New Classrooms

Innovation Partners for Learning

The Iceberg Problem

How Assessment and Accountability Policies Cause Learning Gaps in Math to Persist Below the Surface . . . and What to Do About It

2019

Contents

Acknowledgments

We are grateful to the many policy makers, educators, researchers, and education-sector leaders who contributed to our thinking on this project. Their willingness to set aside time for robust and thoughtful discussions on policy, research, and practice substantially enriched this work. We're also particularly grateful to the teachers and school leaders who voluntarily participated in focus groups, both live and virtually.

We are also grateful to Barr Foundation, Carnegie Corporation of New York, and the Nellie Mae Education Foundation for their generous support of this publication and their thought partnership on these issues. Many people within New Classrooms contributed to the publication, especially Joel Rose, Chris Rush, Jennifer Stillman, Charles Voltz, and Michael Watson. Team members from New Classrooms' Board of Directors, External Wing, Leadership Team, and Program Wing served as vital reviewers and editors throughout the process. New Classrooms contracted with Bellwether Education Partners to provide research and advisory support and conduct the focus groups.

None of this would be possible without our partner schools, teachers, and students, from whom New Classrooms Innovation Partners learns every day.

ACKNOWLEDGMENTS

The following interviewees, partners, funders, advisors, reviewers, and external supports helped make this publication a reality, but all views, content, and recommendations expressed here, including any errors of fact or omissions, are those of New Classrooms alone.

Sara Allan
Deputy Director for Education, Bill & Melinda Gates Foundation

Tony Alpert
Executive Director, Smarter Balanced Assessment Consortium

Jay Altman
Co-Founder, FirstLine Schools

Kristen Amundson
Former President and CEO, National School Boards Association

Karim Ani
Founder, Mathalicious

Robert M. Avossa
Ed.D., Senior Vice Presidentat LRP Media Group; Former Superintendent in Fulton County, GA and Palm Beach, FL

John Bailey
Advisor to Walton Family Foundation and Visiting Fellow at AEI

Charlie Barone
Chief Policy Officer, Democrats for Education Reform

Justin Barra
Director, Education at Chan Zuckerberg Initiative

Dr. Damian W. Betebenner
Senior Associate, National Center for the Improvement of Educational Assessment

Jim Blew
Assistant Secretary, U.S. Department of Education

Jo Boaler
Professor of Mathematics Education, Stanford Graduate School of Education

Noah Bookman
Executive Director, CORE Data Collaborative

Jason Botel
Former Assistant Secretary, U.S. Department of Education

Andrea Castañeda
Chief Innovation Officer, Tulsa Public Schools

Chris Cerf
Former Commissioner of Education, State of New Jersey

Stacey Childress
CEO, NewSchools Venture Fund

Cliff Chuang
Senior Associate Commissioner, Massachusetts Department of Elementary and Secondary Education

ACKNOWLEDGMENTS

Peter Coe
Chief Academic Officer, UnboundEd

Ted Coe
Director of Mathematics, Achieve

Michael Cohen
President, Achieve

Larry Cuban
Professor Emeritus of Education, Stanford University

Richard Culatta
Chief Executive Officer, ISTE

Linda Darling-Hammond
President and Chief Executive Officer, Learning Policy Institute

Dale Erquiaga
National President and Chief Executive Officer, Communities in Schools; Former Education Commissioner, Nevada

Deborah Gist
Superintendent, Tulsa Public Schools

Joe Gleberman
Managing Director, The Pritzker Organization

Laura Hamilton
Senior Behavioral Scientist and Distinguished Chair in Learning and Assessment, the RAND Corporation

Leah Hamilton
Director of Education, The Barr Foundation

Donna Harris-Aikens
Senior Director, Education Policy & Practice, National Education Association

Neil Heffernan
Professor of Computer Science, Worcester Polytechnic Institute and Co-Founder ASSISTments

Andrew Hodge
Director of Math Innovation Zones, Texas Education Agency

Michael Horn
Co-Founder, Clayton Christensen Institute

Dr. Gisèle Huff
Executive Director, Jacqueline Hume Foundation

Abby Javurek
Senior Director, NWEA

Shavar Jeffries
President, Democrats for Education Reform

Lindsay Jones
President and CEO, National Center for Learning

David Keeling
Founding Partner, EdNavigator

Angela Kennedy-Toon
Managing Partner, Education Elements

Sal Khan
Founder, Khan Academy

Anthony Kim
Founder and CEO, Education Elements

Kenneth Klau
Executive Office of Education, Commonwealth of Massachusetts

Joel Klein
Former Chancellor, New York City Department of Education

Wendy Kopp
CEO and Co-Founder, Teach for All

Lewis Leiboh
Senior Program Officer, U.S. Education, Bill & Melinda Gates Foundation

David Levin
Co-Founder, KIPP

Art Levine
Senior Fellow and President Emeritus, Woodrow Wilson Foundation

Michael Levine
Chief Knowledge Officer, Sesame Workshop

Michael Magee
Chief Executive Officer, Chiefs for Change

Carmel Martin
Former Assistant Secretary, U.S. Department of Education

Dan Meyer
Chief Academic Officer, Desmos

Chris Minnich
Chief Executive Officer, NWEA

Pedro Noguera
Distinguished Professor of Education at the Graduate School of Education and Information Studies, UCLA

Lillian Pace
Vice President of Policy and Advocacy, KnowledgeWorks

John Pane
Senior Scientist, RAND Corporation

Todd Penner
Portfolio Director, Michael and Susan Dell Foundation

Mike Petrilli
President, Thomas P. Fordham Foundation

Carrie Heath Phillips
Senior Program Director of Student Transitions, Council of Chief State School Officers

Stephen L. Pruitt, Ph.D.
President, Southern Regional Education Board (SREB)

Macke Raymond
Director, Center for Research on Education Outcomes (CREDO) at Stanford University

Jonah Rockoff
Professor of Economics, Columbia Business School

Jason Schweid
Executive Director, Research, Design, and Strategy, UnboundEd

Mora Segal
Chief Executive Officer, The Achievement Network

Joy Silvern
Independent Consultant; Former Deputy Chief of Staff, U.S. Department of Education

Joshua Starr
Chief Executive Officer, PDK International

Diane Tavenner
Co-Founder and Chief Executive Officer, Summit Public Schools

Saskia Thompson
Program Director, New Designs for Schools and Systems, the Carnegie Corporation of New York

Thomas Toch
Director, FutureEd

Philip Uri Treisman
Professor of Mathematics, University of Texas at Austin

Marla Ucelli-Kashyap
Assistant to the President for Educational Issues, American Federation of Teachers

Daniel Weisberg
Chief Executive Officer, TNTP

Joanne Weiss
President, Weiss Associates; Former Chief of Staff to U.S. Secretary of Education Arne Duncan

Richard J. Wenning
Executive Director, Be Foundation

Anne Wicks
Anne Kimball Johnson Director of the Education Reform Initiative, George W. Bush Institute

The Iceberg Problem

The Iceberg Problem refers to the observation that only a very small amount of information is available or visible about a situation or phenomenon, whereas the more comprehensive information or bulk of data remains hidden from view.

It gets its name from the fact that only about one-tenth of an iceberg's mass is seen outside while about nine-tenths of it is unseen, deep underneath the water's surface.

Executive Summary

Ms. Rodriguez has many high hopes for her sixth-grade math students. She hopes they will find joy in learning about the beauty and complexity of mathematical concepts and make connections to the world around them. She hopes they perform well on the end-of-year test so they are set up to succeed in the seventh- and eighth-grade courses designed to prepare them for high school. She hopes they will excel in high school math, enroll and succeed in college, and perhaps pursue a degree in science, technology, engineering, or math. She hopes that these degrees will open up opportunities to pursue rewarding and lucrative careers.

When the school year began, Ms. Rodriguez's students arrived from six different elementary schools. Since she didn't have access to their grades or incoming state test scores when the school year began, she was not sure what to expect. In the first few weeks, she realized that of her class of 30 sixth-grade students, maybe five were keeping up with grade-level work. She is now frustrated but not surprised that some of her lessons don't seem to stick. She tries her best to help her students understand what the sixth-grade work is asking for, but some just seem lost. She wishes she had the time to work with each of them one-on-one, to break down any misunderstandings and figure out what they may have missed in the past.

One day, sensing that many of her students were struggling with Operations on Decimals because they hadn't quite mastered Decimal Place Value in the fifth grade, she taught her students a lesson on Decimal Place Value that she thought might help. (Decimal Place Value was not included in the sixth-grade curriculum that her district adopted, so she found a lesson online that she thought might work). That day her principal also happened to come in for a classroom observation. In her post-observation conference, her principal told her to adhere to the grade-level curriculum since that is what would be covered on the statewide summative test and would thus serve as the basis for the school and district evaluation. There was little time to cover much beyond that.

Ms. Rodriguez has high expectations for all of her students and believes that all of them are capable of being ready for the rigors of high school math. But she does not see how they will ever get there if she is unable to properly address her students' unfinished learning from elementary school. She is beginning to wonder if an exclusive focus on grade-level material is truly what is best for each of her students.

This is just one example of the varied challenges educators experience every day in schools across the country—their hopes for their students are high, but the tools teachers have and the rules they are told to follow often do not yield the results students need.

In our experience, the fastest way to accelerate student learning is to provide opportunities where students are challenged at the appropriate level for their existing skills and knowledge—not too easy, not too difficult. A student might not be able to conquer a brand-new topic on their own, but with the right supports, they can learn and retain something new that was previously out of reach. This insight, known across educational and psychological literature as the "zone of proximal development," undergirds many widely used curricular and instructional strategies.

But policies from district, state, and federal educational authorities signal to them to focus their instruction on grade-level standards each year regardless of their students' zones of proximal development. Grade-based accountability systems are understood as necessary safeguards against inequity, but they may also limit the potential for more effective, student-centric instructional approaches that can better achieve college and career readiness for each student.

In developing this paper, we have drawn upon seven years of experience operating a program called Teach to One: Math in partnership with hundreds of teachers across 15 states, serving more than 40,000 students. Our work has enabled us to operate in schools governed by public school districts, charter school boards, and independent entities in urban, suburban, and rural settings. We have worked with students who are behind grade-level expectations and with students who are ahead; with students who qualify for special services; with English learners; and with students from across a variety of racial and ethnic groups.

Our perspective is further informed by a concerted research and development effort we conducted that is focused on how best to accelerate students through middle grade math standards. As part of that effort, we have meticulously investigated the standards and underlying concepts reflected at each grade level, explored and tested the mathematical relationships among those concepts, and reviewed tens of thousands of lessons that relate to those concepts. We also analyzed the results of over 100,000 summative and formative assessments, administered over six million assessments of our own, and partnered with universities and research firms in order to advance our collective understanding of how students learn math.

This paper is not only based on the experiences of our day-to-day work; it draws upon existing research, policies, and literature. We conducted extensive interviews with policy leaders, math experts, advocates, and researchers, including those with perspectives that differ from our own. We analyzed publicly available data and our own internal data on student progress. We examined results from focus groups with middle school math teachers in three cities, in schools both within and outside of our partner network, to hear directly about teachers' instructional strategies when students come in with unfinished learning from prior years as well as teachers' experiences with curriculum, assessment, and accountability.

Working directly with districts and schools across the country to address this challenge has given our organization a firsthand perspective on the challenges faced by educators to improve these outcomes. In some communities, there are particular challenges in recruiting, developing, and retaining high-quality math teachers, many of whom might have more attractive employment opportunities in other sectors. In other communities, ongoing leadership transitions at the school or district level can lead to continual shifts in organizational direction. Poverty-related issues such as trauma, violence, and nutrition are all, of course, highly relevant to student academic performance. So too are the expectations that adults have for students.

While these and other factors undoubtedly contribute to the challenges of preparing more students for high school math, we believe there is another consideration at play that has gone relatively unnoticed by policy makers: the underlying policy landscape itself and its ultimate impact on teacher practice.

> We believe there is another consideration that has gone unnoticed: the underlying policy landscape itself.

Based on our experience and publicly available research and data about middle school math, we argue:

1. **Math is cumulative.** Unfinished learning from prior years makes it harder for students to master more advanced concepts.
2. **Policies incentivize an exclusive focus on grade-level instruction.** Current education policies signal to educators to focus their instruction on annual grade-level standards regardless of individual student needs.
3. **This approach is hindering college and career readiness.** An instructional focus on grade-level instruction keeps students from addressing the unfinished learning from prior school years that is required to master more advanced concepts.

To be clear, this is not a call to reverse the principles of standards, accountability, rigor, transparency, and equity that undergird the Every Student Succeeds Act (ESSA). They are essential elements for building a school system worthy of the students they serve. Our education system gained significantly from the development of these systems, and they are substantial accomplishments.

But these accomplishments cannot be the end. Even under the most optimistic of circumstances, it would take decades for our schools to ultimately achieve the vision of every child succeeding. If our nation is to ever have an educational system that can enable all students to unlock their full potential, we will need new ideas and approaches to get there.

This is a call to federal, state, and local leaders to create the space within ESSA for more innovative approaches to learning and measurement that allow for students to take different paths to the same outcome of college and career readiness. While ESSA provides states with far more flexibility than was permitted under No Child Left Behind (NCLB), the primary growth measures used for purposes of accountability are limited by the fact they are confined to the narrow band of each grade's standards and assessments. So long as that single path defines the benchmark of success, it is unlikely that approaches to learning that accelerate students from their unique performance levels can be successful.

To accelerate math achievement, opportunity, and equity, this paper urges federal, state, and local education leaders to:

- ✓ Measure learning growth through the use of assessments that cover standards from across multiple grade levels.
- ✓ Modify accountability systems in order to incentivize instructional practices that best support each student's ability to accelerate to grade level and beyond.
- ✓ Launch Math Innovation Zones.
- ✓ Make available high-quality instructional supports and strategies that account for unfinished learning from prior school years.
- ✓ Advance a future vision for assessment and accountability that incorporates more precise measures of student learning growth.

Our goals for this paper are to push foward the conversation about assessment, accountability, innovation, and student learning, and to find ways to resolve growing tensions between grade-based accountability systems and more personalized approaches to instruction. The resolution must prioritize high, rigorous standards and protect against systemic bias so that students from every community can benefit from all the opportunities that come with a college- and career-ready education. We welcome perspectives from others in the field, including those who may disagree with our assessment of the problem and potential solutions. And we recognize that there are real tensions here, as there are in almost every policy-related educational issue.

Nonetheless, it is time to honestly confront the challenges facing too many schools and work together toward an educational system where every student can reach his or her full potential.

This paper is NOT arguing:

1. that high expectations and academic rigor are unimportant. They are essential.
2. that our nation's system of schooling is free from systemic biases. These biases are pervasive, and our educational policies must target overcoming them.
3. that standards-based reform and related accountabilities should be eliminated. They are key building blocks to future progress.
4. that proficiency doesn't matter. It does. However, for some, achieving college and career readiness in the long term requires building key foundational skills in order to get there.
5. that students should not learn any grade-level content. Grade-level exposure matters, but an exclusive focus on grade-level material can keep some students from filling critical pre-grade gaps and others from accelerating beyond grade-level expectations.
6. that the recommendations in this paper apply to anything other than middle grade math. That's just what we know best.

Introduction

Jobs requiring science, technology, engineering, and math (STEM) skills are among the fastest growing and highest paying in the country—average wages in STEM fields are double that of non-STEM occupations.[1]

By 2022, the Bureau of Labor Statistics predicts that there will be more than one million open STEM jobs added to the US economy.[2] These jobs are likely to be the engines of economic growth and opportunity for decades to come, a vital American bulwark against offshoring, and they require a strong understanding of math starting as early as possible.

But in too many schools, students regularly miss out on learning the mathematical skills and concepts required to attain these positions.[3] National and international measures show mediocre performance and minimal systemic growth for American students in math. These results are exacerbated for students of color and low-income students, for whom large, persistent achievement gaps indicate systemic failures and unacceptable inequities.

National and international measures show mediocre performance and minimal systemic growth for American students in math.

Struggles with math are particularly acute in middle school grades, where millions of students enter with unfinished learning from elementary school that can lead to their falling further behind as material becomes more complex and learning gaps accumulate.[4]

> Struggles with math are particularly acute in middle school grades.

Once students fall behind in math, most schools are ill-equipped to help them rebuild what they missed and simultaneously keep up with grade-level standards. As a result, unfinished learning accumulates such that by the time they enter high school, many are unprepared for mathematical coursework that will keep them on track for high school graduation and postsecondary success.

On the most recent administration of the National Assessment of Educational Progress (NAEP), only 33 percent of American students in eighth grade met the proficiency level or above in math, with historically disadvantaged student groups performing at less than half the rate of white students.[5]

Standards-based reform and accountability strategies arose in recognition of these unacceptable inequities in education achievement and opportunity. Federal legislation adopted in 2001 under NCLB and amended in 2015 under the ESSA requires

Figure 1
2017 National Assessment of Educational Progress
8th Grade Math Proficiency

Students	% At or Above Proficient
White	44%
Black	13%
Hispanic	20%
Asian/Pacific Islander	62%
Native Hawaiian/Other Pacific Islander	25%
American Indian/Alaskan Native	18%
Two or More Races	37%
Free/Reduced Lunch	18%
Overall	**33%**

each state to administer annual math and reading tests aligned with grade-level standards for all students annually in grades three through eight and at least once in high school. The cumulative impact of recent decades of assessment, standards, and accountability reforms has yielded progress in several areas, including

- far greater transparency into achievement gaps between student subgroups;
- increased clarity for teachers on what students need in order to stay on a college and career-ready trajectory;
- improved consistency in the setting of high expectations for every student based on college and career readiness; and
- more objective information for families on whether students are reaching key educational milestones.

More important, NAEP scores in math improved in the early years following the adoption of NCLB.[6] But over the last decade, NAEP scores in math have been relatively flat while equity gaps have remained wide and fairly static.[7] ACT math scores recently hit a twenty-year low[8] while SAT scores reflect a similar decline.[9] Policy makers can fairly debate the myriad factors that go into student performance trends and the overall impact of the law itself, but few could credibly argue that our nation's current approach to teaching students math is systematically succeeding.

