



A Beacon, a Barometer, and a Bridge

Imagining More Equitable, Student-Centered
State Standards and Assessments in Minnesota



About this report

This paper builds on a prior Education Evolving report, “*Defining and Measuring Student-Centered Outcomes*” (www.educationevolving.org/outcomes), which lays out a foundational argument and framework. This paper goes a step further, with concrete recommendations for Minnesota’s system of statewide assessments and academic standards.

WITH DEEP GRATITUDE

We are immensely thankful to the more than 70 students, parents, teachers, administrators, policymakers, and researchers we spoke to for this paper. We do not list them by name because we spoke with them on conditions of anonymity, but each contributed to the final product. We extend a special note of gratitude to the team at the Minnesota Department of Education, for their openness to conversation and their generosity of time.

ABOUT THE AUTHORS

The research team for this report included Lars Esdal (of Education Evolving); Addy Welch (an independent researcher); and Laura Potter, Mengchen Su, Sara Kemper, and Alyssa Parr (of the Center for Applied Research and Educational Improvement or CAREI at the University of Minnesota). While Lars is the primary author—and assumes full responsibility for the words and recommendations within—the ideas are a joint product of the team.

ABOUT EDUCATION EVOLVING

We are a Minnesota-based nonprofit, nonpartisan organization focused on advancing equitable, student-centered learning—that is learning that honors each student’s unique assets, interests, identities, and aspirations. To that end, we support teachers designing and leading schools, and policies that catalyze community-led innovation.

Learn more at www.educationevolving.org.

Contents

| | |
|--|-----------|
| Introduction | 4 |
| Rethinking the what and how of assessment | 4 |
| State assessments and standards matter | 4 |
| Naming opportunities (and issues) with current statewide assessments | 6 |
| Background | 8 |
| Minnesota’s academic standards | 8 |
| Minnesota’s statewide assessments | 8 |
| Federal and state assessment laws | 8 |
| A complex history and relationship with equity | 9 |
| RECOMMENDATION AREA 1 | |
| State Academic Standards as a Beacon | 10 |
| <i>Imagine a More Focused Model of Learning Progress</i> | |
| 1A. Write standards and benchmarks to model “learning progressions” | 10 |
| 1B. Increase the rigor and focus of standards, while reducing overall quantity | 11 |
| 1C. Create a state “Portrait of a Graduate” | 12 |
| RECOMMENDATION AREA 2 | |
| Grade 3-8 Assessments as a Barometer | 13 |
| <i>Imagine a More Equitable, Growth-Focused “System Check”</i> | |
| 2A. Involve youth in test development and prioritize cultural relevance | 15 |
| 2B. Accelerate the release timeline (and user-friendliness) of test results | 15 |
| 2C. Prioritize measuring GROWTH in the design of the Grade 3-8 MCA | 17 |
| SIDEBAR: Five priorities for growth model design | 18 |
| RECOMMENDATION AREA 3 | |
| High School Assessments as a Bridge | 20 |
| <i>Imagine “Badges” as Causes for Celebration (and Credit and Jobs)</i> | |
| 3A. Launch a high school “badging” system | 20 |
| 3B. Expand acceptance of badges by colleges and employers | 22 |
| 3C. Ultimately seek federal waiver to use badges as the high school MCAs | 23 |
| SIDEBAR: Badging - A concept already widely in use | 24 |
| Conclusion | 25 |
| Appendix | 26 |
| <i>A Psychometric Concepts Glossary</i> | |

Introduction

Rethinking the what and how of assessment

To get the equitable, student-centered education system we want and need—one which honors each unique student, and prepares them for the changing 21st century world—we need to rethink not only the design of learning, but also the what and how of assessment.

Why? Assessments matter because they express our educational values. Our choices about which knowledge and skills we measure, and how we measure them, set the “goal posts” for learning. As goes the old adage, what gets measured gets done.

Assessments are used by people in many different roles and in many different ways (see Table 1). This paper focuses on statewide assessments, the final row in the table, and in particular the Minnesota Comprehensive Assessments (MCAs).

This paper’s purpose is to explore how statewide assessments and academic standards could better align with, and help push our state toward, more equitable and student-centered learning.

Statewide assessments are developed based on state academic standards, which are state-adopted statements about what students should know and be able to do, across subjects and grade levels. Because assessments and standards are so inextricably linked, this paper’s focus also includes standards.

This paper’s purpose is to explore how statewide assessments and academic standards could better align with, and help push our state toward, more equitable and student-centered learning.

State assessments and standards matter

The central premise of this paper is that statewide assessments and standards matter and deserve attention. We assert three main reasons why:

1. They’re required.

State assessments are required under federal law, namely the Every Student Succeeds Act (ESSA).¹ Nobody we spoke with saw signs that this will change soon (and, it’s beyond the scope of this paper to consider *federal policy*, and whether it should). Regardless of one’s feelings on statewide assessments, the reality is they’re here to stay for the foreseeable future.

2. They’re influential.

State assessments are influential, shown to be linked with things like the oversight and support schools receive;² family perceptions of quality and enrollment choices;³ and even property values.⁴ And, largely because test scores are so visible and scrutinized, they shape what is taught in schools and how.⁵

3. They do (or at least in theory could) play a helpful role.

We heard many families, educators, and policymakers voice their value as a “benchmark” or “system check”—for monitoring equity in opportunities and outcomes; tracking progress over time; comparing schools and districts; and evaluating local programs, curricula, and assessments.

This paper asks: given the reality that statewide assessments are both required and influential, how can we lean into the third reason, and make them as helpful as they can be?

This paper asks: given the reality that statewide assessments are both required and influential, how can we lean into the third reason, and make them as helpful as they can be? Even more specifically, how can they best support and push education to be more equitable and student-centered?

Our research for this paper began with asking dozens of students, educators, and parents these questions. The next section summarizes what we heard.

| ASSESSMENT TYPE | WHO USES AND HOW | EXAMPLE(S) |
|--|---|---|
| Student Self-Assessments | Students, to support and adapt their own learning. | A self-diagnostic assessment of a particular math skill |
| Classroom Formative Assessments | Educators and students, to take a pulse check for understanding and adapt learning. | A mid-unit quiz; a student presenting an interim work product (like a paper outline) |
| Screener, Intervention, and Monitoring Assessments | Educators including specialists (for example, reading specialists), to identify specific students and their areas of needed support. | FastBridge CMBreading/math (for academics) or mySAEBRS (for social and emotional learning) |
| Interim or Benchmark Assessments | Educators and administrators, to identify overall progress, particular grade levels or subjects needing attention, gaps in curricula, etc. | NWEA MAP, or FastBridge aReading/aMath |
| School or Classroom Summative Assessments | Educators and administrators, to gather evidence of learning, identify standards met, etc. | End-of-unit or end-of-course exams, presentation/defense of a completed project or capstone |
| Statewide Summative Assessments (the focus of this paper) | Educators, administrators, and policymakers, for reflection, planning, policymaking, and evaluation of programs and curricula; families, to make informed decisions and push for needed improvements. | The Minnesota Comprehensive Assessment (MCA), as well as assessments designed for English language learners and students with significant cognitive disabilities as detailed later in the Background section. |

Table 1. There are many types of assessments, used in many different ways to support students and learning. Statewide assessments—the final row in this table—are the focus of this report.

Naming opportunities (and issues) with current statewide assessments

As detailed on the next page, our methodology for this report involved formal interviews, focus groups, and unstructured conversations with students, parents, educators, and others.

