

TESTING OUR PATIENCE:

HOW WISCONSIN LOWERED STANDARDS,
WIDENED THE ACHIEVEMENT GAP,
AND BUSTED ITS STATE EXAMS

HOW TO FIX TEST SCORES



A REPORT BY



INSTITUTE FOR
REFORMING GOVERNMENT

OCTOBER 2024



INSTITUTE FOR
REFORMING GOVERNMENT



ABOUT THE INSTITUTE FOR REFORMING GOVERNMENT CENTER FOR INVESTIGATIVE OVERSIGHT

The Institute for Reforming Government is a 501(c)(3) non-profit that wants to give all Wisconsinites an opportunity to succeed and prosper. IRG does this by developing policy reforms, engaging communities, and holding government accountable.

IRG's Center for Investigative Oversight supports rigorous, independent, and objective oversight of state government and its agencies.

CIO Director Jake Curtis joined the IRG team in 2024. Throughout his career he has focused on government relations matters for both public and private sector clients, representing dozens of units of government and handling administrative matters before WERC, DWD, ETF, DNR, DFI, DATCP, and DPI.

Jake is a graduate of the University of Wisconsin – Eau Claire, B.A., summa cum laude, 2005, the University of Wisconsin Law School, cum laude, 2008, the United States Air Force, Officer Training School, 2021, and the United States Air Force, The Judge Advocate General's Corps, 2022.



ABOUT THE PROCESS

On April 25, 2024, IRG submitted a public record request for the period January 1, 2023, to April 25, 2024. IRG requested DPI release records concerning the 2024 standard setting of cut scores for the Forward Exam and ACT as well as records using the term "cut scores." DPI released records to IRG on September 4, 2024. For transparency's sake, IRG has published all received records on its website.

IRG has selected records for this report that inform the public about the significant problem Wisconsin's education leadership has created, how the problem arose, and how the problem may affect educators, officials, parents, and kids.

ReformingGovernment.org

TABLE OF CONTENTS

Key Findings	5
Key Quotations	7
Key Questions	9
Initiating Test Score Changes	11
Early Inequity	12
Trouble Understanding	15
Changing the Tests	18
Widening the Achievement Gap	20
Aftermath	23
Email Appendix	26



Key Findings

Within her first year as Wisconsin's educational leader, State Superintendent Jill Underly published a series of essays on equity. She wrote rousing about Wisconsin's infamous education achievement gaps for disadvantaged groups and what it would take to fix them. "We need to question what we are doing as an education system that results in these disparities... The question is again whether we have the collective will to reckon with these disparities and the need for belonging in our classrooms. I believe we do, and we must."¹

She also included this. "This is the reality of the achievement gap in Wisconsin. It cannot continue this way; we are failing our students of color, and one factor in that failure is in the language we use to describe it."²

Two years later, the language describing the achievement gap became the center of a boiling debate. On June 6, 2024, the Department of Public Instruction publicly announced two substantial changes to the Forward Exam, PreACT Secure, and ACT that students take at the end of the school year.³ The names of the four categories into which students sort would switch to softer, "growth-based" terms. Students would no longer demonstrate Below Basic, Basic, Proficient, or Advanced skills, but be Developing toward, Approaching, Meeting, or Advanced past Wisconsin state standards.⁴

More consequentially, DPI altered each test's "cut scores," the minimum scores a student needs to reach each of those tiers of achievement.⁵ Citing revisions to state standards, DPI announced that Wisconsin families would no longer be able to compare their children's results to their standing before the pandemic.⁶ Additionally, it soon became clear that DPI had lowered the standards for each performance level.⁷ These changes have major consequences for Wisconsin's schools, local leaders, state officials, and families.

¹ State Superintendent Doctor Jill Underly, "[Racial Disparities and Our Collective Will](#)," April 19, 2022.

² State Superintendent Doctor Jill Underly, "[Racial Disparities and Our Collective Will](#)," April 19, 2022.

³ Wisconsin Department of Public Instruction, "[DPI Focuses on Growth-Based Mindset in Updated Student Performance Level Descriptors for Standardized Assessments](#)," June 6, 2024.

⁴ Wisconsin Department of Public Instruction, "[Updated Asset-Based Performance Levels](#)," June 6, 2024.

⁵ Wisconsin Department of Public Instruction, "[DPI Focuses on Growth-Based Mindset in Updated Student Performance Level Descriptors for Standardized Assessments](#)."

⁶ Wisconsin Department of Public Instruction, "[What You Need To Know: Standard Setting and Forward Exam Updates](#)," August 14, 2024.

⁷ Alan J. Borsuk, "[Does Lowering Cut Scores and Changing Terminology on Standardized Tests Better Serve Wisconsin Students?](#)" *Milwaukee Journal Sentinel*, August 23, 2024.

Given that test scores release this month, it has become important to know when the state test changes were decided, what about them has changed, how much the tests have changed, and how changes could affect schools' pursuit of excellence.

Based on public records:

- » **State Superintendent Underly initiated the changes to test score standards and wanted them lowered to match other states'.**
- » **State Superintendent Underly struggled to interpret the implications of new test score standards.**
- » **DPI staff provided information to decisionmakers based on either 50% or 55% proficiency in 2024, up from approximately 40% in 2023.**
- » **DPI staff projected those proficiency levels would widen Wisconsin's achievement gaps along racial, poverty, and disability lines.**
 - » **DPI staff projected those proficiency levels would increase gaps between 4 to 10 points for Black students, 3 to 4 points for economically disadvantaged students, and 5 to 10 points for special-needs students.**
- » **DPI staff projected those proficiency levels would increase ratings on federally mandated state report cards, student performance being equal.**
 - » **DPI staff projected those proficiency levels eventually would cause 71% of districts to score 4 or 5 stars on state report cards, up from 36%, and 63% of schools to score 4 or 5 stars, up from 46%.**
- » **DPI staff projected those proficiency levels would benefit low-poverty schools more than high-poverty schools on state report cards due to how they are calculated.**
- » **DPI staff, to balance out the benefits to low-poverty schools, suggested realigning aspects of the report cards that favor high-poverty schools, benefiting them.**
 - » **DPI staff projected both beneficial changes together eventually would cause 93% of districts to score 4 or 5 stars on state report cards as well as 80% of schools.**
- » **DPI staff, to prevent this drastic shift, suggested changing report card cut scores for the 2024 to 2025 school year.**

Whether State Superintendent Underly accepted the standards suggested by DPI staff and standard setting participants this summer is currently unknown, as are 2024 state test scores. These public records end in April and test scores are private until later in October.

However, if test scores rise between 10% and 15%, if achievement score gaps increase statewide for the disadvantaged, if state report card ratings sharply rise, especially for low-poverty schools, and if Wisconsin politicians spend 2025 fixing report card ratings instead of reading instruction, the teacher supply, special-needs funding, or barriers to expanding high-quality schools, as early district results suggest, the record shows that DPI knew all of it would happen and did it, anyway.

Do we have the collective will to reckon with the massive disparities among Wisconsin students? Or will other states continue to pass us by, taking new residents, new businesses, and new possibilities with them?

Key Quotations

“Dr. Underly is interested in taking another look at how we establish our cut scores for the Forward Exam. Given that we’re amid some other change in that area, we’re asking for a decision paper on what it would take to change the cut scores, as well as what timeline we’d be working off.”

- then-Executive Director Thomas McCarthy, January 23, 2023, initiating changes to test score standards

“Using comparison data from 2019 as the basis for adopting these cut scores reflects that expectations of performance have not changed over time.”

- Office of Educational Accountability, May 30, 2023, setting PreACT score standards for 2023 that DPI would later lower in 2024

“I need help with this. I obviously trust Viji, but I truly don’t understand what I am looking at for my approval. And with all this other nonsense going on with literacy I want to make sure we’re not throwing more fuel onto this fire. The crummy thing is, I am an educator and I don’t understand it - so how are parents supposed to understand this too? If we could set the standards and the cut scores, but then have some kind of way to interpret it to parents and educators as a companion, that would be great. For example, what does Proficient mean vs. Advanced? That they are at grade level vs. the next grade level? I just hate this stuff so much.”

- Superintendent Jill Underly, June 13, 2023, reviewing potential PreACT score standards, reviewing statements saying lowering cut scores would increase report card star ratings more for low-poverty districts than high-poverty districts, questioning the meaning of student performance levels, and commenting on a matter related to literacy during the same period as Act 20 negotiations

“I’m going to need a primer on this, or a tutoring session. I still don’t understand it, and it’s just a learning block that I have on my end.”

- Superintendent Jill Underly, July 6, 2023, reviewing potential PreACT score standards again 1 day before approving them

“After equating, DRC Research staff will provide the impact data to DPI for approval. The impact data shows the percentage of students in each performance level for each grade and content.”

- Data Recognition Corporation, October 26, 2023, referring to performance data DPI would receive by June 10, 2024, before educators reset test score standards

“I also said that NAEP only tests 10,000 nationally, and that’s really an awful way to make generalizations about how kids are doing, and that when we use state assessments to compare kids state by state, it’s not an apples to apples comparison - so if the feds want to get in the assessment game, they should really make it so that we all take the same assessment and have the same cut scores. Because otherwise it’s not very useful information. ...But my feelings on high stakes tests are pretty well known.”

- Superintendent Jill Underly, December 19, 2023, criticizing the Department of Education’s methodology for the NAEP, the national test commonly analyzed by researchers

“Before I jumped off- yes, I’d like to have a conversation about our cut scores being the highest in the country and how we communicate what that translates to with NAEP.”

- Superintendent Jill Underly, February 20, 2024, following up on lowering test score standards

“I also want to, and maybe this exists as a possibility, to norm our levels to be similar to other states... as I hear that we have some of the highest cut scores nationally. I would like them to be looked at so that we aren’t judged negatively when we have direct standards.” - Superintendent Jill Underly, March 7, 2024, taking a tangent on lowering test score standards while discussing changing student performance level names

“Assuming these standard setting activities raise proficiency rates statewide, it is reasonably likely that: (1) proficiency-based achievement gaps will increase, (2) report card Overall Scores will increase year-over-year, and (3) increases in Overall Score and Overall Rating category will impact schools disproportionately (report card rating increases will benefit low-poverty schools more than high-poverty schools).” - Office of Educational Accountability, April 3, 2024, explaining to Superintendent Underly and her cabinet that lowering test score standards would lead to inequitable outcomes

“Our analysis projects that student groups who have been traditionally underserved will see smaller gains in proficiency rates as compared to more advantaged groups, widening the gaps between them.”

- Office of Educational Accountability, April 3, 2024, explaining to Superintendent Underly and her cabinet that lowering test score standards would widen the visibility of achievement gaps

“If proficiency rate gains are substantial, we project report card ratings will increase to such a degree that there will be little difference between schools in terms of their ratings.” - Office of Educational Accountability, April 3, 2024, explaining to Superintendent Underly and her cabinet that lowering test score standards would make school accountability report cards increasingly uniformly high and therefore useless

“In addition, these report card rating increases will benefit low-poverty schools more than high-poverty schools due to variable weighting of Achievement and Growth priority areas, as required by state law. ...As a result, rising proficiency and PBP rates will raise report card scores and ratings to a greater degree for schools and districts with lower percentages of ECD students.” - Office of Educational Accountability, April 3, 2024, explaining to Superintendent Underly and her cabinet that lowering test score standards would disproportionately benefit low-poverty schools, significantly increasing the number whose star ratings would rise

“While the impacts of the assessment standard setting on achievement gaps cannot be addressed, OEA would be able to address some of the challenges to the report card system to a certain extent... Potential next steps: 1. Keep report card calculations and rating thresholds unchanged for the 2023-24 school year as we anticipate relatively moderate increases this year. 2. For the 2024-25 school year, make the following changes to report card calculations: a. Include only two years of assessment data in achievement calculations, including the Achievement priority area, as well as the achievement components of the TGO and On-Track to Graduation priority areas... b. Adjust Growth and TGO rescaling formulas to align these priority area score distributions with the new Achievement priority area score distribution. 3. Prior to release of 2024-25 report cards, conduct a report card standard setting to establish new report card rating thresholds.” - Office of Educational Accountability, April 3, 2024, proposing to Superintendent Underly and her cabinet that DPI significantly change state accountability report cards in 2025 to attempt to balance out the effect of lowering test score standards in 2024

“Does Cabinet have a lower or upper threshold for recommendations? Forward? ACT?”

- Office of Educational Accountability, April 9, 2024, ending a meeting on test score standards before teachers had a chance to set those standards in July

Key Questions

These records raise numerous questions about the process, results and effects of DPI's choices. Asking these questions would create accountability.

What in January 2023 prompted Superintendent Underly to want to change test score standards?

Could 2024 cut scores have been calibrated to the same distributions as 2023 cut scores?

Did DPI leadership reconsider lowering cut scores when told in May 2023 that doing so would benefit low-poverty districts more than high-poverty districts?

If DPI set PreACT scores in 2023 that reflected expectations of performance had not changed over time, then lowered them in 2024, have DPI's expectations of students lowered over time?

Was "this other nonsense going on with literacy" referring to the Act 20 legislation being worked on then? If so, was there an aspect of Act 20 that Superintendent Underly considered nonsense, or was it all nonsense?

Did Superintendent Underly tell legislators during the Act 20 negotiations that DPI would change test cut scores?

In simple terms that parents can understand, how would Superintendent Underly define "Proficient" under 2023 standards? How would she define "Meeting" under 2024 standards?

Why did Superintendent Underly want to change cut scores before she understood how they had been calibrated?

Who guided Superintendent Underly through understanding PreACT test score setting?

Which state's test score standard does Superintendent Underly prefer we match?

Did the DPI cabinet recommend cut score floors or ceilings for the Forward or ACT before teachers participated in the score-setting process months later?

Does Superintendent Underly believe the Department of Education's NAEP test, used to track state and national progress for decades, is an illegitimate measure of student learning?

Did DPI provide educators setting Forward Exam cut scores contextual information like past test scores?

Did DPI provide educators setting ACT cut scores contextual information like past test scores?

Did Superintendent Underly accept the cut scores as each group of participants recommended, or did she adjust them?

Does Superintendent Underly feel like a 19 on the ACT is an appropriate minimum for proficiency for 11th graders, given ACT predicts 37% of students at that level would get a D or F in college algebra and 18% at that level would get a D or F in college English composition?

Does DPI possess “impact data” for the Forward Exam sent in May and June 2024, as the contract states? Does that data show 2024 proficiency rates before June standard setting took place?

Given Superintendent Underly’s bewilderment when trying to understand the effects of PreACT score setting, how did she know if Forward and ACT score setting was the right match for Wisconsin students?

Are the current cut scores “reasonable and defensible in terms of the process, public and educator perception, and related other data,” as Director Visalakshi Somasundaram stated?

Did Superintendent Underly’s approved changes increase proficiency achievement gaps on the Forward Exam, PreACT, or ACT? Were such changes equitable?

Did academic- or equity-focused members of Superintendent Underly’s cabinet agree that cut score changes were equitable?

Did DPI enact any cut score changes or adjustments to 2024 report cards?

What is the distribution of star ratings to districts on 2024 report cards?

Did low-poverty schools disproportionately benefit on 2024 report cards?

Will DPI enact any cut score changes or adjustments to 2025 report cards?

Is it equitable to ask parents to get their hands dirty to understand whether their child is doing well in school or if their child’s school is successful?

Was DPI telling the truth in August when they said the effects of cut score changes on state report cards were unknown?

Does Governor Evers feel like these test score and report card changes unfairly diminish high-poverty schools and unfairly commend low-poverty schools?

Initiating Test Score Changes

Superintendent Jill Underly initiated test score standard changes as early as January 2023.

On January 23, 2023, an email arrived in the inboxes of high-level officials within Wisconsin's Department of Public Instruction. "Dr. Underly is interested in taking another look at how we establish our cut scores for the Forward Exam. Given that we're amid some other change in that area, we're asking for a decision paper on what it would take to change the cut scores, as well as what timeline we'd be working off."

This came from Thomas McCarthy, Executive Director at the time, and was sent to Student and School Success Assistant State Superintendent Keona Jones and Policy Initiatives Advisor Aastha Ranabhat as well as Executive Director McCarthy's peers in leadership, Deputy State Superintendent John Johnson and Equity and Inclusion Executive Director Demetri Beekman.

Deputy State Superintendent Johnson and Office of Educational Accountability Director Visalakshi Somasundaram quickly agreed that it was a good time for changes given that DPI had revised state standards for English language arts and mathematics during the pandemic.⁸ Director Somasundaram explained, "Standard settings are typically done when there are changes to the standards or assessments. We plan to assess the revised ELA and mathematics standards on the Forward Exam beginning in spring 2024 (as John Johnson mentioned). Therefore, it could be appropriate to do a new standard setting in the summer of 2024. We are putting together a more detailed brief with options and a timeline for all the assessed content areas."

As DPI discussed the matter, students would ultimately take 2023 Forward Exams under the old content and cut scores, despite having new standards in place since 2021.⁹

What matters? How high or low DPI set test score standards was a separate issue from the need to set up standards for refreshed state tests. Though the Forward Exam included new content that needed to be calibrated by educators, Superintendent Underly appeared to have initiated a review of cut scores independent of that fact.

⁸ Wisconsin Department of Public Instruction, "[State Standards](#)," May 21, 2024.

⁹ Wisconsin Department of Public Instruction, "[2024 Wisconsin Forward Exam Performance Level Standard Setting](#)," June 29, 2024.

Early Inequity

Superintendent Underly and her cabinet knew as early as May 2023 that lowering test cut scores would disproportionately help advantaged groups, widen achievement gaps, and inflate report card ratings.

Director Somasundaram replied back about cut score changes on May 30, 2023, focusing on the PreACT Secure. Students took the PreACT instead of the ACT Aspire beginning in 2023.¹⁰ Because of this, DPI had to set that test's cut scores for the first time. These changes would set the tone for Forward and ACT changes in 2024.

For those readers who do not know, Wisconsin students in district, charter, and private schools on a voucher are required to take state tests every spring from 3rd grade to 11th grade. 3rd to 8th graders take the Forward Exam, which primarily tests reading and mathematics knowledge and skills.¹¹ High school freshmen and sophomores take the PreACT, which also focuses on those subjects.¹² High school juniors take the similar ACT.¹³ Both the PreACT and ACT are the same tests students in other states take, but Wisconsin uses them to fulfill state and federal accountability requirements.¹⁴ Students who take the test receive their scores, which fall into one of four categories: Below Basic, Basic, Proficient, and Advanced. (Students who miss the test are also recorded.)¹⁵ Two things are important about those ratings. First, Proficient and Advanced students have met Wisconsin's standards and are considered "proficient." Second, cut scores determine how many students fall into each category. High cut scores will cause fewer students to score proficiently, and lower cut scores will cause more students to score proficiently.¹⁶

Readers also need one more critical piece of background. All three tests factor into "state report cards," federally required rating systems for districts and schools.¹⁷ Each receives a score between 1 and 5 stars, and school officials and parents alike often rely on those ratings as measures of quality. Report card ratings include four factors: Achievement (test score proficiency), Growth (how much students grow on tests in 1 year), Target Group Outcomes (how well the lowest performers do on test proficiency, attendance, and graduation), and On-Track to Graduation (attendance, graduation, and certain test score proficiencies). Critically, to evaluate how much schools help students and not just to judge schools based on how disadvantaged or advantaged their students are, schools have certain factors count for more toward their star rating. Achievement (test score proficiency) receives more consideration when scoring low-poverty districts. Growth (yearly test score improvement) receives more consideration when scoring high-poverty districts.¹⁸

¹⁰ Wisconsin Department of Public Instruction, "[PreACT](#)," May 2, 2024.

