WestEd
WestEd.org

A First Look at Early Literacy Performance in Massachusetts

Results of Initial Analysis Based on State Grantee Literacy Screening Assessments

Mariann Lemke, Dan Murphy, Aaron Soo Ping Chow, Hayley Spencer, and Angela Zhang

Fall 2023

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

Beginning with the 2020/21 school year, the Massachusetts Department of Elementary and Secondary Education (DESE) began collecting literacy screening assessment data from schools and districts participating in certain state grants. Screening assessment benchmark scores were available for just over 35,000 unique students across 2020/21 and 2021/22, representing about 10 percent of the state's population of grade K–3 students and including students from 43 districts and 159 schools.

The screening assessments in use in the state vary in how they were designed, how they are administered, and how they define risk of reading difficulty. Most, but not all, screening assessments also include several levels of risk. These differences, along with the fact that students take these assessments several times per year, mean that multiple methods are used to calculate numbers of students at risk of reading difficulty across assessments in the state and that the meaning of "risk" can differ somewhat across assessments. The screening assessment data include more economically disadvantaged and English learner students and fewer grade 3 students than the state as a whole, suggesting some caution may be needed in generalizing from results of these analyses to the state.

Nonetheless, these literacy screening assessment data provide an initial opportunity to examine student performance in the early grades and assess the extent to which intended policy outcomes in literacy are being achieved. This report provides a first look at the literacy performance of students in grades K–3 in Massachusetts based on available data from both the 2020/21 and 2021/22 school years.

Key findings include the following:

- Across all time periods combined (beginning, middle, and end of year), 52 percent of students were ever classified as below benchmark or showing any level of risk of reading difficulty. About 27 percent of students were at the highest levels of risk as defined by each assessment. Between 12 and 18 percent of students were classified as potentially at risk of dyslexia on the screening assessments that provide such a measure. In addition:
  - Fewer kindergarten and grade 3 students were identified as at any level of risk than were grade 1 and grade 2 students.
  - More economically disadvantaged students, English learner students, Hispanic students, Black students, and students receiving special education services were identified as at any level of risk than were their peers without those characteristics.

  - More students were identified as being at any level of risk in the beginning of year (BOY) assessment time period compared to other time periods.

- Most students identified as at risk of reading difficulty at BOY are still at risk at the end of the year (EOY); conversely, most students meeting screening assessment benchmark at BOY also met benchmark at EOY. However, 40 percent of students who were at risk at BOY were able to meet benchmark by EOY, and students who got on track by the middle of the year (MOY) were significantly more likely to meet benchmark at EOY than those who were classified at risk at MOY. Most students who were at risk in one grade level were also at risk in the next grade level, but there were substantial differences by grade. Many more kindergarten students got on track between kindergarten and grade 1 than did students moving from grade 1 to grade 2 or grade 2 to grade 3.
- Although English learner students were more likely to be identified as at risk than non–English learners overall, English learner students with low English proficiency as measured by ACCESS were not necessarily always below benchmark on screening assessments. A little over one third of English learner students who met benchmark scored at the two lowest English proficiency levels on ACCESS.
- DIBELS 8th Edition, i-Ready, mCLASS, and Star Early Literacy were the most commonly used literacy screening assessments among grantees reporting data to DESE.
- Linking grade 3 EOY literacy screening assessment benchmark cut scores to the grade 3 Massachusetts Comprehensive Assessment System (MCAS) English language arts (ELA) state assessment scale allows comparisons of the benchmarks. Results indicate that screening assessment benchmark cut scores that identify students at risk cluster in the MCAS Partially Meeting Expectations performance level, though there is variation among screening assessments.
- Literacy screening assessments can be useful predictors of MCAS ELA performance. However, the predictive value of literacy screening assessment benchmarks used to identify students who are on track to MCAS ELA proficiency varies. The percentage of students classified as on track who did not reach MCAS ELA proficiency standards ranged from 15 percent to 69 percent across screening assessments. Students who were classified by their grade 3 EOY literacy screening assessment performance to be at any level of risk of reading difficulties rarely met MCAS Grade 3 ELA proficiency standards.

The following sections provide more detail on these analyses.

Introduction

We know reading matters. Research has shown, time and again, that children who are not proficient early readers are at risk of poor social, educational, and economic outcomes (e.g., National Research Council, 1998). Many states, such as Massachusetts, have embarked on ambitious projects to identify and address potential reading difficulties and improve outcomes for students in the early grades.

Education Week reports that as of October 2022, 46 states had enacted reading- and/or dyslexia-related legislation, including provisions for student screening, teacher professional development, and other initiatives (Heubeck, 2023). In October 2018, the Massachusetts legislature enacted a new statute that added provisions to existing law requiring DESE to guide districts to develop screening procedures to identify students with potential indicators of "neurological learning disability including, but not limited to, dyslexia." Additionally, in September 2022, the Board of Elementary and Secondary Education in Massachusetts adopted an amendment to state regulations that requires all 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 beginning in July 2023 (see box). Many of these laws, including Massachusetts's law, include requirements for screening all students for risk of reading difficulties, often using commercial literacy assessment tools that states approve specifically for these purposes. The goal of such screening is to identify students who may be at risk of reading difficulty and proactively intervene with instructional supports that ensure that all students have the foundational skills needed to be successful readers.

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)

Beginning with the 2020/21 school year, DESE began collecting literacy screening assessment data from schools and districts participating in certain state grants. Although limited in scope, these literacy screening assessment data provide an important opportunity to examine student performance in the early grades and assess the extent to which intended policy outcomes are being achieved. On behalf of DESE, WestEd, a nonprofit research and development organization, has begun developing a longitudinal database of early literacy assessment scores and carrying out preliminary analyses. In collaboration with DESE and a group of stakeholders in the state, WestEd developed an initial set of research questions to be addressed with data from the 2020/21 and 2021/22 school years. (See table 1; appendix A provides the full set of initial research questions from which questions to be addressed using the first set of available data were drawn.)

Initial questions are aimed at providing summary information about assessments and student performance, including information about how benchmarks used to identify risk for different screening assessments compare. Some research questions are aimed at helping schools and DESE target efforts and resources based on patterns of performance—by knowing at what time periods or grade levels more students may be at risk, for example. The goal of analysis focused on English learners is to provide information about how well screening assessments work with English learners and how the performance of students learning English may be similar to or different from native English-speaking peers. Research questions will be reviewed and updated each year.

Table 1. Research questions

Topic	Research Question
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? • 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 by time period?
Student performance and progress	• How many students are identified as at any level of risk by time period, grade, and student group? • How many students are identified as at significant risk of reading difficulty by time period, grade, and student group? • How many students are identified as at potential risk for dyslexia? • How does student performance change as the school year progresses? Do students identified as at risk remain at risk? • How does student progress vary by risk level, grade, and student background characteristics? • How does student performance change across grade levels? Do students at risk remain at risk across years?
English learner student performance and English proficiency	• How many literacy screening assessment benchmark scores are available for English learner students? To what extent does the sample of English learner students represent the overall English learner student population in the state? • What is the relationship of English language proficiency to screening assessment performance and achieving benchmarks? • Do students at early stages of English proficiency achieve benchmarks? • Are students with higher levels of English proficiency classified as at risk? • How does performance and benchmark attainment vary for English learners by specific domains of English proficiency?
English learner student performance, native language, and program of instruction	• How does performance and benchmark attainment vary for English learners by native language and English language education program?
Comparing screening assessment benchmarks and predicting later outcomes	• How do literacy benchmark scores from different screening assessments compare to each other in Massachusetts Comprehensive Assessment System (MCAS) score terms? • What is the relationship between grade 3 literacy screening assessments and MCAS performance?

This report provides the results of the initial analysis of the approved early literacy universal screening assessment data collected from state grantees. 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.[1] 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 (MCAS) data, and
- Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) data.

Early Literacy Universal Screening Assessment Data

In the 2020/21 and 2021/22 school years, recipients of certain state grants were required to provide their students' early literacy universal screening assessment data to the state. These grants include:

- the Early Grades Literacy grant (EGL; FC734),
- the Early Literacy Screening Assessment and Professional Development grant (FC576), and
- the Growing Literacy Equity Across Massachusetts grant (GLEAM; FC509/510).[2]

[1] School- and district-level data can be found at https://profiles.doe.mass.edu/.

[2] Districts receiving the GLEAM grant provided screening assessment data for the 2021/22 school year, while districts receiving EGL and FC576 grants provided screening assessment data for the 2020/21 and 2021/22 school years.

Each of these literacy-related grants required participating schools and districts to prioritize the adoption and administration of valid and reliable early literacy screening assessments to inform instructional decision-making and planning in the early elementary grades.[3] In future years, data collection may also include other grants or department activities (e.g., schools and districts participating in Multi-Tiered Systems of Support Academies) or voluntary sharing of data by schools or districts.

As part of early efforts to encourage screening of students for potential reading difficulties, DESE approved 14 commercial early literacy screening assessments for use in the state, although their use was not required except for certain grantees. These screening assessments included the following:

- Acadience Reading
- aimswebPlus
- Amira Learning
- DIBELS 8th Edition
- EarlyBird
- easyCBM
- FastBridge assessments (aReading, AUTOreading, CBMreading, earlyReading)
- i-Ready
- ISIP ER
- iSTEEP
- Lexia RAPID
- MAP Growth
- mCLASS
- Star assessments (CBM, Early Literacy, Reading)

In 2022, the list of approved assessments was updated to better reflect recent Massachusetts Dyslexia Guidelines.[4] A state panel including researchers, teachers, administrators, and other specialists reviewed information from publishers about their assessments according to criteria including (1) constructs measured, (2) technical adequacy, (3) attention to linguistic diversity, and (4) administration usability and support. A summary of the 2022 universal screening assessment criteria and descriptions of approved assessments can be found in appendix B.[5]

[3] Additional information about DESE's grant funding opportunities can be found at https://www.doe.mass.edu/grants/default.html.

[4] The Massachusetts Dyslexia Guidelines can be viewed on DESE's Special Education page at https://www.doe.mass.edu/sped/links/dyslexia.html.

[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 at https://www.doe.mass.edu/instruction/screening-assessments.html.

Vendors may submit assessments for review using the current criteria until December 31, 2024, meaning the list of approved assessments may continue to evolve.

Currently, eight early literacy screening assessments are approved for use in elementary schools in Massachusetts, with three that were rated during the review process as "Meet Expectations" (DIBELS 8th Edition, EarlyBird, and mCLASS) and five that "Partially Meet Expectations" (Acadience Reading, the FastBridge suite of assessments, i-Ready, MAP Reading Fluency, and the Star suite of assessments). See appendix B for DESE's brief description of each of these approved assessments.

For this report, data from the following 11 screening assessments—including some from current and previously approved lists—were included in analyses:

- Acadience Reading (2020/21 only)
- DIBELS 8th Edition
- mCLASS
- EarlyBird (2021/22 only)
- FastBridge (aReading, CBMreading, earlyReading)
- i-Ready (Diagnostic and Literacy Tasks)
- ISIP ER (2021/22 only)
- Lexia RAPID
- MAP Growth (2021/22 only)
- MAP Reading Fluency (2021/22 only)
- Star (CBM, Early Literacy English and Spanish, Reading) (2021/22 only)

DESE collected the grades K–3 screening assessment data from individual schools and districts 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/BOY, winter/MOY, spring/EOY), test administration date, and test name, along with screening assessment data such as composite scores, benchmark levels, national percentile ranks, subtest scores, subtest benchmark levels, and dyslexia screening assessment scores 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 so that support services

could be provided to them. However, most of the approved screening 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 or below benchmark, at benchmark, above benchmark; see 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.[6] Many vendors 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.

Some screening assessments also include additional flags for students who may be at risk of severe reading difficulties, including dyslexia. These indicators are often based on overall performance, with additional criteria related to performance on specific subtests (e.g., rapid automatized naming tasks) or based on performance on specific subtests alone (see table 2).

For the analyses in this report, we use the generic terms "at risk" or "did not meet benchmark" to differentiate students with any level of risk of reading difficulty from those classified as having little or no risk using each screening assessment's definitions and benchmarks for risk. Both categories of students may include multiple performance or benchmark levels. We also include analyses that focus on students at the lowest performance or highest risk levels to describe students "at significant risk." 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. Scores or flags intended specifically to describe risk of severe reading difficulties or dyslexia are not included in the analysis describing students meeting or not meeting benchmark levels overall but are analyzed separately to describe students at potential risk of dyslexia. Note that DESE did not review or approve screening assessments specifically for dyslexia flagging.

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 can establish their own local benchmarks or could have calculated them manually. To ensure as much comparability as possible in the data, benchmark levels and reading risk flags were recalculated according to the technical documentation provided by the assessment vendors. 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 5% of records). The rules used to calculate benchmark

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

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 and additional reading risk or dyslexia flags for early literacy screening assessments included in 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. Levels include: • 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)*, **	Not available
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 vendor suggests that risk on the Letter Naming Fluency and Phonemic Segmentation Fluency subtests in kindergarten and grade 1 and Nonsense Word Fluency in grade 1 through grade 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.

Early Literacy Screening Assessment	Years with Data	Benchmark or Risk-Level Descriptions	Dyslexia and/or Additional Reading Risk Flag Descriptions****
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. 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. Indicator scores are as follows: • Low Risk • At Risk***

EarlyBird	2021/22 (available for kindergarten only)	EarlyBird provides a metric at each time period to identify students at risk of reading difficulties. In BOY, a flag identifies students performing "below expectations." 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 vendor, "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 only available 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 the end of the school year (provided the student doesn't receive appropriate remediation). According to the vendor, "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.). 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. 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 (WRCPM) 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 [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 • Low Risk: 40th–70th percentile • Some Risk: 15th–39th percentile*	Not available

Early Literacy Screening Assessment	Years with Data	Benchmark or Risk-Level Descriptions	Dyslexia and/or Additional Reading Risk Flag Descriptions****
		• 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. 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. Levels include: • 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 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 and typical growth measures to determine what 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****
ISIP 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 (above the 40th percentile) are on track and performing at grade level • Tier 2 (between the 21st and 40th percentile) are at some risk, are performing moderately below grade level, and are in need of intervention* • Tier 3 (20th percentile and below) are at risk, are performing seriously below grade level, and are in need of intensive intervention*, **	Not available
Lexia RAPID	2020/21 2021/22	Lexia RAPID reports three performance levels based on its Reading Success Probability score: "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." Levels include: • 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 vendor 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." 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, Exceeding) that are designed to describe which students are on or off track to meeting MCAS 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 that includes 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, which are established based on normative data. 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." 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 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) *, **	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." ***Triple asterisks 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. ****Note that DESE did not review or approve screening assessments specifically for dyslexia flagging.

Student-Level State Education Data (SIMS, MCAS, and ACCESS)

In addition to the K–3 early literacy screening assessment data, other student-level data were used for analysis, including Student Information Management System (SIMS) data, Massachusetts Comprehensive Assessment System (MCAS) data, and Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) data.[7] 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, MCAS, and ACCESS data collections. A brief description of each dataset follows.

SIMS

SIMS collects data pertaining to various student characteristics (e.g., gender, race/ethnicity, English learner status, immigrant status, native language, attendance) for the majority of students across the state. In the SIMS dataset, students can be identified using their unique statewide identifier (SASID) or their locally assigned student identifier, which is unique at the district level. SIMS data are submitted three times per school year (i.e., in October, March, and June) by districts across Massachusetts. For this analysis, data from the June collection was used to provide background characteristics for students in the screening assessment data unless it was missing. In those cases, October data was used. Only 270 observations (0.3 percent of cases and 134 students) were missing from both the October and June SIMS data, or state student IDs were not available in the screening assessment data to match to SIMS data.

For this analysis, only the variables that were relevant to grade 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 and Version 20.7 for the 2021/22 school year.[8]

MCAS

The MCAS data for the 2020/21 and 2021/22 school years provide student assessment scores in mathematics and English language arts (ELA) for the grade 3 students in the sample. This data provides a standardized measure of ELA achievement for most grade 3 students, allowing for an analysis of 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

[7] 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. See https://www.doe.mass.edu/research/researchers.html.

[8] The current SIMS Data Handbook can be viewed on DESE's SIMS page at https://www.doe.mass.edu/infoservices/data/sims/.

the United States are exempted from taking the MCAS ELA assessment, and these students are therefore also excluded from MCAS-related analysis.

ACCESS for ELLs

English learner students in grades K–12 in Massachusetts are tested annually using the ACCESS assessment to satisfy federal and state laws that require measuring the English proficiency of these students each year. ACCESS for ELLs is used to measure student proficiency in reading, writing, listening, and speaking typically in January and February of each school year.[9]

The ACCESS assessment provides 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 from 1 to 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 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 2020/21 and 2021/22 were retrieved from DESE's school and district profiles website to provide contextual data about the sample of students used in the analysis. Overall, the data pertain to the following four main categories of information: (1) educator characteristics (e.g., teacher race/ethnicity, teacher retention rate, percentage of experienced teachers), (2) student performance (e.g., MCAS scores, MCAS achievement levels), (3) student enrollment and demographic characteristics (e.g., grade-level

[9] Information regarding the ACCESS for ELLs assessment (and its associated performance levels and interpretations) can be viewed on DESE's MCAS page at https://www.doe.mass.edu/mcas/access/.

enrollment, race/ethnicity, gender, percentage of low-income or economically disadvantaged students, attrition rate, retention rate), and (4) 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 and 2021/22 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., fall, winter, and spring), 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 data suppression of student demographic and assessment data. DESE uses a minimum sample size of six students for reporting any student demographic information and a minimum sample size of 10 students for reporting student assessment outcomes. A dash (-) represents suppressed data in this report.

Analysis and Findings

As described in the introduction, this report is intended to provide a first look at the literacy performance of students in grades K–3 in Massachusetts based on available data. In this section, we describe available data and discuss how well the data represent students in grades K–3 in the Commonwealth as a whole. We also provide information about the early literacy screening assessments in use and how their benchmarks compare, as well as initial information about student performance and progress, with an additional look at English learner students. More specifically, in this section we describe the methods and results of analysis that address the research questions shown in table 1. The following sections provide more detail on these analyses.

Screening Assessment Data Overview and Student Performance

This section provides an overview of the literacy screening assessment data available for analysis and summarizes key trends in student early literacy attainment. Table 3 shows the specific research questions addressed in this section.

Table 3. Research questions about overall performance and progress

Topic	Research Question
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? • 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 by time period?
Student performance and progress	• How many students are identified as at any level of risk by time period, grade, and student group? • How many students are identified as at significant risk of reading difficulty by time period, grade, and student group? • How many students are identified as at potential risk for dyslexia? • How does student performance change as the school year progresses? Do students identified as at risk remain at risk? • How does student progress vary by risk level, grade, and student background characteristics? • How does student performance change across grade levels? Do students at risk remain at risk across years?

Screening Assessment Data Overview

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

Early literacy screening assessment data for the 2020/21 and 2021/22 school years combined include a total of 92,222 records with benchmark scores (i.e., levels that identify whether students are at risk of reading difficulty given their grade level and time of testing, as described in table 2).[10] This number includes all assessments and time periods, since most assessments are administered several times per year.[11] Eighty-nine percent of records were from the 2021/22 school year. Because there were so many fewer scores available from 2020/21, in the

[10] 11,117 records were removed from the data due to missing benchmark scores or other data issues.
[11] 2,499 students have multiple scores from different screening assessments in the same time period in the same school year.

analyses that follow we focus on all available data (i.e., data from 2020/21 and 2021/22 combined).

The most commonly used assessments were DIBELS 8th Edition (25%), i-Ready (22%), Star Early Literacy (17%), and mCLASS (14%) (see table 4). These four screening assessments together represent three quarters of all observations. DIBELS 8th Edition and mCLASS, both of which are based on the same assessment tasks, together made up approximately 40 percent of all scores. The remaining screening assessments each made up less than 5 percent of the observations across both years, except Lexia RAPID, which made up 7 percent.

As noted in table 2, EarlyBird and Star CBM were excluded from primary analyses of student risk. Star CBM only provides scores for each specific subtest/skill included in the assessment (e.g., letter naming); there is no overall description of student risk. Approximately 2,700 Star CBM observations across three districts in the 2021/22 school year were therefore excluded. EarlyBird uses separate metrics for each time period to identify students at risk of reading difficulties (see table 2). However, due to delays in obtaining details about benchmark levels and lack of any BOY data, approximately 1,400 MOY and EOY EarlyBird scores from one district in the 2021/22 school year were therefore excluded. However, EarlyBird is included in the discussion of students flagged as at risk of dyslexia based on its dyslexia risk indicator.

Table 4. Number of literacy screening assessment benchmark scores by year and assessment

Early Literacy Screening Assessment	2020/21 and 2021/22		2020/21		2021/22	
	Number	%	Number	%	Number	%
Acadience Reading	326	<1%	326	3%	0	0%
DIBELS 8th Edition	22,650	25%	964	9%	21,686	26%
mCLASS	13,258	14%	3,875	37%	9,383	11%
FastBridge aReading	1,342	1%	0	0%	1,342	2%
FastBridge CBMreading	1,604	2%	575	6%	1,029	1%
FastBridge earlyReading	875	<1%	457	4%	418	1%
i-Ready	20,226	22%	3,800	37%	16,426	20%
ISIP ER	1,902	2%	0	0%	1,902	2%
Lexia RAPID	6,575	7%	372	4%	6,203	8%
MAP Growth	1,908	2%	0	0%	1,908	2%

Early Literacy Screening Assessment	2020/21 and 2021/22		2020/21		2021/22	
	Number	%	Number	%	Number	%
MAP Reading Fluency	314	<1%	0	0%	314	<1%
Star Early Literacy	15,869	17%	0	0%	15,869	19%
Star Early Literacy Spanish	2,247	2%	0	0%	2,247	3%
Star Reading	3,126	3%	0	0%	3,126	4%
Total	**92,222**	**100%**	**10,369**	**100%**	**81,853**	**100%**

Source: 2020/21 and 2021/22 district-provided screening assessment data

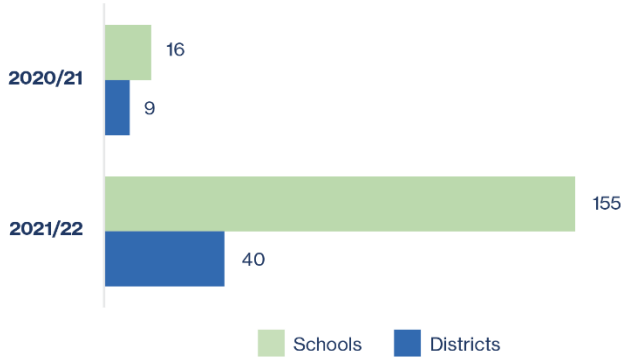
How many students, schools, and districts are represented by the data?

Most screening assessments are administered at least three times per school year (i.e., beginning of year [BOY], middle of year [MOY], and end of year [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 92,222 observations shown in table 4 represent 3,772 unique students across 16 schools in nine districts in 2020/21 and 31,598 unique students across 155 schools in 40 districts in 2021/22.[12] The median number of students per district was 519, with five districts comprising about one third of the data overall (see figure 1, which shows numbers of schools and districts in the sample in each year).

[12] 129 students (about 3% of the 2020/21 screening assessment data) had no benchmark scores associated with their screening assessment data during the school year, and 3,539 students (about 10% of the 2021/22 screening assessment data) had no benchmarks associated with their screening assessment data during the school year. These observations are excluded from the analyses in this report.

Figure 1. Between school years 2020/21 and 2021/22, numbers of schools and districts included in analysis increased



Source: 2020/21 and 2021/22 district-provided screening assessment data

These students represent about 1.5 percent of the total grade K–3 student population in the state in 2020/21, about 12.4 percent in 2021/22, and about 10 percent across years.[13] About 5 percent of students (1,779 students) across the sample have at least one benchmark score in 2020/21 and 2021/22.

Table 5 provides the number of unique students with data from each literacy screening assessment in either 2020/21 or 2021/22. 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 (i.e., Early Literacy English, Early Literacy Spanish, and Reading) or FastBridge assessments (i.e., aReading, CBMreading, and earlyReading).

Table 5. Number of students by year and assessment

Early Literacy Screening Assessment	2020/21 and 2021/22		2020/21		2021/22	
	Number	%	Number	%	Number	%
Acadience Reading	113	<1%	113	3%	0	0%
DIBELS 8th Edition	8,716	23%	354	9%	8,362	24%
mCLASS	4,813	13%	1,454	38%	3,359	10%
FastBridge aReading	629	2%	0	0%	629	2%

[13] Enrollment data for the 2020/21 and 2021/22 school years were retrieved from DESE's school and district profiles website at https://profiles.doe.mass.edu/statereport/enrollmentbygrade.aspx.

Early Literacy Screening Assessment	2020/21 and 2021/22		2020/21		2021/22	
	Number	%	Number	%	Number	%
FastBridge CBMreading	848	2%	229	6%	619	2%
FastBridge earlyReading	447	1%	162	4%	285	<1%
i-Ready	7,951	21%	1,324	34%	6,627	19%
ISIP ER	675	2%	0	0%	675	2%
Lexia RAPID	2,391	6%	220	6%	2,171	6%
MAP Growth	992	3%	0	0%	992	3%
MAP Reading Fluency	164	<1%	0	0%	164	<1%
Star Early Literacy	7,671	20%	0	0%	7,671	22%
Star Early Literacy Spanish	1,352	4%	0	0%	1,352	4%
Star Reading	1,738	5%	0	0%	1,738	5%
Total	**38,500**	**100%**	**3,856**	**100%**	**34,644**	**100%**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: 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.

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

Key Findings

- Demographic composition of the early literacy sample is similar to the state, suggesting that results are broadly generalizable.
- Sample includes a higher percentage of K–1 students and a lower percentage of grade 2–3 students than the state. Sample has slightly higher percentages of economically disadvantaged students, English learners, students receiving special education services, and Hispanic students.
- Background characteristics of students differ by screening assessment.

The early literacy screening assessment data have a greater percentage of K–1 students and a smaller percentage of grade 2–3 students than the state overall (see figure 2). About 46 percent of the screening assessment sample is classified as economically disadvantaged, 17 percent as receiving special education services, and 23 percent are classified as English learner students. Additionally, 78 percent are White, 27 percent are Hispanic, 13 percent are Black, 10 percent are Asian, 5 percent are American Indian or Alaskan Native, and less than 1 percent are Native Hawaiian or Pacific Islander.

Compared to the state as a whole, the sample includes more economically disadvantaged students and English learner students.[14] The racial/ethnic distribution of the sample is generally similar to that of the state, with slightly larger percentages of American Indian or Alaska Native and Hispanic students and slightly smaller percentages of Black students. Students receiving special education services are slightly overrepresented in the sample—17 percent compared to 15 percent in the state—as well.[15]

Appendix D 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 D.1.1 and D.1.2 in appendix D). For example, i-Ready has about half as many economically disadvantaged students as the state overall, while other screening assessments (e.g., mCLASS) include more. Appendix D also provides a comparison of the MCAS scores of grade 3 students across screening assessments, showing that grade 3 students in the sample have slightly smaller percentages of students meeting MCAS proficiency standards than the state average (see table D.2 in appendix D).

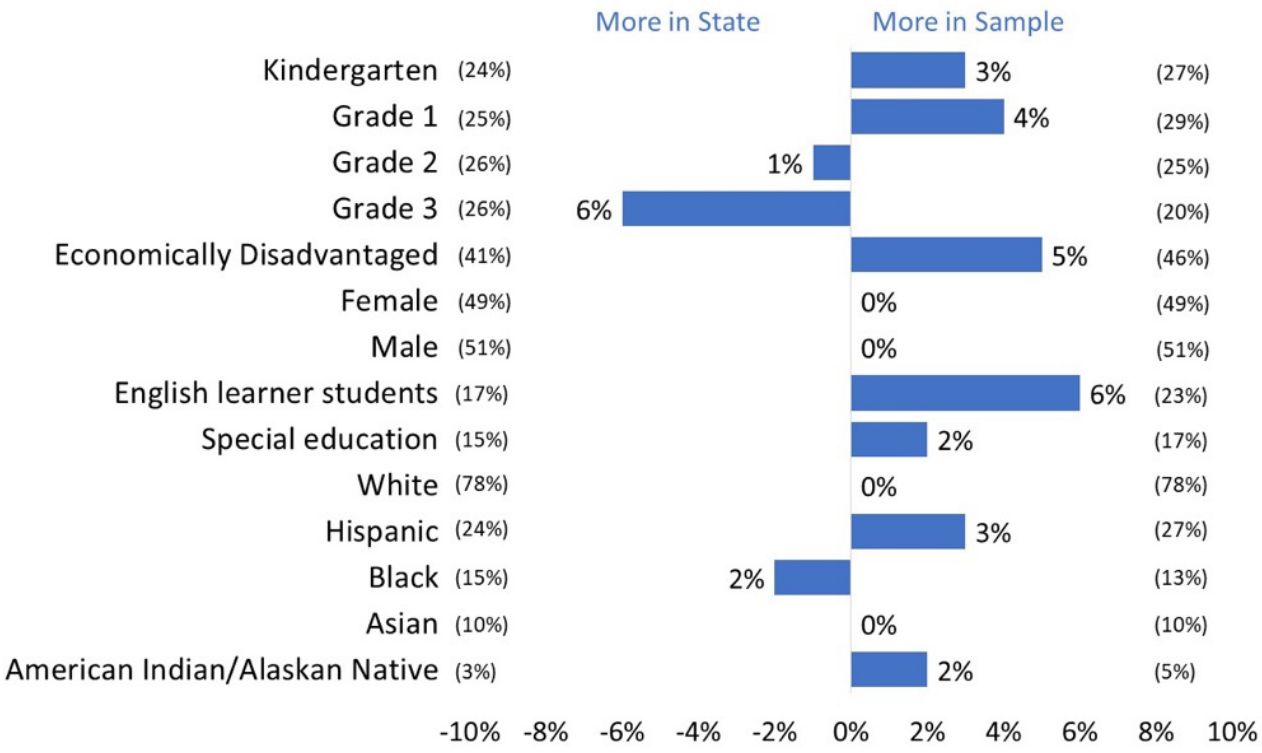
Overall, while some caution should be used, the relatively small magntitude of observed differences between students included in the literacy screening assessment data and the state

[14] Note that the state prioritizes grant funding for schools and districts with high-need students.