Those dialogs surfaced several themes—about both opportunities, and current realities that stand in the way. In sum, they asked us to imagine statewide assessments and academic standards that:

- **Centered youth in design and development.**
While there have been good strides made in recent years by the Minnesota Department of Education (MDE) to increase youth involvement, the development of standards and assessments is still largely adult-driven and adult-centering.
- **Included *different ways of speaking, writing, and knowing.***
Many we spoke to were concerned that test content still fell short in representing the rich diversity of experiential, cultural, racial, and linguistic backgrounds of youth in our state.
- **Captured growth, especially of those who need more support.**
Current statewide assessments and growth measures effectively show “no growth” for students who are 2+ years behind, even if they make major learning gains. What’s more, students and families don’t get results until months later, at which point some don’t even remember taking it.
- **Were *more focused and learning science-aligned.***
While progress has been made in the latest revision cycle, state standards and benchmarks are still too “mile wide, inch deep,” which reduces the tightness of their focus on key elements of the science of learning and literacy.
- **Valued the *deep, applied skills necessary for true college and career readiness (especially in high school).***
Statewide assessments capture little at the “strategic thinking” level, nothing at the “extended thinking” level⁶ (Webb’s Depth of Knowledge Level 3 and 4),⁷ and miss other key skills like effective communication and creativity.⁸
- **Provided *occasions for celebration, and stepping stones to what’s next.***
In contrast, nearly all of the students we spoke to experienced statewide assessments as a substantial source of anxiety, and one with little relevance to their life and learning.

The remainder of this paper explores what it would take to seize these opportunities.

The list above is long, and does stretch our imagination. While we understand the concern that asking statewide assessments to “do too much” is unrealistic, too often this is used as a reason not to make *any* big changes.

While [these recommendations] may prove challenging and require Minnesota to stretch into being a national leader at the cutting edge of assessment, that is precisely where we need to be. Our young people deserve it.

Further, this concern often doesn't fully take into account recent field developments (for example, in learning progression theory, artificial intelligence, test designs that gauge grade level proficiency and also adapt far above/below grade level, and more).

The recommendations in this paper take this concern seriously, and still we believe all recommendations offered here are realistic. While they may prove challenging and require Minnesota to stretch into being a national leader at the cutting edge of assessment, that is precisely where we need to be. Our young people deserve it.

In the rest of this paper, we first set context—on state standards and assessments, the laws governing them, and their complex relationship with equity. Then, we offer recommendations in three areas: academic standards, grade 3 to 8 assessments, and high school assessments.

A NOTE ON OUR METHODOLOGY

The findings and recommendations in this paper draw on both conversations and desk research, in particular:

- Focus groups, interviews, and conversations with 50+ people in Minnesota who experience or work with statewide assessments—including students, teachers, administrators, parents, policymakers, and researchers;
- Interviews with dozens of the country's leading assessment experts—including state assessment directors outside of Minnesota, and others involved in research, advising, and technical assistance on statewide assessments;
- An extensive, 50-state landscape scan of innovative assessment approaches being used; and finally
- A deep literature review on relevant policy and psychometric topics uncovered in our conversations and landscape scan.

Background

Minnesota’s academic standards

Minnesota law requires statewide standards in language arts, math, science, social studies, and physical education.⁹ The content of the standards are not written in law; rather, statute requires MDE to draft and adopt standards in rule, in consultation with communities and educators.

Within each subject, standards consist of “strands” and (within strands) “anchor standards,” both of which are consistent across grades. For each anchor standard, MDE publishes one or more “benchmarks” at each grade level. Benchmarks are concrete statements of what students should know or be able to do, and are also used to develop items for statewide assessments.

Within each subject, each elementary and middle school grade level has around 40 *total* benchmarks, across all anchor standards. For high school, there is one “bucketed” set of benchmarks for each anchor standard (rather than distinct benchmarks for each grade level).

Minnesota’s statewide assessments

The Minnesota Comprehensive Assessments (MCAs) are the most widely taken statewide assessments in our state and the focus of this paper. The English language arts and math MCAs are given each year in grades 3 through 8 and once in high school; the science MCA is given once in 5th grade, 8th grade, and high school.

The MCA tests for ELA and math are each around 45 questions long and take a majority of students one to two hours to complete.¹⁰ In most cases, the test is taken on a computer. The test is “adaptive”, meaning a student will get progressively more difficult questions as they demonstrate more grasp of the material, and vice versa.¹¹

While, due to limitations on scope, this paper is focused on the MCA there are three other annual state assessments given in Minnesota.¹² The ACCESS test is given to all students identified by the state as English Language Learners; it gauges student progress in learning academic English. Finally, two assessments serve as alternatives to the MCA and the ACCESS, respectively, for students with significant cognitive disabilities: the Minnesota Test of Academic Skills (MTAS) and the Alternate ACCESS.

Federal and state assessment laws

Federal laws (currently Every Student Succeeds Act, or ESSA) require that states test students each year between grades 3 through 8, and once in high school for ELA and math; and once in elementary, middle, and high school for science.¹³ There are no testing requirements for other subject areas.

State law codifies many of these federal requirements, primarily in MN Statutes 120B.30 through 35. In general, the statute leaves details of assessment design to the Minnesota Department of Education (MDE).

A complex history and relationship with equity

We've attempted to take an equity lens throughout this paper. Still, and additionally, an explicit acknowledgment is merited on the complex history (and present-day relationship) the discipline of assessment, and state assessments in particular, have with educational justice and racial equity.

In some of their early uses, standardized assessments were employed to evaluate and rank people with explicitly racist intentions and effects.¹⁴ Even as the explicitness of this purpose faded, most assessments were (and still are) designed from an implicitly white-normed frame of reference.¹⁵ This legacy and these challenges hover over assessments still today.

And, there is another side to this complex relationship. We spoke with many people who pointed out the perpetual systemic failures of education to equitably serve all students—with unacceptable patterns along income, racial, and ethnic groups. Many articulated that having some common, comparable statewide assessment that can help to expose and challenge systemic inequities in opportunity and biases in expectations is critically important.¹⁶

Going forward, our actions must acknowledge both perspectives, and the tension among them. That is, we must explicitly disrupt inequitable patterns in assessment design, content, and use—while also maintaining a comparable benchmark on unacceptable systemic inequities that exist.

As we move next to recommendations, we keep this charge front of mind.

We must explicitly disrupt inequitable patterns in assessment design, content, and use—while also maintaining a comparable benchmark on unacceptable systemic inequities that exist.

RECOMMENDATION AREA 1

State Academic Standards as a Beacon



Imagine a More Focused Model of Learning Progress

As described above, state standards set the “blueprint” for assessments—the topics covered, the form and quantity of questions needed, etc. This first recommendation area addresses the standards themselves.

1A. Write standards and benchmarks to model “learning progressions”

CONTEXT AND RATIONALE:

- The idea of “learning progressions” has been gaining attention in both assessment and learning science communities in recent years.¹⁷
 - In essence, learning progression theory says that discrete bits of knowledge and skill build upon one another, and can be sequenced in a way that mirrors the way a person develops their overall understanding of a content area.
 - For example, in a learning progression for “place value”, students move—across grades K through 5—from counting objects, to decomposing numbers into 10s and 1s, to writing numbers to the 100ths place, to writing numbers to the 1000s place and in expanded form (i.e. $60.37 = 6 \times 10 + 3 \times 1/10 + 7 \times 1/100$).¹⁸
 - A learning progression approach is in contrast to what one might call a “laundry list approach.” That is, where longer lists of knowledge and skills are created for a given grade level and/or content area, with less attention paid to their larger sequence relative to one another, within and across grades.
- Having standards that explicitly reflect learning progression is helpful because it means everything built on those standards (including state assessments, but also curriculum created by curriculum companies and local districts) draws on the underlying learning progression-based theories about how students learn.

- This is particularly important given our state’s current (and overdue) push to embrace the science of reading. In short, literacy development is a “learning progression,” and our state standards must reflect that.

In short, literacy development is a “learning progression,” and our state standards must reflect that.



AND SO: MDE academic standards review committees should use a “learning progressions” lens as they draft revised standards and benchmarks—both within and across grades.

In practice, this means grouping benchmarks that build upon one another in strands and anchor standards, within and across grades. And second, making sure benchmarks are laid out and presented in such a way as to make clear their connections to other benchmarks with which they form a progression.