This paper examines accountability and policy challenges affecting instruction through the lens of middle school math (usually, grades six through eight, and often including algebra). Middle grade math is an important focal point because middle schools are responsible for educating students who may have already fallen substantially behind, and educators have just a few years to lift all their students to a high school level. In some districts, middle school performance helps determine high school choices for students, further amplifying the importance of a high-quality middle school experience for all students.

As we have engaged in this work, what has become clearer to us is an acute tension between an instructional program that is best for each student to ensure they are ready for college and career and an underlying policy context rooted in grade-level expectations. The mathematical skills required for students to engage with grade-level material in middle school and high school are built upon a deep, conceptual understanding from previous years. Yet while many students arrive at middle school without these foundational skills, state and federal policy systems incentivize teaching to

grade-level expectations in order to curtail low expectations and inequitable outcomes. We do not see strong evidence in the field or in the research that in math, strict adherence to grade-level content in order to accelerate learning works for all students.

> Policy can set the stage for even greater transformation.

Moreover, we have seen in our work that individualized instruction and high expectations can go hand in hand, and that if we are able to identify and address unfinished learning from prior years, students can advance more quickly and successfully toward college- and career-ready goals.

Although our current education policy arrangements evolved in response to profound educational inequities, they require refinement so that the resulting incentives and measures do not create new barriers to student success. While policy alone will not solve all challenges in the classroom, policy can set the stage for even greater transformation. Our hope is for a new perspective on accountability that preserves rigor, transparency, and equity, while also creating the space for new approaches to learning that have the potential to achieve better results for all students.

It is important to note that the findings and recommendations in this paper are exclusive to the grade span and domain we are most familiar with: middle school math instruction. The cumulative nature of math, combined with unfinished learning that many students bring with them from elementary school, create a unique challenge for our nation's middle school math teachers. We encourage others with experience in other grade spans or domains to explore how these findings and recommendations apply in other areas, if at all.

Figure 2
Demographic Profile of Students Participating in Teach to One: Math 2018–19

Students	%
White	14%
Black	28%
Hispanic	49%
Asian/Pacific Islander	2%
American Indian/Alaskan Native	4%
Two or More Races	3%
Free/Reduced Lunch	76%

New Classrooms' Teach to One: Math

New Classrooms' Teach to One: Math is a personalized learning model that schools adopt as a core or supplemental mathematics program for students in grades 5–11. Teach to One: Math leverages the power and potential of multiple educators working together to give each student a targeted, individualized learning experience that enables them to learn with different groups of students and in a variety of instructional modalities, including teacher-led instruction, collaborative learning, and independent study.

Each year, the program generates a personalized curriculum that includes a set of mathematical skills and concepts connecting where students are to high academic standards. Students experience that curriculum through a combination of teacher-led, collaborative, and independent learning modalities designed to both build mathematical fluency and habits for lifelong success. Each day, a scheduling algorithm leverages up-to-date student performance data and other research in order to generate a recommended daily schedule for each student. This allows students to progress at their own pace and to take more ownership over their own learning.

More information about Teach to One: Math can be found at newclassrooms.org.

Key Insight #1:

Math is cumulative—unfinished learning from prior years makes it harder for students to master more advanced concepts.

Mathematical skills build upon one another over time.

Mathematical concepts and skills build upon one another as students advance through middle school. The instruction that students receive reflects a coherent body of knowledge made up of interconnected concepts and designed around coherent progressions from grade to grade so that students can build new understanding onto foundations built in previous years.[10]

Through our work, we have continually leveraged academic research on how students effectively progress through the K–12 mathematics landscape in order to map specific mathematical concepts and skills to the college- and career-ready grade-level standards.[11] In doing so, we see that each grade's set of skills require students to have knowledge of prerequisite skills from prior years.

For example, as reflected in Figure 3, when a seventh grader learns about Operations with Rational Numbers, the instruction builds upon predecessors from prior grades, including Understanding Integers and Rational Numbers (sixth grade), Operations with Decimals (fifth and sixth grade), and Operations with Fractions (fifth grade).

Figure 3

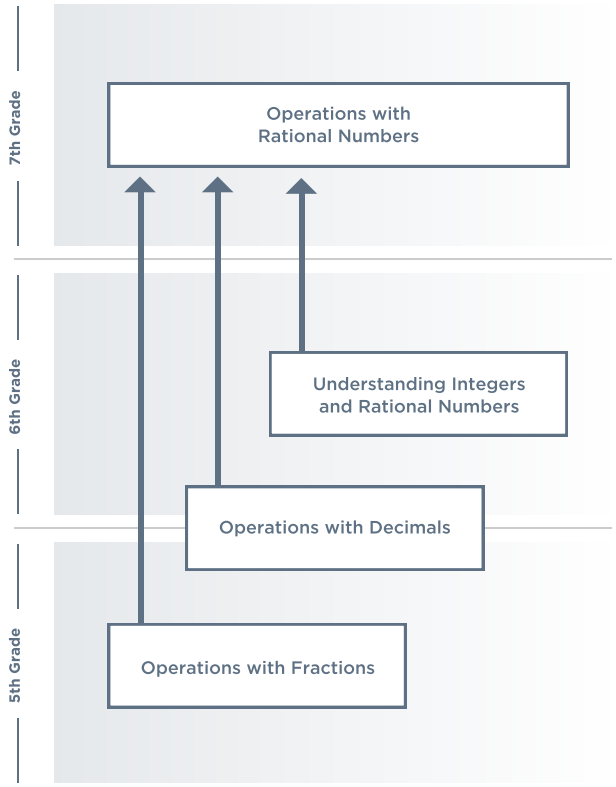
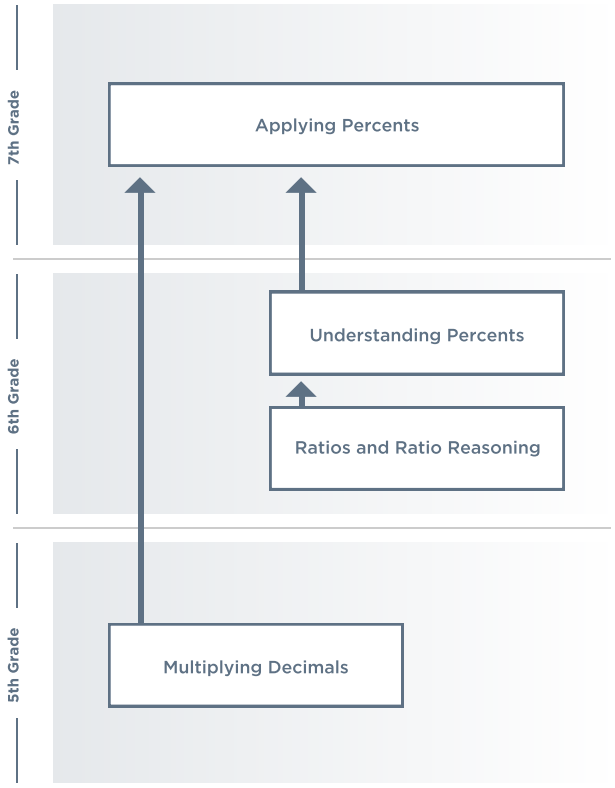


Figure 4



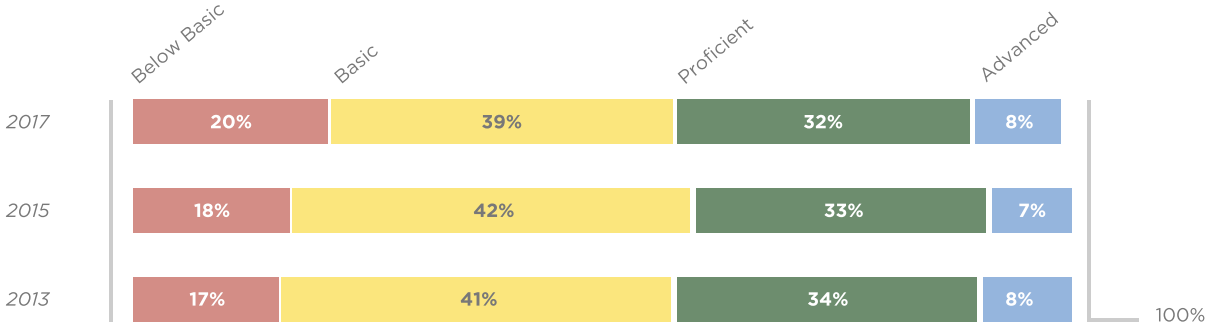
Similarly, as reflected in Figure 4, Applying Percents (also a seventh-grade skill) builds upon predecessor skills from prior grades such as Understanding Percents (sixth grade), Ratios and Ratio Reasoning (sixth grade), and Multiplying Decimals (fifth grade).

A link to the Major Concept Map that undergirds Teach to One: Math can be found in the endnotes.[12] This map reflects just one way that college and career standards can be converted to more distinct and interrelated mathematical concepts and skills, and after extensive review and analysis, has been validated by Professor Neil Heffernan at Worcester Polytechnic Institute.

Many students enter middle school with significant learning gaps from their time in elementary school.

Middle school educators must confront the reality that students come into their classrooms with a broad range of math skills and knowledge and often with substantial unfinished learning from prior school years. Figure 5 below shows fourth-grade NAEP math results for the past three test administrations. In 2017, one in five fourth graders fell into the lowest tier of math performance, well below grade level.

Figure 5
NAEP Fourth-Grade Math Achievement
2013–17



NAEP does not administer tests in fifth or sixth grades, which would allow for nationally representative estimates of math performance when students typically enter middle school course work, but fourth grade performance is a useful reference point for the large number of students nationally who enter middle school in need of significant academic supports.

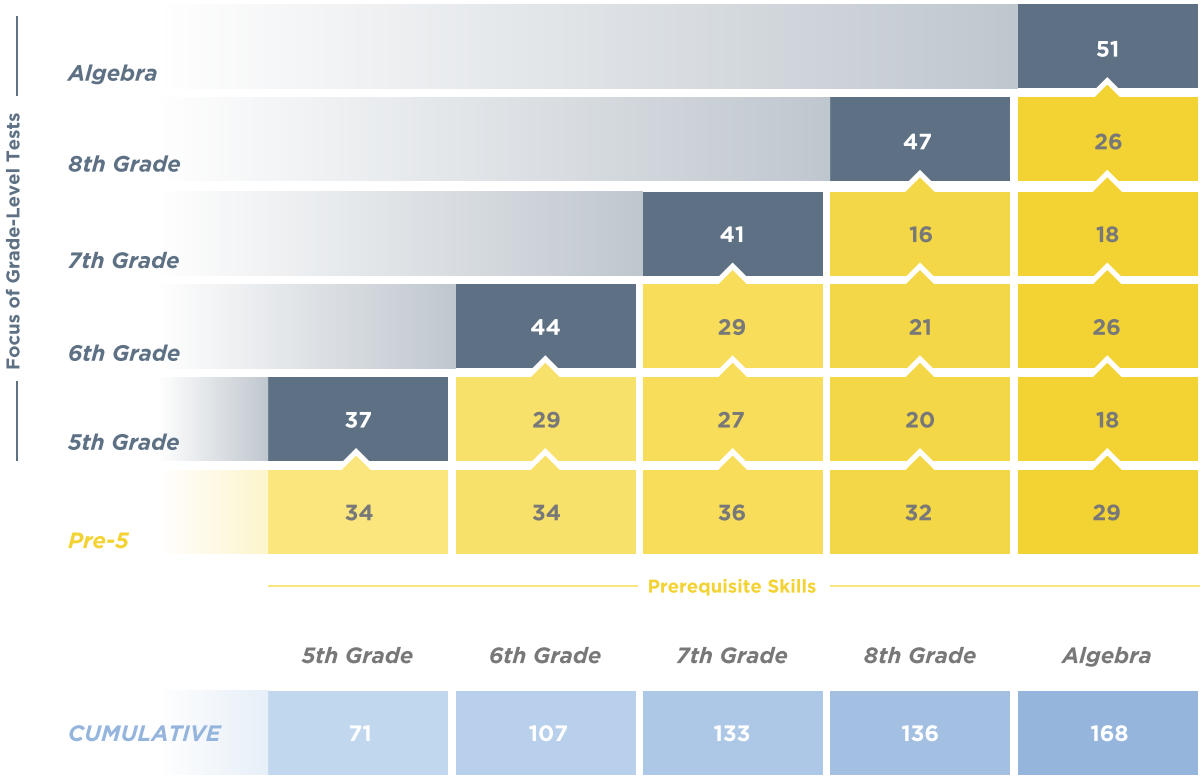
There are many reasons why large numbers of students fall behind in math by the time they enter middle school, through no fault of their own. Research on elementary school math instruction found that different schools and individual teachers vary widely in how they spend time teaching different areas of math content.[13] Significant numbers of elementary school educators (and educators in other grades) struggle with math themselves and thus report difficulties teaching math to their students.[14] While these data suggest the need for deeper analysis and reform in elementary school math that falls outside the scope of this paper, middle schools must nonetheless serve the students who arrive at their doors, regardless of their starting point.

Achieving grade-level proficiency requires both filling pre-grade gaps and mastering grade-level material.

In traditional school models, students spend most of their time working as a group on the same skills, and success is measured based on annual state tests. Figure 6 demonstrates just how problematic that approach can be for students who fall far below grade level.

The chart reflects another way of presenting data from the Teach to One: Math Major Concept Map. The darker boxes at the top of each column represent math skills that are included on the state-mandated summative test for each grade level. In the sixth grade, the summative test can cover approximately 44 skills, including, for example, how to evaluate numerical expressions. But students won't know how to evaluate numerical expressions if they never learned how to multiply and divide large numbers in prior grades, as reflected by the underlying yellow boxes. Mastering the 44 sixth-grade skills, for example, requires knowing 29 out of the 37 fifth-grade skills as well as 34 skills from before fifth grade.

Figure 6



Challenges arise when students don't fully master prerequisite skills in previous school years. This unfinished learning makes it difficult for students to learn more complex

concepts in subsequent years. It is not a simple matter of catching up from the previous year either; in some cases, critical prerequisite skills may have been introduced several years prior and are revisited in increasingly complex applications over time. When students do not master the simpler applications, these learning gaps can accumulate if students' future classroom experiences fail to recognize this fundamental challenge.

> Unfinished learning makes it difficult for students to learn more complex concepts in subsequent years.

Many middle school math teachers have too little time in a single year to both cover grade-level material and address students' unfinished learning from prior years.

The comprehensiveness of the math standards themselves generally require teachers to dedicate the full school year to fully cover all of the grade-level topics. For example, the 44 skills and concepts reflected in the sixth grade include topics such as Ratio Reasoning, Understanding Percents, and Understanding and Solving Simple Equations—each of which can take anywhere from three to four days to adequately cover (for a total of 132 to 176 school days). Over a 180-day school year, teachers may have 160 to 170 of those days for regular classroom instruction when one considers schoolwide events, testing, teacher absences, snow days, and other matters that can disrupt regular instruction. Even in a class with mostly on-grade students, teachers can struggle to cover all of grade-level topics during that time period.

But when a student starts the school year with unfinished learning from prior years, the challenge of both covering grade-level material and strategically addressing unfinished learning can become even more daunting. For example, eighth graders are expected to learn about Multi-Step Equations during the course of the school year, though some students begin the school year not having mastered critical predecessor skills such as Simple Equations, Operations on Rational Numbers, or Adding and Subtracting Algebraic Expressions. Each of these topics could readily take three to four days to sufficiently cover—and they are only the predecessors for one grade-level skill. Without sufficient time to both adequately address unfinished learning and cover grade-level material, more learning gaps can readily accumulate.

> The ground some students must cover in a single year in order to attain grade-level proficiency is substantial and improbable.

For students who enter a school year having demonstrated proficiency in prior years, achieving proficiency in the subsequent year's curriculum is both appropriate and reasonable. But for students who come into middle school multiple years behind, the teachers' challenge of both addressing unfinished learning while also successfully teaching all grade-level content is profound.

To illustrate the magnitude of this challenge, we reviewed the NWEA MAP assessment scores of incoming sixth-grade students served through Teach to One: Math. The MAP assessment is an adaptive assessment that measures student performance in ways agnostic to students' enrolled grade levels, but includes benchmarks for learning growth and performance at each grade level. MAP uses a measure called the RIT scale, which spans across grades to allow for more straightforward growth comparisons.[15]

For the purposes of this report, we divided student scores into four categories based on NWEA's norms and linking studies with statewide assessments aligned to college- and career-ready standards.[16] We define the achievement categories as students who began their sixth-grade school year:

- four or more years below grade-level standards;
- two and three years below grade-level standards;
- one year below grade-level standards, at grade-level standards, and one year above grade-level standards; and
- two or more years above grade-level standards.

Among the population of students who have participated in Teach to One: Math over the past four years in grades six through eight, at the beginning of the sixth grade school year, 9 percent of students began the school year four or more years below, 56 percent began the school year two or three years below, 33 percent of students began either one year below, at grade level, or one year above, and only 2 percent begin the school year two or more years above grade level.[17] More details on these categories and calculations are available in Appendix I.

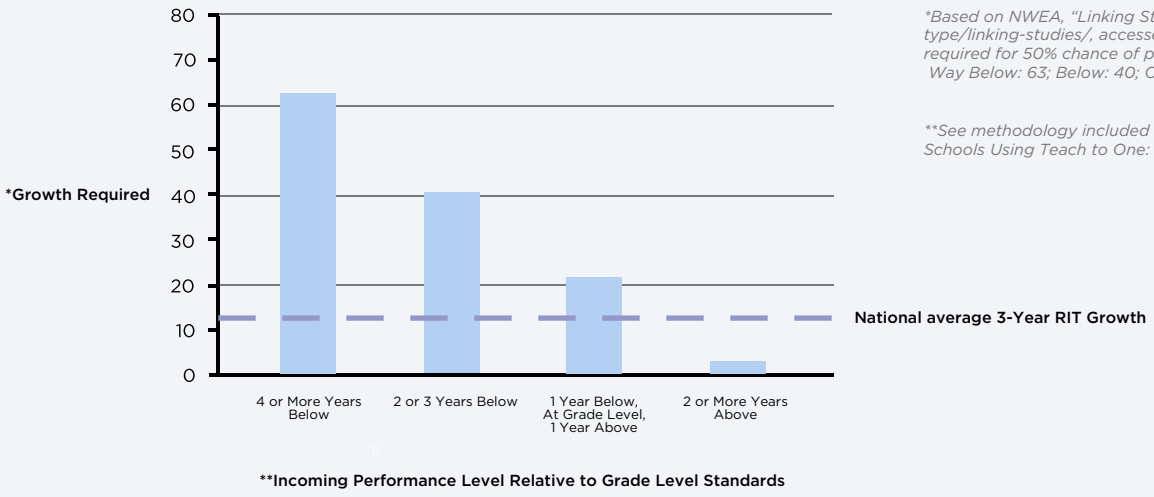
How much does a student who is four or more years below grade level standards need to progress during their middle school years to be ready for high school? To

calculate this, we used the concordance tables provided by NWEA that link students' eighth-grade spring RIT score on MAP to the corresponding PARCC readiness levels (similar studies are available for other state assessments).[18] Because the students we serve all take the MAP when they enter the sixth grade, we are able to calculate the three-year RIT gains they must achieve in order to reach proficiency on the eighth-grade PARCC.