[15] 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.

(generally 5 percent or less) suggest that the results may be broadly generalizable to the state population of K–3 students.

Figure 2. Screening assessment data includes fewer grade 3 students, more economically disadvantaged students, and more English learner students than K–3 students in the state overall



Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: 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., 24 percent of students in the state are in kindergarten compared to 27 percent in the sample, or there are 3 percent more kindergarten students in the sample than in the state overall). Race and ethnicity are not exclusive. Student records can indicate more than one. The percentage of students classified as economically disadvantaged for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state developed a different measure (i.e., the "low-income" metric) for use in the 2021/22 school year, but it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy. The percentage of students classified as special education students include those classified as special education at the time of SIMS reporting and those who were previously classified during the current school year.

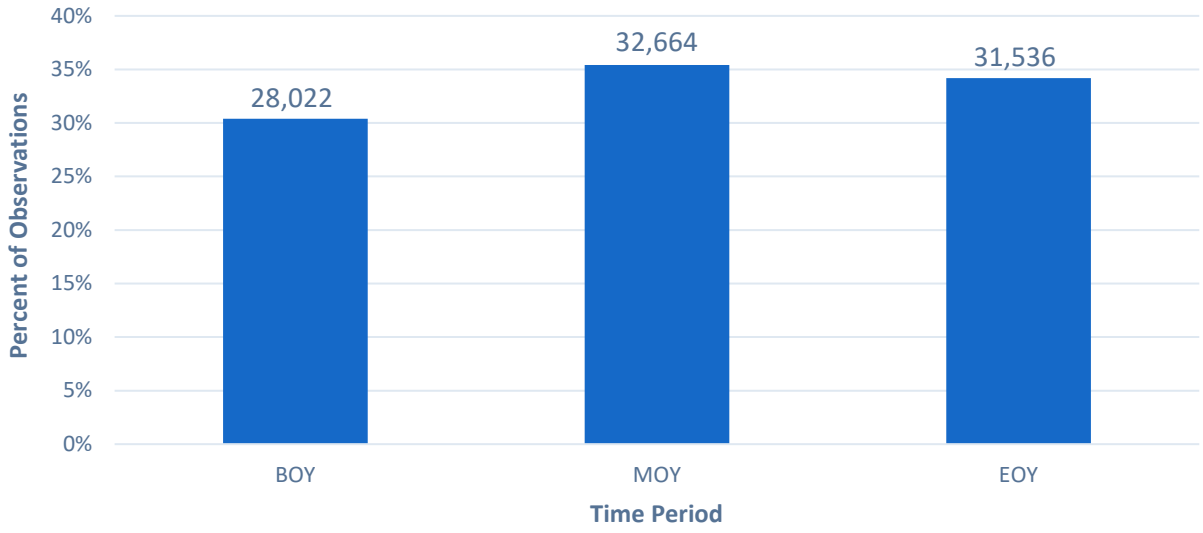
How many benchmark scores are available by time period?

Key Findings

- The number of records with benchmark scores differs by time period (35% in MOY, 34% in EOY, and 30% in BOY).
- The number of records in each time period differs by screening assessment.
- Only 54 percent of students have a benchmark score in all three time periods for the same assessment; 31 percent have a score in two of the three time periods; 15 percent have a score in a single time period.

Of the 92,222 records with benchmark scores in 2020/21 and 2021/22, about 35 percent were from the MOY time period, 34 percent from the EOY time period, and 30 percent from the BOY time period (figure 3). Some screening assessments had many more or fewer data points in a given time period (see table D.3 in appendix D). For example, the Star suite of assessments (i.e., Star Early Literacy, Early Literacy Spanish, and Reading) had significantly fewer BOY scores than MOY or EOY scores; MAP Growth had approximately equal numbers of BOY and MOY scores but double the number of EOY scores. MAP Reading Fluency has no EOY scores, and i-Ready and the FastBridge suite of assessments (i.e., FastBridge aReading, CBMreading, and earlyReading) had fewer EOY scores than BOY or MOY scores.

Figure 3. BOY time period has smallest number of scores, across screening assessments



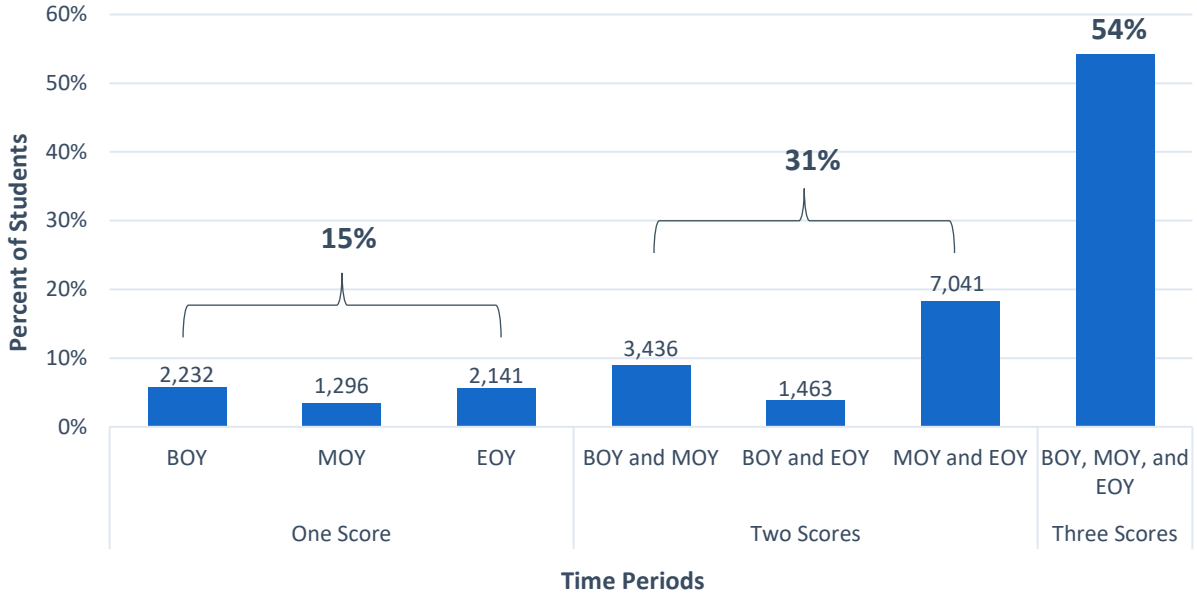
Source: 2020/21 and 2021/22 district-provided screening assessment data

Just over half of the K–3 students in the screening assessment data (54%) had benchmark scores from the same screening assessment in all three time periods (see figure 4). About 31

percent only had benchmark scores from two time periods, and 15 percent only had benchmarks from a single time period.

Of the students with two benchmark scores, most had MOY and EOY scores (see figure 4). Most students with only one benchmark had that score at the BOY or EOY.

Figure 4. Just over half of students have benchmark scores for all three time periods



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Students with multiple screening assessments per time period are counted for each screening assessment.

Students with all three scores included a higher percentage of White students and a lower percentage of economically disadvantaged, English learner, Black, Hispanic, and American Indian/Alaska Native students than students with only one or two scores (see table D.5 in appendix D). For example, 42 percent of students with three scores were economically disadvantaged compared to 54 percent of students with two scores and 60 percent of students with one score.

Student Performance and Progress

There are multiple approaches to describing the numbers of students who may be at risk of reading difficulty based on the literacy screening assessment data. Different stakeholders may be interested in answering different questions for different purposes—for example, who is at the highest level of risk? How many students are at any level of risk? How many students are consistently at risk? In the following sections, we provide information that addresses questions of student performance and progress using several approaches.

First, as described earlier, most screening assessments provide multiple performance levels that identify risk of reading difficulty (e.g., "some risk," "high risk"). This section first describes students at any level of risk, collapsing categories such as "some risk" and "high risk" into a single "at risk" category and categories such as "no risk" or "low risk" into a single "not at risk" category (see table 2 for details on categories for each assessment). Later analysis focuses on students in the highest risk categories or flagged as potentially at risk of dyslexia.

Additionally, students typically take screening assessments at multiple time periods during the year, and not all students have scores from all time periods. In the analysis that follows, we provide information for each time period separately to address questions about student performance at the BOY, MOY, and EOY and summarize information across time periods for students with one, two, and three scores, describing how often students are classified as at risk (i.e., once, twice, or three times or never, ever, or always).

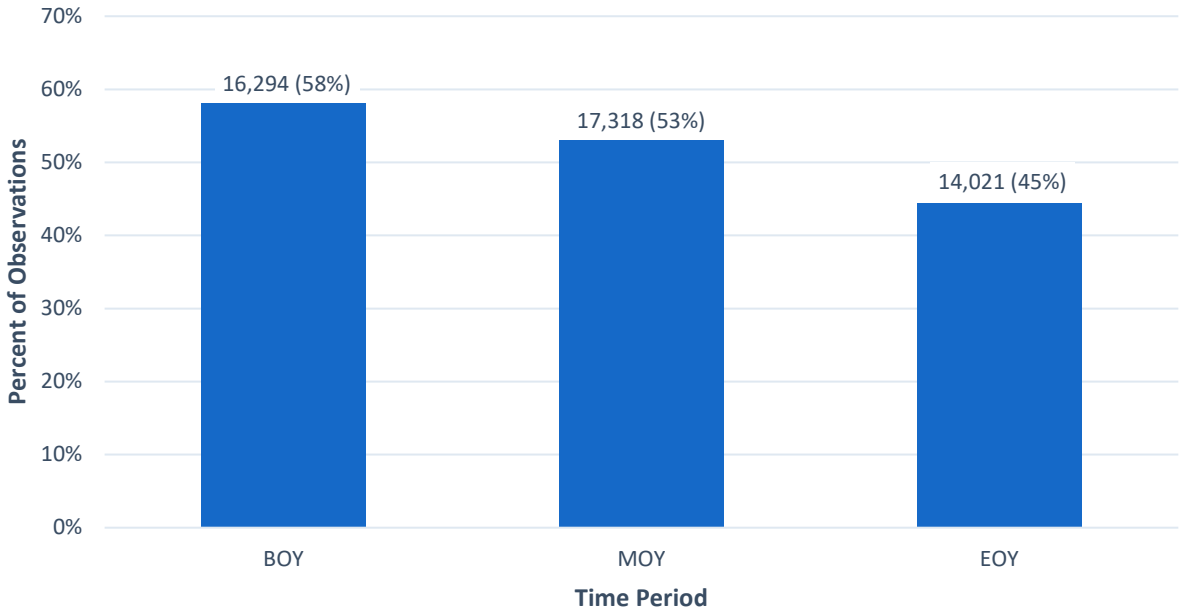
How many students are identified as at any level of risk by time period, grade, and student group?

The percentage of students identified as having at least some risk of reading difficulty ranged from 45 percent to 58 percent of students, depending on the time period (see figure 5). More students were classified as at risk at BOY (58%) than MOY (53%) or EOY (45%). Across time periods, 52 percent of students were below screening assessment benchmark levels of performance.

Key Findings

- Across all time periods combined, 52 percent of scores fell below benchmark level according to the screening assessments.
- More students were classified as at risk at BOY (58%) than MOY (53%) or EOY (45%).
- More grade K–2 students improved over time than grade 3 students.
- Economically disadvantaged 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 at risk at least once during the school year. Asian students were less likely than their peers to ever be classified as at risk.

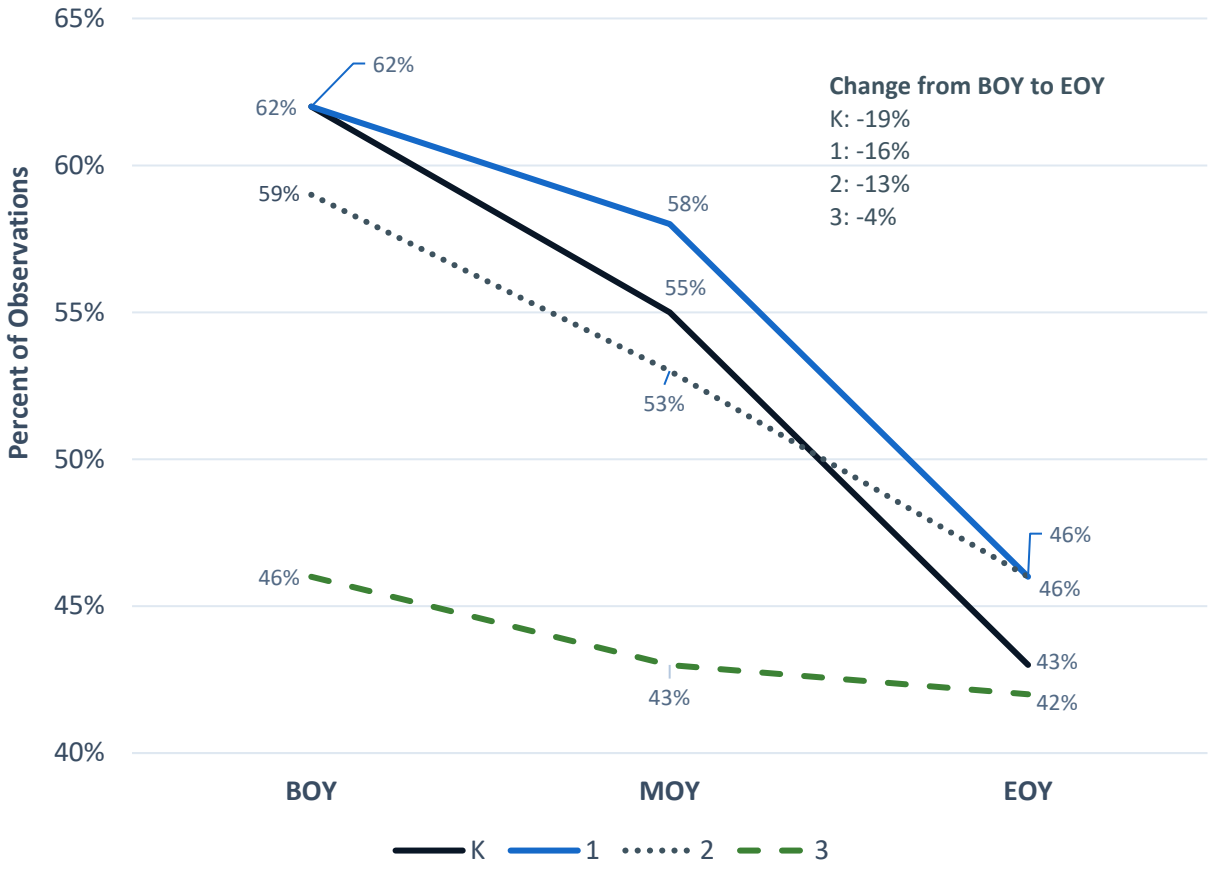
Figure 5. Fewer students were identified as at risk at EOY compared to earlier time periods



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments.

This pattern of decreasing numbers of students at risk over time periods is seen across grade levels, although there was much greater change among students in grades K–2 than grade 3 (see figure 6). While 62 percent of kindergarten students were classified as at risk at BOY, that percentage was 43 percent at EOY. Fewer students in grade 3 were classified as at risk in any time period than students in other grade levels; however, the percentage of students at risk decreased less than lower grade levels (i.e., from 46% to 42%).

Figure 6. Fewer students were at risk at EOY than at BOY across grade levels, but there was a greater change in the percentage of students at risk over time in grades K–2 than in grade 3



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. Includes students with one, two, or three scores.

Figures 5 and 6 show the percentages of students at risk during each time period separately, meaning they include students who may have one, two, or three scores. Among students with scores from BOY, MOY, and EOY, the decrease in the percentage of students at risk from BOY to EOY follows the same pattern, with greater changes for lower grades. But changes are larger for all grades (e.g., decrease of 26% of students at risk for grade K compared to 19% when including students with one, two, or three scores).[16]

[16] Changes by grade from BOY to MOY to EOY for students with all three scores are as follows: grade K: 63% to 51% to 37% (–26%), grade 1: 63% to 53% to 40% (–23%), grade 2: 57% to 48% to 39% (–18%), grade 3: 51% to 41% to 37% (–14%).

Examining individual student performance across time periods shows that approximately 64 percent of students were classified as below benchmark at least once during the school year (see table 6). A later section discusses how individual students progress from BOY to EOY.

Students with fewer scores were more likely to be classified as always at risk in a school year than those with more scores—47 percent of students with two scores were always below benchmark compared to 34 percent of students with three scores, for instance. Eighteen percent of all students were at risk in all three time periods (34% of the 54% of students who had scores in all three time periods). Conversely, more than one third of students were never classified as at risk in the school year, no matter how many scores they had.

Table 6. Percentage of students identified as at risk in school year by number of benchmark scores available

Benchmark Scores Available	Never Below Benchmark in School Year	Below Benchmark One Time	Below Benchmark Two Times	Below Benchmark Three Times
One Benchmark Available	40% (n = 2,264)	60% (n = 3,405)	N/A	N/A
Two Benchmarks Available	34% (n = 4,085)	19% (n = 2,254)	47% (n = 5,601)	N/A
Three Benchmarks Available	35% (n = 7,376)	16% (*n* = 3,291)	15% (n = 3,200)	34% (n = 7,024)
Total	**36% (n = 13,725)**	**23% (n = 8,950)**	**23% (n = 8,801)**	**18% (n = 7,024)**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per row if they were administered multiple screening assessments.

Only 54 percent of the sample had three benchmark records available; some students in the "Below Benchmark One Time" category only had one record; some students in the "Below Benchmark Two Times" category only had two records.

A greater percentage of students in grades 1 and 2 were classified as at risk at EOY compared to students in kindergarten and grade 3, and more grade 1 and grade 2 students were also below benchmark in all three time periods (i.e., always below benchmark in the school year) than students in kindergarten and grade 3 (see table 7). Fewer grade 3 students were ever below benchmark within the school year than students in other grades, and more were never below benchmark within the school year.[17]

[17] Note that this analysis focuses on student performance within the school year; a later section analyzes student performance across school years. Therefore, it is possible that students who were never below benchmark in their current grade were below benchmark in a prior school year.

Table 7. Percentage of students never below benchmark, always below benchmark, ever below benchmark, and below benchmark at EOY by grade level

Grade	Never Below Benchmark in School Year	Always Below Benchmark (Below Benchmark in BOY, MOY, and EOY)	Ever Below Benchmark in School Year	Below Benchmark at EOY
Kindergarten	32%	15%	68%	43%
Grade 1	31%	22%	69%	46%
Grade 2	35%	21%	64%	46%
Grade 3	48%	14%	53%	42%
All Grades	**36%**	**18%**	**64%**	**45%**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per row if they were administered multiple screening assessments. "Never Below Benchmark in School Year," "Always Below Benchmark," and "Ever Below Benchmark in School Year" include students with one, two, or three scores. "Below Benchmark at EOY" includes students with EOY scores, which could include students with one, two, or three scores.

Economically disadvantaged 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 score below benchmark at least once during the school year (see table 8). For example, 76 percent of economically disadvantaged students were ever below benchmark in the school year compared to 53 percent of non–economically disadvantaged students. The relative risk of economically disadvantaged students ever being classified as below benchmark was 1.4 times that of non–economically disadvantaged students. Asian students were less likely to ever be below benchmark. About the same proportions of female and male students were classified as at risk, as were White students and students of color. Across the four most commonly used screening assessments (DIBELS 8th Edition, i-Ready, mCLASS, and Star Early Literacy), the students ever classified as at risk during a school year vary from 55 percent to 76 percent (see table D.7 in appendix D); however, these differences are likely due to differences in the ways each screening assessment's benchmark scores are calculated and in the background characteristics of students participating in each assessment.

Table 8. Percentage of students never below benchmark, always below benchmark, ever below benchmark, and below benchmark at EOY by student group

Demographic	Never Below Benchmark in School Year	Always Below Benchmark (Below Benchmark in BOY, MOY, and EOY)	Ever Below Benchmark in School Year	Below Benchmark at EOY	Relative Risk of Ever Being Below Benchmark in School Year
Economically disadvantaged	24%	23%	76%	60%	1.4
Non-economically disadvantaged	47%	14%	53%	30%	
Female	37%	18%	63%	43%	1
Male	35%	19%	65%	46%	
Nonbinary	--	--	--	--	
English learner students	16%	21%	84%	69%	1.4
Non-English learner students	42%	17%	58%	36%	
Students receiving special education services	18%	33%	82%	67%	1.3
Students not receiving special education services	39%	15%	61%	40%	
Hispanic	19%	22%	81%	65%	1.4
Black	27%	21%	73%	54%	1.2
White	36%	20%	64%	42%	1
Asian	57%	11%	43%	26%	0.6
American Indian/Alaskan Native	18%	5%	82%	71%	†
Native Hawaiian/ Pacific Islander	36%	18%	64%	46%	†
Total	**36%**	**18%**	**64%**	**45%**	

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: 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 classified as at risk compared to students not in that group (e.g., economically disadvantaged students were 1.4 times more likely to be classified as ever at risk compared to non-economically disadvantaged students). Students in the "Never Below Benchmark in School Year," "Always Below Benchmark," "Ever Below Benchmark in School Year," and "Below Benchmark at EOY" categories can have different numbers of scores. Student groups with fewer than 10 students are not shown to protect student privacy.
† Not computed because groups are 5 percent or less of the sample.

How many students are identified as at significant risk of reading difficulty by time period, grade, and student group?

As noted, most of the early literacy screening assessments (except MAP Growth and MAP Reading Fluency) classify students into multiple levels of risk. For example, DIBELS 8th Edition risk categories include "Below Benchmark" and "Well Below Benchmark," indicating differences between students who may be at some risk and those who are at significant risk and may need more intensive support. That means that of the 52 percent of observations not meeting benchmark levels in the 2020/21 and 2021/22 school years, some of these cases indicate more moderate risk than others.

Key Findings

- Across all time periods combined, 30 percent of benchmark scores fell into the highest risk level category.
- Thirty-nine percent of students were classified as ever being at significant risk during the school year, compared to 64 percent ever at any level of risk.
- Among students with scores in each time period, 17 percent were classified as at significant risk in all three time periods, and the percentage at significant risk decreased from BOY to MOY to EOY.
- Economically disadvantaged students, English learners, students receiving special education services, and Hispanic students were about twice as likely as their peers not in those groups to be classified as at significant risk

Focusing just on each screening assessment's highest risk level shows that about 30 percent of scores demonstrate the highest level of risk for each screening assessment (see table 9). As seen in previous analyses using all levels of risk, fewer students are classified as at high risk as the school year progresses (i.e., 32% are at a high level of risk in BOY, 30% in MOY, and 26% in EOY).

Most of the early literacy screening assessments classify about one quarter to one third of benchmark scores as indicating high or significant levels of risk. However, there are differences between screening assessments (see table 9), with Acadience Reading identifying only 2 percent and Lexia RAPID identifying 65 percent. Again, however, these differences may be related to the manner in which each screening assessment set its cut scores to identify students at risk and/or differences in the background characteristics of participating students in each assessment. For example, only 7 percent of the Acadience Reading sample was classified as economically disadvantaged, while 59 percent of the Lexia RAPID sample was economically disadvantaged. Note that, because MAP Growth and MAP Reading Fluency provide only one risk level, the percentages of students described as at risk in this section are the same as those reported in the previous section.

Table 9. Percentage of students identified as at significant risk at BOY, MOY, and EOY by assessment

Early Literacy Screening Assessment	Highest Risk Level	BOY	MOY	EOY	Total
Acadience Reading	Well Below Benchmark	2%	2%	3%	**2%**
DIBELS 8th Edition	Well Below Benchmark	38%	34%	23%	**32%**
mCLASS	Well Below Benchmark	30%	29%	21%	**27%**
FastBridge aReading	High Risk	28%	21%	23%	**24%**
FastBridge CBMreading	High Risk	33%	35%	35%	**34%**
FastBridge earlyReading	High Risk	27%	42%	39%	**35%**
i-Ready	Flagged according to iRDI	12%	12%	12%	**12%**
ISIP ER	Tier 3	32%	28%	28%	**29%**
Lexia RAPID	Low Likelihood of EOY Grade-Level Success	80%	64%	51%	**65%**
MAP Growth	Intensive Intervention	49%	26%	23%	**30%**
MAP Reading Fluency	Flagged—Universal Screener Outcome	35%	30%	N/A	**32%**
Star Early Literacy	Urgent Intervention	36%	37%	36%	**37%**
Star Early Literacy Spanish	Urgent Intervention	50%	37%	39%	**38%**
Star Reading	Urgent Intervention	19%	26%	30%	**26%**
Total		**32%**	**30%**	**26%**	**30%**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: No student taking MAP Reading Fluency had three benchmark records because only BOY and MOY scores were provided. MAP Growth and MAP Reading Fluency only have two risk levels. Therefore, the percentage of students at "significant risk" will equal the percentage previously discussed for these two screening assessments.

Thirty-nine percent of students were classified as ever being at significant risk during the school year (see table 10), compared to 64 percent of students classified as ever at any level of risk in the previous section (see table 6). Among the students who had benchmarks available in all three time periods, 34 percent were at significant risk at BOY, 28 percent were at significant risk at MOY, and 22 percent were at significant risk at EOY. Seventeen percent of students who had three scores were classified as at significant risk in all three time periods. These students were more likely to be economically disadvantaged, be English learners, and receive special education services than the overall sample of students with three available scores. For example, only 18 percent of students with three scores were English learners, while 36 percent of students who were at significant risk at BOY, MOY, and EOY were English learners.

Table 10. Percentage of students never at high risk and at high risk one, two, or three times by number of benchmark scores available

Benchmark Scores Available	Never at High Risk in School Year	High Risk One Time	High Risk Two Times	High Risk Three Times
One Benchmark Available	62% (n = 3,507)	38% (n = 2,162)	N/A	N/A
Two Benchmarks Available	61% (n = 7,315)	15% (n = 1,736)	24% (n = 2,889)	N/A
Three Benchmarks Available	60% (n = 12,563)	13% (n = 2,681)	10% (n = 2,026)	17% (n = 3,621)
Total	**61% (n = 23,385)**	**17% (n = 6,579)**	**13% (n = 4,915)**	**9% (n = 3,621)**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per row if they were administered multiple screening assessments.

Only 54 percent of the sample had three benchmark records available; some students in the "High Risk One Time" category only had one record; some students in the "High Risk Two Times" category only had two records.

As with students identified as at any level of risk, economically disadvantaged students, English learner students, students receiving special education services, and Hispanic students were more likely than their peers not in those groups to be classified as at significant risk at least once during the school year (see table D.10 in appendix D, which shows students identified as at significant risk by grade and other background characteristics). These students were about twice as likely to be classified as at significant risk than their peers not in those groups.

How many students are identified as at potential risk for dyslexia?