¹¹ Wisconsin Department of Public Instruction, "[Wisconsin Forward Exam](#)," September 2, 2024.

¹² Wisconsin Department of Public Instruction, "[PreACT Secure](#)," September 2, 2024.

¹³ Wisconsin Department of Public Instruction, "[The ACT with Writing](#)," September 2, 2024.

¹⁴ Wisconsin Department of Public Instruction, "[2022-23 Report Card Guide](#)," November 14, 2023.

¹⁵ Wisconsin Department of Public Instruction, "[Forward Exam Data and Results](#)," December 2, 2023.

¹⁶ Wisconsin Department of Public Instruction, "[Student Performance Levels](#)," December 2, 2023.

¹⁷ Wisconsin Department of Public Instruction, "[Report Cards Home](#)," November 14, 2023.

¹⁸ Wisconsin Department of Public Instruction, "[2022-23 Report Card Guide](#)," November 14, 2023.

Director Somasundaram used this opportunity to explain the various ways DPI could set PreACT cut scores and how that would set the tone for the Forward and ACT. She began by explaining that cut scores are “ultimately a matter of informed judgment rather than discovering objective performance standards. Recommended cut scores are submitted to the State Superintendent for review, adjustments may be made, and formal approval is documented. Cut scores must be reasonable and defensible in terms of the process, public and educator perception, and related other data.” She also noted that having a revised Forward test and new PreACT test could require resetting the ACT cut scores, too.

Director Somasundaram laid out the facts on cut score standards. Test scores account for a large part of report card star ratings. The more students who score well on the examinations, the higher a report card score is. However, because higher proficiency disproportionately helps districts and schools with fewer economically disadvantaged students, “These impacts will not be felt equally by all schools and districts... Schools and districts with the highest percentages of economically disadvantaged (ECD) students will see little change...”

Before providing recommendations, Director Somasundaram stated a truth that senior leadership would later echo. Previous cut scores set by then-State Superintendent Tony Evers “were intentionally made comparable to the cut scores for the National Assessment of Educational Progress (NAEP) in response to ‘mapping studies’ showing many states with ‘lower standards’ than NAEP.” The NAEP is a national, federally mandated test whose main result comes out every two years.¹⁹ Its proficiency standard represents a higher standard than “being at grade level.”²⁰ This higher standard, and Wisconsin’s along with it, might most closely be described as “college readiness.”²¹ Director Somasundaram noted that, because proficiency scores are so visible to educators, families, policymakers, media, and voters, “DPI must consider whether the current proficiency rates accurately reflect students’ level of knowledge and skills related to the grade-appropriate academic standards.”

Director Somasundaram settled on matching 2023 PreACT cut score standards to match the distribution of students on the 2019 ACT Aspire. She considered matching 2022 ACT Aspire distributions (reflecting post-pandemic decline), using the college readiness standards that the ACT company uses, or reverse engineering the cut scores by drawing a line from the 2023 Forward and ACT proficiency levels.

The effects of this decision are less important than the specifics and the scores. According to Director Somasundaram, “Using comparison data from 2019 as the basis for adopting these cut scores reflects that expectations of performance have not changed over time.” As for scores, DPI set the cut scores for English language arts as a 1, 12, 16, and 24 in 9th grade English language arts, as 1, 13, 18, and 26 in 10th grade English language arts, as 1, 15, 17, and 24 in 9th grade mathematics, and as 1, 16, 19, and 26 in 10th grade mathematics.²²

¹⁹ National Center for Education Statistics, “[About NAEP](#),” March 21, 2024.

²⁰ National Center for Education Statistics, “[Scale Scores and NAEP Achievement Levels](#),” December 9, 2021.

²¹ Burhan Ogut, George Bohrnstedt, and Markus Broer, “[Updated College Enrollment Benchmarks for the Grade 12 NAEP Mathematics Assessment](#),” American Institutes for Research, October 19, 2023.

²² Wisconsin Department of Public Instruction, “[PreACT Secure Data and Results](#),” November 29, 2023.

What matters? Superintendent Underly and DPI leadership knew lowering proficiency standards on state tests would raise report card ratings across the board, inequitably helping low-poverty schools more than high-poverty schools. DPI chose a PreACT cut-score standard it considered a high expectation consistent with its expectations before the coronavirus pandemic. “Proficiency” on Wisconsin and national exams did not mean “percentage of students at grade level,” but something akin to “percentage of students at a college-ready level.”

Trouble Understanding

Despite wanting to lower cut score standards, Superintendent Underly did not understand the standard-setting process or its effects.

Superintendent Underly soon replied privately to her closest reports, Executive Director McCarthy, Deputy State Superintendent Johnson, and Executive Director Beekman. “I need help with this. I obviously trust Viji, but I truly don’t understand what I am looking at for my approval. And with all this other nonsense going on with literacy I want to make sure we’re not throwing more fuel onto this fire. The crummy thing is, I am an educator and I don’t understand it - so how are parents supposed to understand this too? If we could set the standards and the cut scores, but then have some kind of way to interpret it to parents and educators as a companion, that would be great. For example, what does Proficient mean vs. Advanced? That they are at grade level vs. the next grade level? I just hate this stuff so much. ...I just wish I could understand it better as parent and a professional.”

A couple of weeks later, in July 2023, leadership held a meeting to discuss Director Somasundaram’s findings and recommendations. Sometime that day, Deputy State Superintendent Johnson emailed that DPI should lower the 10th grade mathematics cut scores by 1 point for Basic and Proficient. DPI ultimately did not take his suggestion.²³ In reply, Superintendent Underly wrote, “I’m going to need a primer on this, or a tutoring session. I still don’t understand it, and it’s just a learning block that I have on my end.” She had approved the PreACT cut scores by the next day.

This communication reveals very important things. Superintendent Underly struggled to grasp the options put before her that would affect parents’ and educators’ understanding of their kids’ preparedness. She appeared not to know how a regular person should understand the meaning of proficiency. Most importantly, she requested changes to cut scores before understanding the basics of what she would be changing.

Additionally, the “nonsense” Superintendent Underly referred to coincides with a monumental reading law, Act 20, introduced that same month by Republican legislators in collaboration with DPI.²⁴ Governor Evers signed it into law a month later.²⁵ Act 20 copies leading literacy states like Florida, Massachusetts, and Mississippi in hopes of raising Wisconsin’s overall academic competitiveness and closing horrific achievement gaps.²⁶ Given previous cooperation between the Legislature and DPI, it is not clear whether she called Act 20 itself nonsense, whether she considered specific facets of Act 20 nonsense, like the provisions that connect to state test scores, or whether she was referring to some other literacy matter entirely.

²³ Wisconsin Department of Public Instruction, “[PreACT Secure Data and Results](#),” November 29, 2023.

²⁴ Wisconsin State Legislature, “[2023 Wisconsin Act 20](#),” July 19, 2023.

²⁵ Corrinne Hess, “[Gov. Tony Evers Signs Sweeping Reading Literacy Bill into Law](#),” *Wisconsin Public Radio*, July 19, 2023.

²⁶ Kymyona Burk, “[Thompson Center Summit on Early Literacy](#),” March 6, 2023.

Over the next months, however, Superintendent Underly and DPI made clear their stance on Wisconsin's cut scores: they were too high, misleading parents, underrating children, and making DPI look bad. When 2023 test scores released, DPI's press release included a paragraph enforcing that point. "I am also tired of politicians claiming that our children aren't learning because they aren't reaching a proficiency score. Instead of using test scores as a cudgel, we should all take the time to learn what a high bar proficiency on this test represents, because the truth is that our proficiency cut scores are very high in comparison to every other state in the country," Underly argued.²⁷ The day before Milwaukee Public Schools' board discussed its test scores at a committee meeting, then-Communications Director Abigail Swetz emailed the district links to Superintendent Underly's comment and NAEP's research showing that Wisconsin's proficiency standards most closely matched NAEP's high proficiency standards.²⁸ Director Swetz also notified journalist Ruth Conniff that Superintendent Underly would address measurements of success at her Wisconsin State Education Convention speech. A few months later, after DPI released test scores, Superintendent Underly wrote to now-Associate Deputy State Superintendent McCarthy, Deputy State Superintendent Johnson, Executive Director Beekman, and new Executive Director Sachin Chheda, "Before I jumped off- yes, I'd like to have a conversation about our cut scores being the highest in the country and how we communicate what that translates to with NAEP." Soon afterward, addressing 2024 Forward cut score changes, Superintendent Underly wrote, "I also want to, and maybe this exists as a possibility, to norm our levels to be similar to other states... as I hear that we have some of the highest cut scores nationally. I would like them to be looked at so that we aren't judged negatively when we have direct standards." Finally, in a lengthy presentation to Superintendent Underly and her cabinet, the Office of Educational Accountability displayed Wisconsin's state test scores compared to other states'. Wisconsin, which has nationally aligned, higher standards than nearly all states, regularly appeared near the bottom of the slides' charts. That led to the discussion question for the cabinet, "Does Cabinet have a lower or upper threshold for [cut score] recommendations? Forward? ACT?"

In the longest available record, Superintendent Underly also appeared to misunderstand how NAEP works and distrusted statistical sampling. After calling for more federal spending on education to address student disadvantages, Superintendent Underly mused, "I also said that NAEP only tests 10,000 nationally, and that's really an awful way to make generalizations about how kids are doing, and that when we use state assessments to compare kids state by state, it's not an apples to apples comparison - so if the feds want to get in the assessment game, they should really make it so that we all take the same assessment and have the same cut scores. Because otherwise it's not very useful information. ...But my feelings on high stakes tests are pretty well known."

Again, Superintendent Underly criticized high-stakes tests. Despite this, she did not wish Congress abolished state tests. She posed that Congress could institute a unified, national test given to all students, as countries with national academic standards instead of differing state ones often do.

²⁷ Wisconsin Department of Public Instruction, "[DPI Releases Wisconsin Student Assessment System Results](#)," October 10, 2023.

²⁸ National Center for Education Statistics, "[Data Tool: Mapping State Proficiency Standards onto NAEP Scales, 2007-2019](#)," September 22, 2021.

The Superintendent's NAEP criticisms are of more interest. NAEP is produced, administered, and analyzed by the Department of Education.²⁹ They extensively document its sampling methods,³⁰ its scaling methods,³¹ and its documentation methods,³² they simply answer frequent questions,³³ and they illustrate its methods as well.³⁴ The Department of Education can answer any questions state education leaders have about the largest, longest-term, nationally representative assessment of student performance. Additionally, NAEP quizzes approximately 110,000 students nationally in each grade to compare states, not 10,000.³⁵ They do require only 10,000 to 20,000 students for their national sample, but that produces an overall score for America, not the individual state scores to which Superintendent Underly refers.³⁶ A national political poll, for instance, requires fewer respondents than one comparing all 50 states' results to each other.

No leader is a subject-matter expert on every facet of his or her job. But a pattern appears: a rush to leave behind the standard that every child will be prepared for college, but insufficient understanding of what those standards mean and why they may be significant.

What matters? Months after moving to lower test score standards, Superintendent Underly needed significant assistance to understand what standards were and how changing them would affect the number of students grouped into each performance tier. Superintendent Underly disapproved of Wisconsin's cut scores, which align with national NAEP standards and college-readiness standards, and preferred to lower them to match other states'. Superintendent Underly thought the NAEP, administered by experts at the Department of Education, lacks legitimacy. Superintendent Underly referred to "nonsense going on with literacy" at the same time DPI was cooperating with the Legislature and Governor Evers to pass transformational changes in how students read, emulating the states with the highest-performing schools.

²⁹ National Center for Education Statistics, "[About the NAEP Reading Assessment](#)," September 16, 2022.

³⁰ National Center for Education Statistics, "[Statistical Significance and Sample Size](#)," October 14, 2021.

³¹ National Center for Education Statistics, "[NAEP Analysis and Scaling](#)," March 28, 2024.

³² National Center for Education Statistics, "[NAEP Data Collection](#)," March 6, 2024.

³³ National Center for Education Statistics, "[NAEP State Assessment Sample Design Frequently Asked Questions](#)," November 30, 2023.

³⁴ National Center for Education Statistics, "[Assessment Literacy](#)," July 1, 2024.

³⁵ National Center for Education Statistics, "[Appendix Tables for 2022 Reading Report Card](#)," September 16, 2022.

³⁶ National Center for Education Statistics, "[Frequently Asked Questions](#)," February 20, 2024.

Changing the Tests

New standards for the Forward and ACT tests had clear implications for students, and DPI had the final word on cut scores.

In October and November of 2023, DPI began to receive information about setting standards for the 2024 Forward Exam. Data Recognition Corporation, the company that produces the Forward Exam, would run a summer standard-setting session from June 11th to June 14th, 2024. They would select and guide teachers to help DPI establish cut scores that “reflect the updated state content standards, link students’ scores on the tests to the state’s expectations for students in each performance level, and are well-articulated across grades.”

96 educators would split into six groups of 16, each covering either reading or mathematics across two grade levels. Participants would receive descriptions of the knowledge and skills students should have at each grade level, increasing over the four performance tiers of Below Basic, Basic, Proficient, and Advanced. After orientation and training, individuals would write descriptions of the minimum requirements a Basic student, a Proficient student, and an Advanced student should have in order to set the floor for their scoring. After discussing and editing descriptions within their groups and the room, teachers would review the test questions students saw, which DRC ordered from easiest to hardest. Educators would mark the first question they think a Proficient student would know, but a Basic student would not. They would repeat that for Advanced and Basic, then write justifications for their decisions based on the expectation descriptions. DRC would show each group’s median cutoff question for each performance tier. Based on the groups’ chosen cutoffs, DPI would have the option for DRC to show the groups their 2024 benchmarks compared to 2023’s. Either way, groups would debate if they liked the current cutoffs, then would reset them if they would desire. DRC would show each group’s new median cutoff question for each performance tier. Based on the groups’ chosen cutoffs, DPI would have the option for DRC to show the groups the proficiency percentages their cutoffs would produce. Individuals would return to their groups one last time to debate the cutoffs, reset them if desired, and set their “final draft” of cut score recommendations. The whole group would look at the cut scores and proficiency percentages from grades 3 through 8 and provide feedback on such things as unevenness, extremism, or inconsistency. To conclude, one representative from each group would meet with DRC to debate whether they should tweak any cut scores, using the final draft feedback. However, DPI ultimately would decide cut scores based on the groups’ recommendations, “internally and with stakeholders.”

In March 2024, DPI began to receive information about setting standards for the 2024 ACT. ACT, the company formerly known as American College Testing that produces the ACT, would run a summer standard-setting session from July 10th to July 11th. The ACT standard-setting process would be very similar to the Forward’s.

30 educators would split into three groups of 10, each covering reading, mathematics, or science. Participants would consider DPI's descriptions of the four performance tiers of Below Basic, Basic, Proficient, and Advanced as they relate to 11th graders. However, ACT would not share the ACT questions with participants for them to bookmark. Rather, ACT has rigorous data that shows the likelihood a student would receive an A, a B, and a C in college freshman reading or mathematics courses, based on their ACT score. After orientation and training, individuals would pick the percentage chance a student would have of receiving a certain letter grade based on DPI's definition of Proficient, then discuss as groups. ACT would show each group's cutoff definition of Proficient, which ACT score matches their cutoff based on their data, and what percentage of students would be considered Proficient as a result. For example, ACT considers its own college-readiness definition "a 50% chance of getting a B," which corresponds to a 22 cutoff on the ACT mathematics section. Groups could then revise their cut scores based on those results as well as contextual data like ACT's grade projections, past Wisconsin ACT scores, other states' ACT scores, Forward scores, NAEP scores, college enrollment rates by ACT score, and Wisconsin colleges' ACT cutoffs for course placement and scholarships. Groups would repeat the setting and revision process for Advanced and Basic. To conclude, the room would finalize cut score recommendations after reviewing the entire range. However, DPI ultimately would decide cut scores based on the groups' recommendations.

The rigorous processes make clear that DPI fully understood the implications of any cut scores it produced. Even though DRC would work with DPI to reset Forward cut scores later in the year, DRC stated they would send preliminary data to DPI on the percentage of students in each performance tier in May and June, before new standard setting occurred. Additionally, DPI would have impact data after the standard-setting process, as its participants did. On the ACT, DPI would have rigorous data on the likelihood of college success based on each cut score.

What matters? Teachers who set Forward and ACT cut scores understood in real time how their choices would affect proficiency rates. Teachers who set Forward Exam cut scores may have referenced the previous year's proficiency rates, assuming DPI allowed them access to that data during the session. According to the exam contract, DPI would receive Forward performance data before educators suggested new cut scores, though DPI will only release proficiency scores under the changed cutoffs.

Widening the Achievement Gap

Superintendent Underly and her cabinet had clear evidence in April 2024 that lowering test cut scores would disproportionately help advantaged groups, widen achievement gaps, and inflate report card ratings.

In April 2024, an innocuous report became a problem. During the previous months, DPI had debated the merits and legality of changing the names of each student performance tier, which they eventually did. Students would no longer score Below Basic, Basic, Proficient, or Advanced, but be Developing, Approaching, Meeting, or Advanced.³⁷ Much graver concerns would soon press them.

Lowering the bar for student proficiency on state tests would have drastically inequitable effects on achievement gaps and accountability. As Director Somasundaram summarized, “Assuming these standard setting activities raise proficiency rates statewide, it is reasonably likely that: (1) proficiency-based achievement gaps will increase, (2) report card Overall Scores will increase year-over-year, and (3) increases in Overall Score and Overall Rating category will impact schools disproportionately (report card rating increases will benefit low-poverty schools more than high-poverty schools).”

