	33,729	35,222	101,101
Nonbinary	25	27	24	76
English learner	13,525	14,191	14,919	42,635
Special education	11,183	11,521	11,871	34,575
White	50,678	52,998	54,835	158,511
Hispanic/Latino	17,496	18,203	18,707	54,406
Black	8,380	8,843	9,075	26,298
Asian	5,234	5,533	6,652	17,419
American Indian/Alaskan Native	3,248	3,485	3,581	10,314
Native Hawaiian/Pacific Islander	296	335	350	981

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). A cell may contain the same student more than once if they were delivered multiple screening assessments within the school year.

Table E.5 provides the demographic breakdown of students with one benchmark available versus students with two benchmarks available versus students with three benchmarks available.

Table E.5. Comparison of Demographics of Students With One Available Benchmark Versus Two Available Benchmarks Versus Three Available Benchmarks

Grade level or demographic	One benchmark available (percent)	Two benchmarks available (percent)	Three benchmarks available (percent)	State (percent)
Kindergarten	20%	42%	22%	24%
Grade 1	32%	25%	28%	26%
Grade 2	29%	19%	27%	25%
Grade 3	19%	14%	23%	25%
Low income	45%	51%	43%	44%
Female	48%	48%	49%	49%
Male	52%	52%	51%	51%
Nonbinary	--	--	<1%	<1%
English learner	30%	26%	20%	18%
Special education	18%	20%	17%	16%
White	68%	77%	80%	78%
Hispanic/Latino	31%	32%	27%	25%
Black	15%	16%	13%	16%
Asian	18%	9%	8%	10%
American Indian/Alaskan Native	7%	5%	5%	3%
Native Hawaiian/Pacific Islander	<1%	<1%	<1%	<1%

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student records can indicate more than one. Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).

Tables E.6 provides the percentage of students at significant risk by time period and for each assessment, using the benchmark and 25th percentile or below metrics.

Table E.6. Assessment Performance by Time Period and Assessment

Early literacy screening assessment	Benchmark metric—BOY (percent)	25th percentile metric—BOY (percent)	Benchmark metric—MOY (percent)	25th percentile metric—MOY (percent)	Benchmark metric—EOY (percent)	25th percentile metric—EOY (percent)	Benchmark metric—total (percent)	25th percentile metric—total (percent)
Acadience Reading	17%	N/A	16%	N/A	13%	N/A	15%	N/A
aimswebPlus	24%	33%	14%	24%	10%	24%	16%	27%
DIBELS 8th Edition	34%	20%	26%	17%	18%	18%	26%	18%
mCLASS	33%	20%	31%	21%	20%	20%	28%	20%
EarlyBird	45%	N/A	31%	N/A	17%	N/A	27%	N/A
FastBridge aReading	14%	25%	13%	25%	11%	17%	12%	22%
FastBridge CBMreading	14%	25%	14%	28%	13%	27%	14%	27%
FastBridge earlyReading	10%	24%	16%	27%	17%	34%	14%	28%
i-Ready	19%	26%	47%	27%	51%	26%	40%	26%
Lexia RAPID	67%	N/A	43%	N/A	32%	N/A	47%	N/A
Star Early Literacy	52%	53%	41%	42%	40%	41%	44%	45%
Star Reading	39%	40%	31%	31%	29%	30%	33%	34%
Total	36%	32%	34%	27%	29%	26%	33%	28%

Source: District-provided screening assessment data
Note: The Acadience Reading, EarlyBird, and Lexia RAPID data did not contain percentiles that could be used for the 25th percentile or below analysis.

Table E.7 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 E.7. Percent 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 (percent)	At significant risk at MOY (percent)	At significant risk at EOY (percent)	Change from BOY to EOY (percent)	Relative risk at BOY (percent)	Relative risk at MOY (percent)	Relative risk at EOY (percent)
Low income	46%	39%	38%	−8%	2.63*	2.78*	2.69*
Non–low income	17%	14%	14%	−3%	N/A	N/A	N/A
Female	29%	24%	23%	−5%	0.90*	0.91*	0.91*
Male	32%	26%	26%	−6%	N/A	N/A	N/A
Nonbinary	--	--	--	--	†	†	†
English learner	61%	54%	52%	−10%	2.82*	3.07*	2.96*
Non–English learner	22%	18%	17%	−4%	N/A	N/A	N/A
Students receiving special education services	54%	51%	50%	−4%	2.13*	2.53*	2.54*
Students not receiving special education services	25%	20%	19%	−6%	N/A	N/A	N/A
White	28%	23%	23%	−5%	0.75*	0.74*	0.71*
Hispanic/Latino	54%	46%	44%	−9%	2.49*	2.61*	2.54*
Black	37%	32%	32%	−5%	1.29*	1.34*	1.36*
Asian	15%	13%	13%	−2%	0.46*	0.49*	0.50*
American Indian/Alaskan Native	66%	55%	55%	−11%	2.37*	2.35*	2.43*
Native Hawaiian/ Pacific Islander	40%	35%	31%	−10%	†	†	†
Total	30%	25%	25%	−5%	N/A	N/A	N/A

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Risk ratio indicates the likelihood of a student group ever being classified as at risk compared to students not in that group (e.g., students from low income families were 2.69 times more likely than students who are not from low income families to be classified as at significant risk at BOY). Includes only students with three scores. Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--).
† Not computed because groups are 5 percent or less of the sample.
*** Indicates statistical significance at an alpha level less than .05.**

Table E.8 provides the percentage of students at significant risk two or more times (when using the benchmark level and 25th percentile or below metric), by assessment.

Table E.8. Benchmark Performance by Screening Assessment

Early literacy screening assessment	Significantly below benchmark two or more times (percent)	25th percentile or below two or more times (percent)
Acadience Reading	14%	N/A
aimswebPlus	14%	25%
DIBELS 8th Edition	24%	16%
mCLASS	27%	19%
EarlyBird	15%	N/A
FastBridge aReading	11%	20%
FastBridge CBMreading	10%	25%
FastBridge earlyReading	13%	26%
i-Ready	39%	24%
Lexia RAPID	42%	N/A
Star Early Literacy	41%	42%
Star Reading	30%	31%
Total	**31%**	**26%**

Source: District-provided screening assessment data
Note: The Acadience Reading, EarlyBird, and Lexia RAPID data did not contain percentiles that could be used for the 25th percentile or below analysis.

Tables E.9–E.13 show the percentages of students identified as significantly below benchmark multiple times across overlapping student groups.

Table E.9. Likelihood of Students Being Significantly Below Benchmark Multiple Times Controlling for Race/Ethnicity and Low Income Status

Demographic	Low income students significantly below benchmark multiple times (percent)		Low income students relative risk odds ratio		Non–low income students significantly below benchmark multiple times (percent)		Non–low income students relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
White	22%	22%	N/A	N/A	14%	14%	N/A	N/A
Hispanic/Latino	28%	28%	1.27	1.27	18%	18%	1.29	1.29
Black	25%	25%	1.14	1.14	16%	16%	1.14	1.14
Asian	12%	12%	0.55	0.55	7%	7%	0.50	0.50

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table E.10. Likelihood of Students Being Significantly Below Benchmark Multiple Times Controlling for Race/Ethnicity and English Learner Status

Demographic	English learner students significantly below benchmark multiple times (percent)		English learner students relative risk odds ratio		Non–English learner students significantly below benchmark multiple times (percent)		Non–English learner students relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
White	40%	36%	N/A	N/A	14%	14%	N/A	N/A
Hispanic/Latino	47%	43%	1.18	1.19	18%	18%	1.29	1.29

Demographic	English learner students significantly below benchmark multiple times (percent)		English learner students relative risk odds ratio		Non–English learner students significantly below benchmark multiple times (percent)		Non–English learner students relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
Black	44%	40%	1.10	1.11	16%	16%	1.14	1.14
Asian	24%	21%	0.60	0.58	7%	7%	0.50	0.50

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table E.11. Likelihood of English learner Students Being Significantly Below Benchmark Multiple Times Controlling for Race/Ethnicity and Low Income Status