Key Findings

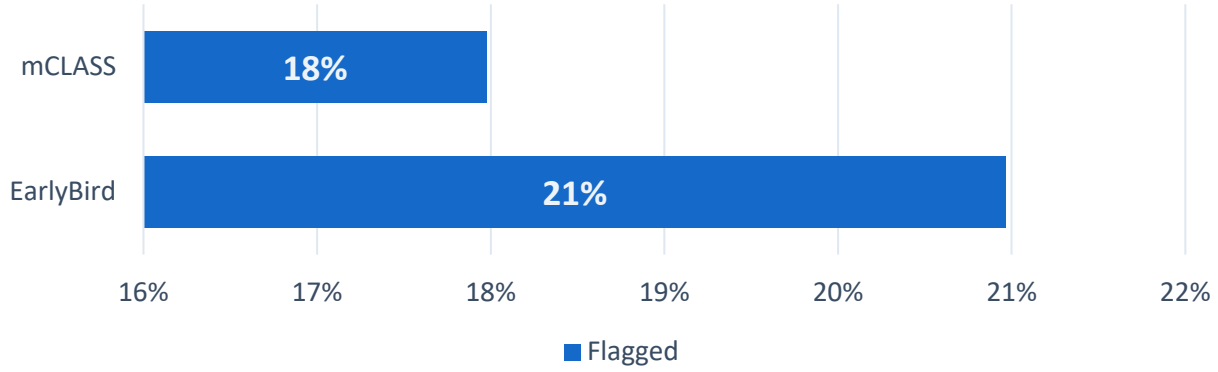
- EarlyBird and mCLASS flag similar percentages of students exhibiting signs of risk related to dyslexia (21% and 18% respectively).
- The percentage of students flagged according to mCLASS increases across time periods, unlike EarlyBird where there is a smaller percentage of students identified at EOY than MOY.
- The percentage of scores flagged by the EarlyBird and mCLASS dyslexia risk flags is similar to the percentage of students classified as at significant risk three times in the school year.

Several early literacy screening assessments (i.e., EarlyBird, MAP Reading Fluency, and mCLASS) provide additional measures designed specifically to identify students potentially at risk of dyslexia (see table 2). These measures are binary flags indicating whether a student exhibits or does not exhibit signs of risk related to dyslexia, although all note that the flags do not indicate a diagnosis of dyslexia (note that DESE did not review or approve screening assessments specifically for dyslexia flagging). The i-Ready Diagnostic assessment also provides an additional risk indicator called the "i-Ready Reading Difficulty Indicator," or iRDI, a score below which "students are considered to have a significant reading difficulty."[18] In Massachusetts, the iRDI score is intended to be used in conjunction with other measures to identify students at risk of dyslexia. However, no data were available for students on the other relevant measures (e.g., rapid automatized naming). As a result, i-Ready scores and flags are not used in this section of the analysis, with the analysis of the iRDI indicator being included in the previous discussion of significant risk. Also, no dyslexia screening assessment data were available for MAP Reading Fluency.

Across the 2020/21 and 2021/22 school years, there are 3,232 observations with an mCLASS risk flag and 1,359 observations with an EarlyBird risk flag. Overall (for both screening assessments and all three time periods), 19 percent of the screening assessment scores were classified as exhibiting a significant reading risk, according to these "dyslexia" risk flags (figure 7). The risk flags for EarlyBird and mCLASS both aim to flag students at risk of dyslexia and identify similar percentages of students (21% and 18%, respectively), which are in line with a widely cited estimated percentage of individuals with dyslexia (i.e., 20%), according to the Yale Center for Dyslexia and Creativity (n.d.).

[18] i-Ready, Dyslexia Screen with i-Ready Assessment, https://i-readycentral.com/download/?res=64534&view_pdf=1.

Figure 7. Approximately one fifth of scores met criteria for potential risk of dyslexia across time periods and assessments with such measures



Source: 2020/21 and 2021/22 district-provided screening assessment data

The percentage of students flagged on the dyslexia risk screening assessment according to mCLASS increases each time period (from 11% in BOY to 18% in MOY to 22% in EOY), unlike EarlyBird, where there is a smaller percentage of students identified at EOY than MOY (see table 11). Overall, however, the percentage of scores flagged as being at significant risk for reading difficulties, according to these "dyslexia" risk flags, is similar to the percentage of students classified as at significant risk three times in the school year among students with three scores (see table 10).

Table 11. Percentage of students identified as at potential risk for dyslexia by time period and assessment

Early Literacy Screening Assessment	BOY	MOY	EOY	Total
mCLASS	11%	18%	22%	**18%**
EarlyBird	N/A	29%	14%	**21%**
Total	**11%**	**21%**	**18%**	**19%**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: No EarlyBird records were available at BOY.

How does student performance change as the school year progresses? Do students identified as at risk remain at risk?

Of students with at least two benchmark scores, most students who were at risk in one time period were also at risk in a later time period, but about 25 to 40 percent of students did get on track after missing an earlier benchmark (see table 12).

Sixty percent of students who did not meet the benchmark at BOY also did not meet the benchmark at EOY, while 40 percent of students who started off at risk at BOY were **not** at risk by EOY.

Conversely, the vast majority of students who started the year at benchmark ended at benchmark. Ninety-one percent of students who met benchmark expectations on a screening assessment at BOY also met the benchmark at MOY, and 93 percent met the benchmark at EOY (see table 12). Similarly, 92 percent of students who met benchmark expectations at MOY met the benchmark at EOY.

Key Findings

- Most students meeting benchmark in one time period also met benchmark in a later time period; similarly, most students at risk in one time period were also at risk in a later time period.
- Some students did get on track (move from at risk to not at risk) during the school year. Forty percent of students who started off at risk at BOY were not at risk by EOY.
- Improving performance between BOY and MOY significantly improves the likelihood of success at EOY. Eighty-seven percent of students who did not meet benchmark at BOY but met benchmark at MOY also met benchmark at EOY.

Table 12. Percentage of students classified as at risk or at benchmark by performance in prior test period

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	92%

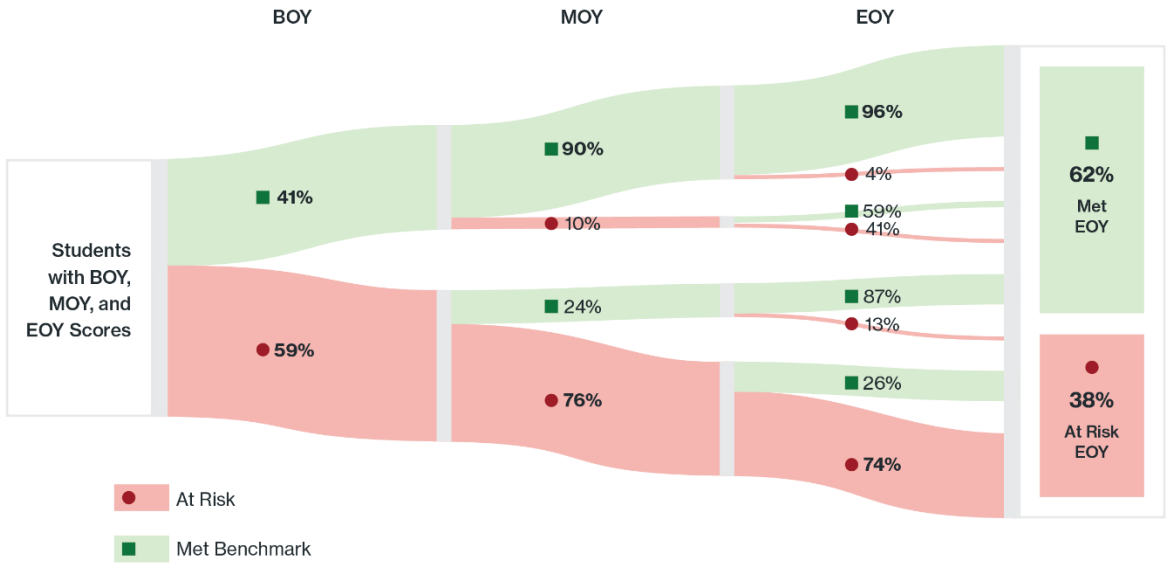
If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	76%	60%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
MOY	N/A	N/A	74%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 12 shows how students with at least two scores (about 85% of students in the sample) progress. Students who had a record during each time period of the school year (54% of the sample) had similar patterns of progress (see figure 8).

Figure 8. Student performance remained relatively consistent throughout the school year but improving performance between BOY and MOY improves likelihood of success at EOY



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (54%) were included in the figure.

Of the students who did not meet benchmark at BOY, most continued to perform below benchmark for the remainder of the school year; 76 percent of these students performed below benchmark at MOY, and 74 percent of the students who performed below benchmark in both time periods also performed below benchmark at EOY. However, about one quarter of students who were classified at risk at BOY were not at risk at MOY, and of these students, the vast majority (87%) were at benchmark by EOY. Improving performance between BOY and MOY significantly improves the likelihood of success at EOY; the odds that a student meets

benchmark at EOY if they scored below benchmark at BOY but above benchmark at MOY is about 18 times the odds that they meet the benchmark at EOY if they scored below benchmark in both previous time periods.

Among the 41 percent of students who performed at or above benchmark at BOY, the vast majority scored above benchmark at MOY and EOY as well. However, students who met the benchmark at BOY but not at MOY also struggled to meet expectations at EOY—41 percent who fell off track at MOY were still at risk at EOY.

How does student progress vary by risk level, grade, and student background characteristics?

Among students with at least two scores, the percentages of students who get on track after being identified as at risk at BOY decrease over grade levels—that is, fewer students at higher grade levels get on track compared to students at lower grade levels. Forty-nine percent of kindergarten students who started at risk end at benchmark compared to 42 percent of grade 1 students and 33 percent of grade 2 and grade 3 students (see tables D.11 through D.14 in appendix D).

Key Findings

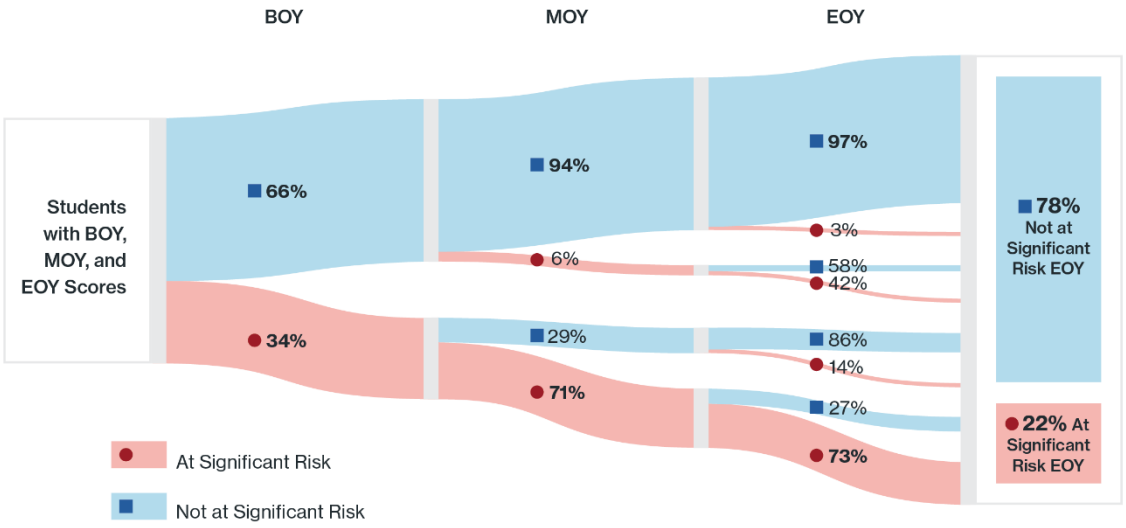
- Fewer students at higher grade levels get on track after being identified as at risk than students in lower grade levels.
- Fewer students at significant risk at BOY achieve benchmark at EOY than students at lower levels of risk.
- Improving performance between BOY and MOY significantly improves the likelihood of not being at significant risk at EOY. Eighty-six percent of students who were at significant risk at BOY but not at MOY were not at significant risk at EOY.

Fewer English learner students, economically disadvantaged students, students receiving special education services, and Black and Hispanic students got on track by EOY after being identified as at risk at BOY or MOY than their peers not in these student groups (see tables D.15 to D.32 in appendix D).

Smaller numbers of students at significant risk (using the higher risk levels described in an earlier section) improve their performance to benchmark levels over time. Although 40 percent of students at any level of risk at BOY were not at risk by EOY, only 24 percent of students at high risk were not at any level of risk by EOY.

Figure 9 shows how students who had scores from all three time periods who were identified as at significant risk (or not) progress over time periods. As with students at any level of risk, students who get on track by MOY are much less likely to be identified as at significant risk at EOY.

Figure 9. Most students at significant risk at BOY remained at significant risk throughout the school year, but improving performance between BOY and MOY improved the likelihood of success at EOY



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. Only students with scores across each time period (54%) were included in the figure.

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

Key Findings

- Fifty-six percent of students who were at risk at BOY at one grade level were still at risk at EOY in the following grade level.
- Students in earlier grades were more likely to get on track across years than students in later grades. Thirty-eight percent of students who started in kindergarten at risk were still at risk at the end of grade 1, compared to 54 percent for students moving from grade 1 to 2, and 69 percent for students moving from grade 2 to 3.

Because the screening assessment data sample for 2020/21 is relatively small, only 1,779 students had benchmark scores in both years. In this section, we examine changes from students with scores from both years, focusing on change from BOY in 2020/21 to EOY in 2021/22 (which provides the most time for student growth). Eight hundred and fifty-five students had BOY scores in 2020/21 and EOY scores in 2021/22 from the same screening assessment, representing about 2 percent of the total number of 38,500 students in the sample (see table 13).

Table 13. Percentage of students with BOY scores in 2020/21 and EOY scores in 2021/22 by grade level

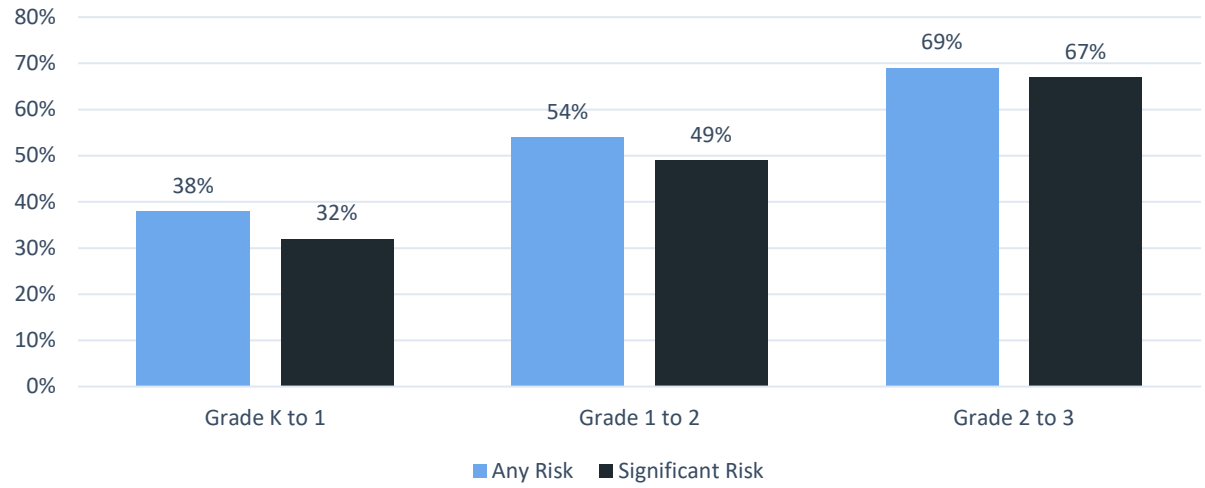
Students With Scores From BOY 2020/21 to EOY 2021/22	Number	% at Risk at BOY	% at Significant Risk at BOY
Grade K to 1	244	38%	25%
Grade 1 to 2	293	30%	13%
Grade 2 to 3	318	42%	22%
Total	855	37%	20%

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Screening assessments with fewer than 10 students in both time periods were excluded from analysis.

Of the students with scores in both time periods, about 37 percent were identified as at risk across grade levels at BOY and about 20 percent at significant risk.

While within a school year about 60 percent of students who started the year at risk were still at risk at EOY, across years that percentage decreases to about 56 percent for all students, but there are substantial differences across grades (see figure 10).

Figure 10. More students at higher grade levels stayed at risk from grade to grade than students at lower grade levels



Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Screening assessments with fewer than 10 students in both time periods were excluded from analysis.

More students in earlier grades got on track across years than did students in later grades. For example, 38 percent of students who started in kindergarten at risk were still at risk in grade 1, compared to 54 percent of students who started at risk in grade 1 and were still at risk in grade 2, and 69 percent of students moving from grade 2 to grade 3. The same patterns in progress appear for students identified as at significant risk.

Although the sample is small, these results suggest that across grades, students who begin at risk can get on track, and this is particularly true for students in kindergarten. Additional data in future years will provide better insight into cross-grade performance.

Discussion of Results

The goal of these analyses is to provide preliminary information about performance and progress of students in the state of Massachusetts based on a sample of state grantees that may inform improvement strategies. For example, by knowing how many and which students are at risk and when, the state may be able to target resources for these students or identify different strategies to use at different time periods.

Overall, data indicate that about 52 percent of students were at risk across time periods. However, the percentages of students at risk vary by time period, with a smaller proportion (45%) at risk at EOY than at BOY. Many students were classified as at risk at least once, but among those with at least two scores, about 35 percent were never classified as at risk and about 38 percent were always at risk. Using benchmark levels that indicate more significant risk across screening assessments, about 30 percent of students are identified as "high risk," with a smaller proportion (about 20%) always at high risk. Economically disadvantaged 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 score below benchmark at least once during the school year.

Students who did not meet benchmark at BOY most often did not meet benchmark at MOY or EOY. However, of students who did not meet benchmark at BOY but did at MOY, a majority went on to meet benchmark at EOY, suggesting that action to get students to benchmark can help them stay on track over time. Students who began the year meeting benchmark most often met benchmark at EOY as well.

Despite having 14 assessments originally approved for use, a few assessments were by far the most used and account for about three quarters of scores: DIBELS 8th Edition, i-Ready, and Star Early Literacy account for over 60 percent of scores, followed by mCLASS with 14 percent.

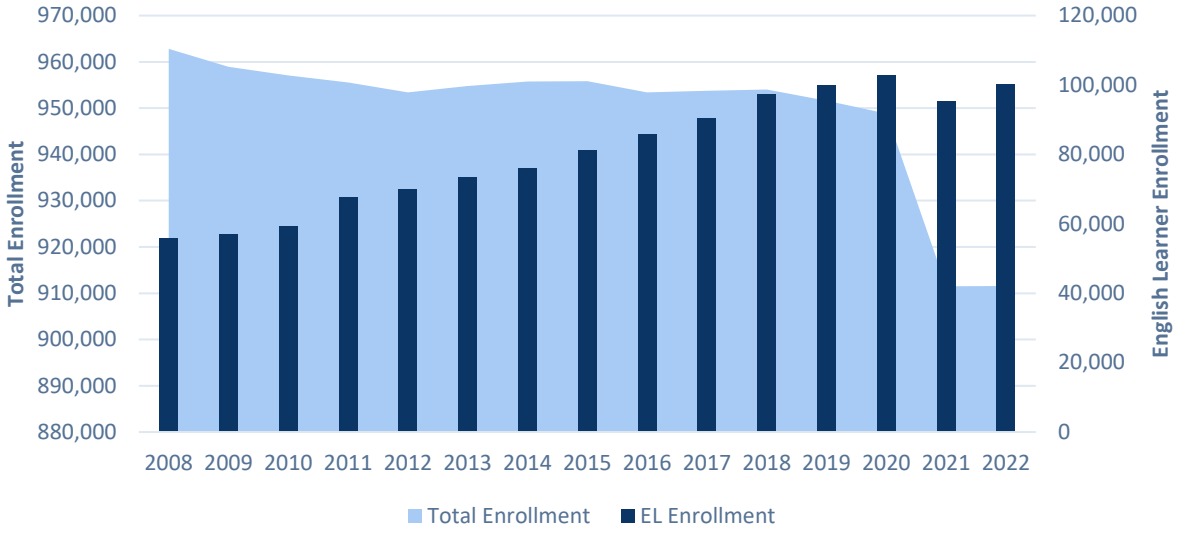
Although the screening assessment data sample described in this section includes about 7 percent of students in grades K–3 across the state, it does not perfectly mirror the composition of the state population in grades K–3 overall. For example, the sample includes a greater number of English learner students. Some caution is therefore needed in extrapolating the

findings from the sample to the larger population. Additionally, many students do not seem to have data from all assessment periods, and in particular, much data is missing from the BOY period. This missing data may indicate that files from schools or districts are incomplete or that they do not assess students in all time periods.

English Learner Student Performance

According to the National Center for Education Statistics (NCES), the nation's English learner student population is increasing; this is also true for Massachusetts. Despite a decrease in total student enrollment in Massachusetts schools in recent years, the English learner student population in the Commonwealth has been increasing, with English learner students comprising 11 percent of total enrollment in the 2021/22 school year (see figure 11).

Figure 11. English learner student enrollment increased steadily over the last 15 years while total student enrollment decreased until COVID-related enrollment disruptions



Source: Author's analysis of Massachusetts Department of Elementary and Secondary Education data (DESE Selected Populations Reports, https://profiles.doe.mass.edu/statereport/selectedpopulations.aspx)

Disentangling the extent to which screening assessment data about English learner students reflect the fact that students are learning English as a language from gaps in reading skills in English that also affect non–English learner students can be challenging. This is especially true if early literacy screening assessment benchmarks are established based on samples of proficient English speakers or without examination of performance differences among English learner students (Cummings et al., 2019; Sandberg & Reschly, 2011).

Despite this fact, some prior research has suggested that screening English learner students for reading difficulty in English can provide useful information about students' skill development (Gersten et al., 2007; Good et al., 2011)—that is, these measures can help educators identify students whose reading skills in English suggest needs beyond instruction and supports to improve their English language proficiency overall.

However, as seen in the previous section, early literacy screening assessments in the sample of data available for 2020/21 and 2021/22 **do** identify more English learner students as at risk than their non–English learner student peers.[19]

Thus, examining how English language proficiency and other student characteristics relate to meeting benchmarks may provide additional information about the importance of English language proficiency for meeting screening assessment benchmarks and about which students within the English learner population may need additional supports to develop reading skills in English.

Table 14 shows the specific research questions to be addressed in this section after discussion of the sample of English learner students and how well it represents English learner students in grades K–3 in the state as a whole.

Table 14. Research questions related to English learner students

Topic	Research Question
English learner student performance and English proficiency	• How many literacy screening assessment benchmark scores are available for English learner students? To what extent does the sample of English learner students represent the overall English learner student population in the state? • What is the relationship of English language proficiency to screening assessment performance and achieving benchmarks? Do students at early stages of English proficiency achieve benchmarks? Are students with higher levels of English proficiency classified as at risk? • How does performance and benchmark attainment vary for English learners by specific domains of English proficiency?
English learner student performance, native language, and program of instruction	• How does performance and benchmark attainment vary for English learners by native language and English Language Education program?

[19] Note that 73 percent of English learner students are also classified as economically disadvantaged, and screening assessment data also identify higher proportions of economically disadvantaged students as at risk than non–economically disadvantaged students.

English Learner Screening Assessment Data Overview

How many literacy screening assessment benchmark scores are available for English learner students? To what extent does the sample of English learner students represent the overall English learner student population in the state?

The screening assessment data sample overall has more English learner students than the state as a whole at the same grade levels. This section reviews the available data for English learner students in more detail to examine how well the English learner group represents the K–3 English learner population in the state (and, as a result, how generalizable findings from these data may be for English learners).

Across the 2020/21 and 2021/22 data, 20,994 benchmarks (about 23% of the sample) are from students classified as English learners.[20] These observations represent 7,939 English learner students overall, with 97 percent of these students (7,663) coming from the 2021/22 school year. English learner students represent about 7 percent of the screening assessment data sample in 2020/21 and almost one quarter of the sample in 2021/22 (24%). Of the 7,939 English learner students in the sample, however, about 15 percent took multiple screening assessments—1,196 took two screening assessments, while 243 took three screening assessments. As a result, students can have multiple benchmarks, which are treated independently in the analysis that follows (see table 15, which shows 9,621 students). The vast majority of the students taking multiple screening assessments took multiple Star screening assessments (i.e., Star Early Literacy, Early Literacy Spanish, and Reading).

Key Findings

- Early literacy screener assessment sample includes about 8,000 English learners.
- Earlier grade levels (K–1) are overrepresented among English learner students in the sample compared to the state, and the sample contains a higher percentage of English learner students in their first year in U.S. schools than the state.
- The distribution of native languages spoken in the sample displays some overrepresentation of native Spanish, Chinese, and Arabic speakers compared to the state.
- The percentage of English learner students in the sample enrolled in each type of English Language Education (ELE) program is roughly similar to the state.
- Performance on ACCESS indicates earlier stages of English language development among students in the sample.

[20] The sample of benchmarks available for analysis is 91,965 (rather than the 92,222 previously used) because some observations did not match with the June or October SIMS collection data.

Table 15. Numbers of English learner students and benchmark scores by assessment

Early Literacy Screening Assessment	Number of Students	Number of Benchmark Scores
Acadience Reading	--	--
DIBELS 8th Edition	1,658	4,345
mCLASS	834	2,238
FastBridge aReading	13	30
FastBridge CBMreading	17	32
FastBridge earlyReading	--	13
i-Ready	590	1,491
ISIP ER	161	429
Lexia RAPID	618	1,620
MAP Growth	318	636
MAP Reading Fluency	--	--
Star Early Literacy	3,465	6,904
Star Early Literacy Spanish	1,263	2,108
Star Reading	674	1,138
Total	**9,621**	**20,994**

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: "Number of Benchmark Scores" 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. Cells with sizes less than 10 are suppressed to protect student privacy.

Four grade levels are included in the screening assessment data—assuming equal distribution of students would mean that each grade level should include 25 percent of the sample of students. However, among English learner students, 30 percent are from kindergarten, 32 percent are from grade 1, 23 percent are from grade 2, and only 15 percent are from grade 3 (see table 16).

Table 16. Percentages of English learner students in the screening assessment data sample and the state

Grade Level	Sample (2020/21 and 2021/22)	State (2020/21 and 2021/22)
Kindergarten	30% (n = 2,348)	24%
Grade 1	32% (n = 2,559)	26%
Grade 2	23% (n = 1,863)	26%
Grade 3	15% (n = 1,169)	24%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Among the English learner students in the sample, the 10 native languages spoken most frequently (see table 17) include eight of the top 10 native languages spoken most frequently by the entire K–3 student population in the state. The top four native languages (i.e., Spanish, Portuguese, Chinese, and Arabic) make up over 80 percent of the English learner students in the early literacy screening assessment sample. As shown, the sample has an overrepresentation of native Spanish, Chinese, and Arabic speakers, and an underrepresentation of native Portuguese speakers.

Table 17. Number and percentage of English learner students in the screening assessment data sample and the state by native language

Language Spoken and Ranking by Enrollment Number (Sample)	Ranking by Enrollment Number (State)	Number of English Learner Students (Sample)	% of English Learner Students (Sample)	% English Learner Students (State)
1. Spanish	1	4,761	60.0%	50.4%
2. Portuguese	2	837	10.5%	14.0%
3. Chinese	3	513	6.5%	4.6%
4. Arabic	4	352	4.4%	3.3%
5. Vietnamese	8	175	2.2%	2.3%
6. Creole (Haitian)	6	149	1.9%	3.2%
7. Twi	11	141	1.8%	0.8%
8. Russian	9	89	1.1%	1.9%
9. Albanian	15	83	1.1%	0.7%
10. Khmer	10	82	1.0%	1.9%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

In Massachusetts, approximately 27 percent of English learner students in the 2020/21 and 2021/22 school years were in their first year in school in the United States at the time of the data collection (i.e., recorded as attending U.S. schools for less than 12 months). However, almost half of the students in the screening assessment data sample (45%) are in their first year in U.S. schools.[21] The percentage of students in their first year in the United States is larger than the state's percentage in each grade level except grade 3. Similarly, the percentage of English learner students classified as immigrants (under the federal definition) is slightly higher in the sample than in the state (23 percent versus 19 percent), while the percentage of English learner students receiving special education services in the sample is equal to that of the state.

In terms of the types of services English learner students receive, the percentage of English learner students in the sample enrolled in each type of program is roughly similar to the state. The sample contains a larger percentage of English learner students whose parent/guardian

[21] Excluding kindergarten students, the percentage of students in their first year in the United States is 27 percent for the grantee screening assessment data sample and 12 percent for the state for grades 1 to 3.

opted out of all English Language Education (ELE) programs offered, though the numbers are small both in the sample and the state (see table 18).