Proposed changes to state tests would make Wisconsin’s achievement gaps even more brutal. Raising state proficiency from 40% to 50% likely would cause “student groups who have been traditionally underserved [to] see smaller gains in proficiency rates as compared to more advantaged groups, widening the gaps between them,” projected Director Somasundaram. Her team predicted new proficiency levels would increase gaps between 3 to 4 points for economically disadvantaged students and between 5 to 10 points for special-needs students. Director Somasundaram broke down the growing disparities by race. In reading, she projected White students would score 59%, Asian students would score 49% (falling 1% further behind), multiracial students would score 46% (1%), Pacific Islander students would score 42% (2%), Hispanic students would score 32% (2%), Indigenous students would score 28% (2%), and Black students would score 18% (4%). In mathematics, she projected White students would score 61%, Asian students would score 51% (falling 3% further behind), multiracial students would score 43% (3%), Pacific Islander students would score 39% (4%), Hispanic students would score 29% (4%), Indigenous students would score 26% (5%), and Black students would score 14% (7%). In addition, comparisons to past proficiency rates would be unhelpful. Director Somasundaram’s team estimates 13% more White students would reach proficiency in mathematics and Black students still adding 6% more.

³⁷ Wisconsin Department of Public Instruction, “[Wisconsin Student Assessment System Updated Asset-Based Performance Levels](#),” June 6, 2024.

The federally required state report cards would rise to the point of uselessness. Director Somasundaram wrote, “School and district report card scores and ratings will also rise if proficiency rates increase. If proficiency rate gains are substantial,” using 55% statewide proficiency, “we project report card ratings will increase to such a degree that there will be little difference between schools in terms of their ratings.” Accompanying graphs predicted a rapid rise over the 3-year phase-in period. Assuming no improvement or decline in student performance, Director Somasundaram estimated the percentage of 4- or 5-star schools would rise from 46% to 53% in 2024, 60% in 2025, and 63% in 2026. Director Somasundaram estimated the percentage of 4- or 5-star districts would rise from 36% to 50% in 2024 and 71% in 2026.

The accountability report cards would score high-poverty schools more harshly than low-poverty schools. Director Somasundaram continued, “In addition, these report card rating increases will benefit low-poverty schools more than high-poverty schools due to variable weighting of Achievement and Growth priority areas, as required by state law. ...As a result, rising proficiency and PBP rates will raise report card scores and ratings to a greater degree for schools and districts with lower percentages of ECD students.” Director Somasundaram predicted that 45% of low-poverty districts like Whitefish Bay, DeForest, and Kohler eventually would gain an additional star on their ratings, while only 23% of high-poverty districts like Cudahy, Beloit, and Green Bay would.

Director Somasundaram provided a potential avenue to fix this lopsided benefit to advantaged students. Simply, DPI would severely change the state report card ratings in 2025. Step 1: “Keep report card calculations and rating thresholds unchanged for the 2023-24 school year as we anticipate relatively moderate increases this year,” referring to a 14% rise in highly rated districts, disproportionately those with lower poverty, that would rate half of all Wisconsin districts 4 or 5 stars. Step 2: “For the 2024-25 school year... include only two years of assessment data in achievement calculations” rather than three, which “would shorten the timeline to full impact of new assessment cuts. Adjust Growth and TGO rescaling formulas to align these priority area score distributions with the new Achievement priority area score distribution.” Unfortunately, Director Somasundaram foresaw those changes alone would result in 80% of schools and 93% of districts scoring 4 or 5 stars on state report cards in 2025. So, step 3: “Prior to release of 2024-25 report cards, conduct a report card standard setting to establish new report card rating thresholds.” While achievement gap growth “cannot be addressed” without raising proficiency cut scores, “OEA would be able to address some of the challenges to the report card system to a certain extent.”

These projections, previewed in 2023 but made explicit while the 2024 Forward Exam was going on, would have mammoth implications for Wisconsin education and politics.³⁸ It would change report card scoring for the second time in four years, likely initiating a similar political battle.³⁹ The visibility of Wisconsin's achievement gap would grow, and districts that set long-term goals to close those gaps would lose ground and be forced to reset expectations. The PreACT cut scores, which DPI had set just one year earlier, would reset to match the Forward and ACT proficiency rates. DPI, whose accountability plan with the federal government already needed renewal, would find any achievement gap closure goal much more difficult to achieve. Massive increases in proficiency, especially in advantaged schools in Wisconsin, would prevent comparisons to student success before the pandemic. Shortly thereafter, Deputy State Superintendent Johnson posed to the team, "Could we push out a decision for a year on - new cut scores - on new nomenclature?" as they grappled with the ripple effects of cut score changes.

What matters? Despite rhetorically treating equity as an essential value, DPI lowering proficiency cut scores on state tests was predicted to increase racial, economic, and disability achievement gaps. Lowering cut scores was projected to cause widespread inflation of state report card ratings, disproportionately favoring low-poverty schools. DPI proposed a 2-step solution: increasing lenience on growth and target group report card scores, primarily favoring high-poverty schools, then significantly altering report card scores. Failing to rescore report cards was projected to result in 80% of schools and 93% of districts scoring 4 or 5 stars.

³⁸ Wisconsin Department of Public Instruction, "[Wisconsin Student Assessment System Updated Asset-Based Performance Levels](#)," February 27, 2024.

³⁹ Rory Linnane, "[Did State Officials 'Rejigger' School Scores to Make Them Look Better, as Lawmakers Say?](#)" *Milwaukee Journal Sentinel*, March 1, 2022.

Aftermath

Early indicators affirm planned changes were carried through, greatly changing the appearance of student achievement in Wisconsin.

After all of these changes, what happened to Wisconsin's students? Test scores and report cards have not yet gone public, and public records ended in April. However, some effects of DPI's cut score changes are apparent.

There is evidence that widespread proficiency increases as well as achievement gap increases on the Forward Exam have come to pass. 13.7% fewer Green Bay middle schoolers scored in the lowest category in 2024.⁴⁰ Appleton's achievement gap for students with disabilities grew 7.9% in English language arts and 7.6% in mathematics, 8.6% and 5.7% for students whose first language is not English, and 4.3% and 4.0% for Black students.⁴¹ White middle schoolers in Madison scored 77% and 73% proficient in reading and mathematics, increasing achievement gaps an additional 5% and 11% for Hispanic students and an additional 7% and 13% for Black students.⁴²

DPI had just set PreACT cut scores in 2023, but lowered them to match cut score changes on the Forward and ACT.⁴³ Proficiency cut scores in English language arts fell from 16 and 18 for 9th and 10th graders to 15 and 17. Proficiency cut scores in mathematics fell from 17 and 19 for 9th and 10th graders to 16 and 17.⁴⁴ In other words, DPI adjusted its expectations for a 10th grader's mathematics skills to match the previous 9th grade standard.

On the ACT, DPI has clearly defined a standard for success in high school that does not align with success in college. Previously, DPI set a 20 in English language arts and a 22 in mathematics as proficiency.⁴⁵ Now, DPI marks proficiency as a 19 and 19.⁴⁶ According to ACT's data, students considered proficient in 2024 would have a -2.5% smaller chance of getting an A, a -2.9% smaller chance of getting a B, and a -1.3% smaller chance of getting a C in freshman English Composition I compared to students considered proficient in 2023. Students considered proficient in 2024 would have an -9.0% smaller chance of getting an A, a -15.5% smaller chance of getting a B, and a -9.7% smaller chance of getting a C in freshman Algebra compared to students considered proficient in 2023.⁴⁷

⁴⁰ Green Bay Area Public Schools, [Board of Education work session](#), August 12, 2024.

⁴¹ Appleton Area School District, [Board of Education meeting](#), September 9, 2024.

⁴² Madison Metropolitan School District, [Board of Education instruction work group](#), September 9, 2024.

⁴³ Wisconsin Department of Public Instruction, "[PreACT Secure Data and Results](#)," September 1, 2024.

⁴⁴ Wisconsin Department of Public Instruction, "[PreACT Secure Data and Results](#)," November 29, 2023.

⁴⁵ Wisconsin Department of Public Instruction, "[ACT Data Proficiency](#)," June 9, 2023.

⁴⁶ Wisconsin Department of Public Instruction, "[The ACT Data and Results](#)," September 1, 2024.

⁴⁷ Jeff Allen, Justine Radunzel, and Joann Moore, "[Evidence for Standard Setting: Probabilities of Success in 'Benchmark' College Courses, by ACT Test Scores](#)," May 12, 2017.

The report cards will change in 2024 and 2025. DPI has attempted to suppress the extreme outcomes of sharply rising test scores this year, and educators will recut report card standards in 2025 “to accurately communicate the success of schools and districts.”⁴⁸

Journalists have focused on the performance tier naming conventions, and DPI has made a strong, student-centric case. “We’re not the only state to make changes to better define and put less blame and shame onto the names,” explained now-Deputy Superintendent McCarthy. “You can call it ‘rose-colored glasses,’ but if I have to explain to my kindergartener that she’s ‘below basic’ in a task? Those words have a sting. They have a specific meaning. The Department feels like there is value in being able to have a conversation that is growth-mindset-oriented. ...So, we’re going to stick on the side of kids on that...”⁴⁹ State Superintendent Jill Underly underlined the mass confusion around whether “proficiency” means “on grade level.” “I’ve often heard confusion from parents, families and legislators on what performance terms on tests meant in regard to where students are at academically.”⁵⁰

However, how DPI has improved information for families is unclear. If students can be Developing, Approaching, Meeting, or Advanced, which one tells parents that their child is prepared with grade-level skills? That is a hard question for Deputy Superintendent McCarthy and now-Senior Policy Advisor Johnson, who said grade level probably falls between Meeting and Approaching.⁵¹ To rephrase that, “Parents now have no simple way to interpret their child’s end-of-year test scores.” DPI continues to deemphasize the importance of the federally required, summative test. Deputy Superintendent McCarthy and Senior Policy Advisor Johnson told media that screener tests, alternative evaluations, or parent-teacher conferences provide richer and deeper insight to parents.⁵² “I think [test scores] are an imperfect tool. I don’t know of a perfect tool for a parent... We’ve engaged in this oversimplification process for a long time to say, ‘Let’s roll up a bunch of data and make some scientific calculations and...I can look at my school and say it’s 4 stars... I think, if you really want to have that conversation, it does require parents to get their hands a little dirty, look at the system, and figure that out.’”⁵³ Is it equitable to ask low-income and single parents, whose school decisions are of the most immense consequence, to come off third shift and dig into aimswebPlus growth graphs? Would Wisconsin officials tell citizens navigating unemployment benefits or BadgerCare, “Get your hands a little dirty, look at the system, and figure it out,” or is that what bureaucrats are paid to do?

What matters? While test scores and report cards release later in the year, early indicators support the surging scores and widening achievement gaps DPI predicted in records. DPI will alter school accountability report cards in 2025. DPI now deems a high school junior proficient who has a 37% chance of earning a D or F in college freshman Algebra. Changes to state tests and report cards may confuse parents more, not less.

⁴⁸ Wisconsin Department of Public Instruction, “[What’s New for the 2023-24 School and District Report Cards?](#),” September 27, 2024.

⁴⁹ Erin Richards, “[Newsmakers: K-12 Education on the Campaign Trail](#),” *WisconsinEye*, August 26, 2024.

⁵⁰ Alan J. Borsuk, “[Are Wisconsin Students Really Doing Better? Or Does It Just Look That Way?](#)” *Milwaukee Journal Sentinel*, August 30, 2024.

⁵¹ Alan J. Borsuk, “[Does Lowering Cut Scores and Changing Terminology on Standardized Tests Better Serve Wisconsin Students?](#)” *Milwaukee Journal Sentinel*, August 23, 2024.

⁵² Alan J. Borsuk, “[Does Lowering Cut Scores and Changing Terminology on Standardized Tests Better Serve Wisconsin Students?](#)” *Milwaukee Journal Sentinel*, August 23, 2024.

⁵³ Erin Richards, “[Newsmakers: K-12 Education on the Campaign Trail](#),” *WisconsinEye*, August 26, 2024.

Three important questions remain for DPI.

Will lower cut scores mislead parents and diminish urgency for instructional and financial improvement? Deputy Superintendent McCarthy said Wisconsin did not benefit from higher cut scores. “We’re returning to something that looks more like our neighboring states.”⁵⁴ The largest long-term academic gains in our lifetimes coincided with the introduction of demographically differentiated, high-stakes state tests.⁵⁵

Will report cards give high marks to a skyrocketing number of schools? DPI said in August, “Any potential impact on state report cards is unknown at this time, as we are working on analyzing data that will help calculate report cards this fall.”⁵⁶ Deputy Superintendent McCarthy promised, “We will make sure [report cards] have the stability...[for] communities to say, ‘This is where we stand.’”⁵⁷ Those statements differ from the projections DPI staff provided in April.

What is the plan to make Wisconsin’s advantaged children nationally competitive and give disadvantaged kids a fighting chance? However we measure students, as Deputy Superintendent McCarthy laid out, “The question becomes, ‘What are we doing with them?’”⁵⁸ Essential programs like reading reforms and teacher apprenticeships are withering. Are criticisms of the prolonged wait for these improvements “partisanship,” too?⁵⁹

One truth is constant amidst all of this change: common citizens ultimately decide the future of young generations. What we accept for our children - our expectations, our commitments, and our opportunities - determine their fates. Will Wisconsin be a place where families settle down and businesses crowd in? Or will we rust and fade, a place people drive past on their way to prosperity?

⁵⁴ Alan J. Borsuk, “[Does Lowering Cut Scores and Changing Terminology on Standardized Tests Better Serve Wisconsin Students?](#)” *Milwaukee Journal Sentinel*, August 23, 2024.

⁵⁵ National Center for Education Statistics, “[Long-Term Trends in Reading and Mathematics Achievement](#),” September 12, 2022.

⁵⁶ Alan J. Borsuk, “[Does Lowering Cut Scores and Changing Terminology on Standardized Tests Better Serve Wisconsin Students?](#)” *Milwaukee Journal Sentinel*, August 23, 2024.

⁵⁷ Erin Richards, “[Newsmakers: K-12 Education on the Campaign Trail](#),” *WisconsinEye*, August 26, 2024.

⁵⁸ Alan J. Borsuk, “[Are Wisconsin Students Really Doing Better? Or Does It Just Look That Way?](#)” *Milwaukee Journal Sentinel*, August 30, 2024.

⁵⁹ Sachin Chheda, [post](#), X, September 8, 2024.

Email Appendix

From: [Johnson, John W. DPI](#)
To: [McCarthy, Thomas G. DPI](#); [Jones, Keona S. DPI](#); [Ranabhat, Aastha DPI](#)
Cc: [Beekman, Demetri A. DPI](#)
Subject: RE: Re-examining our cut scores
Date: Monday, January 23, 2023 9:08:12 AM
Attachments: [image001.png](#)

I will add that the change to the Forward assessment to make it aligned to our new revised standards in ELA, and Math (along with the item development in that) makes this a potentially timely piece.

JJ

From: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Sent: Monday, January 23, 2023 8:34 AM
To: Jones, Keona S. DPI <Keona.Jones@dpi.wi.gov>; Ranabhat, Aastha DPI <Aastha.Ranabhat@dpi.wi.gov>
Cc: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Subject: Re-examining our cut scores

Dr. Jones and Aastha,

Dr. Underly is interested in taking another look at how we establish our cut scores for the Forward Exam. Given that we're amid some other change in that area, we're asking for a decision paper on what it would take to change the cut scores, as well as what timeline we'd be working off. If you feel like we should meet to discuss further, let me know. Thanks!

-tom



Tom McCarthy (he/him/his) | Executive Director
Wisconsin Department of Public Instruction
125 South Webster Street | Madison, WI
608-266-2630 | dpi.wi.gov

From: [Lee, Alanna R. DPI](#)
To: [Jones, Keona S. DPI](#)
Subject: Message from Viji in Teams Chat
Date: Monday, January 23, 2023 3:35:34 PM

Keona,

Below is the message Viji sent me via Teams Chat.

“As we enter a new contract cycle this is an opportune time to discuss setting new cut scores (standard setting) for the Forward Exam. Standard settings are typically done when there are changes to the standards or assessments. We plan to assess the revised ELA and mathematics standards on the Forward Exam beginning in spring 2024 (as John Johnson mentioned). Therefore, it could be appropriate to do a new standard setting in the summer of 2024. We are putting together a more detailed brief with options and a timeline for all the assessed content areas.”

Alanna Lee

Executive Staff Assistant

Division for Student and School Success

Wisconsin Department of Public Instructions

125 South Webster Street

Madison, WI 53703

alanna.lee@dpi.wi.gov

608-266-5199

From: [Johnson, John W. DPI](#)
To: [Underly, Jill K. DPI](#); [Beekman, Demetri A. DPI](#); [McCarthy, Thomas G. DPI](#)
Subject: FW: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval
Date: Thursday, June 1, 2023 1:52:13 PM
Attachments: [WSAS Standard Setting Plan May 30, 2023.docx](#)
[2023 PreACT Secure Cut Scores Approval - May 30, 2023.docx](#)

Please read the attached documents for our assessment folks. We should discuss this for a few minutes at Exec next week.

From: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>
Sent: Tuesday, May 30, 2023 11:50 AM
To: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>
Cc: Jones, Keona S. DPI <Keona.Jones@dpi.wi.gov>
Subject: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

Good morning John,

A few months ago, Dr. Underly requested a paper on the standard setting. Tom McCarthy emailed Keona regarding this request. We have worked on this attached paper, which includes a background overview of the standard settings for the past years, the upcoming couple of years, and the PreACT Cut Score recommendation.

Our students took the Grade 9 and 10 PreACT assessment for the first time during Spring 2023, and we recently received preliminary data files from ACT. As with the ASPIRE assessment, we have statistically established Cut Scores for PreACT. We look forward to the State Superintendent's approval.

Once we receive the approval, we will provide the Cut Scores to our data warehouse team to move forward with the dashboard development work for the PreACT. In early July we will QA the final data file and review these Cut Scores again to see if we need to make any changes. If any changes are needed at that point, I will share the updated Cut Scores for approval and update the data warehouse team to make the necessary changes.

Please let us know if you have any questions.

Thank you,
Viji.



To: *Jill K. Underly, PhD, State Superintendent, DPI*
From: *OEA Staff*
Date: *Tuesday, May 30, 2023*
Subject: *Wisconsin Student Assessment System (WSAS) Standard Setting Plan*

Context:

The Office of Educational Accountability (OEA) seeks approval for the standard setting plan below.

Background:

Standard setting is the process of determining the test scores (cut scores) that divide performance on a test into categories used for reporting student achievement. These performance categories are, in turn, tied to descriptions intended to capture what typical students at each level know and can do in relation to the content standards being assessed. Standard settings are typically conducted when assessments change significantly. The process is ultimately a matter of informed judgment rather than discovering objective performance standards. Recommended cut scores are submitted to the State Superintendent for review, adjustments may be made, and formal approval is documented. Cut scores must be reasonable and defensible in terms of the process, public and educator perception, and related other data (for example, comparisons to trend data, vertical articulation across grades, or data from other assessments, such as the National Assessment of Educational Progress (NAEP)).

Considerations:

Accountability

The School and District Report Cards form the basis of the state accountability system. The current report card design and accountability rating categories have been in place since the 2020-21 report cards.