Demographic	Low income English learner students significantly below benchmark multiple times (percent)		Low income English learner students relative risk odds ratio		Non–low income English learner students significantly below benchmark multiple times (percent)		Non–low income English learner students relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
White	53%	49%	N/A	N/A	40%	36%	N/A	N/A
Hispanic/Latino	61%	57%	1.15	1.16	47%	43%	1.18	1.19
Black	58%	54%	1.09	1.10	44%	40%	1.10	1.11
Asian	35%	31%	0.66	0.63	24%	21%	0.60	0.58

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table E.12. Likelihood of Students Being Significantly Below Benchmark Multiple Times Controlling for Race/Ethnicity and Special Education Status

Demographic	Students receiving special education services significantly below benchmark multiple times (percent)		Students receiving special education services relative risk odds ratio		Students not receiving special education services significantly below benchmark multiple times (percent)		Students not receiving special education services relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
White	56%	46%	N/A	N/A	14%	14%	N/A	N/A
Hispanic/Latino	56%	44%	1.00	0.96	18%	18%	1.29	1.29
Black	57%	48%	1.02	1.04	16%	16%	1.14	1.14
Asian	40%	26%	0.71	0.57	7%	7%	0.50	0.50

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table E.13. Likelihood of students Being Significantly Below Benchmark Multiple Times Controlling for Low Income and Special Education Status

Demographic	Low income students receiving special education services significantly below benchmark multiple times (percent)		Low income students receiving special education services relative risk odds ratio		Non–low income students receiving special education services significantly below benchmark multiple times (percent)		Non–low income students receiving special education services relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
White	68%	60%	N/A	N/A	56%	46%	N/A	N/A
Hispanic/Latino	68%	58%	1.00	0.97	56%	44%	1.00	0.96

Demographic	Low income students receiving special education services significantly below benchmark multiple times (percent)		Low income students receiving special education services relative risk odds ratio		Non–low income students receiving special education services significantly below benchmark multiple times (percent)		Non–low income students receiving special education services relative risk odds ratio	
	Female	Male	Female	Male	Female	Male	Female	Male
Black	70%	61%	1.03	1.02	57%	48%	1.02	1.04
Asian	54%	38%	0.79	0.63	40%	26%	0.71	0.57

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Description of Statistical Model for Intersectional Analysis

In the section examining the extent to which low income status and other factors explain variation in students' likelihood of being significantly below benchmark, 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 2022/23 school year. Use of such a model not only enables exploration of how student and school-level characteristics relate to literacy screening assessment performance, but also can identify schools that may outperform their peers after taking into account those variables.

Students are clustered within schools that may vary in their effects on student performance. To account for this fact, a multilevel logistic regression model, where students are nested within schools, was used to examine the effects of student- and school-level factors on the risk of being identified as significantly below benchmark. The model is shown below.

$$\text{logit}(\pi_{ij}) = \beta_{0j} + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \ldots + \beta_p X_{ijp}$$

This equation represents a two-level multilevel logistic regression model where

- $\text{logit}(\pi_{ij})$ is the natural logarithm of the odds of being significantly below benchmark multiple times for student *i* in school *j*,
- β_{0j} is the intercept, representing the log odds of being at risk within school j when all predictor variables are zero, and
- $\beta_1, \beta_2, \ldots, \beta_p$ are the fixed effect coefficients of predictor variables $X_{ij1}, X_{ij2}, \ldots, X_{ijp}$.

The model controls for student- and school-level predictor variables found to be associated with students' likelihood of being significantly below benchmark. 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 not to be statistically significant, as were variables examining the amount of student mobility, rates of student discipline, and teacher experience within schools. These variables were therefore removed from the final model. Student-level variables retained in the final model were

- gender,
- race/ethnicity,
- low income status,
- english learner status,
- whether the student received special education services, and
- early childhood (EC) program experience.

School-level variables were

- percentage of low income students,
- student attendance rate, and
- teacher retention rate.

The final model results are shown in Table E.14.

Table E.14. Results of Multilevel Logistic Regression Model Showing the Effects of Student- and School-Level Predictor Variables on Students' Risk of Being Significantly Below Benchmarks Multiple Times in 2022/23

Predictor	*B*	*SE*	*t*	*p*	OR	OR 95% CI— LL	OR 95% CI— UL
Intercept	−1.826	0.051	−36.05	<.001	0.161	0.146	0.178
Student-level predictors							
Asian	−0.745	0.118	−6.31	<.001	0.475	0.377	0.598
Black	0.187	0.086	2.19	0.029	1.206	1.020	1.426

Predictor	*B*	*SE*	*t*	*p*	OR	OR 95% CI— LL	OR 95% CI— UL
Hispanic/Latino	0.318	0.055	5.76	<.001	1.375	1.234	1.532
English learner × female	1.404	0.104	13.48	<.001	4.073	3.321	4.995
English learner × Male	1.243	0.104	11.99	<.001	3.465	2.828	4.245
Low income	0.548	0.028	19.82	<.001	1.729	1.638	1.825
EC program							
Formal	−0.111	0.031	−3.57	<.001	0.895	0.841	0.951
Informal	0.203	0.122	1.66	0.096	1.226	0.964	1.557
Formal × English learner × Asian	−0.266	0.171	−1.56	0.119	0.703	0.767	0.549
Informal × English learner × Asian	−0.376	0.491	−0.77	0.444	1.062	0.687	0.262
None × English learner × Asian	0.115	0.174	0.66	0.508	1.118	1.122	0.798
Formal × English learner × Black	−0.823	0.165	−4.98	<.001	0.429	0.439	0.318
Informal × English learner × Black	−1.042	0.604	−1.72	0.085	0.419	0.353	0.108
None × English learner × Black	−0.151	0.177	−0.85	0.393	0.820	0.860	0.608
Formal × English learner × Hispanic/Latino	−0.298	0.115	−2.61	0.009	0.678	0.742	0.593
Informal × English learner × Hispanic/Latino	−0.279	0.250	−1.12	0.264	0.926	0.757	0.464
None × English learner × Hispanic/Latino	0.051	0.108	0.47	0.637	1.022	1.052	0.851
Formal × English learner × White	−0.403	0.142	−2.84	0.005	0.668	0.506	0.883
Informal × English learner × White	−1.516	0.485	−3.12	0.002	0.220	0.085	0.568

Predictor	*β*	*SE*	*t*	*p*	OR	OR 95% CI— LL	OR 95% CI— UL
Special education services							
Yes × Asian × Female	2.165	0.265	8.16	<.001	8.710	5.180	14.646
Yes × Asian × Male	1.543	0.177	8.71	<.001	4.678	3.305	6.621
No × Asian × Female	−0.046	0.137	−0.34	0.735	0.955	0.730	1.248
Yes × Black × Female	1.926	0.188	10.25	<.001	6.864	4.749	9.921
Yes × Black × Male	1.548	0.138	11.23	<.001	4.703	3.590	6.163
No × Black × Female	0.137	0.107	1.29	0.198	1.147	0.931	1.415
Yes × Hispanic/Latino × Female	1.734	0.111	15.58	<.001	5.661	4.552	7.041
Yes × Hispanic/Latino × Male	1.265	0.074	17.11	<.001	3.544	3.066	4.097
No × Hispanic/Latino × Female	0.121	0.067	1.81	0.070	1.129	0.990	1.286
Yes × White × Female	2.047	0.059	34.66	<.001	7.747	6.900	8.697
Yes × White × Male	1.682	0.047	35.71	<.001	5.378	4.904	5.899
School-level predictors							
Low income percentage	0.005	0.002	2.51	0.012	1.005	1.001	1.008
Attendance rate	−0.071	0.028	−2.50	0.012	0.932	0.882	0.985
Teacher retention rate × Asian	−0.032	0.006	−5.50	<.001	0.969	0.958	0.980
Teacher retention rate × Black	−0.015	0.005	−3.13	0.002	0.985	0.976	0.994
Teacher retention rate × Hispanic/Latino	−0.015	0.003	−4.41	<.001	0.985	0.978	0.991
Teacher retention rate × White	−0.014	0.006	−2.57	0.010	0.986	0.975	0.997

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model. *β* = regression coefficient. *SE* = standard error. *t* = t statistic. *p* = p-value. CI–LL: confidence interval–lower limit. CI–UL: confidence interval–upper limit.