Figure 7 shows that students who began sixth grade four or more years below grade level would have to grow by 63 RIT points over three years in order to reach a level where they are 80 percent likely to pass a rigorous state test such as PARCC at the end of eighth grade. Students who were two or three years behind must grow 40 RIT points over three years in order to meet the same bar. **As the average student typically gains 13 RIT points during the same three-year period, achieving these benchmarks would require growing at a rate three to five times the national average.**[19]

Figure 7

Three-Year Growth Required for 80% Probability of Reaching 8th Grade Proficiency [20]



**Based on NWEA, "Linking Studies," nwea.org/resource/type/linking-studies/, accessed November 1, 2018. Growth required for 50% chance of proficiency are comparable: Way Below: 63; Below: 40; On: 22; Above: 3.*

***See methodology included in "Three-Year Growth at Schools Using Teach to One: Math" in Appendix I.*

While the population of students in our partner schools is not representative of the nation as a whole (76 percent of students qualify for a free or reduced lunch versus 52 percent nationally[21]), the need for our nation to develop viable pathways that enable students to catch up and achieve college and career readiness is no less profound. If policy and instructional systems are built around the assumption that proficiency can always be reached in a single year and few students will need to substantially revisit and rebuild prior year's content, gaps in high school readiness will persist for many years to come.

The profound challenge and improbability of students catching up in middle school math was further reinforced in a recent policy brief published by the Institute for Education Policy at Johns Hopkins University. Using publicly available state assessment data, researchers analyzed sixth- and eighth-grade cohort data from 1,651 schools across six states and the District of Columbia and found less than 1 percent of schools were able to improve their proficiency levels by at least 30 percentage points.[22] The study also found that less than 1 percent of schools were able to both consistently improve their proficiency levels (by any amount) between sixth and eighth grade while also reducing the number of students in the lowest performing category.[23]

Teachers feel the tension between grade-level instruction and individual student needs.

"It's hard for a student who has shown progress and is growing to not be able to see that growth on the test that matters most. It affects our students who are consistently scoring the lowest and struggling the most." —Middle school math teacher

In our experience, the fastest way to accelerate student learning is to provide opportunities where students are challenged at the appropriate level for their existing skills and knowledge—not too easy, not too difficult. A student might not be able to conquer a brand-new topic on their own, but with the right supports, they can learn and retain something new that was previously out of reach. This insight, known across educational and psychological literature as the "zone of proximal development," undergirds many widely used curricular and instructional strategies.[24] Examples of strategies that can bring a new skill into students' ideal learning zone include "scaffolding" different forms of assistance around a concept until students can eventually complete the problem on their own[25] or "spiraling" curriculum so that teachers revisit and deepen fundamental concepts frequently.[26]

> With the right supports, students can learn something new that was previously out of reach.

But what happens when that ideal zone for learning falls outside the menu of topics in a students' grade-assigned level? Or when the scaffolding that would be required for students to access grade-level material is both so vast and nuanced to each student's unique circumstances that these strategies are effectively impractical? How do teachers approach this challenge, and how (if at all) do district, state, and federal policies support them?

"Students' feelings about math are tied to their success in math. As soon as you start giving them opportunites to succeed, and they see the growth, their feelings change."
—Middle School Math Teacher

As part of this project, New Classrooms contracted with Bellwether Education Partners to conduct focus groups with teachers in and out of the New Classrooms network to hear directly from educators about their experiences with standards, assessment, and accountability systems, and to learn how these systems might play out in different school contexts.[27] In each of the focus groups, we heard consistently that teachers shared a strong belief in the need for students to reach high expectations, but many feel caught between an expectation to focus only on grade-level materials and the kinds of skills and knowledge they think their students may need. One teacher expressed the tension this way: "The curriculum we were given says the kids should already know everything up to their grade, and they don't. I was even told a couple of times when I first started that I was teaching below-grade work, and I should be doing on-grade work. But the students weren't ready for that yet."

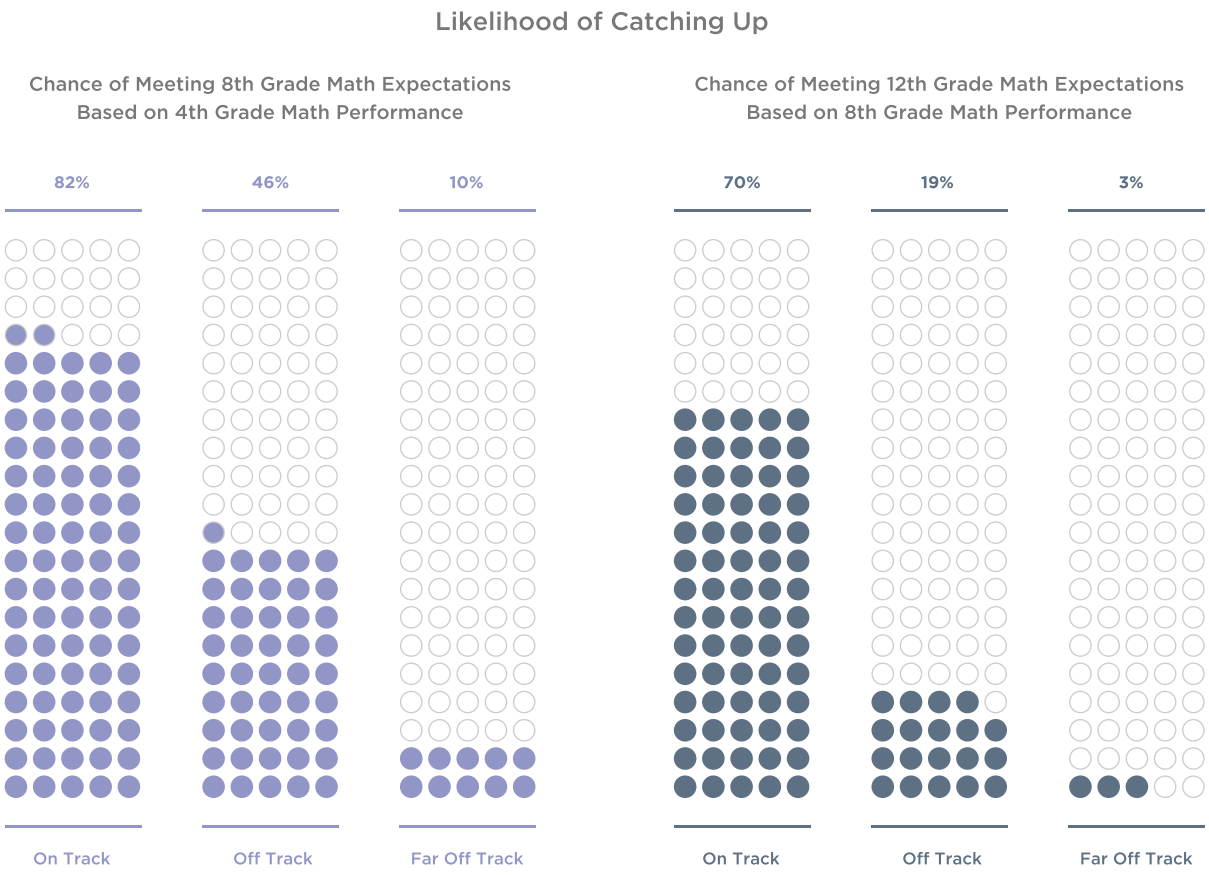
Through the course of this research, one barrier to effective personalized learning approaches that surfaced is access to high-quality materials from across multiple grade levels and supports in service of accessing grade-level content. Many participating teachers did not feel fully confident teaching skills outside of the grade-level standards in which they were most familiar. Teachers generally receive little training in how to diagnose and address unfinished learning. Focus group participants pointed to two primary strategies in response to large gaps in student knowledge: stick exclusively to on-grade content and hope for the best, or spend substantial time independently searching for off-grade instructional resources that may or may not be of high quality.

Teachers were rarely fully satisfied with available resources on how to help students who are far behind. Because many states and educational advocates define exclusive adherence to grade-level standards as a primary determinant of curriculum quality, teachers who want to provide instruction on pre-grade skills often fend for themselves to find materials for their classroom or devise their own strategies that may not be effective.

Failing to adequately address unfinished learning in middle school has long-term consequences

Longitudinal studies of individual students over time can show more precisely how students who fall behind are likely to stay behind. For example, Figure 8 shows some of the findings from a 2012 study conducted by ACT.[28] Researchers tracked results from

tens of thousands of students, one group from fourth grade to eighth grade, and another from eighth grade to twelfth grade. In both groups, the vast majority of students who started out behind in math stayed behind. Chances of catching up were even worse in the upper grades: a student who was "far off track" in eighth grade math had only a 3 percent chance of reaching college readiness by the end of high school.[29] This same study also found that students in the "far off track" group were much more likely to attend high-poverty schools.

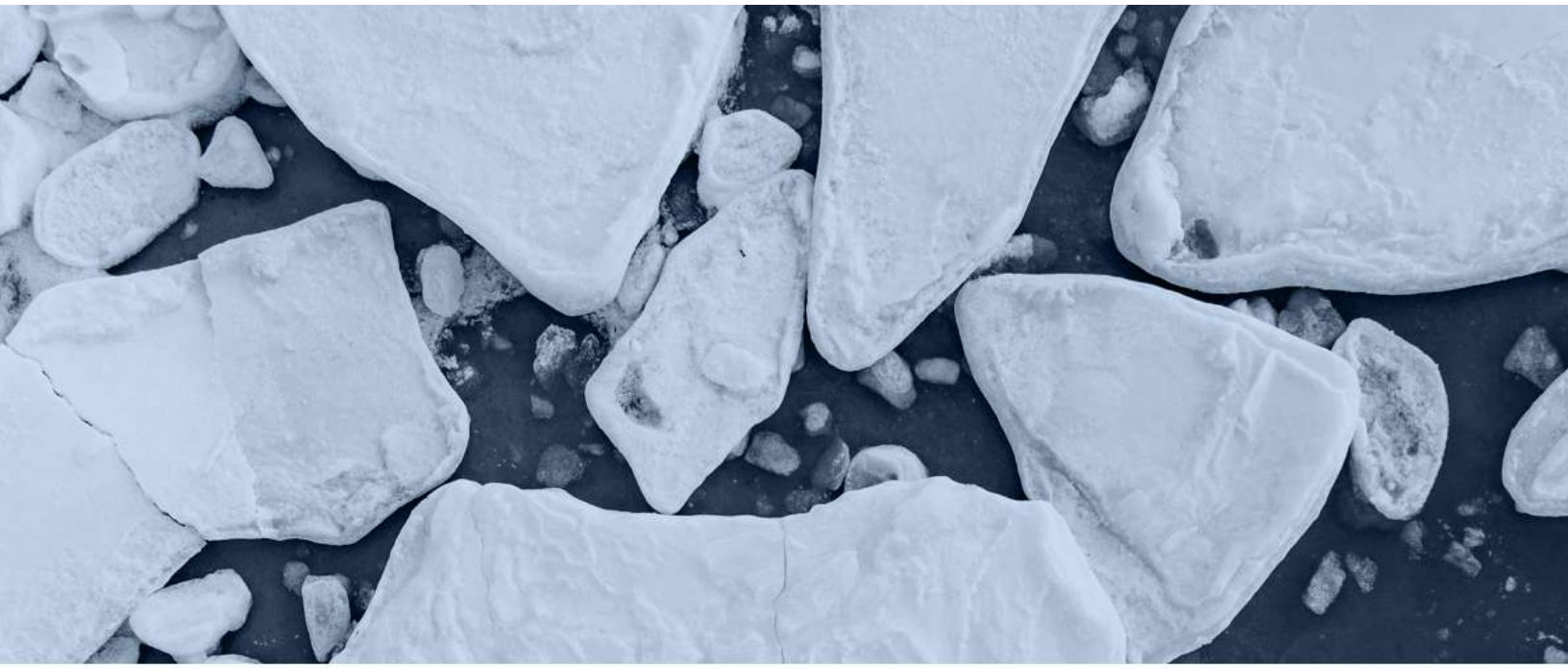


Source:Chrys Dougherty and Steve Fleming, "Getting Students on Track to College and Career Readiness: How Many Catch up from Far behind?," ACT, November 2012.

One reason why it may be difficult for students who come to high school off track to catch up to their peers may have to do with the way in which high school coursework is aligned to graduation requirements. The courses that students must pass as a condition to high school graduation generally require predecessor knowledge from middle school. However, high schools may not have these courses available since they generally would not qualify for high school credits. As a result, students have little choice but to take high

school courses they are not yet ready for—often multiple times. A 2014 transcript study by Phil Daro for the SERP Institute found that in one district, students commonly repeated courses such as Algebra I or Algebra II two or three times and that only 5 percent of students took a presumably normal sequence of courses from Algebra I through Calculus.[30]

As students struggle to pass credit-bearing high school courses, they are often allowed to participate in online credit recovery programs. The quality of these programs varies widely, and researchers have begun to question whether their popularity is playing a key role in the fact that national high school graduation rates have increased while objective indicators of student performance remain relatively unmoved.[31]



Key Insight #2

Current educational policies favoring grade-level instruction are hindering many students' longer-term success.

Education policies are oriented around annual grade-level expectations.

At the core of today's federal and state educational policies for K–12 schools is a system oriented around annual expectations at each grade level and a set of standards, assessments, and accountabilities designed to drive instructional behavior toward meeting those annual expectations. Appendix III provides a brief overview of how assessment and accountability policies have evolved since *A Nation at Risk* was published 1983.

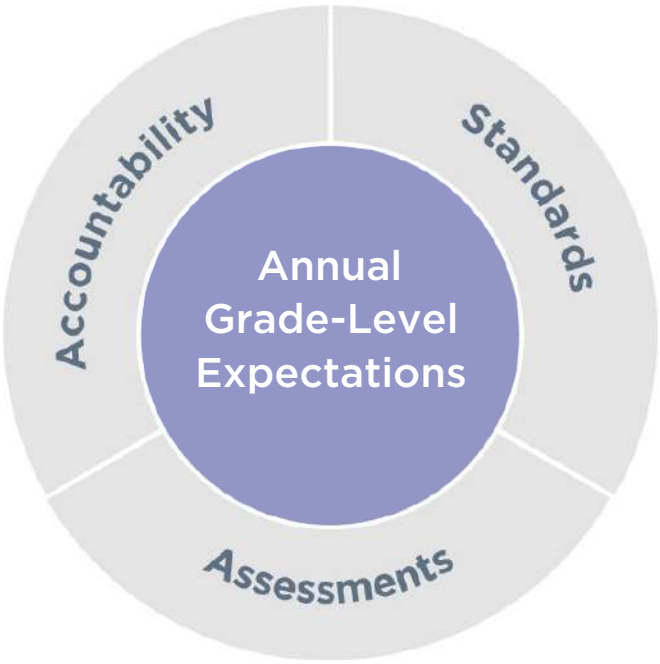
These policies aim to advance several important principles, including

- providing every student with access to rigorous grade-level instruction that will prepare them for the future;
- closing achievement gaps and combating systemic bias against historically marginalized student groups by providing equitable learning opportunities and setting clear and common expectations for success;
- holding adults in school systems accountable for ambitious, measurable learning outcomes rather than process-oriented inputs;

- giving families annual information they need to understand their student's progress, make educational decisions for their student, and advocate for change; and
- giving policy makers and system leaders information they need on an annual basis to evaluate school success and address areas for improvement.

The grade-level standards that lie at the heart of these policies are anchored in college and career readiness. ESSA requires states to adopt "challenging state academic standards" that apply equally to all public school students in the state in math, English language arts, science, and any other subjects the state designates.[32] States should demonstrate the rigor of their standards by aligning with entrance requirements for credit-bearing postsecondary coursework.

The shift to more rigorous college- and career-ready standards has been one of the biggest education policy developments of the past decade. Well-designed college- and career-ready standards show a logical progression of skills and knowledge from one grade to the next, which, in an ideal world, all children would follow to the letter. They are also intended to guide the instruction that students receive each year.



But in math, when students miss key steps along the way in this progression or learn at a pace that is faster or slower than the state standards anticipate, the standards alone do not provide guidance to teachers on where to focus instruction. They signal to a seventh-grade teacher, for example, that all seventh-grade students should be taught seventh-grade content—whether they happen to be performing two years behind grade level or two years ahead.

The grade-level expectations embedded in policy reflect a single path to college and career readiness for all students. They leave little room for other instructional paths to that same goal that are more attuned to each student's incoming performance level.

Summative assessments are not structured to measure the full range of learning growth.

Under ESSA, all students in grades three through eight must take a statewide summative assessment aligned to their enrolled grade level. All sixth graders take the sixth-grade test, all seventh graders take the seventh-grade test, and so forth. ESSA allows for only a narrow set of exceptions to these guidelines.[33]

The tests must assess student performance against the standards for their grade level, and only on-grade measures can fulfill ESSA's requirements for accountability purposes (as discussed in the next section of this paper). This means there must be enough assessment items aligned to grade-level standards so that each student's performance can be measured relative to grade-level expectations. The assessments themselves, which generally comprise 30–50 questions of varying complexity, must be valid, reliable, and administered statewide.

