



# Dyslexia Working Group

## Assessment Comparison

May 2025

---

# Agenda

- **Assessment types & uses**
- **Comparison of interim assessments**
- **Takeaways**
- **Reactions & Questions**



---

# Assessment Types & Uses

---

# Validity Evidence

**Common question:**  
***Is the assessment  
valid and reliable?***

**Our response:**  
***What purpose are you  
trying to fulfill?***

According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014):

- **Validity:** The degree to which evidence and theory support the interpretations of **test scores** for proposed **uses of tests**
  - Evidence based on test content
  - Evidence based on response processes
  - Evidence based on internal structure
  - Evidence based on relations to other variables
  - Evidence for consequences of testing
- **Reliability:** The extent to which an instrument produces measures that are consistent or repeatable
  - Internal consistency
  - Test-Retest
  - Alternate form
  - Interrater



# Assessment Types

	Purpose	Frequency	Questions Addressed
<b>Summative</b>	Evaluates whether students have met grade level standards	Once, typically toward the end of the school year	Have these students met grade-level standards?
<b>Interim</b>	Evaluates whether students are advancing toward achievement of grade-level standards	At key points throughout the year	Are students on track to meet grade-level standards by the end of the year?
<b>Screeners</b>	Identifies those who may need extra support to attain desired learning outcomes	Typically 1–3 times per school year	Do students require additional support or further evaluation?
<b>Diagnostic</b>	Determines the eligibility of students for specialized programming or services	As needed, typically based on the results from other assessments	What are students' strengths and areas of specific need? Can learning needs be diagnosed by additional focused assessment?
<b>Formative</b>	Checks students' understanding during the course of instruction to guide teaching and learning	Daily or ongoing	Are students learning what was planned for them to learn? If not, how can understanding be improved to meet learning goals?

# READ Act Assessment Types and Definitions

Assessment Type	Purpose/Definition	Frequency
Interim**	<p>To determine whether a student has a significant reading deficiency (SRD) in grades K–3.</p> <p>Defined by READ Act as a universal screening assessment administered to all students to identify those who may experience lower than expected reading outcomes who may be at risk for reading challenges.</p>	Administered three times per year
Diagnostic	<p>To identify a student's specific reading deficit(s).</p> <p>Defined by READ Act to use for students identified through screening as possibly having a significant reading deficiency so as to pinpoint a student’s specific area(s) of weakness and provide in-depth information about students’ skills and instructional needs.</p>	Within 60 calendar days of the initial administration of the interim assessment, if score is below the SRD cut score
Summative	<p>To inform taxpayers and state policy makers, support identification of successful programs, and serve a variety of state and federal accountability needs.</p> <p>Defined by READ Act as an end-of-year comprehensive measurement of student mastery.</p>	Once, at the end of the school year. Not required.

