Lincoln Public Schools Fall Literacy and Math Data Roll-Up
October 2022

Introduction

Included in this report are:

1. Beginning of Year (BoY) data for math and literacy in grades 1-5
2. i-Ready math data for grades K-8
3. Comparisons in math and literacy between different data sources including BoY, MCAS, and i-Ready.

As we work to further develop our approach to Key Yearly Measures, we expect winter and spring opportunities to update this data on the progress we see from our students. At the end of the year we anticipate reporting Key Yearly Measures regarding:

- % of students meeting or exceeding expectations in spring in ELA data roll-up
- Progress from fall to spring in ELA data roll-ups
- % of students at or above grade level on spring i-Ready
- % of students who reached their typical and stretch growth goals in i-Ready

This report was written with the intention of being read in totality but in recognition that different stakeholders might care more about specific chapters of the report, the following table of contents provides page numbers for those seeking to learn about a specific component.

Beginning of Year (BoY) Data Grades 1-5

Literacy

Teachers utilize a variety of common assessment tools to gauge students literacy skills throughout a year and across grade levels. In grades 1-5 we collect screening data three times a year (fall, winter, spring) using DIBELS 8 (Dynamic Indicators of Basic Early Literacy Skills) a normed screening assessment, and a developmental spelling inventory (grades 2-5 in the fall), or on-demand writing sample (grade 1 in the fall). There are six subtests within DIBELS 8, and all children in grades K-5 are assessed with either two, three, or four of the screening subtests depending on their grade level and the benchmark interval. Each of the one- minute assessments is administered by a trained member of our screening team which is comprised of literacy specialists, special educators, EL teachers, and speech and language pathologists. Students in grades 2-5 complete an additional three-minute DIBELS screener within their classroom that is administered in a whole-class setting.

Subtest	Skills
Letter Naming Fluency (LNF)	Risk Indicator/Rapid Automatic Naming
Phoneme Segmentation Fluency (PSF)	Phonemic Awareness
Nonsense Word Fluency (NWF)	Alphabetic Principle and Phonics, Decoding/Blending
Oral Reading Fluency (ORF)	Alphabetic Principle and Phonics, Accuracy and Fluency
MAZE (Cloze Passage Assessment)	Accuracy, Fluency, Comprehension
Word Reading Fluency (WRF)	Alphabetic Principle and Phonics, Accuracy and Fluency

For students whose screening data indicates a need for more information (either from low scores or for conflicting information), the literacy specialist team conducts a range of other assessments to dig deeper and this information is shared with classroom teachers as they seek to understand their students and when considering overall levels for students.

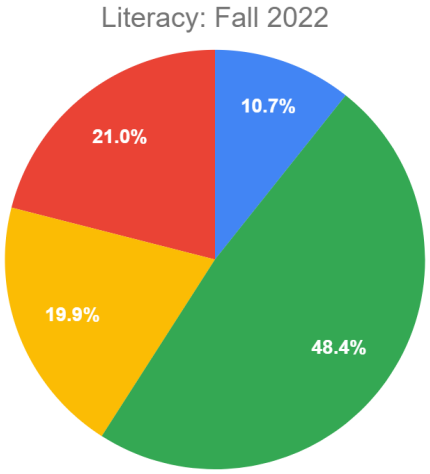
This fall report is based on DIBELS, deeper dive data for students who flagged, and teacher's observations of students in literacy work throughout the first three weeks of school. Through the fall and into the winter, teachers will also utilize the Fountas and Pinnell Benchmark Assessment System (BAS), on-demand writing tasks about their reading and about other topics, as well as ongoing snap/trick word assessments, sentence dictation, running records, and conferences. Kindergarten students are also assessed using Early Bird, a digital tool with 12 subtests across critical foundational skill areas to predict reading ability. For these reasons, our fall data comes from a limited sample but our winter and spring cycles will be informed by many more data points.

Literacy specialists and teachers work together to determine an overall level for each student. We are using the same terms that MCAS uses, though we recognize that this is more an exercise in finding approximate synonyms than in precision and exact translation.

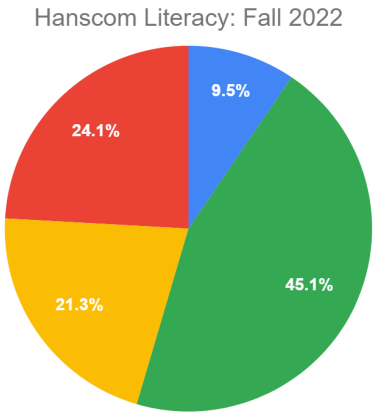
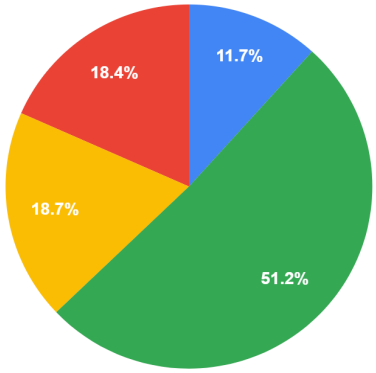
Not meeting expectations	Partially meeting expectations	Meeting expectations	Exceeding expectations

Overall across the district

Using our screening assessments and what teachers observed in students across the first few weeks of school, we found that nearly 50% of the 552 students in grades 1-5 were meeting grade level benchmarks with 10% exceeding benchmark expectations. Of the remaining 40% of students across the district, nearly half of those were slightly below benchmark and the other half more significantly below benchmark.



The charts below illustrate the percentage of students in each of these four campuses, disaggregated by campus. While we often see differences across our two campuses, interestingly this set of fall data looks quite similar.

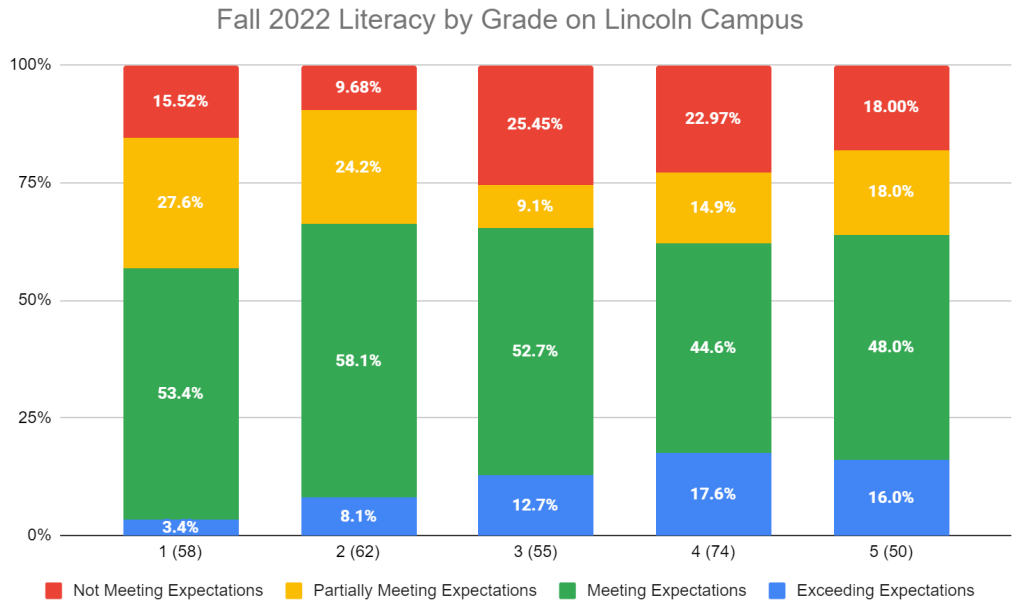
Hanscom Campus (253)	Lincoln Campus (299)
	

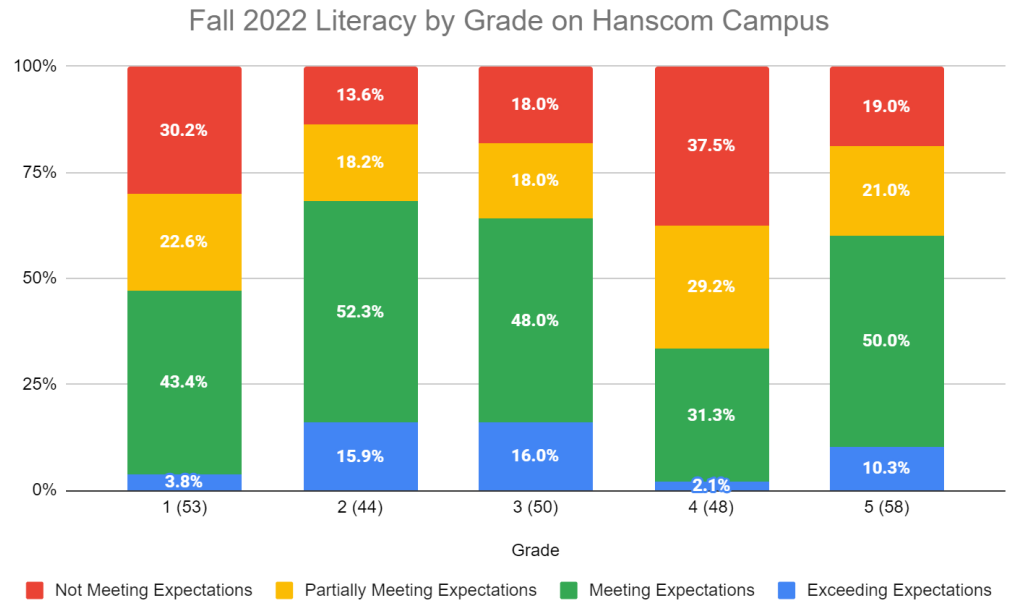
When scores are low and students "flag" it is an invitation to ask more questions to understand a child more deeply. Teachers and specialists collaborate and to identify which students need intervention support in the classroom from their teacher, which students will get intervention services by a specialist outside the classroom, and if any classes require a co-teaching approach between a classroom teacher and literacy specialist. In addition, teachers have been crafting learning goals for each of the students in their classroom, regardless of overall level.

In addition to looking at individual student data, principals and the literacy team also disaggregate data by grade level, racial identity, and other demographics to identify areas of need. Examining the data this way allows principals and the literacy specialist team to consider where resources are best deployed, such as intervention services, the use of Boost Block, and the allocation of classroom assistants.

Distribution Across Grades

The charts below show the distribution of overall levels across grades on each campus. On the Lincoln campus there is a higher percentage of students not meeting benchmark expectations in grades 3 and 4, but there are also a smaller percentage of students partially meeting expectations in those two groups, leaving the overall percentage of students meeting or exceeding benchmark expectations quite consistent across grades 1-5.