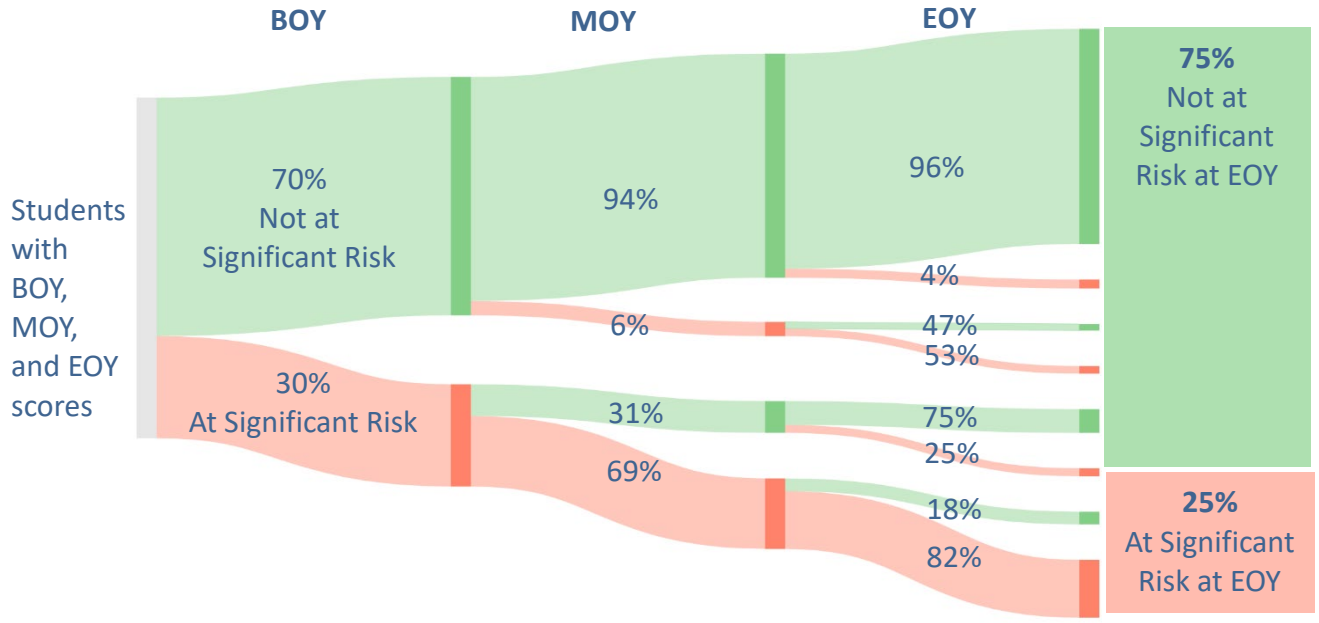
The reference group in the model is composed of white, male students who are not English learner students, not from low income families, do not receive special education services, did not attend an EC program, and who attended schools with the average percentage of low income students and average attendance and teacher retention rates. Such students had an 11 percent likelihood of being significantly below benchmark multiple times during the 2022/23 school year. Model results show several interactions among the predictor variables, particularly among gender, race/ethnicity, and English learner status, which means the way in which these variables affect a students' risk can depend on the groups and combinations of groups to which students belong. For example, schools with high teacher retention rates primarily benefit Asian students, and formal EC programs seem to provide the most benefit for English learner students who are Black.

Not only does this model allow for exploration of how student and school background characteristics relate to the risk of being identified as significantly below benchmark multiple times, but, because it includes separate effects for schools, it also enables identification of schools that are associated with the largest decreases in risk for the students compared to other schools after controlling for multiple student- and school-level variables. Schools identified in this model were also consistently identified as top performers in different models that were explored during the process of developing the final model. DESE will use this information as an opportunity to identify potentially useful strategies at those schools.

Appendix F. Student Progress

Figure F.1 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.

Figure F.1. Progression of students at 25th percentile or below across school year



Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period were included in the figure.

Tables F.1–F.24 provide the progression of students significantly below benchmark from one time period—beginning of year (BOY), middle of year (MOY), and end of year (EOY)—to the next, by student subgroup.

Table F.1. Significant Risk Performance Progression for Kindergarten Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	91% (*n* = 6,787)	87% (*n* = 6,966)
MOY	N/A	N/A	93% (*n* = 9,990)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	52% (n = 2,732)	37% (*n* = 1,916)
MOY	N/A	N/A	59% (*n* = 2,664)

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.2. Significant Risk Performance Progression for Grade 1 Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	80% (*n* = 9,311)	85% (*n* = 9,510)
MOY	N/A	N/A	94% (*n* = 10,520)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	71% (*n* = 3,744)	56% (*n* = 2,777)
MOY	N/A	N/A	67% (*n* = 4,184)

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.3. Significant Risk Performance Progression for Grade 2 Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	89% (*n* = 9,381)	88% (*n* = 9,093)
MOY	N/A	N/A	94% (*n* = 10,091)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	78% (*n* = 4,475)	70% (*n* = 3,764)
MOY	N/A	N/A	82% (*n* = 4,624)

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.4. Significant Risk Performance Progression for Grade 3 Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	93% (*n* = 8,313)	89% (*n* = 7,788)
MOY	N/A	N/A	90% (*n* = 8,424)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	81% (*n* = 3,721)	78% (*n* = 3,468)
MOY	N/A	N/A	85% (*n* = 3,741)

Source: District-provided screening assessment data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.5. Significant Risk Performance Progression for Female Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 16,930)	87% (*n* = 16,711)
MOY	N/A	N/A	93% (*n* = 19,465)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	70% (*n* = 7,016)	58% (*n* = 5,559)
MOY	N/A	N/A	71% (*n* = 7,137)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.6. Significant Risk Performance Progression for Male Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	87% (*n* = 16,825)	87% (*n* = 16,611)
MOY	N/A	N/A	93% (*n* = 19,510)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	70% (*n* = 7,637)	61% (*n* = 6,355)
MOY	N/A	N/A	74% (*n* = 8,056)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.7. Significant Risk Performance Progression for Low Income Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	84% (*n* = 10,639)	82% (*n* = 10,392)
MOY	N/A	N/A	90% (*n* = 13,086)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	75% (*n* = 10,101)	65% (*n* = 8,385)
MOY	N/A	N/A	77% (*n* = 10,085)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.8. Significant Risk Performance Progression for Non–Low Income Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	90% (*n* = 23,133)	89% (*n* = 22,946)
MOY	N/A	N/A	94% (*n* = 25,909)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	61% (*n* = 4,557)	49% (*n* = 3,532)
MOY	N/A	N/A	66% (*n* = 5,111)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.9. Significant Risk Performance Progression for English Learner Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	78% (*n* = 3,335)	75% (*n* = 3,249)
MOY	N/A	N/A	86% (*n* = 4,328)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	81% (*n* = 6,529)	72% (*n* = 5,558)
MOY	N/A	N/A	81% (*n* = 6,495)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.10. Significant Risk Performance Progression for Non–English Learner Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	89% (*n* = 30,437)	88% (*n* = 30,089)
MOY	N/A	N/A	94% (*n* = 34,667)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	63% (n = 8,129)	52% (*n* = 6,359)
MOY	N/A	N/A	68% (*n* = 8,701)

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.11. Significant Risk Performance Progression for Students Receiving Special Education Services

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	76% (*n* = 3,248)	75% (*n* = 3,195)
MOY	N/A	N/A	87% (*n* = 3,817)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	84% (*n* = 5,043)	76% (*n* = 4,403)
MOY	N/A	N/A	83% (*n* = 5,260)

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.12. Significant Risk Performance Progression for Students Not Receiving Special Education Services

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	89% (*n* = 30,524)	89% (*n* = 30,143)
MOY	N/A	N/A	94% (*n* = 35,178)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	65% (*n* = 9,615)	53% (*n* = 7,514)
MOY	N/A	N/A	69% (*n* = 9,936)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table.

