

# None of the Above

**A NEW APPROACH TO  
TESTING AND ASSESSMENT**

**August 2014**



**“Teachers have a vital perspective on testing and assessment. As front-line observers, we experience how state assessments work with our specific student populations. As a result, we have valuable insight on how to use testing in schools.”**

**SURAJ GOPAL**, *ninth-grade STEM special education teacher, Hudson High School of Learning Technologies*

# Contents

Executive Summary	<b>1</b>
Introduction	<b>5</b>
Design: Improve the accuracy of standardized assessments	<b>7</b>
Culture: Create and maintain a positive testing environment in schools	<b>12</b>
Teaching: Use data to improve instruction	<b>16</b>
Accountability: Include data in critical decisions	<b>20</b>
Conclusion	<b>27</b>
Teacher Policy Team Process and Methodology	<b>30</b>
Notes	<b>31</b>
Teacher Policy Team and Acknowledgements	<b>34</b>



Jada



# Executive Summary

Standardized testing can be deeply beneficial to students, teachers, and schools by providing an important measure of progress, as well as meaningful feedback about areas of success and areas of growth. As teachers, we know the costs and benefits of assessments. This leaves us between two sides of an often-heated debate, but this is where the evidence leads us. In short, tests have value—so let's take advantage of them. Here is how:

## **DESIGN: IMPROVE THE ACCURACY OF STANDARDIZED ASSESSMENTS**

A large body of research shows that well-designed standardized tests can provide valuable information about students' knowledge and teachers' performance. In fact, such tests are often predictive of long-term life outcomes. It is essential to ensure that all standardized tests are well-designed and that feedback from teachers is solicited during all stages of the testing process.

A common concern is that the accuracy of assessments is undermined by excessive “teaching to the test,” which does not contribute to meaningful learning. However, there is little evidence that test preparation even produces significantly higher test scores when tests are well-designed and focused on higher-order skills. Teachers and principals should be strongly discouraged from teaching to the test because it neither raises test scores nor results in genuine learning.

Computer-adaptive testing is an important tool for improving the accuracy of assessments. Such tests do a better job than traditional assessments of measuring both high- and low-achieving students, and should be made widely available for adoption.

Finally, ensuring the quality of state-created tests is an iterative process. The vast majority of state test items should be released publicly so that stakeholders, such as teachers and parents, can offer feedback on the exams.

## **CULTURE: CREATE AND MAINTAIN A POSITIVE TESTING ENVIRONMENT IN SCHOOLS**

In some schools, the negative culture surrounding standardized testing is pervasive, undermining the value of assessments and harming teachers' morale and students' motivation. A truly pernicious culture can lead to cheating. As educators, we must work within our schools to create a positive culture that recognizes the value of testing for learning and growth. Best practices should be instituted to deter, detect, and investigate potential instances of cheating.

Policymakers must address the negative impact of excessive testing by getting an accurate measure of time spent on assessment and eliminating unnecessary tests. Moreover, the use of alternate assessments, including holistic, portfolio-based exams, should be studied to determine whether they are compatible with data-driven improvement and accountability.

## **TEACHING: USE DATA TO IMPROVE INSTRUCTION**

The data from standardized tests can serve as an important tool for teachers and administrators. Research suggests that both teachers and schools benefit from thoughtful use of data. Data-driven instruction can be improved in a variety of ways, including: ongoing professional development for teachers; a dedicated data specialist in each school; and data that is returned to teachers in a timely, disaggregated, and accessible manner.

“Throughout our team’s research, a ‘positive culture’ of assessments and data-driven instruction was a key recurring theme for school success. That culture starts with each of us, in our own classrooms and buildings, and will only happen if teachers are invested as active participants in the process of shaping changes to testing and assessment.”

**Trevor Baisden**, *founding fifth-grade ELA and history lead teacher,*  
*Success Academy Bronx 2 Middle School*

## ACCOUNTABILITY: INCLUDE DATA IN CRITICAL DECISIONS

Because test scores are important reflections of student learning, assessment data should be a part of consequential decisions. In fact, there is a large body of literature showing the benefits of using tests as part of a multiple measure accountability framework. However, tests should never be the sole basis for any high-stakes decision. For example, the current system of denying graduation to any student who does not pass all Regents exams is misguided and should be revised to incorporate multiple measures.

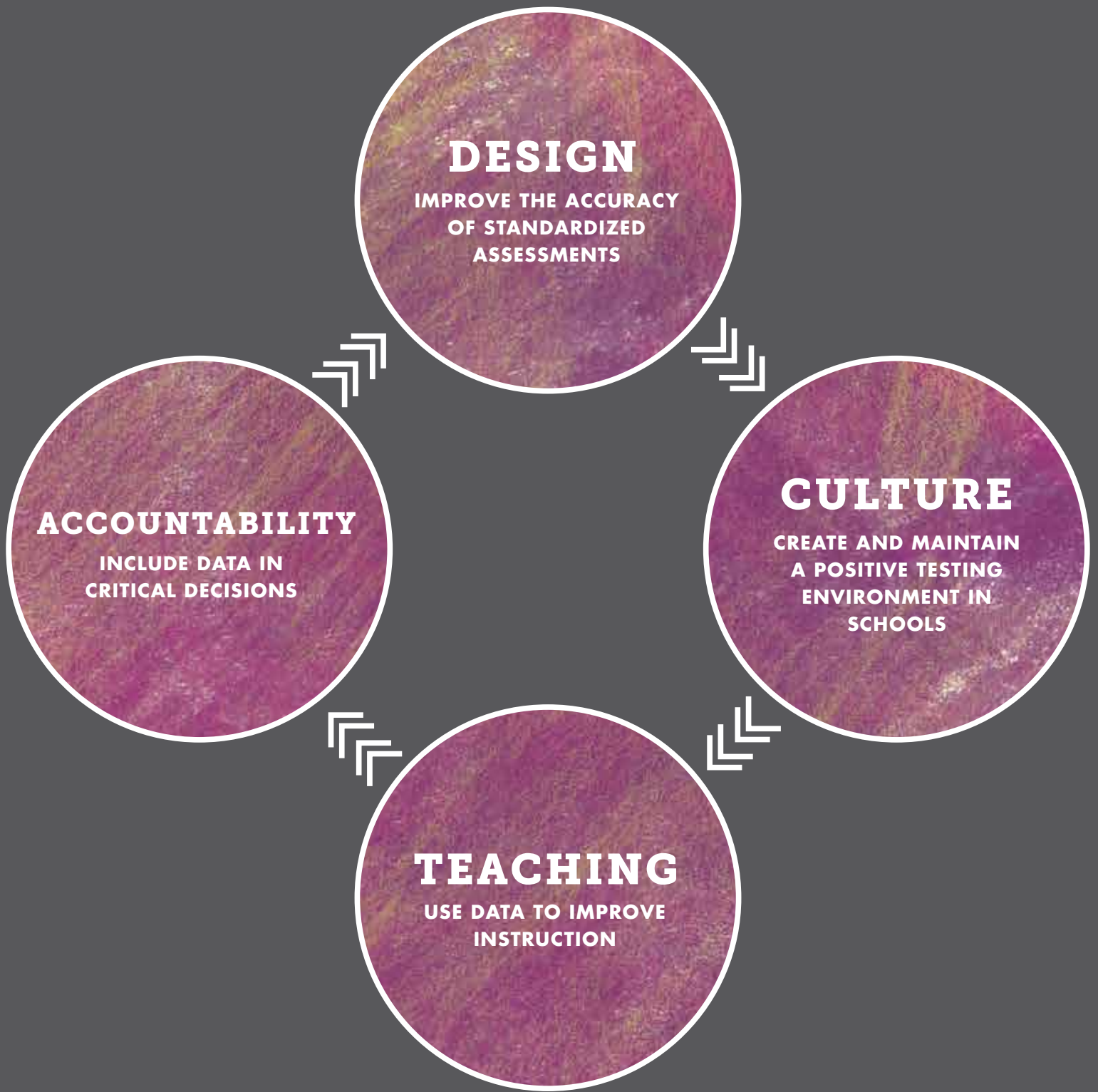
Furthermore, when connecting student test scores to teacher performance, special care must be taken to isolate the effect of teachers and exclude the multitude of factors outside teachers’ control that affect student performance.

Teachers of traditionally non-tested subjects should be evaluated using growth measures or student learning objectives on assessments that are designed with significant input from educators.

## CONCLUSION

We believe in the value of standardized assessments when they are used carefully. They can be a critical tool for teachers and students alike, and we would be unwise to discard them. At the same time, policymakers, administrators, and teachers must invest the time, money, reflection, and work necessary to realize the value of assessments.