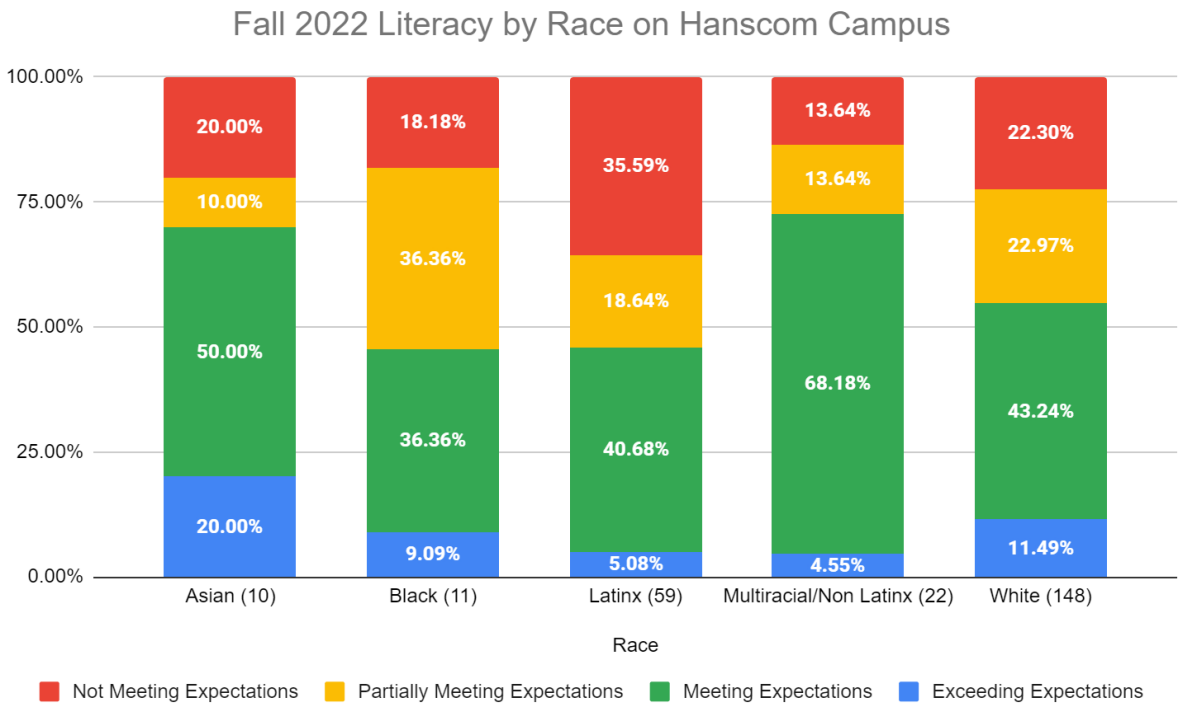




Differences Across Racial Identities -- Hanscom Campus

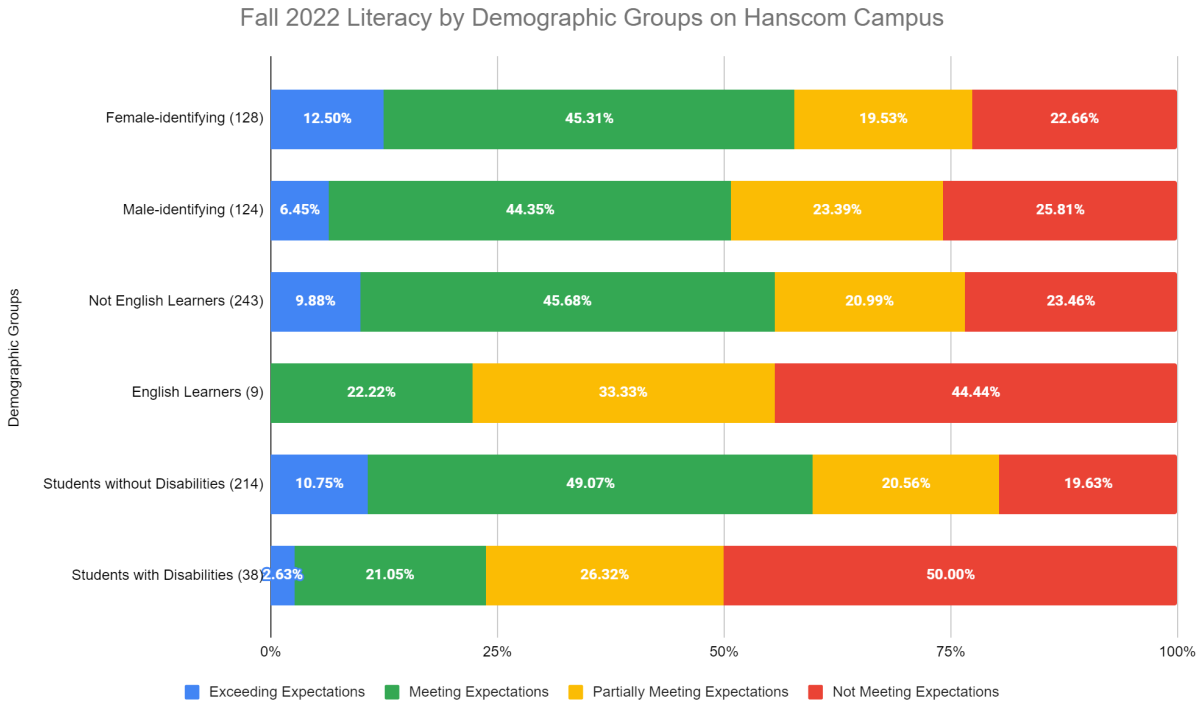
When we disaggregate our fall diagnostic data in literacy by racial identities at Hanscom, we can see differences across groups, just like many of our other data sources, including MCAS and annual surveys. Note that data was suppressed for racial groups whose n-size was below 5 and thus individually identifiable.

It is challenging to draw significant meaning when n-sizes are small since the data is statistically less reliable. For example, for Asian and Black students in grades 1-5 at Hanscom, a single student represents 9-10 percentage points and can significantly impact the overall trend. With higher numbers of Latinx, Multiracial/Non-Latinx, and White students, the results are more reliable. The higher percentage of Latinx students not meeting benchmark expectations stands out as an area for teachers to focus their support on over the next few months.



Differences Across Other Demographics -- Hanscom Campus

We also disaggregated by gender, English learner (EL) status, and special education status. Note that data was suppressed for demographic groups whose n-size was below 5 and thus individually identifiable.

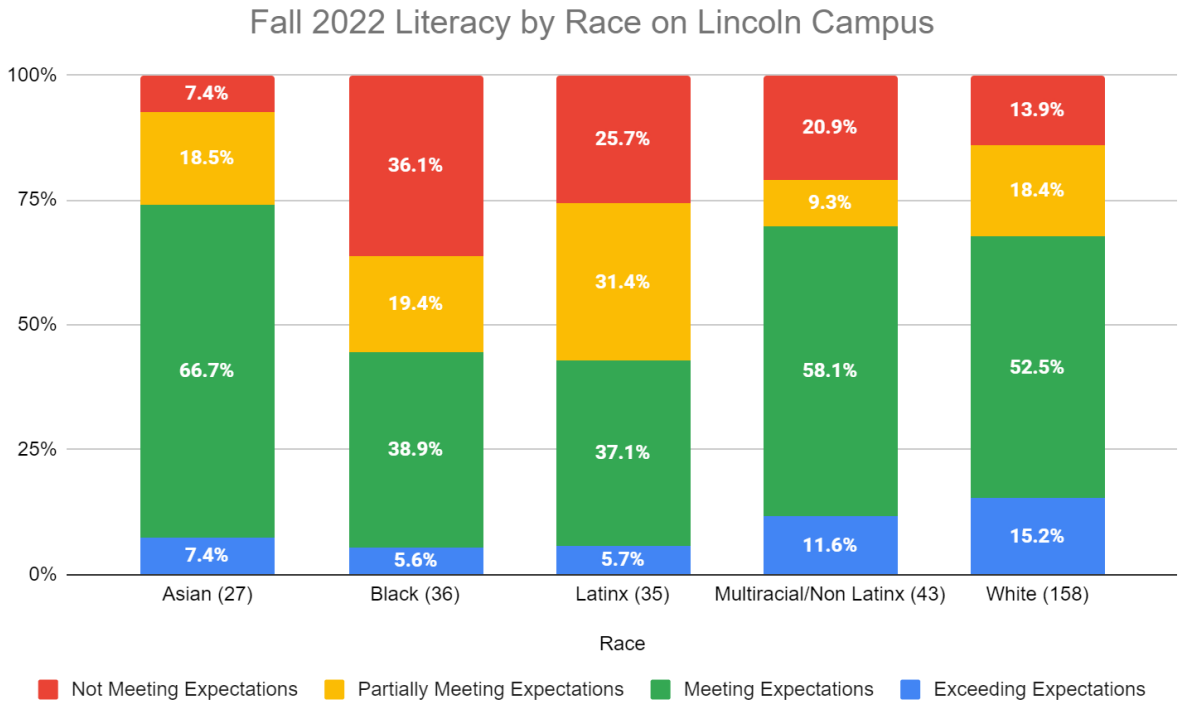


Female-identifying students showed slightly stronger literacy skills than male-identifying students (which we typically see on MCAS data, as well.) Visually, two demographic groups' data stands out: EL students and students with disabilities. In this set of diagnostics, EL students were all assessed in English. For some students who are farther in their language development, this might not have had any significant impact. For others who are starting their English language acquisition, especially if they are newcomers, this task is more challenging and does not illustrate the strength of their literacy skills in their home language. What is even more important to highlight is the n-size for ELs at Hanscom: this group only includes 9 students, so each student constitutes 11% of the total group. For example, one single student not scoring at the level of not meeting expectations and instead scoring at a level that meets expectations would mean that 33% of ELs meet expectations, 33% partially meet expectations, and 33% do not meet expectations, significantly neutralizing the gap.

The other data point that stands out is that 50% of students with disabilities at Hanscom do not meet benchmark expectations in literacy. This group of 38 students includes students with a wide range of disabilities, some related to language and literacy and others for different diagnoses. The n-size is not as small as English Learners but significantly smaller than the other groups listed. 5 students scoring in one category versus another would tip the percentages 13 points. We would expect that students with some disabilities, particularly around language, processing, or executive functioning would flag on a timed screener. In some ways, this data is confirmation that we have the right students on our radar.

Differences Across Racial Identities -- Lincoln Campus

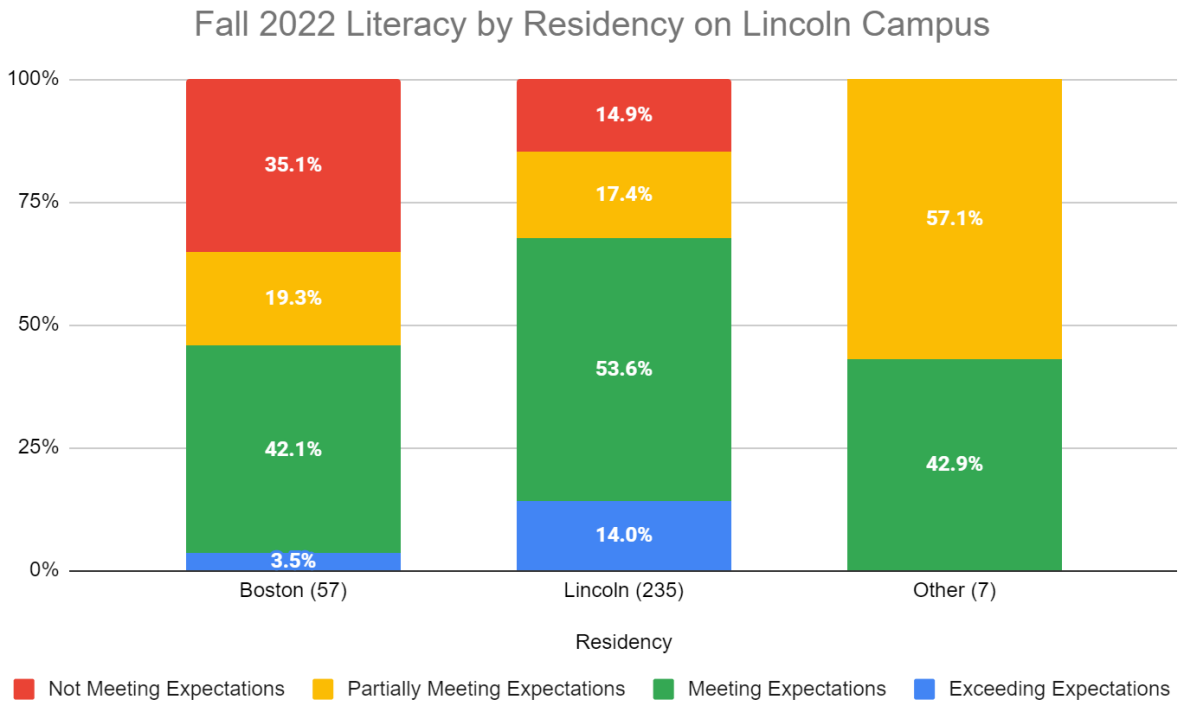
Compared to the Hanscom campus, the number of students in each racial group are more similar, which makes the data more reliable and easier to compare across groups. Note that data was suppressed for racial groups whose n-size was below 5 and thus individually identifiable.