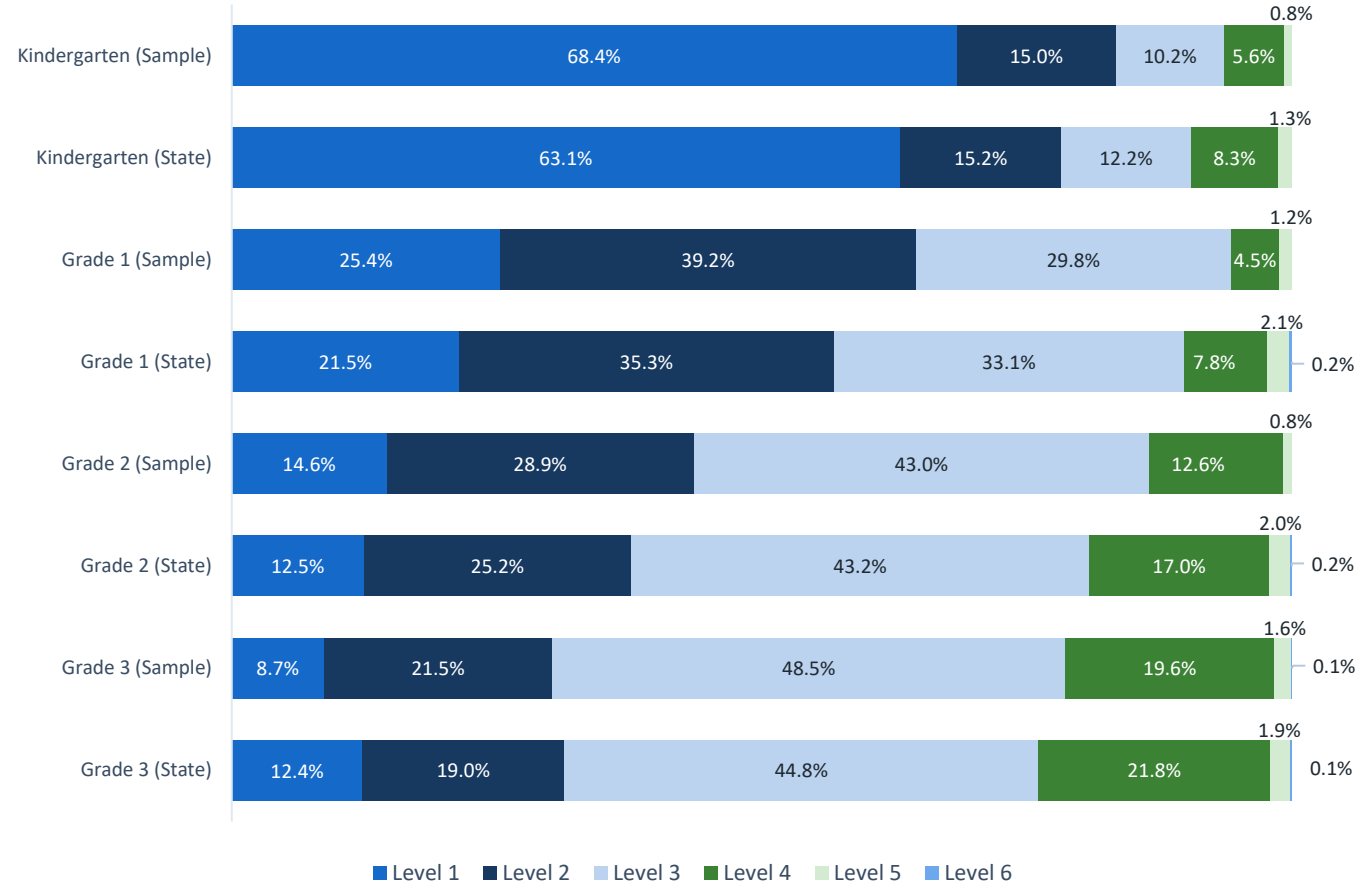
Table 18. Number and percentage of English learner students in the screening assessment data sample and the state by type of English learner services received

English Language Education (ELE) Program	Sample (2020/21 and 2021/22)	State (2020/21 and 2021/22)
Not Enrolled in an English Language Education Program	0% (n = 0)	0.8%
Sheltered English Immersion	94.7% (n = 7,515)	93.0%
Dual Language Education	2.0% (n = 160)	4.3%
Oher Bilingual Programs	0.5% (n = 37)	0.1%
Parent/Guardian Opted Out of All ELE Programs	2.8% (n = 223)	1.8%
Transitional Bilingual Education	--	0.04%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: Cells with sizes fewer than 10 are suppressed to protect student privacy.

More English learner students in grades K–2 in the sample scored in the bottom three proficiency levels on ACCESS than students across the state, indicating earlier stages of English language development among students in the sample (see figure 12). Performance in the sample across grade levels does follow the state pattern of students performing at higher proficiency levels on ACCESS at higher grade levels.

Figure 12. At grades K–2, more English learner students in the screening assessment data sample perform at lower composite ACCESS proficiency levels than English learners in the state as a whole



Source: 2020/21 and 2021/22 district-provided screening assessment data and state-provided ACCESS data
Note: Sample excludes 290 English learner students (of the 7,939 English learner students in the screening assessment sample data) due to missing ACCESS proficiency levels.

Overall, these data suggest that English learner students in the sample may be somewhat different from English learner students across the state in terms of their English language proficiency, age, and time in school in the United States. Such differences could be due to the specific schools and districts providing screening assessment data. (Note that the state does prioritize districts with higher proportions of high-need students for some grant funding.)

English Learner Student Performance and Progress

What is the relationship of English language proficiency to screening assessment performance and achieving benchmarks? Do students at early stages of English proficiency achieve benchmarks? Are students with higher levels of English proficiency classified as at risk?

Despite research showing that early literacy screening assessments can provide valid information for students just learning English, educators may still be concerned about how reliable or useful information from screening assessments is for these students. In this section, we examine the relationship between English proficiency as measured by ACCESS scores and screening assessment performance to determine how well students at different levels of English proficiency perform.

Key Findings

- Sixty-nine percent of English learner students did not meet benchmark at EOY compared to 36 percent for non–English learner students.
- More English learner students with higher composite proficiency levels on ACCESS met screener assessment benchmarks than those with lower levels.
- English learner students with an ACCESS composite proficiency level score of 4 or above met benchmark expectations at higher rates than their non-EL peers.
- Students with the least English proficiency (according to ACCESS) were able to demonstrate benchmark-level performance on screening assessments.
- Eighty-eight percent of students at risk according to the English language version of Star Early Literacy were also classified as at risk on the Spanish version.

Given that ACCESS is typically administered during January or February of the school year, the timing of the MOY screening assessment results is likely most in line with the ACCESS administration compared to BOY and EOY screening assessment results. However, as there are likely differences in the timing of assessments across schools and districts due to district calendars and testing schedules, we compare ACCESS results to both the MOY and EOY screening assessment data.

As described in the earlier section on available data, ACCESS scores provide four composite levels that range from 1 through 6 and that consider student performance on each of the four language domains. Scores are derived from a weighted combination of domain scaled scores, as follows:

- Overall/Composite: 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%)

Overall, English learner students with higher composite proficiency levels on the ACCESS assessment more frequently met benchmark expectations than those with lower performance on the ACCESS assessment (see table 19). However, English learner students with an overall ACCESS proficiency level score of 4 or above met benchmark expectations at higher rates than their non–English learner peers. This finding is in line with previous research (e.g., Gersten et al., 2007) that suggests that English learner students can perform at proficiency levels comparable to their native-English-speaking peers on literacy-related measures, although in this case English learners are not performing comparably but are outperforming their non–English learner peers.[22]

As with non–English learners, English learner students improve their performance over the year, with fewer students at risk at EOY than at BOY across ACCESS proficiency levels.

Table 19. Percentage of students who did not meet benchmarks at MOY or EOY by ACCESS composite performance level

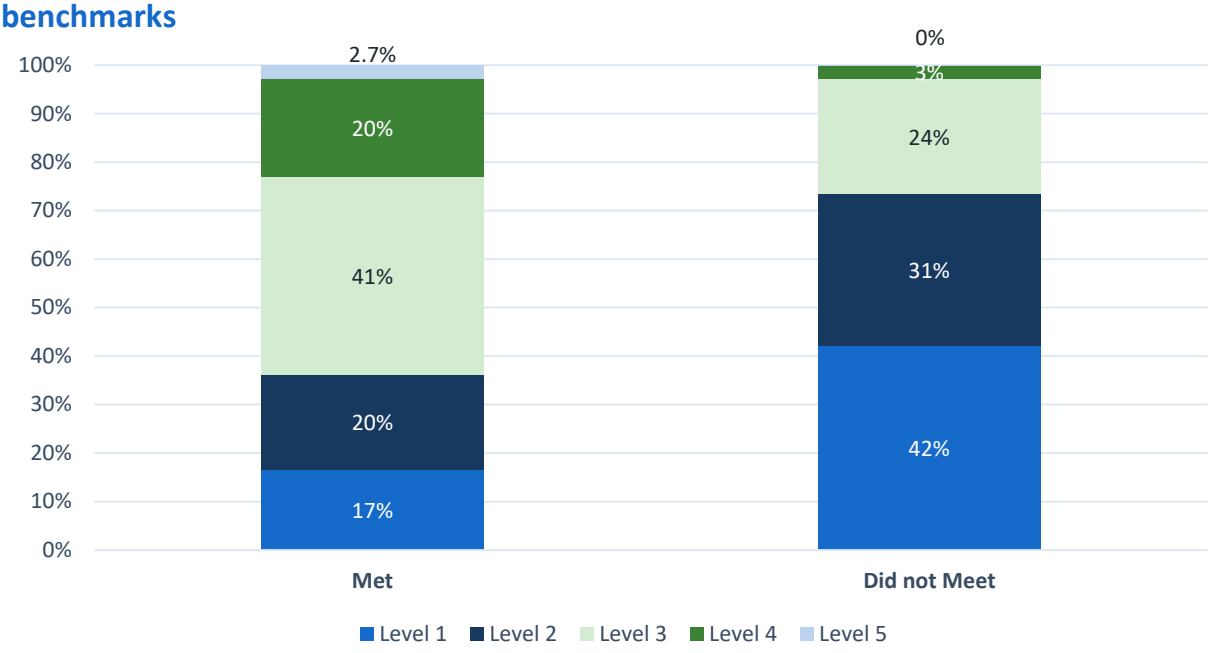
ACCESS Composite Performance Level	Did Not Meet at MOY	Did Not Meet at EOY
Level 1	91%	85%
Level 2	86%	78%
Level 3	67%	56%
Level 4	32%	23%
Level 5	11%	4%
Level 6	--	--
Average English Learners	77%	69%
Average Non-English Learners	45%	36%

Source: 2020/21 and 2021/22 district-provided screening assessment data, October and June SIMS collection data, and state-provided ACCESS data
Note: Sample excludes 290 English learner students (of the 7,939 English learner students in the screening assessment sample data) due to missing ACCESS proficiency levels. Eleven students were classified both as English learner and non–English learner students in the 2021/22 school year. These students received different classifications in different schools/districts. Cells with sizes fewer than 10 are suppressed to protect student privacy.

[22] In Massachusetts, the ACCESS requirement for exit from English learner status is an overall composite level score of 4.2 and a literacy composite level score of 3.9. About half of the students in level 4 met exit criteria.

Students who met benchmark at EOY included students at all levels of ACCESS proficiency—that is, even students with very little English proficiency were able to demonstrate benchmark-level performance on early literacy screening assessments (see figure 13). Thirty-seven percent of students who met EOY benchmarks were in the lowest two ACCESS levels. Conversely, English learner students classified as at risk at EOY included some students at higher levels of proficiency (24% from level 3 and about 3% from level 4 and above).

Figure 13. Students who meet benchmark include students at all levels of English proficiency, but few students at higher levels of English proficiency did not meet benchmarks



Source: 2020/21 and 2021/22 district-provided screening assessment data, October and June SIMS collection data, and state-provided ACCESS data

As described earlier, most of the screening assessments include multiple levels of risk. Focusing on students at the most significant risk (i.e., the lower benchmark levels) shows that about 51 percent of the English learner student scores in the sample were classified as significantly at risk as compared to about 77 percent at any level of risk. English learner students at significant risk also include students from across ACCESS proficiency levels (though most are at earlier levels of English proficiency), suggesting that students at any level of English proficiency can be classified at high levels of risk.

Just over 1,000 students took screening assessments in both Spanish and English (mostly Star assessments, as noted in an earlier section), 939 of whom were classified as English learner students. Of these students, at both MOY and EOY, the vast majority of students who were classified as at risk on the English language version were also classified as at risk on the Spanish

language version (see table 20). Eighty-eight percent of students at risk in English at EOY were also classified as at risk when assessed in Spanish, suggesting that there is a consistent set of students who may experience reading difficulties that cannot be attributed to the language acquisition process alone. At the same time, however, about half of students who met the benchmark in Spanish did not meet it in English.

Table 20. Number and percentage of English learner students who met and did not meet EOY benchmarks on English and Spanish language versions of assessments

End of Year	Did Not Meet Benchmark (Spanish)	Met Benchmark (Spanish)
Did Not Meet Benchmark (English)	88% (n = 666)	12% (n = 87)
Met Benchmark (English)	56% (n = 105)	44% (n = 81)

Source: 2020/21 and 2021/22 district-provided screening assessment data
Note: Of the 939 students, seven students are recorded as speaking Portuguese as their native language and one student is recorded as speaking Creole (Haitian) as their native language.

How does performance and benchmark attainment vary for English learners by specific domains of English proficiency?

The ACCESS assessment measures students' speaking, listening, reading, and writing skills in English and, as described earlier, provides scores that combine information across these skills. Research about literacy development among English learners has found that oral language development can be an important scaffold to developing literacy (see, for example, Kieffer, 2012; Gutierrez et al., 2010).

Key Findings

- At least half of English learner students who met benchmark at EOY had reading, speaking, and writing skills at the three lowest ACCESS levels.
- Only a quarter of English learner students meeting benchmark scored in the bottom three levels of the listening domain, which is much smaller than the reading, writing, and speaking domains; 53 percent meeting benchmark scored in the highest level on the listening domain.

Screening assessment data suggest that English learner students can meet EOY screening assessment benchmarks with basic writing, speaking, and reading skills (see table 21). At least half of students who met benchmarks had reading, speaking, and writing skills at the three lowest levels on ACCESS. However, few students with less developed listening skills met screening assessment benchmarks. This pattern may reflect emphasis of screening assessments on prereading skills, such as identifying letter names and sounds or recognizing common words, rather than reading

longer texts or comprehension or providing longer oral or written responses. Figure E.1 in appendix E provides additional detail on ACCESS composite scores combining domains and screening assessment performance.

Table 21. Percentage of English learner students who met and did not meet benchmarks at EOY by ACCESS domain performance levels

Domain and Performance	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Reading: Met Benchmark	29%	13%	16%	13%	18%	11%
Reading: Did Not Meet Benchmark	47%	27%	15%	7%	4%	<1%
Listening: Met Benchmark	7%	6%	12%	7%	15%	53%
Listening: Did Not Meet Benchmark	24%	12%	18%	7%	15%	24%
Speaking: Met Benchmark	10%	30%	29%	13%	5%	14%
Speaking: Did Not Meet Benchmark	29%	33%	21%	8%	3%	6%
Writing: Met Benchmark	26%	26%	39%	9%	–	–
Writing: Did Not Meet Benchmark	58%	20%	20%	1%	–	–

Source: 2020/21 and 2021/22 district-provided screening assessment data and state-provided ACCESS data
Note: Cells with sizes less than 10 are suppressed (shown with –) to protect student privacy.

How does performance and benchmark attainment vary for English learners by native language and English Language Education program?

As mentioned in the beginning of this section, Spanish is by far the most common native language in the sample of English learner students (see table 17). Over three quarters of English learner students with Spanish as a native language do not meet benchmark expectations at EOY (81% when data is aggregated across all three time periods). After native Portuguese-speaking English learners, native Spanish-speaking English learners have the lowest rate of achieving benchmark expectations when compared to peers with other native languages (see table 22). Chinese, Russian, and Albanian native speakers have the highest rates of benchmark attainment, with about 65 percent of Chinese and Russian speakers and about half of Albanian

speakers meeting expectations at EOY. Most notably, more Chinese native speakers met benchmark at EOY and overall than non–English learner students in the sample.

However, this pattern may be more related to other student characteristics than native language. For example, there is significant variation among these groups of English learner students in terms of the percentage classified as economically disadvantaged. Spanish and Portuguese native speakers, while having the highest percentage of students not meeting expectations throughout the school year, also have two of the highest rates of economically disadvantaged students. Conversely, Chinese and Russian native speakers have two of the lowest rates of economically disadvantaged students. Students' racial and ethnic backgrounds may also intersect with native language, educational experiences, and screening results (Dabach, 2014; Umansky et al., 2020).

Key Findings

- Fewer native Portuguese speakers and native Spanish speakers met benchmark compared to Chinese, Russian, and Albanian native speakers.
- More Chinese native speakers met benchmark at EOY than non–English learner students.
- Differences in performance by native language may be due to other student characteristics such as socioeconomic status.
- Ninety-five percent of English learner students were enrolled in sheltered English immersion programs.
- About the same percentages of students were at risk at the end of the year in dual language and sheltered English immersion programs (67% compared to 70%).
- Eighty percent of English learner students enrolled in other bilingual programs did not meet benchmark at EOY compared to only 54 percent of EL students not participating in services.
- Differences in performance by ELE program may be related to other student characteristics (e.g., fewer English learner students whose parent/guardian opted them out of ELE programs were economically disadvantaged than those enrolled in other programs).

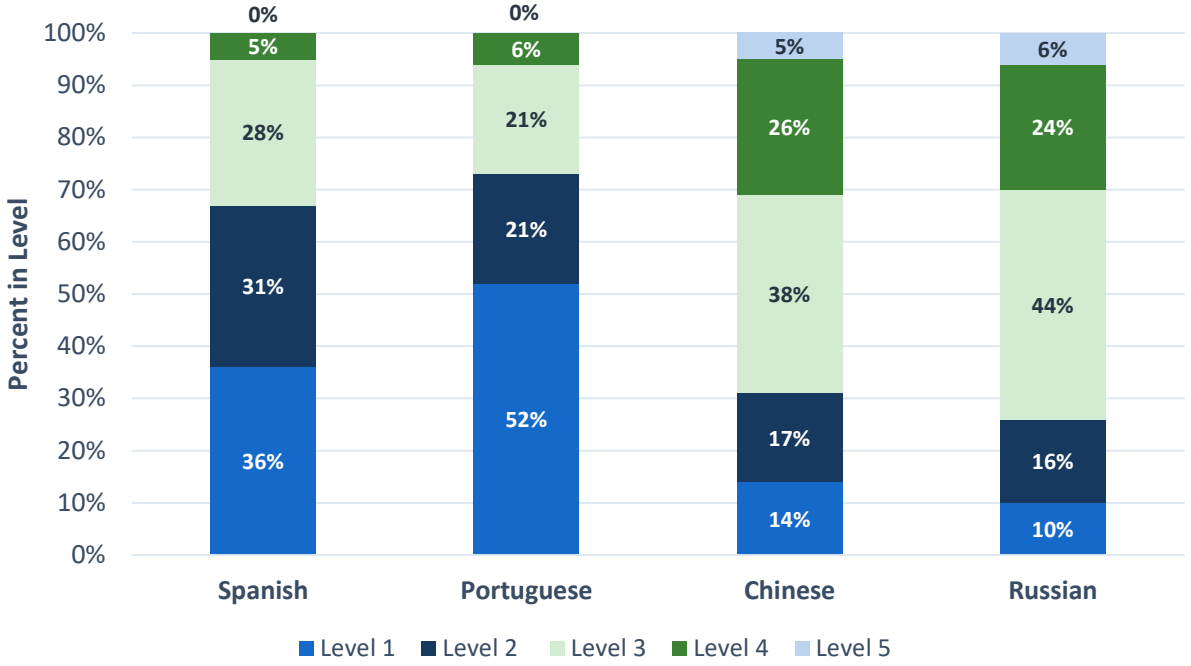
Table 22. Percentage of students not meeting benchmarks by time period and overall by native language

Language Spoken and Ranking by Enrollment Number (Sample)	% Not Meeting BOY	% Not Meeting MOY	% Not Meeting EOY	% Not Meeting Overall	% Economically Disadvantaged
1. Spanish	86.5%	82.0%	76.4%	80.6%	80.9%
2. Portuguese	87.3%	86.3%	78.1%	83.3%	78.6%
3. Chinese	51.7%	47.8%	34.7%	44.1%	46.2%
4. Arabic	75.2%	62.8%	50.9%	62.9%	78.7%
5. Vietnamese	70.0%	61.7%	52.1%	60.1%	53.7%
6. Creole (Haitian)	83.7%	85.4%	70.8%	79.4%	76.5%
7. Twi	86.2%	64.0%	54.3%	66.6%	76.6%
8. Russian	77.5%	59.8%	37.8%	58.9%	34.8%
9. Albanian	64.4%	55.6%	50.6%	56.2%	55.4%
10. Khmer	76.9%	79.5%	66.2%	73.5%	53.7%
Non-English Learner Students	54.0%	45.0%	36.0%	45.0%	38.0%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: Findings are restricted to the top 10 native languages most frequently cited among the sample. The percentage of students classified as economically disadvantaged for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state developed a different measure (i.e., the "low-income" metric) for use in the 2021/22 school year, but it was unavailable for the 2020/21 school year.

A similar pattern can be observed when comparing the overall ACCESS proficiency level scores of these groups of English learner students (see figure 14). More native speakers of Spanish and Portugese score in the bottom two proficiency levels of ACCESS compared to Chinese- and Russian-speaking English learners. As shown in figure 14, while students across ACCESS proficiency levels achieve literacy benchmarks, more students at lower levels are classified as at risk than students at higher levels.

Figure 14. ACCESS performance varies by native language



Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Nearly all English learner students (about 95%) in the screening assessment data sample were enrolled in sheltered English immersion programs. There were some differences between benchmark attainment of these students and students in other programs, although numbers of students in other types of programs are small. About the same percentages of students were at risk at the end of the year in dual language and sheltered English immersion programs (67% compared to 70%; see table 23). English learner students enrolled in other bilingual programs had the highest percentage of students not meeting benchmark at the end of the school year (80%), while students not participating in services had the smallest percentage of students at risk (54%).

As with native language, however, some of these differences may also be related to other student characteristics as well as to the design and goals of these programs. For example, fewer English learner students whose parent/guardian opted them out of ELE programs were economically disadvantaged than those enrolled in other programs. Similarly, other bilingual and dual language programs may not emphasize English reading skills as much as sheltered English immersion programs. For example, multiple studies that have investigated the relative effects of bilingual and English immersion programs found that there is a small to moderate effect of bilingual instruction on English literacy outcomes (Cheung & Slavin, 2012; Slavin et al., 2011). However, findings suggest that the advantages associated with bilingual education take several years to emerge (Umansky et al., 2020), likely due to the initial focus placed on home language literacy as a foundation for the teaching and development of English literacy (Umansky & Reardon, 2014).

Table 23. Percentage of English learner students not meeting benchmarks by type of English language services received

Program Type	% Not Meeting EOY	% Economically Disadvantaged
Sheltered English Immersion	70%	73%
Dual Language Education	67%	79%
Other Bilingual Program	80%	95%
Parent/Guardian Opted Out of All ELE Programs	54%	61%
Transitional Bilingual Education	--	--

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: The percentage of students classified as economically disadvantaged for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Cells with sizes less than 10 are suppressed to protect student privacy.

Discussion of Results

Although the sample of English learner students in the screening assessment data has some differences from the overall English learner population in the state (notably, more students from lower grades, more students in their first year in the United States, and more students at lower levels of English proficiency), analysis of benchmark performance appears to provide at least some support for prior research on the utility of literacy screening assessments with English learners.

Overall, English learners are more likely than their non–English learner peers to not meet expected proficiency levels on their early literacy screening assessments during the school year. However, among English learner students, students at both higher and lower levels of English proficiency do meet literacy benchmarks—i.e., students with low levels of English proficiency are not always classified as at risk. The fact that students who were classified as at risk in English were most often also classified as at risk based on Spanish-language assessments (for English learner students with a native language of Spanish) also provides some evidence that screening assessments are identifying students who may need additional support, while potentially still over-identifying English learner students overall. DESE has recently released guidance to assist educators in screening multilingual learners and interpreting multilingual learners' screening results in a culturally and linguistically sustaining manner. This guidance includes recommendations to screen in the native language where possible; consider screening results alongside other sources of data, including family input and educator observations of progress; and consider students' progress in comparison to peers with a similar linguistic background.

There were differences in benchmark attainment for English learner students with different native languages and instructional programs; however, these differences may also be related to other background characteristics such as economic disadvantage or English proficiency rather than native language or program design alone.

Future analyses with additional data for English learner students over time may be able to provide additional insight into the relationship of screening assessment performance and English proficiency.

Using MCAS to Compare Assessment Benchmarks and Predictive Power

One of the challenges of the screening assessment data being collected is that it includes data from assessments that were designed differently and use different approaches to determine whether students are at risk or on track. However, there is one common assessment that all students in grade 3 in the state do take: MCAS. The MCAS English Language Arts (ELA) assessment therefore provides a mechanism to look across screening assessments using a single yardstick.

In this section, we use Massachusetts public school student test scores from 2021/22 to link benchmark cut scores from grade 3 literacy screening assessment EOY results to MCAS Grade 3 ELA scale scores. Table 24 shows the specific research questions to be addressed in this section.

Table 24. Research questions related to comparing screening assessment benchmarks and relationship to MCAS

Topic	Question
Comparing screening assessment benchmarks and predicting later outcomes with screening assessment data	• How do literacy benchmark scores from different screening assessments compare to each other in Massachusetts Comprehensive Assessment System (MCAS) score terms? • What is the relationship between grade 3 literacy screening assessments and MCAS performance?

As noted, the purpose of this analysis is to facilitate interpretation of the different literacy screening assessment benchmark performance levels by comparing their relative stringency to proficient reading performance as measured by the MCAS Grade 3 ELA assessment. It is hoped that a clearer understanding of the alignment between screening assessment benchmarks and MCAS Grade 3 ELA cut scores will help educators better understand where their students are in their growth toward becoming proficient readers. We use two different analytic approaches: equipercentile linking and predictive linking, which address the general research questions

above but also provide somewhat different information about how screening assessments compare.

Both approaches use a single-group design, in which examinees take both assessments that are to be linked. An advantage of the single-group design is that it controls for differences in examinee proficiency (Brennan, 2006). In addition, to minimize the effects of distance between the literacy screening assessment and MCAS Grade 3 ELA administrations, we use the grade 3 EOY early literacy screening assessment to create the link. The closer in time the two test administrations are, the less error is associated with each linkage. As data become available for more students in the future, we will be able to expand the scope of analyses as well as verify and increase the precision of the links to the MCAS scale.

It is important to note that tests that are linked are not equated. For two tests to be equated, they must be written to the same set of specifications, because equating adjusts for differences in difficulty, not for differences in content (Kolen & Brennan, 2004). Literacy screening assessments have different specifications, different score scales and distributions, and varying degrees of relationship with the MCAS ELA assessment. It should be understood when interpreting the results that the linked tests are considered concordant rather than equated, and the linked scores are not interchangeable (Pommerich et al., 2000). Concordance provides a substantially weaker form of linkage than equating. The linkages in this section can be interpreted as, for example, students who score 203 or higher on the MAP Growth screening assessment will most likely score 500 or higher on the MCAS ELA assessment.

The equipercentile linking procedure (Kolen & Brennan, 2004) is a statistical method that links scores from tests *X* and *Y* that have the same percentile rank within the linking sample or samples. Concordant cut scores in this study, therefore, are comparable in terms of the passing percentage on either test within the linking sample. This procedure will be used to answer the following research questions:

- How do literacy screening assessment benchmark cut scores that classify students as having high, moderate, and negligible risk of reading difficulties align with MCAS Grade 3 ELA cut scores that classify students as Not Meeting, Partially Meeting, Meeting, and Exceeding Expectations?
- How do literacy screening assessment benchmark cut scores vary in their alignment to the MCAS scale?

The equipercentile method depends entirely on the data in the sample; therefore, the accuracy of the linkages depends on the representativeness of the sample test score distribution with respect to the population test score distribution. According to Ho (2012), because cut scores linked using the equipercentile method have the same passing percentage on either scale, the relationship is based only on the stringency of the cut score, which means the benchmark cut score corresponds to the linked MCAS ELA scale score.

Predictive linking uses regression-based methods to link scales from different tests by using a focal test to predict scores on an outcome test (Thng, 2019). These methods are commonly used during evidence-based standard-setting procedures to provide information to panelists by linking performance standards to a future outcome test (McCarty et al., 2012). In this case, the focal tests are the early literacy screening assessment benchmarks, and the outcome test is the MCAS ELA assessment. This procedure will be used to answer the following research questions:

- How much do predictive cut scores vary when compared to equipercentile-linked cut scores?
- How well do literacy screening assessments discriminate between students who are and are not on track to meeting proficiency standards on the MCAS ELA assessment?
- How well do different literacy screening assessment benchmark cut scores discriminate between students who are and are not on track to meeting proficiency standards on the MCAS ELA assessment?