Implementation of new ELA and Mathematics cut scores for the Forward Exam in 2024 and PreACT Secure in 2023 will impact report card scores and ratings. The direction and magnitude of those impacts will depend on the magnitude of change in the percentage of students in each of the performance levels (Below Basic, Basic, Proficient, and Advanced). Generally speaking, for report cards statewide, if the standard setting process increases the number of students in the Proficient and Advanced categories, report card scores will increase and some schools and districts will see their ratings increase as well. Conversely, if the standard setting decreases those Proficient and Advanced percentages, scores and ratings will go down.

These impacts will not be felt equally by all schools and districts, however. Schools and districts with the highest percentages of economically disadvantaged (ECD) students will see little change, because of the statutory requirement to decrease the weight of the Achievement score relative to the Growth score as percent ECD increases. Achievement scores are sensitive to performance level percentage changes, while Growth scores are not. In addition, high schools will not see their scores and ratings change as much, because this standard setting would not affect ACT performance level cuts. Furthermore, if the PreACT Secure standard setting holds performance level percentages relatively steady, then high schools will see few score or rating changes.

Finally, because report cards use three years of assessment data to arrive at the Achievement score, any impacts will be increasing in magnitude for three years before they stabilize. Stabilization would not occur until the 2025-26 report cards when all three years of assessment data used in the Achievement priority area utilize the updated standards.

Potential Impact on ACT Cut Scores

ACT Aspire and ACT with writing cut scores were set after the first administrations in the spring of 2015. As new cut scores are established on the Forward Exam for grades 3-8 and the PreACT Secure for grades 9-10 (PreACT Secure has now replaced the ACT Aspire

3



WISCONSIN DEPARTMENT OF
Public Instruction

Jill K. Underly, PhD, State Superintendent

assessment), it may be necessary to revisit the cut scores for the ACT to ensure coherence across all grade levels.

Perceptions of Reasonableness

Overall proficiency rates, which are the percentage of students who achieve either proficient or advanced status, are one of the key metrics by which educators and the public evaluate the performance of the schools in Wisconsin. Therefore, DPI must consider whether the current proficiency rates accurately reflect students' level of knowledge and skills related to the grade-appropriate academic standards.

Forward Exam cut scores in ELA and Mathematics were last set in 2016. These cut scores were informed by the cut scores set for the Wisconsin Knowledge and Concepts Examinations (WKCE) in 2012-13, which were intentionally made comparable to the cut scores for the National Assessment of Educational Progress (NAEP) in response to "mapping studies" showing many states with "lower standards" than NAEP.

PreACT Secure options:

1. Use the same proficiency rate distribution as Aspire.
2. Use one or more of the PreACT Secure college-readiness benchmarks.
3. Use some combination of using the college-readiness benchmark for Proficient and extrapolation/interpolation using the proficiency data from the grade 8 Forward Exam and the ACT.

Option 1: A and B - Use ACT Aspire Results from Prior Years to Establish Cut Scores

ACT Aspire results from spring 2019 and spring 2022 show a difference in performance across proficiency levels. The distribution of results from these administrations is matched with spring 2023 performance and corresponding cut scores have been identified. The ACT Aspire results columns identify the percent of students scoring at the respective performance level for these examinations. The PreACT Secure based columns display the potential cut scores that DPI could establish if the desire is to match performance to results from these administrations. The 2023 impact data columns display results if the cut scores would be set at the corresponding levels.

Math*						
	Option A - 2022 Aspire as Baseline			Option B - 2019 Aspire as Baseline		
	ACT Aspire 2022 Results	PreACT Secure Based off of 2022 Aspire	2023 Impact Data	ACT Aspire 2019 Results	PreACT Based off of 2019 Aspire	2023 Impact Data
Grade 9	% at each level	Performance Level Scores	% at each level	% at each level	Performance Level Scores	% at each level
Advanced	13.0%	23 - 35	12.6%	11.1%	24 - 35	10.4%
Proficient	24.8%	17 - 22	26.9%	34.6%	17 - 23	29.1%
Basic	27.4%	15 - 16	35.4%	26.6%	15 - 16	35.4%
Below Basic	34.9%	1 - 14	25.0%	27.7%	1 - 14	25.0%
Grade 10						
Advanced	12.5%	25 - 35	12.6%	9.8%	26 - 35	10.3%
Proficient	23.7%	19 - 24	21.5%	28.5%	19 - 25	23.8%
Basic	24.4%	16 - 18	28.9%	28.3%	16 - 18	28.9%
Below Basic	39.4%	1 - 15	37.0%	33.4%	1 - 15	37.0%

* Preliminary data - final data file to arrive from vendor on May 27, 2023.

ELA*						
	Option A - 2022 Aspire as Baseline			Option B - 2019 Aspire as Baseline		
	ACT Aspire 2022 Results	PreACT Secure Based off of 2022 Aspire	2023 Impact Data	ACT Aspire 2019 Results	PreACT Based off of 2019 Aspire	2023 Impact Data
Grade 9	% at each level	Performance Level Scores	% at each level	% at each level	Performance Level Scores	% at each level
Advanced	8.4%	25 - 35	8.7%	9.5%	24 - 35	10.6%
Proficient	33.4%	16 - 24	36.5%	35.6%	16 - 23	34.6%
Basic	34.1%	13 - 15	26.1%	33.0%	12 - 15	37.4%
Below Basic	24.2%	1 - 12	28.6%	21.9%	1 - 11	17.3%
Grade 10						
Advanced	10.0%	26 - 35	11.4%	10.5%	26 - 35	11.4%
Proficient	30.9%	18 - 25	30.3%	32.4%	18 - 25	30.3%
Basic	33.2%	13 - 17	35.1%	32.1%	13 - 17	35.1%
Below Basic	26.0%	1 - 12	23.2%	25.0%	1 - 12	23.2%

* Preliminary data - final data file to arrive from vendor on May 27, 2023.

Option 2 - Use PreACT Secure Readiness Benchmarks to Establish Proficient Cut Score

The PreACT Secure On Target benchmark is the score associated with a 50% chance of meeting the ACT College Readiness Benchmark in 11th or 12th grade and is the minimum score of the On Target range.

7



Jill K. Underly, PhD, State Superintendent

For math, the On Target benchmark is set at 18 for the spring of grade 9 and 19 for the spring of grade 10. Preliminary spring 2023 administration data is found in the table below.

ACT has not established a combined ELA benchmark on the PreACT Secure. English and Reading benchmarks have been established however, and OEA staff have combined these benchmarks to create an ELA benchmark. For spring of grade 9, the On Target benchmark for English is 13 and the On Target benchmark for Reading is 18. Averaging these scores results in an ELA On Target benchmark of 16. For spring of grade 10, the On Target benchmark for English is 15 and the On Target benchmark for Reading is 20. Averaging these scores results in an ELA On Target benchmark of 18. Preliminary spring 2023 administration data is found in the table below.

PreACT Secure College Readiness Benchmarks				
	Math		ELA	
	On Target Benchmark	% At or Above Benchmark - Spring 23	On Target Benchmark	% At or Above Benchmark - Spring 23
Grade 9	18	31.2%	16	45.3%
Grade 10	19	34.1%	18	41.7%

Option 3 - Use some combination of using the college-readiness benchmark for Proficient and extrapolation/interpolation using the proficiency data from the grade 8 Forward Exam and the ACT.

Option 3 was evaluated and not considered due to the new Forward Exam administration and standard setting in spring 2024.

Recommendation:

OEA staff recommend adoption of option 1B for the Basic, Proficient and Advanced cut scores for the PreACT Secure for both math and ELA. Using comparison data from 2019 as the basis for adopting these cut scores reflects that expectations of performance have not changed over time. A similar proportion of students in each grade and subject will be identified as Proficient and Advanced using this recommendation. For Basic performance, the results are not nearly as consistent across grades and subjects. However, large

8



Jill K. Underly, PhD, State Superintendent

percentages of students are receiving scores either directly at or one score point below the recommended cut scores. As a result, there will be a large change in percent of students at both the Below Basic and Basic performance levels if the Basic cut score were to be adjusted.

Recommended PreACT Secure Cut Scores				
	Below Basic	Basic	Proficient	Advanced
Math Grade 9	1 - 14	15 - 16	17 - 23	24 - 35
Math Grade 10	1 - 15	16 - 18	19 - 25	26 - 35
ELA Grade 9	1 - 11	12 - 15	16 - 23	24 - 35
ELA Grade 10	1 - 12	13 - 17	18 - 25	26 - 35

% of Students at Each Performance Level Based on Recommended Cut Scores - Spring 2023*				
	Below Basic	Basic	Proficient	Advanced
Math Grade 9	25.0%	35.4%	29.1%	10.4%
Math Grade 10	37.0%	28.9%	23.8%	10.3%
ELA Grade 9	17.3%	37.4%	34.6%	10.6%
ELA Grade 10	23.2%	35.1%	30.3%	11.4%

* Preliminary data - final data file to arrive from vendor on May 27, 2023.

From: [Underly, Jill K. DPI](#)
To: [Johnson, John W. DPI](#); [Beekman, Demetri A. DPI](#); [McCarthy, Thomas G. DPI](#)
Subject: RE: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval
Date: Tuesday, June 13, 2023 10:58:00 AM

Thanks, and just to clarify, I think Viji does a great job. I just wish I could understand it better as parent and a professional. That's my own shortcoming.

From: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>
Sent: Tuesday, June 13, 2023 10:57 AM
To: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Subject: RE: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

This is actually coming to us in a B week. In early July

From: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>
Sent: Tuesday, June 13, 2023 10:56 AM
To: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Subject: RE: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

I need help with this. I obviously trust Viji, but I truly don't understand what I am looking at for my approval. And with all this other nonsense going on with literacy I want to make sure we're not throwing more fuel onto this fire.

The crummy thing is, I am an educator and I don't understand it – so how are parents supposed to understand this too?

If we could set the standards and the cut scores, but then have some kind of way to interpret it to parents and educators as a companion, that would be great. For example, what does Proficient mean vs. Advanced? That they are at grade level vs. the next grade level? I just hate this stuff so much.

From: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>
Sent: Thursday, June 1, 2023 1:50 PM
To: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Subject: FW: WSAS Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

Please read the attached documents for our assessment folks. We should discuss this for a few minutes at Exec next week.

From: [Underly, Jill K. DPI](#)
To: [McCarthy, Thomas G. DPI](#)
Subject: RE: July 6th Exec meeting - Wisconsin Student Assessment System (WSAS) Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval
Date: Thursday, July 6, 2023 10:08:00 AM

I'm going to need a primer on this, or a tutoring session. I still don't understand it, and it's just a learning block that I have on my end.

From: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Sent: Thursday, July 6, 2023 9:22 AM
To: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>; Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Cc: Jones, Keona S. DPI <Keona.Jones@dpi.wi.gov>
Subject: RE: July 6th Exec meeting - Wisconsin Student Assessment System (WSAS) Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

[9:21 AM] Johnson, John W. DPI

The only thought I have is 10th grade lower the prof/basic cut and the basic/below basic cut by one point. I agree to use Plan B for this year and see alignment with Forward next year.

[9:21 AM] Johnson, John W. DPI

10th math

From: [Somasundaram, Visalakshi DPI](#)
To: [McCarthy, Thomas G. DPI](#)
Cc: [Jones, Keona S. DPI](#)
Subject: RE: July 6th Exec meeting - Wisconsin Student Assessment System (WSAS) Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval
Date: Friday, July 7, 2023 12:04:13 PM

Awesome, thank you!

Have a wonderful weekend,
Viji.

From: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Sent: Friday, July 7, 2023 12:01 PM
To: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>
Subject: RE: July 6th Exec meeting - Wisconsin Student Assessment System (WSAS) Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

Thanks for the reminder. Attached!

-tom

From: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>
Sent: Friday, July 7, 2023 11:45 AM
To: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Subject: RE: July 6th Exec meeting - Wisconsin Student Assessment System (WSAS) Standard Setting Plan and 2023 PreACT Secure Cut Scores for Approval

Good morning Tom,

I hope you are having a great Friday. I am writing regarding the PreACT approval with Dr. Underly's signature.

Thank you,
Viji.

From: [Swetz, Abigail V. DPI](#)
To: [Thiel, Christopher J](#)
Subject: cut scores links
Date: Wednesday, November 8, 2023 9:45:00 AM
Attachments: [image001.png](#)

<https://dpi.wi.gov/news/releases/2023/student-assessment-results-forward-exam>

<https://nces.ed.gov/nationsreportcard/subject/publications/studies/pdf/2021036.pdf>



Abigail Swetz (she/her)
Communications Director
Wisconsin Department of Public Instruction
608-224-6164 | dpi.wi.gov

From: [Underly, Jill K. DPI](#)
To: [Chheda, Sachin K. DPI](#); [McCarthy, Thomas G. DPI](#); [Johnson, John W. DPI](#); [Beekman, Demetri A. DPI](#)
Subject: CCSSO Assessment Discussion
Date: Tuesday, December 19, 2023 3:41:06 PM
Attachments: Outlook-Wisconsin .png

At the board meeting a bit ago, we had a couple of prompts about assessment and whether CCSSO should help pursue an Assessment Waiver from US DE.

Not sure if we want to find time to talk about this?

Here were the prompts:

1. What is your vision for assessment in the next 3-5 years.
2. What role do you want to the federal government to play or not play?
3. Given all the experts and organizations doing work in the assessment space, what is CCSSO's unique value add? And what do state leaders need?

My feedback is that if the federal government is going to mandate assessments they need to invest more in public education so that we can truly raise the lowest achieving schools that are in those spaces because of things outside of the role of public educators - mainly civil rights, segregation by poverty, and disproportionality of discipline, identification of special education students, etc - based on race and poverty and ability.

I also said that NAEP only tests 10,000 nationally, and that's really an awful way to make generalizations about how kids are doing, and that when we use state assessments to compare kids state by state, it's not an apples to apples comparison - so if the feds want to get in the assessment game, they should really make it so that we all take the same assessment and have the same cut scores. Because otherwise it's not very useful information. We get better information from our local assessments which cost a lot less.

We should probably come up with an assessment vision - so add it to the list. But my feelings on high stakes tests are pretty well known.



Dr. Jill Underly | State Superintendent
Wisconsin Department of Public Instruction
125 South Webster Street | Madison, WI
608-266-8687 | dpi.wi.gov

From: [Ruth Conniff](#)
To: [Swetz, Abigail V. DPI](#)
Subject: Re: One more question...
Date: Wednesday, January 10, 2024 1:14:48 PM

You don't often get email from ruth@wisconsinexaminer.com. [Learn why this is important](#)

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

Thanks!

Is there anything else you feel I should include? We didn't talk a lot about assessments.

On Jan 10, 2024, at 1:00 PM, Swetz, Abigail V. DPI <Abigail.Swetz@dpi.wi.gov> wrote:

Good afternoon, Ruth.

Thanks for the question. From a technical standpoint, it bears understanding that assessment scores are just a part of what goes into determining report card scores, and that for districts with high percentages of economically disadvantaged students, growth is weighted more significantly than achievement, as is [required by state law](#). When it comes to proficiency cut scores, I would point you in the direction of Dr. Underly's statement in the [release we did on assessment](#) results in October. Dr. Underly would also like to add this: "Our public education system should be about getting *every* kid what they need – in the way they need it – in order to achieve success."

Also, just wanted to make sure you knew that Dr. Underly will be delivering a speech at the State Education Conference next week, and I know she plans to address measurements of success in her address.

<image001.png> **Abigail Swetz** (she/her)
Communications Director
Wisconsin Department of Public Instruction
608-224-6164 | dpi.wi.gov

From: [Underly, Jill K. DPI](#)
To: [Johnson, John W. DPI](#); [McCarthy, Thomas G. DPI](#); [Beekman, Demetri A. DPI](#); [Chheda, Sachin K. DPI](#)
Date: Tuesday, February 20, 2024 9:03:26 AM

Before I jumped off- yes, I'd like to have a conversation about our cut scores being the highest in the country and how we communicate what that translates to with NAEP

From: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>

Sent: Thursday, March 7, 2024 10:19 AM

To: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>

Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>

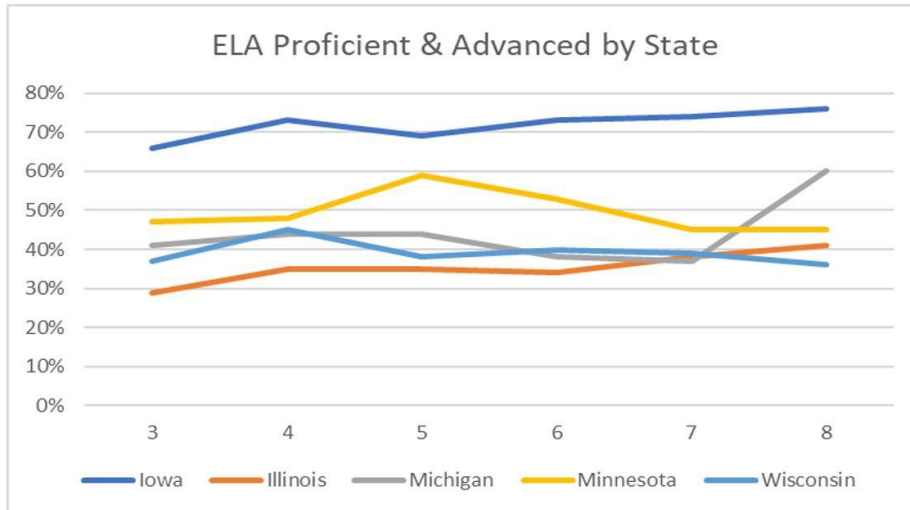
Subject: Re: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

This is a change I want. I have always hated the wording of basic, etc.

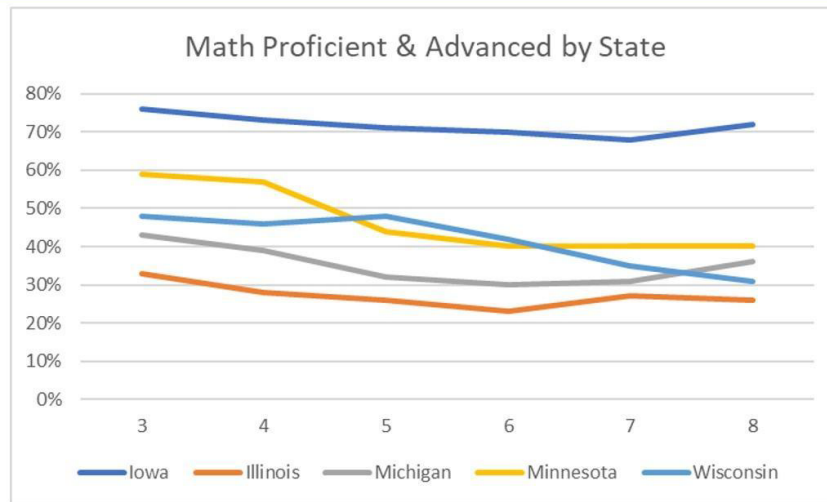
I also want to, and maybe this exists as a possibility, to norm our levels to be similar to other states... as I hear that we have some of the highest cut scores nationally. I would like them to be looked at so that we aren't judged negatively when we have direct standards. That's a digression.