Table F.13. Significant Risk Performance Progression for White Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 27,710)	87% (*n* = 27,440)
MOY	N/A	N/A	93% (*n* = 31,999)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	69% (*n* = 11,112)	58% (*n* = 8,902)
MOY	N/A	N/A	72% (*n* = 11,561)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.14. Significant Risk Performance Progression for Non-White Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 6,062)	87% (*n* = 5,898)
MOY	N/A	N/A	91% (*n* = 6,996)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	74% (*n* = 3,546)	65% (*n* = 3,015)
MOY	N/A	N/A	76% (*n* = 3,635)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.15. Significant Risk Performance Progression for Hispanic/Latino Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	82% (*n* = 5,477)	79% (*n* = 5,363)
MOY	N/A	N/A	88% (*n* = 6,970)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	77% (*n* = 7,213)	68% (*n* = 6,112)
MOY	N/A	N/A	79% (*n* = 7,144)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.16. Significant Risk Performance Progression for Non-Hispanic/Latino Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	89% (*n* = 28,295)	89% (*n* = 27,975)
MOY	N/A	N/A	94% (*n* = 32,025)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	64% (*n* = 7,445)	53% (*n* = 5,805)
MOY	N/A	N/A	68% (*n* = 8,052)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.17. Significant Risk Performance Progression for Black Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	85% (*n* = 3,761)	84% (*n* = 3,654)
MOY	N/A	N/A	90% (*n* = 4,431)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	73% (*n* = 2,378)	64% (*n* = 1,949)
MOY	N/A	N/A	75% (*n* = 2,508)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.18. Significant Risk Performance Progression for Non-Black Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 30,011)	87% (*n* = 29,684)
MOY	N/A	N/A	93% (*n* = 34,564)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	70% (*n* = 12,280)	59% (*n* = 9,968)
MOY	N/A	N/A	73% (*n* = 12,688)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.19. Significant Risk Performance Progression for Asian Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	92% (*n* = 3,642)	91% (*n* = 3,612)
MOY	N/A	N/A	95% (*n* = 3,997)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	66% (*n* = 625)	52% (*n* = 477)
MOY	N/A	N/A	62% (*n* = 680)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.20. Significant Risk Performance Progression for Non-Asian Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	87% (*n* = 30,130)	86% (*n* = 29,726)
MOY	N/A	N/A	93% (*n* = 34,998)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	70% (*n* = 14,033)	60% (*n* = 11,440)
MOY	N/A	N/A	74% (*n* = 14,516)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.21. Significant Risk Performance Progression for American Indian/Alaskan Native Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	84% (*n* = 838)	81% (*n* = 800)
MOY	N/A	N/A	83% (*n* = 1,138)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	76% (*n* = 1,549)	73% (*n* = 1,439)
MOY	N/A	N/A	83% (*n* = 1,530)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.22. Significant Risk Performance Progression for Non–American Indian/Alaskan Native Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 32,934)	87% (*n* = 32,538)
MOY	N/A	N/A	93% (*n* = 37,857)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	69% (*n* = 13,109)	58% (*n* = 10,478)
MOY	N/A	N/A	72% (*n* = 13,666)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Table F.23. Significant Risk Performance Progression for Native Hawaiian/Pacific Islander Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	84% (*n* = 123)	81% (*n* = 124)
MOY	N/A	N/A	91% (*n* = 161)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	73% (*n* = 99)	55% (*n* = 71)
MOY	N/A	N/A	71% (*n* = 100)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

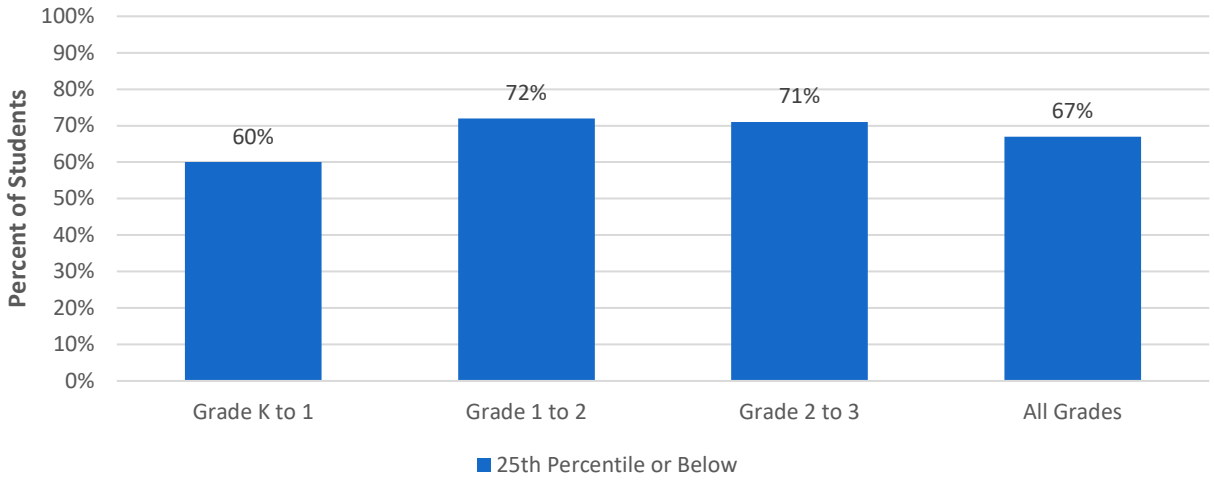
Table F.24. Significant Risk Performance Progression for Non–Native Hawaiian/Pacific Islander Students

If not significantly below benchmark	Then not significantly below at BOY	Then Not Significantly Below at MOY	Then not significantly below at EOY
BOY	N/A	88% (*n* = 33,649)	87% (*n* = 33,214)
MOY	N/A	N/A	93% (*n* = 38,834)
If significantly below benchmark	**Then significantly below at BOY**	**Then significantly below at MOY**	**Then significantly below at EOY**
BOY	N/A	70% (*n* = 14,559)	60% (*n* = 11,846)
MOY	N/A	N/A	73% (*n* = 15,096)

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times per time period if they were administered multiple screening assessments. Students with only one score during the school year were not included in the table. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Figure F.2 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.

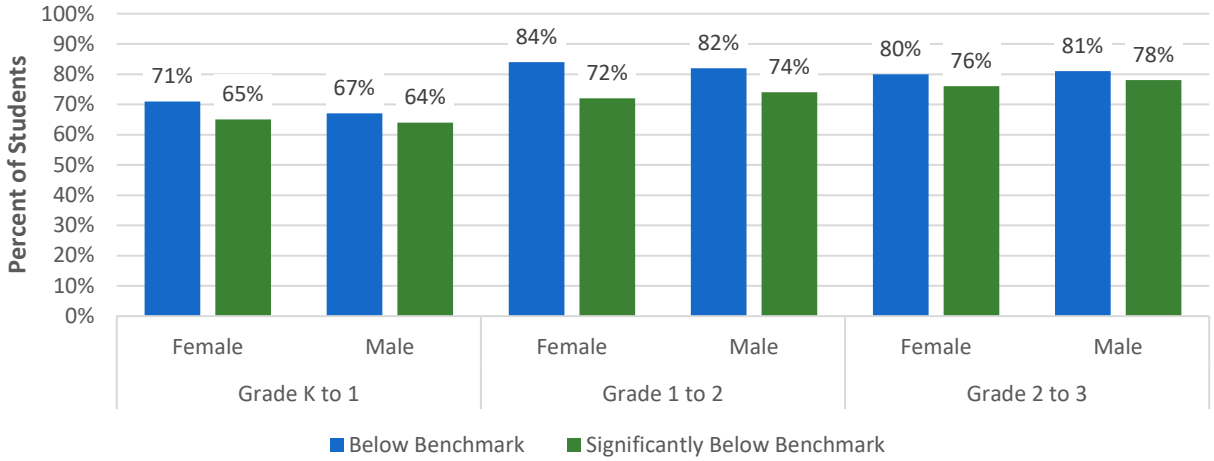
Figure F.2. Progression of Students at 25th Percentile or Below Across School Years



Source: District-provided screening assessment data
Note: Some students may appear multiple times if they were administered multiple screening assessments.