The accuracy of predictive linking depends on the stringency of the cut score and the association (i.e., correlation) between the focal and outcome tests (Ho, 2012). In other words, the benchmark cut score predicts the linked MCAS ELA scale score, and the higher the correlation between literacy screening assessment and MCAS ELA assessments, the more accurate its predictive link will be. Classification accuracy and Receiver Operating Characteristic (ROC) analyses are used to measure how well the literacy screening assessments discriminate between students who are and are not on track to meeting proficiency expectations.

Preliminary results of equipercentile and predictive linking studies suggest the following:

- Benchmark cut scores used to classify students who are at high risk of reading difficulties link to scores near the threshold between the Not Meeting and Partially Meeting Expectations performance levels.
- Benchmark cut scores used to identify students who are on track to meeting MCAS ELA proficiency standards link to MCAS ELA scores that range from the middle of the Partially Meeting Expectations performance level to the beginning of the Meeting Expectations performance level.

Data Used in Equipercentile and Predictive Linking Analysis

The 2021/22 grade 3 EOY literacy screening assessment concordance samples include 4,066 students from 63 schools within 28 districts and test scores from seven early literacy screening assessments (table 25). Additional samples from FastBridge CBMreading (25), Star Early Literacy (93), and Star Early Literacy Spanish (34) 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 25 were considered sufficiently large and representative to produce reasonable preliminary estimates. That said, there are some limitations to the representativeness of the samples that are worth noting.

Key Findings

- The screening assessment concordance sample included 4,066 grade 3 students across seven screening assessments in 2021/22.
- Low-income students, English learners, and Hispanic students are overrepresented in the sample compared to the state, while White students are underrepresented.
- Linking 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.

Table 25. Grade 3 EOY literacy screening assessment data collected during the 2021/22 school year

Early Literacy Screening Assessment	Districts	Schools	n
DIBELS 8th Edition	9	20	920
FastBridge aReading	2	3	180
i-Ready Diagnostic	2	6	707
Lexia RAPID	2	5	543
MAP Growth	4	11	624
mCLASS	4	6	429
Star Reading	5	12	663
Total	28	63	4,066

Source: 2021/22 district-provided screening assessment data

The screening assessment data were collected by convenience in that screening assessments were chosen by districts that participated in a grant intended to improve early literacy as opposed to being randomly assigned for research purposes. Further, the samples at this stage of the project are smaller than samples that would be recommended to conduct a stronger form of linkage, such as an equating study. With small samples come challenges that accompany sparse data such as gaps in number distributions, especially near the tails of the distribution. Equipercentile linking is dependent on the data in the sample, which is a potential source of bias when samples are small or unrepresentative. 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 increase and linking estimates are expected to improve as the project develops over time.

For this study, the sampled literacy screening assessment scale score distributions are expected 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 2021/22 school year shows (table 26):

- Hispanic students are overrepresented in the concordance sample while White students are underrepresented.
- Low-income students are overrepresented in the concordance sample.
- English learners are overrepresented in the concordance sample.

Table 26. Student characteristics for the state of Massachusetts and grade 3 EOY literacy screening assessment concordance sample

Student Group	State (n = 67,718)		Concordance Sample (n = 4,066)	
	Number	%	Number	%
Female	33,157	49.0	1,991	49.0
Male	34,535	51.0	2,074	51.0
Asian/Native American/ Pacific Islander	5,515	8.1	404	9.9
Black	6,223	9.2	288	7.1
Hispanic	16,017	23.7	1,301	32.0
Multiracial	3,145	4.6	155	3.8
White	36,809	54.4	1,918	47.2
Low Income: No	37,312	55.1	1,824	44.9
Low Income: Yes	30,406	44.9	2,242	55.1
English Learner: No	56,880	84.0	3,203	78.8
English Learner: Yes	10,838	16.0	863	21.2
Student With IEP: No	54,164	80.0	3,260	80.2
Student With IEP: Yes	13,554	20.0	806	19.8

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data
Note: Students can identify as nonbinary, but samples were not large enough to include in these tables. The current "low-income" metric used in the state was used rather than the "economically disadvantaged" metric used previously in the analysis as only 2021/22 scores were examined.

Student demographic characteristics for the literacy screening assessment samples vary (see tables D.1.1 and D.1.2 in appendix D); therefore, the demographic differences seen in table 26 do not necessarily represent the demographic characteristics for any specific literacy screening assessment sample. For example, in contrast with the overall concordance sample, White students are overrepresented in the FastBridge and i-Ready concordance samples, and Asian students are overrepresented in the MAP Growth sample (see tableD.1.2 in appendix D).

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 27 and figures F.1 through F.8).

Table 27. 2021/22 MCAS Grade 3 ELA scale score mean, median, 5th, and 95th percentiles for the state and grade 3 literacy screening assessment concordance samples

Sample	n	M	SD	5th %ile	Median	95th %ile	Minimum	Maximum
State	63,639	496.50	22.93	457	497	533	440	560
DIBELS 8th Edition	920	489.67	21.95	452	490	524	443	559
FastBridge	180	492.61	20.11	454	494	524	444	544
i-Ready	707	500.14	19.84	466	499	533	444	560
Lexia RAPID	543	489.66	22.29	457	488	524	444	560
MAP Growth	624	496.58	22.32	457	497	533	444	560
mCLASS	429	497.12	21.23	463	497	533	444	560
Star Reading	663	483.14	21.06	447	484	516	443	551

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

As shown in table 27 and figure F.1, when compared with the state in the middle of the distribution:

- Mean and median MCAS ELA scores are similar for the MAP Growth and mCLASS samples.
- Mean and median MCAS ELA scores are slightly higher (<5 points) for the i-Ready sample.
- Mean and median MCAS ELA scores are slightly lower (<10 points) for the DIBELS 8th Edition, FastBridge, and Lexia RAPID samples.
- Mean and median MCAS ELA scores are more than 10 points lower for the Star Reading sample.

When compared with the state at the tails of the distribution:

- Sample distributions appear to have slightly less coverage at the extremes of the scale score distribution.
- The DIBELS 8th Edition, FastBridge, and Star Reading samples are missing scale scores at the top of the scale score range.
- All concordance samples are missing scores at the bottom of the scale score range.

Taken together, results in table 27 and figure F.1 suggest linking 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. These estimates and the methods used to produce them are discussed next.

How do literacy screening assessment benchmark scores from different screening assessments compare to each other in MCAS terms?

Key Findings

- Benchmark cut scores used to identify students who are on track to meeting MCAS ELA proficiency standards link to MCAS ELA scores that range from the middle of the Partially Meeting Expectations performance level to the beginning of the Meeting Expectations performance level.
- Benchmark cut scores used to classify students who are at high risk of reading difficulties link to scores near the threshold between the Not Meeting and Partially Meeting Expectations performance levels.

This section describes equipercentile and predictive linking methods used to link literacy screening assessment scale scores to MCAS ELA scale scores and the associated findings. The equipercentile linking section aligns literacy screening assessment benchmark cut scores that classify students who are at high risk, moderate risk, and low risk of reading difficulties with the MCAS ELA scale, enabling benchmark cut scores to be directly compared. The predictive linking section enables literacy screening assessment benchmark cut scores to be directly compared using a different linking method and examines how well the different screening assessments discriminate between students who are and are not on track to meet MCAS ELA proficiency goals.

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. At a high level, the equipercentile linking procedure 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 on the MCAS Grade 3 ELA scale was found 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 28 and figure 15. As a final check on the reasonability of the equipercentile linking estimates, we linked literacy screening assessment cut scores to the MCAS scale by district to examine the variability of the estimates (see appendix F). The estimates presented in tables F.1–F.7 are similar to the overall estimates, with variability in expected ranges, and support the reasonableness of the results shown in table 28 and figures 15 and 16. It should be noted, however, that the Star Reading national percentile is more than 20 points lower than sample estimates at each benchmark cut score. Those results imply that the students who took Star tended to perform at the lower end of the scale, which could mean that the true linked scores are higher than the estimates shown in table 28.

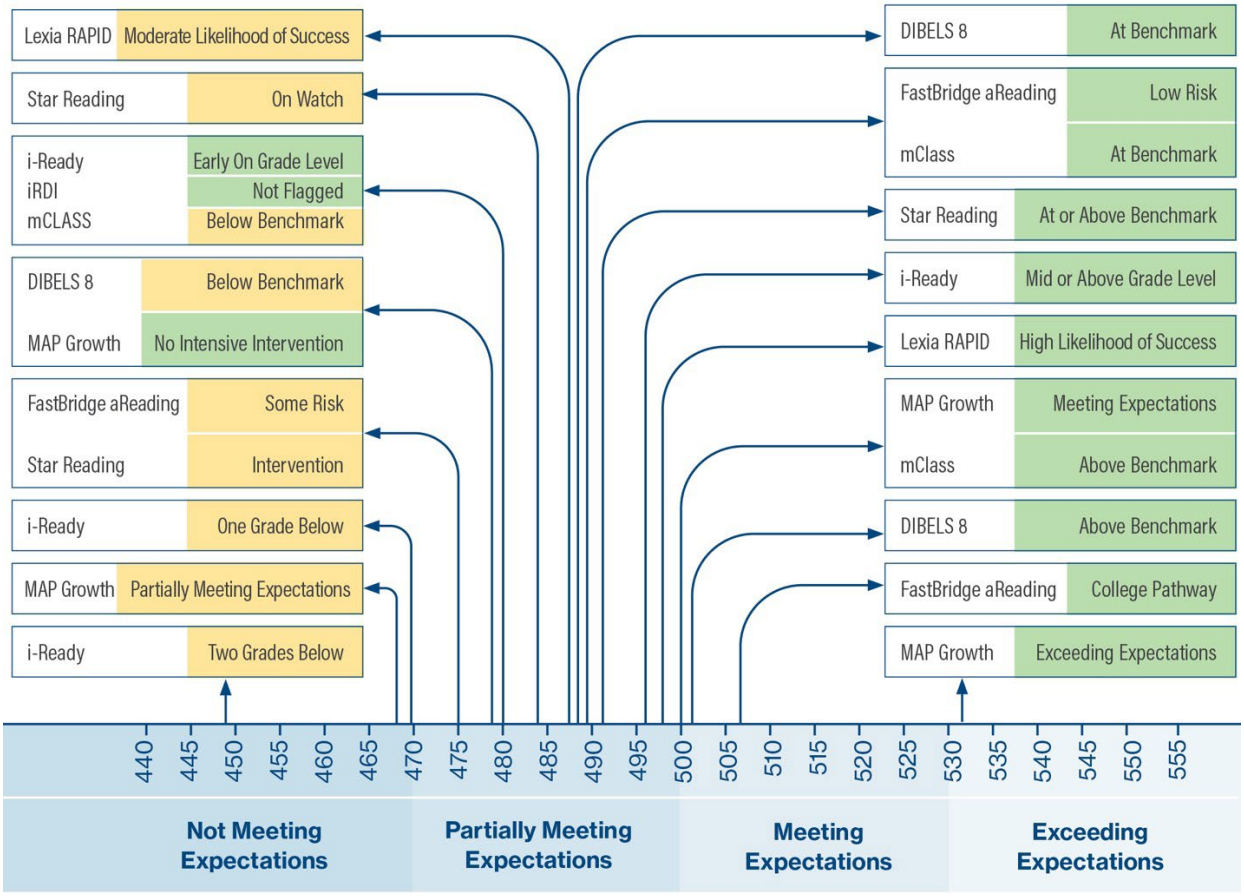
Table 28. Grade 3 literacy screening assessment benchmark cut scores linked to MCAS scale scores and performance levels using equipercentile linking

Early Literacy Screening Assessment	Cut Score	Benchmark	Sample %ile Rank	National %tile Rank	MCAS Scale Score	95% CI LL	95% CI UL	MCAS Performance Level
DIBELS 8th Edition	424	Below Benchmark	33	29	479	477	481	Partially Meeting
DIBELS 8th Edition	442	At Benchmark	48	45	488	486	492	Partially Meeting
DIBELS 8th Edition	467	Above Benchmark	70	65	501	499	503	Meeting
FastBridge	483	Some Risk	21	15	475	472	478	Partially Meeting
FastBridge	503	Low Risk	37	40	489	484	494	Partially Meeting
FastBridge	517	College Pathway	76	70	506	503	509	Meeting
i-Ready	417	Two Grades Below	1	5	449	438	460	Not Meeting
i-Ready	474	One Grade Below	7	15	470	466	474	Partially Meeting
i-Ready	511	Early on Grade Level/iRDI Not Flagged	16	31	480	478	482	Partially Meeting
i-Ready	545	Mid or Above Grade Level	42	59	496	494	498	Partially Meeting
Lexia RAPID	31	Moderate Likelihood of Success	47	31	487	485	489	Partially Meeting
Lexia RAPID	70	High Likelihood of Success	65	70	498	496	500	Partially Meeting
MAP Growth	180	Partially Meeting	12	15	468	466	470	Not Meeting
MAP Growth	189	No Intensive Intervention	23	31	479	476	482	Partially Meeting
MAP Growth	203	Meeting	56	64	500	498	502	Meeting

Early Literacy Screening Assessment	Cut Score	Benchmark	Sample %ile Rank	National %tile Rank	MCAS Scale Score	95% CI LL	95% CI UL	MCAS Performance Level
MAP Growth	222	Exceeding	95	94	531	529	533	Exceeding
mCLASS	424	Below Benchmark	20	21	480	477	483	Partially Meeting
mCLASS	442	At Benchmark	35	41	489	486	492	Partially Meeting
mCLASS	467	Above Benchmark	53	64	500	498	502	Meeting
Star Reading	902	Intervention	38	10	475	473	477	Partially Meeting
Star Reading	943	On Watch	52	23	484	482	486	Partially Meeting
Star Reading	970	At or Above Benchmark	65	40	491	489	493	Partially Meeting

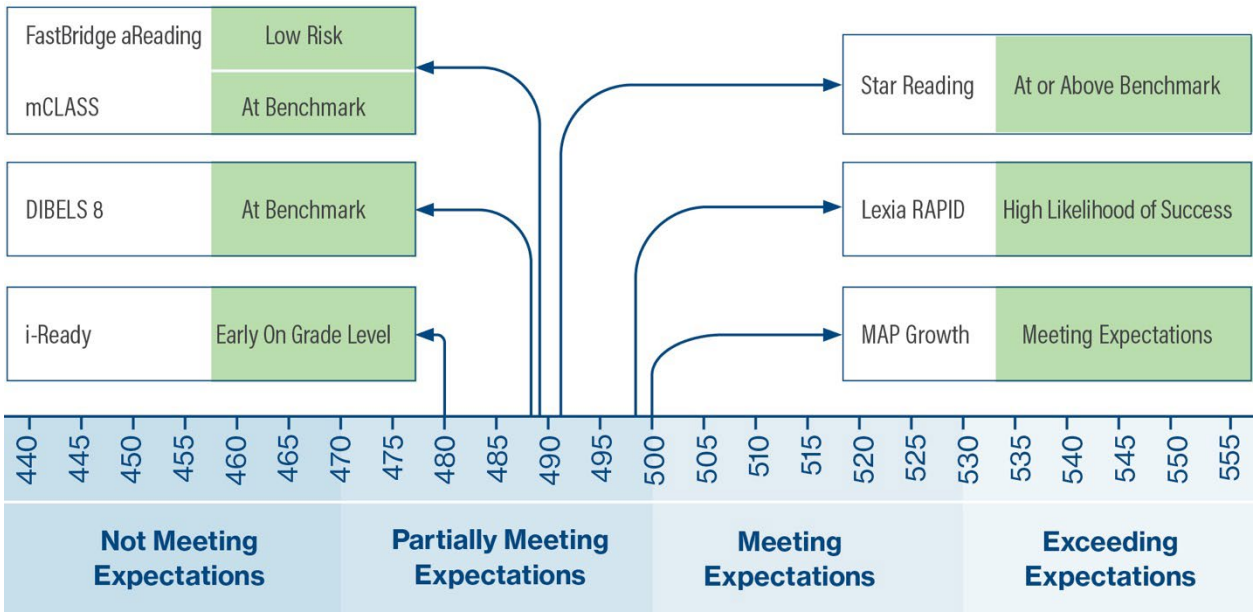
Source: 2021/22 district-provided screening assessment data and state-provided MCAS data
Note: i-Ready has a Reading Difficulty Indicator (iRDI) cut score of 507 that is very close to the Early on Grade benchmark cut score of 511 and produces the same linking results. The analyses for MAP Growth include the benchmarks developed by the linking study (Not Meeting, Partially Meeting, Meeting, Exceeding) conducted by the assessment vendor that links MAP Growth cut scores to the MCAS ELA assessment. These benchmarks were not used in previous sections since they are only available for grade 3 students.

Figure 15. Grade 3 literacy screening assessment benchmark cut scores vary in how they map to MCAS ELA scores using equipercentile linking



Source 2021/22 district-provided screening assessment data and state-provided MCAS data
Note: Benchmarks highlighted in green indicate levels that students must meet/exceed to be classified as meeting benchmark. Benchmarks highlighted in yellow indicate levels that will result in a student being classified as at risk. Students scoring below the yellow benchmarks for Lexia RAPID, mCLASS, DIBELS 8, FastBridge aReading, and MAP Growth are classified as being at significant risk according to the respective assessment; students scoring below the "Star Reading Intervention" benchmark or i-Ready's "iRDI Not Flagged" benchmark are classified as being at significant risk according to the respective assessment.

Figure 16. Grade 3 literacy screening assessment scores that classify students as at any risk or not at any risk link to scores that range from the middle of the MCAS ELA Partially Meeting Expectations performance level to the beginning of the Meeting Expectations performance level



Source: 2021/22 district-provided screening assessment data and state-provided MCAS data
Note: Benchmarks indicate the levels that students must meet/exceed on the respective assessment to be classified as meeting benchmark.

Benchmark cut scores to classify students who are on track to meeting proficiency expectations include:

- DIBELS 8th Edition At Benchmark
- FastBridge Low Risk
- i-Ready Early on Grade Level
- Lexia RAPID High Likelihood of Success
- MAP Growth Meeting Expectations
- mCLASS At Benchmark
- Star Reading At or Above Benchmark

These benchmarks link to scores that range from the middle of the MCAS ELA Partially Meeting Expectations performance level to the beginning of the Meeting Expectations performance level. The 95 percent confidence interval estimates of MCAS ELA scores linked to Lexia RAPID and MAP Growth include the MCAS cut score for Meeting Expectations, suggesting the true linked score includes the MCAS ELA Meeting Expectations cut score. The MAP Growth Meeting Expectations aligns to the MCAS Meeting Expectations scale score by design, as the MAP benchmark cut scores shown in table 28 and figure 15 are the result of an equipercentile linking study by NWEA (2020) to link MAP Growth scores to the MCAS ELA scale. The results of this study provide support for that linking study, as the 95 percent confidence interval estimates for each of the MAP Growth linked scores include the corresponding MCAS ELA cut scores for Partially Meeting, Meeting, and Exceeding Expectations (see table 28).

Key Findings

- Correlations between literacy and MCAS ELA assessments are in the strong to very strong ranges.
- Equipercentile linking estimates (in previous section) and predictive linking estimates are similar and typically predict scores within the same MCAS ELA performance level.
- Predictive linking estimates near the middle of the scale can be viewed with more confidence than those nearer the extremes.
- Classification accuracy based on OLS predictions are in the good range for all screening assessments; FastBridge aReading, i-Ready, MAP Growth, and Star Reading meet the NCII highest-rated level for classification accuracy.
- AUC estimates for all screening assessments are in the good to excellent range.

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

In this section, we describe the ordinary least squares (OLS) regression method used to link literacy screening assessment composite scores to the MCAS scale and predict student performance on MCAS ELA. The

accuracy with which predictions discriminate between students who did and did not meet proficiency standards are examined. In addition, logistic regression is used to create Receiver Operating Characteristic (ROC) plots, which are used to examine the accuracy with which each screening assessment's benchmark cut scores discriminate between students who met and did not meet proficiency standards on the 2021/22 MCAS Grade 3 ELA test.

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 29 shows correlations between literacy and MCAS ELA assessments are in the strong (0.60–0.79) to very strong (0.80–1.00) ranges.

Table 29. Correlation between MCAS Grade 3 ELA and grade 3 EOY literacy screening assessment scale scores

Early Literacy Screening Assessment	n	Composite SS: M	Composite SS: SD	MCAS ELA SS: M	MCAS ELA SS: SD	Correlation
DIBELS 8th Edition	920	442.01	40.99	489.67	21.95	0.67
FastBridge	180	501.77	27.26	492.61	20.11	0.72
i-Ready	707	546.31	44.66	500.14	19.84	0.82
Lexia RAPID	543	46.09	37.84	489.66	22.29	0.72
MAP Growth	624	199.62	15.19	496.58	22.32	0.85
mCLASS	429	450.07	30.76	497.12	21.23	0.64
Star Reading	663	919.52	102.90	483.14	21.06	0.79

Source: 2021/22 district-provided screening assessment data and state-provided MCAS data

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 (see table 29).

OLS Regression and Predictive Linking Results

We created a predictive link from literacy screening assessments to the MCAS ELA scale using OLS regression Equation 3.

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

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 (see table 30).

Table 30. Literacy screening assessment benchmark cut scores linked to MCAS ELA scale scores and performance levels using predictive linking

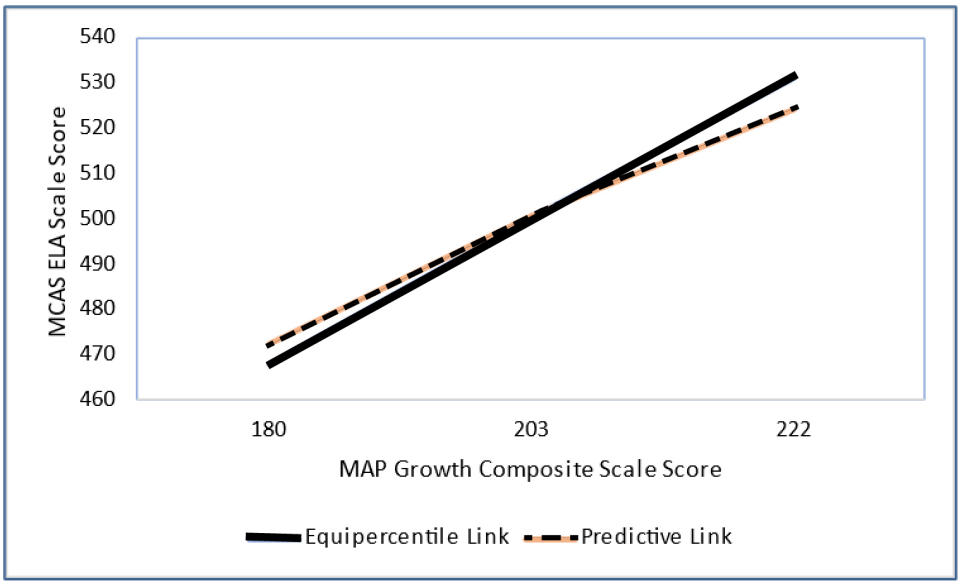
Early Literacy Screening Assessment	Cut Score	Benchmark	MCAS Scale Score	95% CI LL	95% CI UL	MCAS Performance Level
DIBELS 8th Edition	424	Below Benchmark	483	451	515	Partially Meeting
DIBELS 8th Edition	442	At Benchmark	490	458	522	Partially Meeting
DIBELS 8th Edition	467	Above Benchmark	499	467	531	Partially Meeting
FastBridge	483	Some Risk	483	455	510	Partially Meeting
FastBridge	503	Low Risk	493	466	521	Partially Meeting
FastBridge	517	College Pathway	501	473	528	Meeting
i-Ready	417	Two Grades Below	453	431	476	Not Meeting
i-Ready	474	One Grade Below	474	451	497	Partially Meeting
i-Ready	511	Early on Grade Level/iRDI Not Flagged	487	465	510	Partially Meeting
i-Ready	545	Mid or Above Grade Level	500	477	522	Partially Meeting
Lexia RAPID	31	Moderate Likelihood of Success	483	453	513	Partially Meeting
Lexia RAPID	70	High Likelihood of Success	500	470	530	Meeting
MAP Growth	180	Partially Meeting	472	449	496	Partially Meeting
MAP Growth	189	No Intensive Intervention	483	460	507	Partially Meeting
MAP Growth	203	Meeting	501	477	524	Meeting
MAP Growth	222	Exceeding	524	501	548	Meeting
mCLASS	424	Below Benchmark	486	454	518	Partially Meeting
mCLASS	442	At Benchmark	494	462	526	Partially Meeting
mCLASS	467	Above Benchmark	505	473	537	Partially Meeting

Early Literacy Screening Assessment	Cut Score	Benchmark	MCAS Scale Score	95% CI LL	95% CI UL	MCAS Performance Level
Star Reading	902	Intervention	480	455	506	Partially Meeting
Star Reading	943	On Watch	487	461	512	Partially Meeting
Star Reading	970	At or Above Benchmark	491	466	517	Partially Meeting

Source: 2021/22 district-provided screening assessment data and state-provided MCAS data
Note: i-Ready has a Reading Difficulty Indicator (iRDI) cut score of 507 that is very close to the Early on Grade benchmark cut score of 511 and produces the same linking results.

Comparing tables 28 and 30 shows 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. Estimates from the MAP Growth screening assessment provide an illustrative example (see figure 17).

Figure 17. Some differences in predictive and equipercentile estimates linking MAP Growth grade 3 EOY cut scores to MCAS grade 3 scale scores



Source: 2021/22 district-provided screening assessment data and state-provided MCAS data

The mean MCAS ELA scale score for the MAP Growth sample is 496.58 (see table 29). At the MAP Growth Partially Meeting Expectations cut score of 180 shown in figure 17, where the linked score is below the MCAS ELA mean, the predictive linking estimate is higher than the equipercentile estimate. At the Meeting Expectations score of 203, where the linked score is near the mean, the predictive linking estimate is very close to the equipercentile estimate, and at the Exceeding Expectations cut score of 222, where the linked score is above the mean, the predictive linking estimate is lower than the equipercentile estimate. 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. 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.

OLS Regression and Classification Accuracy Results

Using OLS regression Equation 3, each student's predicted MCAS ELA scale score was estimated and saved, and students with a rounded score greater than or equal to 500 were classified as meeting MCAS ELA proficiency standards. The accuracy with which those classifications discriminated between students who truly met or did not meet proficiency standards was then examined (see table 31).

Table 31. Accuracy with which literacy screening assessment OLS regression predictions discriminate between students who meet and do not meet proficiency standards on MCAS Grade 3 ELA by assessment

Early Literacy Screening Assessment	At or Above Cut Score n	Below Cut Score n	Classification Accuracy %	TP	FN	TN	FP
DIBELS 8th Edition	298	622	75.87	0.66	0.34	0.79	0.21
FastBridge	67	113	80.56	0.77	0.23	0.83	0.17
i-Ready	353	354	81.89	0.89	0.11	0.75	0.25
Lexia RAPID	171	372	80.84	0.68	0.32	0.88	0.12
MAP Growth	273	351	84.94	0.81	0.19	0.88	0.12
mCLASS	191	238	71.80	0.77	0.23	0.68	0.32
Star Reading	151	512	87.18	0.89	0.11	0.87	0.13

Source: 2021/22 district-provided screening assessment data and state-provided MCAS data
Note:
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

Classification accuracy based on OLS predictions ranged from ~72 percent to ~87 percent, which can be considered in the good range (70%–90%). The four additional columns in table 31 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 proficiency standards, 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. 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 help and don't get it.

These two errors are assigned different thresholds within the National Center on Intensive Intervention's (NCII, n.d.) tools charts, which 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 less than 0.30 and an FP rate less than 0.20 for the highest rated screening tools. FastBridge aReading, i-Ready, MAP Growth, and Star Reading meet the NCII highest rated criteria based on the results of this analysis.

Logistic Regression and ROC Analysis Results

Logistic regression was used to create ROC plots and conduct an ROC analysis, which was used to evaluate the diagnostic accuracy of the different benchmark cut scores. 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. Classification accuracy is used as a measure of performance in ROC analysis, as is another measure of association known as area under the curve (AUC). As the name indicates, AUC is the area under the curve in the ROC plots produced using Equation 4 (see figures F.10 through F.16 in appendix F). 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). 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 32).