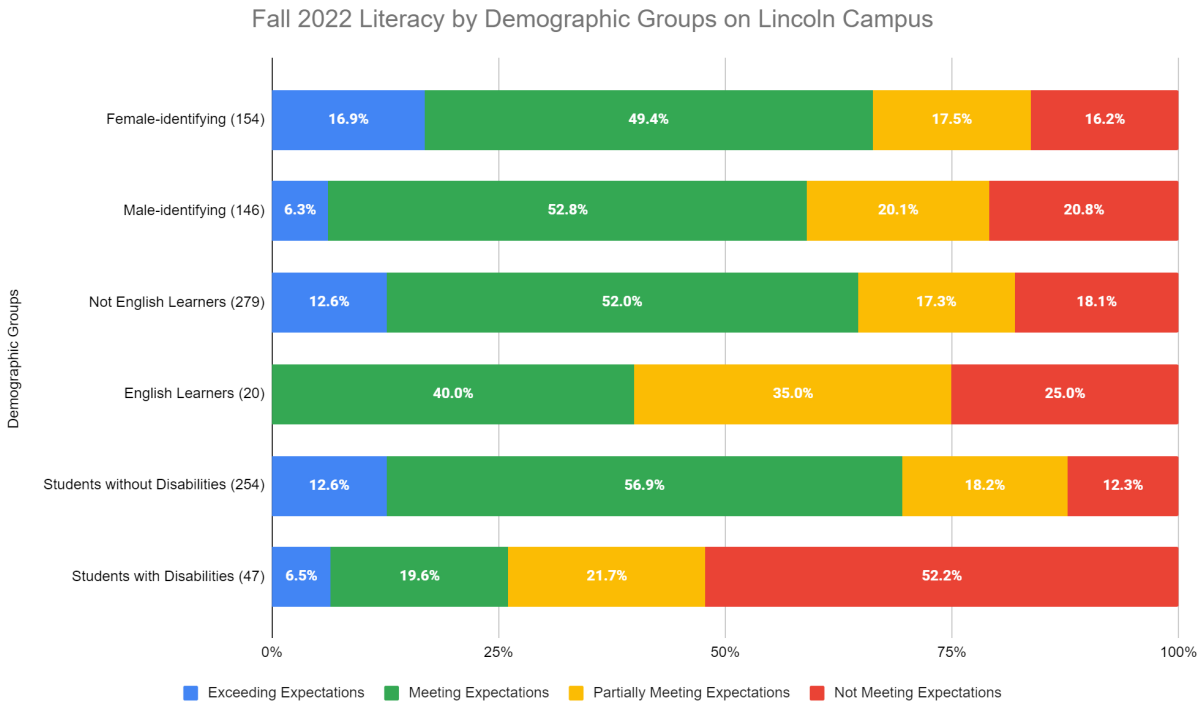
The lower percentages of Black and Latinx students who meet or exceed benchmark expectations on the beginning of year screeners, coupled with the higher percentages of students within these groups who partially met or did not meet benchmark expectations immediately stands out. This is something we will continue to monitor across the fall and into the winter as we examine data from other sources and assessment intervals. We have recently hired a new K-4 Academic Tutor on the Lincoln campus who is working alongside classroom teachers to ensure each student gets what they need to grow and succeed.

Differences Across Other Demographics -- Lincoln Campus

While all the students at Hanscom live on the Base, our students in Lincoln come from three different sets of communities including Boston, Lincoln, and the various towns or cities that staff children reside, each of which has diversity across various spectrums within themselves. The size of the staff children group in grades 1-5 is very small, so the data is significantly less reliable. But, when we compare Boston-resident students to Lincoln-resident students we notice a sizable difference particularly in the percentage of students who are not meeting benchmark expectations.



The difference between female-identifying and male-identifying students on the Lincoln campus is larger than it is on the Hanscom campus. Small n-sizes must be taken into account when examining the scores of EL students. The 25% of students scoring not meeting expectations equate to only 5 students. If just one or two students from each category scored one level higher, then this data would look drastically different. Another way of thinking about this is that 5 EL students make up the group not meeting expectations, while 50 students collectively equal the 18% of students who are not English Learners and are not meeting benchmark expectations in literacy.



Math

For the past few years, teachers in grades 1-5 have used an assessment that is built off of the End-of-Year expectations from the prior grade level to understand students' math skills and knowledge. Essentially, the assessment helps to identify which students are entering, for example, 4th grade with the skills and knowledge they should have left 3rd grade with in the spring. These assessments are scored by a small group of math specialists who use a common set of scoring criteria to increase the validity of the data. If questions arise about a student's skills, math specialists often observe students engaging in math work or conduct additional tasks with them to better understand what areas students have strong foundations in and are facile with applying their knowledge, and where the edge of their understanding lies. This is particularly helpful when math specialists determine which students could benefit from being grouped together for intervention services.

Examining our data collaboratively, with classroom teachers, special educators, math specialists, principals, and other educators all working together, allows each school to consider where resources are best deployed, such as intervention services, the use of Boost Block, and the allocation of classroom assistants, or which strategies should be utilized more, like differentiated small group instruction or specific math games.

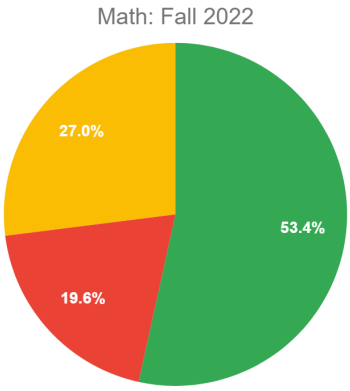
One important distinction between the math and literacy BoY data is that the literacy assessments are nationally normed while our math data is not. The determination of how far away a student is from being on grade level or meeting expectations is far less precise. Connectedly, since the assessments illustrate mastery with the prior grade level's skills, there is essentially no way to demonstrate if a student is exceeding expectations, outside of teacher observations. This is why our BoY math data can only be shown in three categories, rather than four. It is also a contributing factor in why the district adopted i-Ready math diagnostic assessment for grades K-8. Since we needed time to learn i-Ready this fall, we kept our BoY math assessments in place and are reporting on them here, but in future years we expect to eliminate the BoY assessments and only use i-Ready.

Similar to ELA, the data points utilized to determine a student's overall level will grow throughout the year. By the winter cycle, teachers will use the BoY assessment, fall and winter i-Ready, unit assessments, classwork and classroom observations, and small group or intervention progress monitoring when considering holistically what a student knows and can do.

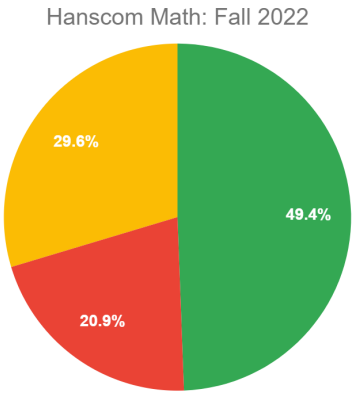
Overall

Using the BoY assessments, slightly more than 53% of the 552 students in grades 1-5 showed they were meeting the expectations required for them to ready for grade level work, 27% partially met expectations and the remaining 19.6% did not meet expectations in skills and knowledge for their grade level. The charts below illustrate the percentage of students across the district and at each campus. We see slightly more students on the Lincoln campus who met expectations for grade level work compared to the Hanscom campus.

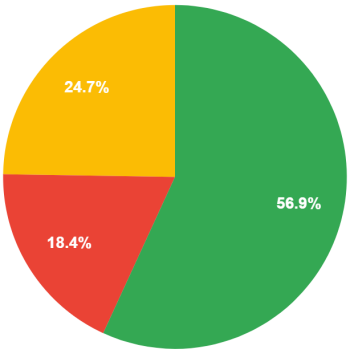
District (352)



Hanscom Campus (253)



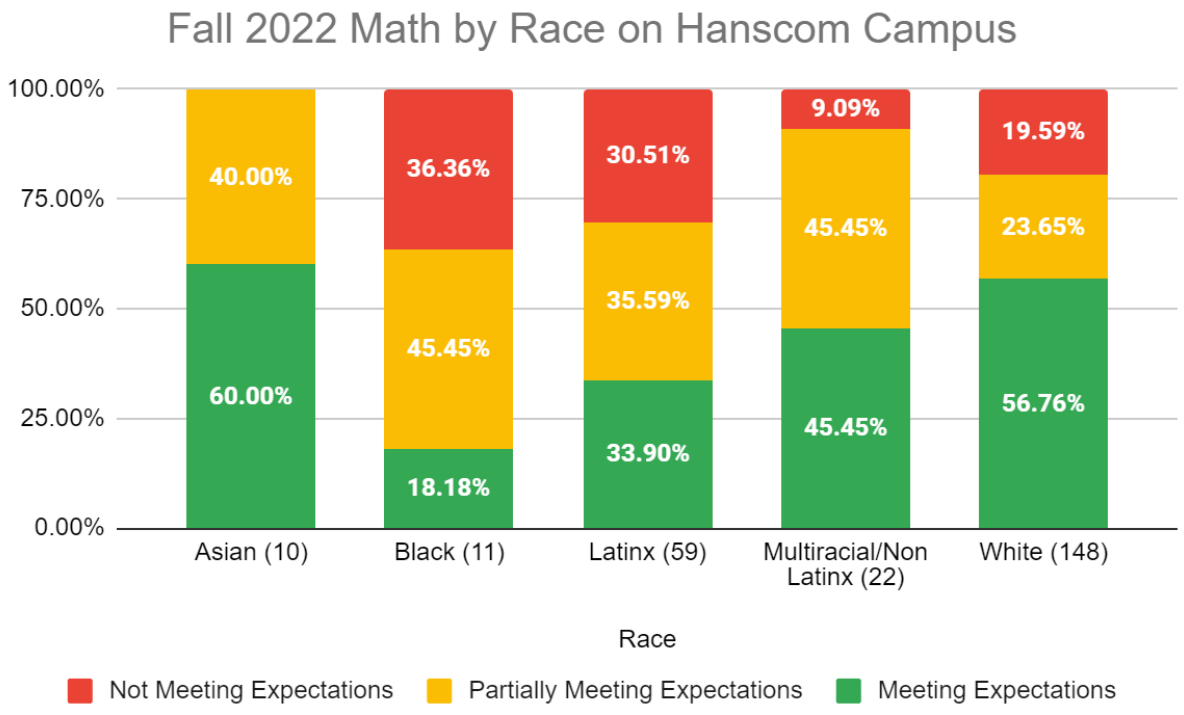
Lincoln Campus (299)



Not meeting expectations	Partially meeting expectations	Meeting expectations

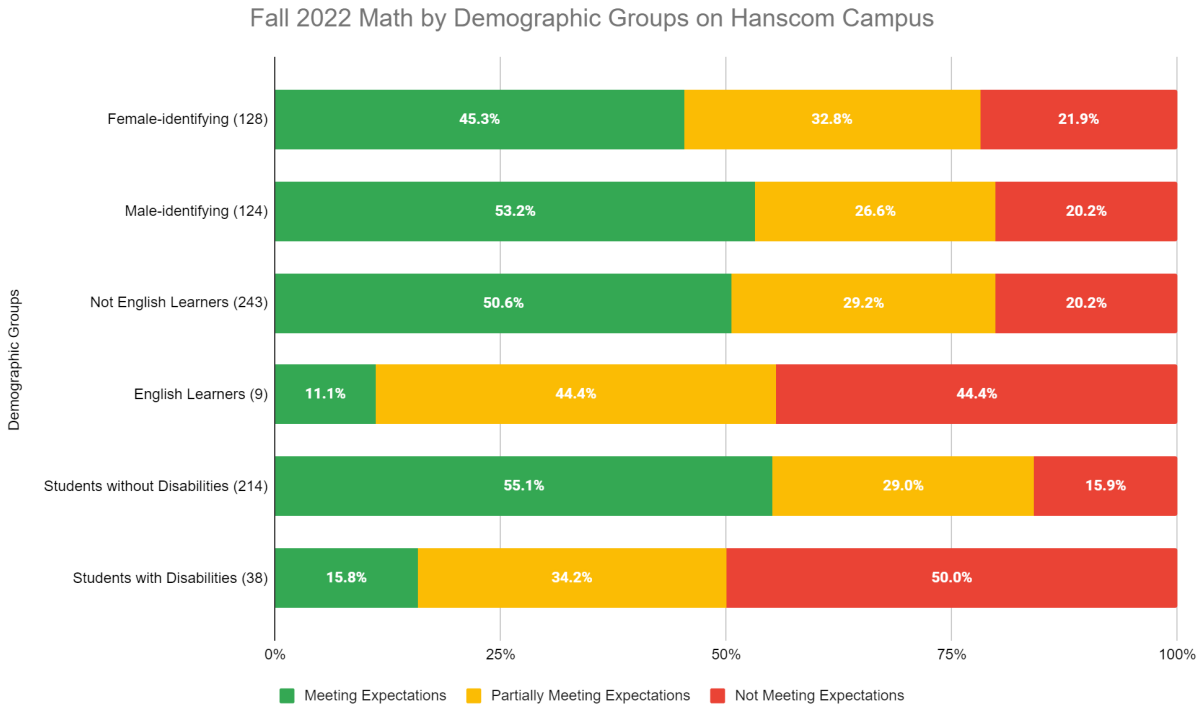
Differences Across Racial Identities -- Hanscom Campus

Similarly to our literacy data, it is critical that we take into account the impact of low n-sizes when considering the reliability of data, particularly for Asian and Black students, each of whom accounts for 9-10 percentage points within each of their respective columns below.