The latest draft of the ELA and math standards mark progress in this direction, through the consistent use of strands and anchor standards, which span grades. Future standards revision committees should keep up and build upon this work.

1B. Increase the rigor and focus of standards, while reducing overall quantity

CONTEXT AND RATIONALE:

- Academic standards in the United States have long been critiqued as a “mile wide and inch deep,” relative to most other industrialized nations.¹⁹
- Per Recommendation 1A above, we need to be sure that our standards and benchmarks are carefully designed to model learning progressions that build literacy and core skills.²⁰ Simultaneously, we need to let go of benchmarks that are not critical parts of a learning progression or useful in students’ future lives, but rather inherited from a “laundry list.”
- Further, many of our standards and benchmarks are low “Depth of Knowledge” (DOK), focused more on recall and rote skills²¹ rather than applied skills, which is an issue particularly for middle and high school standards.²²
- In sum, having too many laundry-list and low Depth of Knowledge standards harms the overall rigor and focus of our standards. It hinders development of the core skills that matter most (especially in early grades), and deep, applied learning (especially in later grades).

In sum, having too many laundry-list and low Depth of Knowledge standards harms the overall rigor and focus of our standards.



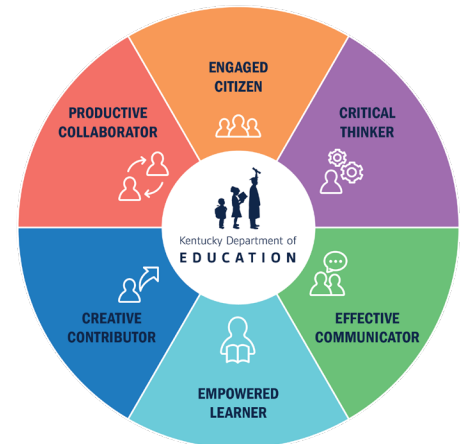
AND SO: MDE academic standards review committees should set specific numeric targets for fewer, more consolidated—but more rigorous and applied—standards and benchmarks during each standards revision cycle.

It will be a major challenge to “see the forest over the trees” on this, as each individual benchmark considered in isolation seems important—especially to experts who know that content area well. But ultimately this must be done; this recommendation had truly universal support from every person we spoke to.

1C. Create a state “Portrait of a Graduate”

CONTEXT AND RATIONALE:

- Even when state standards have been improved by the prior two recommendations, they still don’t explicitly center skills critical for college, careers, and life—for example collaboration, creativity, communication, and resilience.²³
- At least 17 states²⁴ have created “Portraits of a Graduate”²⁵ to fill this gap. In essence, these portraits are statewide frameworks of what it means to be a well-rounded and future-ready graduate (see figure to the right for an example). They’re often created through an extended community engagement process.
- Having a state portrait of a graduate is valuable because it asserts a state’s support, leadership, and vision that a deeper set of skills are important. It can also provide a helpful common framework and reference point for developing more hands-on and applied assessments, and for schools and districts to develop their own locally-adapted portraits.
- Minnesota does have a College and Career Readiness framework,²⁶ which is a nice starting point, though a larger community engagement and drafting process is needed to create a portrait for our state.



KENTUCKY PORTRAIT OF A LEARNER

AND SO: The Minnesota Legislature should charge a formal working group—led by an organization or a coalition—to develop a Portrait of a Graduate for Minnesota.



The working group should involve youth and be explicitly inclusive of Minnesotans with diverse backgrounds and experiences. In the spirit of authentic community co-creation, the working group should be run in partnership with MDE, but not explicitly led by the state agency.

We close this recommendation with a caveat. Most states have struggled to mesh their Portraits of a Graduate with their more granular state academic standards. Schools are asked to “do both”, and the two feel basically separate. It’s prudent to explore how a portrait and state standards could eventually be seamlessly integrated—for example, the Portrait of a Graduate could define higher-order skills, which then also appear as an organizing framework within future revisions of state standards, across all subject areas.

The Portrait of a Graduate could define higher-order skills, which then also appear as an organizing framework within future revisions of state standards across all subject areas.

RECOMMENDATION AREA 2

Grade 3-8 Assessments as a Barometer



Imagine a More Equitable, Growth-Focused “System Check”

We entered this project really hoping to find a fundamentally better approach to annual grade 3-8 ELA and math assessments. In the “landscape scan” portion of this project, we looked for examples across all 50 states. Unfortunately, we found no such panacea.

Table 2 describes the approaches we explored. We don’t feel we can recommend any of these, at least for now. Rather, we suggest keeping grade 3-8 MCAs focused on being a “benchmark” or “system check.” *But*, within that focus, make important changes to better align with and support serving each student—i.e. make that check more equitable, student-centered, and growth-focused. Our recommendations in this section suggest how.

Wait, seriously? Not something bolder?

At least currently, annual ELA and math tests in grades 3-8 are required by federal law, which alone presents a stark reality. Further, we did hear some arguments against a more dramatic departure from the current approach, on two levels:

- **1.** While standardized tests have limitations, for grades 3-8 they’re reasonably good at assessing and “system checking” core skills in ELA and math that lay a base for future learning.²⁷ As Recommendation Area 1 above is acted on (i.e. standards become more focused and aligned with learning science), this would be even more true.
- **2.** As described above, our education system unfortunately manifests systemic inequities, along lines of income, race, English learner and special education status, and other factors. Especially given the groundwork-laying nature of the skills assessed in younger grades, keeping a system check on these inequities remains important.

Ultimately, we need to both acknowledge the limitations in continuing with an approach that relies on standardized tests, and also recognize they’re a reasonable check against systemic inequities in core skill development. Our recommendations below represent our attempt to straddle this tension. (And—spoiler alert—we do feel a bolder approach is needed for high school, as described later, in Recommendation Area 3.)

| APPROACH | CENTRAL IDEA | OUR CONCLUSIONS |
|-----------------------------------|---|--|
| Matrix Sampling | Still use standardized tests but give a smaller subset of questions to each student, or test only a subset of students. ²⁸ | Aggregate results yield conclusions at the school/district/state level, but not the student level. While this seems to fit better the paradigm of MCA as a “system check,” this isn’t possible under current federal law. We also heard concerns from some that this approach could leave particular groups (for example, by grade level, income, race, and special education status) without a “system check” specific to that group. |
| Through-Year Assessments | Use a “benchmark” assessment (like a version of the NWEA MAP) as the state’s summative test. Some states have piloted this, such as Nebraska and Georgia (though the latter has since abandoned their pilot). | A challenge is that every interim assessment throughout the year then becomes high stakes. The net result has been more high-stakes test fatigue for schools and students. |
| Common End-of-Course Exams | Exams given at the end of a unit or course are common (or drawn from common items) across the state. Results across exams are aggregated for a given student and used in place of a standalone MCA. | Similar to through-year assessments, every end-of-course exam becomes higher stakes. This approach could also stifle innovative (for example, competency- and project-based) models. Finally, maintaining item security is tough and this would also be hard to do under current federal law. |
| Performance Assessments | Use teacher-scored performance assessments in place of standardized tests. | A challenge with performance assessments in grades 3-8 is covering the full domain of standards as required by federal law. It’s also a reasonable argument that, especially in elementary, many of those standards are important to cover (and assess) to build core skills. We do think this approach holds promise for high school (see Recommendation Area 3). |

Table 2. Other approaches we explored for Grade 3-8 state ELA and math assessments.

WHAT ABOUT SCIENCE?

While the vast majority of state tests taken in grade 3-8 are in ELA and math, Minnesota does give a science MCA in grades 5 and 8. We decided not to offer recommendations for science here, instead focusing on ELA and math.

The badging/performance assessment-based approach described later, in Recommendation Area 3, might be feasible to apply for the science MCAs, even in grades 5 and 8, since science is particularly conducive to applied performance assessment. However, a detailed discussion of this is out of scope for this paper.