ESSA includes provisions that permit states to adopt assessments that also "measure academic proficiency and growth using items above or below the student's grade level."[34] However, recently issued guidance from the Department of Education specifies that if states design tests to include additional measures of off-grade performance, they must still measure and score students' on-grade performance accurately, and any off-grade measures would not apply to primary academic indicators under ESSA.[35] As a practical matter, due to the broad set of standards at each grade level and pressures to reduce test time and length, most summative state assessments are almost exclusively focused on grade-level content.

> State assessments are unlikely to detect many of the learning gains that students might make from pre-grade skills.

As a result, state assessments are unlikely to detect many of the learning gains that students might make from pre-grade skills. The gains, for example, made by a seventh-grade student entering the school year three years behind her peers and finishing the school year one year behind would likely not be fully reflected on the seventh-grade assessment, since the test itself would likely not ask her to demonstrate knowledge on topics beyond her assigned grade level.

The same holds true for advanced students: Learning gains made by a seventh-grade student who happens to progress into eighth grade are also not likely to be detected as growth, since only seventh-grade items will appear on her state summative assessment.

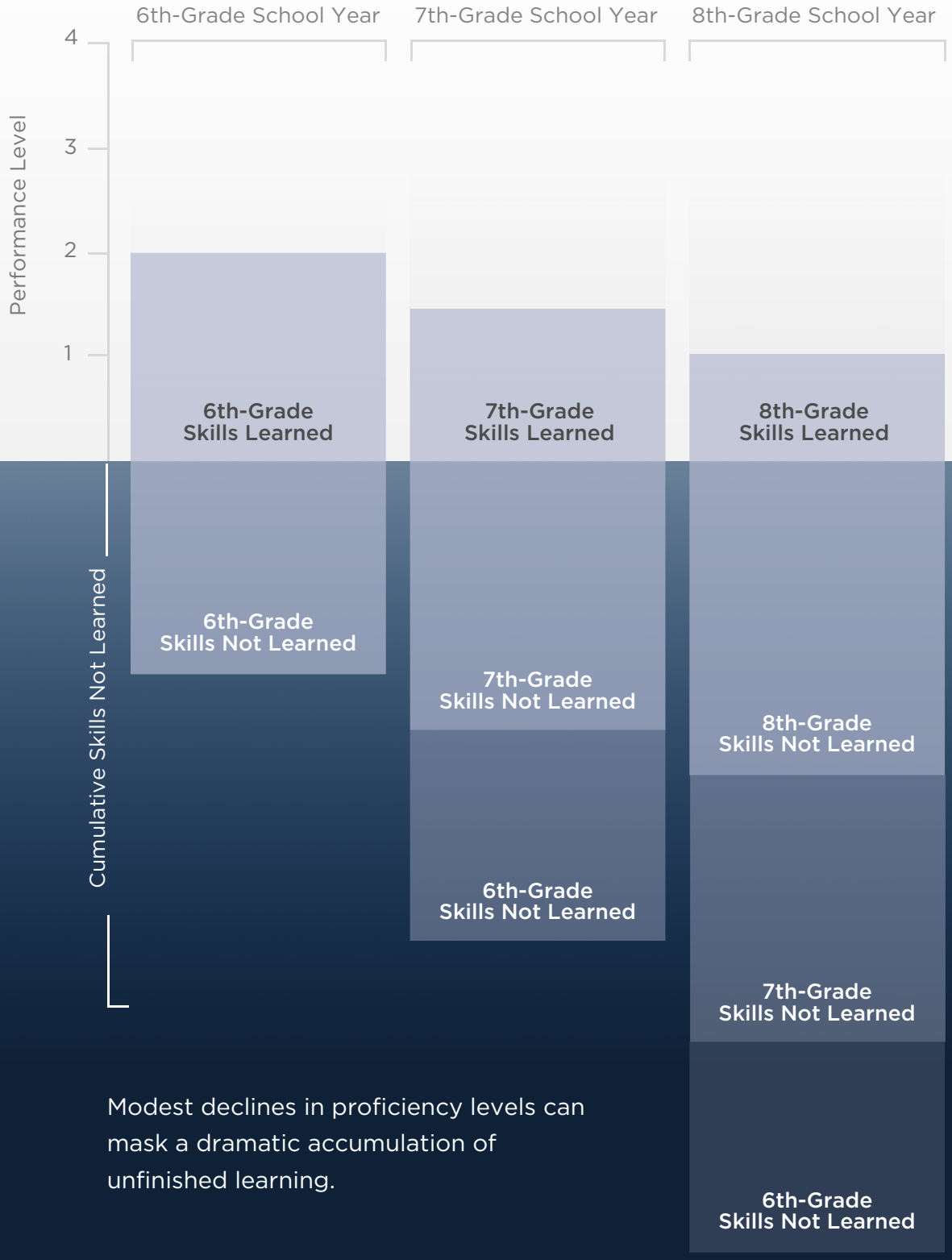
Moreover, while many of the skills that appear on each grade-level test build upon knowledge from the previous year, entire sets of new skills are also introduced. For example, a test focused on seventh-grade standards would generally include topics related to Probability that have little to do

with what is required to succeed in the eighth grade. Similarly, there are a number of eighth-grade skills, such as those related to Geometric Transformations and Scientific Notation, that do not necessarily build on anything covered in the seventh grade. While the standards are written to be cumulative, not all math skills within a given grade are necessarily applicable to more advanced concepts in subsequent grades, as Figure 10 demonstrates.

Responsible technical decisions in test design can ensure that proficiency designations from year to year follow a logical pattern, link well to the tests before and after, and reflect standards, but they cannot change the fact that many of the underlying tested skills are different year to year. This reality can lead to misinterpretation about the learning progress that students actually make from one year to the next.

Figure 10 reflects how an annual focus on grade-level proficiency can cause skill gaps to grow in ways that make it harder for some students to achieve college and career readiness. When a sixth-grade student is taught sixth-grade material, some of those skills will be learned and some will go “unlearned” for a variety of reasons (e.g., lack of predecessor knowledge, uneven teacher quality, student absences). The next year, as the focus of accountability shifts to the seventh-grade assessment, the unlearned skills from sixth grade remain unaddressed, even though those very skills may be essential to mastering seventh-grade content. By eighth grade, even more learning gaps accumulate so that by the time student enters high school, he is simply unprepared for more advanced mathematical topics.

Figure 10
How Learning Gaps Accumulate Over Time



Modest declines in proficiency levels can mask a dramatic accumulation of unfinished learning.

We call this phenomenon the Iceberg Problem

because, like an iceberg, only a very small amount of information (the tip) is visible while the more comprehensive information remains hidden from view.

In K–12 education, while the gaze of policy makers is focused on how students are performing relative to grade-level assessments, learning gaps continue to accumulate below the surface, making longer-term success harder to achieve. There is an educational path for each student to unlock their full potential, but it requires seeing students more as individuals than as a homogeneous group enrolled in a particular grade level.

Accountability systems are limited by the constraints of the underlying assessments.

Federal law requires states to use statewide assessments as the basis for accountability systems and to set goals for increasing the share of students who meet state standards in reading and mathematics, accelerating progress of underperforming subgroups, and improving graduation rates.[36] States must also identify their lowest-performing schools for varying levels of support and intervention.[37] Many states use these systems to assign ratings (such as A to F school grades) to all schools as a means of communicating performance publicly rather than just for identifying the lowest performers, but a school rating is not required under ESSA.

States have flexibility in the specifics of their identification system for low-performing schools but must include certain types of indicators:

Three academic indicators:

- Academic performance in ELA and math based on state assessment results at each grade level in grades three through eight and at least once in high school.
- Student growth or another valid and reliable indicator that allows for meaningful differentiation in student performance.
- High school graduation rates based on the four-year adjusted cohort rate.

Two additional types of indicators:

- English language proficiency, which applies only to English learners.
- At least one additional indicator of school quality or student success, such as chronic absenteeism or school climate (also referred to as “the fifth indicator”).[38]

States determine how each indicator factors in the determination of which schools require intervention and for school ratings where states elect to assign them, but the three academic indicators must each carry "substantial weight," and together carry "much greater weight" than the indicator(s) of school quality or student success.[39]

The stakes can be high.

Under ESSA, schools identified as low performing under these accountability systems may be subject to intensive interventions in certain states, such as state takeover or sanctions.

These incentives further reinforce an instructional focus on grade-level material regardless of students' unique starting points, which make it even harder for students to ultimately achieve college and career readiness.

Beyond ESSA school accountability systems, states, districts, and schools use assessment metrics as a key component in many other evaluation and decision-making activities. For example, 39 states use a measure of student growth on summative tests as part of their teacher evaluation systems.[40] Many district and school administrators believe their positions are dependent upon success on annual summative assessments. And for many charter schools, meeting test-based goals for proficiency and student growth can mean the difference between closure and continued existence.

In these cases, the required academic performance and growth measures are based on grade-level assessments. These incentives further reinforce an instructional focus on grade-level material regardless of students' unique starting points, which make it even harder for students to ultimately achieve college and career readiness.

Growth metrics used for accountability do not provide a full picture of student learning gains.

Under ESSA, growth metrics using state tests have taken on greater importance. State ESSA plans have included a variety of approaches to capturing growth. Some states use "criterion-based" growth metrics to measure the degree to which students are closer to meeting grade-level expectations than they were the previous year. Others use "normative" approaches, such as Student Growth Percentiles, to measure how students' test scores compare to those of students with similar past performance. Others use a

hybrid of the two. Readers can find links to several detailed analyses of these approaches in this paper's endnotes.[41]

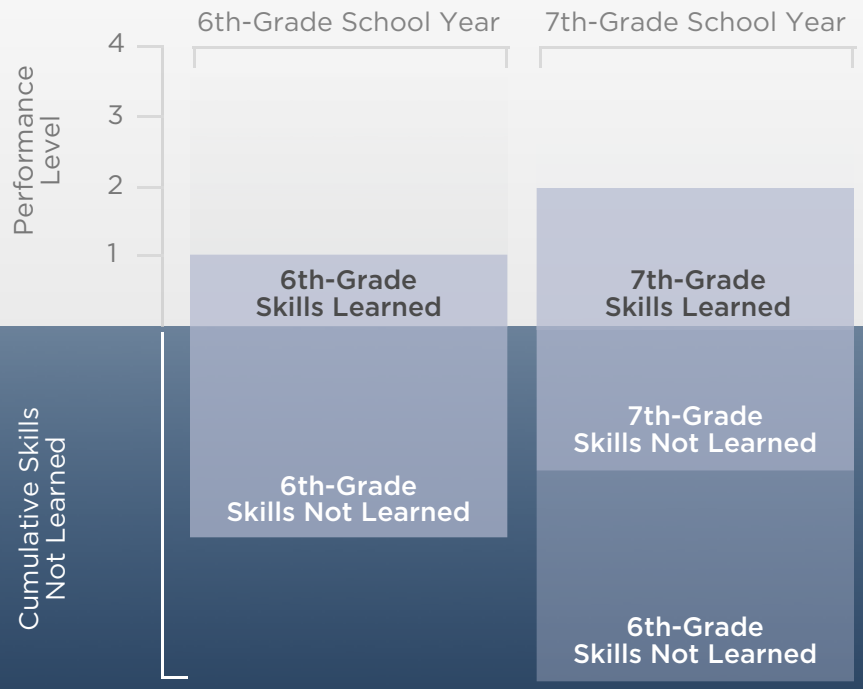
The merits, strengths, and weaknesses of all these metrics have been widely debated at both the state and national level. But these debates rarely acknowledge what all growth measures described above cannot do as long as they are based on traditional grade-level tests: They cannot precisely measure learning and knowledge gains that the tests do not assess—namely, content above or below grade level. This means that "growth" measures currently possible with state assessments only detect a narrow slice of potential learning.

Figure 11 shows how the current accountability system can give educators and parents the misimpression that a student is growing. In this example, an individual student scored at Level One on the sixth grade, as many of the grade-level skills he or she was taught over the school year went unlearned.

Those skill gaps then carried over into seventh grade, where instruction was focused on seventh-grade material (some of which built off of sixth-grade material, and some of which built off of content from earlier grades).

At the end of the year, the student may have learned enough to score a Level Two on the state test, but much of his or her unfinished learning has only compounded along the way. Even though it might appear for accountability purposes that the student is growing, the accumulated gaps will prevent the student from successfully learning more advanced concepts in subsequent years.

Figure 11
The Misimpression of Learning Growth



Often, we find that school leaders and policy makers alike do not fully understand these limitations, and speak about state growth metrics as if they are synonymous with student learning progress. In fact, the "growth" that districts and states generally report on could more accurately be described as "changes in relative performance," because the content of each grade's assessment can be quite different.

Similarly, state and district leaders may be under the misimpression that scale scores from summative assessments that linked to one another (known as vertical scaling) can be used to precisely measure student learning growth. While such comparisons are strongly supported when tests are comparable to one another, comparing grade-level summative assessments—which have varied content from one year to the next—is far weaker and inappropriate for use in higher-stakes contexts.[42] Recent guidance published by the National Education Policy Center was even more explicit about the misuse of vertically aligned instruments to measure student growth in high-stakes contexts.[43] According to the study's author, Madhabi Chatterji, Director of the Assessment and Evaluation Research Initiative at Teachers College, "growth, measured using IRT-based scale score metrics should be avoided altogether in accountability contexts, as these scales are too limited to allow inferences about student learning from grade to grade."[44]

States want to value growth in accountability systems, but their options are limited.

Under ESSA, 47 states incorporated one or more of the growth measures described above as accountability measures in elementary and middle school.[45] There are several often-cited policy rationales behind emphasizing growth measures in an accountability system.[46]

- Provide a fairer opportunity for schools whose students enter with relatively low performance to demonstrate success
- Encourage schools to support students across the achievement spectrum
- Reward schools moving student achievement in the right direction
- Identify schools where students exhibit flat or declining performance for extra support or corrective action

According to a summary of state ESSA plans by the Data Quality Campaign, only two states (California and Kansas) do not plan to include a growth measure in their ratings for grade 3–8 schools.[47] An analysis by the Fordham Institute found that 18 states weighted growth metrics for more than half of their school ratings system, and 24 states count growth at between 33 and 47 percent of a school's overall rating.[48]

The uses of growth measures go beyond ESSA accountability. Any metric used in accountability must also be reported to the public in state report cards,[49] and several states include growth indicators on student-level test reports that go to parents.[50] **Regardless of which approach to measuring growth is used, the fact that the underlying assessments are nearly exclusively focused on grade-level items dramatically limits the kinds of learning growth that can be detected and rewarded in statewide accountability systems.** These limitations may hold states back from encouraging other valuable kinds of growth and progress.

Federal policy limits the ability of computer-adaptive tests to measure learning growth across grades for accountability purposes.

Theoretically, the most comprehensive way to measure growth would be to test every student on every skill, every year, which is obviously not practically feasible or a good use of students' time. But with computer adaptive test (CAT) designs, which ESSA explicitly allows, more sophisticated growth measures that cross multiple grades may be achievable. CAT tests hone in quickly on a students' skill levels by adjusting the items students see according to their prior answers. If a student excels on the initial questions, they will receive more difficult questions in order to test the upper bounds of their knowledge. If a student gets a question wrong, they might get an easier question next. The selection of items for each student is governed by a test algorithm. As a result, CATs can pinpoint students' performance more quickly and precisely, especially for students at the low or high extremes of the performance spectrum, and potentially support more nuanced performance measures.

The most widely used example of a summative CAT test currently used at the state level is Smarter Balanced, a test created by a multistate test consortium. An original intent of the assessment was to measure "maximally accurate" results for each student by sampling content above and below grade level.[51] However, the design of the algorithm used to select items for the assessment limits the test's ability to measure gains on skills included in earlier and later grade-level content standards because federal law requires states to measure the full breadth of grade-level standards for every student. Smarter Balanced therefore prioritizes grade-level content by requiring that the first two-thirds of the assessment (approx. 22 questions) assess grade-level material using items written specifically to measure grade-level content. The pool of items may then expand to include easier or harder items that are also aligned to grade-level content but were initially written for a different grade level. The purpose of the expanded pool of items is to add precision to the measure of grade-level content knowledge and skills, not to provide information on above- or below-grade performance or overall learning growth.[52]

Another CAT, the MAP assessment from NWEA, is an adaptive assessment that measures student performance agnostic

to grade levels. Because it does not currently measure proficiency against grade-level expectations, states may not use the MAP to meet ESSA accountability requirements. NWEA has, however, developed studies that link MAP performance with scores on a variety of state summative tests.[53] We explore the implications of those studies in further detail in Figure 7.

Computer-adaptive assessments carry great potential to shape the future of student assessment. But as testing algorithms become more prevalent, it is important they are developed with a keen eye on equity so that they do not reinforce preexisting biases.

As testing algorithms become more prevalent, it is important that [computer adaptive assessments] are developed with a keen eye on equity so they do not reinforce preexisting biases.

There are barriers to innovative assessment pilots yielding more comprehensive measures of growth.

Recognizing the possibility for innovative states and emerging technologies to enhance assessments, ESSA includes an innovative assessment demonstration authority, commonly called innovative assessment waivers. These waivers allow up to seven states to design and pilot different types of summative tests that could roll out statewide after several years.

However, the requirements of this program limit states' ability to address the challenges around learning growth and grade levels described above in several ways:

- Innovative assessments must still align with grade-level standards.[54] Test items can extend beyond grade-level standards (just as the current assessments can), but innovative assessments face the same practical limitations of test time and length.
- Participating states must still generate valid, reliable, and comparable results including annual summative determinations for all students and subgroups of students.[55] Not only would the comparability provision suggest a need to focus on grade-level content, but guidance issued by the Department of Education through the Assessment Peer Review Process further reinforces the federal requirement to report against grade-level proficiency.[56]

- ESSA requires that the innovation be brought statewide within three years and ultimately meet the same peer review requirements as traditional assessments.[57] This does not give states much time to design and iterate on innovative approaches to assessing student learning.
- ESSA only allows up to seven states to participate. Innovation for the remaining states would have to wait until the next renewal of the Elementary and Secondary Education Act.