Table 32. ROC analysis of accuracy with which 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	Met Proficiency Standards		Classification Accuracy %	TP	FN	TN	FP	AUC
		Yes	No						
DIBELS 8th Edition	Below Benchmark	285	277	61.09	0.96	0.04	0.45	0.55	0.822
DIBELS 8th Edition	At Benchmark	259	393	70.87	0.87	0.13	0.63	0.37	0.822
DIBELS 8th Edition	Above Benchmark	180	520	76.09	0.60	0.40	0.84	0.16	0.822
FastBridge	Some Risk	67	34	56.11	1.00	0.00	0.30	0.70	0.896
FastBridge	Low Risk	66	63	71.67	0.99	0.01	0.56	0.44	0.896
FastBridge	College Pathway	39	104	79.44	0.58	0.42	0.92	0.08	0.896
i-Ready	Two Grades Below	353	9	51.20	1.00	0.00	0.03	0.97	0.913
i-Ready	One Grade Below	352	52	57.14	1.00	0.00	0.15	0.85	0.913
i-Ready	Early on Grade Level/iRDI Not Flagged	351	110	65.21	0.99	0.01	0.31	0.69	0.913
i-Ready	Mid or Above Grade Level	316	261	81.61	0.90	0.10	0.74	0.26	0.913
Lexia RAPID	Moderate Likelihood of Success	160	239	73.48	0.94	0.06	0.64	0.36	0.881
Lexia RAPID	High Likelihood of Success	128	309	80.48	0.75	0.25	0.83	0.17	0.881
MAP Growth	Partially Meeting	273	69	54.81	1.00	0.00	0.20	0.80	0.931
MAP Growth	No Intensive Intervention	272	135	65.22	1.00	0.00	0.38	0.62	0.931
MAP Growth	Meeting	233	297	84.94	0.85	0.15	0.85	0.15	0.931
MAP Growth	Exceeding	34	351	61.70	0.12	0.88	1.00	0.00	0.931
mCLASS	Below Benchmark	186	79	61.77	0.97	0.03	0.33	0.67	0.792
mCLASS	At Benchmark	167	126	68.30	0.87	0.13	0.53	0.47	0.792
mCLASS	Above Benchmark	139	171	72.26	0.73	0.27	0.72	0.28	0.792
Star Reading	Intervention	149	247	59.73	0.99	0.01	0.48	0.52	0.929
Star Reading	On Watch	145	336	72.55	0.96	0.04	0.66	0.34	0.929
Star Reading	At or Above Benchmark	138	416	83.56	0.91	0.09	0.81	0.19	0.929

Source: 2021/22 district-provided screening assessment data and state-provided MCAS data
Note:
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
i-Ready has a Reading Difficulty Indicator (iRDI) cut score of 507 that is very close to the Early on Grade benchmark cut score of 511 and produces the same ROC analysis results.

Results of the ROC analysis indicate the Meeting Expectations benchmark cut score for MAP Growth and At or Above Benchmark cut score for Star Reading meet NCII (n.d.) criteria (see table 32). In addition, results in table 32 can be interpreted in terms of the proportion of students at or above each benchmark cut score who also meet MCAS ELA proficiency standards.

To help illustrate these interpretations, figure 18 compares the MCAS ELA performance of students who met on-track to proficiency standards with those who did not for each literacy screening assessment. Students with moderate to high risk of not meeting proficiency standards are all categorized as at risk in figure 19.

The first column under Met Proficiency Standards in table 32 shows 259 students scored at or above the DIBELS 8th Edition At Benchmark cut score and met proficiency standards on MCAS ELA. These students are depicted in the Meeting and Exceeding sections of the DIBELS 8th Edition bar in figure 19. The TP rate shows that these students represent 0.87 or 87 percent of the 298 students who met proficiency standards (see table 31). The FN column shows 13 percent (0.13) or 39 of 298 students scored below the DIBELS 8th Edition At Benchmark performance standard but still met proficiency standards on MCAS ELA. Those students are depicted in the Meeting and Exceeding sections of the DIBELS 8th Edition bar in figure 18.

Figure 18. Few students identified as at risk on grade 3 literacy screening assessments meet or exceed on MCAS ELA at grade 3

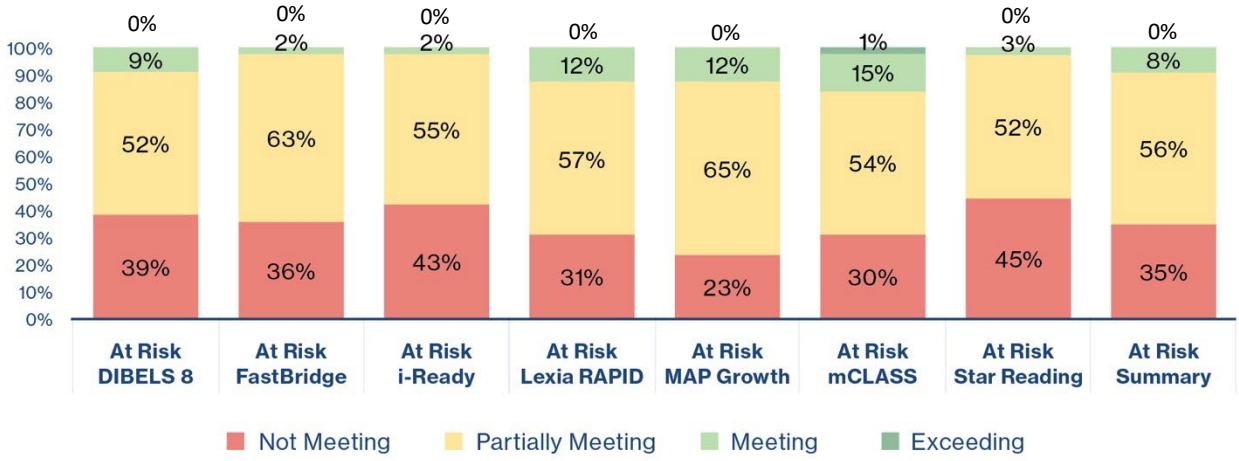


Figure 19. Meeting literacy screening benchmarks at grade 3 does not necessarily indicate MCAS proficiency

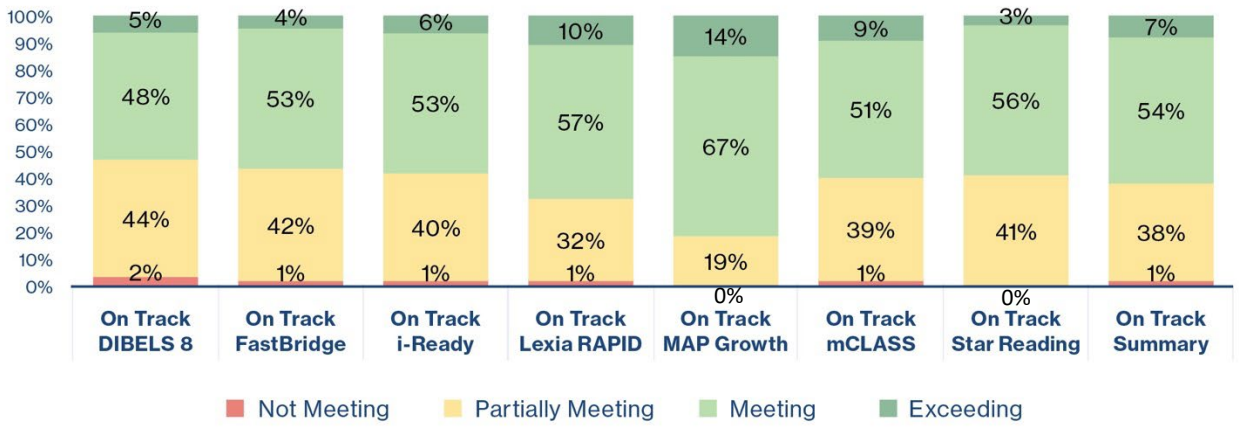


Table 32 shows 393 students scored below the DIBELS 8th Edition At Benchmark performance standard and below the MCAS ELA Meeting Expectations standard. The TN column shows these students represent 0.63 of the 622 students who performed below the MCAS ELA Meeting Expectations standard (see table 31). These students are depicted in the Not Meeting and Partially Meeting sections of the DIBELS 8th Edition bar in figure 18. Finally, the FP column shows that 0.37 or 229 students met the DIBELS 8th Edition At Benchmark standard and did not meet the MCAS ELA Meeting Expectations standard. These students are depicted in the Not Meeting and Partially Meeting sections of the DIBELS 8th Edition bar in figure 19.

The FN column in table 32 indicates that students who were classified below the on-track performance levels on the early literacy assessments during the spring of grade 3 were unlikely to meet MCAS ELA proficiency standards. For example, no students who scored below the FastBridge Some Risk threshold, the i-Ready One Grade Below threshold, or the MAP Growth Partially Meeting Expectations threshold met MCAS ELA proficiency standards. The FP column in table 32 also shows students who reached an early literacy on-track benchmark did not always meet MCAS ELA proficiency standards. The FP rates describing students who did not meet proficiency standards on MCAS ELA ranged from 15 percent to 69 percent across the on-track benchmarks (see table 32).

The FP rates of the Lexia RAPID, MAP Growth, and Star Reading screening assessments' on-track benchmark cut scores were under the NCII FP highest rating threshold of 0.20. The DIBELS 8th Edition, FastBridge, i-Ready, and mCLASS screening assessments also had benchmark cut scores under the NCII FP threshold, but those cut scores are more stringent than the cut scores used to indicate on track to proficiency in the study. For example, the DIBELS 8th Edition Above Benchmark cut score has an FP rate of 0.16, but the At Benchmark cut score used to designate on track has a higher FP rate of 0.37 (see table 32).

Discussion of Results

Measures of association between literacy and MCAS ELA assessments, including correlation, classification accuracy, and ROC AUC estimates, indicate strong to very strong association between the tests, suggesting early literacy screening assessments can be useful predictors of MCAS ELA performance. Preliminary results linking grade 3 EOY literacy screening assessment benchmark cut scores to the MCAS Grade 3 ELA scale show cut scores link to various regions of the MCAS scale depending on how risk of reading difficulty is defined. Across the screening assessments, benchmarks with similar definitions of risk link to similar regions of the MCAS scale.

For example, literacy screening assessment benchmark cut scores that identify high risk of reading difficulties link to MCAS ELA scale scores near the threshold between Not Meeting Expectations and Partially Meeting Expectations performance levels, cut scores that identify moderate risk link to scale scores in the middle of the Partially Meeting Expectations performance level, and cut scores that identify negligible risk link to scale scores near the beginning of the MCAS ELA Meeting Expectations performance level.

In this study, we dichotomized risk to discriminate between students who are and are not on track to MCAS ELA proficiency standards. Dichotomizing risk in this way emphasized how definitions of on track and at risk can play an important role in determining which screening assessment benchmarks educators might use when making decisions about instruction and interventions.

For example, in this study, the DIBELS 8th Edition At Benchmark cut score was used to identify students who are and are not on track to meeting MCAS ELA proficiency standards. That cut score seems reasonable as an overall on-track/at-risk indicator as it separates students with moderate to high risk from those with little risk of reading difficulties. As an on-track indicator of MCAS Grade 3 ELA proficiency, however, that cut score is too lenient, as the FP rate shows 37 percent of the students classified as on track did not meet proficiency standards (see table 32). The results of this study suggest the Above Benchmark cut score, which has an FP rate of 16 percent, could be a better on-track/at-risk indicator for educators designing instruction and interventions geared toward MCAS ELA proficiency.

On the other hand, if educators are primarily interested in designing instruction and intensive interventions geared toward the students who are most in need, then the At Benchmark cut score would be too stringent, meaning too many students would be identified as at risk given the purpose. If the goal were to target students most in need of assistance, the Below Benchmark cut score would be a better on-track/at-risk indicator, as indicated by its FN rate of 0.04 (see table 32).

Overall, the results show students who are classified by literacy screening assessments as at moderate or high risk of reading difficulties during the spring of their grade 3 year rarely meet

proficiency standards on the MCAS Grade 3 ELA test. This is perhaps expected given the narrow window of time to improve in areas of weakness between the grade 3 EOY literacy screening assessment and MCAS Grade 3 ELA assessments and overall low rates of proficiency on MCAS. Future studies may link literacy assessment performance in earlier grades to MCAS Grade 3 ELA to provide a better understanding of the ways in which students classified as having moderate and high risk of reading difficulties progress over time.

Discussion and Next Steps

This report provides a first look at literacy performance in grades K–3 in the state of Massachusetts. As discussed, the sample of data available to examine performance is not perfectly representative of the state as a whole and may provide a slightly different picture of the numbers of students at risk than would a statewide dataset.

As more data are collected over time, the extent to which the screening assessment data can provide useful information will likely improve. Additionally, some issues with the initial set of data (such as missing BOY scores) may be resolved over time as the collection and processing of screening assessment data becomes routine in the state. DESE is also planning to work with assessment vendors to potentially collect data about grantees directly, thus reducing the burden on schools and districts and improving the consistency of the data collected.

In the meantime, the analyses in this report may provide some useful insights for policymakers and practitioners. For example, while many students can be classified as at risk at least once during the school year, the percentage who are always at risk is much smaller. Perhaps even more importantly, analysis of student performance across time periods within the school year shows that students identified as at risk can and do get back on track, and focusing on providing effective interventions and supports to get students on track can make a difference. Those types of supports are also relevant for English learner students—data in this report suggest that screening assessments may identify students at risk of reading difficulties even among students just learning English, though there is also evidence that screening assessments may at least somewhat over-identify these students. Finally, the linking study provides information about how benchmark cut scores compare across assessments, which could identify metrics that would be consistent across districts for purposes of targeting resources and supports.

Future research will be able to track more students across grades and extend the analyses in this initial report. Additionally, new questions—for example, analysis of student, school, and district characteristics associated with performance or of districts or schools outperforming peers, or potentially about specific reading skills as measured by reading subtests—can also be investigated, with the ultimate goal of continuing to inform policy and practice that will advance literacy in the state.

References

Amplify Education, Inc. (2019). mClass dyslexia screening measures technical manual.

Brennan, R. L. (Ed) (2006). *Educational Measurement,* 4th ed. Westport, CT: Praeger Publishers.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., ... & Holzinger, A. (2021). *Deep ROC analysis and AUC as balanced average accuracy to improve model selection, understanding and interpretation*.

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. (2022). *Massachusetts Department of Elementary and Secondary Education early literacy screening assessment RFP# 22MACISMB1*.

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. (2019). *FAST autoreading technical report*.

FastBridge Learning. (2019). *FastBridge Learning: Benchmarks and norms interpretation and use guidelines (version 4)*.

FastBridge Learning. (2019). *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). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. 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., … Wallin, J. (2011). *DIBELS Next assessment manual.* Eugene, OR: Dynamic Measurement Group.

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.

Heubeck, E. (2023, February 23). Universal screening for dyslexia is not enough. *EdWeek.*

Ho, A., (2012, October). *Off track: Problems with "on track" inferences in empirical and predictive standard setting*. Working paper.

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. New York, NY: Springer.

Lexia Learning. (2020). *Lexia RAPID assessment technical manual for RAPID version 5.1: Grades K–2*.

Lexia Learning. (2022). *Lexia RAPID assessment technical manual for RAPID version 5.1: Grades 3–12*.

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*.

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 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*. The National Academies Press. https://doi.org/10.17226/6023.

NWEA. (2019). *MAP growth technical report March 2019*.

NWEA. (2020). *Linking study report: Predicting performance on the Massachusetts Comprehensive Assessment System (MCAS) based on NWEA MAP Growth scores*. Portland, OR: Author.

NWEA. (2021). *Spanish MAP growth reading technical report March 17, 2021*.

NWEA. (2021). *English MAP reading fluency technical report August 2021*.

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).

Renaissance Learning. (2021). *Star assessments for Spanish: Reading technical documentation*.

Renaissance Learning. (2021). *Star CBM reading technical manual*.

Renaissance Learning. (2022). *Defining benchmarks in star assessments*.

Renaissance Learning. (2022). *Star assessments for early literacy technical manual*.

Renaissance Learning. (2022). *Star assessments for early literacy technical manual*.

Renaissance Learning. (2022). *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

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

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 (2018–2020). *8th edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Technical manual*. https://dibels.uoregon.edu

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

Appendix

Appendix A. Initial Research Questions

Table A.1. Initial research questions

Topic	Question	2020/21 and 2021/22
Summarize	How many scores are available overall and for each assessment, by grade and subgroup? To what extent do the scores represent the overall K–3 student population in the state?	x
What	What are the relationships between subtests and meeting or not meeting benchmarks within assessments? Can we identify profiles of students with particular combinations of skills who meet or do not meet benchmarks? Are there any patterns across assessments, keeping in mind that the definitions of skills may not be the same for each assessment?	
What	What are the relationships between subtests/subskills and MCAS performance? Can we identify profiles of students with particular combinations of skills who are proficient or not proficient on MCAS?	
What—ELs	What are the relationships between subtests/components skills for English learners? What are the relationships between subtests and meeting or not meeting benchmarks?	
When	What are the trends in student early literacy attainment within grades K–2? How many students are meeting/not meeting benchmarks at each time period at each grade level? How do the proportions of students meeting/not meeting benchmarks compare across grades and time periods? Are there particular point(s) when more or fewer students are off track? At what grades does the proportion of students failing to meet benchmarks peak?	x
When	For each assessment and for the assessments taken together, which, if any, test periods (beginning, middle, or end of year in grades K–3) explain the most variation in grade 3 MCAS results? How much additional explanatory power is provided by using data from each time period?	
Who	What proportion of Massachusetts students who do not meet literacy benchmarks in kindergarten or grade 1 achieve benchmark status, and when?	
Who	In what ways, if at all, do Massachusetts students who do not meet literacy benchmarks in kindergarten or grade 1 and who achieve benchmark status within the next 2 years on the same assessment, differ systematically from students who never meet benchmark?	

Topic	Question	2020/21 and 2021/22
Who/Why	What student factors are associated with variation in student reading achievement and benchmark status within grades K–3? Examples include race/ethnicity, language status, IEP status, socioeconomic status, gender, and mobility for students.	
Why	Are there schools or districts where students are "outperforming" expectations, based on the demographics of students and/or schools or other factors?	
Why	What are the primary school or district factors associated with variation in student reading achievement and benchmark status within grades K–3? Examples include urban/suburban/rural, size, accountability status, and aggregated student-level demographics such as socioeconomic status.	
Why	What are the primary instructional factors associated with variation in student reading achievement and benchmark status within grades K–2? Examples include use of core curricular materials, interventions in use, level of MTSS implementation, teacher knowledge/skill in reading (subject to availability of data).	
Why—EL focus	What is the relationship of English language proficiency to screening assessment performance and achieving benchmarks?	**x**
Why—EL focus	How does performance and benchmark attainment vary for English learners by first (native) language and type of services or program?	**x**
Compare benchmarks, predict later outcomes	How do literacy benchmark scores compare to each other in MCAS score terms? What percentile in the distribution of students taking each screening assessment is represented by each benchmark score? What is the MCAS score that corresponds to that percentile? [Equipercentile linking, predictive linking] What is the predicted grade 3 MCAS score for a student on each literacy benchmark assessment, given screening assessment scores? [Predictive linking]	**x**

Appendix B. Assessment Review Criteria and Summaries

The table below provides the established criteria that 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 B.1. Massachusetts's early literacy universal screening assessment criteria

Criteria Category	Assessment Details

Constructs Measured	• Alphabetic Knowledge
	  ○ Letter Identification (Kindergarten)
	  ○ 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 learners • Normed for English learners 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 National Center on Intensive Intervention (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 B.2 through B.9 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 B.2. 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 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: Uses a Letter Naming Fluency task with distinct cut scores 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 B.3. MA DESE's summary of EarlyBird, from EarlyBird Education

Approval Status	Approved: Meets Expectations
Grades Covered	K 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 ☒ RAN: 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 $8.00
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/ Sue Bonaiuto Susan.bonaiuto@earlybird.com 617-462-4779

Table B.4. 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 ☒ RAN: Uses a Letter Naming Fluency task with distinct cut scores 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 B.5. 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 ☒ RAN: 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 PD: $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 B.6. 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 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 ☒ RAN: 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 $8.00; 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 PD: $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 B.7. 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
Skillsd 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: 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 six hours. Six-hour virtual sessions may be split among several days; onsite sessions are one 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 B.8. 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
SkillsAssessed 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: 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 $9.00 (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/ Mary Ellen LaCamera Maryellen.lacamera@nwea.org

Table B.9. MA DESE's summary of STAR Elementary Bundle (Early Literacy, Reading, CBM) from Renaissance

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 ☒ RAN: 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 $9.00 (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/ Mary Ellen LaCamera Maryellen.lacamera@nwea.org

Appendix C. Business Rules and Data Processing Specifications

This report draws on data from multiple sources, including extant student-level data provided by the 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 the Early Grades Literacy grant (EGL; FC734), the Early Literacy Screening Assessment and Professional Development grant (FC576), and/or the Growing Literacy Equity Across Massachusetts grant (GLEAM; FC509/510);
- 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

One hundred and fifty-three unique files with early literacy universal screening assessment data were provided by DESE—26 files were from districts receiving the EGL grant, 22 files were from districts receiving the GLEAM grant, and 108 files were from districts receiving the FC576 grant.[23] Of the 153 files provided, five files were not used; four files, which contained Acadience Reading, DIBELS 8, Star CBM Reading, and Star Early Literacy scores, were not used as they did not contain a student identifier that could be used to link the screening assessment data to the state-level data, and one file, which contained FastBridge AUTOreading scores, was not used as K–3 scores were not included.

Cleaning of the 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, Lexile measures); renaming variables to create a standardized format across screening assessments; creating variables to contain school and

[23] Three files were associated with both the EGL and GLEAM grants.

district codes for the merging process; creating variables containing the composite benchmark level 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], end of year [EOY]); and selecting one score per student per time period per screening assessment.

The analytic team used vendor-defined cut scores (obtained through the assessment technical manuals or communication with the vendors) to create a composite benchmark variable and reading risk flag variable. Although these variables were typically available in the DESE-provided files, it was possible for districts/schools to manually adjust 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 the 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 screening assessment flag and the EarlyBird dyslexia risk flag 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 EarlyBird flag is generated by the EarlyBird team using "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."

The time period corresponding to each score was typically determined by using a variable within the file or was indicated in the file name. However, in some cases the time period was missing (in 23 files: nine for i-Ready, one for Star Early Literacy, one for Star Reading, and 12 for ISIP ER). The time period was determined for the 11 i-Ready, Star Early Literacy, and Star Reading files by using the administration date provided within the file and the default testing periods provided by the vendors. ISIP ER was typically delivered each month during the school year. The analytic team used the September scores as the BOY scores, January scores as the MOY scores, and the May scores as the EOY scores. If students were missing September scores, October scores were used as the BOY scores, February scores were used as the MOY scores if January scores were missing, and June scores were used as the end of year scores if May scores were missing.

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

- If one observation had more data or one record had a composite score and another did not, the observation with more data or a composite score was selected.
- 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 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 CBM Reading 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 rushing flag if they answered questions in less than 11 seconds on average per item and a yellow rushing 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, 1,416 DIBELS 8th Edition composite scores (provided by districts) were replaced with composite scores generated by the analytic team using the composite score formulas in the technical manual. This recalculation was done for three main reasons. First, some of the observations in the district or school-provided files were missing composite scores, and some subtest scores were also missing. If the student met the gating rules (i.e., the student was high performing and did not take all subtests as a result) or the discontinue rules (i.e., the student significantly struggled and did not take all subtests), vendor-provided rules were used to estimate the student's composite score. Second, 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). Third, some of the observations had composite scores although the student did not complete the necessary grade-level subtests and did not meet the gating/discontinue rules.

Finally, three mCLASS 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 (SIMS, MCAS, ACCESS)

In addition to the K–3 early literacy screening assessment data, other student-level data from the 2020/21 and 2021/22 school years were also used for the analysis, including October and June Student Information Management System (SIMS) data, Massachusetts Comprehensive

Assessment System (MCAS) data, and Assessing Comprehension and Communication in English State-to-State for English Language Learners (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., 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 assessment data with the state education data. First, the combined screening assessment file was merged with the SIMS data using the June collection where possible. The analytic team first attempted to match student screening assessment scores with their June SIMS data using the state assigned student identifier (SASID), grade level, school code, and district code. As students may appear multiple times within the SIMS collection 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 screening 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 screening assessment data with their appropriate June SIMS observation using the student's SASID, grade level, and district code (i.e., without the school code as a matching variable). The process was repeated with the October collection 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 student's state identifier; rather, the files only contained 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*[24] to determine the appropriate screening assessment-SIMS match to keep. Typically, we used the enrollment status variable, which describes the enrollment status of a student within the school

[24] The 2022 researcher's guide is available on DESE's website: https://www.doe.mass.edu/research/researchers-guide.docx.

(e.g., enrolled, dropout, transferred in-state public) 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 103,339 screening assessment observations for the 2020/21 and 2021/22 school years, 103,065 (99.7%) matched with a corresponding observation in the SIMS collection data. Of the observations that matched, 102,751 (99.7%) matched with an observation in the June SIMS data and 314 (0.3%) matched with an observation in the October SIMS data. Six thousand six hundred and fifty screening assessment observations did not match exactly with the SIMS data (i.e., did not match using the grade level, school code, and district code); 16 screening assessment observations had a different grade level than their corresponding SIMS observation, 44 had different district codes, and 6,646 had different school codes. Ninety-eight percent of the school differences (6,491 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 dataset 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 and 2021/22 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: (1) educator characteristics, (2) student performance, (3) student enrollment and demographics, and (4) financial. These data were 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 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, 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 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. Additional Tables: Student Performance and Progress

Tables D.1.1 and D.1.2 provide the demographic breakdown of the early literacy screening assessment sample (for 2020/21 and 2021/22) by screening assessment.

Table D.1.1. Comparison of student demographics of early literacy screening assessment sample to the state (by early literacy screening assessment)

Demographic	Acadience Reading	DIBELS 8th Edition	mCLASS	FastBridge aReading	FastBridge CBMreading	FastBridge earlyReading	i-Ready	State
Kindergarten	100%	31%	27%	2%	2%	47%	16%	24%
Grade 1	0%	29%	28%	7%	16%	47%	28%	25%
Grade 2	0%	22%	29%	51%	45%	3%	28%	26%
Grade 3	0%	18%	16%	41%	37%	3%	28%	26%
Economically disadvantaged	7%	44%	42%	60%	59%	56%	23%	41%
Female	51%	49%	49%	50%	50%	47%	49%	49%
Male	49%	51%	51%	50%	50%	53%	51%	51%
Nonbinary	0%	--	--	0%	0%	0%	--	<1%
English learner students	--	19%	17%	2%	2%	--	7%	17%
Special education designation	12%	18%	19%	23%	23%	22%	17%	15%
White	81%	85%	76%	93%	94%	96%	84%	78%
Black	--	10%	8%	14%	11%	6%	9%	15%
Hispanic	--	28%	10%	7%	6%	4%	10%	24%
Asian	20%	8%	19%	2%	2%	--	13%	10%
American Indian/Alaskan Native	--	3%	<1%	4%	3%	--	<1%	3%
Native Hawaiian/Pacific Islander	0%	<1%	<1%	--	--	--	<1%	<1%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.1.2. Comparison of student demographics of early literacy screening assessment sample to the state (by early literacy screening assessment)

Grade Level/Demographic	ISIP ER	Lexia RAPID	MAP Growth	MAP Reading Fluency	Star Early Literacy	Star Early Literacy Spanish	Star Reading	State
Kindergarten	31%	23%	12%	<1%	38%	33%	1%	24%
Grade 1	35%	24%	11%	99%	37%	34%	14%	25%
Grade 2	33%	27%	13%	0%	22%	29%	40%	26%
Grade 3	0%	27%	65%	0%	2%	4%	45%	26%
Economically disadvantaged	41%	59%	61%	70%	68%	80%	58%	41%
Female	51%	48%	53%	57%	49%	50%	48%	49%
Male	49%	52%	47%	43%	51%	50%	52%	51%
Nonbinary	0%	0%	0%	0%	--	0%	--	<1%
English learner students	24%	26%	32%	--	45%	93%	39%	17%
Special education designation	17%	19%	14%	21%	16%	9%	14%	15%
White	81%	77%	56%	93%	67%	42%	58%	78%
Black	20%	21%	16%	12%	22%	9%	14%	15%
Hispanic	21%	51%	50%	8%	48%	96%	51%	24%
Asian	3%	4%	26%	--	6%	--	7%	10%
American Indian/Alaskan Native	--	<1%	7%	--	14%	59%	29%	3%
Native Hawaiian/Pacific Islander	--	<1%	2%	0%	<1%	--	--	<1%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.2 compares the sample of grade 3 students with the state's grade 3 population by comparing the percent of students meeting or exceeding expectations on MCAS (for each screening assessment) with the state average.