2A. Involve youth in test development and prioritize cultural relevance

CONTEXT AND RATIONALE:

- Currently, MCA test items are developed primarily by educators with high levels of subject matter expertise. While MDE does make efforts to be conscious of many forms of diversity among the educators who partake in this process, and is piloting work to involve students more in statewide assessments, assessments are still largely adult-driven.
- Further, there is a process (namely, through “differential item functioning” or DIF analysis²⁹) by which test questions are checked to see if different groups of students (considering gender, race, and English Learner status) perform differently.
- There is still room for improvement in the creation and review of test items themselves, so that they draw on the cultural context, perspectives, and lived experiences of youth—in particular, youth of the many different racial and cultural backgrounds present in our state.
- This is important so that the tests are valid measures of *learning*, without tripping up a student on cultural or linguistic context they may not have.³⁰ Particularly when tests are used in an accountability context, that validity is a moral imperative.

AND SO: Youth should be involved in designing and reviewing test items, to ensure they are rooted in youth perspectives and lenses.



This will require intentionality and care for the young people who would be entering a mostly adult-driven process. Adults staffing this work with youth should see themselves as coaches and guides, not just facilitators of a technical process. Further, youth should be supported and coached specifically in applying the lens of cultural relevance³¹—so that tests invoke examples and language reflective of their own backgrounds, perspectives, and lived experiences.

2B. Accelerate the release timeline (and user-friendliness) of test results

CONTEXT AND RATIONALE:

- The current schedule and format of MCA results releases could be improved to make the data more useful as a “system check” or “benchmark.”
- Schools and districts would benefit from having fast access in the spring to preliminary results that are (1) easy to view, in aggregated form, to inform their reflection and planning; and (2) in a format that can be smoothly or even automatically imported into their student data systems (a process already at least partially in place with many technology vendors).

- Further, individual students and parents described a desire to see their results more quickly. This would help maintain the relevance to, and thus participation rates of, individuals. “By the time I got the results, I didn’t remember what was on the test—or even that I had taken it,” commented one student.

“By the time I got the results, I didn’t remember what was on the test—or even that I had taken it,” commented one student.

- Other states have different timelines and technology tools for the release of results. Take Hawaii, for example:
 - Districts, schools, and even individual teachers can access a dashboard to see preliminary, continually updated, automatically-aggregated results in real time at the school, district, and state level, even while testing is still in progress.
 - Families receive notification via email of their students’ official results typically in early August. Many parents and students we spoke to for this paper said that getting results before a new school year begins would be a big improvement.

AND SO: Preliminary MCA results should be continuously aggregated in an online dashboard for administrators and teachers, and easily or automatically imported into district data systems. Official student-level results should be released to families before the new school year begins.



We realize there is no simple solution here. Districts’ internal processes and technology systems differ substantially. And, MDE doesn’t have educator or family contact information to “bypass” districts and communicate results directly. Rather, it will be critical that MDE continue to invest in understanding, integrating with, and accommodating districts’ processes and systems.

This is extra important with respect to individual student score report distribution, which we’ve heard is an onerous task often involving many manual steps. Could MDE provide files in bundles or formats that would make districts’ jobs easier? Or even create a system whereby districts could provide family contact info as part of their student roster (along with perhaps a customized introductory message), with the understanding that MDE would then automatically send results to families once individual reports were ready, ideally in early August?

Finally, we recognize the extra steps that go into releasing the *final results* at the district and school level, which are used for accountability calculations. We urge that this continue to be as soon as possible, though keeping a target of September or October seems realistic (and is the timeline used across all other states we looked at).

2C. Prioritize measuring GROWTH in the design of the Grade 3-8 MCA

CONTEXT AND RATIONALE:

- Almost universally, the educators that we spoke to said that they believe it's equally or more important for a "system check" to capture information about growth than proficiency status.
- To be clear, we heard that "proficiency" is absolutely important as an end goal, but that the best (and truly only) way to get there is to focus on growth. Strong growth will only ever lead to higher levels of proficiency.
- There is also evidence that prioritizing growth reduces the economic and racial bias inherent in using proficiency scores to make school-level "quality" conclusions.³² Growth is a better and more equitable indicator of school contributions to learning.
- As described further below, measuring growth is complicated.³³ Conceptually it requires assumptions about an underlying "learning progression" along which growth happens, which isn't always fully there in terms of standards or curriculum. Even when imperfect, growth measures still tell us important things about student development over time.

DEFINING GROWTH

For the purposes of this paper, we define growth as a student's progress within an interconnected domain of knowledge or along a learning progression.

It's equally or more important for a "system check" to capture information about growth... "proficiency" is absolutely important as an end goal, but the best (and truly only) way to get there is to focus on growth.

AND SO: Going forward, we need to prioritize growth in our state assessment system design—namely, in grades 3-8 where standards within and across grades more clearly represents a learning progression.



Detailed technical recommendations for assessments and growth models are beyond the scope of this paper. Rather we suggest five specific priorities Minnesota policymakers and state assessment administrators must prioritize to accomplish this growth recommendation:

SIDEBAR: Five priorities for growth model design

i. Root in intuitive concepts of growth (i.e. a vertical scale)

Nearly everyone’s conceptual understanding of “academic growth” draws heavily from physical examples. For example, as a child grows taller they gradually climb a growth chart at their pediatric visits.

In order for growth measures to make intuitive sense to folks without an assessment background—and ultimately, to be trusted, used, and resilient politically—they should lean on this physical growth metaphor as much as possible. Minnesota learned this the hard way several years ago, with its foray into “z-score” growth measures, which we repeatedly heard were unintuitive and mathematically confusing.

Ideally, Grade 3-8 MCAs would be scored on a “vertical scale”, i.e. yield scores along a single, continuous scale that is used across grades. For example, in Iowa (which scores its state assessments on a vertical scale), a student’s math score might grow from 403 in third grade, to 437 in fourth grade, 458 in fifth grade, and so on. With a vertical scale, a student’s scores are able to be compared, within or across grades, to gauge their “amount” of growth. (See Appendix for more on vertical scales.)

A lot of psychometrically-respected states use vertical scales, and it’s critical that Minnesota takes this path.

To reiterate, while we acknowledge an underlying vertical scale is not essential for measuring growth from a psychometric perspective, we do believe it’s critical for clearly and sustainably communicating growth to general audiences.

Building a score scale, consistent in interval across grades, is technically difficult and unlikely to be perfect. But a lot of psychometrically-respected states use vertical scales, and it’s critical that Minnesota takes this path.

ii. Aim for low correlation with prior proficiency and SES (i.e. not a value table)

Right now Minnesota uses a “value table” as its growth measure. In this approach, a student’s year-over-year change in proficiency status (i.e. moving from partially meets to meets standards) determines whether and how much they “grew”.

The problem with value tables is that they are highly correlated with prior proficiency, and with socio-economic status.³⁴ Further, value tables don’t equitably capture any growth for students who are far behind, made substantial progress, but are not yet near “proficient” at their grade level—arguably those whom our assessment system should most push us to serve.

iii. Measure “content-referenced” growth (i.e. progress through material)

A theme we heard over and over again in our conversations with students and educators is that, on a philosophical level, “growth” should be defined in terms of an individual’s progress in grasping material—rather than in comparison to other students.

“Growth” should be defined in terms of an individual’s progress in grasping material—rather than in comparison to other students.

For example, a growth measure could be defined as student progress (for example, change in vertical scale score) relative to a “yearly growth target” at a given grade. Targets for

each grade could be set by committees that pour over state standards, individual item performance, and student score results (across various prior-year scores). Ultimately, they would determine what score change would represent rigorous, but also realistic, progress through the learning progression defined by state standards. We recognize the imperfect assumptions required in this approach (in particular, regarding the spacing and interpretation of a score change at different score levels), but feel they are worth making to keep the focus of a growth measure on progress through material.