Georgia, Louisiana, New Hampshire, and North Carolina have been approved to participate in the innovative assessment pilot, though most are still focused on grade-level assessments.[58] Louisiana's plan is focused on a combined test for English and social studies, New Hampshire's plan expands on the use of performance assessments,[59] and North Carolina's plan calls for the use of multiple formative assessments, all of which would be aligned to grade-level standards.[60]

Currently, the most promising approach to mitigate the Iceberg Problem comes from Georgia's plan. As part of its plan, participating districts will administer three formative assessments over the school year with items from multiple grade levels. The state will then aggregate those items that relate to grade level standards to determine a state summative score.[61] If successfully executed, Georgia's approved plan holds the most promise for integrating comprehensive measures of learning growth while still generating grade level performance.

> The instructional focus these policies incentivize may be at odds with what may be truly best for each student given their unfinished learning from prior years.

The instructional implications of grade-based assessments and accountability

The federal policies that undergird statewide assessment and accountability systems send an unmistakable signal to middle grade math teachers: focus your instruction on the grade-level standards.

Without clear and common expectations and grade-based reporting every year, achievement gaps and systemic inequities might flourish in hiding. And the most disadvantaged students are the most likely to suffer under these conditions—as happened frequently in the era before standards-based accountability, when disbelief that

disadvantaged students could reach high levels of achievement was an accepted norm among many educators and political leaders. Today's teachers have more clarity into the level of rigor required for students to be on a college and career trajectory so they can better maintain high expectations.

But at the same time, we must acknowledge a real cost to a policy orientation focused on grade-level expectations—that the instructional focus these policies incentivize may be at odds with what may be truly best for each student given his or her unfinished learning from prior years. What may have been intended by some policy makers as the equivalent of an educational "dipstick" to gauge how students are performing may instead be driving an instructional experience that can cause some students to fall further behind.

The unfortunate truth is that millions of students, including the vast majority of students from historically disadvantaged communities, are coming to middle school with unfinished learning from elementary school.[62] This places an immense burden on the middle grade math teacher to not only cover grade-level material, but to also diagnose and fill each student's unique pre-grade gaps that relate to grade-level content all within a single school year—a tall order for even the most talented of teachers. This challenge becomes even more daunting as students progress through middle school and learning gaps continue to accumulate.

Students arriving into middle school multiple years behind grade-level standards need a viable instructional bridge that enables them to catch up and move ahead. This requires a strategic mix of pre- and on-grade skills, often for more than a single school year, given the unfinished learning that has accumulated over time. Today's assessment and accountability policies oriented around annual grade-level proficiency make it far harder to take this kind of instructional approach.

Key Insight #3

Balancing pre-grade level, on-grade level, and post-grade level skills to each student's needs can better support their long-term success.

Evidence is emerging on the relationship between student performance and the level of instructional content.

While ESSA requires an accountability system that focuses on grade-level mastery, districts and schools are free to develop and apply their own parallel accountability systems so long as they continue to meet state and federal requirements.

Most districts generally measure growth in the ways reflected in their state's ESSA plan. But some focus on different growth measures based on NWEA MAP, partly because it includes items from across multiple grade spans. These schools still take the state test and, for purposes of state accountability, their growth is measured in ways aligned with federal policy. But under their own district- or school-based accountability system, they are able to focus on comprehensive learning growth.

Over the last six years, our program has operated in districts and schools with different philosophies around teaching students grade-level material. In schools where

accountability systems imposed at the district or school level are based on growth measures that cross multiple grades, we can tailor a personalized curriculum for each student that includes a mix of pre-grade, on-grade, and post-grade material, depending on their unique starting points. For many students who enter sixth grade multiple years behind, this means spending meaningful time addressing unfinished learning in service of grade-level material. This can also mean not necessarily exposing students to all grade-level standards in a single year, given the time it takes to fill multiple years of unfinished learning.

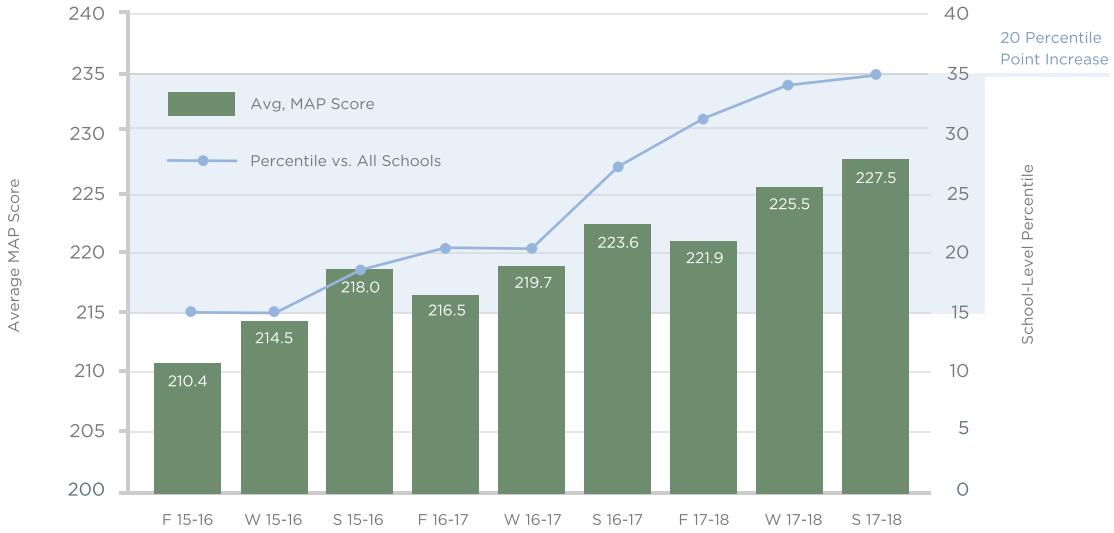
By contrast, in partner schools judged primarily on annual state assessment growth measures, school leaders have often asked us to weight students' personalized curricula more heavily toward grade-level content for the year. This can mean leaving important pre-grade gaps unaddressed.

The impact of accountability policy on student performance in the context of Teach to One: Math was explored by MarGrady Research in a 2019 study funded by the Bill & Melinda Gates Foundation. The study included several correlational analyses that compared the MAP growth of students participating in Teach to One: Math over three years to national norms.

First, the study found that, overall, students served through Teach to One: Math over their three years in middle school grew 20 percentile points (from the 15th percentile to the 35th percentile).[63]

Figure 12

Teach to One: Math Average MAP Score and School-Level Percentile (Consistently-Enrolled Students)

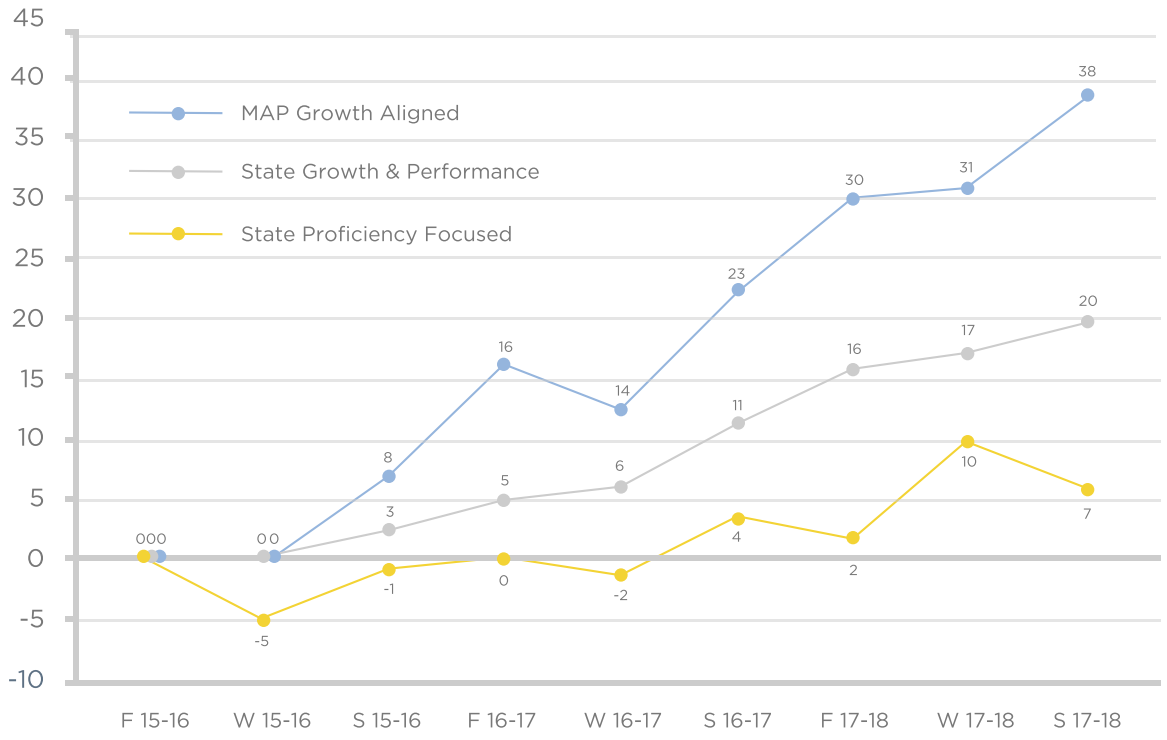


Students	739	713	730	694	718	735	707	709	739

Note: Includes only students who were 6th-graders in 2015–16, 7th graders in 2016–17, 8th-graders in 2017–18, enrolled in the same school for all three years, and had both a fall 2015 and a spring 2018 MAP score. Students are not required to have a test score in every period to be included.

Second, the study found that schools that operated within accountability systems that valued learning growth (as reflected on the MAP) grew 38 percentile points over the three-year period while those focused largely on state proficiency grew 7 points.[64]

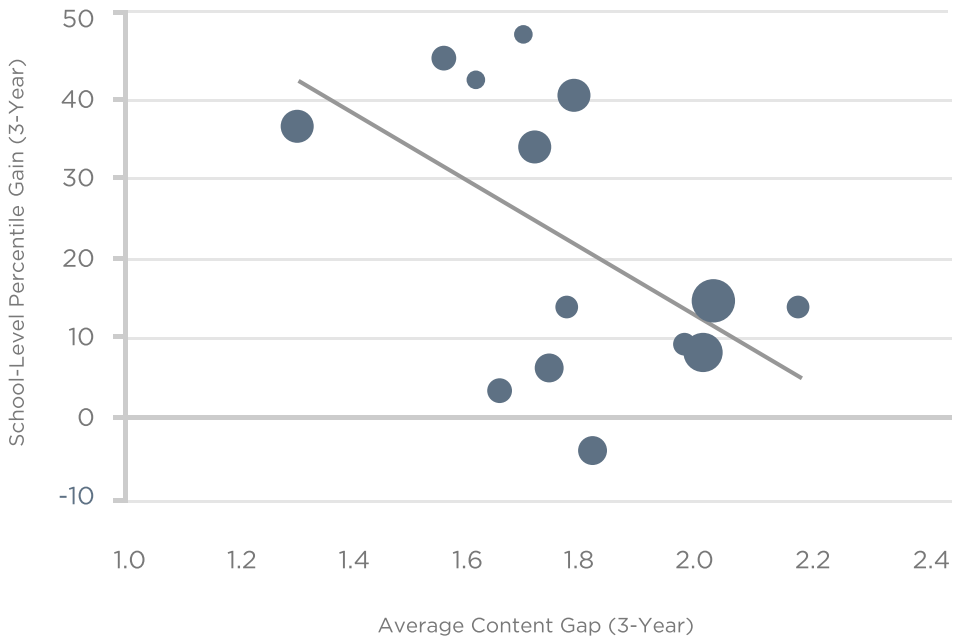
Figure 13
Change in School-Level Percentile by School Category



Note: Figure shows percentile gain for consistently-enrolled students. The MAP Growth Aligned category includes schools 4, 5 and 6. The State Growth & Performance category includes schools 1, 2, 3 and 7. The State Proficiency Focused category includes schools 8–14.

And third, the study also included unique analysis that juxtaposed the learning growth that students made on MAP to the academic level reflected in their daily lessons. To do so, MarGrady included a concept called the "content gap," which it defined as the difference between the incoming performance level of students and the average level of instructional content they received over the school year. For example, a sixth-grade student who enters the school year on a fourth-grade level and receives content that averages at a sixth-grade level would have a content gap of two for that year.

Figure 14
Relationship Between Content Gap and School-Level Percentile Gains (3-Year)



Note: Includes only consistently-enrolled students. Bubble size is proportional to the number of consistently-enrolled students.

In Figure 14, each bubble reflects a school (and the size of bubble corresponds to the size of the school). The x-axis reflects the average three-year content gap for each school (effectively recalculating the annual content gap each year and then averaging the results over three years), and the y-axis reflects the three-year gains students made on MAP. The study found suggestive evidence that schools with a smaller content gap—those where the math content presented better matched students' tested grade level from the beginning of the year—tended to see stronger gains.[65]

As with any school-based program, there are a number of factors—including curriculum implementation, teacher quality, and staffing, and factors external to the school—that can influence outcomes. The MarGrady study was not structured to establish causality. We expect future studies can control for many of these variables in order to more conclusively determine the impact that underlying accountability systems may be having on student acceleration over the longer term.

Additionally, we do not yet have sufficient data to determine the impact that these different progression strategies have on eighth-grade state test performance. Our partner schools to date have asked that we focus student progression strategies either on annual measures of growth (as measured by MAP) or a blend of learning growth and annual

grade-level skills, and have sometimes shifted this focus in different school years and within different student cohorts. This has prevented third-party studies from drawing generalizable conclusions about the impact of Teach to One: Math on state tests.[66]

Comprehensively testing the impact of this approach on eighth-grade state tests would require comparing the impact of students who receive a prioritized mix of pre- and on-grade skills over three years—all in service of eighth-grade performance—to students who receive grade-level content for each of three years, controlling for factors including those listed above. Future studies (both for our work and elsewhere) may wish to focus here.

There is little evidence to suggest that for students who are far below grade level, focusing instruction exclusively on grade-level content is effective.

While evidence of the value of balancing pre- and on-grade skills is still emerging, it should be viewed against the evidence base for what policy currently incentivizes—providing all students with grade-level content.

Many studies have found that too many students in the past were subjected to repetitive, ineffective instruction that did little to prepare them for college and a career.[67] Schools where low expectations and ineffective instruction reigned disproportionately were those that served low-income students, students of color, and other historically disadvantaged students. Students spent years in remedial math courses that got them no closer to graduating. This unacceptable state of affairs has yielded policy and curricular remedies that focus on grade-level instruction and assessment as key drivers for ensuring instructional rigor.

At the same time, we find little evidence in the research to support the notion that in middle grade math, grade-level content for students who have already fallen far behind works without careful strategies to build up key missing skills. This is a difficult question to study, in part because curricula vary between districts and schools, and curricula on paper do not indicate how teachers spend their time and energy under current policies. Even in schools with a grade-level curriculum focus, the best teachers may still be doing a lot to spiral in prerequisite skills that students need.

One way that researchers have explored the effects of giving students content far outside

their current skills is by looking at students taking algebra in eighth grade instead of high school. A policy push in the early 2000s placed many eighth-grade students in algebra who would have otherwise taken a pre-algebra eighth-grade math course. One study found that students enrolled in advanced eighth-grade algebra with low incoming math skills performed about seven grade levels below their peers on NAEP and struggled with questions testing elementary-level understanding.[68] Another study found that low-achieving students pushed into algebra did less well in subsequent math courses through high school, especially in geometry.[69] This is possibly because students never had a chance to build up prerequisite skills that would have helped them later on.

The above gives us reason to believe that in math, low-performing students pushed into grade-level content without appropriate support and attention to prerequisite skills may not be better off in the long run.

Some districts are taking the lead in comprehensively measuring learning growth.

As previously discussed, a number of districts and schools incorporate instruments that measure learning growth across multiple grade levels into their accountability systems, even though their states continue to measure growth through changes in statewide summative performance. First, in Chicago Public Schools (CPS), the district has created a local school accountability system where three out of four academic success metrics are growth-based and the primary assessment to measure growth is the MAP.[70] This system, enacted in 2014, is separate from the state's ESSA tests and accountability systems, but it fulfills CPS's responsibilities under Illinois state law to identify and support struggling schools. CPS's ratings are used to track school performance over time, share data publicly with families and stakeholders, and identify schools in need of intervention. Growth data are also used to inform instructional planning in schools and are used in some grades as a factor in decisions about grade promotion and admissions to selective high schools.[71] In addition, CPS uses accountability ratings to offer struggling schools more standards-based professional development, coaching, and support, showing that a growth focus and a commitment to rigorous standards can complement one another.[72]

More recently, CPS announced that graduates attaining college- and career-ready credentials grew from 31 percent in 2014 to 47 percent in 2018.[73] Average math

test scores, course grades, and passing rates have improved over the past four years among middle school students, especially in schools receiving the most support from the district to understand and implement rigorous math standards.[74] Performance in grades three through eight math improved from 45 percent of students meeting national norms at their grade level in 2013 to 57 percent of students in 2018.

If policymakers want to create the space for schools to better meet the unique needs of their individual students, they must create new policy frameworks that enable and encourage these practices.

Another interesting example of district-level accountability and instructional uses of data is underway in Tulsa, Oklahoma.[75] Tulsa Public Schools developed data dashboards that allow educators to view, sort, and analyze data, including multiple years of academic and non-academic measures. Dashboards like these could form the stepping-stones to setting and understanding multiyear growth goals. Like CPS, Tulsa uses MAP in addition to state tests. The district uses growth as its primary internal success metric and regularly reviews performance measures like proficiency to check that students are progressing to college and career readiness. Tulsa considers its own growth metrics as the most useful measure of student and system performance. The district supports school principals to set their own rigorous and attainable goals for MAP growth at the school level. The district is also planning a large-scale high school redesign process that aims to align assessments, accountability, and internal systems away from grade- and time-based models for student achievement.[76]

Districts such as Chicago Public Schools and Tulsa Public Schools are outliers when it comes to measuring learning growth across multiple grades. In general, districts are more likely to accede to math measures and tests adopted by their states rather than develop parallel systems. If policy makers want to create the space for schools to better meet the unique needs of their individual students, they must create new policy frameworks that enable and encourage these practices.