Table D.2. Comparison of percent of students meeting/exceeding expectations on MCAS

Early Literacy Screening Assessment	Sample (% Meeting/Exceeding Expectations)	State (% Meeting/Exceeding Expectations)
All Screening Assessments	43%	47%
DIBELS 8th Edition	39%	47%
mCLASS	47%	47%
FastBridge aReading	37%	47%
FastBridge CBMreading	38%	47%
i-Ready	56%	47%
Lexia RAPID	30%	47%
MAP Growth	44%	47%
Star Early Literacy	3%	47%
Star Reading	27%	47%

Source: 2020/21 and 2021/22 district-provided screening assessment data and state-provided MCAS data

Note: No grade 3 students took Acadience Reading, ISIP ER, or MAP Reading Fluency. FastBridge earlyReading and Star Early Literacy Spanish were not included in the table above as only 11 and 51 grade 3 students with MCAS achievement levels, respectively, took these assessments.

Table D.3 provides the number of observations with available benchmarks by time period, for each of the early literacy screening assessments.

Table D.3. Number of observations with benchmarks by time period and screening assessment

Early Literacy Screening Assessment	BOY	MOY	EOY	Total
Acadience Reading	107	108	111	326
DIBELS 8th Edition	7,813	7,091	7,746	22,650
mCLASS	4,186	4,507	4,565	13,258
FastBridge aReading	478	462	402	1,342
FastBridge CBMreading	612	678	314	1,604
FastBridge earlyReading	344	280	251	875
i-Ready	7,005	7,456	5,765	20,226
ISIP ER	637	645	620	1,902
Lexia RAPID	2,173	2,136	2,266	6,575
MAP Growth	461	466	981	1908
MAP Reading Fluency	158	156	0	314
Star Early Literacy	3,492	6,239	6,138	15,869
Star Early Literacy Spanish	60	1,100	1,087	2,247
Star Reading	496	1,340	1,290	3,126
Total	28,022	32,664	31,536	92,222

Source: 2020/21 and 2021/22 district-provided screening assessment data

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

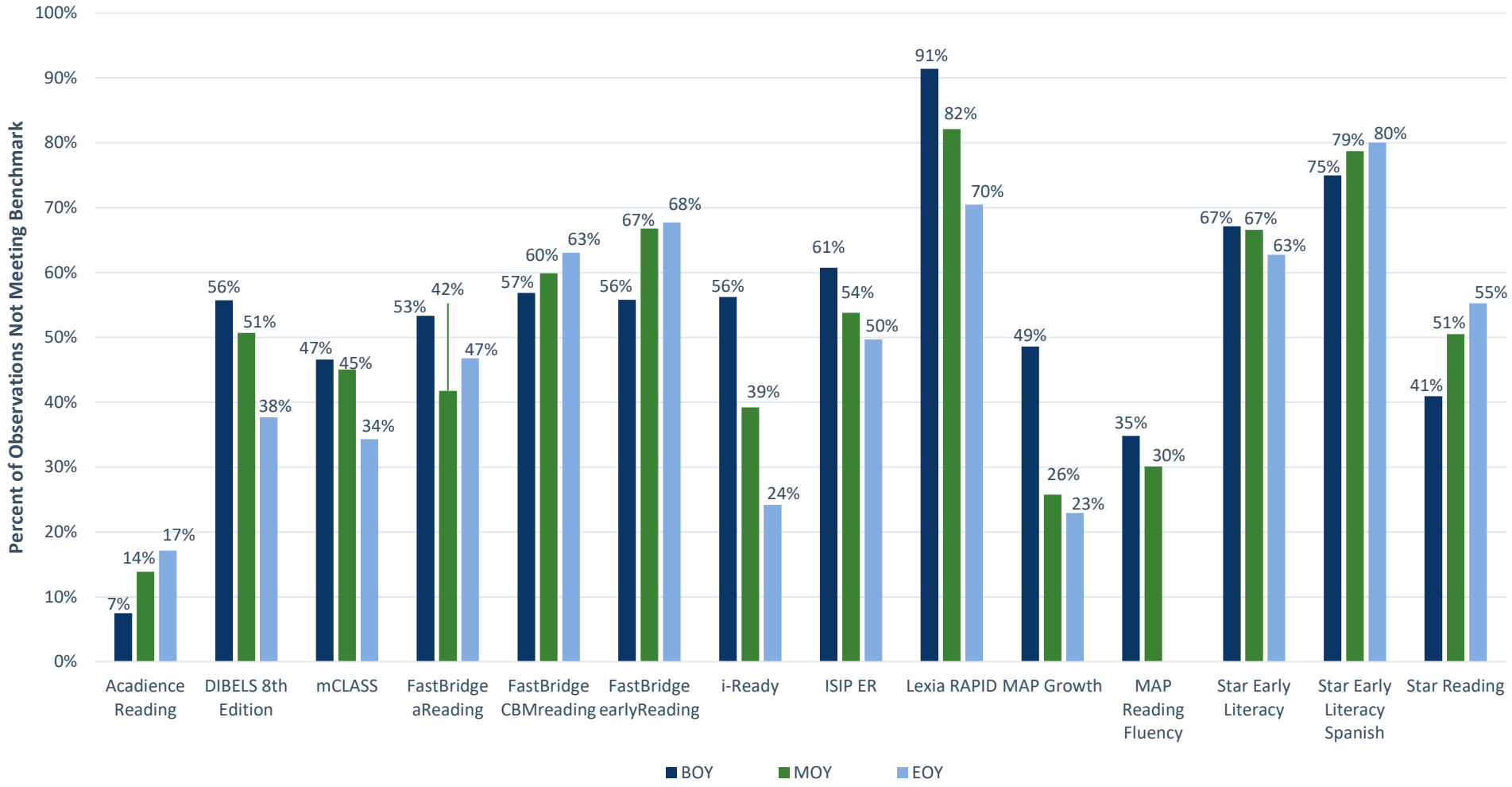
Table D.4. Number of benchmark scores by time period and grade level and demographic characteristics

Grade Level or Demographic	BOY	MOY	EOY	Total
Kindergarten	5,800	8,456	8,857	23,113
Grade 1	8,741	9,610	9,447	27,798
Grade 2	7,861	8,364	7,940	24,165
Grade 3	5,620	6,234	5,292	17,146
Economically disadvantaged	11,538	15,489	15,392	42,419
Female	13,718	16,061	15,500	45,279
Male	14,217	16,532	15,914	46,663
Nonbinary	--	--	--	23
English learner students	4,668	8,036	8,290	20,994
Special education designation	5,076	5,495	5,280	15,851
White	23,132	25,028	23,788	71,948
Black	3,285	4,253	4,213	11,751
Hispanic	6,179	9,805	9,880	25,864
Asian	2,760	3,020	2,989	8,769
American Indian/Alaskan Native	560	2,572	2,524	5,656
Native Hawaiian/Pacific Islander	142	182	183	507

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. The percentage of students classified as economically disadvantaged for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. A cell may contain the same student more than once if they were delivered multiple screening assessments within the school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Figure D.1 Benchmark performance by time period and screening assessment



Source: 2020/21 and 2021/22 district-provided screening assessment data

Note: Some students may appear multiple times per time period if they were administered multiple screening assessments. There are no records available for MAP Reading Fluency at EOY.

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

Table D.5. Comparison of student demographics of students with one available benchmark versus two available benchmarks versus three available benchmarks

Grade Level or Demographic	1 Benchmark Available	2 Benchmarks Available	3 Benchmarks Available	State
Kindergarten	26%	31%	23%	24%
Grade 1	25%	21%	34%	25%
Grade 2	28%	23%	27%	26%
Grade 3	21%	25%	16%	26%
Economically disadvantaged	60%	54%	42%	41%
Female	49%	49%	49%	49%
Male	51%	51%	51%	51%
Nonbinary	--	--	--	<1%
English learner students	35%	33%	18%	17%
Special education designation	18%	16%	18%	15%
White	71%	70%	82%	78%
Black	14%	15%	12%	15%
Hispanic	35%	38%	24%	24%
Asian	10%	7%	10%	10%
American Indian/Alaskan Native	11%	16%	2%	3%
Native Hawaiian/Pacific Islander	<1%	<1%	<1%	<1%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.6. Benchmark performance by screening assessment and grade level

Early Literacy Screening Assessment	Kindergarten		Grade 1		Grade 2		Grade 3		All Grades	
	Did Not Meet	Met	Did Not Meet	Met	Did Not Meet	Met	Did Not Meet	Met	Did Not Meet	Met
Acadience Reading	42	284							42	284
DIBELS 8th Edition	3,753	3,429	3,070	3,338	2,280	2,879	1,767	2,134	10,870	11,780
mCLASS	1,769	1,842	1,571	2,209	1,450	2,331	758	1,328	5,548	7,710
FastBridge aReading	0	10	--	39	411	299	223	358	636	706
FastBridge CBMreading	--	--	170	104	465	249	310	291	952	652
FastBridge earlyReading	263	157	275	152	--	12	--	--	549	326
i-Ready	856	1,677	3,280	2,918	2,634	3,473	1,488	3,900	8,258	11,968
ISIP ER	339	264	419	247	284	349			1,042	860
Lexia RAPID	1,246	249	1,276	297	1,493	267	1,323	424	5,338	1,237
MAP Growth	31	303	110	200	109	256	319	580	569	1,339
MAP Reading Fluency	0	--	102	211					102	212
Star Early Literacy	3,188	2,576	4,320	2,446	2,599	489	246	--	10,353	5,516
Star Early Literacy Spanish	567	230	663	74	500	141	51	21	1,781	466
Star Reading	10	12	128	177	510	681	945	663	1,593	1,533
Total	12,071	11,042	15,386	12,412	12,739	11,426	7,437	9,709	47,633	44,589

Source: 2020/21 and 2021/22 district-provided screening assessment data

Note: Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.7. Benchmark performance by screening assessment

Early Literacy Screening Assessment	Never Below Benchmark in School Year	Always Below Benchmark (Below Benchmark in BOY, MOY, EOY)	Ever Below Benchmark in School Year	Below Benchmark at EOY
Acadience Reading	71%	--	29%	17%
DIBELS 8th Edition	40%	23%	60%	38%
mCLASS	45%	22%	55%	34%
FastBridge aReading	49%	20%	51%	47%
FastBridge CBMreading	36%	8%	64%	63%
FastBridge earlyReading	32%	19%	68%	68%
i-Ready	43%	13%	57%	24%
ISIP ER	30%	33%	70%	50%
Lexia RAPID	8%	54%	92%	70%
MAP Growth	64%	8%	36%	23%
MAP Reading Fluency	57%	N/A	43%	N/A
Star Early Literacy	24%	13%	76%	63%
Star Early Literacy Spanish	15%	<1%	85%	80%
Star Reading	41%	1%	59%	55%
Total	**36%**	**18%**	**64%**	**45%**

Source: 2020/21 and 2021/22 district-provided screening assessment data

Note: Some students may appear multiple times per row if they were administered multiple screening assessments. "Never Below Benchmark in School Year," "Always Below Benchmark," and "Ever Below Benchmark in School Year" include students with one, two, or three scores. "Below Benchmark at EOY" includes students with EOY scores, which could include students with one, two, or three scores.

Table D.8. Benchmark performance by time period and grade level and demographic characteristics

Grade Level or Demographic	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Kindergarten	3,613	62%	4,659	55%	3,799	43%	12,071	52%
Grade 1	5,439	62%	5,567	58%	4,380	46%	15,386	55%
Grade 2	4,668	59%	4,440	53%	3,631	46%	12,739	53%
Grade 3	2,574	46%	2,652	43%	2,211	42%	7,437	43%
Economically disadvantaged	8,254	72%	10,468	68%	9,170	60%	27,892	66%
Non-economically disadvantaged	7,988	49%	6,797	40%	4,800	30%	19,585	40%
Female	7,904	58%	8,355	52%	6,682	43%	22,941	51%
Male	8,347	59%	8,921	54%	7,292	46%	24,560	53%
Nonbinary	--	--	--	--	--	--	--	--
English learner	3,698	79%	6,154	77%	5,734	69%	15,586	74%
Non-English learner	12,554	54%	11,124	45%	8,242	36%	31,920	45%
Special education	4,025	78%	4,135	74%	3,592	67%	11,752	73%
Non-special education	12,269	54%	13,183	49%	10,429	40%	35,881	47%
White	13,614	59%	12,877	51%	10,108	42%	36,599	51%
Black	2,248	68%	2,630	62%	2,262	54%	7,140	61%
Hispanic	4,656	75%	7,024	72%	6,385	65%	18,065	70%
Asian	1,063	39%	1,011	33%	765	26%	2,839	32%
American Indian/Alaska Native	362	65%	1,948	76%	1,802	71%	4,112	73%
Native Hawaiian/Pacific Islander	80	56%	92	51%	85	46%	257	51%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. The percentage of students classified as economically disadvantaged was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. A cell may contain the same student more than once if they were delivered multiple screening assessments within the school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Tables D.9.1 through D.9.14 provide a breakdown of the performance of students for each of the screening assessments in each time period. The screening assessment-specific benchmark category names are used. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.9.1. Benchmark performance by time period, Acadience Reading

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Above Benchmark	90	84%	62	57%	48	43%	200	61%
At Benchmark	--	--	31	29%	44	40%	84	26%
Below Benchmark	--	--	13	12%	16	14%	35	11%
Well Below Benchmark	--	--	--	--	--	--	--	--

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.2. Benchmark performance by time period, DIBELS 8th Edition

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Above Benchmark	1,580	20%	1,633	23%	2,579	33%	5,792	26%
At Benchmark	1,878	24%	1,863	26%	2,247	29%	5,988	26%
Below Benchmark	1,357	17%	1,183	17%	1,131	15%	3,671	16%
Well Below Benchmark	2,998	38%	2,412	34%	1,789	23%	7,199	32%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.3. Benchmark performance by time period, mCLASS

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Above Benchmark	1,159	28%	1,144	25%	1,662	36%	3,965	30%
At Benchmark	1,076	26%	1,332	30%	1,337	29%	3,745	28%
Below Benchmark	692	17%	737	16%	586	13%	2,015	15%
Well Below Benchmark	1,259	30%	1,294	29%	980	21%	3,533	27%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.4. Benchmark performance by time period, FastBridge aReading

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Advanced/College Pathway	117	24%	153	33%	100	25%	370	28%
Low Risk	106	22%	116	25%	114	28%	336	25%
Some Risk	122	26%	98	21%	96	24%	316	24%
High Risk	133	28%	95	21%	92	23%	320	24%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.5. Benchmark performance by time period, FastBridge CBMreading

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Advanced/College Pathway	131	21%	113	17%	50	16%	294	18%
Low Risk	133	22%	159	23%	66	21%	358	22%
Some Risk	147	24%	170	25%	88	28%	405	25%
High Risk	201	33%	236	35%	110	35%	547	34%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.6. Benchmark performance by time period, FastBridge earlyReading

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Low Risk	152	44%	93	33%	81	32%	326	37%
Some Risk	100	29%	69	25%	71	28%	240	27%
High Risk	92	27%	118	42%	99	39%	309	35%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.7. Benchmark performance by time period, i-Ready

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Mid or Above Grade Level	1,687	24%	2,871	39%	3,183	55%	7,741	38%
Early on Grade Level	1,379	20%	1,660	22%	1,188	21%	4,227	21%
1 Grade Level Below	3,117	44%	2,471	33%	1,192	21%	6,780	34%
2 Grade Levels Below	728	10%	403	5%	181	3%	1,312	6%
3 or More Grade Levels Below	94	1%	51	1%	21	0%	166	1%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.8. Benchmark performance by time period, ISIP ER

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Tier 1	250	39%	298	46%	312	50%	860	45%
Tier 2	186	29%	166	26%	137	22%	489	26%
Tier 3	201	32%	181	28%	171	28%	553	29%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.9. Benchmark performance by time period, Lexia RAPID

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
High Likelihood of EOY Grade-Level Success	186	9%	382	18%	669	30%	1,237	19%
Moderate Likelihood of EOY Grade-Level Success	256	12%	388	18%	445	20%	1,089	17%
Low Likelihood of EOY Grade-Level Success	1,731	80%	1,366	64%	1,152	51%	4,249	65%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.10. Benchmark performance by time period, MAP Growth

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
No Intensive Intervention	237	51%	346	74%	756	77%	1,339	70%
Intensive Intervention	224	49%	120	26%	225	23%	569	30%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.11. Benchmark performance by time period, MAP Reading Fluency

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
Not Flagged	103	65%	109	70%			212	68%
Flagged	55	35%	47	30%			102	32%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.12. Benchmark performance by time period, Star Early Literacy

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
At/Above Benchmark	1,147	33%	2,084	33%	2,285	37%	5,516	35%
On Watch	444	13%	740	12%	733	12%	1,917	12%
Intervention	647	19%	1,084	17%	912	15%	2,643	17%
Urgent Intervention	1,254	36%	2,331	37%	2,208	36%	5,793	37%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.13. Benchmark performance by time period, Star Early Literacy Spanish

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
At/Above Benchmark	15	25%	234	21%	217	20%	466	21%
On Watch	--	--	152	14%	175	16%	333	15%
Intervention	--	--	309	28%	269	25%	587	26%
Urgent Intervention	30	50%	405	37%	426	39%	861	38%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.9.14. Benchmark performance by time period, Star Reading

Benchmark	BOY: Did Not Meet		MOY: Did Not Meet		EOY: Did Not Meet		All Time Periods: Did Not Meet	
	n	%	n	%	n	%	n	%
At/Above Benchmark	293	59%	663	49%	577	45%	1,533	49%
On Watch	56	11%	158	12%	149	12%	363	12%
Intervention	54	11%	177	13%	175	14%	406	13%
Urgent Intervention	93	19%	342	26%	389	30%	824	26%

Source: 2020/21 and 2021/22 district-provided screening assessment data

Table D.10. Significant risk performance by grade level and demographic characteristics

Grade Level or Demographic	Never At Significant Risk in School Year	At Significant Risk Three Times	Ever At Significant Risk in School Year	Significant Risk at EOY	Relative Risk of Ever Being at Significant Risk
Kindergarten	59%	7%	41%	23%	
Grade 1	61%	10%	39%	27%	
Grade 2	60%	12%	40%	28%	
Grade 3	65%	9%	35%	29%	
Economically disadvantaged	47%	13%	53%	38%	2.0
Non-economically disadvantaged	74%	6%	26%	15%	
Female	63%	9%	37%	25%	0.9
Male	59%	10%	41%	28%	
Nonbinary	--	--	--	--	
English learner students	38%	14%	62%	46%	1.9
Non-English learner students	68%	8%	32%	19%	
Students receiving special education services	36%	22%	64%	50%	1.9
Students not receiving special education services	66%	7%	34%	22%	
White	62%	10%	38%	25%	0.9
Black	53%	11%	47%	32%	1.2
Hispanic	42%	14%	58%	42%	1.9
Asian	78%	5%	23%	14%	0.5
American Indian/Alaskan Native	47%	3%	53%	43%	†
Native Hawaiian/Pacific Islander	57%	10%	43%	27%	†

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Note: Race and ethnicity are not exclusive. Student records can indicate more than one. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

†Not computed because groups are 5 percent or less of the sample.

Table D.11. Performance progression for kindergarten students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	84%	89%
MOY	N/A	N/A	88%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	72%	51%
MOY	N/A	N/A	68%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.12. Performance progression for grade 1 students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	89%	92%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	77%	58%
MOY	N/A	N/A	73%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.13. Performance progression for grade 2 students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	94%	96%
MOY	N/A	N/A	94%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	79%	67%
MOY	N/A	N/A	80%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.14. Performance progression for grade 3 students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	94%	93%
MOY	N/A	N/A	91%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	73%	67%
MOY	N/A	N/A	82%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.15. Performance progression for EL students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	84%	86%
MOY	N/A	N/A	82%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	86%	74%
MOY	N/A	N/A	83%

Source: 2020/21 and 2021/22 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 D.16. Performance progression for non-EL students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	92%	94%
MOY	N/A	N/A	93%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	73%	56%
MOY	N/A	N/A	69%

Source: 2020/21 and 2021/22 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 D.17. Performance progression for economically disadvantaged students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	87%	88%
MOY	N/A	N/A	87%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	84%	71%
MOY	N/A	N/A	80%

Source: 2020/21 and 2021/22 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 are not included in the table. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year.

Table D.18. Performance progression for non–economically disadvantaged students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	92%	95%
MOY	N/A	N/A	94%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	68%	49%
MOY	N/A	N/A	65%

Source: 2020/21 and 2021/22 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 are not included in the table. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year.

Table D.19. Performance progression for students receiving special education services

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	83%	87%
MOY	N/A	N/A	87%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	88%	80%
MOY	N/A	N/A	86%

Source: 2020/21 and 2021/22 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 D.20. Performance progression for students not receiving special education services

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	92%	93%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	72%	54%
MOY	N/A	N/A	71%

Source: 2020/21 and 2021/22 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 D.21. Performance progression for White students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	76%	59%
MOY	N/A	N/A	73%

Source: 2020/21 and 2021/22 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 D.22. Performance progression for non-White students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	90%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	77%	61%
MOY	N/A	N/A	78%

Source: 2020/21 and 2021/22 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 D.23. Performance progression for Black students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	87%	89%
MOY	N/A	N/A	85%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	80%	67%
MOY	N/A	N/A	75%

Source: 2020/21 and 2021/22 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 D.24. Performance progression for non-Black students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	75%	59%
MOY	N/A	N/A	74%

Source: 2020/21 and 2021/22 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 D.25. Performance progression for Hispanic students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	88%	88%
MOY	N/A	N/A	84%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	84%	71%
MOY	N/A	N/A	82%

Source: 2020/21 and 2021/22 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 D.26. Performance progression for non-Hispanic students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	94%
MOY	N/A	N/A	93%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	73%	55%
MOY	N/A	N/A	69%

Source: 2020/21 and 2021/22 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 D.27. Performance progression for Asian students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	93%	96%
MOY	N/A	N/A	96%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	69%	49%
MOY	N/A	N/A	64%

Source: 2020/21 and 2021/22 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 D.28. Performance progression for non-Asian students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	90%	92%
MOY	N/A	N/A	91%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	77%	61%
MOY	N/A	N/A	75%

Source: 2020/21 and 2021/22 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 D.29. Performance progression for American Indian/Alaskan Native students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	88%	88%
MOY	N/A	N/A	81%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	77%	63%
MOY	N/A	N/A	87%

Source: 2020/21 and 2021/22 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 D.30. Performance progression for non–American Indian/Alaskan Native students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	76%	60%
MOY	N/A	N/A	73%

Source: 2020/21 and 2021/22 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 D.31. Performance progression for Native Hawaiian/Pacific Islander students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	97%	91%
MOY	N/A	N/A	88%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	78%	69%
MOY	N/A	N/A	76%

Source: 2020/21 and 2021/22 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 D.32. Performance progression for non–Native Hawaiian/Pacific Islander students

If This Benchmark Met:	Then This Benchmark Met:		
	BOY	MOY	EOY
BOY	N/A	91%	93%
MOY	N/A	N/A	92%

If This Benchmark Missed:	Then This Benchmark Missed:		
	BOY	MOY	EOY
BOY	N/A	76%	60%
MOY	N/A	N/A	74%

Source: 2020/21 and 2021/22 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 D.33. Significant risk performance progression, all grades

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	56%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.34. Significant risk performance progression for kindergarten students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	92%	93%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	61%	39%
MOY	N/A	N/A	59%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.35. Significant risk performance progression for grade 1 students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	93%	94%
MOY	N/A	N/A	93%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	72%	58%
MOY	N/A	N/A	69%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.36. Significant risk performance progression for grade 2 students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	95%	97%
MOY	N/A	N/A	96%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	79%	66%
MOY	N/A	N/A	76%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.37. Significant risk performance progression for grade 3 students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	94%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	73%	65%
MOY	N/A	N/A	80%

Source: 2020/21 and 2021/22 district-provided screening assessment 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 D.38. Significant risk performance progression for EL students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	87%	89%
MOY	N/A	N/A	88%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	79%	65%
MOY	N/A	N/A	75%

Source: 2020/21 and 2021/22 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 D.39. Significant risk performance progression for non-EL students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	95%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	68%	52%
MOY	N/A	N/A	66%

Source: 2020/21 and 2021/22 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 D.40. Significant risk performance progression for economically disadvantaged students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	90%	91%
MOY	N/A	N/A	90%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	75%	62%
MOY	N/A	N/A	73%

Source: 2020/21 and 2021/22 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 are not included in the table. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.41. Significant risk performance progression for non–economically disadvantaged students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	96%	96%
MOY	N/A	N/A	97%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	64%	47%
MOY	N/A	N/A	63%

Source: 2020/21 and 2021/22 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 are not included in the table. Economically disadvantaged status for both the 2020/21 and 2021/22 school years was determined using the state's "economically disadvantaged" metric. The state returned to a different measure (i.e., the "low-income" metric) in the 2021/22 school year; however, this was not used as it was unavailable for the 2020/21 school year. Student groups with fewer than 10 students are not shown to protect student privacy.

Table D.42. Significant risk performance progression for special education students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	85%	85%
MOY	N/A	N/A	87%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	84%	74%
MOY	N/A	N/A	82%

Source: 2020/21 and 2021/22 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 D.43. Significant risk performance progression for non–special education students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	95%	96%
MOY	N/A	N/A	95%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	65%	49%
MOY	N/A	N/A	65%

Source: 2020/21 and 2021/22 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 D.44. Significant risk performance progression for White students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	95%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	56%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 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 D.45. Significant risk performance progression for non-White students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	93%	94%
MOY	N/A	N/A	93%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	74%	57%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 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 D.46. Significant risk performance progression for Black students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	90%	91%
MOY	N/A	N/A	92%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	74%	60%
MOY	N/A	N/A	69%

Source: 2020/21 and 2021/22 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 D.47. Significant risk performance progression for non-Black students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	55%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 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 D.48. Significant risk performance progression for Hispanic students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	90%	91%
MOY	N/A	N/A	89%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	76%	62%
MOY	N/A	N/A	74%

Source: 2020/21 and 2021/22 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 D.49. Significant risk performance progression for non-Hispanic students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	96%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	68%	52%
MOY	N/A	N/A	66%

Source: 2020/21 and 2021/22 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 D.50. Significant risk performance progression for Asian students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	96%	97%
MOY	N/A	N/A	98%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	70%	46%
MOY	N/A	N/A	59%

Source: 2020/21 and 2021/22 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 D.51. Significant risk performance progression for non-Asian students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	93%	94%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	57%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 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 D.52. Significant risk performance progression for American Indian/Alaskan Native students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	92%	91%
MOY	N/A	N/A	87%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	72%	58%
MOY	N/A	N/A	76%

Source: 2020/21 and 2021/22 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 D.53. Significant risk performance progression for non–American Indian/Alaskan Native students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	95%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	56%
MOY	N/A	N/A	69%

Source: 2020/21 and 2021/22 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 D.54. Significant risk performance progression for Native Hawaiian/Pacific Islander students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	95%	96%
MOY	N/A	N/A	93%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	74%	55%
MOY	N/A	N/A	62%

Source: 2020/21 and 2021/22 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 D.55. Significant risk performance progression for non–Native Hawaiian/Pacific Islander students

If Not at Significant Risk:	Then Not at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	94%	95%
MOY	N/A	N/A	94%

If at Significant Risk:	Then at Significant Risk:		
	BOY	MOY	EOY
BOY	N/A	71%	56%
MOY	N/A	N/A	70%

Source: 2020/21 and 2021/22 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.

Appendix E. Additional Tables: English Learner Student Performance

Table E.1. Most common native language sample representation, 2021

Language Spoken and Ranking by Enrollment Number (Sample)	Ranking by Enrollment Number (State)	Number of English Learner Students (Sample)	% of English Learner Students (Sample)	% of English Learner Students (State)
1. Spanish	1	154	55.8%	50.0%
2. Portuguese	2	22	8.0%	13.2%
3. Arabic	4	21	7.6%	3.3%
4. Polish	30	20	7.3%	0.2%
5. Twi	11	18	6.5%	0.8%
6. Swahili	20	--	--	0.5%
7. Chinese	3	--	--	4.9%
8. Creole (Haitian)	6	--	--	3.2%
9. Ga	49	--	--	0.1%
10. Afrikaans	71	--	--	0.02%

Source: 2020/21 and 2021/22 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.

Table E.2. Most common native language sample representation, 2022

Language Spoken and Ranking by Enrollment Number (Sample)	Ranking by Enrollment Number (State)	Number of English Learner Students (Sample)	% of English Learner Students (Sample)	% of English Learner Students (State)
1. Spanish	1	4,607	60.1%	50.4%
2. Portuguese	2	815	10.6%	14.0%
3. Chinese	3	508	6.6%	4.6%
4. Arabic	4	331	4.3%	3.3%
5. Vietnamese	8	173	2.3%	2.3%
6. Creole (Haitian)	6	146	1.9%	3.2%
7. Twi	11	123	1.6%	0.8%
8. Russian	9	88	1.2%	1.9%
9. Albanian	15	81	1.1%	0.7%
10. Khmer	10	81	1.1%	1.9%

Source: 2020/21 and 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table E.3. Rate of achieving MOY and EOY benchmark by ACCESS composite proficiency level, 2021

ACCESS Composite Proficiency Level	Did Not Meet in MOY	Did Not Meet in EOY
Level 1	82%	72%
Level 2	87%	81%
Level 3	68%	66%
Level 4	--	36%
Level 5	--	--
Level 6		
Average English learners	69%	66%
Average non–English learners	48%	37%

Source: 2020/21 and 2021/22 district-provided screening assessment data, October and June SIMS collection data, and state-provided ACCESS data

Note: Student groups with fewer than 10 students are not shown to protect student privacy.