Differences Across Other Demographics -- Hanscom Campus

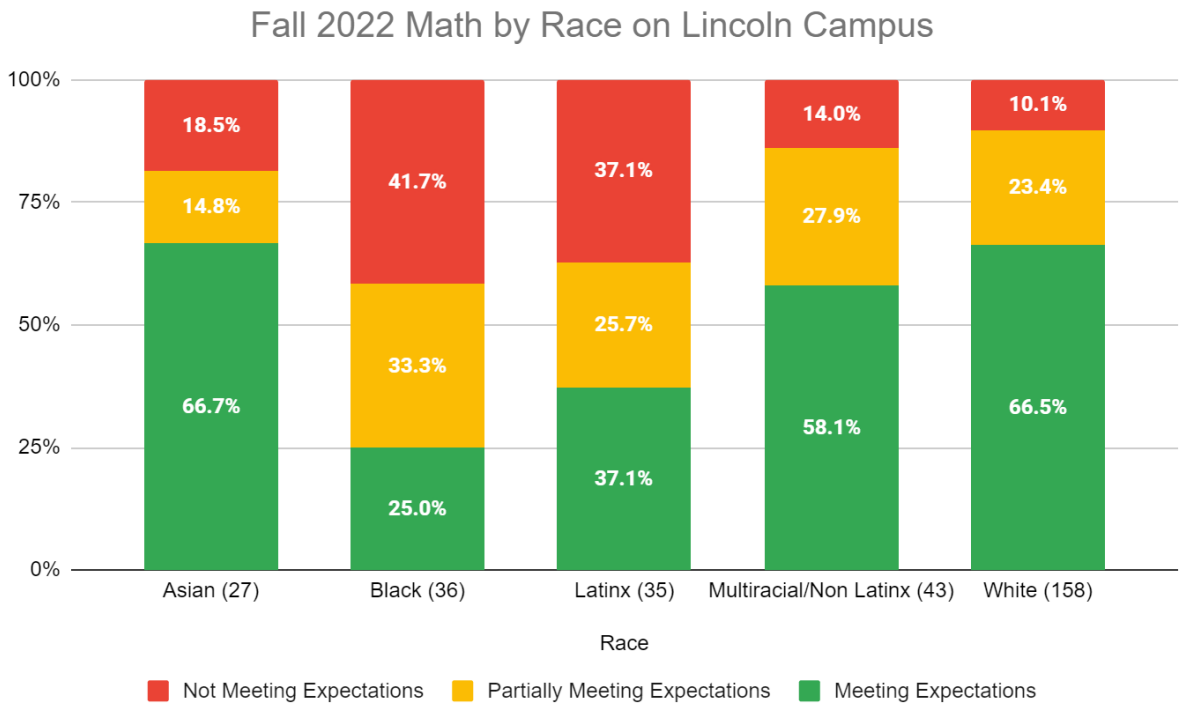
In literacy, we typically see in our internal data and on MCAS that female-identifying students slightly outperform male-identifying students; the inverse is true in math and our BoY data below illustrates this. We see that English Learners had the highest percentage of students partially meet or not meet expectations, but we also have to keep in mind that with such a small group of students this data is highly influenced by each individual child. If two students from each of the partially met or did not meet levels had shown they met expectations, this bar would look remarkably similar to the other subgroups.



Just like in literacy, we notice that 50% of students with disabilities did not meet expectations on the math skills expected for their grade level. On the Hanscom campus we did not notice a significant compounding impact of disability and race when we dug into this group of students.

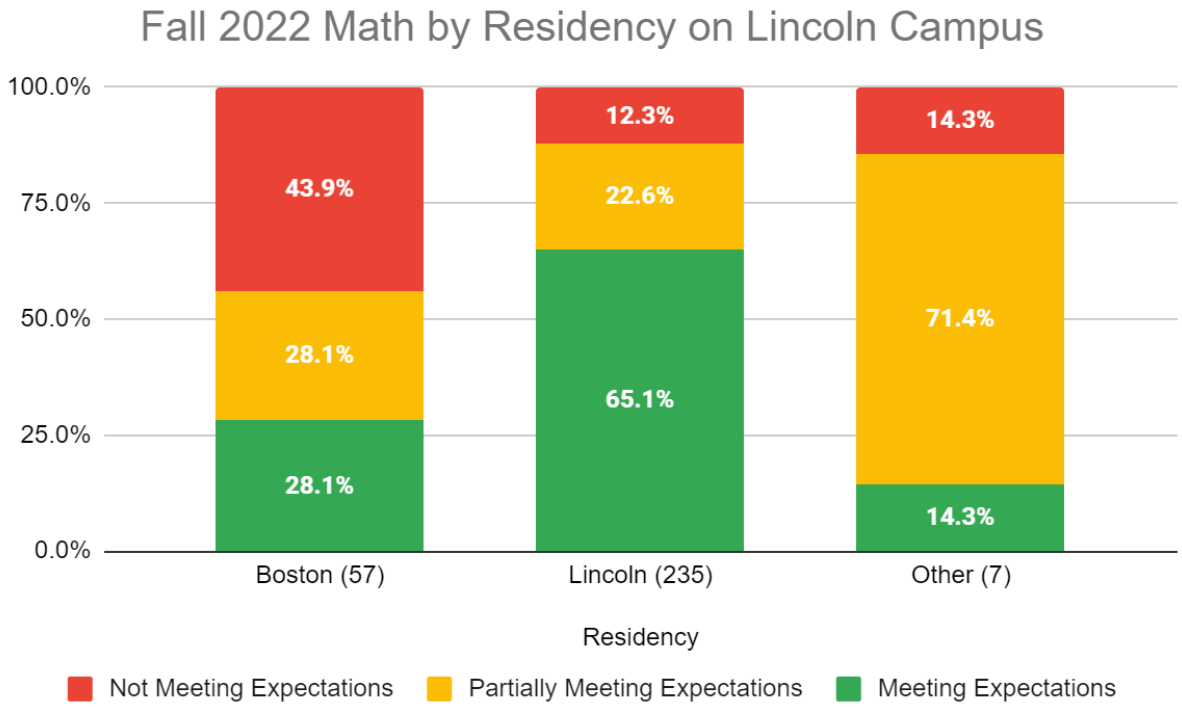
Differences Across Racial Identities -- Lincoln Campus

When we examine data across racial groups we see that there are not significant differences between Asian, Multiracial/Non-Latinx, and White students. However, we do see disparities for Latinx and Black students, in which each of these groups has a lower percentage of students meeting expectations for their grade level and higher percentages of students not meeting those expectations.

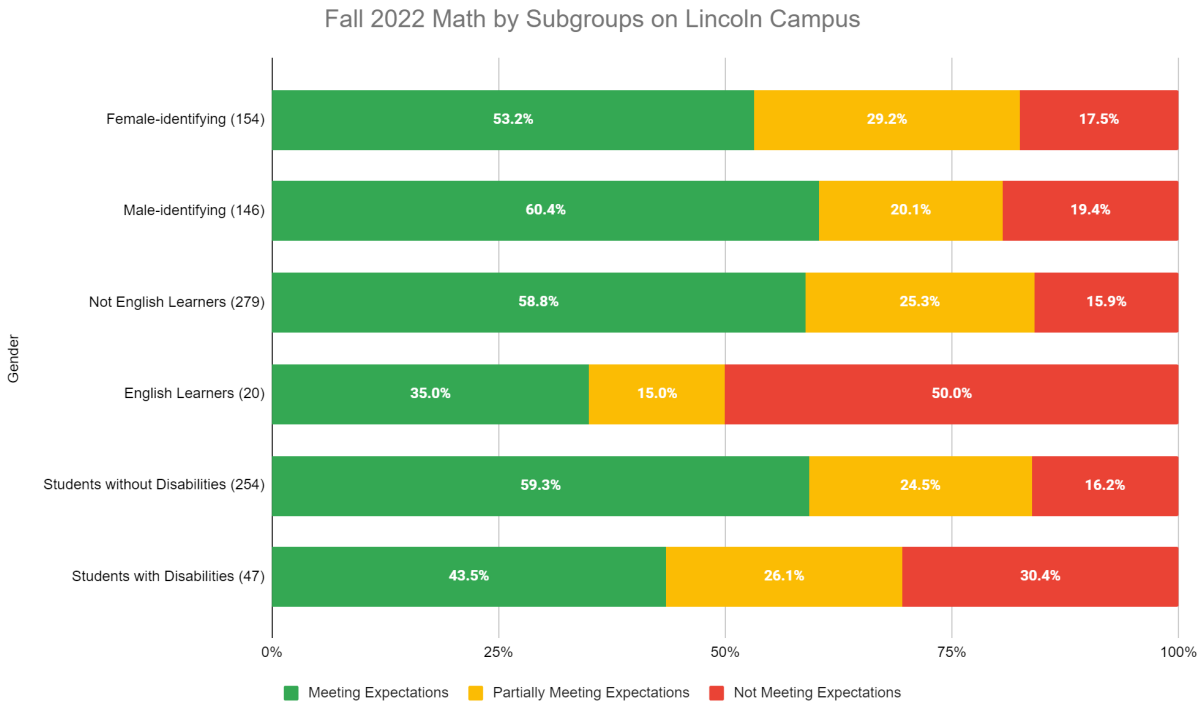


Differences Across Other Demographics -- Lincoln Campus

Besides looking across racial groups, we also look at residency, gender, English learner (EL), and disability status on the Lincoln campus. A much lower percentage of Boston-resident students met grade level expectations compared to Lincoln-resident students. While 25 Boston-resident students and 29 Lincoln-resident students were identified as not meeting expectations, there is a significant difference in the overall percentage of each group, since there are four times as many Lincoln-resident students in grades 1-5 than Boston-resident students.



In math at Lincoln, the gap between students with disabilities and students without disabilities was smaller than at Hanscom, and smaller than the gap that exists in literacy. While a significant 16 percentage point gap still exists, it suggests an area for further inquiry.