A focus on true learning growth can also motivate teachers and students to reach high expectations.

Many teachers see students demotivated when they are taught or assessed on skills that

are far beyond their zone of proximal development. As one teacher said, "We want students to feel confident and competent, and to stay excited. If the only test they have is a grade-level test that they bomb every time, then what kind of motivation do they have to keep trying?" On the other hand, when students have opportunities to see learning growth and achieve small victories along the way, math seems more engaging and accessible. "[Students'] feelings about math are tied to their success and achievement in math. As soon as you start giving them opportunities to succeed, and then they see the growth, their feelings change." Teachers with whom we spoke were passionate about the importance of students understanding where they stood relative to long-term goals and having opportunities to meet high expectations. More nuanced growth measures can motivate more students and teachers to successfully meet college- and career-ready expectations.

Accountability models that run parallel to ESSA and emphasize more comprehensive growth metrics that incorporate students' starting points can allow students and teachers to track continuous progress and see rewards for positive change. In a recent national survey, 74 percent of teachers said that measures of student academic growth should be a part of their school evaluations, and 64 percent said student growth measures should be a part of teacher evaluations.[77]

Teachers told us that growth metrics help students see their progress and serve as a motivator. When schools and teachers are judged on certain goals, students are likely to set similar goals for themselves, and they may be more motivated by goals emphasizing ambitious but attainable levels of growth over a seemingly arbitrary or faraway proficiency bar.[78] In schools and districts that balanced pre-grade and on-grade instruction and emphasized learning growth, teachers felt that students were set up to achieve success.

Recommendations

We believe deeply in the impetus behind the shift to standards-based accountability—a commitment to equity, rigor, and transparency conceptualized to ensure that all students have access to the full range of opportunities that achieving college and career readiness affords. Our education system gained significantly from the development of these systems. We now have strong rigor and alignment in academic standards in most states. We now have access to significantly more information about schools' performance, and we are shining a light on the gravity and persistence of gaps in learning among traditionally underserved students. These are substantial accomplishments.

We must candidly acknowledge the trade-offs and costs a policy orientation focused on grade-level expectations creates.

But at the same time, we must candidly acknowledge the trade-offs and costs a policy orientation focused on grade-level expectations creates. For far too many students, these costs are substantial given their unfinished learning from prior years. This is especially

problematic for middle grade mathematics given the cumulative nature of math itself, the fact that students are entering the middle grade years with unfinished learning from elementary school, and that high schools are generally ill-equipped to catch students up.

Rigor, equity, clarity, and transparency remain foundational principles for high-quality accountability systems. These systems also should encourage instructional practices that meet students' needs and effectively advance them to college and career readiness. The current, standards-centric accountability system—under which middle grade math instruction falls—is oriented around measuring performance annually. While this can create the superficial appearance of equity, this approach creates instructional incentives that prioritize covering grade-level material over meeting students in their zone of proximal development. For students with unfinished learning from prior years, this can actually keep them from ultimately achieving college and career readiness.

In the current political climate, there appears to be little appetite at the federal level to revisit the core tenets of ESSA anytime soon. As a result, states and districts that want to explore new pathways for assessment and accountability may find their aspirations are limited by a) current federal policy that requires an accountability system based on the statewide measurement of performance against grade-level expectations and b) practical limitations in the frequency and length of the assessments themselves.

Nonetheless, policy makers can still take action in response to the challenges we describe.

Recommendation #1: Measure Comprehensive Learning Growth

Statewide assessments are almost exclusively focused on grade-level standards and thus any learning gains made on pre- or post-grade skills are unlikely to be reflected on these instruments. While future assessments may be able to effectively measure both learning growth and grade-level proficiency, state and districts looking to better capture learning growth now may wish to complement current statewide assessments with the adaptive assessments that incorporate standards from multiple grade levels. The MAP assessment (NWEA), i-Ready (Curriculum Associates), and STAR Math (Renaissance Learning) are three products schools use to achieve these goals. Some of these instruments have studies that help users to better understand each student's current performance relative to national norms and to predict performance on statewide assessments.

Recommendation #2: Modify Accountability Systems

ESSA places many restrictions (both explicit and practical) on the ability of state accountability systems to incorporate measures of comprehensive learning growth and

incentivize more strategic approaches to catching students up. Nonetheless, several strategies that states may want to consider when addressing the challenges outlined above include:

- **Incorporate multiyear growth metrics in school ratings systems.** While any measure in a state accountability system must be calculated annually, that need not preclude multiyear measures. States may choose to compare changes in proficiency rates over multiple school years as part of their growth calculations within their accountability systems. For example, schools could be rewarded for the degree to which student proficiency levels changed between the start of sixth grade and the end of eighth grade, thereby encouraging schools to adopt a long-term view of performance to strategically catch students up and move them ahead.

- **Weight key transition points more heavily.** Some states, such as Arizona, place extra weight on critical benchmarks, such as third-grade reading. A state could, for example, weigh the performance of fifth- and/or eighth-grade students on their respective summative assessments more heavily than other middle grade levels. This approach would emphasize the importance of high school math readiness and also allow schools to strategically take a longer-term view.

- **Explore the use of adaptive assessments that span multiple grades into measures of school quality and student success.** ESSA restricts academic measures to grade-level tests, but states could explore options under ESSA's "fifth indicator" to use adaptive assessments in order to capture learning growth beyond a single grade level. The fifth indicator is broadly defined by ESSA as a measure of school quality and student success and is a required piece of school accountability determinations. Many states have chosen chronic absenteeism or other non-academic indicators for this category, but nontraditional academic measures are an additional option. For example, in states such as Nebraska, where all students take adaptive tests that span multiple grades, changes in school growth percentiles on the MAP could be a component of a fifth indicator score.

- **Create supplemental growth indicators.** Several states dissatisfied with the structures of NLCB and ESSA have created their own parallel or supplementary systems governed only by state law. These systems cannot overrule what is required under federal law, but they can provide additional information to help inform administrators, teachers, and parents about other aspects of school performance. If these systems were targeted at innovative or alternative schools

not served well under ESSA, they could play a valuable role in defining and encouraging a different vision of success. Some states have already begun to incorporate supplemental indicators in other domains (e.g., social-emotional learning and English Learners).

States may be able to implement many of these recommendations without federal approval. Other recommendations may need to be incorporated into future statewide ESSA plans (which can be revised at any time) and approved at the federal level before they can be fully implemented.[79]

Districts have more flexibility to serve as laboratories for accountability innovation, with the backstop of state accountability systems providing an additional layer of assurance and transparency. They can take many of the steps described above for states even further, given their governance responsibilities. The cases of Chicago and Tulsa show what is possible when districts offer their own visions for accountability and align their support systems toward those goals. Districts such as Lindsay Unified in California have taken this even further, building competency-based learning progressions that include standards from across multiple grade levels into each student's instructional experience.[80]

Recommendation #3: Launch Math Innovation Zones

The State of Texas launched a statewide effort to incubate high-quality blended learning programs aligned with state standards to "dramatically impact the life trajectory of students with a focus on eighth-grade algebra."[81] Participating districts must use an independent and state-approved growth metric three times a year (fall, winter, spring) as one component of Math Innovation Zones is to measure the effectiveness of the personalized learning program. These results on learning growth are monitored by the state on a quarterly basis and districts who have been selected for this program are eligible to receive a designation on the state accountability system.[82] Other states could introduce similar, math-specific, innovation initiatives.

Recommendation #4: Make Available High-Quality Instructional Supports and Strategies That Account for Unfinished Learning

Policy changes around accountability systems are necessary but not sufficient to improve outcomes for students. Change happens at the classroom level, when educators have the supports, incentives, and tools they need to guide and accelerate student learning. This paper's purpose is not to prescribe one curricular approach or learning model. There are support systems and resources that could improve teaching and learning on a broader scale that states and districts could offer right away:

- **Implement strategies to build skills in support of grade-level standards.** One way to help reconcile the tension between grade-level expectations and students' starting points is by ensuring that students are always either accessing grade-level content or advancing toward it by focusing on the unfinished learning required to meet grade-level standards. This would mean that pre-grade learning is not aimless or based on low expectations—it should enable acceleration toward meeting a particular grade-level skill. This strategy can help students access and master grade-level content even if they begin with significant gaps.

- **Find more time.** The rigor and depth reflected in today's college- and career-ready standards generally require a full school year of instruction. Any time spent helping students address unfinished learning is less time available to focus on grade-level material. If districts and schools expect all students to successfully learn grade-level material, they must find and make effective use of additional instructional time (e.g., double blocks, after-school programs, summer school) and ensure teachers understand and implement strategies that promote efficient learning to successfully accomplish both objectives.

- **Promote job-embedded professional learning.** Teachers will benefit from training in the content standards from multiple grade levels. Such trainings can focus on the subject matter itself, the relevant learning progressions, and common misconceptions. Simply offering this form of professional development will signal to teachers a deeper awareness of the inherent challenges they face.

- **Provide high-quality instructional materials anchored in grade level but inclusive of pre-grade predecessors.** High-quality instructional materials are critical to effective instruction, and, first and foremost, schools and districts should invest in quality and consistency across grades. However, even districts that have purchased high-quality, grade-aligned material can end up leaving teachers on their own to search for high-quality, off-grade content. A more systematized way to source and leverage instructional material that includes content from multiple grade levels (either through models such as Teach to One: Math or through other innovative approaches) will both drive quality and save teachers time.

- **Ensure time for collaboration.** Teachers will benefit from planning time with those who teach students enrolled in other grade levels, where they can explore in more detail strategies and approaches for teaching content that may not be explicitly included in grade-level standards but are necessary predecessor skills for student mastery of grade-level content.

Recommendation #5: Advance a Future Vision for Assessment and Accountability

Given the varied incoming mathematical skill levels that students begin the year with and the limited instructional time teachers have with students, one can imagine a future assessment and accountability system that is both aligned with the foundational principles of rigor, equity, clarity, and transparency and also with what is best instructionally for each student. Such a system could include

- **Adaptive Assessments.** Assessments that span multiple grades and can adapt to student responses in order to give credit for learning gains that are outside of students' enrolled grade level.

- **Competency-based Assessments.** School success could be determined based on the ability for students to reliably demonstrate mastery of key academic skills and concepts throughout the school year, as opposed to a single statewide summative assessment.

- **Curriculum-embedded Assessments.** Innovative learning models and next-generation curriculum offerings can ultimately embed reliable assessments that, with appropriate controls and comparability capabilities, could one day replace annual summative assessment with far greater levels of precision.

No doubt other possible hallmarks of what a future assessment and accountability system might look like will emerge, and healthy debates will be had along the way about what is best. It will be incumbent upon policy makers, innovators, advocates, academicians, researchers, parents, and others to begin to design what these solutions might look like and to ensure there is space within the current policy landscape for them to evolve.

Conclusion

Our aspiration in publishing this paper is to catalyze robust conversations and debates about how accountability systems can support strong instructional practices and advance equity for the most disadvantaged student groups and where our current systems fall short of their goals. This is not just one conversation or a problem with a single answer—these issues require discussion at the local level, in schools and districts, across states, and nationally among educators, policy makers, advocates, and instructional support organizations. There are important issues and legitimate competing points of view on the best paths forward for instruction, accountability, and equity. Ignoring these issues out of fear of disrupting the status quo is unacceptable.

Policy makers cannot simply ignore the fact that math learning is cumulative.

Policy makers cannot simply ignore the fact that math learning is cumulative—the skills and knowledge gained in one year provide an essential foundation for accessing and mastering skills in multiple grade levels down the line. When students do not fully master foundational skills, this unfinished learning accumulates over time, making it increasingly challenging for the student to catch up. The instructional incentives and pressure to deliver exclusively grade-level content created by the predominant assessment and

accountability structures is fundamentally at odds with the needs we see in the high proportions of students who enter middle school multiple grade levels behind.

This is not about having low expectations of students; it is a fact inherent in the sequential nature of mathematics itself and the need to honestly confront the unfinished learning gaps that have accumulated throughout the elementary school years in service of students' longer-term success.

In our work with partner schools, we have found that with the right set of supports and structures, students who enter middle school below grade-level standards can catch up and move ahead. For some, the path to getting there requires addressing unfinished learning that may take more than a single year. But, in the end, readiness for high school is achievable.

However, today's assessment and accountability policies fixated solely on grade-level content make that journey more difficult. Policies push teachers to focus on grade-level material to the exclusion of individual growth, which may be causing some of the most disadvantaged students to fall even further behind. The policies may also be preventing advanced learners from progressing to skills beyond their assigned grade level, even when they have the ability to do so.[83] Even the greater emphasis on growth metrics emerging under ESSA cannot fully address the challenge because those metrics are still based on grade-level assessments. The resulting blind spot in accountability threatens the equity and transparency these systems were designed to protect.

Those who disagree with our position may argue that effective teaching practices can enable students to learn both grade-level skills and the applicable pre-grade learning gaps. While this may be true in other subjects or in lower grade spans, we see little evidence in middle grade math to suggest that such an expectation should serve as the basis for national policy. A teacher delivering a lesson on Quadratic Equations cannot simply provide a side lesson on Exponents for those students who happen to have that learning gap. Each topic requires dedicated time and attention in order for them to be properly covered. Simply expecting all teachers to achieve what few (if any) teachers can reasonably accomplish is not a recipe for systematic success.

We similarly reject the notion that teaching students pre-grade content is somehow deemed not rigorous. From an instructional perspective, teaching pre-grade skills does not equate to teaching procedural approaches only. In fact, access to grade-level content often requires students to apply deep, conceptual understanding of foundational skills that require rigorous instructional practices to develop.

From a content perspective, equating grade-level expectations with academic rigor also ignores the fact that students have different starting points. One eighth grader may not be able to successfully access a lesson in Systems of Equations because they lack the predecessor knowledge, while another may

find that skill not rigorous enough based on what they already know. Equating grade-level instruction with "rigorous instruction" mistakenly assumes all students within a grade level have the same starting point.

We are calling on the field—practitioners, policy makers, philanthropists, and other stakeholders—to come together to honestly and productively grapple with the tensions in our current system, to understand where policies may create barriers to our most ambitious goals for students, and to create a better educational system for all students.

We hope to bring our perspective together with leaders and practitioners of all levels, both those who may agree with us and those who may not, for a productive dialogue focused on our shared goal of enabling all students to achieve college and career readiness and success. We also hope our perspective helps to spur the development of more innovative learning models that seek to effectively balance this tension.

We are also encouraging states and districts to create the space for developing what one day might be a new accountability system that preserves high expectations, transparency, and equity while also incorporating more robust measures of learning growth that incentivize more effective instructional practices. Our recommendations include a set of near-term recommendations for like-minded states and districts to consider within the current framework of ESSA in order to advance these efforts.

We fully recognize that policy is not the sole source of the problem and that policy changes do not provide a complete solution. There is work to be done to ensure that schools and teachers have the instructional resources, skills, and knowledge to provide the nuanced and effective math instruction students need to rapidly advance and excel. We believe this is an area ripe for both additional research and instructional innovation.

Appendix I: Methods

In this paper, New Classrooms sought to understand whether and how standards, assessments, and accountability systems ensure that students develop the necessary skills to demonstrate learning in math, enable teachers and schools to provide high-quality math instruction that properly serves students at all skill levels, and enable emerging efforts in personalized and mastery-based learning.

To answer these questions, New Classrooms drew on internal expertise and external sources and used a variety of quantitative and qualitative research methods:

Literature review:

- History, evolution, and impact of standards-based accountability policies
- Common features of state policies on math standards, assessment, and accountability systems
- Landscape of standards and assessments in states
- District- and state-level accountability systems
- Secondary mathematics curriculum, instructional methods, and student learning progressions
- Assessment methods, design, and metrics, including measures of performance and growth

Data analysis:

- Reviewed publicly available NAEP, state assessment, and longitudinal data on student math outcomes and progress toward college and career readiness, particularly at key transition points like eighth grade
- Analysis of New Classroom's NWEA MAP and state assessment data from Teach to One classrooms across multiple years and sites

Expert interviews:

- Internal interviews with New Classrooms leaders on their experiences with partner schools and districts, and emerging findings from Teach to One sites

- More than 20 external interviews with experts on standards, assessment, curriculum, and math instruction to understand various perspectives and experiences on the impact of standards-based reform and the role it should play in math instruction, assessment, and accountability

    - These included policy experts and leaders at the federal level and in various states, leading researchers in assessments and accountability, curriculum developers, competency-based learning experts and advocates, including many people whose views and approaches on math instruction and policy differ from those of New Classrooms

Teacher focus groups:

- Convened six focus groups in multiple cities consisting of middle school math teachers to gather perspectives on their own experiences with math instruction and how the current system of standards and assessments influences their day-to-day work

- Supplemented with interviews of principals and/or instructional leaders at those school sites

- More detail on teacher focus groups below

Data analysis for Figures 12, 13, and 14 were based on four years of NWEA MAP results among Teach to One: Math students, from 2014–2015 through 2017–2018. Only students with a fall and spring score in each year were included. New Classrooms used the NWEA MAP to PARCC linking study to create performance categories.[84] Among the summative test linking studies available from NWEA, we selected PARCC based on its evidence of college and career rigor as an appropriate benchmark.[85] Not all students in this data set live in states where PARCC is administered. The process to calculate performance relative to assigned grade level is as follows:

- The PARCC linking study establishes a MAP RIT score that predicts a 30 percent chance of proficiency on each grade's end-of-year exam. Those scores were used as the cut point for "effective grade-level bands" in the MAP vertical scale (e.g., the grade-level band for sixth-grade math based on this study was 224 to 233).

- Each student's Fall RIT score was assigned to one of these effective grade-level bands.

- We calculated the difference between each student's effective grade performance and their assigned grade. For instance, a student assigned to seventh grade with

performance in the fifth-grade band would have a difference of -2. A student assigned to fifth grade with performance in the sixth-grade band would have a difference of +1.