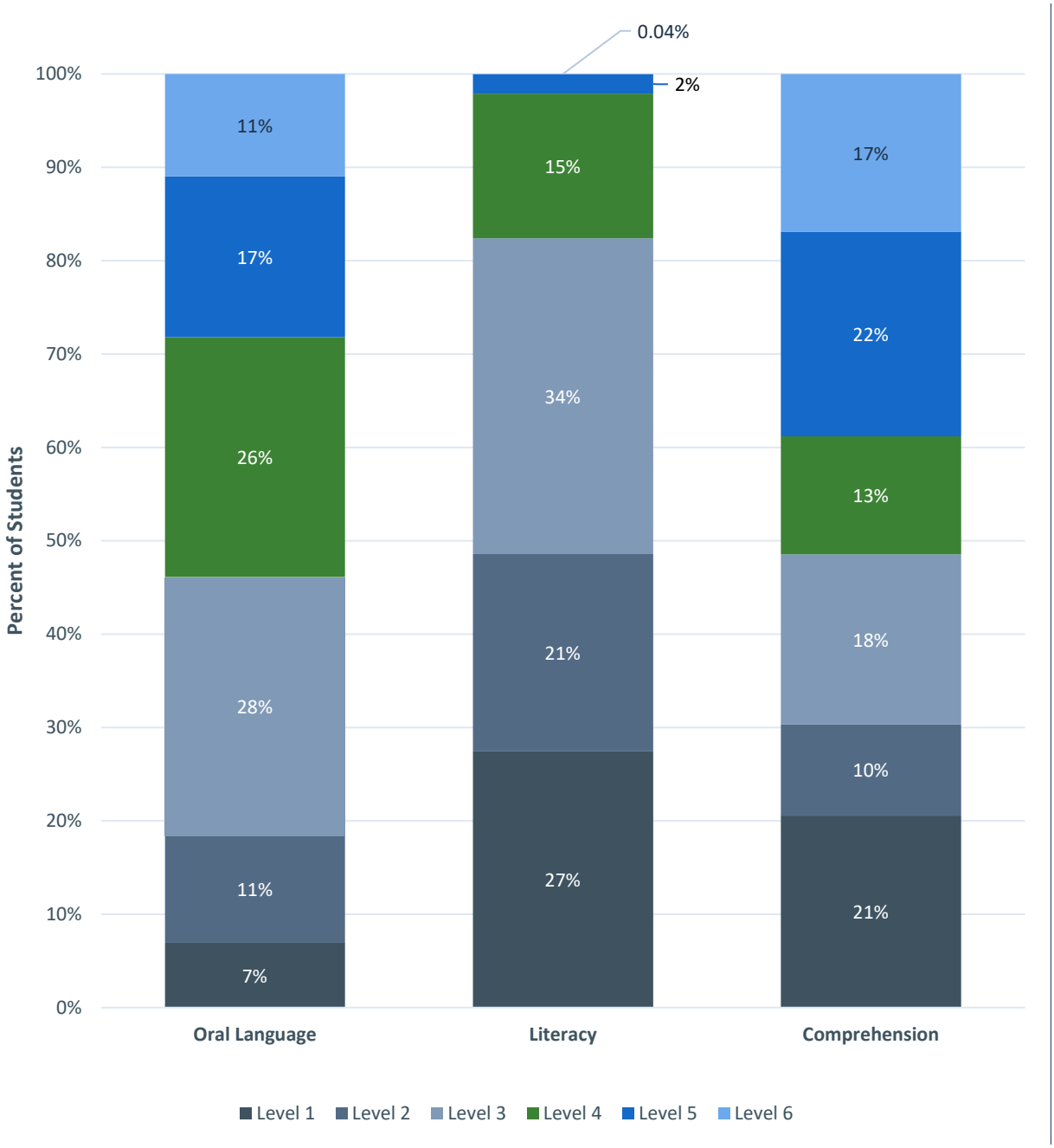
Table E.4. Rate of achieving EOY benchmark by ACCESS composite proficiency level, 2022

ACCESS Composite Proficiency Level	Did Not Meet in MOY	Did Not Meet in EOY
Level 1	91%	85%
Level 2	86%	78%
Level 3	67%	56%
Level 4	32%	23%
Level 5	--	--
Level 6	--	--
Average English learners	77%	69%
Average non–English learners	45%	35%

Source: 2020/21 and 2021/22 district-provided screening assessment data, October and June SIMS collection data, and state-provided ACCESS data

Note: Student groups with fewer than 10 students are not shown to protect student privacy.

Figure E.1. ACCESS level representation within meeting benchmark: Students meeting EOY benchmarks came about equally from lower and higher levels of ACCESS oral language and comprehension measures



Appendix F. Additional Tables: Linking

Tables F.1 through F.8 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 F.1. DIBELS 8th Edition (n = 920) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	453	49.2%
Male	467	50.8%
Asian/Native American/ Pacific Islander	39	4.2%
Black	74	8.0%
Hispanic	321	34.9%
Multiracial	42	4.6%
White	444	48.3%
Low Income: No	336	36.5%
Low Income: Yes	584	63.5%
English Learner: No	742	80.7%
English Learner: Yes	178	19.4%
Student With IEP: No	707	76.8%
Student With IEP: Yes	213	23.2%

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table F.2. FastBridge aReading (n = 180) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	86	47.8%
Male	94	52.2%
Asian/Native American/ Pacific Islander	--	--
Black	--	--
Hispanic	18	10.0%
Multiracial	30	16.7%
White	120	66.7%
Low Income: No	71	39.4%
Low Income: Yes	109	60.6%
English Learner: No	--	--
English Learner: Yes	--	--
Student With IEP: No	148	82.2%
Student With IEP: Yes	32	17.8%

Source: 2021/22 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.

Table F.3. i-Ready (n = 707) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	335	47.4%
Male	372	52.6%
Asian/Native American/ Pacific Islander	38	5.4%
Black	44	6.2%
Hispanic	58	8.2%
Multiracial	35	4.9%
White	532	75.3%
Low Income: No	539	76.2%
Low Income: Yes	168	23.8%
English Learner: No	674	95.3%
English Learner: Yes	33	4.7%
Student With IEP: No	592	83.7%
Student With IEP: Yes	115	16.3%

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table F.4. Lexia Rapid (n = 543) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	263	48.4%
Male	280	51.6%
Asian/Native American/ Pacific Islander	18	3.3%
Black	62	11.4%
Hispanic	270	49.9%
Multiracial	12	2.2%
White	181	33.3%
Low Income: No	216	39.8%
Low Income: Yes	327	60.2%
English Learner: No	425	78.3%
English Learner: Yes	118	21.7%
Student With IEP: No	415	76.4%
Student With IEP: Yes	128	23.6%

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table F.5. MAP Growth (n = 624) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	328	52.6%
Male	296	47.4%
Asian/Native American/ Pacific Islander	242	38.8%
Black	44	7.1%
Hispanic	161	25.8%
Multiracial	16	2.6%
White	161	25.8%
Low Income: No	246	39.4%
Low Income: Yes	378	60.6%
English Learner: No	439	70.4%
English Learner: Yes	185	29.7%
Student With IEP: No	526	84.3%
Student With IEP: Yes	98	15.7%

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table F.6. mCLASS (n = 429) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	217	50.6%
Male	211	49.2%
Asian/Native American/ Pacific Islander	17	4.0%
Black	14	3.3%
Hispanic	42	9.8%
Multiracial	11	2.6%
White	345	80.4%
Low Income: No	258	60.1%
Low Income: Yes	171	39.9%
English Learner: No	405	94.4%
English Learner: Yes	24	5.6%
Student With IEP: No	337	78.6%
Student With IEP: Yes	92	21.5%

Source: 2021/22 district-provided screening assessment data and October and June SIMS collection data

Table F.7. Star Reading (n = 663) student characteristics of grade 3 EOY assessment concordance sample

Student Group	Number	Percent
Female	309	46.6%
Male	354	53.4%
Asian/Native American/ Pacific Islander	46	6.9%
Black	--	--
Hispanic	431	65.0%
Multiracial	--	--
White	135	20.4%
Low Income: No	158	23.8%
Low Income: Yes	505	76.2%
English Learner: No	342	51.6%
English Learner: Yes	321	48.4%
Student With IEP: No	570	86.0%
Student With IEP: Yes	93	14.0%

Source: 2021/22 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.

Figure F.1. Distribution of MCAS Grade 3 ELA scale score for statewide population

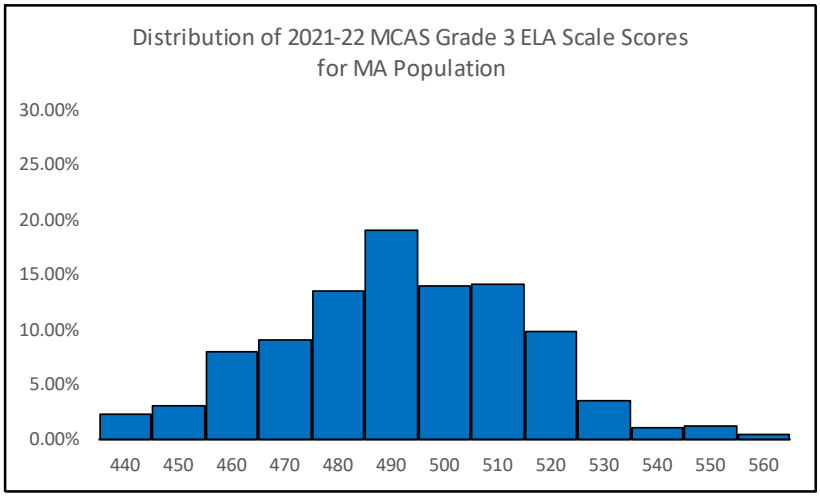


Figure F.2. Distribution of MCAS Grade 3 ELA scale score for DIBELS 8 grade 3 EOY sample

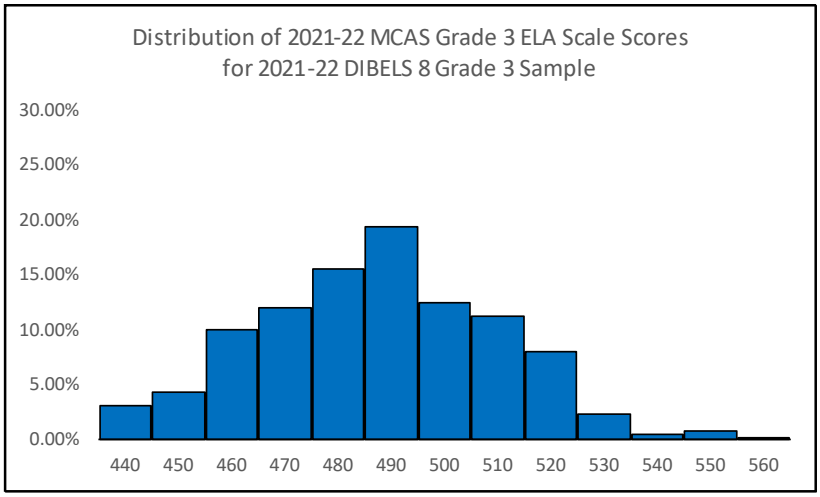


Figure F.3. Distribution of MCAS Grade 3 ELA scale score for FastBridge aReading grade 3 EOY sample

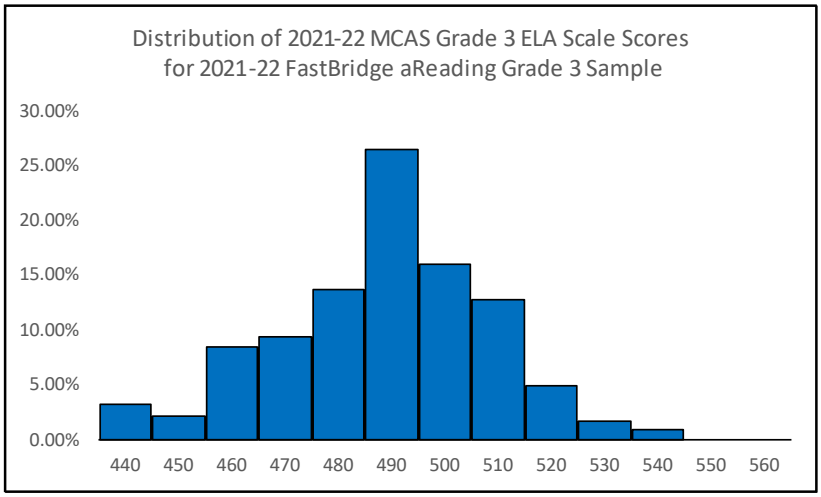


Figure F.4. Distribution of MCAS Grade 3 ELA scale score for i-Ready grade 3 EOY sample

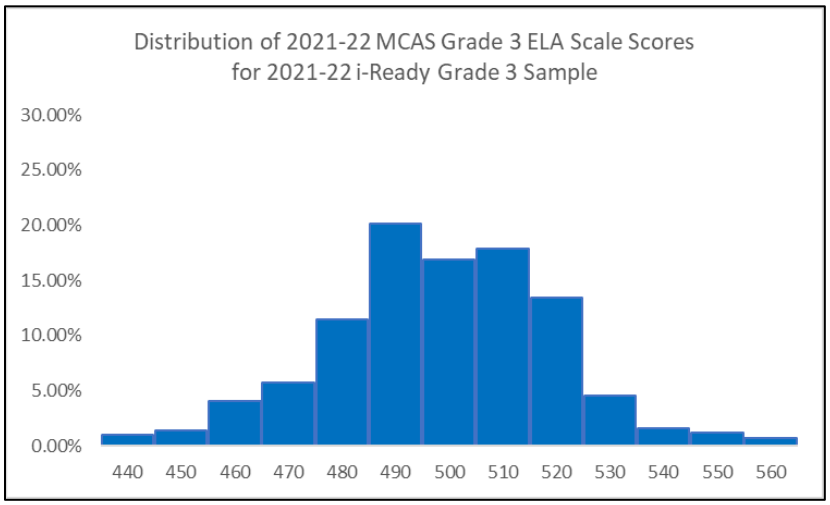


Figure F.5. Distribution of MCAS Grade 3 ELA scale score for MAP Growth grade 3 EOY sample

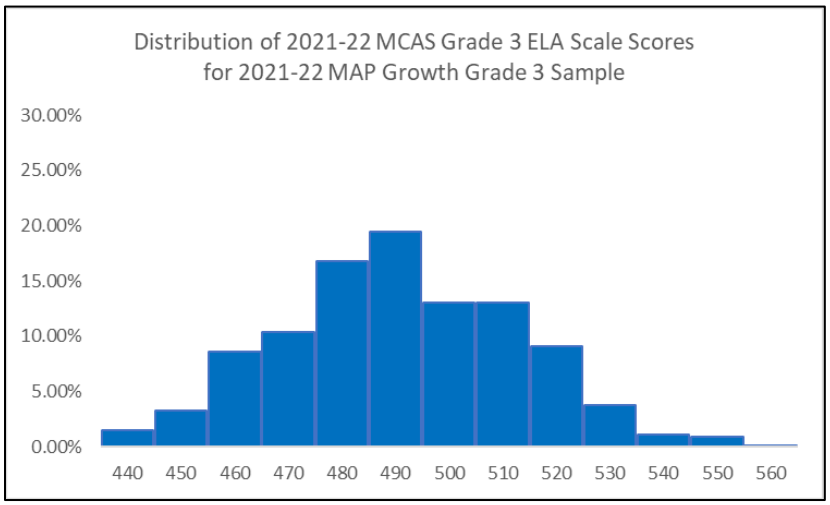


Figure F.6. Distribution of MCAS Grade 3 ELA scale score for Lexia RAPID grade 3 EOY sample

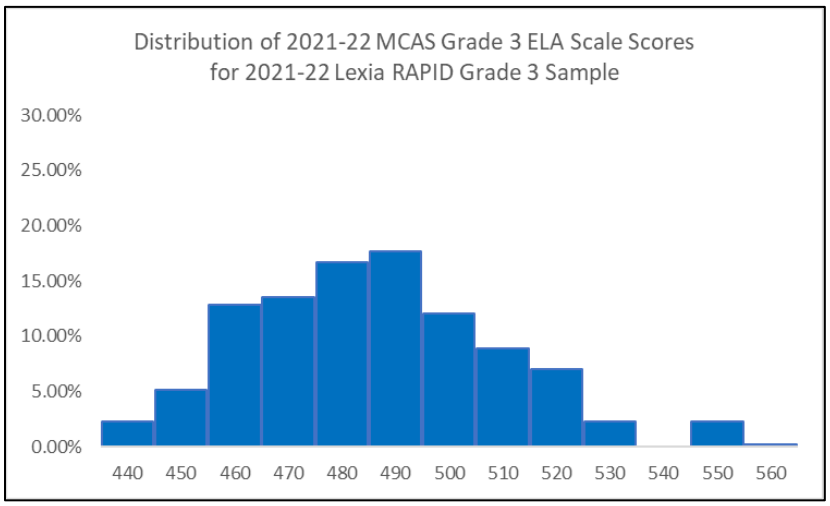


Figure F.7. Distribution of MCAS Grade 3 ELA scale score for mCLASS grade 3 EOY sample

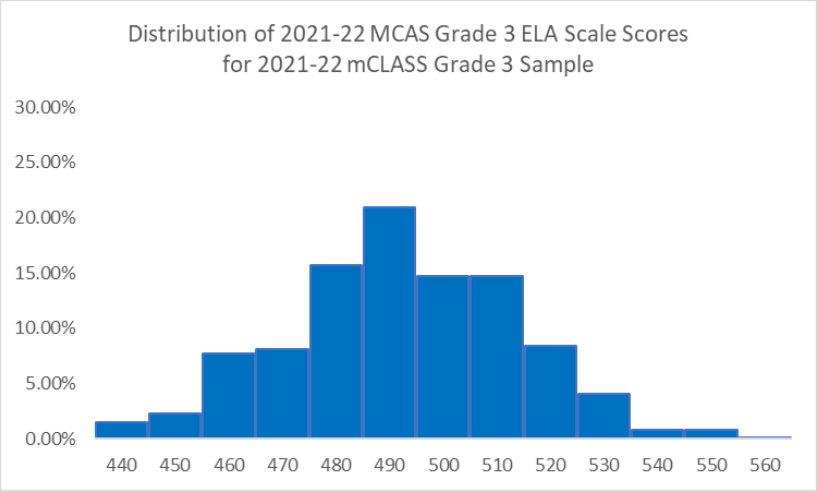


Figure F.8. Distribution of MCAS Grade 3 ELA Scale Score for Star Reading grade 3 EOY sample

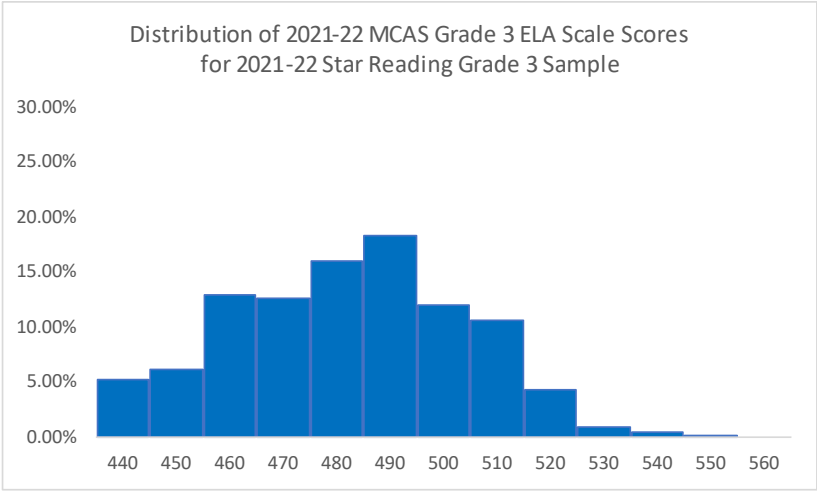
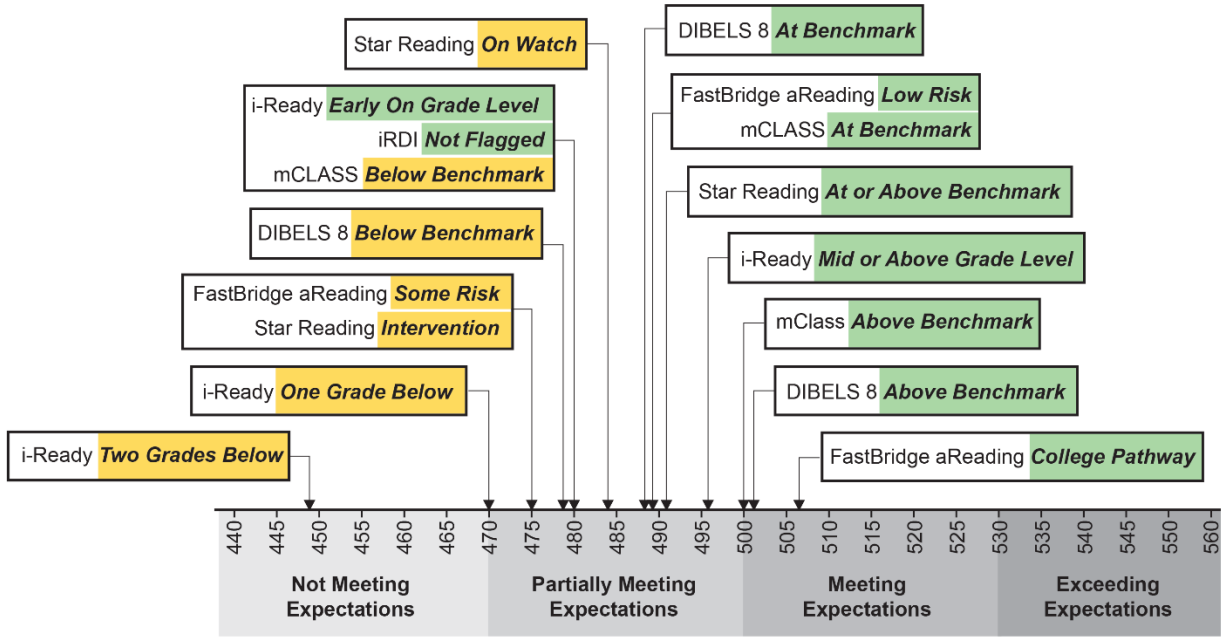


Figure F.9. Mapping of grade 3 early literacy screening assessment benchmark cut scores to MCAS ELA Scores for assessments approved in 2022 review



Source: 2021/22 district-provided screening assessment data and state-provided MCAS data

Note: Benchmarks highlighted in green indicate levels that students must meet/exceed to be classified as meeting benchmark. Benchmarks highlighted in yellow indicate levels that will result in a student being classified as at risk. Students scoring below the yellow benchmarks for mCLASS, DIBELS 8, and FastBridge aReading are classified as being at significant risk according to the respective assessment; students scoring below the "Star Reading Intervention" benchmark or i-Ready's "iRDI Not Flagged" benchmark are classified as being at significant risk according to the respective assessment.

Figure F.10. DIBELS 8th Edition ROC curve for early literacy assessments

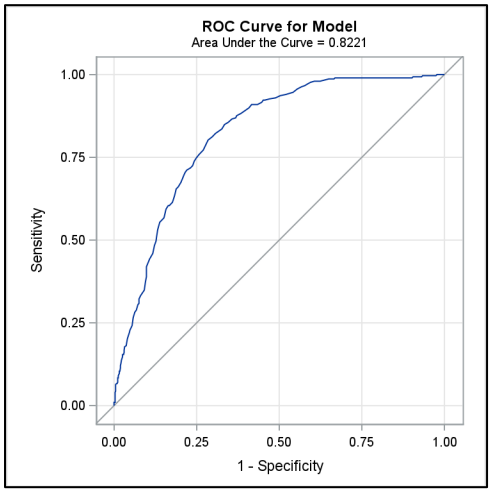


Figure F.11. FastBridge aReading ROC curve for early literacy assessments

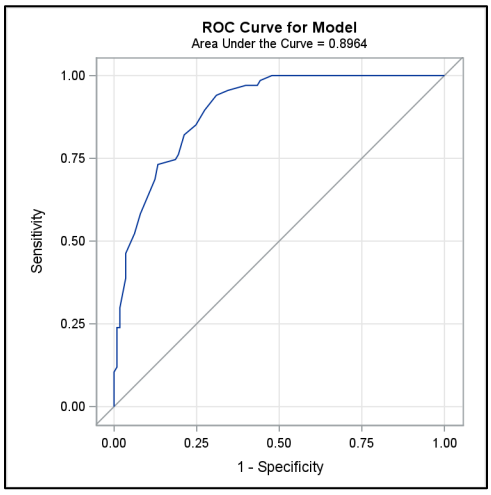


Figure F.12. i-Ready ROC Curve for early literacy assessments

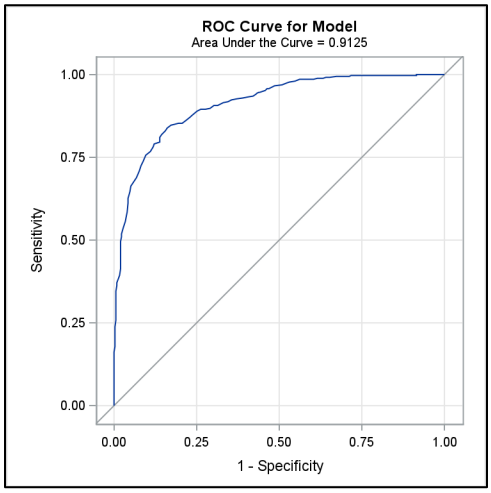


Figure F.13. Lexia RAPID ROC Curve for early literacy assessments

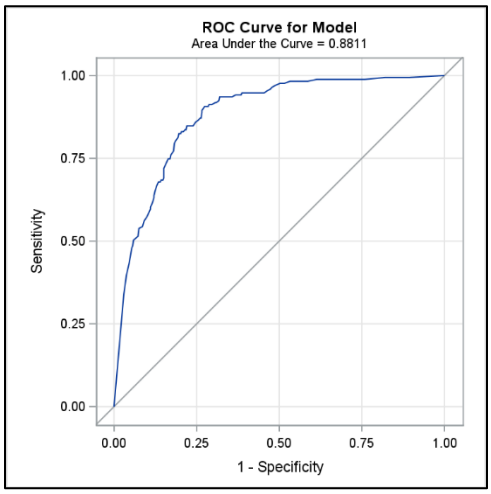


Figure F.14. MAP Growth ROC Curve for early literacy assessments

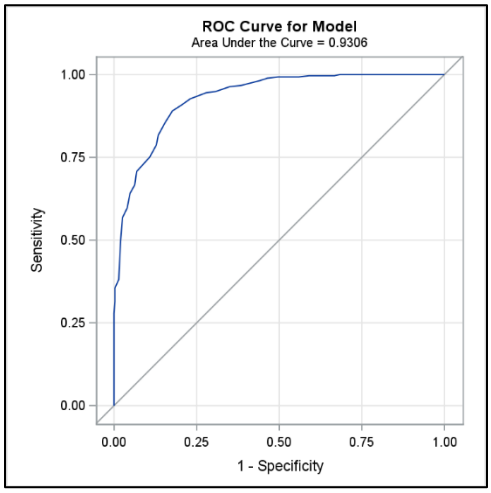


Figure F.15. mCLASS ROC Curve for early literacy assessments

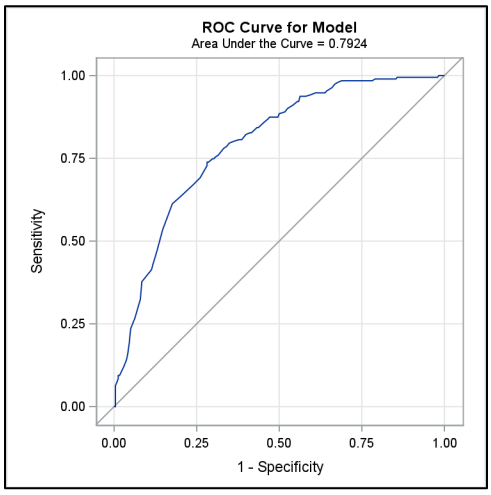


Figure F.16. Star Reading ROC Curve for early literacy assessments