I'd like to get this release and change in place for next school year. But if it must wait a year maybe that gives us more time to plan the release.

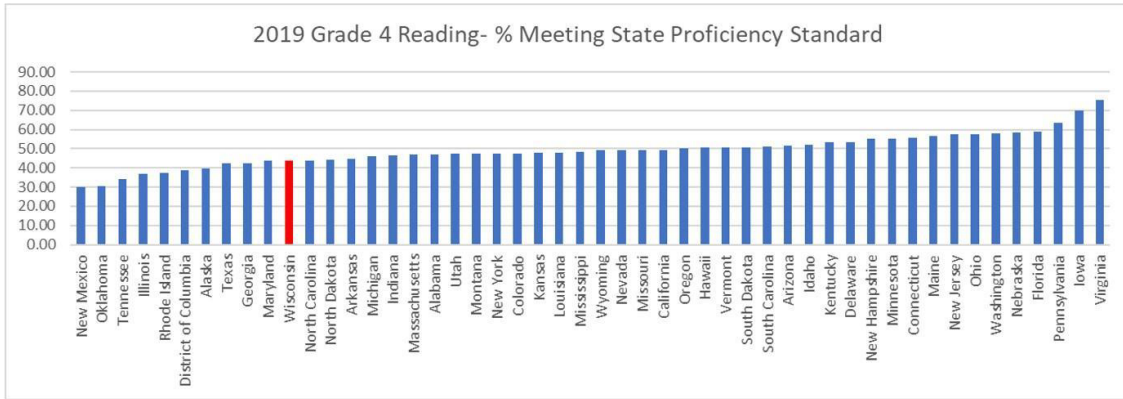
Upper Midwest State Proficiency Rates



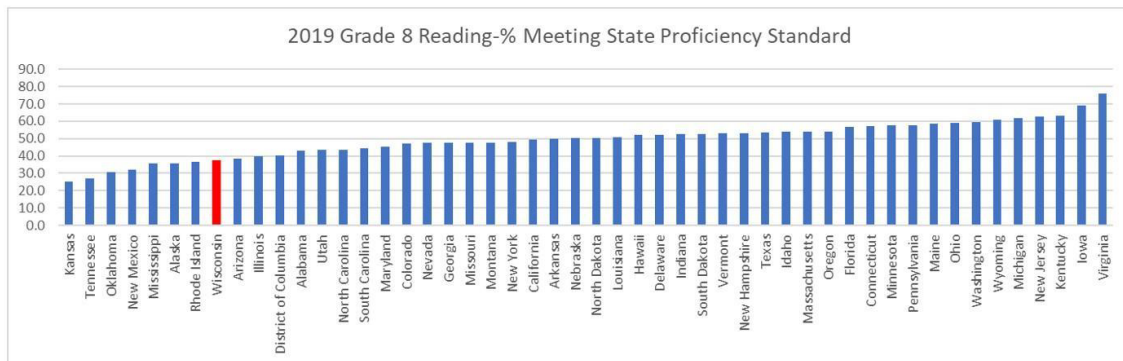
Upper Midwest State Proficiency Rates



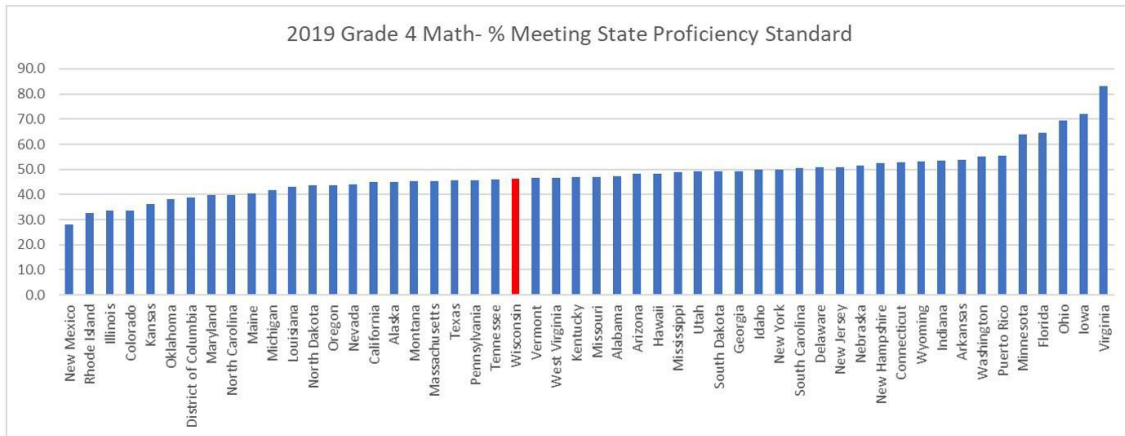
Grade 4 ELA/Reading Proficiency State Comparisons



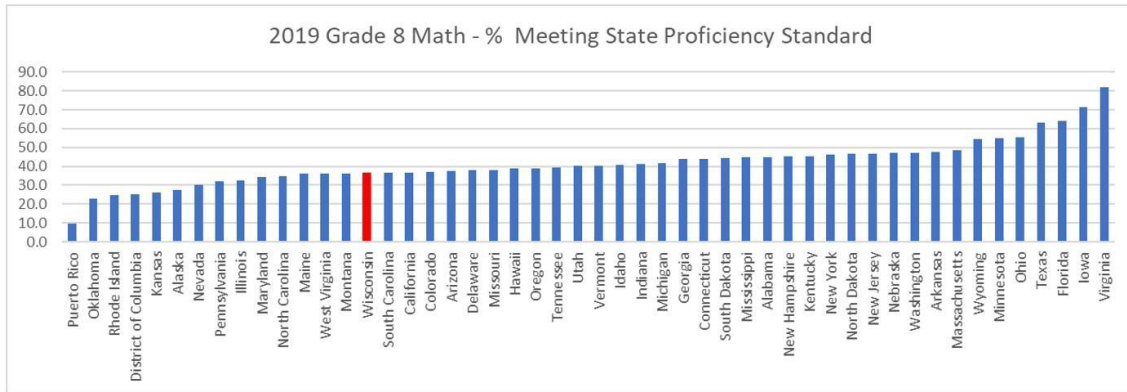
Grade 8 ELA/Reading Proficiency - State Comparisons



Grade 4 Math Proficiency - State Comparisons



Grade 8 Math Proficiency - State Comparisons



Standard Setting Discussion

Does Cabinet have a lower or upper threshold for recommendations?

- Forward
- ACT

DPI seeks to establish cut scores for the assessments which: (a) reflect the updated state content standards, (b) link students' scores on the tests to the state's expectations for students in each performance level, and (c) are well articulated across grades. For each assessment, three cut scores will be established to define four performance levels: *Below Basic*, *Basic*, *Proficient*, and *Advanced*.

Four Standard Setting Groups

Participants will work in six groups, each group consisting of approximately 16 Wisconsin educators. Each group will focus on two tests, as shown here:

- Grades 3–4 ELA, approximately 16 participants
- Grades 5–6 ELA, approximately 16 participants
- Grades 6–8 ELA, approximately 16 participants
- Grades 3–4 mathematics, approximately 16 participants
- Grades 5–6 mathematics, approximately 16 participants
- Grades 6–8 mathematics, approximately 16 participants

Accordingly, the committee will consist of approximately 96 participants. In each group, participants will be divided into four *tables* of approximately four participants each for small-group discussions.

Participant Recruitment

DRC will invite participants to register for the standard setting using DRC's online meeting management system. To do so, DRC will use a list of potential participants provided by DPI.

DRC suggests implementing these minimum requirements for participants:

- Possess a four-year college degree
- Have knowledge of the tested content
- Hold an affiliation with a Wisconsin school
- Be able to implement the standard setting process as planned
- Have knowledge of the tested population and instructional environment

The majority (if not all) of participants should be Wisconsin teachers with current (or very recent) experience in the classroom. Educators with content expertise who help classroom teachers (e.g., content coaches) are also recommended. Should DPI choose to include other stakeholders (e.g., school-level administrators, district-level administrations, testing coordinators), these stakeholders should meet the minimum requirements for participants.

The standard setting committee should reflect the diversity of Wisconsin. As feasible, participants should have the same balance of demographic characteristics as the state in terms of:

- Gender
- Race and ethnicity
- Region (e.g., education districts)

Workshop Schedule

Three groups of Wisconsin educators will convene to consider the performance standards for the ELA and mathematics tests. Table 1 provides a high-level daily agenda of the workshop.

Table 1. High-level schedule for the 2024 Wisconsin ELA & math standard setting

Day	Time	Activity
Day 1	AM	Orientation, discuss performance level descriptors (PLDs) for grade 3/5/7
	PM	Study OIB for grade 3/5/7
Day 2	AM	Complete OIB, Round 1 for grade 3/5/7
	PM	Rounds 2 and 3 for grade 3/5/7
Day 3	AM	Discuss performance level descriptors (PLDs) for grade 4/6/8
	PM	Study OIB for grade 4/6/8
Day 4	AM	Complete OIB, Rounds 1 and 2 for grade 4/6/8
	PM	Round 3 for grade 4/6/8, review recommendations, articulation discussion

Workshop Staff

DRC will administer the training for the workshop, provide general facilitation, and consult with DPI throughout the workshop. Each group will be overseen by a facilitator from DRC Psychometric Services. Facilitators' main roles are to guide participants through the standard setting activities as planned and to present workshop data. Facilitators will be supported by content experts from DRC Test Development

who will answer questions about the items and tested content. These staff will be supported by data analysts from DRC Psychometric Services who will record participants' judgments.

The workshop staff works to help standard setting participants engage with the process and make well-informed, defensible judgments. Workshop staff members are not standard setting participants and do not contribute to the committees' standard setting recommendations.

Performance Level Descriptors (PLDs)

Performance level descriptors (PLDs) are a key input into the standard setting process. PLDs summarize the knowledge, skills, and abilities expected of students in each performance level. Egan et al. (2012) suggest a framework of four types of PLDs, described here.

- 1) *Policy PLDs* summarize the state's definition for each performance level, providing information to stakeholders on the state's suggested interpretation of each level. They are typically not specific to any given grade or content area.
- 2) *Range PLDs* summarize the knowledge, skills, and abilities expected of students in a given performance level on a specific test. The range PLDs show the types of content, as informed by the state content standards, that should be mastered by students in each performance level on the test at hand.
- 3) *Threshold PLDs* are based on the range PLDs and summarize the knowledge, skills, and abilities expected of students who are at the point-of-entry (the *threshold*) of each performance level. For any given test, these descriptors show the types of skills needed just to be classified in a given performance level (e.g., just to be classified as *Proficient*).
- 4) *Reporting PLDs* are the version of the PLDs used for score reporting. Typically, a version of the policy or range PLDs are used, and the language in the reporting PLDs is adjusted to be accessible to a wide audience that may not have in-depth content knowledge. (Note: Reporting PLDs are not part of the scope of this workshop.)

PLDs Developed Prior to the Standard Setting

Policy and range PLDs will be developed prior to the standard setting. These PLDs will be presented to standard setting participants during the workshop. At the workshop, participants will use these PLDs to develop threshold descriptors.

Benchmarks and Impact Data

Most standard settings involve elements of content- and policy-based information. At the standard setting, Wisconsin educators will spend most of their time considering content-based information: they

will study the PLDs, examine test items, and consider the content-based expectations for students.

DPI may also wish participants to consider some policy-based information, such as *impact data*, the percentage of students who would be classified in each performance level if the recommended cut scores were applied. To give participants context, DPI may also wish to provide *benchmarks*.

Benchmarks refer to any policy-based information that is presented to participants that help participants make their cut-score recommendations. The use of benchmarks at standard setting is well established (Phillips, 2012; McClarty et al., 2013), especially in the Bookmark Procedure (Lewis et al., 2012; Ferrara et al., 2021). Thoughtful use of benchmarks can bring policy- and content-based information together in a meaningful way.

Potential Benchmarks

Potential sources of benchmarks include:

- Previous years' performance on the Wisconsin Forward Exam
- State performance on NAEP in grades 4 and 8 2022
- State and national performance of the college readiness benchmark on ACT

When benchmarks are shown to participants, DRC recommends the messages below be communicated to participants.

- 1) The cut scores should be tied to the state content standards and PLDs. The content-based expectations for students in each performance level should drive participants' judgments.
- 2) Impact data represent a snapshot of performance, and that performance may be different than past (or future) years.

Benchmarks may be presented to participants in several forms, including as pages in the ordered item booklet (*OIB Benchmarks*), as impact data, or both.

Potential Use of OIB Benchmarks

The proposed standard setting process is primarily based on content. To help participants stay focused on content, DPI may wish to present one set of benchmarks (e.g., historical test results) in terms of *ordered item booklet* page number (often described during the standard setting as an *OIB benchmark*).

For example, DPI may have an expectation that the percentages of students classified in each performance level will be similar to those observed in 2023. Should this be the case, participants may be shown benchmarks in the ordered item booklet (OIB) after Round 1. To do so, participants given an OIB page position (e.g., a “benchmark on page 19” of the OIB), and participants will be told that if they place their bookmarks on those pages, their recommended cut scores will be consistent with the existing cut scores. However, participants will also be told that the content should be their guide: if there is a content-based reason to do so, they may place their bookmark on any appropriate page in the OIB.

To calculate these benchmarks, DRC will find the percentages of students classified in each performance level on the tests in 2023, and then calculate cut scores on the 2024 tests which yield similar

percentages. The benchmarked OIB pages for the tests will be associated with those calculated cut scores. Participants will be asked to consider the benchmarks as they make their cut score recommendations. Participants will consider the knowledge and skills that students would need to demonstrate on the assessment to meet the cut score associated with the benchmarked cut scores.

Potential Use of Impact Data Benchmarks

When impact data are shown, participants may also be shown *benchmark impact data* to help put the impact data in context. For example, when examining 2024 impact data based on their Round 2 bookmarks, participants may also be shown 2023 impact data for context.

Construction of the Ordered Item Booklets (OIBs)

A separate OIB will be prepared for each test. Items ascend in terms of difficulty throughout the OIB. An OIB is comprised of the test items which are ordered in terms of difficulty. Item difficulty, as quantified by scale location given a response probability (RP) value, is based on data from Wisconsin students.

Participants use the OIB to recommend cut scores. Accordingly, it is important that the items included in the OIB span the difficulty continuum—from easy to hard—and that items are found around the points on the test scale where cut scores are likely to appear. The items in the OIB should also reflect the test blueprint, mirroring the range of skills measured by the test.

Response Probability (RP) Criterion

To order the items, an RP criterion must be selected. For example, with a guess-adjusted RP criterion of 0.67, the location for each item is defined as the IRT scale value associated with a 0.67 chance of answering the item correctly after guessing is factored out. This criterion, often abbreviated as RP67GA, is associated with the Bookmark Procedure (e.g., Lewis et al., 1996). However, the choice of RP criterion is a policy decision, and other states have selected different RP criteria for various purposes.

At previous standard settings for the Wisconsin Forward Exam, an RP criterion of RP50 has been used. Specifically, RP50 was used at the 2022 standard setting for social studies, and it was used at the 2019 standard setting for science (Data Recognition Corporation, 2022). To promote consistency with these previous standard settings, DPI may choose to continue using RP50 at the 2024 standard setting for ELA and mathematics. As an alternative, DPI may wish to base its decision on data: if so, DRC will create

several hypothetical OIBs using Wisconsin’s test data, each using different RP criteria, and let DPI consider the implications of each.

Workshop Procedure

On the first morning of the workshop, participants from all six groups will convene in a single session. Participants will be welcomed by DPI and DRC. The workshop will then begin with an orientation.

Orientation. DPI will begin the workshop by welcoming participants and providing a brief overview of why the standard setting is being held. This important session sets the tone for the workshop, and DRC recommends DPI use the time to describe the types of outcomes expected from the standard setting.

Workshop training. During the training session, DRC will provide an overview of the purpose of the standard setting and will describe the implementation of the standard setting methodology. The participants will use “training versions” of the following materials.

The **ordered item booklet (OIB)** is comprised of the test items, which are ordered in terms of difficulty. The ordering is straightforward in that easier items are placed earlier in the book and harder items follow.

The **item map** summarizes the materials in the OIB. The item map indicates the order of difficulty, scale score location, item identification number, scoring key, and content standard that each item measures.

DRC will also describe the roles and responsibilities of facilitators, table leaders, and participants to the committee. Participants will be reminded that it is important they keep confidential all the test items, student data, and cut score recommendations they see during the workshop.

Convene in groups. Participants will then be dismissed into their pre-assigned groups and tables. At their tables, DRC will provide a laptop computer for each participant to use during the workshop. On an access-restricted, purpose-built website (termed *the Hub*), participants will be given access to materials they will use during the workshop (e.g., item maps, evaluations).

In the breakout rooms, facilitators from DRC will welcome participants and will show participants how to use the Hub to access workshop materials.

Begin with grades 3, 5, and 7. Participants will then begin the Bookmark Procedure for either grade 3, 5, or 7. (Later, they will repeat the process for grade 4, 6, or 8.)

Review the content standards and range PLDs. Participants will briefly review the content standards and range PLDs for their first test. All participants will be asked to consider the knowledge, skills, and abilities detailed in the state content standards and how they are reflected in the range PLDs.

Discuss the threshold students. Participants will discuss their expectations for the three *threshold students*. A threshold student is a theoretical student whose level of knowledge, skills, and abilities is at the point-of-entry for a given performance level. There are three threshold students for each test, one

per cut score.

At their tables, participants will use the range PLDs to discuss the content-based expectations for each student. Then working as a table, they will write bulleted lists of these expectations on large pieces of paper to post on the wall of the breakout room. After each table has developed its own bulleted list, representatives from each table will report-out their lists. DRC will instruct participants to listen carefully as each table presents its lists. As needed, participants will update their lists, and these lists will remain on the walls of the breakout room throughout the process. By engaging in this discussion, participants will gain a richer, shared understanding of the content-based expectations for each threshold student.

Examine the student test. DRC will provide an opportunity for participants to review the test items as students may have experienced them. By reviewing the test, participants will understand how the test is structured and how students might view the test items on test day.

To review the student test, participants will be given the online practice form of the test for their assigned grade. During this process, participants will be encouraged to pay special attention to how items are structured and how the various item types are presented to students.

Study the ordered item booklet (OIB). Participants will study the items in the OIB and take notes on their item maps. By studying the items in the OIB, participants will gain an understanding of the knowledge, skills, and abilities that are measured by the test.