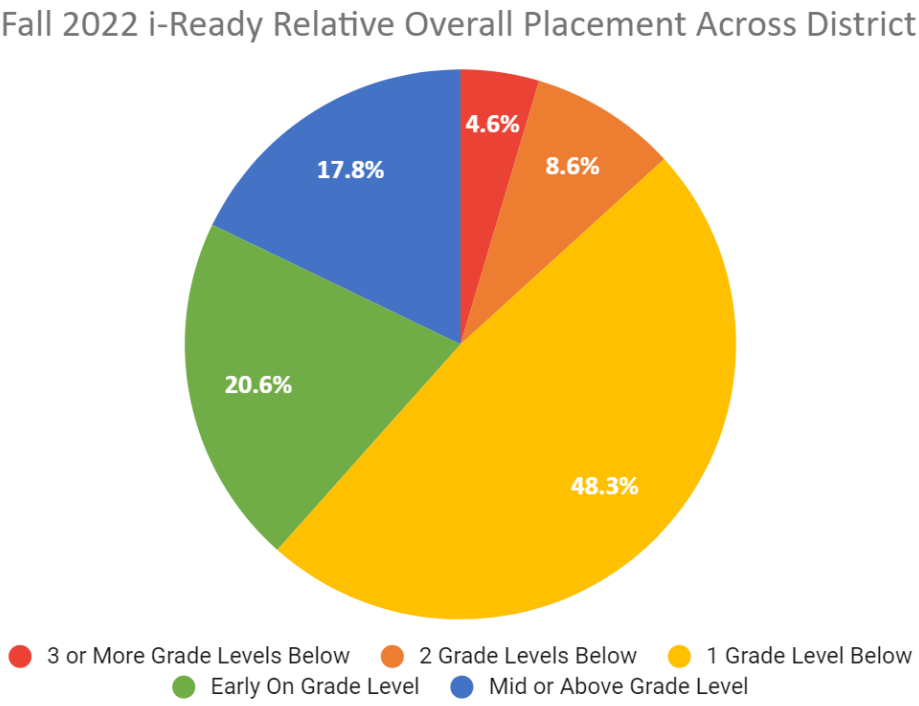


i-Ready Grades K-8

This fall, we adopted the i-Ready math diagnostic assessment for all students in kindergarten through eighth grade. i-Ready is taken on an iPad and is adaptive, meaning it matches the difficulty of questions to the proficiency level of each student. As students answer questions correctly, they are given more difficult questions. As students answer questions incorrectly, they are given easier questions. This means that by design each student will get about 50% of the questions they see correct and about 50% of the questions incorrect. We coached students to expect this experience, and our catchphrase was, "try your best and move on," since we didn't want students to get stuck and spend longer on the assessment than is intended or overly stress about the experience. The assessment provides teachers with a set of data on each student as well as a variety of suggestions for instructional groupings and resources for teaching these groups.

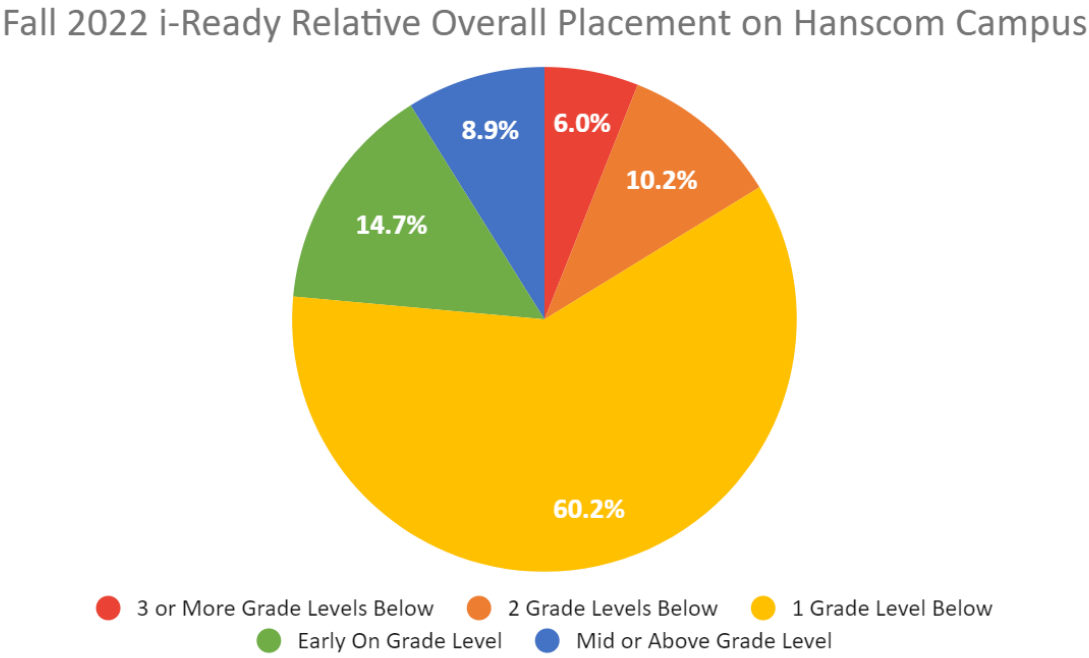
Overall across the district

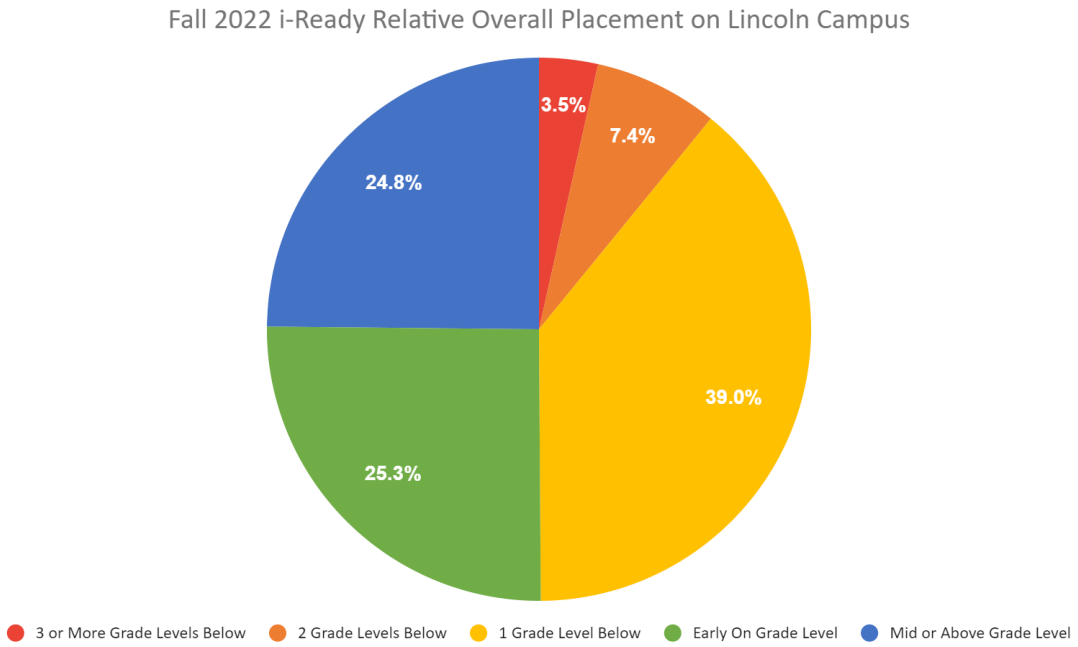
The chart below illustrates the relative overall placement for K-8 students across the district:



The biggest notable difference between our BoY assessment data and i-Ready data is that nearly half of our students tested one grade level below. After looking at raw scores, it is clear that some of these students are very close to being on grade level while others are farther behind. Only 13% of students scored 2 or 3 grade levels below.

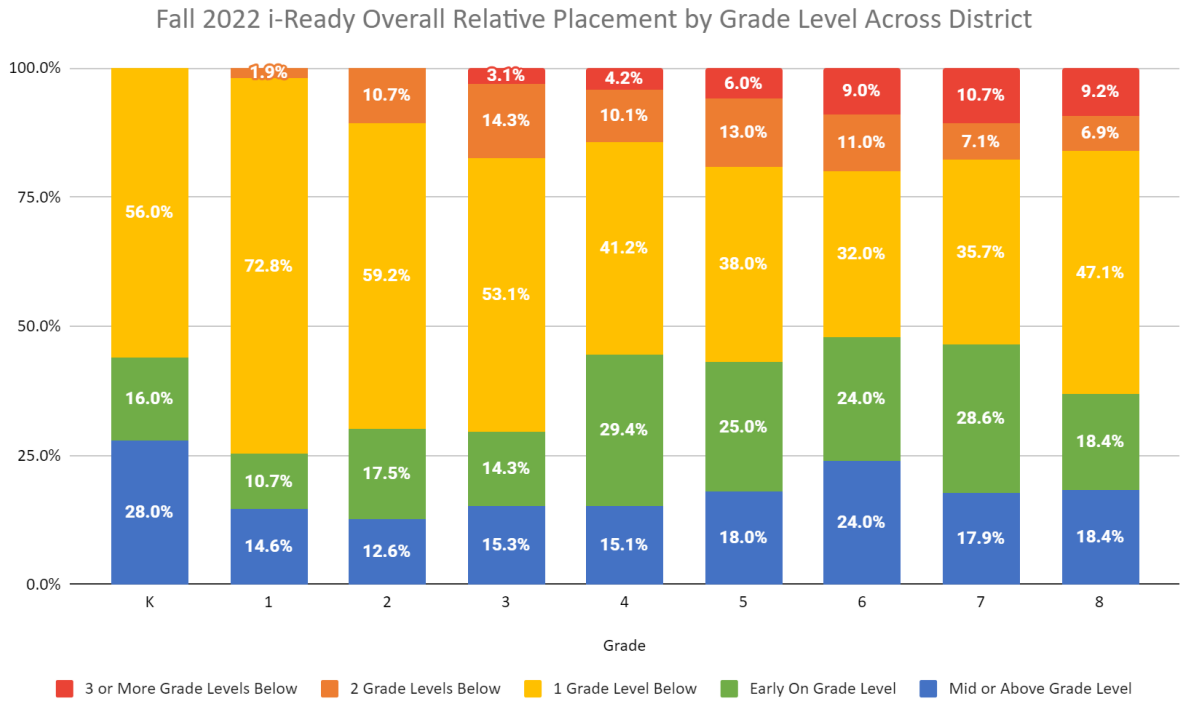
The breakdown of relative overall placements differs between the two campuses. While the percentages of students scoring 2-3 grade levels below are not significantly different, the percentages of those 1 grade level below, early on grade level, and those mid or above grade level are quite different across campuses, with far more students on the Lincoln campus assessing at higher levels.