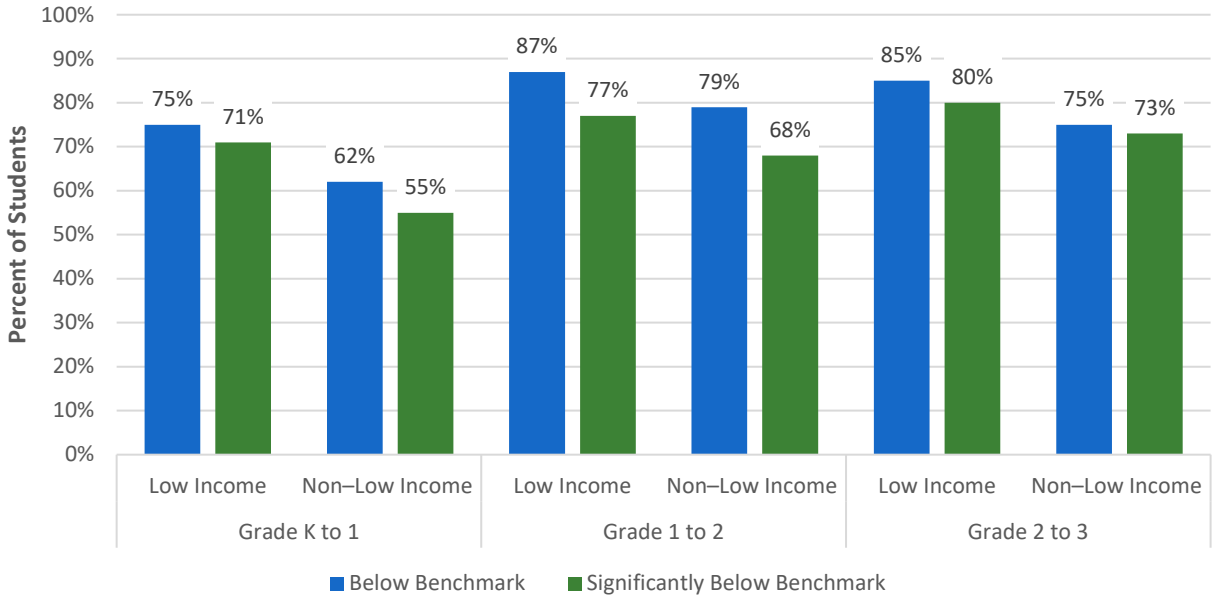
Figures F.3–F.11 show the percentage of students who were below or significantly below benchmark in one grade level and in the subsequent grade level, by student subgroup.

Figure F.3. Progression of Female and Male Students Below or Significantly Below Benchmark Across School Years



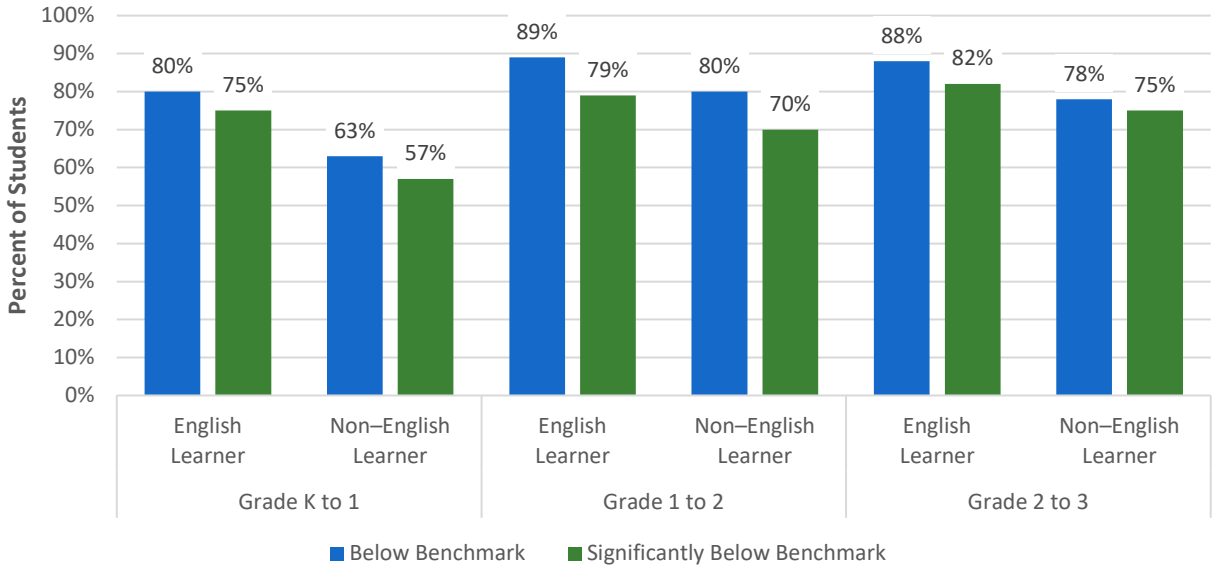
Source: District-provided screening assessment data and October and June SIMS collection data
Note: Some students may appear multiple times if they were administered multiple screening assessments.

Figure F.4. Progression of Low Income and Non–Low Income Students Below or Significantly Below Benchmark Across School Years



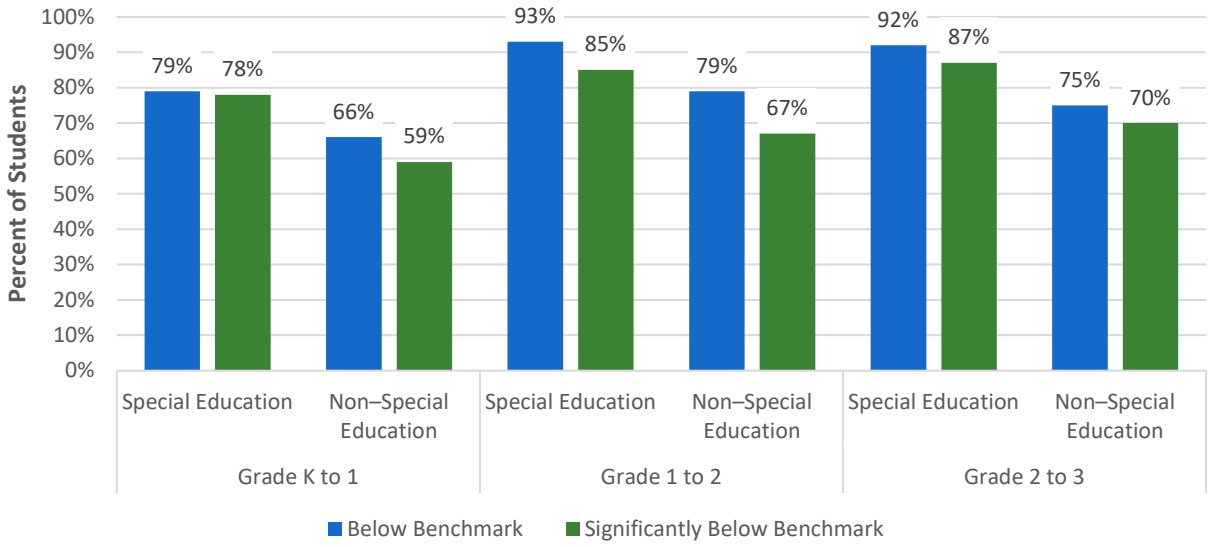
Source: District-provided screening assessment data and October and June SIMS collection data
Note: Some students may appear multiple times if they were administered multiple screening assessments.