*\*\*Opportunity for dyslexia screening.*



---

# Comparing Interim Assessments

---

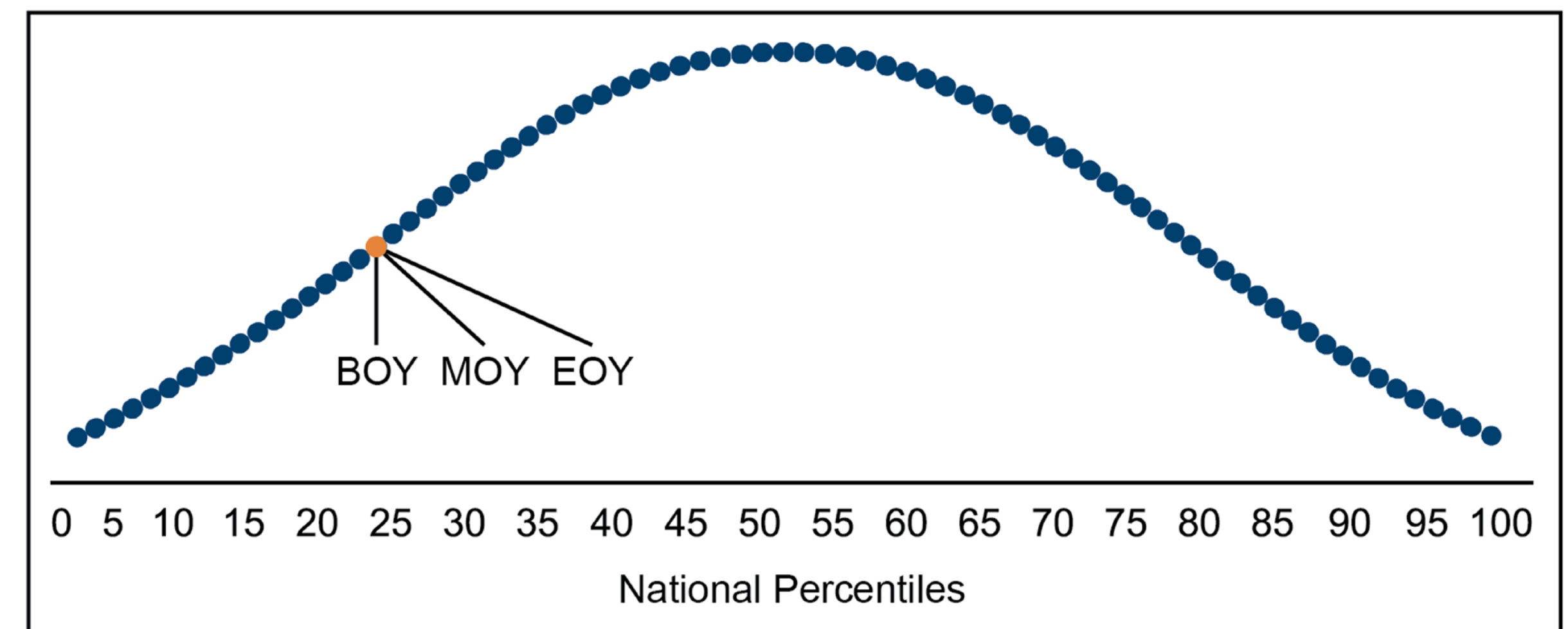
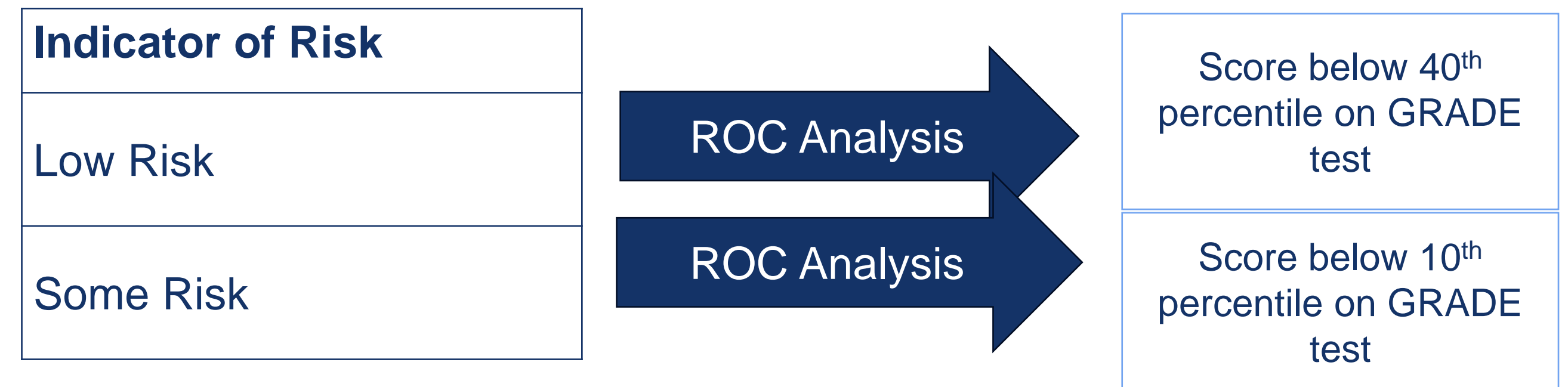
# Approved Interim Assessments

- **Five (5) interim assessments approved for 2023–24 school year and beyond**
  - Acadience Reading
  - mCLASS: DIBELS 8th Edition (English and Spanish)\*
  - i-Ready Assessment for Reading
  - ISIP Reading (English) and ISIP Lectura Temprana (Spanish)
  - Star Early Learning (English and Spanish)
- **All assessments, except Acadience Reading, include an approved diagnostic assessment**
- **All assessments, except ISIP Lectura Temprana, claim the test has been validated for dyslexia screening**
  - Some assessments may use SRD cut scores to determine risk of dyslexia