# MAKING GOOD USE OF STANDARDIZED TESTS







LORRAINE D'AMICO

Executive Director  
NYC Department of Education

Join the #ENY conversation  
#EduExcellence  
@NYCSchools

Wednesday, November 11, 2015





# Introduction

Standardized assessments have increasingly become a part of life for schools across the country. Since No Child Left Behind became law in 2001, there has been a growing attention to measuring districts', schools', and students' progress, with a particular focus on historically disadvantaged students.

Critics of this trend suggest doing away with standardized tests entirely, while many proponents argue that we simply need to stay the course. As a team of 14 teachers, committed to elevating our profession and ensuring students succeed, our response is “none of the above.” We are unified in the belief that testing has significant value, with the understanding that the way tests are currently designed and used must be improved. In this paper, we lay out a new vision for testing and assessment, beginning with the design of assessments and ending with the important decisions that test results should inform.

In New York, testing has dominated the conversation about the implementation of new teacher evaluation programs and the Common Core State Standards. We find ourselves firmly in the middle between those who would do away with testing altogether and those who do not acknowledge any flaws in the current system. But we are comfortable in the rational middle—comfortable with the view that as educators we can benefit from the information these tests provide. We are comfortable with the idea that our students' growth on tests can be one part of our evaluations, while using that same data to inform

our teaching decisions. Finally, we believe that a standard measure can be critical in ensuring equality in education. We believe that disaggregated assessment data shines a light on populations of students who are not getting the education they deserve.

We all have a part to play in changing the substance and the culture of testing. “None of the Above” has something for everyone: teachers and principals, state and district administrators, elected officials and policymakers. In June 2014, in response to concerns about the role of standardized tests in teacher evaluation, the New York State legislature passed a so-called “safety net” that removes the impact of state assessments on teachers with the lowest evaluation ratings for two years. Let us say no to the all-or-nothing approaches, and make the most of this time to get these tests right.

---

“Teachers see the impact that testing and assessment have on our practice and our students. Teachers know firsthand what is best for our students and our practice. It’s important for us to have a voice in the testing and assessment debate because it has a direct impact on the daily actions of teachers and students.”

**Christine Montera**, *social studies teacher, East Bronx Academy for the Future*



# Design

## IMPROVE THE ACCURACY OF STANDARDIZED ASSESSMENTS

### SUMMARY OF RECOMMENDATIONS

- When designing tests, follow best practices such as ensuring alignment to standards, testing higher-order thinking, and actively soliciting teacher input. ■ ▲
- Prioritize higher-order instruction, and eliminate excessive test preparation that does not contribute to meaningful learning. ▲ ●
- Use computer-adaptive assessments, which improve tests' accuracy by measuring the growth of low- and high-performing students. ■ ▲
- Release the vast majority of state test items publicly after the assessment window has closed so that all stakeholders can monitor the quality of the exams. ■

■=STATE ▲=DISTRICTS ●=SCHOOLS

### KEY TAKEAWAYS FROM RESEARCH AND EXPERIENCE

At the core of the debate on testing is a critical question: Are standardized assessments reflective of students' learning? Test opponents, on one end of the spectrum, claim they are not indicative of student learning or achievement.<sup>1</sup> At the other extreme are those who argue that a single assessment on a single day is the only measure that we should use to make high-stakes decisions.<sup>2</sup> As classroom teachers, we think the truth falls somewhere in between.

There is abundant research showing that standardized tests are meaningful. Such assessments can predict with moderate accuracy individuals' first-year college GPA,<sup>3</sup> cumulative college GPA,<sup>4</sup> post-college income,<sup>5</sup> and success in graduate school.<sup>6</sup> Aggregate international test scores are also predictive of the economic prosperity of countries.<sup>7</sup> Additionally, teachers whose students' standardized test scores grow produce an increase in those students' adult incomes and rates of college attendance.<sup>8</sup> This research shows that standardized tests are able to capture important information about what is happening in our classrooms.

Standardized tests, however, are not the be-all and end-all; they do not measure everything that matters. There are many students who do not test well and end up leading

happy, successful lives. Research indicates that certain subjective evaluations of teachers are only modestly correlated with their students' test-based success,<sup>9</sup> suggesting what many teachers know: that tests cannot measure the full value of an educator. Indeed, the teachers who have the greatest positive effects on students' social and behavioral skills are not always the ones who produce the highest test score gains.<sup>10</sup> This is why past E4E-New York papers on teacher evaluation<sup>11</sup> and Common Core implementation<sup>12</sup> have insisted on multi-measure evaluation and decision-making for teachers and students.

There are other limitations to standardized tests, which we will discuss later in this paper, but, in short, tests are meaningful but don't measure everything.



“Improving tests will only be effective with the active participation of teachers in testing design on a district and state level. Our ability to share insights from the classroom, as well as the cultural and socioeconomic backgrounds of our students, will undoubtedly help create high-quality assessments.”

**Blackfoot U-Ahk**, fourth- and fifth-grade teacher of students with severe emotional disabilities,

Coy L. Cox School P.369k

### RECOMMENDATION:

WHEN DESIGNING TESTS, FOLLOW BEST PRACTICES SUCH AS ENSURING ALIGNMENT TO STANDARDS, TESTING HIGHER-ORDER THINKING, AND ACTIVELY SOLICITING TEACHER INPUT.

All tests are not created equal. Anecdotally, as teachers, all of us have experience with assessments that were poorly written or were not aligned with the academic standards. We also all have experience with many well-designed tests that were fair assessments of our students’ learning and our teaching, and that gave us important data that we were able to use to improve our instruction.

We were heartened to learn about the process that New York State test questions (technically called “items”) go through before they are ever used on an official exam. It takes a full two years for each item to be approved through a process that includes extensive field testing, statistical

validation, and input from a committee of teachers.<sup>13</sup> It is disconcerting, however, that even after such a thorough process, there are still concerns from educators about the quality of these tests.<sup>14</sup>

We are glad the New York State Education Department uses a committee of teachers to validate testing items. The opportunity to join such a committee should be widely disseminated so that as many teachers as possible have the chance to share their voice.

We also believe that there should be a formal system for soliciting and receiving teacher commentary so that all educators can share feedback after a test has been given. We recommend that the State Education Department send a survey to all teachers who administered tests to gather feedback on positive and negative aspects of the assessments.

## DESIGNING QUALITY ASSESSMENTS

When designing all tests, the following practices must be followed:

- Classroom teachers need to provide input throughout the process, from the creation of the tests to feedback after the tests are given. This feedback must be taken into account and meaningfully acted upon.
- Tests must be aligned to standards and assess higher-order thinking skills.<sup>15</sup>
- The diversity of students’ backgrounds—including differences in geography, socioeconomic status, racial identity, disability status, etc.—must be considered in test development in order to avoid potential bias.
- Test items should be worded to make sure each item measures the specific standard being assessed, as opposed to students’ ability to understand a tricky question.
- The amount of time given for assessments and the number of assessments given in a single day need to be age-appropriate.

## RECOMMENDATION:

### PRIORITIZE HIGHER-ORDER INSTRUCTION AND ELIMINATE EXCESSIVE TEST PREP.

One of the most serious critiques of standardized assessments is that excessive “teaching to the test” can effectively negate the validity of an exam, as students learn how to score well without learning meaningful skills or content. Teaching to the test, or “drill and kill,” tends to take valuable time away from rich, higher-order instruction. No teacher gets into the profession for this kind of mechanized work, and it undermines teachers’ and students’ love of school.

But contrary to the notion that tests can be gamed by excessive preparation, research suggests that the best way to prepare for most standardized assessments is through challenging, authentic work focused on content and skills.<sup>16</sup> One study that examined students’ preparation for the ACT found that “improvements from [an ACT pre-test] to the ACT are smaller the more time teachers spend on test preparation in their classes and the more they use test preparation materials.” Moreover, “the focus on testing strategies and practice diverts students’ and teachers’ efforts from what really matters—deep analytic work in academic classes.”<sup>17</sup> In other words, at least for well-designed assessments, excessive test preparation may actually lead to worse results.

This aligns with our experience, as well as recent statements from education leaders. As New York City Schools Chancellor Carmen Fariña said, “If we do good teaching, that’s the best test prep.”<sup>18</sup> Similarly, New York State Education Commissioner John King stated, “The best preparation for testing is good teaching.”<sup>19</sup> We agree.

Since there is scant evidence that excessive teaching to the test will lead to higher assessment results, teachers and principals need to be shown this research. When educators realize that test prep is counterproductive, more time will be spent on authentic teaching and learning.

## RECOMMENDATION:

### USE COMPUTER-ADAPTIVE ASSESSMENTS.

One valid concern about traditional tests is that they cannot adequately capture the growth of students who are significantly above or below grade level. The good news is that technology offers a solution to this problem—computer-adaptive testing adjusts question difficulty based on students’ demonstrated skill level. This sort of assessment, which is already relatively widely used, including by the Graduate Record Examinations (GRE)<sup>20</sup> and the Graduate Management Admission Test (GMAT),<sup>21</sup> would help teachers get a better sense of students’ growth from year to year.<sup>22</sup> Similarly, computer-adaptive tests give more accurate information to students and parents. We therefore strongly support the use of computer-adaptive testing whenever available, and encourage investment in this alternative where it does not exist.