Figure F.5. Progression of English Learner and Non–English Learner Students Below or Significantly Below Benchmark Across School Years



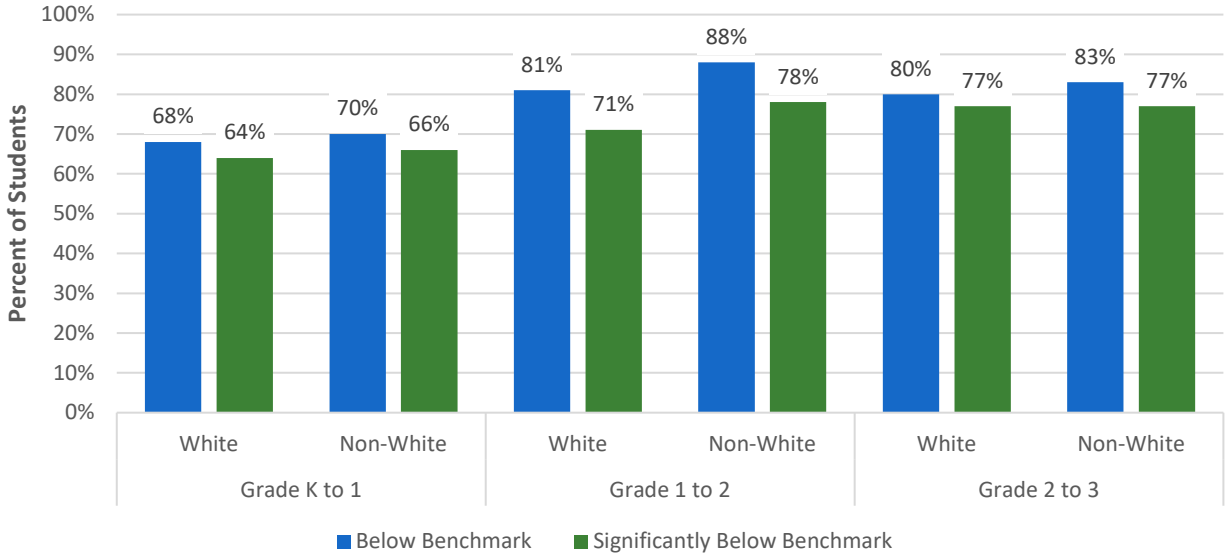
Source: District-provided screening assessment data and October and June SIMS collection data.
Note: Some students may appear multiple times if they were administered multiple screening assessments.

Figure F.6. Progression of Students Receiving and Not Receiving Special Education Services Below or Significantly Below Benchmark Across School Years



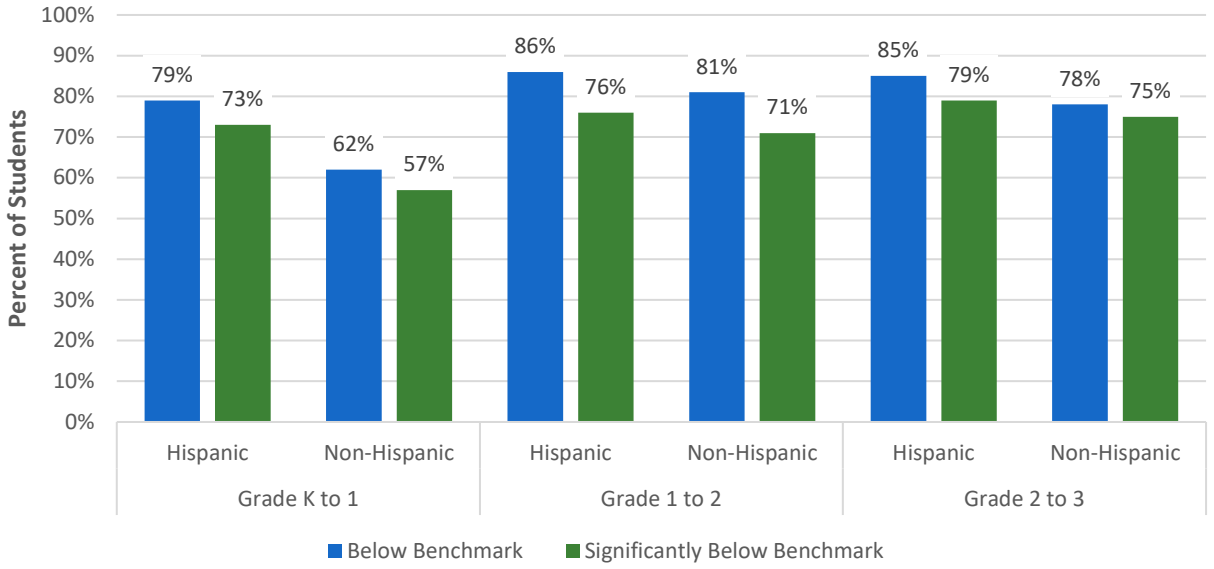
Source: District-provided screening assessment data and October and June SIMS collection data
Note: Some students may appear multiple times if they were administered multiple screening assessments.

Figure F.7. Progression of White and Non-White Students Below or Significantly Below Benchmark Across School Years



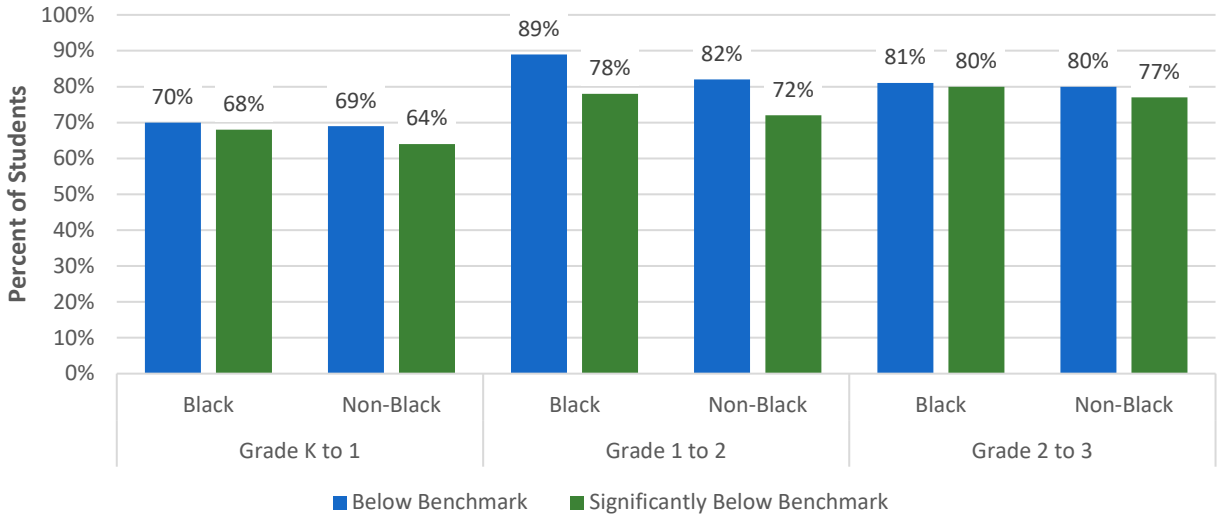
Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Figure F.8. Progression of Hispanic/Latino and Non-Hispanic/Latino Students Below or Significantly Below Benchmark Across School Years