# What does “significant reading deficiency” mean?

- Different content assessed
- Different testing methods (fixed-form vs. computer adaptive testing)
- Different methods of setting benchmark cut scores
  - Logistic regression / ROC analysis
  - Normative
  - Criterion-based standard setting approach
- Differences can lead to different students being identified



---

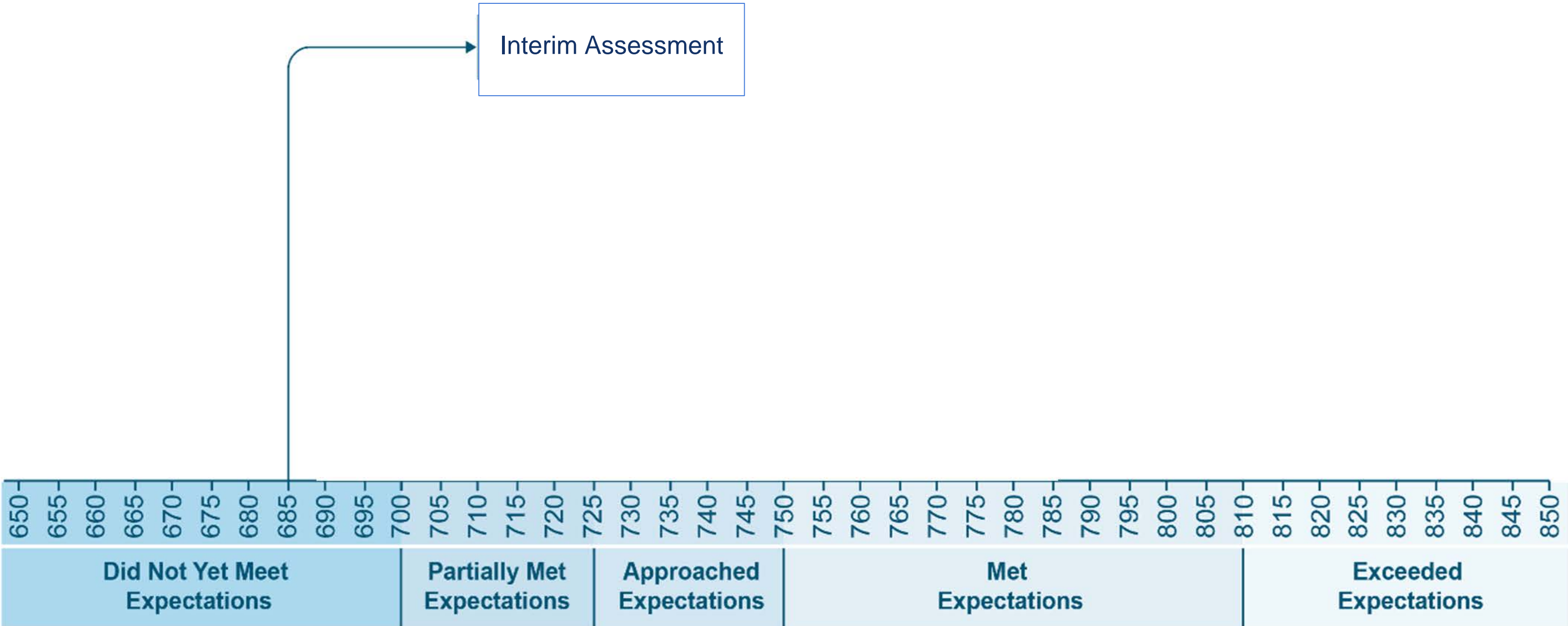
# How do reading difficulty cut scores differ across assessments?