Questions have been raised regarding whether computer-adaptive testing will lead to low expectations for struggling students.<sup>23</sup> We understand these concerns, but ultimately disagree: We are not aware of evidence that educators will lower expectations for their students simply because tests focus on academic growth. If, for example, data show that a certain school’s students are not making progress, efforts can be made to help those students and ensure that teachers are held accountable. In that sense, more accurate data will help rather than hinder the improvement and accountability process. Moreover, there is no clear alternative—students who are far behind or far ahead need a meaningful gauge of their progress, and computer-adaptive tests provide this.

**“Computer-adaptive testing is absolutely crucial because many of my students are far behind and would benefit from a test scaled to their abilities.”**

**Rachael Beseda**, *first-grade special education teacher, Global Community*

*Charter School*

## WHAT IS COMPUTER-ADAPTIVE TESTING?

Computer-adaptive assessments start all students at the same level—in this case, at their grade levels. However, questions on the test become progressively harder as the test-taker gets more questions right or progressively easier as the test-taker gets more questions wrong. That does not mean that if a student gets the first few questions wrong, the remainder of the test will be below grade level. Instead, the test continuously adapts based on the student's responses. For example, if a student gets the first few questions wrong, but the next several questions right, the difficulty level will begin increasing as more correct answers are given. This process allows assessments to meet students where they are at in order to get an accurate measure of their learning and growth.

That being said, it is important that computer-adaptive assessments give all students a fair opportunity to engage with grade-level content. All tests should begin with grade-level questions, and only move down once it becomes clear that students are not at grade level. Furthermore, such tests should attempt to push all students to demonstrate higher-order thinking skills. For example, a student reading below grade level can still be given the chance to show the same skills as her grade-level peers, but do so with a less-challenging text.

### RECOMMENDATION:

RELEASE THE VAST MAJORITY OF STATE TEST ITEMS PUBLICLY AFTER THE ASSESSMENT WINDOW HAS CLOSED.

All tests, especially those used for making high-stakes decisions, need to undergo careful scrutiny both before and after administration. We believe there is a healthy process in place to ensure quality in the creation of New York State exams. At the same time, it has been frustrating for many educators that state tests prohibit teachers and students from discussing the contents of the exam.<sup>24</sup> Right now, with low public confidence in tests,<sup>25</sup> the state needs

to allocate funds to significantly increase the transparency of state assessments,<sup>26</sup> except for field test items, which, by design, cannot be publicly released. These funds will allow for the printing of additional “forms” of state assessments that will give the state the ability to field test more items, decreasing the need to reuse (and thus keep hidden from public view) previous items. This will allow for the elimination of the widely criticized<sup>27</sup> stand-alone field tests.

Increased transparency will let educators, parents, and students give feedback on state tests, which is particularly important as the Common Core standards are being implemented. This also ensures that teachers and students have a better understanding of what to expect on future exams. We believe that this will not only improve the assessments themselves by holding test designers and the New York State Education Department accountable to the public, but will also help restore public trust in the exams.

“Schools’ and teachers’ emphasis should always be on high-quality rigorous instruction. Both research and experience suggest that this is the best method for preparing for well-designed assessments.”

**Vivett Hemans**, *English and language arts teacher, Eagle Academy for Young Men of Southeast Queens*



## ADDITIONAL BENEFITS OF TESTING

Our paper is organized around the two main benefits of standardized assessments: using them for improvement and as a factor in important decisions. However, we would be remiss if we did not discuss some smaller, but important additional benefits of testing.

- Assessments provide evidence of achievement and opportunity gaps. Using both the NAEP and state tests mandated by No Child Left Behind, policymakers and concerned citizens have quantitative evidence of the inequities that persist in our country. Testing not only shows that this is the case, but also helps quantify the gap and determine whether it is expanding, contracting, or staying constant. While qualitative evidence is also important in this regard, test scores can provide the hard data necessary to bring light to the shameful inequities that persist in our country.
- Standardized tests are important to prepare students for success in adult life. Not only must college-bound students take the SAT or ACT, but all those who aspire to graduate school must take additional exams. Potential lawyers must take the LSAT and the bar exam; would-be doctors must do well on the MCAT and board exams. The list goes on and includes most professions. That is not to say that the purpose of K–12 education should be to prepare students for assessments, but we would be doing a disservice if we limit students' exposure to the types of high-stakes tests they need to do well on later in life.
- There is some evidence that assessments do not simply measure learning, but actually enhance it. A variety of studies<sup>28</sup> have found that students retain information better after being tested on it. At this point, it is not clear that this research applies to standardized tests, but it is a potential value that points to the necessity of aligning standards—what is taught in class—to what is tested.

# Culture

## CREATE AND MAINTAIN A POSITIVE TESTING ENVIRONMENT IN SCHOOLS

### SUMMARY OF RECOMMENDATIONS

- Measure time spent, by both students and teachers, on testing and eliminate unnecessary and redundant exams. ■ ▲
- Implement best practices, such as administering tests in controlled environments and monitoring for test irregularities, to prevent and detect cheating. ■ ▲ ●
- Create or expand pilot programs of schools using nontraditional tests to determine whether they lead to positive results for students, and can be used to evaluate and support teachers and schools. ■ ▲

■=STATE ▲=DISTRICTS ●=SCHOOLS

### KEY TAKEAWAYS FROM RESEARCH AND EXPERIENCE

Many of our experiences suggest that in too many instances, the culture of testing and assessment in New York has turned toxic. No doubt this is not the case in all schools, but for too many of us, testing has become something to be feared and avoided.

But it does not have to be that way. The negative culture of testing that permeates some schools must change. We believe that part of this shift has to come from us as teachers: We should be focusing on the value that assessments have to offer. We cannot be surprised that a pessimistic culture exists in schools if the adults in those buildings have counterproductive attitudes about testing.

Teachers cannot solve this problem alone, however. We need principals to do their part, by setting a positive building-wide tone about assessments. Moreover, as discussed earlier, we need principals to communicate clearly to teachers that excessive test prep will not raise test scores. Currently, though, it is often principals who

mandate that teachers engage in this counterproductive practice, feeding a negative cycle that harms student engagement.

As we will discuss further in a subsequent section, teachers also need to be given the tools to use test results to improve instruction. When teachers are supplied with what we need to make tests valuable, our outlook will change for the better. Moreover, part of the anxiety that surrounds testing comes from the feeling that a single test can determine our students' futures. A commitment to using multiple measures for all high-stakes decisions—another topic we will elaborate on in a later section—will go a long way toward eliminating this fear.

Finally, accountability must be paired with support throughout the year. What if teachers and students did not feel that low test scores would lead to punishments or poor ratings, but that they would lead to increased support and resources? To be clear, we do believe in accountability, but accountability should always go hand-in-hand with support and resources. Tests should be instructive, as well as evaluative. It is outside the scope of this paper to address what such support should look like specifically, but this should be a core tenet of any accountability system.

## RECOMMENDATION:

MEASURE TIME SPENT, BY BOTH STUDENTS AND TEACHERS, ON TESTING AND ELIMINATE UNNECESSARY AND REDUNDANT EXAMS.

One cause of the general frustration directed at standardized tests is the widespread feeling that there are simply too many of them. We certainly feel that way. As we have elaborated, we believe there is value in assessment, but any such value must be weighed against the time and effort invested in testing.

The first and most important step must be to accurately gauge how much time is being spent on testing. We were glad that New York State Governor Andrew Cuomo's Common Core Implementation Panel attempted to

address the underlying problem by recommending a 2 percent limit on school time spent on local and state assessments combined, and a 2 percent limit on test prep.<sup>29</sup> These suggested changes were subsequently implemented in the State Budget.<sup>30</sup>

The goal here is laudable, but we are skeptical of an arbitrary percentage that does not vary by grade. That is why we need a genuine figure for just how much time and money are spent on testing—this should include time spent preparing, administering, and grading these assessments for teachers; money spent developing the test; time spent by students taking tests (including field tests); and instructional time lost on days when tests are administered.

We think the state took a step in the right direction by requiring an audit of assessments to make sure districts are not giving unnecessary assessments based on the assumption that they are mandated by the state.<sup>31</sup> It is important that this audit is prioritized so that excessive testing is reduced as soon as possible.

Once these two audits are complete, districts can make smart decisions, with the input of teachers, about which tests are worthwhile and which are not.