The OIB will be presented online using DRC INSIGHT. Each participant will be given a single-use test ticket to access the test items. To record their thoughts about the items, participants will take notes on electronic item maps using Google Sheets. Each participant will have their own item map, and participants will be encouraged to record the skills that students need to answer each item correctly.

Refresher training. Participants will then reconvene in a refresher training session. During this presentation, DRC will remind participants how to consider their bookmark placement and how bookmarks represent cut scores. Participants will be asked to keep the threshold students in mind as they place their bookmarks.

At the end of this training session, participants will be given a mid-process evaluation to gauge their understanding of the process. Participants are also asked to signal whether they are ready to proceed with the process; or, if they are not ready, by writing down any questions they may have. These questions can then be addressed with the entire group before participants continue with the process.

Round 1. Individually, participants will consider the knowledge, skills, and abilities measured by the items in the OIB, and will compare these against the content standards and PLDs. Using their understanding of the threshold students, participants will place their Round 1 bookmarks for all three cut scores without discussion.

Participants will be reminded that their primary task is to make bookmark placements in the OIB that are consistent with the PLDs, with the tested content, and with their expectations for students. To place their bookmarks, participants will use this process.

- 1) Study the PLDs to understand the expectations for students in each performance level.
- 2) Consider the knowledge and skills of students at the threshold of each performance level.
- 3) Identify the set of items (toward the beginning of the OIB) that measure skills the just *Proficient* student would definitely have mastery of, and then identify the set of items (toward the back of the OIB) that measure skills the just *Proficient* student would not have mastery of. The items between these sets represent the *Proficient bookmark range*.
- 4) Start with a hypothetical bookmark within the *Proficient* bookmark range. Determine whether there is good correspondence between the content represented by the items before their bookmark and the expectations for the threshold student.
- 5) If there is good correspondence, keep that bookmark placement and move on to the *Advanced* cut score, followed by *Basic*.
- 6) If there is not good correspondence, move the bookmark forward or backward in the OIB (one page at a time) until there is good correspondence.
- 7) Repeat the process for *Advanced* and for *Basic*.

Participants will be asked to develop content-based rationales for each of their bookmarks. These rationales explicitly link the content measured by the items before each bookmark with participants' threshold student conceptualizations. Participants will record their bookmark placements electronically.

Review Round 1 recommendations. DRC will then tabulate participants' Round 1 bookmark placements and will calculate each group's median cut score recommendations. The group's cut score recommendation is associated with the median bookmark placements.

If desired by DPI, participants will also be shown OIB benchmarks that are based on the 2023 test results. Participants will be told that if they place their bookmarks on (or near) the OIB benchmarks, their recommendations will be consistent with performance of students in 2023.

Participants will then subdivide into their tables. In their tables, participants will share their Round 1 bookmark placements with their colleagues. One at a time, in a discussion led by the table leader, participants will share their bookmark placement along with the content-based rationale behind their judgment. Participants will be encouraged to refer to the OIB, item map, content standards, PLDs, benchmarks (if used), and threshold student descriptions throughout this discussion.

Round 2. Following the discussion, participants will again individually consider their bookmark placements. Participants are free to either keep their bookmark placements from Round 1 or change their bookmark placements. All participants will make their bookmark placements individually.

Review Round 2 recommendations. DRC will then tabulate participants' Round 2 bookmark placements, calculate each group's cut score recommendations, and calculate *impact data* to present to the committee. Impact data are the percentages of students who would be classified in each performance level given a set of cut scores. The impact data will be based on the Round 2 median recommendations.

DRC will describe to the group how the cut scores and impact data were calculated. It is recommended that DPI listen to this presentation and subsequent conversation with participants, as participants may

have important policy-based questions. If used, DRC will present the benchmarks (e.g., prior-year test performance) alongside the impact data to help participants put these data in context.¹ DRC will caution participants against over-relying on the impact data or benchmarks—participants will be reminded that content should be their primary guide—but will be told that the impact data and benchmarks represent additional information for the committee.

Round 3. DRC will then lead each group in a discussion of the Round 2 recommendations. In this group-wide discussion, participants will be asked to share the rationales behind their bookmark recommendations. Whenever possible, participants will be asked to make explicit reference to workshop materials, such as the OIB and PLDs.

Participants will take part in a group-wide discussion about their Round 2 bookmark placements. This discussion allows all the participants in the group to discuss their bookmark placements together. During this discussion, participants will summarize the conversations they had at their tables before Round 2 bookmark placement, will describe their reaction to the presentation of Round 2 recommendations, and will share their content-based rationales behind their bookmarks.

DRC will then instruct participants to make their Round 3 bookmark placements individually. Participants are free to either keep their bookmark placements from Round 2 or change them.

Review Round 3 recommendations and repeat the process. After all groups finish Round 3, participants will be shown the recommendations for grades 3, 5, and 7 for their content area. Participants will be encouraged to consider the consistency (*articulation*) of the recommendations across grades. Then participants will repeat the process for grade 4, 6, or 8.

Review all recommendations across grades. After the Bookmark Procedure is complete for all grades, participants will be shown the cut score recommendations (and associated impact data) across grades. Participants will be asked to consider whether the performance standards form a reasonable, explainable pattern across grades; whether they see their own group’s recommendations as well-articulated with the other grades and whether they have any questions about the recommendations made by other groups.

If a group sees that its cut score recommendations are not well articulated with the other grades, they will be encouraged to consider how much flexibility they have around their recommendations. For example, the group’s median Round 3 bookmark may have been on OIB page 20, but the group may recognize that bookmarks between pages 18–23 are still consistent with the PLDs and tested content.

Table leaders will be asked to take notes on their participants’ opinions about the across-grade articulation, as well as any cut score flexibility. Table leaders will be asked to bring these notes to the

¹ If benchmarks are presented after *both* rounds 1 and 2, DRC recommends that they be based on the same reference data. For example, if 2023 test results are used for reference, then *OIB benchmarks* based on these data may be presented after round 1, and *impact data* based on the same data may be presented after round 2. Structured in this way, participants would be able to put their bookmark placements and impact data in context with the reference data.

articulation discussion.

Workshop evaluation. All participants will complete an evaluation of the standard setting. Results from the evaluation will be included in the standard setting technical report and can be used to document how fair and valid the participants perceived the standard setting process, and whether participants support their cut score recommendations.

Participants will then be dismissed from the workshop with the thanks of DPI and DRC. Table leaders will be asked to remain and take part in the articulation discussion.

Articulation Discussion

Vertical articulation, or across-grade consistency, is an important quality of any set of performance standards. In this sense, performance standards are consistent when the impact data form a meaningful pattern across grades. Throughout the standard setting, participants will consider the across-grade articulation of the impact data. At the end of the standard setting, the table leaders will review the across-grade articulation one last time.

Articulation process. Two separate across-grade articulation discussions will be held: one for ELA and one for mathematics. Within each discussion, the 12 table leaders will participate in a structured conversation with a facilitator from DRC.

Within each content area, the table leaders will be shown a presentation of recommended cut scores and associated impact data, just as shown to the entire committee at the end of the main standard setting. In a discussion led by DRC, the table leaders will consider these questions:

- 1) Do the performance standards form a reasonable, explainable pattern across grades?
- 2) Are there any cut score recommendations which give the committee pause?
- 3) Are there any cut scores the group recommends adjusting to promote articulation?

To answer the first question, table leaders will discuss the progression of cut scores and impact data across grades. They will be asked to describe the pattern and consider whether it is reasonable.

To answer the second question, participants will discuss any cut scores which appear particularly high or low (when compared to other grades) and discuss why the differences might exist.

To answer the third question, the table leaders will consider whether any of the cut score recommendations should be adjusted to promote better across-grade articulation. If the group chooses to recommend adjusting one or more cut scores, the committee will be charged with providing its rationale for doing so. The rationale may be content-based, or it may include a blend of content- and policy-based information. Consensus will drive the recommendations. However, if the committee cannot come to consensus, majority and minority opinions will be elicited. All recommendations will be provided to DPI for review and evaluation, including the recommendations from the main standard

Follow-up and Technical Documentation

After the standard setting, DRC will work with DPI as it reviews the committees' recommendations, considers the cut scores for the assessments. Afterwards, DRC will work to prepare the standard setting technical report.

DPI Consideration of Committee Recommendations

Within one working day following the conclusion of the standard setting, DRC will send DPI a memorandum describing the committee-recommended cut scores for the tests, including a brief narrative description of the standard setting process. This memo is designed to help DPI as it discusses the recommendations internally and with stakeholders.

Standard Setting Technical Report

The standard setting technical report is designed to provide comprehensive, detailed documentation of the process used to establish the performance standards for the assessments.

The standard setting technical report will include a narrative description of the standard setting process, detailed presentations of participants' recommendations, information about the psychometric analysis of the test items, results of the end-of-workshop evaluations, and copies of selected training materials used during the workshop.

The final report will be submitted to DPI 12–16 weeks following the standard setting. DRC will submit the report to DPI in PDF format.

Equating	Deliver Impact Data_Math_Science_Social Studies_DPI Review	3 days	Fri 5/10/24	Tue 5/14/24
Workplan	Workplan_DPI creates Draft 2	5 days	Mon 5/20/24	Fri 5/24/24

Data Files	Deliver Multiple Student Upload by District	1 day	Tue 5/21/24	Tue 5/21/24
Planning Meeting	Planning Meeting in Madison [Spring 2025]	2 days	Wed 5/22/24	Thu 5/23/24
Data Files	Deliver Unlock Report	1 day	Fri 5/24/24	Fri 5/24/24
Workplan	Workplan_DRC provides tables for Final Draft	4 days	Tue 5/28/24	Fri 5/31/24
Master Calendar	DPI reviews Master Calendar - Final	5 days	Tue 5/28/24	Fri 5/31/24
EPIC	Deliver Site & User Files [Spring Update File DPI to DRC EIS]	1 day	Fri 5/31/24	Fri 5/31/24
Workplan	Workplan_DPI creates Final with signatures	5 days	Mon 6/3/24	Fri 6/7/24
Equating	Deliver Impact Data_ELA_DPI Review	2 days	Fri 6/7/24	Mon 6/10/24
Report Interpretation Guide	Report Interpretation Guide - DPI Reviews Draft 2	5 days	Mon 6/10/24	Fri 6/14/24
Standard Setting Meetings	Conduct Standard Setting Meeting [ELA]	4 days	Tue 6/11/24	Fri 6/14/24
Standard Setting Meetings	Conduct Standard Setting Meeting [Math]	4 days	Tue 6/11/24	Fri 6/14/24
Standard Setting - ELA and Math	Deliver Cut Scores for Approval	1 day	Fri 6/14/24	Fri 6/14/24
Standard Setting - ELA and Math	DPI Approves Cut Scores	5 days	Mon 6/17/24	Fri 6/21/24
Final Student Data File [GRF, All Content Areas]	Deliver GRF_Preliminary GRF	1 day	Fri 6/21/24	Fri 6/21/24
Report Interpretation Guide	Report Interpretation Guide - DPI Approves Final	3 days	Mon 6/24/24	Wed 6/26/24
Manual Rescore Appeals Process	Manual Rescore Appeals Process_DPI reviews Form	5 days	Mon 6/24/24	Fri 6/28/24
End-of-Year Report	End-of-Year Report_DPI reviews Draft 1	20 days	Tue 6/25/24	Tue 7/23/24
Report Interpretation Guide	Report Interpretation Guide - DRC updates and makes 508 compliant	15 days	Thu 6/27/24	Thu 7/18/24
Technology Updates	June_Technology Updates / System Requirements_DRC sends to DPI	0 days	June	June
Final Student Data File [GRF, All Content Areas]	Deliver GRF_Final GRF	1 day	Fri 7/12/24	Fri 7/12/24
Technical Report - 2024	Technical Report 2024_TEST MAP APPENDICES	1 day	Fri 7/12/24	Fri 7/12/24
Quarterly Invoices	Quarter 4 Invoice_DRC sends to DPI	1 day	Fri 7/12/24	Fri 7/12/24
Final Student Data File [GRF, All Content Areas]	Deliver GRF_Final GRF_DPI Review and Approve	3 days	Mon 7/15/24	Wed 7/17/24
Data Review Meetings	Conduct Data Review Meeting [Science_2024]	1/2 day	Tue 7/16/24	Tue 7/16/24
Data Review Meetings	Conduct Data Review Meeting [Social Studies_2024]	1/2 day	Tue 7/16/24	Tue 7/16/24
Data Review Meetings	Conduct Data Review Meeting [ELA_2024]	1 day	Wed 7/17/24	Wed 7/17/24
ISR Cover Letter - School	Deliver Finished ISR Cover Letter - School_DPI Review	5 days	Wed 7/17/24	Tue 7/23/24
Data Review Meetings	Conduct Data Review Meeting [Math_2024]	1 day	Thu 7/18/24	Thu 7/18/24
Report Interpretation Guide	Report Interpretation Guide - EPM Posts Final PDF	1 day	Fri 7/19/24	Fri 7/19/24
ISRs	ISR Live data mockup_DPI Review and Approve by 2pm	2 days	Wed 7/24/24	Thu 7/25/24

Interactive Reporting	Interactive Reporting_DPI Review and Approve by 2pm	2 days	Wed 7/24/24	Thu 7/25/24
Final Student Data File [GRF, All Content Areas]	Deliver GRF_Final GRF_Access Granted to Districts	1 day	Fri 7/26/24	Fri 7/26/24
Data Files	Deliver Private School Ordering_EPM Delivers File to DPI and CESA 6	1 day	Fri 7/26/24	Fri 7/26/24
Manual Rescore Appeals Process	Manual Rescore Appeals Process_Window for Appeals	67 days	Fri 7/26/24	Tue 10/29/24
ISRs	Electronic ISRs available in DRC INSIGHT Portal_Before Noon	1 day	Fri 7/26/24	Fri 7/26/24
Interactive Reporting	Interactive Reporting_Access granted to Districts before Noon	1 day	Fri 7/26/24	Fri 7/26/24
Item Review Meetings	Conduct Item Review Meeting [Math_2024]	3 days	Tue 7/30/24	Thu 8/1/24
Item Review Meetings	Conduct Item Review Meeting [Science_2024]	2 days	Tue 7/30/24	Wed 7/31/24
Item Review Meetings	Conduct Item Review Meeting [ELA_2024]	4 days	Mon 8/5/24	Thu 8/8/24
Item Review Meetings	Conduct Item Review Meeting [Social Studies_2024]	3 days	Tue 8/6/24	Thu 8/8/24
End-of-Year Report	End-of-Year Report_DPI reviews Draft 2	10 days	Wed 8/7/24	Tue 8/20/24
ISRs	Electronic ISR_Split ISRs posted on DRC INSIGHT Portal	1 day	Mid-August	Mid-August
End-of-Year Report	End-of-Year Report_DPI reviews Draft 3_Word version with tables	5 days	Fri 8/23/24	Thu 8/29/24
ISRs	Ship REPORTS [ISRs]	1 day	Mon 8/26/24	Mon 8/26/24
ISRs	REPORTS Due in District [ISRs]	1 day	Tue 8/27/24	Tue 8/27/24
End-of-Year Report	End-of-Year Report_DRC creates Final Version	11 days	Fri 8/30/24	Mon 9/16/24
Data Files	Deliver Forensic Report	1 day	Wed 9/4/24	Wed 9/4/24
Equating [Norm Tables]	Deliver Norm Tables_DPI will release after Embargo	1 day	Wed 9/11/24	Wed 9/11/24
Technical Report - 2024	Technical Report 2024_DPI reviews Draft 1	10 days	Tue 9/24/24	Mon 10/7/24
Technical Report - 2024	Technical Report 2024_DPI reviews non-compliant Draft 2	8 days	Wed 10/30/24	Fri 11/8/24
Manual Rescore Appeals Process	Manual Rescore Appeals Process_DPI approves requests	2 days	Thu 10/31/24	Fri 11/1/24
Standard Setting - ELA and Math	DRC Delivers Standard Setting Report	1 day	October, 2024	October
Manual Rescore Appeals Process	Manual Rescore Appeals Process_DRC sends reports to districts	1 day	Mon 11/11/24	Mon 11/11/24
Technical Report - 2024	Technical Report 2024_DRC creates Final Compliant Report	11 days	Mon 11/11/24	Mon 11/25/24
Manual Rescore Appeals Process	Manual Rescore Appeals Process_DRC delivers invoices to districts	1 day	Mon 11/25/24	Mon 11/25/24

9.11 Preliminary Impact Data

After equating, DRC Research staff will provide the impact data to DPI for approval. The impact data shows the percentage of students in each performance level for each grade and content.

DRC will deliver the following scoring tables and analyses to DPI in the Technical Report:

- Scale Score to Local Percentile Tables – DPI posts these to their website annually
- Number and percentage of students at LOSS (lowest obtainable scale score) and at HOSS (highest obtainable scale score) table
- Scale Score to Standard Error of Measurement tables
- Raw score to scale score tables
- Scale score cut score ranges and Standard Performance Index (SPI) cut score ranges
- Research will calculate cut scores for each Operational SPI
- Test Integrity Analysis: Prepare and deliver a test integrity analysis to DPI, included in a Forensic Report.
- Delivery of data from research per assessment administration including classical item statistics, and Item Response Theory (IRT) statistics.

Standard Setting Methodology:

ACT uses an empirical standard setting approach to set cut scores on the ACT for states using the ACT for state and federal accountability. This approach links ACT scores directly to academic performance in college, in contrast to content-based approaches that focus on item content and difficulty. We have used this approach in several states that are currently using the ACT for federal accountability.

The primary focus of the standard setting is the probabilities of earning a grade of A, B or higher, or C or higher in a first-year credit-bearing college course. These probabilities are based on a national sample of ACT-tested students whose college grades were obtained via partnership with colleges and universities, using the same methodology that was used to create the ACT College Readiness Benchmarks. The Benchmarks reflect the ACT scores associated with a 50% chance of earning a grade of B or higher (or a 75% chance of earning a C or higher) in common first-year credit-bearing college courses. The Benchmarks were first developed in 2005 and updated in 2013, and were largely stable across the two development samples. While the Benchmarks are provided as ACT's definition of college readiness, panelists will review the probabilities across the ACT score scale when making their recommendations of the cut scores that correspond to each performance level.

Panelists review information about the probabilities of success, impact data, and other comparative data such as performance in previous years and other grade levels, performance on the National Assessment of Education Progress (NAEP), ACT scores used for scholarships or course placement by colleges in the state, and state-level college enrollment rates by ACT score. Panelists engage in multiple rounds of discussion and multiple rounds of ratings, resulting in a final set of recommended cut scores.

The rating task focuses on the state's definition of each performance level. For Round 1, panelists are asked to consider a minimally Proficient student in their subject area, and to indicate the probability (e.g., 50%, 60%) and criterion (e.g., B or higher, C or higher college course grade) that best reflects their judgment of a minimally Proficient student. Subsequent rounds of ratings are tied to the ACT scores that correspond to the probabilities of earning a grade of A, B or higher, or C or higher as well as impact data. Panelists will provide their recommended Basic and Advanced cut scores and will have the opportunity to adjust their ratings in subsequent rounds.