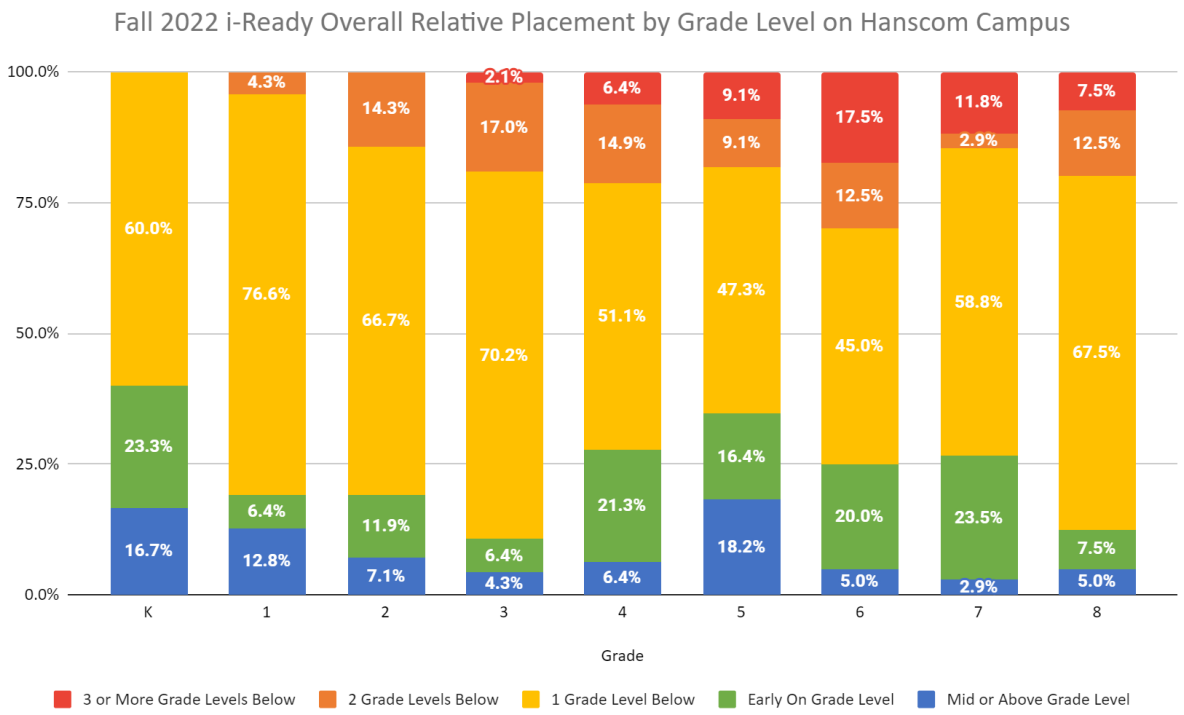


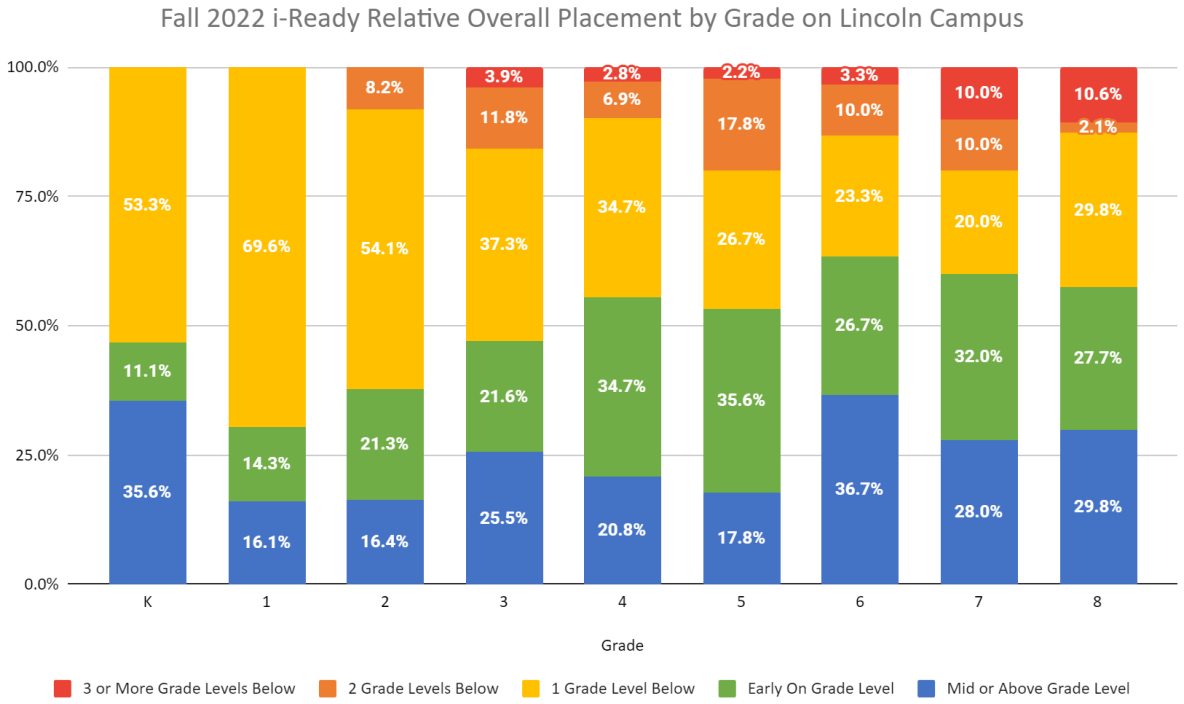
Distribution Across Grades

We examined the scores of students across grade levels. Students in the upper grades 4-8 scored Early On Grade Level and Mid or Above Grade Level at higher rates than students in grades 1-3. Students in Kindergarten can only be considered one grade level below if they are not early on grade level. Similarly, students in first grade can only be one or two grade levels below if they are not early on grade level.



Just like with our overall results, differences by grade level between campuses can be seen, with students at Lincoln performing early on grade level or above at significantly higher percentages.

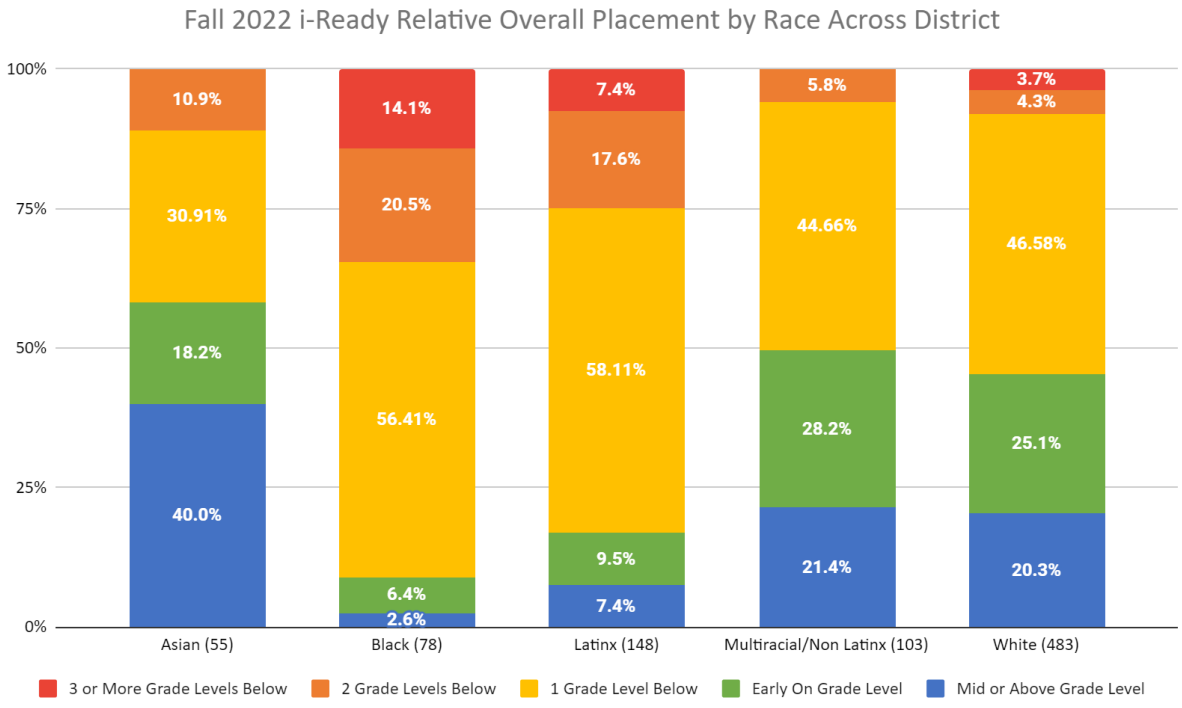


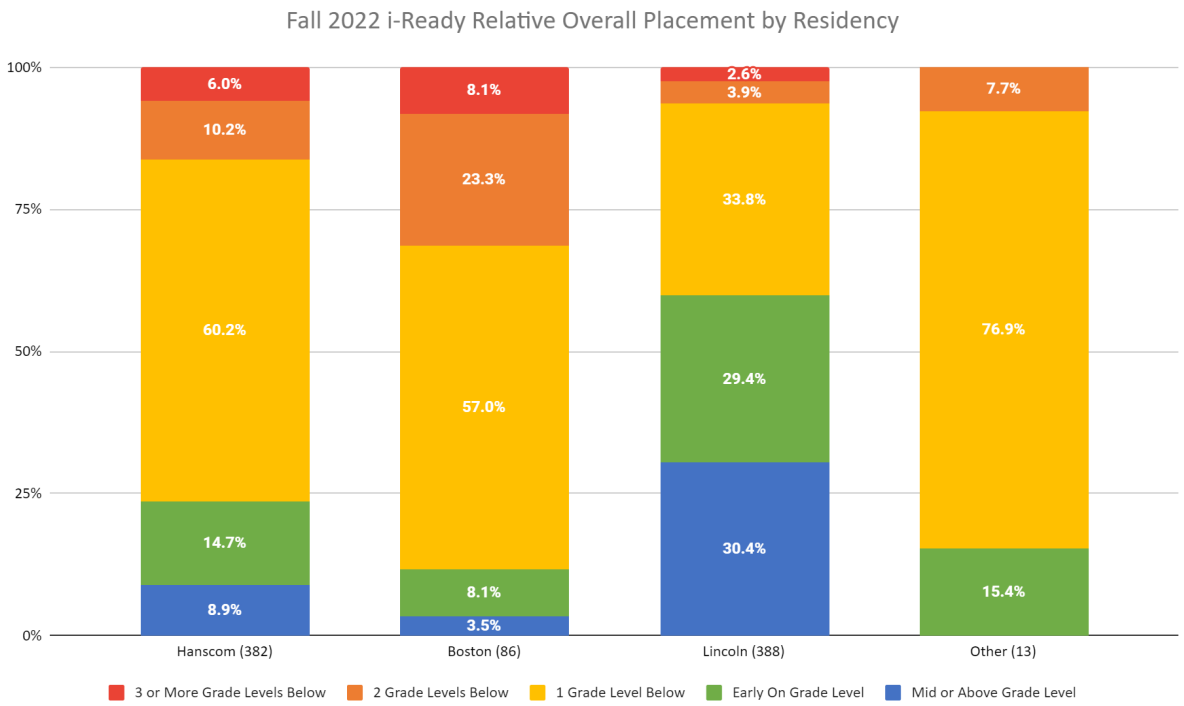


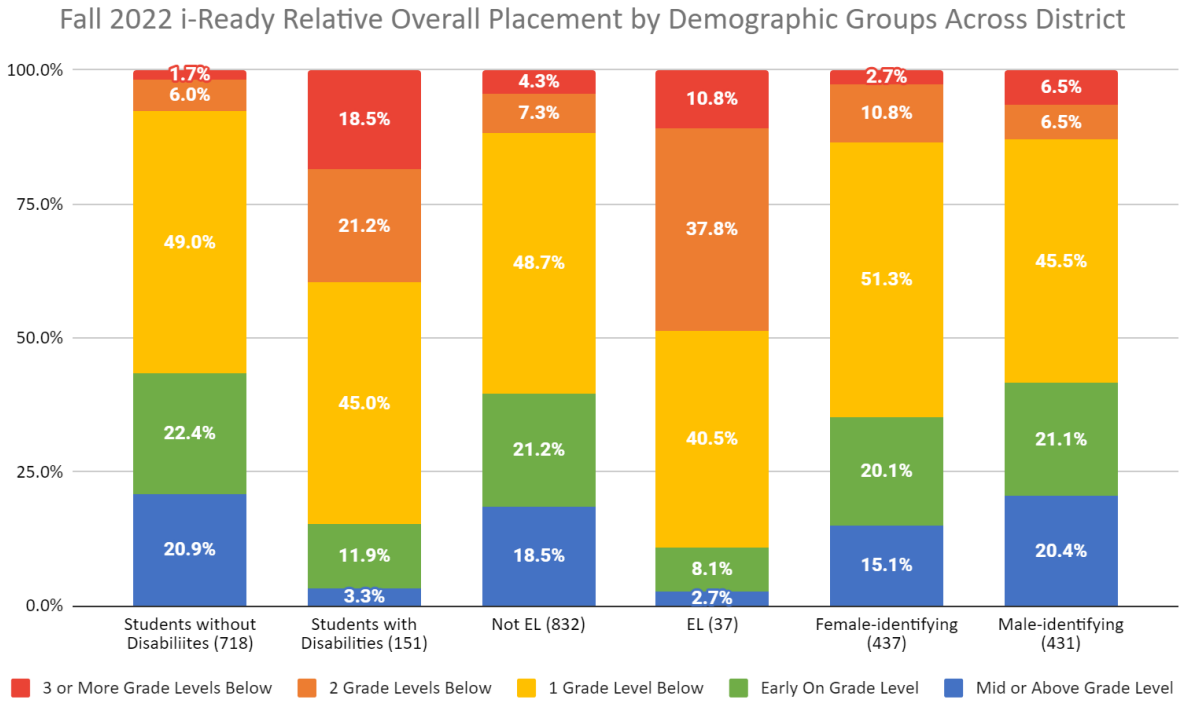
One thing we were concerned about going into i-Ready was if students would rush through the assessment; we had heard from other districts that when this happens it makes it harder to believe in the reliability of the data. Fortunately, our teachers did an excellent job framing the assessment for students and supporting them while they completed it. Across the entire district's 869 K-8 students, only 11 received a "yellow" rush flag, meaning they may have been going through the assessment very quickly and only 4 received a "red" rush flag, meaning they were likely going very quickly. When we examined these students to see if these students scored very low and thus, perhaps their data was not accurate, we discovered that these 16 students scored in all 5 placement levels, ranging from 3 or more grade levels below up to mid or above grade level.