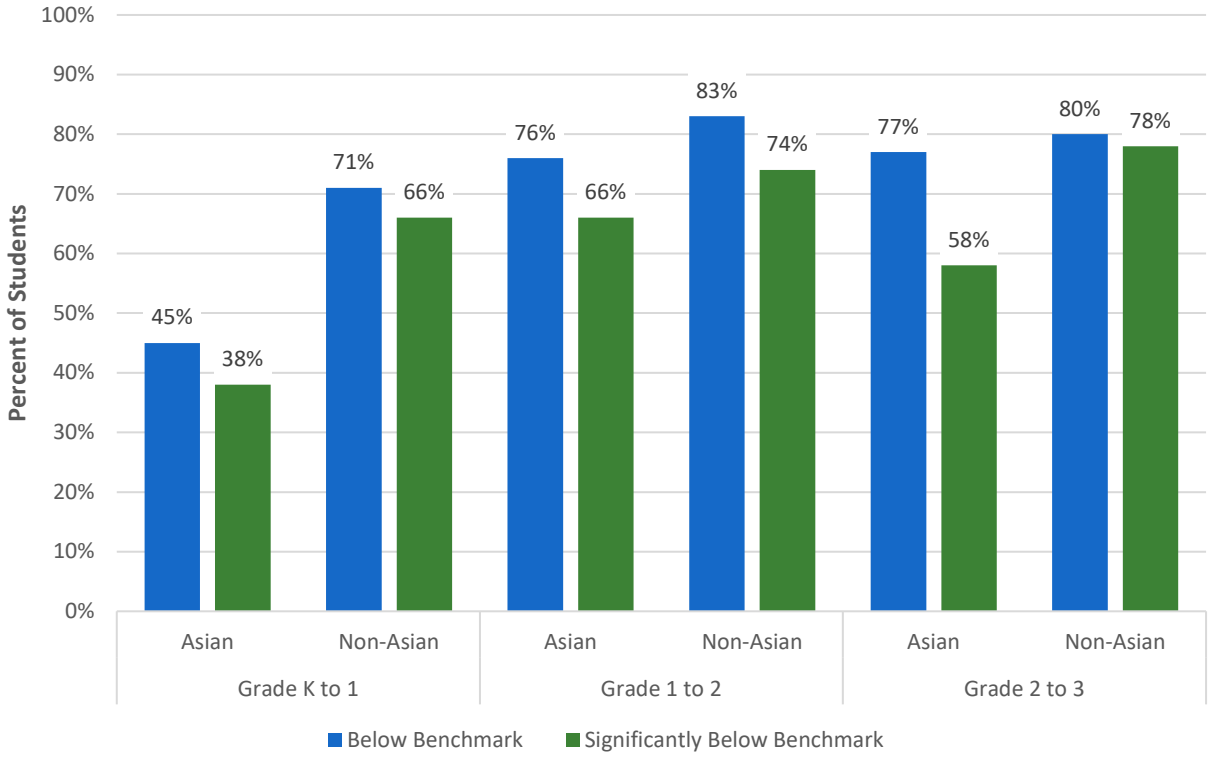
Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Figure F.9. Progression of Black and Non-Black Students Below or Significantly Below Benchmark Across School Years



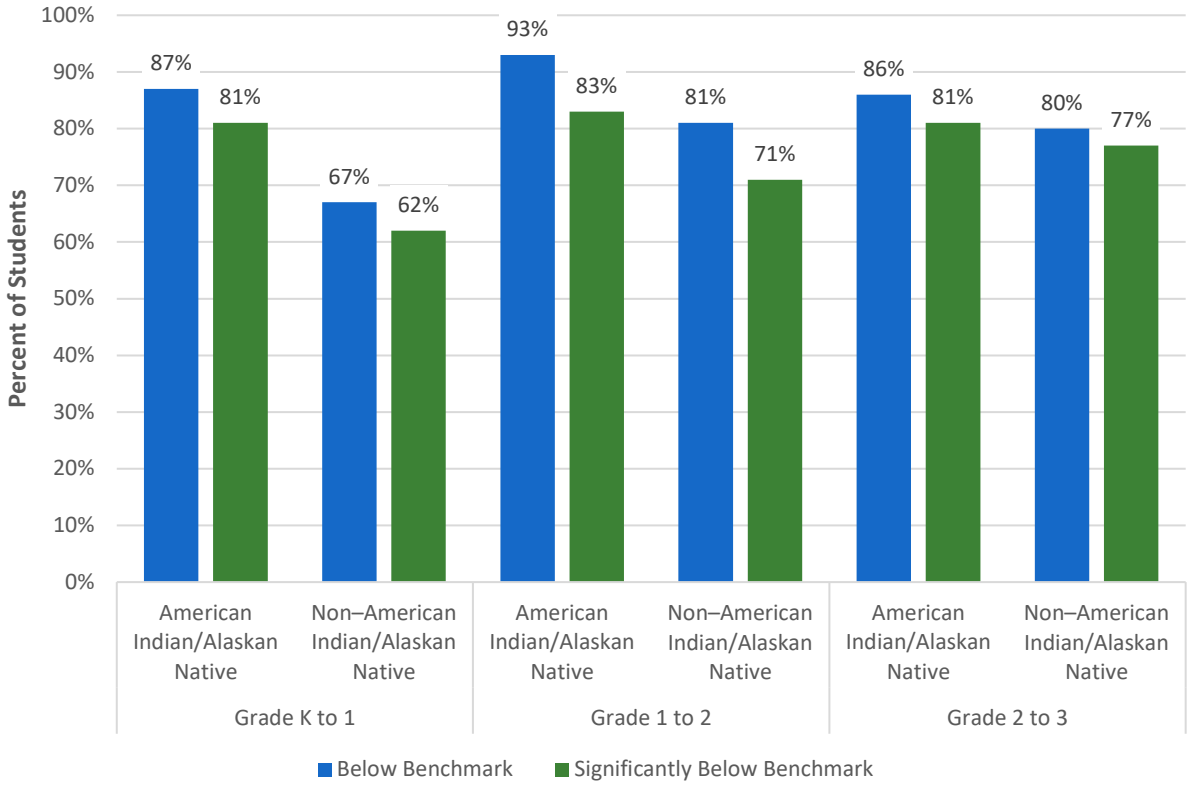
Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may appear multiple times if they were administered multiple screening assessments. Students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories).

Figure F.10. Progression of Asian and Non-Asian Students Below or Significantly Below Benchmark Across School Years



Source: District-provided screening assessment data and October and June SIMS collection data
Notes: 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).

Figure F.11. Progression of American Indian/Alaskan Native and Non–American Indian/Alaskan Native Students Below or Significantly Below Benchmark Across School Years



Source: District-provided screening assessment data and October and June SIMS collection data
Notes: 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).

Appendix G. Early Childhood and English Learner Program Performance

Table G.1 shows the percentage of students attending each early childhood (EC) program by demographic characteristic.

Table G.1. Percentage of Each Demographic Group Participating in Each Early Childhood Experience Program

Demographic	No EC experience	Informal: CFCE	Informal: PCHP	Informal: CFCE & PCHP	Formal: LFCP	Formal: CBP	Formal: LFCP and CBP
Low income	33%	2%	<1%	<1%	8%	46%	10%
Non–low income	14%	<1%	<1%	<1%	8%	69%	8%
Female	23%	1%	<1%	<1%	9%	57%	9%
Male	22%	1%	<1%	<1%	8%	60%	9%
English learner students	45%	2%	1%	<1%	8%	34%	10%
Non–English learner students	16%	1%	<1%	<1%	9%	66%	8%
Students receiving special education services	15%	1%	--	--	6%	71%	7%
Students not receiving special education services	24%	1%	<1%	<1%	9%	57%	9%
White	20%	1%	<1%	<1%	9%	61%	9%
Hispanic/Latino	38%	1%	<1%	<1%	8%	40%	11%
Black	23%	1%	--	1%	8%	57%	10%
Asian	21%	1%	--	1%	8%	62%	8%
American Indian/Alaskan Native	52%	2%	--	--	7%	31%	7%
Native Hawaiian/ Pacific Islander	30%	--	--	--	--	57%	--
Total	**23%**	**1%**	**<1%**	**<1%**	**9%**	**59%**	**9%**

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). CFCE: Coordinated Family and Community Engagement. PCHP: Parent-Child Home Program. LFCP: Licensed Family Child Care Provider. CBP: Center Based Program. Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--). Rows may not add up to 100 percent due to rounding.

Table G.2 shows the demographic breakdown of each EC program.