On a related note, we did consider Student Growth Percentiles (or SGPs) as a growth measurement approach. Ultimately, we decided the approach is in conflict with the values we heard resoundingly from students, of wanting their growth to be gauged as progress in building knowledge and skills, *not* as comparison to peers. We also note that, under a SGP approach, in a circumstance where every school and student in the state improved greatly (really, what we most hope happens), the number of students showing strong growth would remain the same—i.e. *nobody* would show better growth.³⁵

iv. Measure learning for low and high performing students (i.e. highly adaptive)

For any growth measure, it's important that the underlying *assessments* on which it is based are able to capture learning across a wide range of performance levels—including those far ahead and far behind grade level. Minnesota's current state tests—which adapt the difficulty of items *within* grade—do a reasonable job of this.

Longer term, we envision an MCA that captures enough information about grasp of grade level standards to make proficiency determinations as required by federal law, but is fully grade adaptive in its questions, to inform a more accurate scale score for the purposes of growth measures. This will only get easier as Minnesota's academic standards become more focused and better represent learning progressions (see Recommendation 1A and 1B above); tests of the future could adapt not just in terms of general "item difficulty", but more smartly among and along learning progressions.³⁶

Some states are already doing this. For example, Nebraska uses a ~40-item test, about two-thirds of items are on-grade and one-third are off. While only the on-grade questions are used for proficiency determinations, all questions are used in their growth calculation.³⁷

v. Consider adjustments to individual-level growth reporting

We acknowledge that shifting to more fully adaptive tests could potentially reduce the ability to make conclusions at the individual student level. One possible solution is to no longer report individual growth, but instead report school-level, and potentially classroom-level statistics.

We heard mixed feelings on this. Most educators we spoke with supported this, saying that as a "system check" the MCA should focus on yielding data at school, program, or grade level. We spoke with almost no educators who used individual student MCA scores to inform teaching.

Similarly, students we heard from supported this idea. Many did not like the feeling of being "labeled" by a test given at a single point in time. They preferred to get information on how they were doing from classroom tests and benchmark assessments more aligned with recent their learning.

On the other hand, we talked to parents who expressed that knowing where their students are at is important. We do note that literally every district that we talked to gives a benchmark exam (such as the NWEA MAP or FastBridge), and reports that information to families—often in a format and with a frequency that is more helpful for families than the MCA.

One compromise could be to simultaneously eliminate individual growth reporting, while also adding to Minnesota law a requirement that benchmark exam results be released to families, to be sure they have information on their specific student's growth (though, again, that is already the norm).

RECOMMENDATION AREA 3

High School Assessments as a Bridge



Imagine “Badges” as Causes for Celebration (and Credit and Jobs)

A truly universal theme we heard in our conversations—especially those with students and teachers—was that state high school MCAs are not working well. We believe Minnesota needs a fundamentally different approach for high school assessments.

We can and should ask more from our high school assessment system. Imagine if it could:

- Give causes for celebration and help students mark their progress.
- Help students to understand, articulate, and gain confidence in their strengths.
- Cover not just abstract knowledge, but skills like collaboration and problem solving.
- More equitably offer credentials and open doors to whatever is next in a student’s life.
- All while continuing to serve as a “system check” on equitable learning opportunities.

We propose here a set of interconnected recommendations to realize this potential.

3A. Launch a high school “badging” system

CONTEXT AND RATIONALE:

- Measuring many of the deeper, applied skills critical for college, career, and life readiness is out of scope for a standardized assessment like the MCA.³⁸
- The high school students we spoke to for this paper understood the purpose of the MCAs in theory,

but said almost universally they didn't really care about them—introducing low-effort response bias issues that threaten the validity of interpreting their test results.³⁹

- Furthermore and relatedly, some students simply don't take them; opt-out rates are far higher for the high school MCAs, and have been growing in recent years.⁴⁰
- Finally, given in 10th and 11th grade, high school MCAs miss learning that happens in later grades, or by students who follow non-traditional pathways through coursework.



AND SO: Minnesota should launch a “badging” system for high school students. Imagine Girl Scout/Boy Scout badges, but oriented around academic content, applied competencies, and career skills, and with a system for assessing and tracking progress.⁴¹

To be clear, all high school academic standards would still be mapped to a required badge—so a badging system wouldn't mean “bypassing” state standards. Take high school math standards for example; badges could be created for particular strands (like Data Analysis), or for anchor standards more commonly covered in a high school course (like Geometry); or even for specific smaller groupings of benchmarks.

Additionally, badges could be created in other areas and fields such as science (like the “arguing from evidence” substrand); social studies (like the “ways of knowing” standard in the ethnic studies strand); career and technical education areas (like welding); or even linked with existing certificates (like the certificate for Clinical Medical Assistants).

Each badge could have levels of mastery, such as Novice, Competent, Proficient, Advanced; or Bronze, Silver, and Gold. Each level would have specific performance descriptors articulated.

In this badging system, high school assessments would be recast as pathways to badges rather than as isolated events. Ultimately, badges would be awarded by teams of educators after students demonstrate they'd met the performance standards defined for it. To validate learning, those educators would use various assessment techniques: multiple choice tests, performance assessments, exhibitions and presentations scored against rubrics, etc.

A core challenge that statewide assessments solve is uneven (and oftentimes inequitable) expectations across schools/districts. In recognition of this, the badges and the assessment used to award them should be created by educator working groups that cross school and district boundaries. Those working groups would regularly check for consistent and high quality scoring across districts (see the New Zealand example below).

BADGES: ONE OF SEVERAL SYNONYMOUS TERMS

A number of terms are used for this concept, all of which basically refer to the same idea, including: microcredentials, credentials, certificates, and competencies.

Ultimately, badges would be awarded by teams of educators after students demonstrate they'd met the performance standards defined for it.

3B. Expand acceptance of badges by colleges and employers

CONTEXT AND RATIONALE:

- Currently, access to credit- and credential-granting opportunities (such as Advanced Placement course enrollment and credit, and post-secondary enrollment options) is inequitable, often varying based on factors such as a student's prior achievement, income, and race.⁴²
- Even with a badging system the question “Why is this relevant to me?” raised by students taking our current high school assessments is still fair.

AND SO: These badges should be not only “exit tickets” focused on expectations in high school; they should *also* have real-world value in colleges and the workplace.



The state could work with colleges (perhaps starting with the Minnesota State system) to set up credit articulation agreements such that if a student scored at a certain level on a badge, they would be eligible for college credit (as Minnesota Bilingual Seals already offer; see Sidebar).

Further, the state could work with employers, encouraging them to accept badges in the hiring process. For example, a student could bring a badge in welding to get a welding job; a language badge to a translation job; or a collection of badges in English Language Arts and Social Sciences to apply for a researcher job at an education policy organization like Education Evolving.

Ultimately, this approach could go a long way towards addressing deep, systemic inequities in our state. A badging system levels the playing field, giving all youth the opportunity to demonstrate what they know and gain momentum toward wherever they're headed next.

This approach could go a long way towards addressing deep, systemic inequities in our state... giving all youth the opportunity to demonstrate what they know and gain momentum toward wherever they're headed next.

AN OPEN QUESTION: BADGES IN MIDDLE SCHOOL?

This paper breaks up our recommendations into those for grade 3-8 and those for high school—mostly because current assessments look different for each of those grade bands.

In practice, the “badging” system proposed here in Recommendation Area 3 may work well also for middle school students—either in addition to, or potentially in place of, some of what we proposed in Recommendation Area 2.

We don't have a clear answer to offer here, except to say middle school is unique and this question deserves more consideration.

3C. Ultimately seek federal waiver to use badges as the high school MCAs

CONTEXT AND RATIONALE:

- Initially, conventional high school MCAs could run in parallel with the badging system, as the former is still required by federal law.
- This running-parallel would be fine for a while, as Minnesota educators built the case that they were assessing and awarding badges with a reasonable level of reliability and validity.
- However, eventually—with a badging system covering all high school state standards—one could rightfully argue that there was a “system check” in place, and that doing both the badging system and the high school MCA was duplicative.