## RECOMMENDATION:

### IMPLEMENT BEST PRACTICES TO PREVENT AND DETECT CHEATING.

Though the vast majority of educators regularly administer assessments with honesty and fidelity, an extreme outgrowth of a counterproductive school culture manifests itself in cheating scandals, which have occurred throughout the country.<sup>32</sup>

Some have taken these cheating scandals to mean that standardized tests should be eliminated, but this makes no more sense than cancelling final exams because a handful of students tried to cheat on them. Instead, we should institute best practices—based on a U.S. Department of Education symposium on test integrity<sup>33</sup>—to ensure that cheating rarely happens, and how to detect and investigate it if it does.

<p>In order to <b>PREVENT CHEATING,<sup>34</sup></b> the state, districts, and schools must:</p>	<p>In order to <b>DETECT EVIDENCE OF CHEATING,<sup>35</sup></b> the state, districts, and schools must:</p>	<p>In order to <b>INVESTIGATE CHEATING,<sup>36</sup></b> the state, districts, and schools must:</p>
<p>Develop and disseminate a standard definition of cheating.</p>	<p>Monitor test results for irregularities as part of the testing process.</p>	<p>Establish procedures for conducting an investigation if one is necessary.</p>
<p>Train principals and teachers to administer exams.</p>	<p>Ensure that proctors look for evidence of irregularities during assessment administration.</p>	<p>Create standards that will trigger an investigation.</p>
<p>Keep testing windows short.</p>	<p>Use advanced analytic techniques, such as erasure analysis, to check for irregularities.</p>	<p>Provide whistleblower protections.</p>
<p>Administer tests in controlled environments.</p>		<p>Use trained personnel to conduct the investigation.</p>
<p>Establish and monitor a chain of custody for testing materials.</p>		<p>Make the investigation as transparent as possible.</p>
<p>Store and score test materials off-site.</p>		<p>Make use of sanctions when wrongdoing is found.</p>

In sum, these best practices, created by experts in the field, will help stop cheating in the first place, while ensuring a fair process if testing irregularities are found. We emphasize, though, that an ounce of prevention is worth a pound of cure here—a healthy testing culture will go a long way toward eliminating this problem.

## RECOMMENDATION:

### CREATE OR EXPAND PILOT PROGRAMS OF SCHOOLS USING NONTRADITIONAL TESTS.

One serious problem with traditional standardized tests, which often include multiple-choice questions, is that it can be difficult to continually engage students in such exams. For students and teachers, so-called “bubble tests” have become a chore that must be endured. As discussed

earlier, we believe that schools have an important role in changing this culture. At the same time, alternatives to traditional assessments should be explored and tested for their effectiveness.

The New York Performance Standards Consortium is a group of 28 schools that have used “performance assessments” in place of traditional high-stakes tests.<sup>37</sup> The Consortium schools boast impressive results, showing



“Policymakers should embrace a pilot program for portfolio assessment in order to see whether this type of assessment can work. I think that project-based learning and inquiry-based work are things I don’t do nearly enough—I rely on more traditional assessments, and teachers need to think of ways to cater to all students’ needs and strengths in terms of assessment.”

**Charlotte Steel**, seventh-grade math teacher, Booker T. Washington M.S. 54

their students graduate high school at higher rates than other demographically similar New York City students.<sup>38</sup> But the fact that these schools produce strong graduation rates does not mean that performance assessments are the cause. Moreover, legitimate questions have been raised regarding the ability to fairly and efficiently use performance assessments to evaluate teachers and assess student learning.<sup>39</sup>

We therefore propose an expanded pilot program that allows more schools to enter into the Performance Standards Consortium, while also determining whether

such assessments are compatible with data-driven improvement and accountability. We recommend opening up an application for schools interested in joining the program, and conducting a lottery in order to randomly accept half of the eligible applicant schools into the pilot. Under this approach, schools that adopt the performance assessment model can be evaluated against similar schools that do not. If this system gets positive results for teachers and students, it should be expanded to even more city schools.

# Teaching

## USE DATA TO IMPROVE INSTRUCTION

### SUMMARY OF RECOMMENDATIONS

- Offer high-quality training throughout the year for teachers on how to improve instruction using assessment data. ▲ ●
- Provide each school with a teacher who serves as a "data specialist." ▲ ●
- Ensure that teachers and administrators receive timely, detailed, and disaggregated data in a transparent, accessible format. ■ ▲

■=STATE ▲=DISTRICTS ●=SCHOOLS

### KEY TAKEAWAYS FROM RESEARCH AND EXPERIENCE

Research is clear that assessment data can be used as a tool for teachers and schools to improve.<sup>40</sup> It has been found, for example, that schools that make thoughtful use of data often produce significant gains in student achievement.<sup>41</sup> Research also suggests that access to data can increase the quantity and quality of conversations that educators have with colleagues, parents, and students.<sup>42</sup> Data can enhance collaboration among educators<sup>43</sup> and can improve teachers' instruction.<sup>44</sup> There is also evidence that the most successful charter schools make use of data-driven improvement and instruction.<sup>45</sup> Overall, data can and should be used to help schools and teachers improve.<sup>46</sup>

Unfortunately, this is not always happening. One recent study found that a new data system introduced in Cincinnati Public Schools was rarely used by educators and did not lead to observable student gains.<sup>47</sup> A pilot program in Pennsylvania produced similar results.<sup>48</sup> The key, then, is to give teachers the support we need to make good use of testing data.

### RECOMMENDATION:

OFFER HIGH-QUALITY TRAINING THROUGHOUT THE YEAR FOR TEACHERS ON HOW TO IMPROVE INSTRUCTION USING ASSESSMENT DATA.

Teachers and administrators need more training on how to use data effectively. The New York City teachers' contract recently put in place more time for professional development.<sup>49</sup> Some of that time should be dedicated to high-quality training on understanding and using student



“It is particularly important that teachers receive thorough and useful training in data-driven instruction.

Unless the results of assessments are used to move teaching and learning forward, they serve little value.”

**Michelle Kniffin**, *ninth- to 12th-grade math teacher,*

*High School of Telecommunication Arts and Technology*

data. It is worth noting that while we support school-based creation of professional development, this may be an area in which schools need outside support and expertise to design appropriate programs.

### **RECOMMENDATION:**

PROVIDE EACH SCHOOL WITH A TEACHER WHO SERVES AS A “DATA SPECIALIST.”

Teachers need continuous support in using data systems. We need more than a one-time training. We propose that at least one teacher in each school receive the designation of “data specialist.” This role should come with extensive training, as well as the responsibility of supporting and working with staff to use data and integrate this information into their regular assessment of—and feedback for—their students. Additionally, data specialists should receive compensation for this role that is either monetary or in the form of a lighter class load. A final benefit is that this position could potentially serve as an additional rung on a teacher career ladder, a concept that past E4E Teacher Policy Teams have endorsed.<sup>50</sup>

### **RECOMMENDATION:**

ENSURE THAT TEACHERS AND ADMINISTRATORS RECEIVE TIMELY, DETAILED, AND DISAGGREGATED DATA IN A TRANSPARENT, ACCESSIBLE FORMAT.

To make full use of assessments, teachers and administrators need timely, detailed, and disaggregated data in order to tailor their instruction to address their students’ needs. The current system does not supply educators with sufficiently detailed feedback on these exams. Compounding this problem is the fact that the results do not come back until the summer, and thus teachers often cannot act on the data. A high priority must be placed on giving educators actionable, disaggregated, and timely results from standardized assessments. Teachers also need access to a high-quality, easily navigable interface in which we can access all relevant data. Georgia, in particular, is a state that has been highlighted for its success in making data accessible and easy to use for teachers,<sup>51</sup> and New York should follow suit.

## COMMON CORE ASSESSMENTS CONSORTIUM

As the Common Core State Standards are being implemented across the country, new testing consortia are being rolled out that are aligned to the new standards. There are two testing groups: Smarter Balanced Assessment Consortium (SBAC),<sup>52</sup> which has been adopted at least in part by 20 states,<sup>53</sup> and the Partnership for Assessment of Readiness for College and Careers (PARCC),<sup>54</sup> which has been adopted by 14 states and the District of Columbia.<sup>55</sup> Field tests took place in the spring of 2014,<sup>56</sup> and the full assessments will be available for use beginning in the 2014–2015 school year.