- **Used equipercentile linking (Kolen & Brennan, 2004) to link each of the interim assessment cut scores to the CMAS scale to compare how the cut scores vary across assessments**
  - Method is content-neutral (content on the assessments do not need to be identical)
  - Allows comparison of how difficult interim assessments are in terms of identifying students at risk of reading difficulty
  - Scores on the different assessments are considered equivalent when the scores on each test have the same percentile rank
- **Method uses a paired group design (a group of students who took two different assessments)**
  - Our analysis includes only students who took both the interim assessment and CMAS in the spring of grade 3



# Example of Link Calculation

- **Example of linking of interim assessment to CMAS:**
  - 11% of students who took the interim assessment were identified as having an SRD due to their score
  - The score at the 11<sup>th</sup> percentile on CMAS was 685 for this group of students
  - The interim assessment SRD cut score maps to 685 on CMAS since these values were both at the 11th percentile for the same group of students



---

# Important Considerations

- **Accuracy of the linkages depends on how representative the analytic samples are of the statewide population**
- **Analysis just focuses on assessment cut scores; does not take into account differences in content of the assessments or other differences (e.g., differences in administration)**
- **Analysis is only focused on grade 3 students at the end of the year (we only have spring data available)**
  - We can link earlier grade level performance to the CMAS scale but linkages will be less accurate

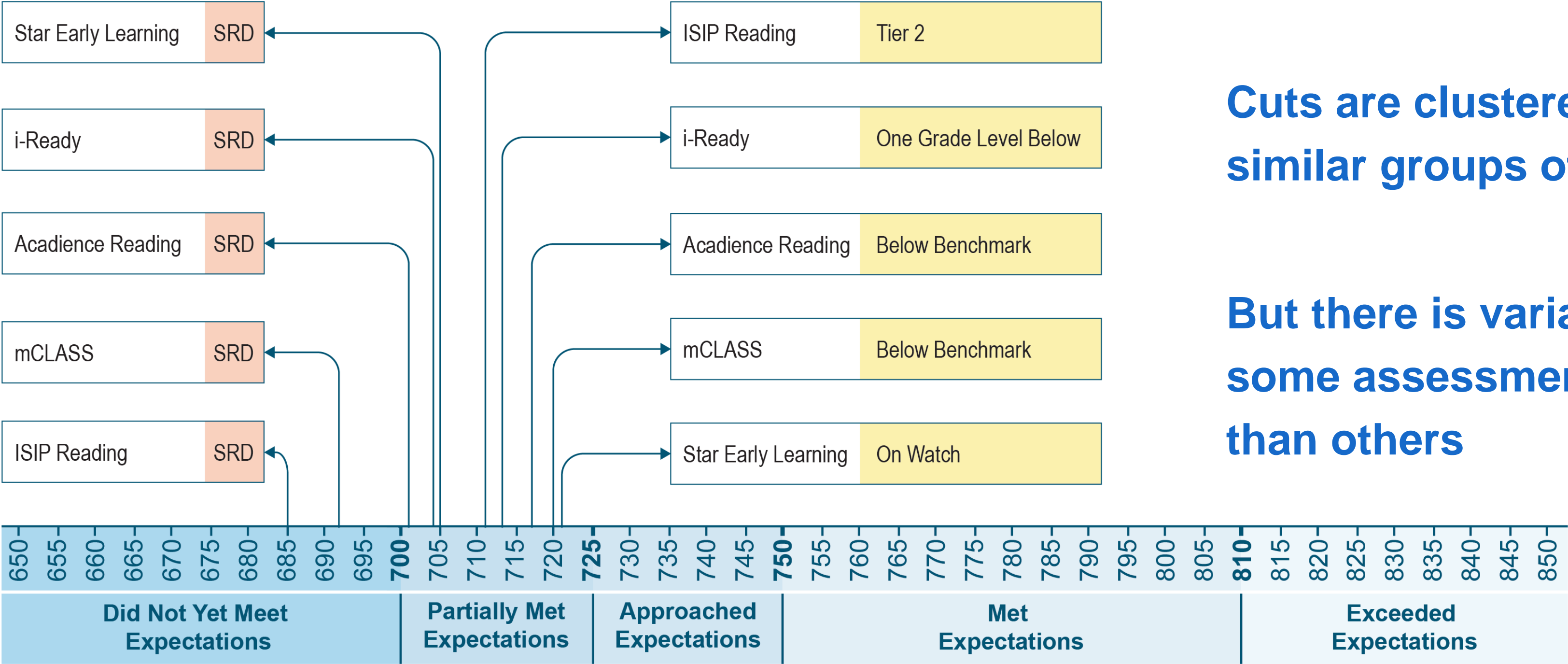


# How do the assessment SRD cut scores compare to the CMAS composite scale?

SRD cut scores (red) primarily link to CMAS scores in the higher end of “Did Not Yet Meet Expectations” range and lower end of “Partially Met Expectations” range