Panelists:

Three panels, one for each subject area (ELA, math, and science). Panelists will include:

- 8-12 panelists per subject area
- HS teachers, administrators, counselors
- Subject matter experts, as well as experience with students with disabilities, English learners
- Higher education faculty, policymakers, workforce development
- Representation from across the state (urban, rural, large and small districts)

Primary data:

- Probabilities of earning an A, B or higher, or C or higher college course grade by ACT score in each subject area (<https://www.act.org/content/dam/act/unsecured/documents/R1648-evidence-for-standard-setting-2017-05.pdf>).
- Wisconsin ACT grade 11 spring 2024 examinees. Percent at/above each ACT score in each subject area.

Supplemental data:

- Wisconsin ACT grade 11 spring 2023 examinees. Percent at/above each ACT score in each subject area. (impact from previous years)
- Aggregate performance in other ACT statewide-tested states, spring 2023 & 2024. Percent at/above each ACT score in each subject area.
- Wisconsin 2023 ACT-tested grad class. Percent at/above each ACT score in each subject area.
- Wisconsin Grade 8 performance on Forward Exam, spring 2023 & 2024: Percent of students at/above each performance level in each subject area. (impact from adjacent grades and previous years)
- NAEP data: Wisconsin and National percent of students at/above each performance level, most recent results for 8th and 12th grade in reading, math, and science.
- State and national college enrollment rates from National Student Clearinghouse by ACT score in each subject area, 2022 and 2023 ACT-tested HS graduates (2021 & 2022 if 2023 data are not available in time)
- Use of ACT scores in Wisconsin post-secondary institutions:
 - ACT scores used for course placement in Wisconsin's colleges & universities
 - Scholarships tied to ACT scores

Standard Setting Agenda

- **Day 1, morning:**
- Introductions, purpose of the meeting.
- WI DPI presents context: Why are new cut scores needed? Any other relevant information.
- ACT presents information about the empirical standard setting approach, ACT College Readiness Benchmarks and the data used to create the Benchmarks.
- Focus is on probabilities of success (e.g., 50%, 60%), criterion levels (e.g., B or higher, C or higher college course grades) and how they relate to the state's definition of Proficient.
- Group discussions (within-subject and whole room).
- Round 1 rating focuses on the Proficient Cut: What criterion and probability of success does each panelist believe represents a minimally Proficient student?
- **Day 1, afternoon:**
- Present Round 1 results.
- Present impact data and supplemental data.
- Group discussions (within-subject and whole room).
- Round 2 rating, Proficient cut.
- **Day 2, morning:**
- Recap, debrief of day 1, answer any questions, address any concerns.
- Present Round 2 results.
- Group discussions (within-subject and whole room).
- Round 3 ratings, Basic and Advanced cuts.
- **Day 2, afternoon:**
- Present Round 3 results.
- Group discussions (within-subject and whole room).
- Final ratings, all 3 cuts.
- Report final results then adjourn.
- Debrief meeting (ACT and WI DPI) to immediately follow the event.

After the Standard Setting

- Results of standard setting will be provided to WI DPI immediately following the event.
- Technical report summarizing the standard setting will be provided approximately two or three weeks after the event.
- Performance Level Descriptor (PLD) work can begin once the cut scores are approved.

Table 2. Probability of Success in English Composition I, by ACT English Score

ACT English Score	N	English Composition I Success Level		
		A	B or higher	C or higher
1	0	0.010	0.098	0.449
2	0	0.012	0.111	0.474
3	1	0.014	0.126	0.500
4	5	0.017	0.141	0.524
5	9	0.020	0.162	0.545
6	33	0.023	0.181	0.561
7	109	0.028	0.200	0.585
8	373	0.032	0.221	0.609
9	498	0.038	0.244	0.625
10	888	0.044	0.266	0.648
11	1,234	0.051	0.295	0.670
12	1,511	0.059	0.328	0.692
13	2,023	0.067	0.356	0.716
14	3,531	0.077	0.393	0.746
15	5,489	0.091	0.425	0.757
16	5,879	0.106	0.460	0.771
17	5,858	0.122	0.496	0.790
18	6,558	0.144	0.528	0.808
19	8,466	0.165	0.559	0.819
20	9,828	0.190	0.588	0.832
21	9,499	0.220	0.618	0.847
22	7,869	0.249	0.648	0.859
23	5,659	0.283	0.675	0.871
24	5,248	0.322	0.706	0.890
25	4,839	0.363	0.730	0.899
26	3,163	0.401	0.756	0.905
27	2,067	0.448	0.790	0.912
28	1,865	0.486	0.814	0.920
29	1,358	0.523	0.834	0.928
30	927	0.560	0.853	0.934
31	628	0.596	0.870	0.941
32	407	0.636	0.886	0.946
33	297	0.681	0.897	0.951
34	306	0.719	0.908	0.955
35	138	0.754	0.919	0.959
36	20	0.784	0.929	0.963

Table 3. Probability of Success in College Algebra, by ACT Mathematics Score

ACT Mathematics Score	N	College Algebra Success Level		
		A	B or higher	C or higher
1	0	0.001	0.014	0.082
2	0	0.002	0.018	0.095
3	0	0.002	0.021	0.112
4	0	0.003	0.026	0.129
5	0	0.003	0.031	0.148
6	0	0.004	0.038	0.171
7	0	0.006	0.045	0.197
8	1	0.007	0.055	0.224
9	2	0.009	0.065	0.254
10	4	0.012	0.078	0.286
11	13	0.016	0.093	0.321
12	65	0.020	0.111	0.346
13	236	0.025	0.133	0.393
14	803	0.033	0.158	0.434
15	2,095	0.042	0.186	0.460
16	4,757	0.053	0.218	0.513
17	6,044	0.067	0.262	0.559
18	5,978	0.086	0.305	0.596
19	6,870	0.106	0.352	0.629
20	6,467	0.134	0.404	0.663
21	6,474	0.161	0.459	0.698
22	5,670	0.196	0.507	0.726
23	5,853	0.235	0.552	0.755
24	5,985	0.287	0.592	0.781
25	4,232	0.340	0.639	0.802
26	3,313	0.392	0.688	0.826
27	2,204	0.451	0.732	0.848
28	1,261	0.509	0.769	0.868
29	877	0.575	0.801	0.885
30	389	0.638	0.830	0.900
31	306	0.697	0.859	0.914
32	199	0.745	0.879	0.922
33	159	0.785	0.900	0.935
34	138	0.827	0.916	0.945
35	55	0.860	0.931	0.953
36	14	0.887	0.942	0.959

New ISR Format



- Goal of new format:
 - More parent friendly
 - Additional Information
- New asset-based performance level category labels.

Wisconsin Forward Exam Student Report

Student Name:	Student Name	School:	Wisconsin School
Local Student ID:	1234567890	District:	Wisconsin District
Date of Birth:	XX/XX/XXXX	Grade:	8
		Test Date:	Spring 2024

About the Wisconsin Forward Exam
The Wisconsin Forward Exam is administered online in the spring of each school year to students in grades 3-8 in English Language Arts (ELA) and Mathematics; grades 4 and 8 in Science; and grades 4, 8, and 10 in Social Studies. The Forward Exam is aligned to the Wisconsin Academic Standards and is designed to gauge how well students are doing in relation to those standards.

How Did My Student Perform?

	Level 1	Level 2	Level 3	Level 4
English Language Arts	▲			
Mathematics	▲			
Science	No Score			
Social Studies	▲			

What Does My Student's Performance Level Mean?
There are four performance levels defined below. Levels 3 and 4 are the state targets for all students.

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
Student demonstrates minimal understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.	Student demonstrates partial understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.	Student demonstrates adequate understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.	Student demonstrates thorough understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.

How Does My Student's Score Compare with Other Students?

Content Area	Year Student's Percentile Rank
ELA	62
Mathematics	79
Science	No Score
Social Studies	92

A percentile rank compares your student's score on the Forward Exam to other student's Forward Exam scores across the state. Percentile ranks range from 1 to 99, with 99 being the highest. The rank is the percentage of students who scored the same as or below your student's score.

Page 1

From: Johnson, John W. DPI

Sent: Friday, February 16, 2024 1:26:38 PM

To: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>

Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>

Subject: FW: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

THIS IS FROM OUR ASSESSMENT FOLKS FOR REVIEW AND APPROVAL – WE HAD THEM DO STAKEHOLDER OUTREACH INCLUDING a Survey

As we have previously updated you, we surveyed our stakeholders on asset-based performance level names. I am attaching a decision paper summarizing our collaborative work across the DPI teams and our recommendation, and we are looking for the State Superintendent's approval to move forward with these new terms. We need to send these final approved names to our vendors for them to start programming for student reports and other reporting to happen this summer.

Thank you,
Viji.

From: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>
Sent: Wednesday, March 6, 2024 5:26 PM
To: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>
Subject: RE: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

Hi there,

I am elevating this up again. It has big impact on all in WI schools (including parents).

Did a decision by Exec happen on this while I was gone?

From: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>
Sent: Thursday, February 22, 2024 9:12 AM
To: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>
Subject: Re: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

Please discuss this and figure out a decision. It is a very public facing change. Keona and Viji are point in this.

From: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>

Sent: Thursday, March 7, 2024 10:09:50 AM

To: Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>

Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>

Subject: RE: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

There is one more issue on this that I am getting an answer on. Namely, would this change require us to change our Fed ESEA Plan (approved years ago with the current nomenclature)? If yes, that is a difficult process and takes an amount of time (and sometimes public comment/feedback periods).

So, at this point I think we should decide the following:

- In general, is this a change we want
- Is this a change we want for release of scores to schools and parents in the late summer and public in the fall? So, timing.
- What is the plan to REALLY get out in front of this big change and win the comms battle (I think there will be blow back)?

The ESEA Plan issue impacts timing and other aspects.

JJ

From: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>

Sent: Thursday, March 7, 2024 10:19 AM

To: Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>

Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>

Subject: Re: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

This is a change I want. I have always hated the wording of basic, etc.

I also want to, and maybe this exists as a possibility, to norm our levels to be similar to other states... as I hear that we have some of the highest cut scores nationally. I would like them to be looked at so that we aren't judged negatively when we have direct standards. That's a digression.

I'd like to get this release and change in place for next school year. But if it must wait a year maybe that gives us more time to plan the release.

From: [Johnson, John W. DPI](#)
To: [Underly, Jill K. DPI](#); [McCarthy, Thomas G. DPI](#); [Chheda, Sachin K. DPI](#); [Beekman, Demetri A. DPI](#)
Cc: [Adams, Laura L. DPI](#); [Lovings, Tacara M. DPI](#)
Subject: RE: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System
Date: Thursday, March 7, 2024 10:47:39 AM

And the cut scores decision will be in front of you this summer due to the changes to the ELA and Math assessments (change to align to their new academic standards).

From: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>
Sent: Thursday, March 7, 2024 10:42 AM
To: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>; Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>
Subject: Re: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

Absolutely... it was a total digression, but that's where I want it to eventually go.

Get [Outlook for iOS](#)

From: McCarthy, Thomas G. DPI <Thomas.McCarthy@dpi.wi.gov>
Sent: Thursday, March 7, 2024 10:30:27 AM
To: Underly, Jill K. DPI <Jill.Underly@dpi.wi.gov>; Johnson, John W. DPI <John.Johnson@dpi.wi.gov>; Chheda, Sachin K. DPI <Sachin.Chheda@dpi.wi.gov>; Beekman, Demetri A. DPI <Demetri.Beekman@dpi.wi.gov>
Cc: Adams, Laura L. DPI <Laura.Adams@dpi.wi.gov>; Lovings, Tacara M. DPI <Tacara.Lovings@dpi.wi.gov>
Subject: RE: FOR APPROVAL: Recommended Asset-Based Performance Levels for the Wisconsin Student Assessment System

To be clear...the name changing can happen quickly. The cut scores, that's gonna take more time and planning (and stakeholder engagement).

-tom

From: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>
Sent: Wednesday, April 3, 2024 4:02 PM
To: Jones, Keona S. DPI <Keona.Jones@dpi.wi.gov>
Cc: Bohrod, Samuel W. DPI <Samuel.Bohrod@dpi.wi.gov>; Lee, Alanna R. DPI <Alanna.Lee@dpi.wi.gov>
Subject: FW: Standard Setting and Cabinet

Good afternoon Keona,

Thank you for meeting with us today! Sam and I worked on the following text regarding ESSA Accountability and updated the attached briefing paper as we discussed.

Best
Viji.

From: Bohrod, Samuel W. DPI <Samuel.Bohrod@dpi.wi.gov>
Sent: Wednesday, April 3, 2024 3:56 PM
To: Somasundaram, Visalakshi DPI <Visalakshi.Somasundaram@dpi.wi.gov>
Subject: Standard Setting and Cabinet

The ESSA accountability system utilizes percentiles ranks to calculate the achievement indicator score and summary score. Therefore, the ESSA accountability system is not affected in the same manner as the report card calculations and is likely to be very minimal.

ESSA requires Wisconsin to set **long-term goals** for academic achievement in ELA and math and Wisconsin "set the ambitious goal of cutting the achievement gap in half within eight years" ([Wisconsin ESSA Plan](#), p. 18) and increasing the statewide ELA and math proficiency rates every year from 2017-18 to 2024-25 ([Wisconsin ESSA Plan](#) pp. 119-120). OEA is working on a proposed plan for the renewal of the ESSA long-term goals due to their expiration in 2024-25 and possible impacts of assessment standard setting. We look forward

to submission of renewed ESSA statewide goals to USED in early 2025.

Sam Bohrod
Assistant Director of Accountability



Memorandum

Date: 4/3/2024

To: Cabinet

From: Office of Educational Accountability

Subject: Reporting and Accountability Impacts of 2024 Assessment Standard Setting Processes

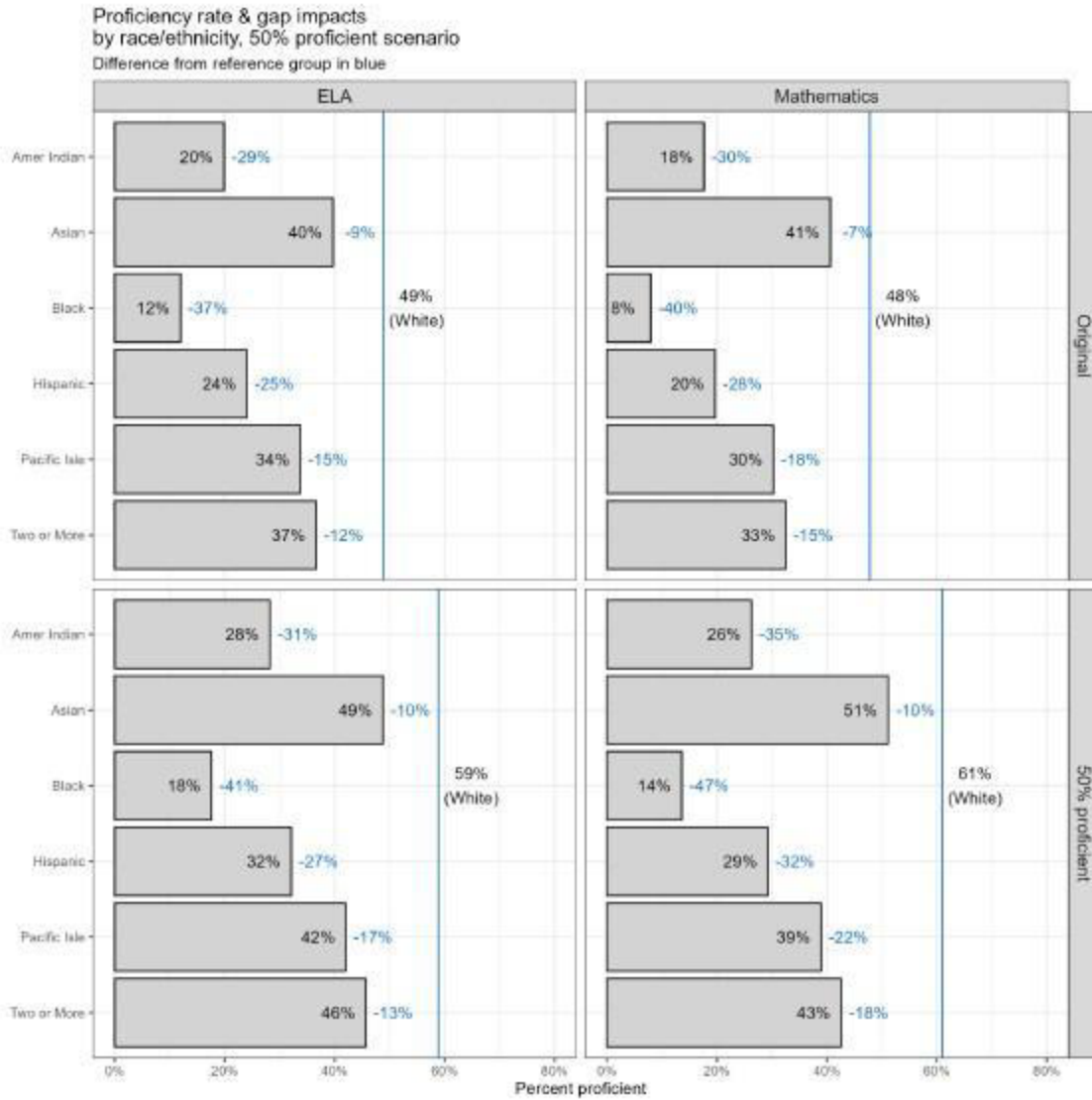
Summary

Standard setting activities for each of the statewide assessments are imminent. DPI is administering new Forward English language arts (ELA) and mathematics assessments for the 2023-24 school year. The new Forward assessments will measure updated academic standards and use a new scale. As a result, the Office of Educational Accountability (OEA) will hold a Forward standard setting to establish new cut scores delineating the four performance levels used in statewide reporting and in state and federal accountability systems. Subsequently, OEA will pursue standard setting processes for the ACT and PreACT Secure assessments. These standard setting activities will culminate in recommendations to the State Superintendent of new cut scores for each of these statewide assessments.

Assuming these standard setting activities raise proficiency rates statewide, it is reasonably likely that: (1) proficiency-based achievement gaps will increase, (2) report card Overall Scores will increase year-over-year, and (3) increases in Overall Score and Overall Rating category will impact schools disproportionately (report card rating increases will benefit low-poverty schools more than high-poverty schools).