New York State has adopted PARCC,<sup>57</sup> but has not yet determined when the new exams will be rolled out.<sup>58</sup> Below, we discuss aspects of PARCC and how they align with our recommendations:

- We are encouraged that PARCC assessments appear to test higher-order thinking skills. Although it is too early to determine for sure, the sample questions<sup>59</sup> leave us optimistic that rigorous skills will be tested, and low-level multiple-choice tests will be deprioritized.
- It is very important that PARCC continuously involves teachers in the creation and revision of the exams. PARCC has already shown evidence of having engaged teachers throughout this process, and we are pleased to see such a clear commitment to teacher input.<sup>60</sup> Moreover, we recommend that PARCC distribute surveys to teachers at the end of each year to garner feedback on the year's assessments.
- Although PARCC tests will be completed using computers, they will not be computer adaptive,<sup>61</sup> with the important exception of optional diagnostic exams. It is disappointing that this valuable technology will not be utilized for the summative assessments, as PARCC is missing an opportunity to get accurate growth measures of high- and low-achieving students. Although a PARCC frequently asked questions document<sup>62</sup> claims that the assessments "will measure the full range of student performance, including the performance of high- and low-achieving students," it is not clear how they will manage to do so. We urge PARCC to consider moving to computer-adaptive assessments, particularly in light of the fact that SBAC will be utilizing this technology.<sup>63</sup>
- An advantage of computer-based assessments is that cheating will be more difficult, since school staff will not handle or transport physical testing materials.<sup>64</sup> However, new threats to testing security—such as access to the Internet—may exist, and PARCC, in partnership with schools and districts, must ensure teachers and school leaders are prepared to administer the tests fairly and monitor for irregularities.
- An additional advantage of using computer-based assessments is timely feedback to schools, teachers, and students. For many questions—ones that have clear right or wrong answers—the data should be available almost immediately. Though for others—performance tasks, essays, or any items that require manual grading—the turnaround will understandably be longer. However, we are glad that PARCC has stated that its "goal is to have data from the performance-based assessments returned before the end of the school year."<sup>65</sup> It is crucial that PARCC ensures that teachers receive timely, disaggregated, and user-friendly data.
- As we have argued, transparency is a necessary aspect for all important exams, in part to ensure that the public is given an opportunity to offer feedback on the content and quality of assessments, and in part to ensure public trust in such assessments. So far, we are encouraged that PARCC has already released sample tests<sup>66</sup> and plans to release 40 percent of test items each year. We hope the commitment to transparency continues and expands as full-scale tests are implemented.





# Accountability

## INCLUDE DATA IN CRITICAL DECISIONS

### SUMMARY OF RECOMMENDATIONS

- Isolate the effects of teachers and schools to ensure that those serving at-risk student populations are not penalized by out-of-school factors. ■ ▲
- Evaluate teachers of non-tested subjects based on authentic assessments—developed and validated by teachers—using growth measures or student learning objectives. ▲
- Make high-stakes decisions based on multiple sources and multiple years of evidence. ■ ▲ ●

■=STATE ▲=DISTRICTS ●=SCHOOLS

### KEY TAKEAWAYS FROM RESEARCH AND EXPERIENCE

There is now abundant evidence that using test score growth as part of a multiple-measure evaluation and accountability system can benefit students. Multiple peer-reviewed studies<sup>67, 68, 69</sup> have found that students benefit when adults are held accountable for results.<sup>70</sup> There is also research showing that teacher evaluation that considers evidence of student learning can be beneficial to students.<sup>71</sup> Finally, and most importantly, evidence suggests that, when designed and implemented well, accountability systems can impact school quality in a way that leads to long-term positive effects on students' adult incomes.<sup>72</sup> All that being said, the current way that test scores are used to make important decisions needs to be improved to ensure they are fair to students, teachers, and schools.

### RECOMMENDATION:

ISOLATE THE EFFECTS OF TEACHERS AND SCHOOLS TO ENSURE THAT THOSE SERVING AT-RISK STUDENT POPULATIONS ARE NOT PENALIZED BY OUT-OF-SCHOOL FACTORS.

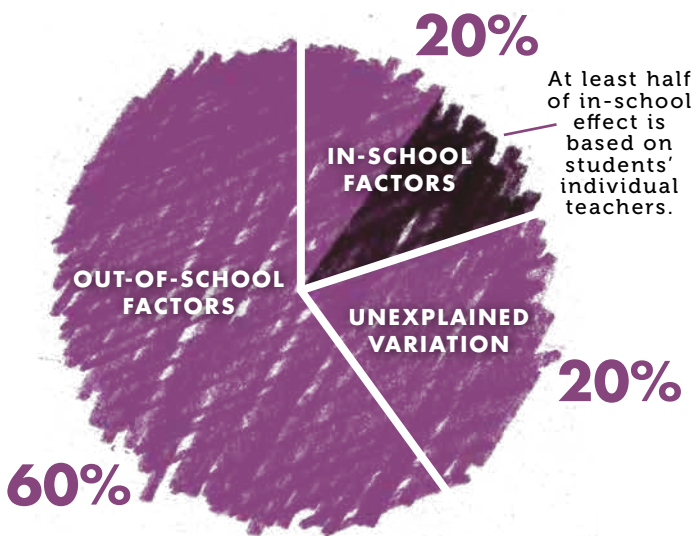
One of the most difficult, but most important, aspects of using student test score growth in an evaluation system is isolating the effects of schools and teachers. After all, many factors—including poverty and parental involvement—

affect a given student's achievement, and only a fraction can be attributed to his teachers or his school. Indeed, only about one-fifth to one-quarter of student test scores are explained by the quality of their schools, and of that, about one-half to two-thirds are the result of the student's individual teachers.<sup>73</sup>

We are *not* saying that teachers and schools do not matter. But we also cannot blame those same teachers and schools for all the factors that can contribute to low student achievement. If we simply look at absolute test scores, as often happens,<sup>74</sup> with no accounting for growth or student background, the schools and teachers working with our most challenging students will be unfairly penalized. Moreover, some struggling schools and teachers who work with high-achieving students will be overlooked.<sup>75</sup>

With the use of “value-added modeling,”<sup>76</sup> we can go a long way toward isolating teachers’ and schools’ effects by controlling for students’ prior tests scores, as well as other factors outside teachers’ control.

## FACTORS CONTRIBUTING TO STUDENT ACHIEVEMENT



Source: Di Carlo, M. (2010, July 14) Teachers Matter, But So Do Words. Shanker Blog. Retrieved from <http://shankerblog.org/?p=74>. (Note that these percentages are approximations.)

## WHAT IS VALUE ADDED?

**Value added is a statistical method that attempts to isolate teachers’ influence on their students’ test score growth.**

**Value-added models can take into account a variety of variables that affect students’ performance, including prior achievement, socioeconomic status, disability status, special education status, attendance, disciplinary record, and class size.<sup>77</sup> Although some critics of value-added measures correctly point out that teachers’ ratings can vary from year to year,<sup>78</sup> others respond that this can be ameliorated through multiple years of data, and that similar variance exists in performance metrics of other professions.<sup>79</sup> Value-added scores are particularly reliable for teachers at the extremes of the distribution.<sup>80</sup> Research also suggests that teachers’ value-added scores predict their effects on students’ long-term outcomes such as income and college attendance.<sup>81</sup>**



## EXAMPLE: TWO-STEP VALUE-ADDED MODEL

In recent years, as New York has started using a student growth model to evaluate teachers, concerns have been raised about the extent to which it fairly accounts for factors outside of educators' and schools' control.<sup>82</sup> A report on the subject found evidence that the 2012–2013 New York State growth measure may have been partially biased against some teachers and principals who serve certain student populations.<sup>83</sup> With New York State likely to use value-added scores as 25 percent of teacher evaluation in the 2014–2015 school year,<sup>84</sup> now is the time to consider the ideal model.

We recommend an approach that more fully accounts for factors outside teachers' and schools' control. This method—known as a two-step value-added model, or proportionality—is designed to make apples-to-apples comparisons.<sup>85</sup> In other words, this model eliminates any correlation between teachers' and schools' value-added scores and the student populations they teach—it guarantees that educators of, for example, students in poverty or students with disabilities will not receive disproportionately low ratings.

This will address the concern that student achievement measures penalize teachers and schools who serve certain student populations. It will also ensure that evaluation measures will not exacerbate persistent inequities in those schools—high-poverty schools will have a tougher time recruiting and retaining teachers if those educators face a higher chance of a low evaluation score.