Cuts are clustered so assessments identify similar groups of students

But there is variation across assessments - some assessments will identify more students than others

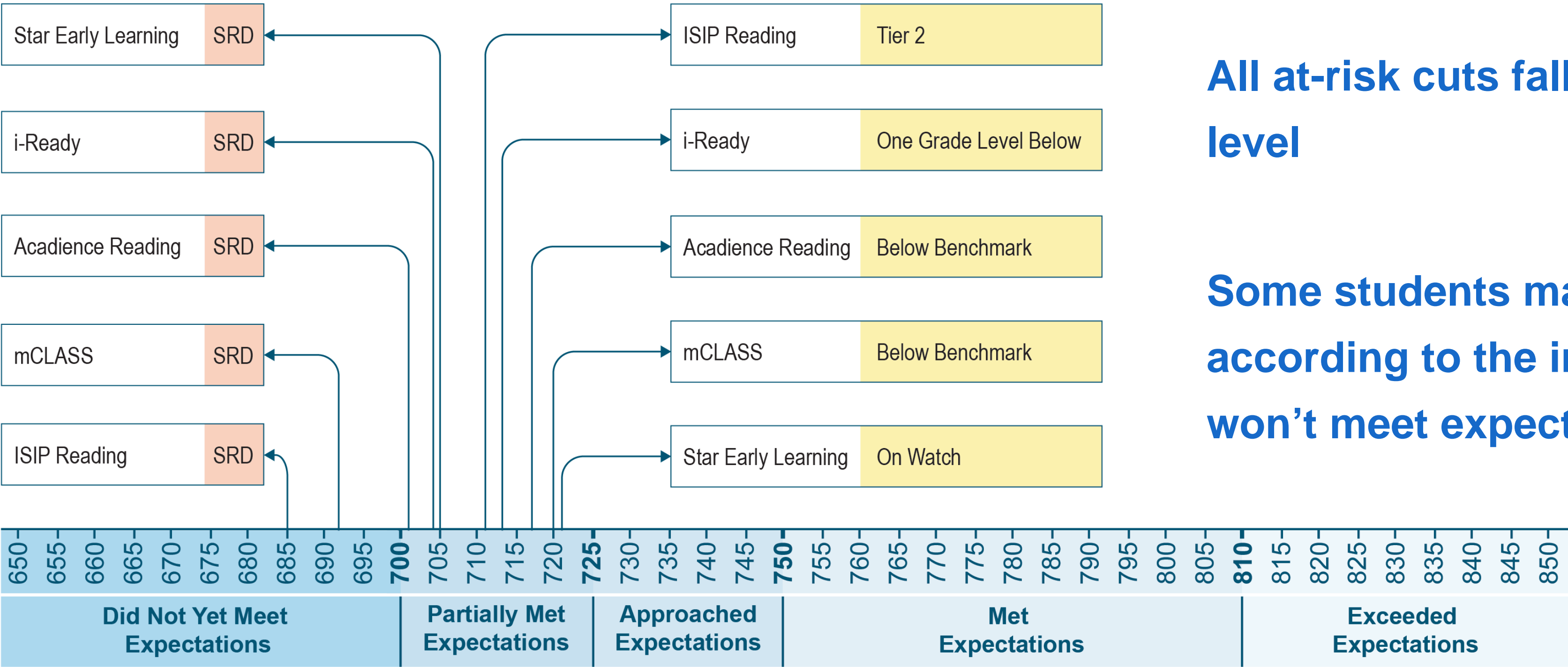


# How do the assessment general at-risk cut scores compare to the CMAS composite scale?

Similar findings for general at-risk cuts (yellow) that identify students at any level of risk of reading difficulty