Differences Across Demographic Groups -- District Level

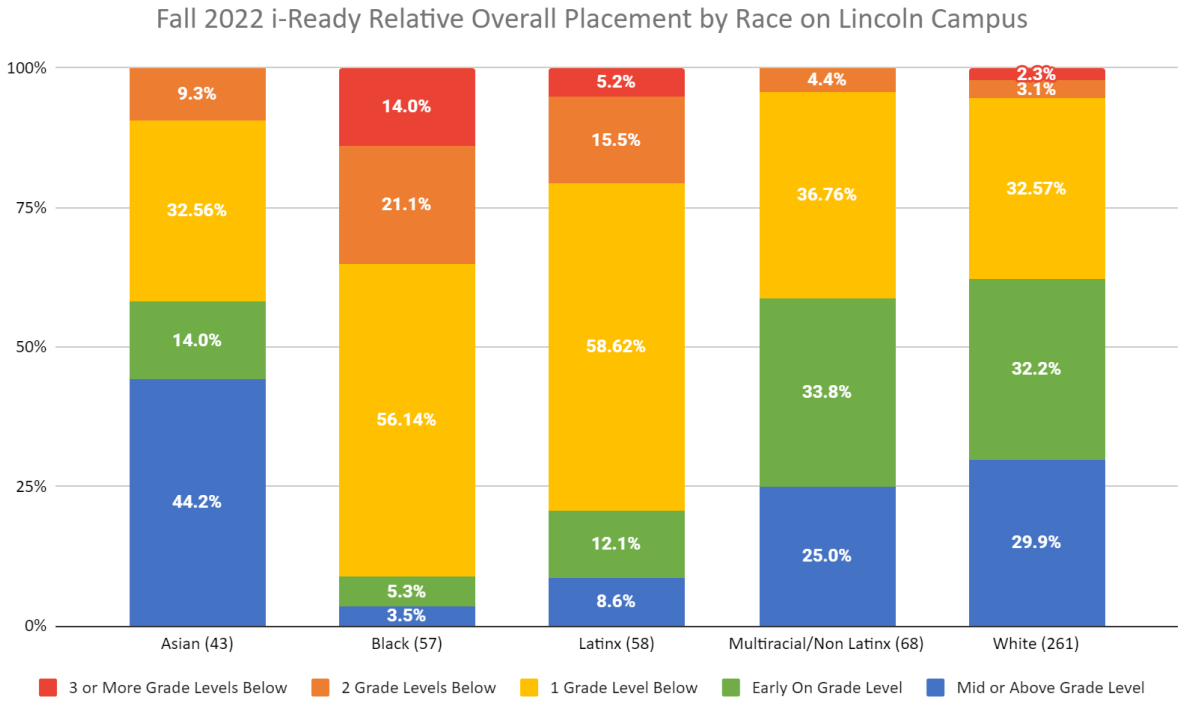
At the district level we see gaps by racial groups and other demographic groups, similar to other assessments.

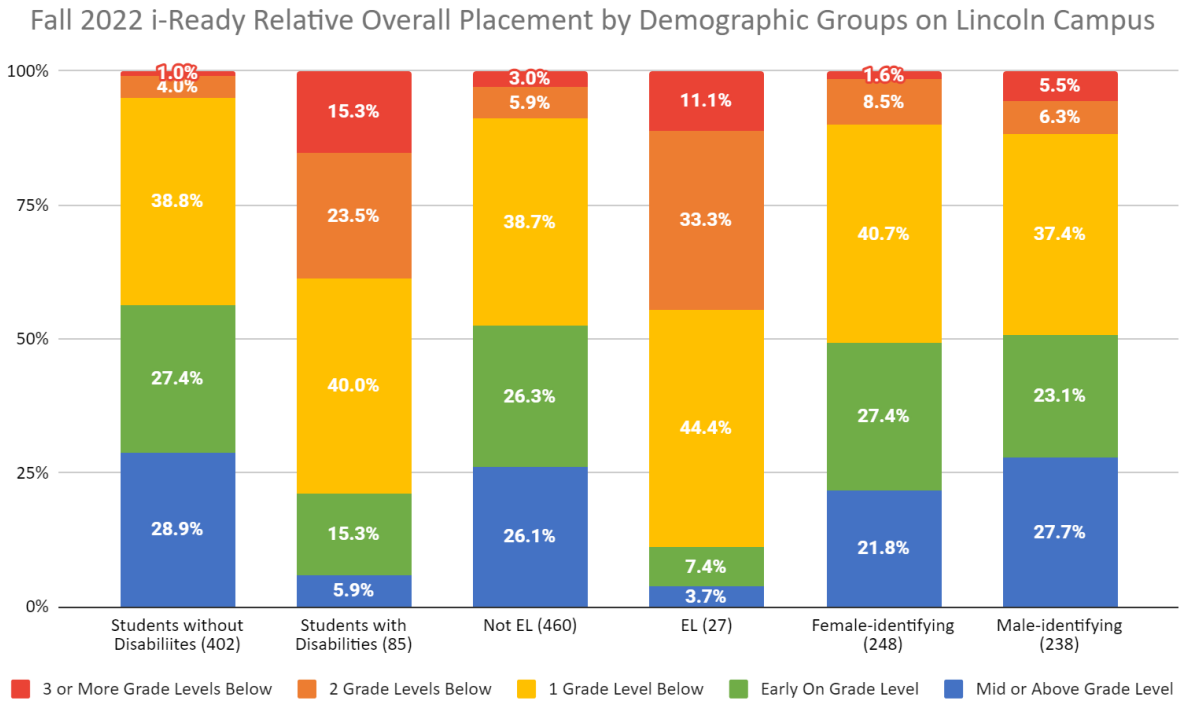



Fall 2022 i-Ready Relative Overall Placement by Residency
100%
75%
50%
25%
0%
6.0%
10.2%
60.2%
14.7%
8.9%
8.1%
23.3%
57.0%
8.1%
3.5%
2.6%
3.9%
33.8%
29.4%
30.4%
7.7%
76.9%
15.4%
Hanscom (382)
Boston (86)
Lincoln (388)
Other (13)
3 or More Grade Levels Below
2 Grade Levels Below
1 Grade Level Below
Early On Grade Level
Mid or Above Grade Level

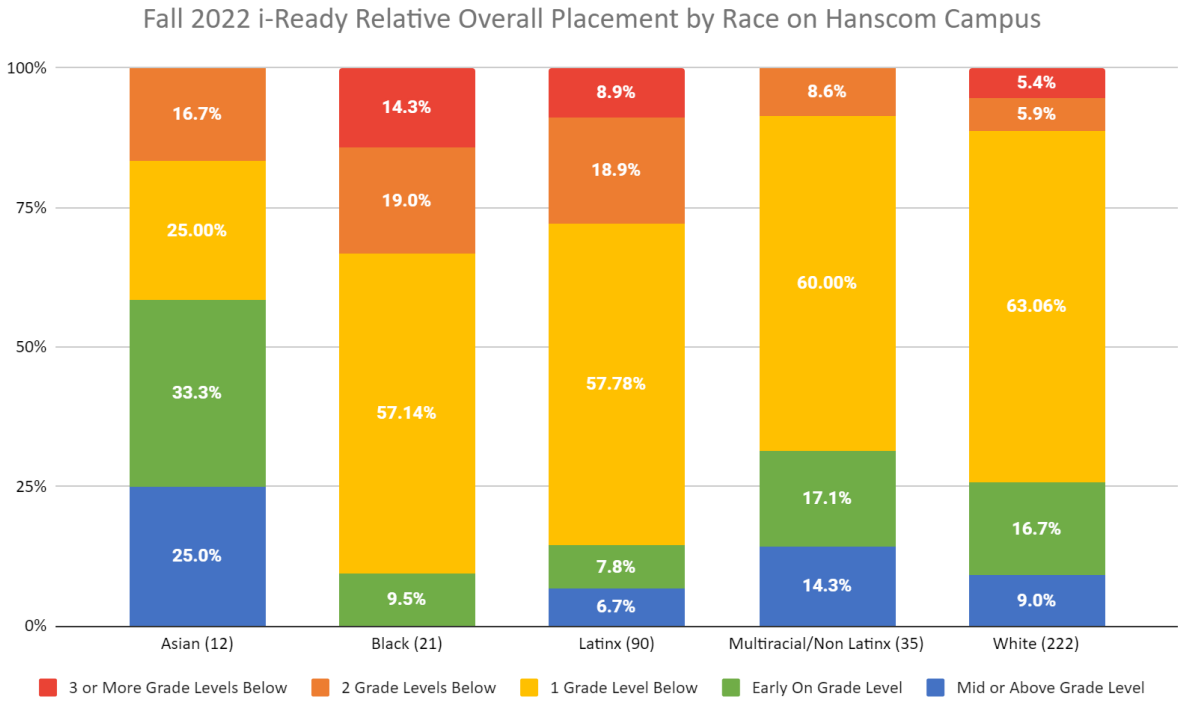

Fall 2022 i-Ready Relative Overall Placement by Demographic Groups Across District
100.0%
75.0%
50.0%
25.0%
0.0%
1.7%
6.0%
49.0%
22.4%
20.9%
18.5%
21.2%
45.0%
11.9%
3.3%
4.3%
7.3%
48.7%
21.2%
18.5%
10.8%
37.8%
40.5%
8.1%
2.7%
2.7%
10.8%
51.3%
20.1%
15.1%
6.5%
6.5%
45.5%
21.1%
20.4%
Students without Disabiliites (718)
Students with Disabilities (151)
Not EL (832)
EL (37)
Female-identifying (437)
Male-identifying (431)
3 or More Grade Levels Below
2 Grade Levels Below
1 Grade Level Below
Early On Grade Level
Mid or Above Grade Level