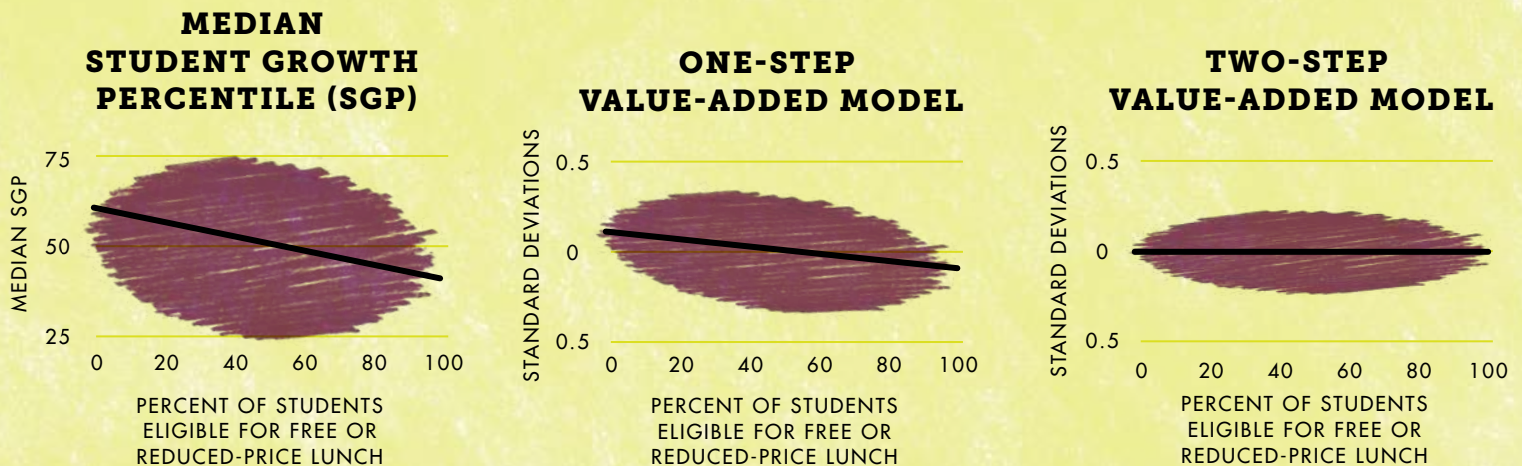
We recognize that genuine inequalities persist between and within our schools,<sup>86</sup> and that correlations between teacher effectiveness scores and student populations likely reflect some genuine differences in teacher quality. But our goal in an evaluation system is not just to get an accurate picture of teacher quality, but also to design a system that provides useful information to support teacher and school improvement, while helping districts and principals make retention and dismissal decisions. We are convinced that the two-step model does just that.<sup>87</sup>





## COMPARING DIFFERENT GROWTH MEASURES

The following graphs show three different ways of measuring schools' student achievement growth. The x-axis is a measure of school poverty, while the y-axis is a measure of school effectiveness based on the given growth measure. The shaded areas are scatter plots showing the range of schools' scores. The line shows the correlation between schools' level of poverty and their level of effectiveness. Note that these are examples based on schools in Missouri, so representations of New York schools may vary in certain ways.



Sample for all growth measures is 1,846 schools

Source: Ehlert, M., Koedel, C., Parsons, E., et al. Selecting Growth Measures for School and Teacher Evaluations: Should Proportionality Matter? National Center for Analysis of Longitudinal Data in Education Research. Retrieved from <http://www.caldercenter.org/publications/upload/wp-80-updated-v3.pdf>

### RECOMMENDATION:

EVALUATE TEACHERS OF NON-TESTED SUBJECTS BASED ON AUTHENTIC ASSESSMENTS—DEVELOPED AND VALIDATED BY TEACHERS—USING GROWTH MEASURES OR STUDENT LEARNING OBJECTIVES.

Many educators do not teach in grades or subjects that have annual state tests, and therefore cannot be evaluated using value-added measures. In order to comply with the new evaluation law, some teachers are being rated based on students or subjects they do not teach—for example, in some cases gym teachers are being rated on English scores.<sup>88</sup> This practice must stop, because it violates a core tenet<sup>89</sup> of any accountability system: Teachers should not be held accountable for outcomes outside of our control.<sup>90</sup> We are glad that the New York City teachers' contract will move accountability in this direction.<sup>91</sup>

Districts need to invest in authentic performance measures for teachers in non-tested subjects, particularly ones

like music, art, and physical education. In many cases, these performance assessments may be combined with more traditional written tests. Results should not only be considered in individual teachers' evaluations, but school evaluations as well. The creation of standardized performance assessments for these subjects has been experimented with,<sup>92</sup> though the evidence is limited on how successful such programs have been. In all non-tested subjects, evaluations should be based on student learning objectives<sup>93</sup> or measures of student growth that ensure fair comparisons are being made across classrooms.

Our top priority is to ensure that any such assessments are designed by and with teachers, and are validated by teachers. Educators should have a hand in the design, the administration, and the revision of these assessments. This is absolutely essential. When teachers are involved in the creation of exams, the tests are more likely to reflect what is being taught in the classroom.

“Using multiple measures for high-stakes decisions is particularly important to me and my students because so many ELLs often struggle on tests but are bright, capable students.”

**Maura N. Henry**, sixth- to 12th-grade English as a Second Language teacher,

*The Young Women’s Leadership School of Astoria*

### **RECOMMENDATION:**

MAKE HIGH-STAKES DECISIONS BASED ON MULTIPLE SOURCES AND MULTIPLE YEARS OF EVIDENCE.

We believe in the value of test scores to inform and evaluate students, teachers, principals, and schools, but we also are convinced that a single test score should not be the sole basis for any high-stakes decision. A broad array of theory and evidence suggests that multiple measures are always preferable in high-stakes circumstances.<sup>94</sup> We are encouraged, then, that New York City—like all districts and states that have adopted the new wave of teacher

evaluation<sup>95</sup>—has used a multiple measure system, with student growth as one factor among others.<sup>96</sup> Similarly, we are glad that the New York City Department of Education recently adopted a multiple-measure system for student promotion and retention decisions.<sup>97</sup> We think the city and state have done a good job ensuring that important decisions are based on multiple sources of evidence. Nevertheless, there is room for improvement.

## **EXAMPLE: HIGH SCHOOL GRADUATION EXAMS**

The current requirement that all New York State students pass a series of exams in order to graduate high school is an example of a policy that fails to consider multiple measures. Under the current system, students will only receive a high school diploma if they pass five state-mandated Regents exams.<sup>98</sup> (Students with disabilities or IEPs have some limited additional options.)

This policy is designed to create high expectations for students—an admirable goal—but it ends up harming some of them. Anecdotal<sup>99</sup> and empirical<sup>100, 101, 102, 103</sup> evidence show that high school graduation exams have little or no positive effects and significant negative consequences for students who fail such tests. There is even alarming research showing that mandated graduation exams can lead to increased incarceration rates.<sup>104</sup>

With this evidence in mind, we take the position that high school graduation exams should never be the sole basis for denying students their diplomas. It is appropriate for such tests to be part of a multiple-measure graduation system, but not as inflexible roadblocks for students trying to graduate. It is outside our scope to discuss what precisely such a system should look like, but we will note that, holistic multi-measure graduation models exist and should be studied.<sup>105</sup>



## UNIQUE STUDENT POPULATIONS

One important aspect of assessment that is not discussed enough is the effect on unique populations of students, including those receiving special education, students with disabilities, English-language learners, and gifted and talented students. A thorough discussion of issues surrounding testing with each of these student populations is beyond the scope of this paper. However, we were very cognizant of these students while crafting our recommendations. Here, we highlight and elaborate on how specific components of our recommendations affect these students.

- In the design of tests, the needs of unique populations of students must be carefully considered. First and foremost, teachers of a variety of student populations should be represented on the panel of educators who design and review assessments. Particular care must be given in writing test items to ensure that certain students are not disadvantaged. For example, math tests should not, in most cases, include idioms that English-language learners might not be familiar with, since such a question would not measure those students' mathematical ability.
- As we have previously articulated, we believe in the value of computer-adaptive testing. These assessments will benefit unique student populations—specifically those who are low- and high-achieving—by gauging their growth accurately. This needs to be a high priority. If we want students and teachers to believe in the value of the assessments, we need to make them useful to all students. Computer-adaptive tests will significantly help in this regard.
- Our recommendation regarding the use of multiple measures in making high-stakes decisions—specifically graduation decisions—will have a positive effect on unique populations of students.<sup>106</sup> English-language learners and special education students have long graduated at lower rates than other students. The move to a multiple-measure system will not solve this problem, but it will give all students multiple avenues to demonstrate their knowledge of the content necessary to graduate.



**“When teacher input is sought out and reflected in assessments and their implementation, tests will become an effective tool to accurately gauge student achievement and growth, as well as an empowering tool for the teachers to improve their teaching practices.”**

**IRIS WON**, *ninth- to 12th-grade mathematics and technology teacher, Renaissance High School for Musical Theater & Technology*



# Conclusion

As teachers, this is our vision for making full use of standardized assessments—for taking advantage of a powerful tool that requires careful execution.

Tests can be a force for good, and we would be unwise to throw them out of our toolbox. At the same time, they cannot be our only tool. We cannot use a hammer when a wrench is necessary, and we will usually need both.

Improving how tests are used is a shared responsibility. As teachers, we must do our part—administer tests with fidelity, use data to improve when it is available, and advocate for better assessments when necessary. But policymakers must also step up—they must provide us with the support we need, and they must make wise decisions about how often tests are administered and how results are used. This will take time, money, reflection, and a lot of work.

Let's get started.

# Summary

## KEY RESEARCH TAKEAWAYS

### DESIGN

IMPROVE THE ACCURACY  
OF STANDARDIZED  
ASSESSMENTS



#### KEY TAKEAWAYS

- Tests are useful, though imperfect, measures of students' learning and teachers' effectiveness.
- The accuracy of tests is directly related to test quality—well-designed assessments provide important information, but poorly designed tests have little to no use.