All at-risk cuts fall below the “Met Expectations” level

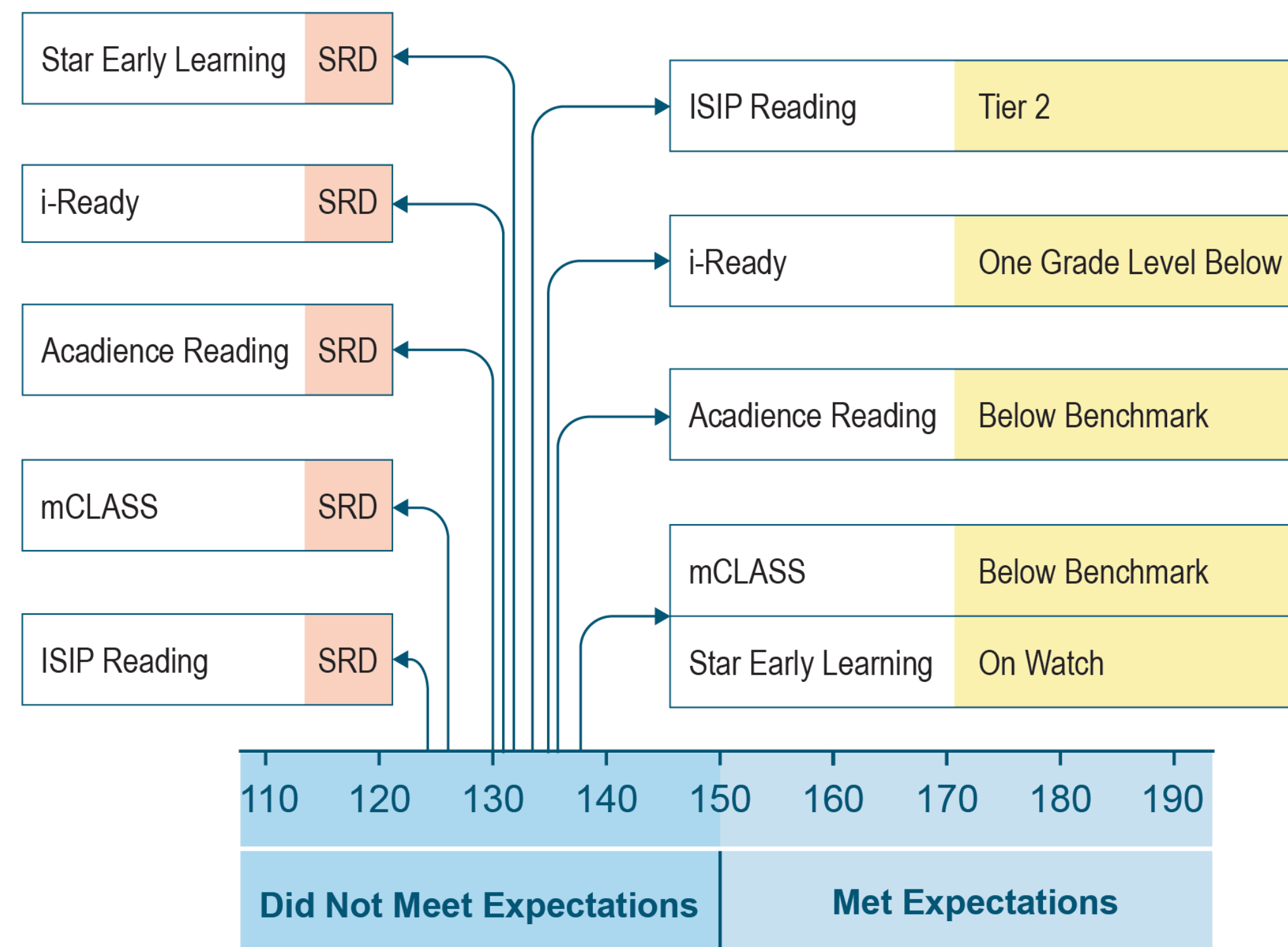
Some students may be considered as on grade level according to the interim assessments but likely won’t meet expectations on CMAS in grade 3