The magnitude of the proficiency rate increase will differ across student groups. Our analysis projects that student groups who have been traditionally underserved will see smaller gains in proficiency rates as compared to more advantaged groups, widening the gaps between them. We project a particularly large difference in gains between Black and White student groups. Differing gains could widen the gap in proficiency rates between these groups somewhere between 4 and 10 percentage points. We also project the proficiency rate gap to widen by 5 to 10 percentage points between students with disabilities and those with no disability, by 3 to 8 percentage points between English learners and students who are English language proficient, and by 3 to 4 percentage points between economically disadvantaged students and students who are not economically disadvantaged.

Figure 1: Proficiency rates compared by race/ethnicity statewide, 2022-23 status quo compared to 50% proficient scenario

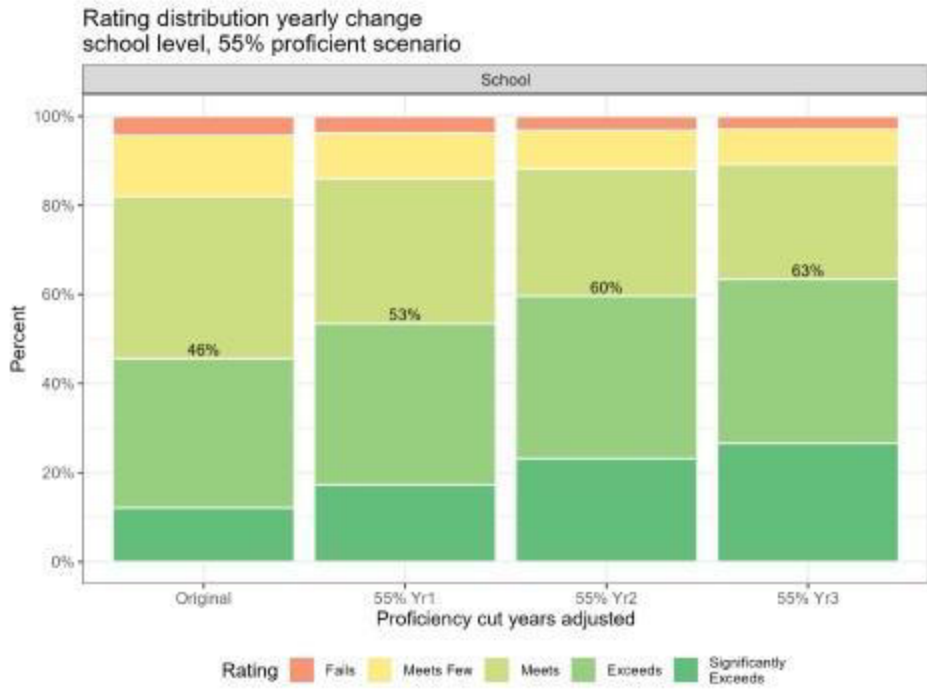


School and district report card scores and ratings will also rise if proficiency rates increase. **If proficiency rate gains are substantial, we project report card ratings will increase to such a degree that there will be little difference between schools in terms of their ratings.** Assuming substantial increases in proficiency rates, when the full effect of the assessment standard settings are felt we project that nearly two-thirds of schools will receive report card ratings in

2

the two highest categories; in 2022-23 by comparison 46% of school report cards had ratings in the two highest categories.

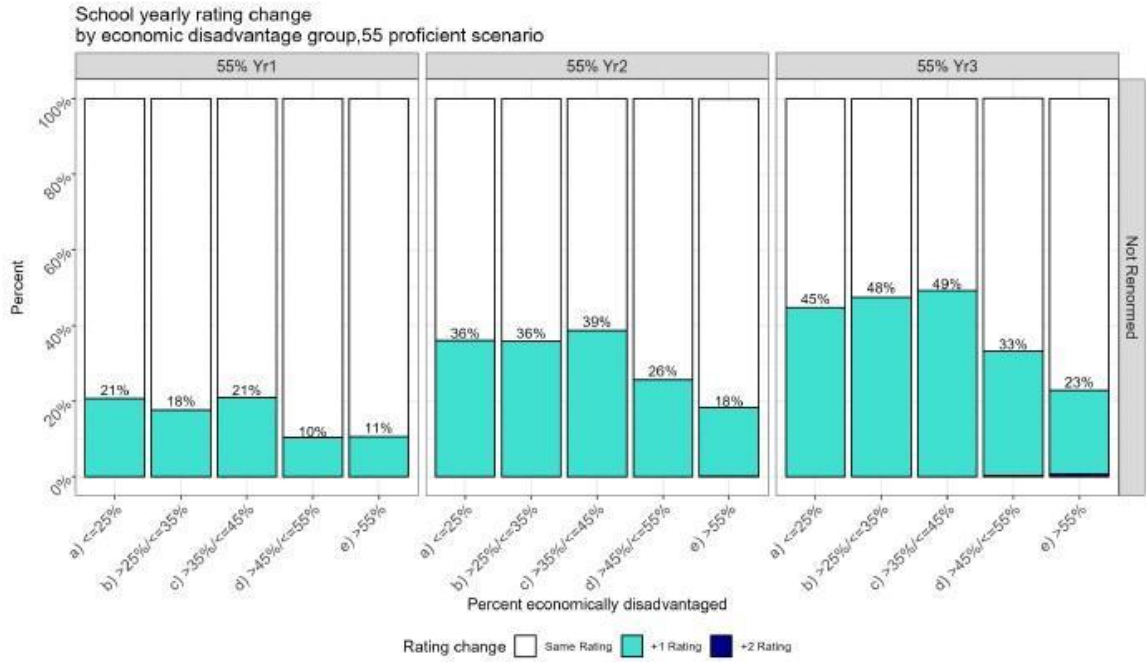
Figure 2: School report card rating increases over three years



Due to the multi-year nature of the achievement scores, report card rating impacts of new assessment cut scores would be incremental and spread across the next three years of report cards. When all years of data included in achievement scores have the new assessment cut scores applied, we project that nearly two-thirds of schools will receive report card ratings in the two highest categories (see Figure 2 above). This projection is based on the 55% proficient statewide scenario.

In addition, these report card rating increases will benefit low-poverty schools more than high-poverty schools due to variable weighting of Achievement and Growth priority areas, as required by state law. State law (Wis. Stat. 115.385(1g)(c)) requires that the relative weighting of the Growth and Achievement priority areas depends on the percentage of students at the school or district who are economically disadvantaged (ECD). Greater weight is given to Growth when the ECD percentage is high, and greater weight is given to Achievement when the ECD percentage is low. As a result, rising proficiency and PBP rates will raise report card scores and ratings to a greater degree for schools and districts with lower percentages of ECD students (see Figure 3 below).

Figure 3: Variable weighting leads to greater gains for low-poverty schools



While the impacts of the assessment standard setting on achievement gaps cannot be addressed, OEA would be able to address some of the challenges to the report card system to a certain extent. OEA's accountability team proposes to conduct a report card standard setting in summer or fall of 2025 when two years of assessment data with the new cuts can be included. The standard setting would define new rating thresholds to arrive at a distribution of report card ratings that appropriately differentiates between schools. As part of that process, we also propose rescaling Growth and Target Group Outcomes (TGO) priority area scores

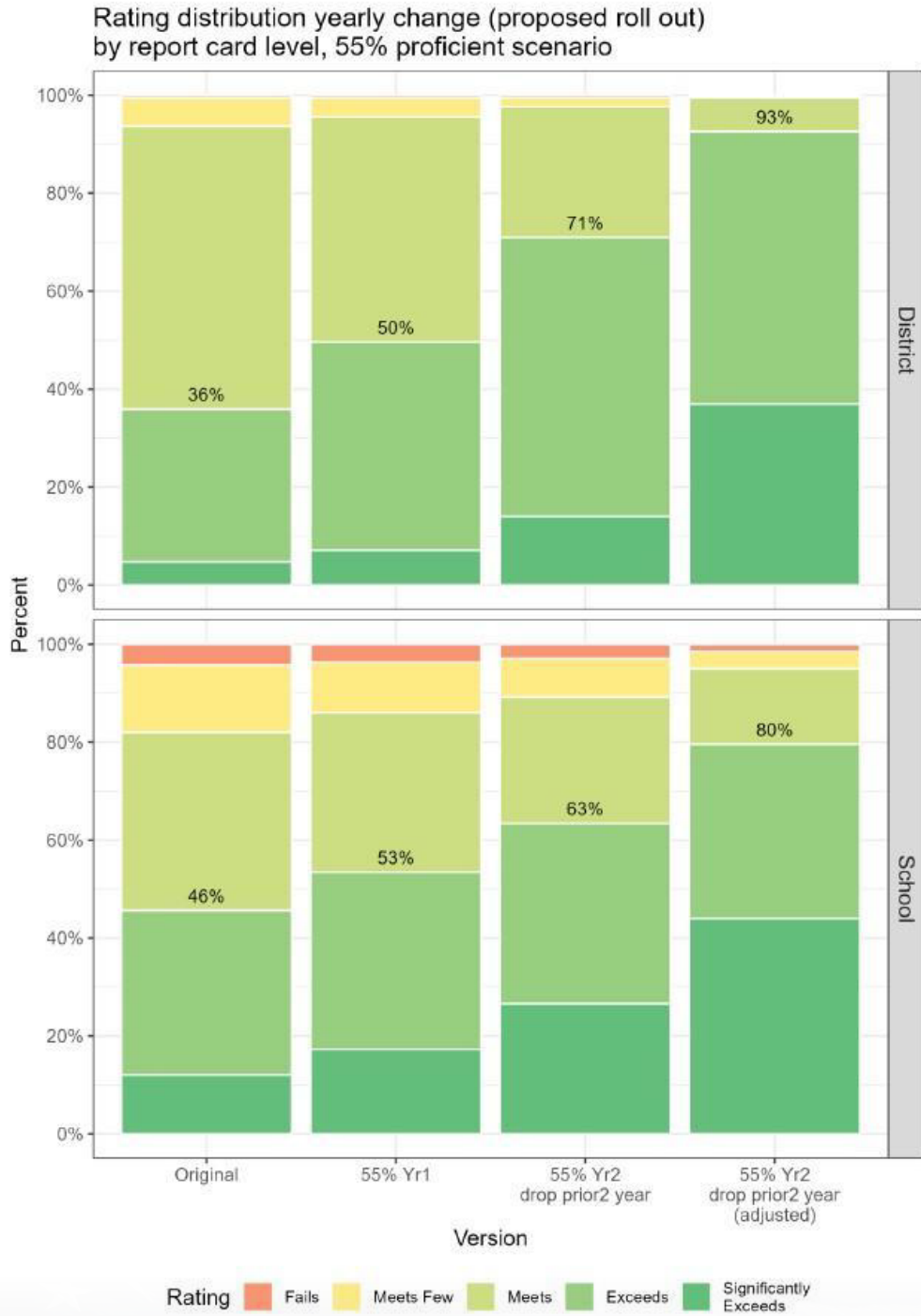
so that they align with Achievement priority area scores. This will ensure that the report card scoring system is fair to schools regardless of poverty level, as well as being fair to schools with and without sufficient data for Growth and TGO scores.

Potential next steps

1. Keep report card calculations and rating thresholds unchanged for the 2023-24 school year as we anticipate relatively moderate increases this year.
2. For the 2024-25 school year, make the following changes to report card calculations:
 - a. Include only two years of assessment data in achievement calculations, including the Achievement priority area, as well as the achievement components of the TGO and On-Track to Graduation priority areas. Achievement typically includes three years of data; under this construction it would take three years for the new assessment cuts to fully impact report card scores. Using only two years of assessment data in 2024-25 would shorten the timeline to full impact of new assessment cuts.
 - b. Adjust Growth and TGO rescaling formulas to align these priority area score distributions with the new Achievement priority area score distribution.
3. Prior to release of 2024-25 report cards, conduct a report card standard setting to establish new report card rating thresholds.

Keeping report card calculations unchanged for the 2023-24 school year gives OEA time to prepare for a report card standard setting. Assessment standard setting activities will conclude in late July of 2024. Given that the increases in points-based proficiency rates in 2023-24 will be tempered by those from the two years prior, OEA anticipates relatively moderate increases in report card scores and ratings in 2023-24. This can be seen by comparing the 'Original' and '55% Yr1' columns in Figure 4 below. The 'Original' columns are projections of report card ratings for districts (top panel) and schools (bottom panel) in 2023-24 if proficiency rates are unchanged.

Figure 4: Report card rating changes based on potential next steps 1 and 2, prior to a report card standard setting (step 3)



Standard Setting Methodology (PreACT Secure)

We will use a statistical method (linear interpolation between Forward Exam and ACT) to update the existing PreACT Secure cut scores and draft new recommendations for the State Superintendent's approval.



Reporting Impacts

Statewide proficiency rate gains:

- Gains for all student groups
- Smaller gains for traditionally underserved groups



ACT with writing - ELA

Score	2023 Admin ELA % At or Above-All Students	ELA - A Probability	ELA - B or Higher Probability	ELA - C or Higher Probability	Advanced Proficient Basic
30	3%	0.58	0.84	0.92	
29	4%	0.53	0.82	0.91	
28	6%	0.49	0.79	0.90	
27	8%	0.45	0.77	0.89	
26	11%	0.40	0.73	0.87	
25	14%	0.36	0.70	0.86	
24	17%	0.32	0.66	0.84	
23	22%	0.29	0.62	0.82	
22	27%	0.25	0.58	0.80	
21	32%	0.22	0.55	0.78	
20	37%	0.19	0.51	0.76	
19	43%	0.16	0.47	0.73	
18	48%	0.14	0.43	0.71	
17	54%	0.12	0.39	0.68	
16	60%	0.11	0.35	0.65	
15	65%	0.09	0.31	0.61	
14	70%	0.08	0.28	0.58	
13	75%	0.07	0.25	0.54	
12	78%	0.06	0.23	0.51	

- Current proficient cut score linked to 50% likelihood of obtaining a B or higher in an entry level *English composition* course in college.
- Adjusting likelihood up or down will dramatically change the percentage of students achieving proficiency.

ACT with writing - Math

Score	2023 Admin Math % At or Above-All Students	Math - A Probability	Math - B or Higher Probability	Math - C or Higher Probability	Advanced Proficient Basic
30	4%	0.64	0.83	0.90	
29	5%	0.58	0.80	0.89	
28	7%	0.51	0.77	0.87	
27	10%	0.45	0.73	0.85	
26	13%	0.39	0.69	0.83	
25	17%	0.34	0.64	0.80	
24	20%	0.29	0.59	0.78	
23	24%	0.23	0.55	0.75	
22	27%	0.20	0.51	0.73	
21	30%	0.16	0.46	0.70	
20	33%	0.13	0.40	0.66	
19	37%	0.11	0.35	0.63	
18	43%	0.09	0.30	0.60	
17	52%	0.07	0.26	0.56	
16	63%	0.05	0.22	0.51	
15	74%	0.04	0.19	0.46	
14	83%	0.03	0.16	0.43	
13	87%	0.02	0.13	0.39	
12	89%	0.02	0.11	0.35	

- Current proficient cut score linked to 50% likelihood of obtaining a B or higher in an entry level *algebra* course in college.
- Adjusting likelihood up or down will dramatically change the percentage of students achieving proficiency.

Reporting Impacts

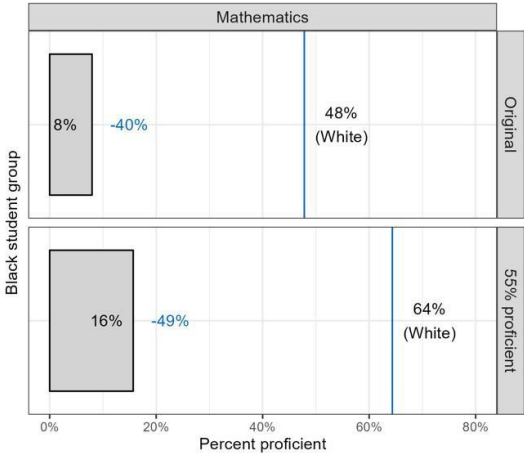
Statewide proficiency rate gains:

- Gains for all student groups
- Smaller gains for traditionally underserved groups



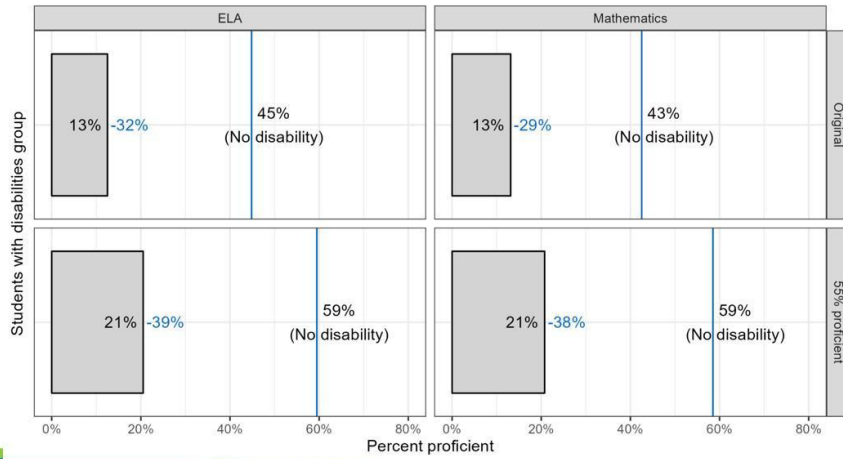
Reporting Impacts by Race/Ethnicity

Proficiency rate & gap impacts
Black student group, 55% proficient scenario
Difference from reference group in blue



Reporting Impacts by SwD

Proficiency rate & gap impacts
 Students with disabilities group, 55% proficient scenario
 Difference from reference group in blue



Accountability Impacts

- Achievement score increases
- Report card ratings increases over multiple years
- Fewer rating increases for high ECD schools
- ESSA accountability impacts minimal



From: [Johnson, John W. DPI](#)
To: [Somasundaram, Visalakshi DPI](#); [Bohrod, Samuel W. DPI](#); [Olsen, Philip DPI](#); [Jones, Keona S. DPI](#)
Subject: Guiding Questions for Wednesday Exec Conversation on Policy Implications
Date: Monday, April 15, 2024 2:41:55 PM

Guiding Questions:

What are the policy implications of these changes?

What parts of DPI are needed at the table?

What decision points are really needed at this time?

What is UEED approval process for – new cut scores and for new nomenclature?

Could we push out a decision for a year on – new cut scores – on new nomenclature?

What exact steps happen for each of these changes – process?

How are you approaching prof level setting with teachers?

What is your recommendation(s) and why?

- FOR ALL – PLEASE BE VERY CLEAR AND CONCISE.

John W. Johnson, PhD
Deputy State Superintendent
Office of the State Superintendent
Department of Public Instruction
State of Wisconsin



INSTITUTE FOR REFORMING GOVERNMENT

ReformingGovernment.org

PO Box 180291
Delafield, WI 53018

info@reforminggovernment.org

 [ReformingGovernment](https://www.facebook.com/ReformingGovernment)
 [@ReformingGovt](https://twitter.com/ReformingGovt)