AND SO: Ultimately, Minnesota should seek a federal waiver to use the badging system in place of the high school MCAs.



There may be the possibility of using the Innovative Assessment Development Authority (IADA) Pilot to get the federal law flexibility to do this. The U.S. Department of Education has announced some changes,⁴³ which signals more willingness to do that.

And ultimately, federal law could evolve to the point where the badging system could just supplant the current high school MCA without needing a waiver; even proponents of state testing have increasingly voiced issues with high school assessments in particular.⁴⁴

SIDEBAR: Badging - A concept already widely in use

Minnesota Bilingual Seals Program

Created by the Learning English for Academic Proficiency and Success (LEAPS) Act in 2014, this program enables Minnesota schools to award “seals” to graduating students who demonstrate proficiency in languages other than English.⁴⁵ The program offers three seal tiers, based on level of proficiency per standards set by the American Council on Teaching Foreign Languages (ACTFL): a Platinum Seal for Advanced Low; Gold Seal for Intermediate High; and a Proficiency Certificate for Intermediate Low.

Students may use these seals to earn college credit (originally at Minnesota State systems, but other colleges have begun accepting them as well) for language. A number of assessments have been approved to certify these skill levels—most of them developed by teams of educators in the target language. The process for both developing and administering the assessments is taken seriously; as the Seals program website asserts, they “are tied to college credit; therefore, the integrity of the assessment administration is important.”

North Dakota Digital Credentials

Announced in August 2022,⁴⁶ the state of North Dakota developed an application for publishing and storing verifiable digital credentials. The state-developed software enables students to “have all their degrees, certifications, badges and skills located in one accessible and secure location. A wide variety of badges can and will be available to be stored in the digital wallet, “ranging from welding to cybersecurity and pathways from high school to community college certificates or bachelor’s degrees.”⁴⁷

A primary goal North Dakota has articulated is enabling individual students and citizens to own a portable record of their skills, and ease transitions among high school, post-secondary education, and the workforce. While the program is just beginning, it’s worth watching as an example.

New Zealand Record of Achievement

Over the course of the 1990s, New Zealand rolled out a “Qualifications Framework” program, part of which included launching a comprehensive, national Record of Achievement (ROA).⁴⁸ Similar to North Dakota, the state-sponsored program is both a system for indexing, verifying, and documenting mastery of learning—as well as a software application to “store” each person’s learning record.

The record stores a wide variety of skills, including “Qualifications” (including National Certificate of Education Achievement, similar to a high school diploma), as well as “Components of learning”, which most closely resemble anchor standards in Minnesota (for example, “Respond critically to significant aspects of visual and/or oral text(s) through close reading, supported by evidence” within English). For the latter, each component is scored: N (Not Achieved), A (Achieved), M (Merit), or E (Excellence).

Level of mastery is determined at the local level by trained local educators using mostly performance assessments. Further, the New Zealand Quality Authority runs a formal quality assurance program auditing the fairness and validity of those local assessments and their uses.

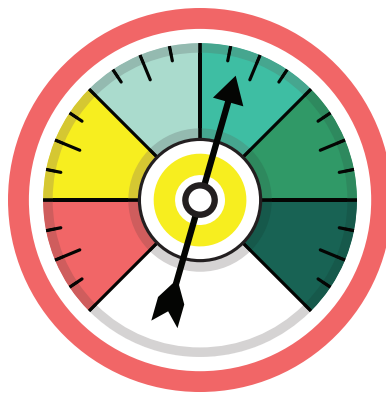
Conclusion

Assessments matter. To repeat the adage from the introduction, what gets measured gets done. Our statewide assessments—and the academic standards upon which they are based—set the highest-level “goal posts” for learning in our state.

If we want more equitable, student-centered learning, we need state standards and assessments that support and push movement in that direction.

We need standards that provide *beacons* for focused, equitable, and real-world relevant learning; grade 3-8 assessments that are timely, useful *barometers*, especially of student growth; and high school assessments that capture and celebrate more relevant learning and serve as *bridges* to opportunity. Across all these areas, students need to be more centrally involved.

These changes won't all be easy. They will require Minnesota to carve new trails, and step into a leading role, nationally. But they are possible. And our young people deserve them.



Appendix

A Psychometric Concepts Glossary

Education Evolving places a high value on simplicity, conciseness, and accessibility. We also value providing concrete recommendations that policymakers can act on. Assessments is an issue where, unfortunately, there is tension between these two values.

We have done our best to keep our language in this report simple. Still, so much of what we feel is needed on the issue of assessments is more dialogue between students and educators who use assessments, and psychometricians and policymakers who have a high level of technical expertise but less context on how assessments are experienced. This “glossary-style” appendix attempts to build that bridge of common understanding.

Types of Test Design and Interpretation

The phrases “criterion-referenced” and “norm-referenced” are commonly used in conversations about tests. Technically, a test itself is not a criterion-referenced test or a norm-referenced test; rather, a test can be designed to make conclusions and interpretations that are norm-referenced and/or criterion-referenced.

Criterion-Referenced Interpretations of Tests

A test designed for “criterion-referenced” interpretations means the test’s score can be used as a reasonable measure of “knowledge, skills, and abilities... in a clearly defined content or behavior domain.”⁴⁹

For example, the third grade math MCA is a test designed primarily for criterion-referenced interpretation. Its criterion is a students’ knowledge, skills, and abilities within the third grade academic standards for math (which were used to develop the test). Thus, a student’s score on a third grade MCA can be interpreted as a measure of the student’s grasp of third grade math standards.

Norm-Referenced Interpretations of Tests

An example of using a test for a “norm-referenced” interpretation would be comparing how one student did on that test relative to the average or “norm” student. For example, saying a student scored at the 70th percentile (i.e. scored higher than 70% of their peers) would be a norm-referenced interpretation.

If one was designing a test to make norm-referenced (rather than criterion-referenced) conclusions, one would try to include questions on which students score differently (to generate a wider distribution of scores), rather than questions which clearly determine how students did within a criterion (for example, their grasp of third grade math standards). The NWEA MAP is an example of a test designed largely for norm-referenced interpretations.

Raw Scores, Scale Scores, and Item Response Theory

Oftentimes an assessment—for example, the third grade math MCA—has more questions than appear on the test that one student takes. For example, a student may take a 40-question test, but those questions are drawn from a pool of hundreds of possible questions about third grade math. The 40 questions a specific student takes are called one “form” of the test.

The difficulty of every form of a test is not equal; some may have harder questions, some easier. The process of scaling basically converts a “raw score” (i.e. the total number of test questions answered correctly) to a “scale score” (i.e. an adjusted score that takes into account how difficult the questions were on a specific test form).

The primary method by which scale scores are produced is through an “Item Response Theory (IRT)” calculation. In essence, with IRT each test question is given a difficulty level; when students answer more difficult questions, the IRT calculation gives them “more points” on a scale score than for an easier question.

Learning Progressions and Conceptualizing Growth

A sort of unspoken universal assumption in education is that as students master material, within and across grades, they are building upon prior knowledge and growing their understanding in a content area. In short, they are “growing within a learning progression.”

Defining growth without an underlying conceptual learning progression makes no sense. For example, if there is no connection between what is learned/tested in grade 4 and grade 5, we may be able to say how a student performed in grade 4 and in grade 5, but we can’t relate the two as “growth.”

We argue that (see Recommendation 1A for details), at least to some extent, Minnesota’s state standards in ELA and math represent a learning progression within and across grades. One could argue about the extent to which that learning progression is reflected in what happens at the school and classroom level. But our state policy establishes a learning progression, and it’s fair for our assessments and growth measures to be designed on that assumption.