# How do the assessment cut scores compare to the CMAS reading subscale?

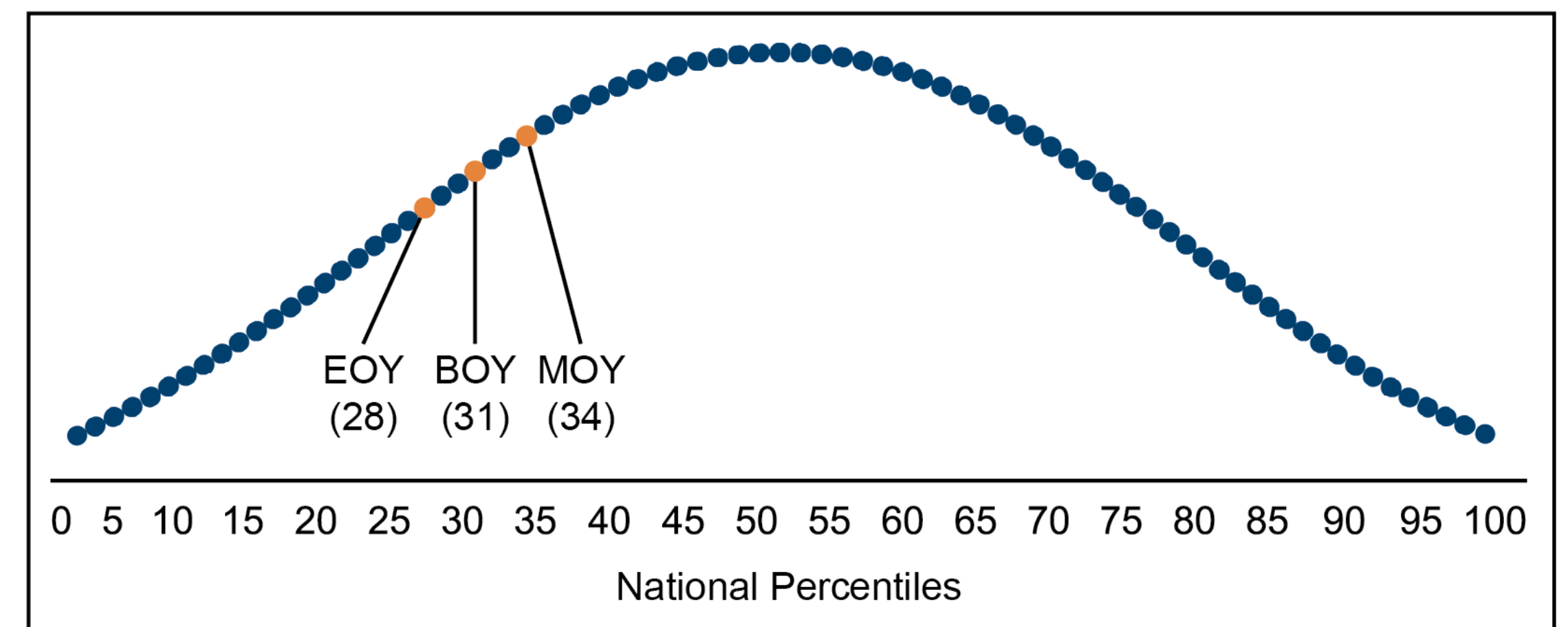
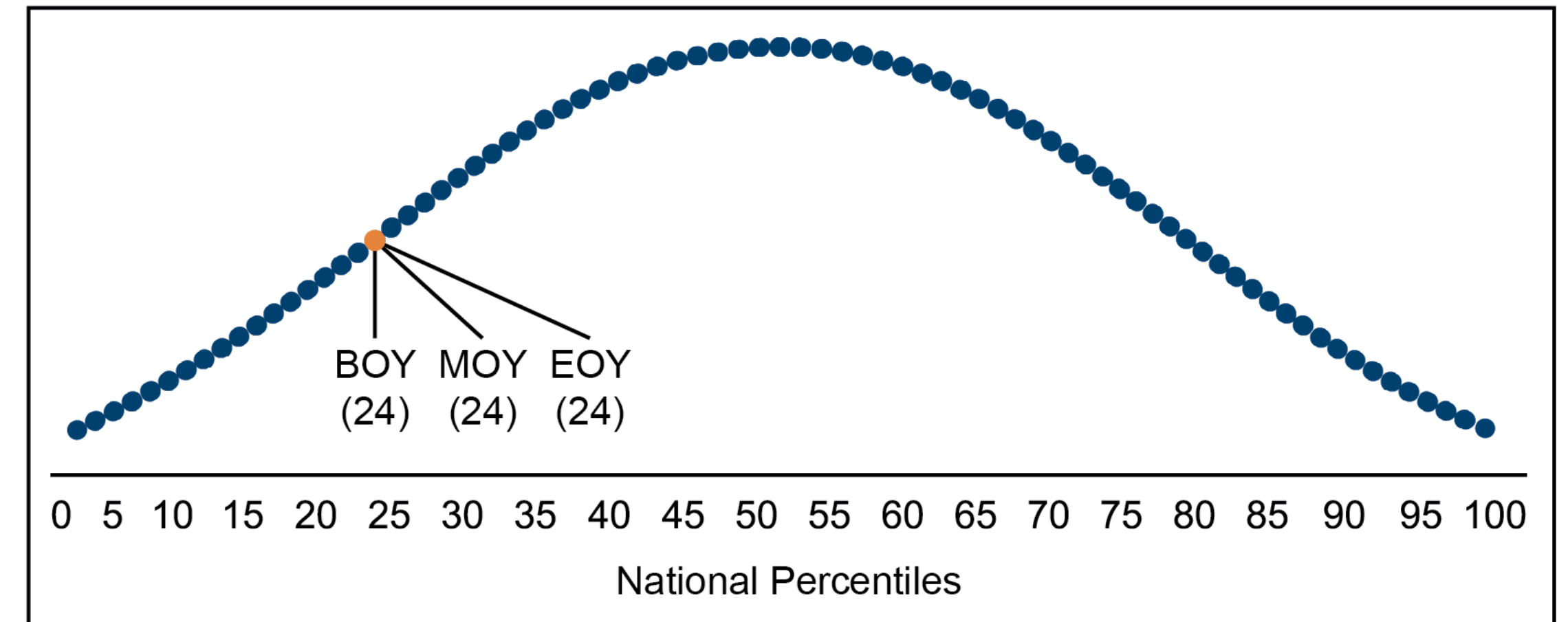
- **Linking interim assessments to the CMAS reading subscale reveal similar findings**
  - Assessments identify similar groups of students but there is variation
  - Not scoring at any level of risk does not mean a student will meet expectations at the end of grade 3