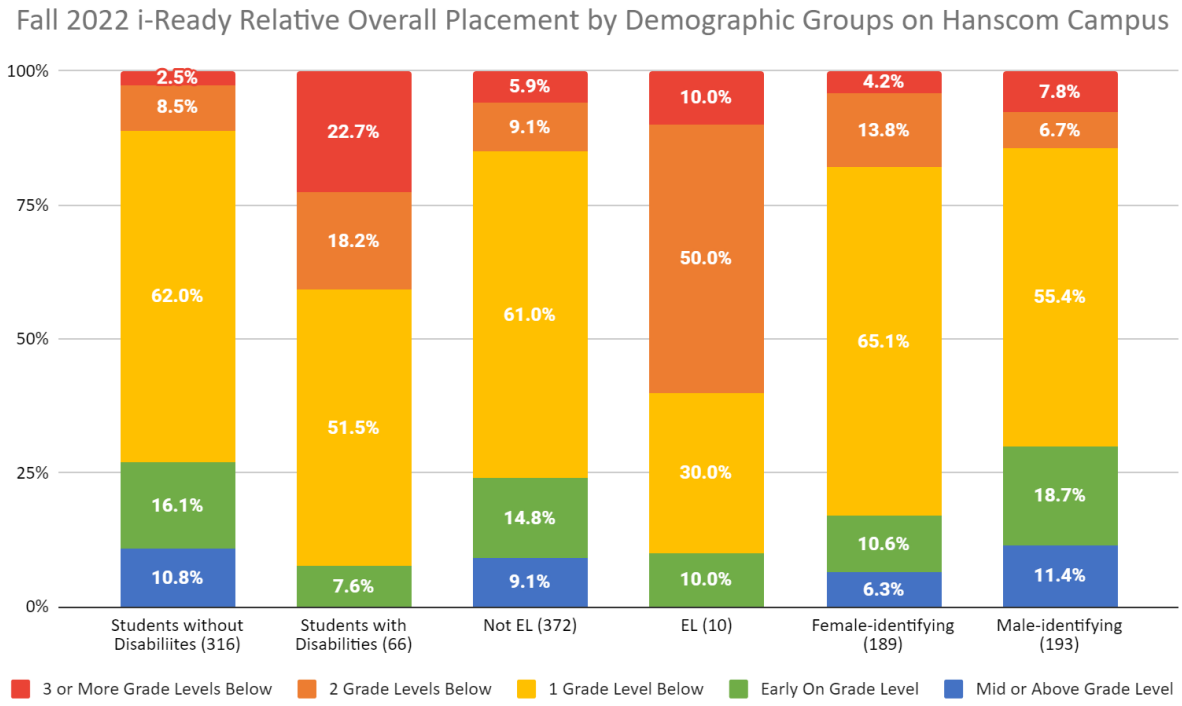
Differences Across Demographic Groups -- Lincoln Campus





Differences Across Demographic Groups -- Hanscom Campus





Duration of Testing Across the District

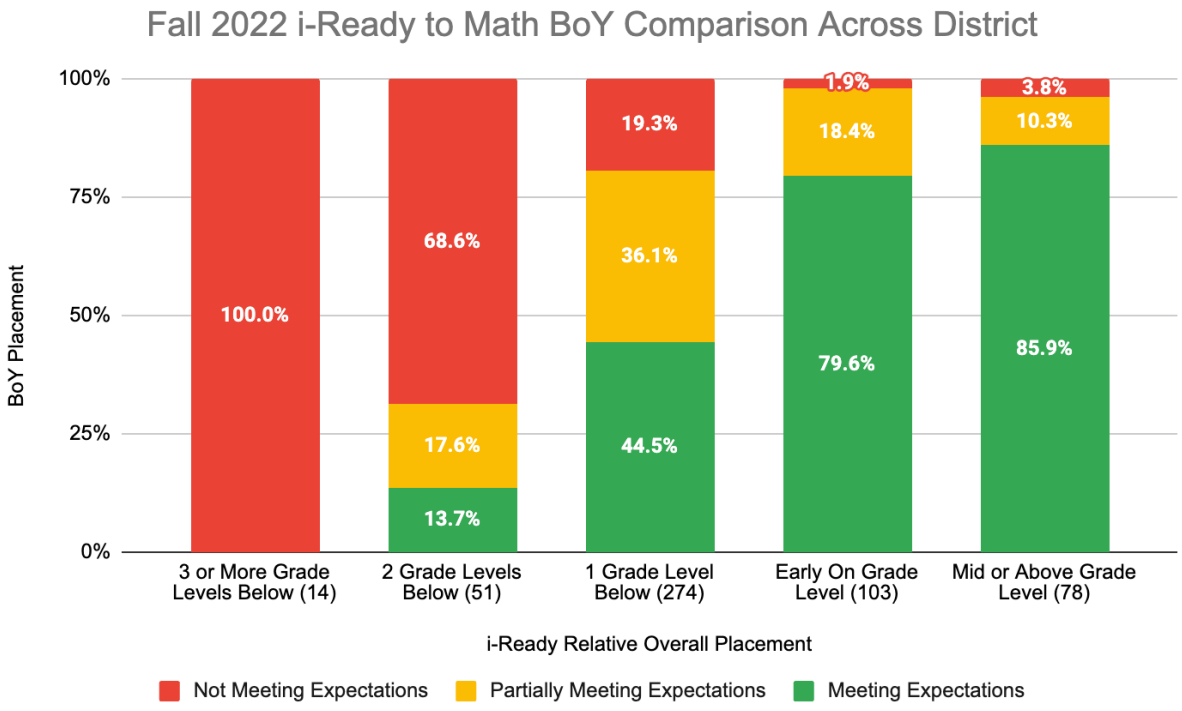
We tried to proactively manage the amount of time that students spent taking i-Ready by preparing students for what the experience would be like and coaching them throughout. There was a very wide range in duration across individual students. On average, kindergartners took the least amount of time and duration increased about 8-10 minutes each increasing grade level.

Grade	Average Duration (min)
K	20
1	27
2	36
3	41
4	52
5	62
6	62
7	70
8	80

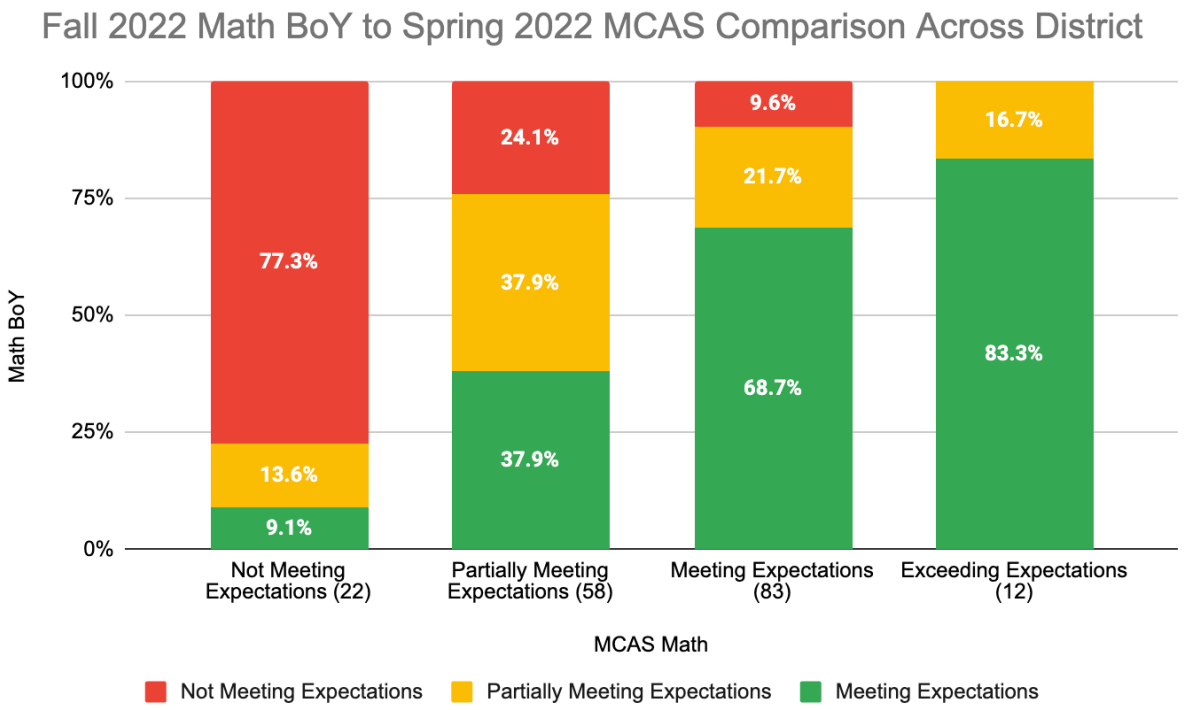
Analysis geared towards understanding if students who took less time performed at lower levels or if students took longer performed at higher levels did not clearly indicate that this was a strong trend across all performance levels. Instead, we are left with questions about if students who took the longest (who were mainly 7th and 8th graders) would have scored the same or relatively similar had they been coached to pace differently. We look forward to the winter when our students will take i-Ready again and improve our processes and our understanding of what is helpful to students.

Comparing Data Across Different Assessments

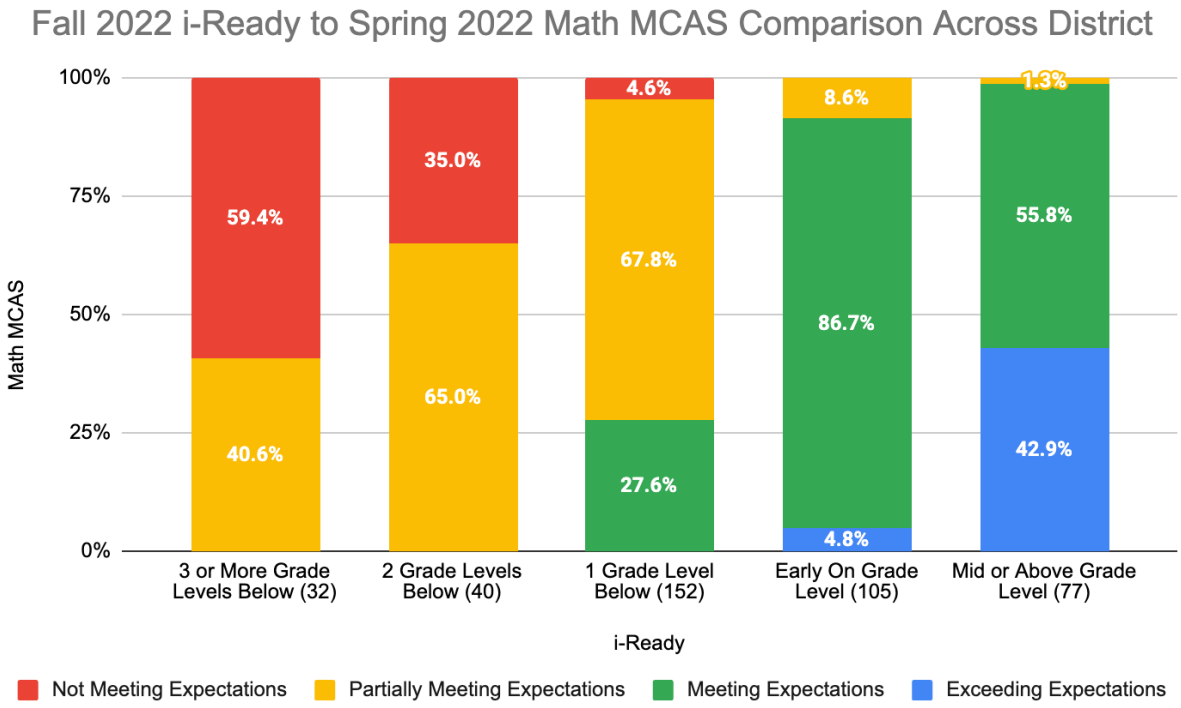
This year, we began analysis comparing data from our various assessments. The chart below shows students along the horizontal axis sorted by their relative overall placement on the fall i-Ready assessment. The colors of each bar indicate how the students in each of these i-Ready levels performed on the BoY assessment and data sort. It makes sense that we see significant overlap on the far ends and less consensus in the middle category, since these assessments are not fully calibrated and one assessment reports 5 levels of skill whereas the other only three.



When we compare last spring's MCAS results with the BoY assessments from this fall we see similar trends. Along the horizontal axis students are sorted into the four placement categories from MCAS. The colors on each bar indicate the percentage of students from that particular MCAS placement and how they scored on the BoY assessment. It is important to keep in mind that MCAS requires significant stamina from students, whereas the math BoY assessment is much shorter in length.



We were curious to compare fall 2022 i-Ready scores to spring 2022 MCAS. The chart below shows i-Ready relative overall placements along the horizontal axis and the colors within each bar indicate how those students scored on MCAS.



Finally, we also compared our literacy BoY data to the ELA spring MCAS.