Vertical Scale Scores

By default, a scale score on a test can only be compared to other scale scores on the same test. For example, a 348 on a third grade math MCA could only be compared to other scores on a third grade MCA—not to, for example, to a 420 on a fourth grade MCA. This is true even if the content in grade 4 builds on the content in grade 3, and the two are connected and theoretically comparable.

To address this, test creators create “vertical scales” which essentially try to enable comparisons of test scores across different tests (such as tests at two grade levels), which each intend to measure learning along a learning progression .

The process of vertical scaling is complicated. One common approach involves having test takers answer test questions from across levels/grades (for example, items from the grade level below and above a student’s grade). Vertical scale creators then use student performance on items from different grade levels to “link” performance across grades.

Some states have created vertical scales that span, for example, grades 3 through 8 on their state

assessments. A key challenge with vertical scales is keeping them “interval scale”—that is, for a change of say 50 points from a score of 320 to 370 to represent the same “amount” of learning or growth as from 500 to 550.

Measuring Growth

For the purposes of our paper, we define growth as student learning and development within a domain over time. Drawing on the concepts and terminology described above, we can ask the question of whether a student has grown in a number of different ways. Imagine a student who took a state test in third grade and then again in fourth grade. Assessment experts often categorize growth measures⁵⁰ into four main categories:

- Did a student change in their proficiency level status (i.e. did their criterion-referenced score change their classification from “meets standards” to “exceeds standards”). This is called a “value table” and is what Minnesota uses.
- Did a student grow at a rate that will put them on track to achieve proficiency by some point of time in the future (how far out varies)? This is called a “growth-to-standard.”
- How much did a student’s score change relative to how much scores of similarly performing students changed? This is called a “student growth percentile.”
- How much did a student’s overall scale score change on an assessment? This is called a “gain score.”
- How much did the scores of students in a particular classroom, school, or district change relative to other students with similar backgrounds (most commonly, students of similar prior achievement levels and levels of socioeconomic status)? This is called “value added.”

Endnotes

1. Technically, a state must administer statewide standards-based assessments in order to get federal education funding. While in theory a state could turn down federal funding, in reality this loss of billions of dollars would leave too large a gap, and all 50 states currently comply. So in practice it's a "requirement."
2. The use of standards-based statewide assessment results in identifying schools for support is a central requirement of federal law; see: US Department of Education. 2016. "Every Student Succeeds Act (ESSA) Accountability, State Plans, and Data Reporting: Summary of Final Regulations." US Department of Education. <https://www2.ed.gov/policy/elsec/leg/essa/essafactsheet170103.pdf>
3. Maddaus, John, and Scott F. Marion. 1995. "Do Standardized Test Scores Influence Parental Choice of High School?" *Journal of Research in Rural Education* 11 (2): 75–83; Houston, David M., and Jeffrey R. Henig. 2023. "The 'Good' Schools: Academic Performance Data, School Choice, and Segregation." *AERA Open* 9 (January).
4. Dhar, Paramita, and Stephen L Ross. 2012. "School District Quality and Property Values: Examining Differences along School District Boundaries." *Journal of Urban Economics* 71 (1): 18–25; Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114 (2): 577–99
5. Koretz, Daniel M. 2008. *Measuring Up*. Cambridge, MA: Harvard University Press; David, Jane L. 2011. "Research Says: High-Stakes Testing Narrows the Curriculum." *Educational Leadership* 68 (6): 78–80
6. See Minnesota Comprehensive Assessments 3 (MCA-III) test blueprints: <https://education.mn.gov/MDE/dse/test/spec/>
7. Webb, Norman L. 1999. "Alignment of Science and Mathematics Standards and Assessments in Four States. Research Monograph No. 18." Wisconsin Center for Education Research, 1025 W.
8. For a review of many of these skills, and evidence supporting their importance, see: Esdal, Lars. 2018. "Defining and Measuring Student-Centered Outcomes." Saint Paul, MN: Education Evolving.
9. Minnesota Statutes, Section 123B.021
10. For more detailed information on the composition of statewide assessments, see MDE's "Testing 1, 2, 3" website at: <https://testing123.education.mn.gov/test/assess/res/>
11. Ibid.
12. For more on these other statewide assessments, see: <https://education.mn.gov/MDE/dse/test/mn/>
13. US Department of Education. 2016. "Every Student Succeeds Act (ESSA) Assessments under Title I, Part A & Title I, Part B: Summary of Final Regulations." US Department of Education. <https://www2.ed.gov/policy/elsec/leg/essa/essaassessmentfactsheet1207.pdf>
14. For a carefully researched summary of this history, see Chapter 2 in Kamenetz, Anya. 2015. *The Test*. New York, NY: Public Affairs. For an explicit example, see: Brigham, Carl Campbell. 1922. *A Study of American Intelligence*. Princeton University Press.
15. Randall, Jennifer. 2021. "'Color-Neutral' Is Not a Thing: Redefining Construct Definition and Representation through a Justice-Oriented Critical Antiracist Lens." *Educational Measurement: Issues and Practice* 40 (4): 82–90; Levitan, Shayna, and Nicholas Munyan-Penney. 2024. "Advocates' Guide to Demanding Racially and Culturally Inclusive State Assessments." Washington, DC: The Education Trust.
16. Mehrotra, Sarah, and Nicholas Munyan-Penney. 2023. "Future of Assessments: Centering Equity and the Lived Experiences of Students, Families, and Educators." Washington, DC: The Education Trust; Lyons, Susan, Mark Johnson, and B. Fiona Hinds. 2021. "A Call to Action: Confronting Inequity in Assessment." Lyons Consulting.
17. Briggs, Derek, and Frederick Peck. 2015. "Rejoinder to Commentaries on Using Learning Progressions to Design Vertical Scales." *Measurement: Interdisciplinary Research and Perspectives* 13 (October): 206–18;; Daro, Phil, Frederic A. Mosher, and Thomas B. Corcoran. 2011. "Learning Trajectories in Mathematics: A Foundation for Standards, Curriculum, Assessment, and Instruction.;" Duncan, Ravit Golan, and Cindy E. Hmelo-Silver. 2009. "Learning Progressions: Aligning Curriculum, Instruction, and Assessment." *Journal of Research in Science Teaching* 46 (6): 606–9; Duschl, Richard, Seungho Maeng, and Asli Sezen. 2011. "Learning Progressions and Teaching Sequences: A Review and Analysis." *Studies in Science Education* 47 (2): 123–82; Peck, Frederick, Raymond Johnson, Derek Briggs, and

Jessica Alzen. 2021. "Toward Learning Trajectory-based Instruction: A Framework of Conceptions of Learning and Assessment." *School Science and Mathematics* 121 (6): 357–68; Shepard, Lorrie A. 2018. "Learning Progressions as Tools for Assessment and Learning." *Applied Measurement in Education* 31 (2): 165–74; Simon, Martin A. 1995. "Reconstructing Mathematics Pedagogy from a Constructivist Perspective." *Journal for Research in Mathematics Education* 26 (2): 114–45; Briggs, Derek C., and Jonathan P. Weeks. 2009. "The Impact of Vertical Scaling Decisions on Growth Interpretations." *Educational Measurement: Issues and Practice* 28 (4): 3–14.

18. Briggs, Derek C., Elena Diaz-Bilello, Fred Peck, Jessica Alzen, Rajendra Chattergoon, and Raymond Johnson. 2015. "Using a Learning Progression Framework to Assess and Evaluate Student Growth." National Center for the Improvement of Educational Assessment.

19. This assertion was first and most famously made in: Schmidt, W. H., Curtis C. McKnight, and S. Raizen. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Springer Science & Business Media.

20. Institute of Education Sciences. 2019. "What Works Clearinghouse: Foundational Skills to Support Reading for Understanding in Kindergarten Through 3rd Grade." Washington, DC: Institute of Education Sciences.