# How do the assessment cut scores change within a school year?

We saw that cut scores vary across assessments – so some assessments may identify more students at risk of reading difficulty than others

But cut scores may also vary within an assessment – there can be variation in the number of students identified as the school year progresses












---

# Takeaways

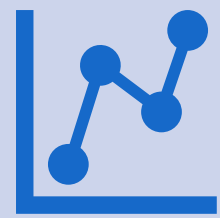
# Summary of Takeaways

-  Assessments measure risk differently but for the same purpose – choose something that meets the proposed goals and needs
-  Be specific with the meaning of “risk” or “SRD”
-  Be aware that that different assessments may not identify the same kids (or same number of kids) as having an SRD
-  Be aware that not being identified as “at risk” doesn’t necessarily imply proficiency on CMAS
-  Know that benchmarks can shift within and across years/grades – has implications for interpreting scores and growth for individual students and groups of students
-  Caution needed when trying to use scores for purposes beyond identifying students in need of support
-  Caution needed when using assessments with bilingual/multilingual students;  
Some vendors have instruments available in Spanish, but they should be validated for use with Spanish-speaking students and Spanish-language standards

No one assessment is necessarily better than another –  
need to consider content, purpose, and score  
interpretation for intended purpose

---

## Ideas to Consider



**Consider whether the approved assessments are being utilized for their intended purpose(s) and that the results are being interpreted appropriately.**



**Consider collaborating with researchers and assessment publishers to validate cut scores and ensure accurate interpretations tailored to specific outcomes.**



**Consider how the approved interim assessments align with screening for characteristics of dyslexia.**



---

# Resources

- **Blog:**
  - [Comparing Early Literacy Assessment: What Really Matters](#)
- **Massachusetts reports:**
  - [Comparison of screener benchmarks in Massachusetts](#)
  - [Report on early literacy performance in Massachusetts](#)
- **Colorado reports:**
  - [Interim assessment comparability analysis](#)
  - [CO Read Act Y4 evaluation report](#)
- **Assessment maker reports:**
  - [Linking study between CMAS and NWEA MAP Growth](#)
  - [Linking study between CMAS and i-Ready](#)

---

# Reactions & Questions