#### RECOMMENDATIONS

- When designing tests, follow best practices such as ensuring alignment to standards, testing higher-order thinking, and actively soliciting teacher input.
- Prioritize higher-order instruction, and eliminate excessive test preparation that does not contribute to meaningful learning.
- Use computer-adaptive assessments, which improve tests' accuracy by measuring the growth of low- and high-performing students.
- Release the vast majority of state test items publicly after the assessment window has closed so that all stakeholders can monitor the quality of the exams.



### ACCOUNTABILITY

INCLUDE DATA IN  
CRITICAL DECISIONS

#### KEY TAKEAWAYS

- Student achievement is a useful measure that should be a part of a multi-measure evaluation framework that holds teachers and schools accountable for student performance.
- Holding schools and teachers accountable for students' performance produces positive results.

#### RECOMMENDATIONS

- Isolate the effects of teachers and schools to ensure that those serving at-risk student populations are not penalized by out-of-school factors.
- Make high-stakes decisions based on multiple sources and multiple years of evidence.
- Evaluate teachers of non-tested subjects based on authentic assessments—developed and validated by teachers—using growth measures or student learning objectives.





## AND OVERVIEW OF RECOMMENDATIONS

### KEY TAKEAWAYS

- The toxic culture of testing that pervades some schools undermines the value of assessments and harms teachers' morale.
- A positive culture begins with viewing assessments as opportunities for growth, and also requires policymakers to create an environment—through support and thoughtful decision-making—that encourages a healthy culture.

### RECOMMENDATIONS

- Measure time spent, by both students and teachers, on testing, and eliminate unnecessary and redundant exams.
- Implement best practices, such as administering tests in controlled environments and monitoring for test irregularities, to prevent and detect cheating.
- Create or expand pilot programs of schools using nontraditional tests to determine whether they lead to positive results for students, and can be used to evaluate and support teachers and schools.



### KEY TAKEAWAYS

- When used properly, assessment data is valuable for improving teachers' practice, and provides helpful information to administrators, parents, and students.
- Teachers and administrators need more support in using data to inform their practice and ensure it is meaningful.

### RECOMMENDATIONS

- Offer high-quality training throughout the year for teachers on how to improve instruction using assessment data.
- Provide each school with a teacher who serves as a "data specialist."
- Ensure that teachers and administrators receive timely, detailed, and disaggregated data in a transparent, accessible format.



# Teacher Policy Team Process and Methodology

## IDENTIFYING E4E'S POLICY FOCUS

E4E surveyed members and held focus groups with E4E-NY members to determine the most important policy issues from teachers' perspective.

## OUR PROCESS

We met for eight weeks to review research on different facets of testing and assessment, particularly as they relate to New York City and State. We considered evidence from different perspectives, held small and large group discussions, and regularly challenged each other's thinking. We ended up with four main categories under which we elaborate upon specific recommendations.

# Notes

- <sup>1</sup> For one example, see: D. Ravitch. (2014, January 18). Do International Test Scores Matter? (Weblog post). Retrieved from <http://dianeravitch.net/2014/01/18/do-international-test-scores-matter/> (“Readers of this blog know that I have repeatedly argued that standardized scores on international tests predict nothing about the future.”)
- <sup>2</sup> Short, A., Campanile, C. (2014, April 9). Bloomberg-era tests no longer top criteria for student promotion: Fariña. *New York Post*. Retrieved from <http://nypost.com/2014/04/09/city-scrap-bloombergs-standardized-tests/>
- <sup>3</sup> Sackett, P.R., Kuncel, N.R., Beatty, A.S., et al. (2012, April 2). The Role of Socioeconomic Status in SAT-Grade Relationships and in College Admissions Decisions. *Psychological Science*, 23(9), 1000-1007. doi: 10.1177/0956797612438732
- <sup>4</sup> Schmitt, N., Keeney, J., Oswald, F.L., et al. (2009, November). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, 94(6), 1479-97. doi: 10.1037/a0016810.
- <sup>5</sup> Robertson, K.F., Smeets, S., Lubinski, D., et al. (2010, December). Beyond the Threshold Hypothesis Even Among the Gifted and Top Math/Science Graduate Students, Cognitive Abilities, Vocational Interests, and Lifestyle Preferences Matter for Career Choice, Performance, and Persistence. *Current Directions in Psychological Science*, 19(6), 346-51. doi: 10.1177/0963721410391442
- <sup>6</sup> Kuncel, N.R., Hezlett, S.A. (February, 2007). Standardized Tests Predict Graduate Students’ Success. 315(5815). DOI: 10.1126/science.1136618
- <sup>7</sup> Hanushek, E.A., Jamison, D.T., Jamison, E.A., et al. (2008, Spring). Education and Economic Growth. *Education Next*, 8(2). Retrieved from <http://educationnext.org/education-and-economic-growth/>
- <sup>8</sup> Chetty, R., Friedman, J.N, Rockoff, J.E. (2011). The Long-Term Impact of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*. Retrieved from [http://obs.rc.fas.harvard.edu/chetty/value\\_added.html](http://obs.rc.fas.harvard.edu/chetty/value_added.html)
- <sup>9</sup> Master, J. (2014, June). Staffing for Success. *Education Evaluation and Policy Analysis*. Retrieved from <http://epa.sagepub.com/content/36/2/207.abstract?rss=1>
- <sup>10</sup> Jennings, J.L., DiPrete, T.A. (2009, March 15). Teacher Effects on Social/Behavioral Skills in Early Elementary School. Retrieved from [http://www.columbia.edu/~td61/Jennings%20and%20DiPrete\\_3\\_15\\_2009\\_Final.pdf](http://www.columbia.edu/~td61/Jennings%20and%20DiPrete_3_15_2009_Final.pdf)
- <sup>11</sup> Adland, J., Braslow, D., Brosbe, R., et al. (Spring, 2011). Beyond Satisfactory: A New Teacher Evaluation System for New York. Retrieved from [http://educators4excellence.s3.amazonaws.com/8/3f/b/1362/E4E\\_Evaluation\\_Paper\\_Final.pdf](http://educators4excellence.s3.amazonaws.com/8/3f/b/1362/E4E_Evaluation_Paper_Final.pdf)
- <sup>12</sup> Barraclough, N., Farnum, C., Loeb, M., et al. (Spring, 2014). A Path Forward: Recommendations from the classroom for effectively implementing the Common Core. Retrieved from [http://educators4excellence.s3.amazonaws.com/8/0b/a/2258/03.24.14\\_TAT\\_CCSS\\_Memo.pdf](http://educators4excellence.s3.amazonaws.com/8/0b/a/2258/03.24.14_TAT_CCSS_Memo.pdf)
- <sup>13</sup> New York State Department of Education. (2014, July 9). New York State Education Department Test Development Process. Retrieved from <http://www.p12.nysed.gov/assessment/teacher/home.html#process>
- <sup>14</sup> See for example: Phillips, E. (2014, April 9). We Need to Talk About the Test: A Problem With the Common Core. *The New York Times*. Retrieved from <http://www.nytimes.com/2014/04/10/opinion/the-problem-with-the-common-core.html>; and Hartocollis, A., (2012, April 20). When Pineapple Races Hare, Students Lose, Critics of Standardized Tests Say. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/04/21/nyregion/standardized-testing-is-blamed-for-question-about-a-sleeveless-pineapple.html?pagewanted=all>
- <sup>15</sup> King, F.J., Goodson, L., Rohani, F. Higher Order Thinking Skills. Center for Advancement of Learning and Assessment. Retrieved from [http://www.cala.fsu.edu/files/higher\\_order\\_thinking\\_skills.pdf](http://www.cala.fsu.edu/files/higher_order_thinking_skills.pdf)
- <sup>16</sup> Newmann, F.M., Bryk, A.S., Nagaoka, J. (2001, January). Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence? Retrieved from <http://ccsr.uchicago.edu/publications/authentic-intellectual-work-and-standardized-tests-conflict-or-coexistence>
- <sup>17</sup> UChicagoNews. (2008, May 27). Intensive ACT test prep during class leads to lower scores; students don’t connect grades, study habits to exam scores. Retrieved from <http://news.uchicago.edu/article/2008/05/27/intensive-act-test-prep-during-class-leads-lower-scores-students-don-t-connect-gr>
- <sup>18</sup> Rafter, D. (2014, January 2). De Blasio picks a schools chancellor. *Queens Chronicle*. Retrieved from [http://www.qchron.com/editions/queenswide/de-blasio-picks-a-schools-chancellor/article\\_687e9c54-a168-54a6-9df7-ebed13034cc2.html](http://www.qchron.com/editions/queenswide/de-blasio-picks-a-schools-chancellor/article_687e9c54-a168-54a6-9df7-ebed13034cc2.html)
- <sup>19</sup> Spector, J. (2014, March 24). John King on upcoming Common Core tests: “The best preparation for testing is good teaching.” *Politics on the Hudson*. Retrieved from <http://polhudson.lohudblogs.com/2014/03/24/john-king-upcoming-common-core-tests-best-preparation-testing-good-teaching/>
- <sup>20</sup> Graduate Record Examinations. How the test is scored. Retrieved from [https://www.ets.org/gre/revised\\_general/scores/how/](https://www.ets.org/gre/revised_general/scores/how/)
- <sup>21</sup> Graduate Management Admission Test. (2010, January 13). The CAT in the GMAT. Retrieved from <http://www.mba.com/us/the-gmat-blog-hub/the-official-gmat-blog/2010/jan/the-cat-in-the-gmat.aspx>
- <sup>22</sup> Smarter Balanced Assessment Consortium. Computer Adaptive Testing. Retrieved from <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Smarter-Balanced-CAT.pdf>
- <sup>23</sup> Brown, E. (2014, March 2). D.C. mulling over Common Core test switch. *The Washington Post*. Retrieved from [http://www.washingtonpost.com/local/education/dc-mulling-over-common-core-test-switch/2014/03/02/29478710-a0b3-11e3-a050-dc3322a94fa7\\_story.html?wpss=rss\\_education](http://www.washingtonpost.com/local/education/dc-mulling-over-common-core-test-switch/2014/03/02/29478710-a0b3-11e3-a050-dc3322a94fa7_story.html?wpss=rss_education)
- <sup>24</sup> Strauss, V. (2014, April 25). AFT asks Pearson to stop ‘gag order’ barring educators from talking about tests. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/04/25/aft-asks-pearson-to-stop-gag-order-barring-educators-from-talking-about-tests/>
- <sup>25</sup> Times Union. (2014). Times Union/Siena College Poll [Data File]. Retrieved from <http://www.timesunion.com/7dayarchive/item/Times-Union-Siena-College-education-poll-30096.php>
- <sup>26</sup> We say this with the understanding that it may not be possible for 100% of all items to be released publicly. We are comfortable with a small number of items—no more than 10%—being held from public view to ensure comparability from tests year to year.
- <sup>27</sup> McIntire, M.E.. (2014, June 11) As Pearson’s annual field testing ends, some want them never to start again. *Chalkbeat*. Retrieved from <http://ny.chalkbeat.org/2014/06/11/as-pearsons-annual-field-testing-ends-some-want-them-never-to-start-again/#.U8Le9FNyjec>
- <sup>28</sup> Roediger, H.L. and Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*. Retrieved from [http://learninglab.psych.purdue.edu/downloads/2006\\_Roediger\\_Karpicke\\_PsychSci.pdf](http://learninglab.psych.purdue.edu/downloads/2006_Roediger_Karpicke_PsychSci.pdf)
- <sup>29</sup> Litow, S.S., Flanagan, J., Nolan, C., et al. (2014, March). Putting Students First: Common Core Implementation Panel Recommendation Report to Governor Andrew M. Cuomo. Retrieved from [http://www.governor.ny.gov/sites/default/files/Common\\_Core\\_Implementation\\_Panel\\_3-10-14.pdf](http://www.governor.ny.gov/sites/default/files/Common_Core_Implementation_Panel_3-10-14.pdf)
- <sup>30</sup> S. 6356D. (2013). Retrieved from <http://open.nysenate.gov/legislation/bill/A8556d-2013>
- <sup>31</sup> *Ibid.*
- <sup>32</sup> Resmovits, J. (2011, August 8). Schools Caught Cheating in Atlanta, Around the Country. *The Huffington Post*. Retrieved from [http://www.huffingtonpost.com/2011/08/08/atlanta-schools-cheating-scandal-ripples-across-country\\_n\\_919509.html](http://www.huffingtonpost.com/2011/08/08/atlanta-schools-cheating-scandal-ripples-across-country_n_919509.html)
- <sup>33</sup> Alpert, T., Amrein-Beardsley, A., Bruce, W., et al. (2013). Testing Integrity Symposium: Issues and Recommendations for Best Practice. Symposium conducted at meeting of U.S. Department of Education.