21. See Minnesota Comprehensive Assessments 3 (MCA-III) test blueprints: <https://education.mn.gov/MDE/dse/test/spec/>

22. National Research Council. 2012. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press.

23. For a review of these skills and evidence supporting them, see our prior publication: Esdal, Lars. 2018. "Defining and Measuring Student-Centered Outcomes." Saint Paul, MN: Education Evolving. www.educationevolving.org/outcomes

24. Stanford, Libby. 2023. "More States Are Creating a 'Portrait of a Graduate.' Here's Why." *Education Week*, December 11, 2023, sec. Policy & Politics, States. <https://www.edweek.org/policy-politics/more-states-are-creating-a-portrait-of-a-graduate-heres-why/2023/12>.

25. For example, see Kentucky's Portrait of a Learner (<https://www.education.ky.gov/school/innov/Pages/Portrait-of-a-Learner.aspx>) or Utah's Portrait of a Graduate (<https://schools.utah.gov/portraitgraduate>)

26. See: <https://education.mn.gov/MDE/dse/ccs/>

27. See, for example, the commonly cited Cunningham, Anne E., and Keith E. Stanovich. 1997. "Early Reading Acquisition and Its Relation to Reading Experience and Ability 10 Years Later." *Developmental Psychology* 33 (6): 934.

28. Marion, Scott. 2023. "Do We Need Testing Every Year?" Center for Assessment. November 1, 2023. <https://www.nciea.org/blog/need-for-annual-testing/>.

29. See "Differential Item Functioning Analysis" heading in: Minnesota Department of Education. 2022. *2021–22 Technical Manual for Minnesota's Statewide Assessments*.

30. This is a central premise in the seminal book: Gay, Geneva. 2010. *Culturally Responsive Teaching: Theory, Research, and Practice*. Teachers College Press.

31. Randall, Jennifer. 2021. "'Color-Neutral' Is Not a Thing: Redefining Construct Definition and Representation through a Justice-Oriented Critical Antiracist Lens." *Educational Measurement: Issues and Practice* 40 (4): 82–90.

32. Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters. 2022. "Race and the Mismeasure of School Quality: Policy Brief." Blueprint Labs. Cambridge, MA: MIT Department of Economics.

33. Castellano, Katherine E, and Andrew D Ho. 2013. "A Practitioner's Guide to Growth Models." Washington, DC: Council of Chief State School Officers.

34. This conclusion is drawn from internal analyses and recommendations done in other states, which were shared with paper authors. The concern of correlation with proficiency (relative to other growth measures) is also discussed in the Data Quality Campaign. 2019. "Growth Data: It Matters, and It's Complicated."

35. For example, see "How Much Growth Is Adequate?" section in Betebenner, D. 2009. "Norm- and Criterion- Referenced Student Growth." *Educational Measurement: Issues and Practice* 28 (4): 42–51.

36. This would require cutting-edge approaches, but in theory an adaptive testing algorithm could model not just overall item difficulty, but also use student answers (and potentially artificial intelligence technologies) to identify specific areas of misunderstanding within

larger constructs, or specifically which lower-level requisite concepts along a learning progression were not yet grasped. A detailed discussion is beyond the scope of this paper but is taking place in the literature. For example Mujtaba, Dena F., and Nihar R. Mahapatra. 2020. "Artificial Intelligence in Computerized Adaptive Testing." In 2020 International Conference on Computational Science and Computational Intelligence (CSCI), 649–54; Wang, Nan, Dongxuan Wang, and Yuting Zhang. 2020. "Design of an Adaptive Examination System Based on Artificial Intelligence Recognition Model." *Mechanical Systems and Signal Processing* 142 (August):106656.

37. Nebraska Department of Education. 2023. "Interpreting NSCAS Growth Scores." https://www.education.ne.gov/wp-content/uploads/2023/07/Interpreting-Student-NSCAS-Growth-Scores_06.08.23.pdf.

38. See Minnesota Comprehensive Assessments 3 (MCA-III) test blueprints: <https://education.mn.gov/MDE/dse/test/spec/>

39. Ziegler, Matthias. 2015. "F*** You, I Won't Do What You Told Me!— Response Biases as Threats to Psychological Assessment." *European Journal of Psychological Assessment* 31 (3): 153–58.

40. Golden, Erin, and MaryJo Webster. 2019. "More Minnesota Students Opting out of State Tests." *Star Tribune*, September 2019. <https://www.startribune.com/more-minnesota-students-opting-out-of-state-tests/560350012/>.

41. Several recent reports have inventoried such systems around the country and world, including: Gagnon, Laurie, Susan Patrick, and Alyssa Weaver. 2023. "Going Beyond the Traditional: Next Gen Credentials and Flexible Learning Pathways." Aurora Institute; Goger, Annelies, and Felix Laniyan. 2022. "Whose Learning Counts? State Actions to Value Skills from Outside the Classroom." Washington, DC: Brookings Institution; Allen, Danielle, David Kidd, and Ariana Zetlin. 2022. "A Call to More Equitable Learning: How Next-Generation Badging Improves Education for All." Edmond & Lily Safra Center for Ethics at Harvard University; McClennan, N. 2023. "Credentialed Learning." *Getting Smart*. <https://www.gettingsmart.com/whitepaper/credentialed-learning/>; Hijden, Peter van der, and Michaela Martin. 2023. "Short Courses, Micro-Credentials, and Flexible Learning Pathways: A Blueprint for Policy Development and Action." Paris, France: UNESCO International Institute for Education and Planning.

42. See the detailed report on this: EdAllies. 2020. "Closing the Rigorous Coursework Gap: Supporting College & Career Readiness for Minnesota's Students of Color." Minneapolis, MN: EdAllies.

43. Klein, Alyson. 2023. "Cardona Says Standardized Tests Haven't Always Met the Mark, Offers New Flexibility." *Education Week*, November 28, 2023, sec. Teaching & Learning, Assessment. <https://www.edweek.org/teaching-learning/cardona-says-standardized-tests-havent-always-met-the-mark-offers-new-flexibility/2023/11>.

44. Jerald, Lynn Olson, Craig. 2020. "Statewide Standardized Assessments Were in Peril Even Before the Coronavirus. Now They're Really in Trouble." *Education Next* (blog). April 21, 2020. <https://www.educationnext.org/statewide-standardized-assessments-were-in-peril-before-coronavirus-bipartisan-backlash/>.

45. For more on this program, see: <https://education.mn.gov/MDE/dse/stds/world/seals/>

46. North Dakota Information Technology. 2022. "ND First State to Launch Credential App". <https://www.ndit.nd.gov/news/nd-first-state-launch-credential-app>.

47. Ibid.

48. See: <https://www2.nzqa.govt.nz/qualifications-and-standards/access-your-results/new-zealand-record-of-achievement/>

49. Nickoderm, Kyle, and Michael C. Rodriguez. 2008. "Criterion-Referenced Interpretation." In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Thousand Oaks, CA.

50. Castellano, Katherine E, and Andrew D Ho. 2013. "A Practitioner's Guide to Growth Models." Washington, DC: Council of Chief State School Officers.

We're grateful to Great Minnesota Schools, Bush Foundation, and other donors of Education Evolving for their generous financial support of our work on defining and measuring student-centered outcomes.

Report design by Jaime Chismar, Super Deluxe Design Studio.

© 2024 EDUCATION EVOLVING



CREATIVE COMMONS

This report carries a Creative Commons license, which permits noncommercial re-use of content when proper attribution is provided. This means you are free to copy, display and distribute this work, or include content from this report in derivative works, under the following conditions:



ATTRIBUTION

You must clearly attribute the work to Education Evolving, and provide a link back to the publication at www.educationevolving.org.



NONCOMMERCIAL

You may not use this work for commercial purposes without explicit prior permission from Education Evolving.



SHARE ALIKE

If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one. For the full legal code of this Creative Commons license, please visit www.creativecommons.org. If you have any questions about citing or reusing Education Evolving content, please contact us at info@educationevolving.org.





educationevolving.org