- These differences determined the achievement bands relative to grade level:
    - four or more years below grade-level standards;
    - two and three years below grade-level standards;
    - one year below grade-level standards, at grade-level standards, and one year above grade-level standards; and
    - two or more years above grade-level standards

We compared this relative grade level to a 80 percent probability of achieving PARCC proficiency in Figure 7. This variable is calculated using a Spring MAP to PARCC Linking Study. There is a minimum Spring RIT score identified for each grade level for a 80 percent chance of proficiency on that grade's PARCC. We subtract the student's Fall RIT score from this RIT score to get the growth needed over the course of the school year to reach that RIT score. We divide that amount of growth by the fall-to-spring national average growth for that grade, which results in the "growth needed" variable.

Appendix II: Focus Groups

New Classrooms contracted with Bellwether Education Partners to conduct five focus groups with middle school math teachers from five schools in different parts of the country: three located in New York City, one in the Atlanta area, and one in Chicago. Focus groups aimed to incorporate educator voices and insights into the paper to support, challenge, or supplement other research findings. Schools were recruited from New Classrooms' and Bellwether's network and partners. School leaders agreed to participate in the project and then extended invitations to their teachers. While these sessions were conducted rigorously, they do not provide a representative sample of educators.

Of the participating schools, two were traditional public schools and three were public charter schools. Two participated in New Classrooms' Teach to One: Math model; three used other models of math instruction. All schools served students who were predominantly from low-income households, but other student demographic trends varied. Among the Teach to One focus groups, most teachers also had experience working in more traditional math teaching environments.

Focus groups were conducted in groups of five to ten teachers, without administrators present. Focus group facilitators used a common, preplanned protocol for each focus group, but the content of each conversation varied somewhat on the school context and participants' areas of interest, and facilitators asked follow-up questions as necessary. The focus groups did not ask for teachers to reach consensus. The goal of the focus groups was to understand teachers' experiences with various models of math instruction and explore the impact of assessment and accountability systems in their work.

The protocol for the focus groups included three areas of inquiry: instructional approaches, assessment and accountability impact, and suggestions for improvement. Example questions below do not fully represent the scope of the focus group protocol.

Instructional approaches

- How do teachers measure students' knowledge and skills at the beginning of the school year and track their progress?
- How do teachers address students' variation in knowledge and skills in their instructional approach?
- How do teachers balance instruction centered on grade-level standards with the

need to differentiate instruction based on students' individual needs?

Assessment and accountability impact

- How do teachers use student assessments, and what do they use them for?
- What impact, if any, do accountability structures linked to student assessment results have on teachers' instructional approaches?
- How do assessments affect how teachers think about their own effectiveness and their students' attitude toward math instruction?

Suggestions for improvement

- How do teachers think their schools could improve the way students' math knowledge and progress are measured?
- What policy changes do teachers think would best support student learning across various skill levels?

All focus group participants received a consent form notifying them that participation was voluntary and informing them of the purpose and content of the discussion. They were notified that any remarks would not be reported back to their schools or linked with their names or schools in this publication. Teachers received a small gift in exchange for participation. New Classrooms ensured that findings resulting from the focus groups removed any potential personally identifying information from quotes included herein.

Appendix III: Evolution of National Policy

Figure 15 — Policy Timeline

Timeline of Standards-Based Reform Milestones

April 1983	The National Commission on Excellence in Education publishes "A Nation at Risk: The Imperative for Educational Reform," a report warning of a "rising tide of mediocrity" in America's educational system.
October 1994	President Bill Clinton signs the Improving America's Schools Act (IASA) into law, requiring every state to adopt academic standards for students in reading and math and regularly administer aligned assessments in certain grade spans.
January 2002	President George W. Bush signs the No Child Left Behind Act (NCLB) into law, requiring states to test students in reading and math annually in grades 3–8 and once in high school, and tying those test results to specific consequences for schools and districts.
December 2008	CCSSO, NGA, and Achieve publish "Benchmarking for Success: Ensuring US Students Receive a World-Class Education," a report encouraging states to adopt common, internationally benchmarked standards in ELA and math to ensure that students have the knowledge and skills to be globally competitive.
May 2009	CCSSO and NGA begin developing the Common Core State Standards.
July 2009	Obama administration announces that states can compete for $4.35 billion in Race to the Top grants to pursue policies like adopting college- and career-ready standards, recruiting and retaining effective teachers and principals, and implementing school turnaround efforts.
June 2010	CCSSO and NGA release the final version of the Common Core State Standards.
September 2010	Obama administration awards roughly $330 million in Race to the Top funding to the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC) to develop assessments aligned to the Common Core State Standards.
September 2011	Obama administration announces that it will grant states waivers from the most burdensome provisions of NCLB in exchange for pursuing certain policies on standards, accountability, and evaluation systems for teachers and school leaders.

continued

December 2015	President Barack Obama signs the Every Student Succeeds Act (ESSA) into law, granting states greater flexibility in how they assess students, design accountability systems, and identify and support schools in need of improvement.
May 2017	16 states and the District of Columbia submit ESSA plans.
September 2017	Remaining 34 states submit ESSA plans.

For decades, standards-based reform and accountability have helped advance equity, transparency, and rigor in America's educational system. In 1983, the National Commission on Excellence in Education, formed by then–Secretary of Education Terrell Bell, published its seminal report, "A Nation at Risk: The Imperative for Educational Reform." The report warned that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people."[86] This report marked the beginning of a bipartisan movement to improve students' academic outcomes via state, federal, and local education policies. Throughout its history, proponents of standards-based reform have emphasized accountability for adults at various levels of education, high academic standards for what skills students should attain by graduation, and particular attention to the needs of low-income students, students of color, and students in persistently low-performing schools.

The first major federal law in this new era of educational reform was the 1994 Improving America's Schools Act (IASA),[87] which reauthorized the Elementary and Secondary Education Act (ESEA)—first enacted in 1965.[88] Among other things, this legislation required that every state adopt academic standards for students in reading and mathematics, and regularly administer assessments aligned to those standards "at some time" in grades three through five, grades six through nine, and grades ten through twelve.[89] However, federal enforcement was weak, and by January 2001, only eleven states were in compliance with IASA's assessment provisions.[90]

The federal role in accountability was expanded further in 2001, when Congress passed the No Child Left Behind Act (NCLB) with strong bipartisan support under President George W. Bush. This legislation put in place the building blocks for modern accountability systems by requiring that states adopt "challenging academic content standards," test students with standards-aligned assessments annually in grades three through eight and once in high school, and link low proficiency rates on those assessments to specific consequences for schools and districts—including state intervention.[91]

NCLB also served an important role in shifting the focus of education policy toward equity. Accountability consequences and interventions were based on not just overall student performance but also the performance of particular subgroups, including economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency. The law also required states to report assessment results in every tested grade and subject, and disaggregate results by these subgroups.[92] For the first time, schools and districts faced meaningful consequences for failing to properly serve all student groups, and families, advocates, and educators had access to consistent data showing the extent and persistence of achievement gaps.

While NCLB helped advance reform in some ways, over time it became clear that its goals were increasingly unrealistic and potentially counterproductive for schools. For example, NCLB required that all students reach the proficient level on state assessments by 2014 (which no state succeeded in meeting).[93] The law's limited requirements on standards also meant that states established vastly different expectations for their students, leading to a patchwork of high- and low-quality standards across the country.[94]

Due to the inconsistent quality and content of academic standards under NCLB, states began to mobilize themselves in response to the challenge of low-quality and inconsistent academic standards. In 2008, the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), and Achieve released "Benchmarking for Success: Ensuring US Students Receive a World-Class Education," a report that recommended states "upgrade state standards by adopting a common core of internationally benchmarked standards in math and language arts for grades K–12 to ensure that students are equipped with the necessary knowledge and skills to be globally competitive."[95] Low-quality standards resulted in countless high school graduates who were not qualified or prepared to advance to postsecondary opportunities. They especially hurt students in low-performing schools, who were more likely to lack access to course content that could support them in college readiness. To address this challenge, state leaders developed a common set of standards that would eventually become known as the Common Core State Standards (CCSS).

The following year, development of CCSS began, incorporating input from state leaders, educators, nonprofits, and content experts, as well as feedback from the general public. Multiple organizations, including ACT and the College Board, released their own versions of college- and career-readiness standards. These influenced the development of CCSS, as did content area standards from groups like the National Assessment Governing Board and the National Council of Teachers of Mathematics, as well as input from states considered to have high-quality standards, including Massachusetts and California. Information from international bodies, such as the Trends International Mathematics and

Science Study (TIMSS) and the Programme for International Student Assessment (PISA), also helped ensure that CCSS set a sufficiently high bar. Additionally, a validation committee, whose members were appointed by a group of governors and chief state school officers, was created to review CCSS's evidence base and development process.[96]

In 2009, while CCSS were being developed, the Obama administration announced that states could compete for $4.35 billion in Race to the Top grants to pursue policies like adopting college- and career-ready standards, recruiting and retaining effective teachers and principals, and implementing school turnaround efforts.[97] When NGA and CCSSO released the final draft of CCSS in 2010,[98] the Race to the Top program played an important role in encouraging adoption of the standards. At their peak, CCSS had been adopted by 46 states.[99]

However, political backlash from both the right and left caused many states to rethink their adoption of CCSS, with some either revising the standards or replacing them with state-developed standards. Conservative opponents of CCSS, primarily associated with the rise of the Tea Party, objected to the federal government's involvement in promoting the standards and shared tests through Race to the Top. Teachers' unions have also expressed opposition, though this is less focused on the standards themselves than on implementation efforts. They cited a lack of adequate time, training, financial and instructional resources, and other support from states and districts to properly implement CCSS in classrooms, and on the use of test results for teacher evaluations.[100]

Both the impact of CCSS and the resulting backlash can be seen in ESSA. Similar to NCLB, ESSA requires states to adopt "challenging academic content standards" and "aligned academic achievement standards" that include at least three level of achievement.[101] In addition, the law requires states to demonstrate that these standards are "aligned with entrance requirements for credit-bearing coursework in the system of public higher education in the state and relevant state career and technical education standards."[102] However, in response to the policies of the Obama administration, ESSA also includes a provision prohibiting the US Secretary of Education from mandating, directing, controlling, coercing, or exercising any direction or supervision over states' standards.[103]

Today, 33 states have maintained their adoption of CCSS (though Minnesota only adopted the ELA standards), and 13 states have adopted standards generally similar to Common Core, while only four states (Alaska, Nebraska, Texas, and Virginia) never adopted Common Core.[104]

According to a recent report from the Thomas B. Fordham Institute,[105] most states not using CCSS have opted for weaker math standards. For example, some states lack fully coherent middle school progressions that make the appropriate connections between

interrelated standards and topics. CCSS has still encouraged some positive trends among those states. The report also identified four "positive trends" in states' math standards, which the authors attribute partly to the continued influence of CCSS. The trends are:

1. A stronger focus on arithmetic in grades K–5;
2. More coherent treatment of proportionality and linearity in middle school, including rates and ratios, slope, and linear relationships and functions;
3. An appropriate balance between conceptual understanding, procedural fluency, and application; and
4. Better organization and teacher supports, including focused introductions for individual grade levels and courses, mathematically coherent organizational approaches that highlight the connections between standards, and helpful ancillary materials.

Along with the shift to common standards, states also began moving toward common assessments. In 2010, again under the Race to the Top program, the Obama administration awarded grants to two consortia of states—the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC)—to develop assessments aligned to CCSS. By 2011, 45 states had joined one or both consortia.[106]

However, the same political backlash that plagued CCSS also caused many states to leave the testing consortia, instead opting to partner with different assessment vendors or create their own assessments. This pushback focused acutely on the perception of "high stakes" tests out of the individual states' control tied to consequences for students, teachers, and schools. As of spring 2018, only 18 states were still administering assessments affiliated with these consortia in some way, and several have announced plans to create their own tests in the next few years.

As states were developing new standards and assessments—aided by the federal Race to the Top program—states began advocating strongly for relief from some of the key requirements of NCLB, particularly consequences for schools not meeting Adequate Yearly Progress (AYP) goals.[107] Congress failed to enact new legislation, and in 2011 the Obama administration began issuing waivers from the law. These waivers granted states flexibility from some of NCLB's requirements—including AYP and the 2014 proficiency goal—in exchange for implementing certain policies, such as the adoption of college- and career-ready standards and teacher evaluation systems based in part on student achievement.[108] Forty-three states, the District of Columbia, and Puerto Rico had their waiver requests approved.[109]

In 2015, Congress finally reauthorized ESEA, and also passed the Every Student Succeeds Act (ESSA) with broad bipartisan support.[110] The law maintains key provisions of NCLB while granting states significantly more flexibility and authority around testing, assessments, and accountability.

- Testing: ESSA preserves many of NCLB's testing provisions, including testing requirements for reading, math, and science, as well as the disaggregated reporting of test results among student subgroups. However, ESSA also clarifies that states may use computer adaptive tests (CAT) and allows tests to "measure academic proficiency and growth using items above or below the student's grade level,"[111] in addition to academic performance at grade level.

- Assessments: Under ESSA, states are now allowed to administer multiple interim tests that add up to a final score rather than using one summative test,[112] and states may apply for innovative assessment waivers for additional flexibility.[113]

- Accountability: Like NCLB, ESSA requires that state assessment results serve as components in school accountability systems. These systems must set goals for increasing the share of students who meet state standards in reading and mathematics, accelerating progress of underperforming subgroups, and improving graduation rates.[114] ESSA requires states to identify their lowest-performing schools for varying levels of support and intervention[115] but provides them flexibility in the specifics of their identification system. While the law requires that states use certain types of indicators, it largely leaves it up to states to decide how each indicator is weighted.[116]

Sixteen states and the District of Columbia submitted their plans for accountability under ESSA to the US Department of Education (ED) for review in May 2017, while the remaining 34 states submitted their plans in September of the same year. The new state accountability plans took effect starting in the 2017–2018 school year.[117]

[1] Stella Fayer, Alan Lacey, and Audrey Watson, "STEM Occupations: Past, Present, and Future," US Bureau of Labor Statistics, 2017, bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf; Anthony P. Carnevale, Ban Cheah, and Andrew R. Hanson, "The Economic Value of College Majors," Georgetown University Center on Education and the Workforce, 2015, cew-7632.kxcdn.com/wp-content/uploads/The-Economic-Value-of-College-Majors-Full-Report-web-FINAL.pdf.

[2] Emily Richards and Dave Terkanian, "Occupational Employment Projections to 2022," Monthly Labor Review, US Bureau of Labor Statistics, December 2013, doi.org/10.21916/mlr.2013.41.

[3] Throughout this paper, we use the word *skills* to include procedural, conceptual, and applied knowledge as a part of a student's overall mathematical understanding.

[4] While the term *gap* might suggest a binary determination of student proficiency, gaps more likely reflect a partial understanding of a particular skill or concept.

[5] National Center for Education Statistics, "The Nations Report Card: Mathematics," National Assessment of Educational Progress, 2017, nationsreportcard.gov/math_2017/#?grade=4.

[6] Michael Hansen et al., The 2018 Brown Center Report on American Education: How Well are American Students Learning? The Brookings Institution, June 2018, brookings.edu/research/2018-brown-center-report-on-american-education-trends-in-naep-math-reading-and-civics-scores/.

[7] Ibid.

[8] Catherine Gewertz, "Math Scores Slide to a 20-Year Low on ACT," *Education Week*, October 17, 2018, edweek.org/ew/articles/2018/10/17/math-scores-slide-to-a-20-year-low.html?r=1611432497&mkey=8026D2C4-0AD1-11E9-A985-EA9FC819EBCD.

[9] Nick Anderson, "College Admission Test Scores Raise Warning Signs about Math Achievement," *The Washington Post*, October 25, 2018, washingtonpost.com/local/education/college-admission-test-scores-raise-warning-signs-about-math-achievement/2018/10/24/ab37ba0a-d7a5-11e8-83a2-d1c3da28d6b6_story.html?utm_term=.51a16adcad10.

[10] Common Core Standards Initiative, "Key Shifts in Mathematics," corestandards.org/other-resources/key-shifts-in-mathematics/. Accessed February 7, 2019.

[11] Resources include: "The Coherence Map," achievethecore.org/coherence-map/. "Principles and Standards for School Mathematics, nctm.org/Standards-and-Positions/Principles-and-Standards/. Progression Documents for the Common Core Math Standards ime.math.arizona.edu/progressions/.

[12] New Classrooms, "Teach to One: Math Major Concepts Map," newclassrooms.org/wp-content/uploads/2016/11/NewClassrooms_TeachtoOneMath_Concept-Map.pdf?pdf=conceptmap.

[13] William H. Schmidt and Curtis C. McKnight, *Inequality for All: The Challenge of Unequal Opportunity in American Schools* (New York: Teachers College Press, 2012).

[14] Ibid.

[15] Yeow Meng Thum and Carl H. Hauser, "2015 MAP Norms for Student and School Achievement Status and Growth," NWEA, November 6, 2015, solutions.nwea.org/research/2015-map-norms-for-student-and-school-achievement-status-and-growth.

[16] NWEA, "Linking Studies," nwea.org/resource/type/linking-studies/. Accessed November 1, 2018.

[17] Based on four years of data across Teach to One sites (2014–2018). N = 9,037.

[18] NWEA, "Linking Studies," nwea.org/resource/type/linking-studies/. Accessed November 1, 2018.

[19] J. Margolis, "Three-Year MAP Growth at Schools Using Teach to One: Math," MarGrady Research, 2019, margrady.com/tto/. Accessed March 5, 2019. Includes summer learning losses between sixth and seventh grades and seventh and eighth grades.

[20] See Methods in Appendix I.

[21] National Center for Education Statistics, "Digest of Education Statistics," nces.ed.gov/programs/digest/d17/tables/dt17_204.10.asp. Accessed April 11, 2019.

[22] Alanna Bjorklund-Young and Jay Plasman, "Reducing the Achievement Gap: An Empirical Analysis of Middle School Math Performance in Six States and Washington, D.C.," The Institute for Educational Policy at Johns Hopkins University, April 2019, edpolicy.education.jhu.edu/wp-content/uploads/2019/04/Achievement-Gap-Policy-Brief.pdf. Accessed April 17, 2019.

[23] Ibid.

[24] The term was coined in the 1930s by psychologist Lev Vygotsky; see more in Seth Chaiklin, "The Zone of Proximal Development in Vygotsky's Analysis of Learning and Instruction," in Vygotsky's *Educational Theory and Practice in Cultural Context*, ed. Alex Kozulin et al. (Cambridge, UK: Cambridge University Press, 2003), people.ucsc.edu/~gwells/Files/Courses_Folder/documents/chaiklin.zpd.pdf.