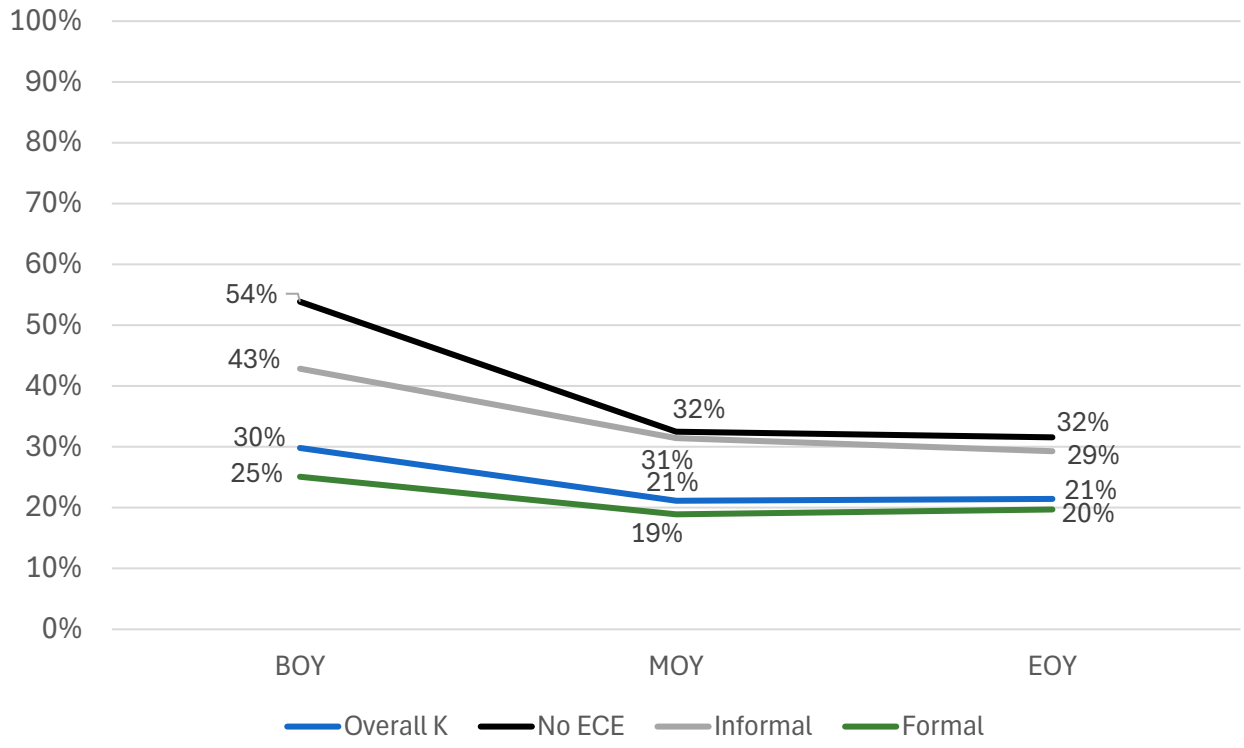
Table G.2. Demographic Breakdown of Early Childhood Experience Programs

Demographic	No EC experience	Informal: CFCE	Informal: PCHP	Informal: CFCE & PCHP	Formal: LFCP	Formal: CBP	Formal: LFCP and CBP
Low income	67%	74%	76%	60%	46%	36%	51%
Non–low income	33%	26%	24%	40%	54%	64%	49%
Female	51%	51%	40%	53%	51%	48%	50%
Male	49%	49%	60%	47%	49%	52%	50%
English learner students	46%	40%	45%	31%	20%	13%	28%
Non–English learner students	54%	60%	55%	69%	80%	87%	72%
Students receiving special education services	9%	12%	--	--	9%	17%	11%
Students not receiving special education services	91%	88%	--	--	91%	83%	89%
White	70%	69%	64%	55%	80%	81%	77%
Hispanic/Latino	49%	45%	40%	31%	28%	20%	38%
Black	14%	22%	--	25%	14%	14%	17%
Asian	9%	9%	--	18%	9%	11%	9%
American Indian/Alaskan Native	13%	10%	--	--	5%	3%	4%
Native Hawaiian/ Pacific Islander	1%	--	--	--	--	1%	--

Source: District-provided screening assessment data and October and June SIMS collection data
Notes: Some students may be included in more than one racial/ethnic category (Hispanic/not Hispanic and multiple race categories). Student groups with fewer than 10 students are not shown to protect student privacy and are represented by two dashes (--). The percentages for opposite demographic characteristics may not add up to 100 due to rounding. CFCE: Coordinated Family and Community Engagement. PCHP: Parent-Child Home Program. LFCP: Licensed Family Child Care Provider. CBP: Center Based Program.

Figure G.1 shows the percentage of kindergarten students at the 25th percentile or below by time period: beginning of year (BOY), middle of year (MOY), or end of year (EOY).

Figure G.1. Percent of Kindergarten Students at Significant Risk by Time Period (Using 25th Percentile or Below Metric)



Source: District-provided screening assessment data and October and June SIMS collection data
Note: Includes students with three scores only.

Table G.3 shows the percentage of kindergarten students at the 25th percentile or below multiple times and by time periods, by EC program.

Table G.3. Percent of Kindergarten Students at or Below the 25th Percentile at BOY, MOY, EOY by Type of Early Childhood Experience

EC experience	BOY (percent at 25th percentile or below)	MOY (percent at 25th percentile or below)	EOY (percent at 25th percentile or below)	Percent at 25th percentile or below multiple times
No EC program experience	53%	37%	36%	37%
Any EC program experience	25%	21%	22%	19%
Informal	43%	35%	30%	29%
Formal	25%	21%	22%	19%
All kindergarten students	30%	24%	24%	22%

Source: District-provided screening assessment data and October and June SIMS collection data
Note: Includes students with two or three scores.

Tables G.4–G.6 show the percentages of students identified as significantly below benchmark multiple times across overlapping student groups.

Table G.4. Likelihood of Students Being Significantly Below Benchmark Multiple Times Controlling for Race/Ethnicity and Early Childhood Experience

Demographic	Students with formal EC experience significantly below benchmark multiple times (%)		Students with formal EC experience relative risk odds (ratio)		Students without formal EC experience significantly below benchmark multiple times (%)		Students without formal EC experience relative risk odds (ratio)	
	Female	Male	Female	Male	Female	Male	Female	Male
White	13%	13%	N/A	N/A	14%	14%	N/A	N/A
Hispanic/Latino	17%	17%	1.15	1.15	18%	18%	1.29	1.29
Black	15%	15%	1.15	1.15	16%	16%	1.14	1.14
Asian	6%	6%	0.46	0.46	7%	7%	0.50	0.50

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table G.5. Likelihood of Low Income Students Being Significantly Below Benchmark Multiple Times Controlling for Early Childhood Experience

Demographic	Low income students with formal EC experience significantly below benchmark multiple times (%)		Low income students with formal EC experience relative risk Odds (ratio)		Low income students without formal EC experience significantly below benchmark multiple times (%)		Low income students without formal EC experience relative risk Odds (ratio)	
	Female	Male	Female	Male	Female	Male	Female	Male
White	20%	20%	N/A	N/A	22%	22%	N/A	N/A
Hispanic/Latino	26%	26%	1.30	1.30	28%	28%	1.27	1.27
Black	23%	23%	1.15	1.15	25%	25%	1.14	1.14
Asian	11%	11%	0.55	0.55	12%	12%	0.55	0.55

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.

Table G.6. Likelihood of English Learner Students Being Significantly Below Benchmark Multiple Times Controlling for Early Childhood Experience

Demographic	English learner students with formal EC experience significantly below benchmark multiple times (%)		English learner students with formal EC experience relative risk Odds (ratio)		English learner students without formal ECexperience significantly below benchmark multiple times (%)		English learner students without formal EC experience relative risk Odds (ratio)	
	Female	Male	Female	Male	Female	Male	Female	Male
White	28%	25%	N/A	N/A	40%	36%	N/A	N/A
Hispanic/Latino	37%	34%	1.32	1.36	47%	43%	1.18	1.19

Demographic	English learner students with formal EC experience significantly below benchmark multiple times (%)		English learner students with formal EC experience relative risk Odds (ratio)		English learner students without formal ECexperience significantly below benchmark multiple times (%)		English learner students without formal EC experience relative risk Odds (ratio)	
	Female	Male	Female	Male	Female	Male	Female	Male
Black	24%	21%	0.86	0.84	44%	40%	1.10	1.11
Asian	18%	15%	0.64	0.60	24%	21%	0.60	0.58

Source: District-provided screening assessment data and October and June SIMS collection data
Note: In this table, racial/ethnic groups are mutually exclusive (i.e., students can only be identified as belonging to a single group). This was used for the purposes of the statistical model.