[25] Jennifer Hammond and Pauline Gibbons, "What Is Scaffolding?" *Teachers' Voices* 8 (2005), ameprc.mq.edu.au/docs/research_reports/teachers_voices/Teachers_voices_8.pdf#page=15.

[26] R. M. Harden and N. Stamper, "What Is a Spiral Curriculum?" *Medical Teacher* 21, no. 2 (1999), faculty.med.virginia.edu/facultyaffairs/files/2016/04/2010-3-23.pdf.

[27] See Appendix I.

[28] Chrys Dougherty and Steve Fleming, "Getting Students on Track to College and Career Readiness: How Many Catch Up from Far Behind?" ACT, November 2012, eric.ed.gov/?id=ED542022.

[29] Ibid.

[30] Jason Zimba, National Council of Teachers of Mathematics 2017 Research Conference, April 7, 2017, San Antonio, Texas.

[31] Amber M. Northern and Michael J. Petrilli, "Credit Recovery: Good Intentions, Poor Execution," The Fordham Institute, edexcellence.net/articles/credit-recovery-good-intentions-poor-execution. Accessed February 7, 2019.

[32] 114th Congress, "S.1177—Every Student Succeeds Act," Section 1111(b).

[33] ESSA text 1111(b)(2)(J); ESSA permits states to exempt eighth-grade students taking advanced math courses from the statewide math assessment used for to eighth-grade students. Students in such courses can instead take the corresponding end-of-course assessment, so long as their results on the end-of-course assessment are included in accountability determinations for that year and they take another more advanced mathematics assessment in high school, and that score is included in accountability determinations for students' high schools.

[34] Exceptions for eighth-grade students in algebra: 1111(b)(2)(C) of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the Every Student Succeeds Act (ESSA).

[35] "A State's Guide to the U.S. Department of Education's Assessment Peer Review Process," US Department of Education, 2018, p. 25, 2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf.

[36] "What's in the Every Student Succeeds Act? Accountability," Education Trust, January 13, 2016, edtrust.org/resource/whats-in-the-every-student-succeeds-act-accountability/.

[37] Susan Lyons, Juan D'Brot, and Erika Landl, "State Systems of Identification and Support under ESSA: A Focus on Designing and Revising Systems of School Identification," CCSSO, November 2017, pp. 4–12, ccsso.org/sites/default/files/2017-12/State%20Systems%20of%20ID%20and%20Support%20-%20Designing%20and%20Revising%20Systems_0.pdf.

[38] Council of Chief State School Officers, "Considerations for Including Growth in ESSA State Accountability Systems," January 2017, p. 4, ccsso.org/sites/default/files/2017-10/CCSSOGrowthInESSAAccountabilitySystems1242017.pdf.

[39] (c)(4)(C)(ii) of Section 1111 of ESSA, congress.gov/bill/114th-congress/senate-bill/1177/text.

[40] Elizabeth Ross et al., "2017 State Teacher Policy Yearbook," NCTQ, 2017, p. 79, nctq.org/dmsView/NCTQ_2017_State_Teacher_Policy_Yearbook.

[41] For more information on growth measures: Juan D'Brot, "Considerations for Including Growth in ESSA State Accountability Systems," CCSSO, January 2017, nciea.org/sites/default/files/pubs-tmp/CCSSO_Growth_Resource.pdf; "Students Can't Wait: Individual Student Growth," Education Trust, edtrust.org/students-cant-wait/individual-student-growth/. Accessed October 20, 2018. "Progress Tables (Value Tables): Another Measure of Student Growth," Virginia Department of Education, January 21, 2015, doe.virginia.gov/boe/committees_standing/accountability/2015/meeting_materials/jan-21_measure_of_student_growth.pdf; Katherine E. Castellano and Andrew D. Ho, "A Practitioner's Guide to Growth Models," Council of Chief State School Officers, February 2013, scholar.harvard.edu/files/andrewho/files/a_pracitioners_guide_to_growth_models.pdf?m=1364611983; "Value-Added Modeling 101: Using Student Test Scores to Help Measure Teaching Effectiveness," RAND Corporation, 2012, rand.org/pubs/corporate_pubs/CP693z4-2012-09.html; "Growth Data: It Matters and It's Complicated," Data Quality Campaign, January 2019, dataqualitycampaign.org/resource/growth-data-it-matters-and-its-complicated/.

[42] Richard J. Patz, "Vertical Scaling in Standards-Based Educational Assessment and Accountability Systems," Council of Chief State School Officers, 2007, pdfs.semanticscholar.org/c1b8/e97bb12142ac842e88da4f825e72ddf9dbc7.pdf.

[43] Madhabi Chatterji, PhD, "A Consumer's Guide to Testing Under Every Student Succeeds Act (ESSA): What Can the Common Core and Other ESSA Assessments Tell Us?" Prepared for the National Education Policy Center, February 21, 2019, nepc.colorado.edu/publication/rd-assessment-guide.

[44] M. Chatterji, personal communication, March 7, 2019.

[45] Julie Woods, "50 State Comparison: States' School Accountability Systems," Education Commission of the States, May 31, 2018, ecs.org/50-state-comparison-states-school-accountability-systems/.

[46] Matt Barnum, "The Growth vs. Proficiency Debate and Why Al Franken Raised a Boring but Critical Issue," The 74 Million, January 18, 2017, the74million.org/article/barnum-the-growth-vs-proficiency-debate-and-why-al-franken-raised-a-boring-but-critical-issue/.

[47] Data Quality Campaign, "Growth Data, It Matters and Its Complicated," dataqualitycampaign.org/resource/growth-data-it-matters-and-its-complicated/. Accessed on January 23, 2019.

[48] Brandon L. Wright and Michael J. Petrilli, "Rating the Ratings: An Analysis of the 51 ESSA Accountability Plans," Thomas B. Fordham Institute, November 2017, edex.s3-us-west-2.amazonaws.com/publication/pdfs/11.15%20-%20Rating%20the%20Ratings%20-%20An%20Analysis%20of%20the%2051%20ESSA%20Accountability%20Plans.pdf.

[49] "Show Me the Data 2017," Data Quality Campaign, 2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2018/01/DQC-Show-Me-the-Data_final.pdf. Accessed October 22, 2018.

[50] "Growth Fact Sheet," Colorado Department of Education, August 2018, cde.state.co.us/accountability/growth-fact-sheet-for-parents-2018.

[51] Smarter Balanced Assessment Consortium, "Race to the Top Assessment Program Application," US Department of Education, 2010.

[52] T. Alpert, personal communication, November 2, 2018.

[53] For example, "Linking the PARCC Assessments to NWEA MAP Growth Tests," NWEA, November 2016, nwea.org/content/uploads/2017/07/PARCC-MAP-Linking-Study_2016.pdf.

[54] Every Student Succeeds Act (ESSA), section 1204, "Innovative Assessment and Accountability Demonstration Authority," 129 STAT. 1885.

[55] Ibid.

[56] US Department of Education Office of Elementary and Secondary Education, "A State's Guide to the U.S. Department of Education's Assessment Peer Review Process," September 24, 2018.

[57] Every Student Succeeds Act (ESSA), section 1204, "Innovative Assessment and Accountability Demonstration Authority," 129 STAT. 1885. Note that USED has permitted an additional two year grace period.

[58] "Innovative Assessment Demonstration Authority (IADA)," U.S. Department of Education, September 28, 2018, ed.gov/admins/lead/account/iada/index.html.

[59] Ibid.

[60] Ibid.

[61] Ibid.

[62] "Explore the Latest NAEP Mathematics Results," The Nation's Report Card, nationsreportcard.gov/math_2017?grade=8. Accessed February 7, 2019.

[63] J. Margolis, "Three-Year MAP Growth at Schools using Teach to One: Math," MarGrady Research, 2019, margrady.com/tto/. Accessed March 5, 2019. Note that a broader group of students, including those not continuously enrolled, showed average three-year gains of 13 percentile points.

[64] Ibid.

[65] Ibid.

[66] D. Ready et al., Final Impact results from the i3 implementation of Teach to One: Math. Columbia University, Consortium for Policy Research in Education, New York, 2018.

[67] Schmidt and McKnight (2012); William Schmidt, "At the Precipice: The Story of Mathematics Education in the United States," *Peabody Journal of Education* 87, no. 1 (February 1, 2012): 133–156, doi.org/10.1080/0161956X.2012.642280; Morgan Polikoff, "The Redundancy of Mathematics Instruction in US Elementary and Middle Schools," *Elementary School Journal* 113, no. 2 (2012): 230–251, journals.uchicago.edu/doi/abs/10.1086/667727.

[68] Tom Loveless, "The Misplaced Math Student: Lost in Eighth Grade Algebra," Brookings Institution, September 2008, brookings.edu/wp-content/uploads/2016/06/0922_education_loveless.pdf.

[69] Charles T. Clotfelder, Helen Ladd, and Jacob Vigdor, "Algebra for Eighth Graders: Evidence on Its Effects from 10 North Carolina Districts," The Calder Center, February 2013, caldercenter.org/sites/default/files/wp87-2.pdf.

[70] "School Quality Rating Policy (SQRP) Handbook," Chicago Public Schools, September 27, 2017, cps.edu/Performance/Documents/SQRPHandbook.pdf.

[71] "2017–18 School/Parent Guide to the Elementary School Promotion Policy," Chicago Public Schools, 2017, cps.edu/SiteCollectionDocuments/ElemPromotionPolicy_English.pdf.

[72] Julia Gwynne and Sarah Cashdollar, "Changes in Math Instruction and Student Outcomes Since the Implementation of the Common Core State Standards in Chicago," University of Chicago, Consortium on School Research, September 2018, consortium.uchicago.edu/sites/default/files/publications/Changes%20in%20Math%20Instruction-Sep2018-Consortium.pdf.

[73] Alexandra Arriaga, "Record Number of CPS Students Graduating with College, Career Credentials," *Chicago Sun-Times*, August 31, 2018, chicago.suntimes.com/news/cps-chicago-public-high-schools-graduating-college-career-credentials-test-scores/.

[74] Gwynne and Cashdollar, "Changes in Math Instruction."

[75] Tara Garcia Mathewson, "Psst! When Teachers Get Useful, Timely Data, They Use It," The Hechinger Report, November 1, 2017, hechingerreport.org/psst-teachers-get-useful-timely-data-use/.

[76] Interview and communications with TPS officials.

[77] "Voices from the Classroom: A Survey of America's Educators," Educators for Excellence, 2018, e4e.org/news/voices-classroom-survey-americas-educators.

[78] Eric M. Anderman et al., "Value-Added Models of Assessment: Implications for Motivation and Accountability," *Journal Educational Psychologist* 45, no. 2 (April 2010): 123–137, doi.org/10.1080/00461521003703045.

[79] Frank Borgan, "Dear Colleague ESEA Letter on State Plan Amendments," November 14, 2018, ed.gov/policy/elsec/leg/essa/dclessaspamendmentprocessltr.pdf.

[80] Chirstina Quattrocchi, "How Lindsay Unified Redesigned Itself from the Ground Up," June 17, 2014, edsurge.com/news/2014-06-17-how-lindsay-unified-redesigned-itself-from-the-ground-up. Accessed February 7, 2019.

[81] Texas Education Agency (TEA), "Math Innovation Zones," tea.texas.gov/Academics/Learning_Support_and_Programs/Math_Innovation_Zones/. Accessed November 1, 2018.

[82] Ibid.

[83] Barshay, "Gifted Classes May Not Help Students Move Ahead Faster," The Hechinger Report, April 15, 2019, hechingerreport.org/gifted-cla sses-may-not-help-talented-st udents-move-ahead-faster/

[84] "Linking the PARCC Assessments to NWEA MAP Tests for Illinois," Northwest Evaluation Association (NWEA), March 2016, nwea.org/content/uploads/2017/06/Illinois_PARCC_Linking_Study_MAR2016.pdf.

[85] Ira Nichols-Barrer et al., "Predictive Validity of MCAS and PARCC: Comparing 10th Grade MCAS Tests to PARCC Integrated Math II, Algebra II, and 10th Grade English Language Arts Tests," Mathematica Policy Research Report, October 5, 2015, mathematica-mpr.com/download-media?MediaItemId=%7B8EF64B21-27FC-48D0-B91F-B8BA3FA95E47%7D.

[86] "A Nation at Risk: The Imperative for Educational Reform: A Report to the Nation and the Secretary of Education, United States Department of Education," National Commission on Excellence in Education, 1983, ed.gov/pubs/NatAtRisk/risk.html.

[87] 103rd Congress, "Improving America's Schools Act of 1994," October 20, 1994, congress.gov/bill/103rd-congress/house-bill/6/text.

[88] 81st Congress, "Elementary and Secondary Education Act of 1965," December 10, 2015, ed.gov/about/offices/list/oii/nonpublic/eseareauth.pdf.

[89] "Summary of the Improving America's Schools Act," *Education Week*, November 9, 1994, edweek.org/ew/articles/1994/11/09/10asacht.h14.html.

[90] Dianne Piche, "Closing the Deal: A Preliminary Report on State Compliance with Final Assessment and Accountability Requirements under the Improving America's Schools Act of 1994," Citizens Commission on Civil Rights, March 1, 2001, files.eric.ed.gov/fulltext/ED460200.pdf.

[91] 107th Congress, "No Child Left Behind Act of 2001," January 8, 2002, congress.gov/bill/107th-congress/house-bill/1/text.

[92] NCLB, 2001.

[93] Alyson Klein, "No Child Left Behind: An Overview," *Education Week*, April 10, 2015, edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html.

[94] David Conley, "The Common Core State Standards: Insight into Their Development and Purpose," Council of Chief State School Officers, June 2014, p. 3, inflexion.org/ccss-development-and-purpose/.

[95] "Ensuring U.S. Students Receive a World-Class Education," National Governors Association, the Council of Chief State School Officers, and Achieve, 2008, p. 28, edweek.org/media/benchmakring%20for%20success%20dec%202008%20final.pdf.

[96] Conley, "The Common Core State Standards," p. 5.

[97] "President Obama, U.S. Secretary of Education Duncan Announce National Competition to Advance School Reform," press release, U.S. Department of Education, July 24, 2009, ed.gov/news/pressreleases/2009/07/07242009.html.

[98] Common Core State Standards Initiative, "About the Standards: Development Process," corestandards.org/about-the-standards/development-process/. Accessed October 23, 2018.

[99] Daniel Hamlin and Paul Peterson, "Have States Maintained High Expectations for Student Performance?" Education Next 18, no. 4 (2018),educationnext.org/have-state s-maintained-high-expectations-student-performance-analysis-2017-proficiency-standards/.

[100] Andrew Ujifisa, "Resistance to the Common Core Mounts," Education Week, April 21, 2014, edweek.org/ew/articles/2014/04/23/29cc-backlash.h33.html.

[101] ESSA text—ESEA 1111(b)(1).

[102] ESSA text—ESEA 1111(b)(1).

[103] ESSA text—ESEA 1111(b)(1)(G)(ii).

[104] State websites.

[105] Solomon Friedberg et al., "The State of the Standards Post-Common Core," Thomas Fordham Institute, August 2018, edex.s3-us-west-2.amazonaws.com/publication/pdfs/(08.22)%20The%20State%20of%20State%20Standards%20Post-Common%20Core.pdf.

[106] Ashley Jochim and Patrick McGuinn, "The Politics of the Common Core Assessments," *Education Next* 16, no. 4 (2016), educationnext.org/the-politics-of-common-core-assessments-parcc-smarter-balanced/.

[107] Alex Johnson, "Majority of States Line Up to Ditch No Child Left Behind," NBC News, September 30, 2011, nbcnews.com/id/44693695/ns/us_news-education_nation/t/majority-states-lining-ditch-no-child-left-behind/#.W1XUfapKjIU.

[108] Michele McNeil and Alyson Klein, "Obama Offers Waivers from Key Provisions of No Child Left Behind," *Education Week*, November 27, 2011, edweek.org/ew/articles/2011/09/28/05waiver_ep.h31.html.

[109] "ESEA Flexibility," Laws & Guidance/Elementary & Secondary Education, US Department of Education, ed.gov/policy/elsec/guid/esea-flexibility/index.html. Accessed October 23, 2018.

[110] "White House Report: Every Student Succeeds Act," The White House Office of the Press Secretary, December 10, 2018, obamawhitehouse.archives.gov/the-press-office/2015/12/10/white-house-report-every-student-succeeds-act.

[111] ESSA text—ESEA 1111(b)(2)(J); ESSA permits states to exempt eighth-grade students taking advanced math courses from the statewide math assessment used for to eighth-grade students. Students in such courses can instead take the corresponding end-of-course assessment, so long as their results on the end-of-course assessment are included in accountability determinations for that year, and they take another more advanced mathematics assessment in high school and that score is included in accountability determinations for students' high schools.

[112] "FAQ: The Every Student Succeeds Act: Assessment Flexibility Under Title I Assessments Section 1111(b)(2)," Knowledge Works, accessed October 23, 2018, knowledgeworks.org/wp-content/uploads/2018/01/essa-faqs-title1-assessments.pdf.

[113] Alyson Klein, "Louisiana, New Hampshire, and Puerto Rico Apply for ESSA Innovative Testing Pilot," *Education Week*, April 3, 2018, blogs.edweek.org/edweek/campaign-k-12/2018/04/ESSA_testing_pilot_louisiana_new_hampshire_and_puerto_rico.html.

[114] "What's in the Every Student Succeeds Act?—Accountability," Ed Trust, January 13, 2016, edtrust.org/resource/whats-in-the-every-student-succeeds-act-accountability/.

[115] Lyons, D'Brot, and Landl, "State Systems of Identification and Support under ESSA."

[116] D'Brot, "Considerations for Including Growth in ESSA State Accountability Systems," p. 4.

[117] Alyson Klein, "ESEA Reauthorization: The Every Student Succeeds Act Explained," *Education Week*, November 30, 2015, blogs.edweek.org/edweek/campaign-k-12/2015/11/esea_reauthorization_the_every.html