# Notes

- <sup>34</sup> Alpert et al. (2013)
- <sup>35</sup> Alpert et al. (2013)
- <sup>36</sup> Alpert et al. (2013)
- <sup>37</sup> New York Performance Standards Consortium. Retrieved from <http://performanceassessment.org/index.html>
- <sup>38</sup> Educating for the 21st Century: Data Report on the New York Performance Standards Consortium. Retrieved from [http://www.nycl.org/files/releases/testing\\_consortium\\_report.pdf](http://www.nycl.org/files/releases/testing_consortium_report.pdf)
- <sup>39</sup> Mathews, J. (2004, Summer). Portfolio Assessment: Can it be used to hold schools accountable? *Education Next*, 4(3). Retrieved from <http://educationnext.org/portfolio-assessment/>
- <sup>40</sup> Wayman, J.C. (2005). Involving Teachers in Data-Driven Decision Making: Using Computer Data Systems to Support Teacher Inquiry and Reflection. *Journal of Education for Students Placed at Risk*, 10(3), 295–308. Retrieved from <http://myclass.nl.edu/tie/tie533/teacherdatause.pdf>
- <sup>41</sup> Wayman. (2005)
- <sup>42</sup> Light, D., Honey, M., Heinze, J. (2005, January). Linking Data and Learning: The Grow Network Study. Center for Children and Technology. Retrieved from <http://cct.edc.org/publications/linking-data-and-learning-grow-network-study>
- <sup>43</sup> Chen, E., Heritage, M., Lee, J. (2005). Identifying and Monitoring Students' Learning Needs with Technology. *Journal of Education for Students Placed at Risk*, 10(3), 309–322. Retrieved from [http://www.tandfonline.com/doi/abs/10.1207/s15327671espr1003\\_6#.U4ijD1Nyjec](http://www.tandfonline.com/doi/abs/10.1207/s15327671espr1003_6#.U4ijD1Nyjec)
- <sup>44</sup> Datnow, A., Park, V., Wohlstetter, P. (2007). Achieving with Data: How high-performing school systems use data to improve instruction for elementary students. Retrieved from <http://www.newschools.org/files/AchievingWithData.pdf>
- <sup>45</sup> Fryer, R.G. (2012, September). Learning from the Successes and Failures of Charter Schools. Retrieved from [http://scholar.harvard.edu/files/fryer/files/haUsingmilton\\_project\\_paper\\_2012.pdf](http://scholar.harvard.edu/files/fryer/files/haUsingmilton_project_paper_2012.pdf)
- <sup>46</sup> Data Quality Campaign. (2012, January). Retrieved from [http://www.dataqualitycampaign.org/files/1357\\_DQC-TE-primer.pdf](http://www.dataqualitycampaign.org/files/1357_DQC-TE-primer.pdf)
- <sup>47</sup> Tyler, J.H. (2013). If You Build it Will They Come? Teachers' Online Use of Student Performance Data. *Education Finance and Policy*, 8(2), 168–207. [http://www.mitpressjournals.org/doi/abs/10.1162/EDFP\\_a\\_00089#.U4indNyjec](http://www.mitpressjournals.org/doi/abs/10.1162/EDFP_a_00089#.U4indNyjec)
- <sup>48</sup> McCaffrey, D.F., Hamilton, L.S. (2007). Value-Added Assessment in Practice: Lessons from the Pennsylvania Value-Added Assessment System Pilot Project. Retrieved from [http://www.rand.org/content/dam/rand/pubs/technical\\_reports/2007/RAND\\_TR506.sum.pdf](http://www.rand.org/content/dam/rand/pubs/technical_reports/2007/RAND_TR506.sum.pdf)
- <sup>49</sup> United Federation of Teachers. Repurposed workday. Retrieved from <http://www.uft.org/proposed-contract/repurposed-workday>
- <sup>50</sup> Consentino, L., D'Amico, J., Fazio, C., et al. (Spring 2014). A Passing Grade: Teachers Evaluate the NYC Contract. Retrieved from <http://www.educators4excellence.org/nycontract/report>
- <sup>51</sup> Data Quality Campaign. (2014, February). Teacher Data Literacy: It's About Time. Retrieved from <http://www.dataqualitycampaign.org/files/DQC-Data%20Literacy%20Brief.pdf>
- <sup>52</sup> Smarter Balanced Assessment Consortium. Retrieved from <http://www.smarterbalanced.org/>
- <sup>53</sup> Smarter Balanced Assessment Consortium. Member States. Retrieved from <http://www.smarterbalanced.org/about/member-states/>
- <sup>54</sup> PARCC. PARCC Online. Retrieved from <http://www.parcconline.org/>
- <sup>55</sup> PARCC. PARCC States. Retrieved from <http://www.parcconline.org/parcc-states>
- <sup>56</sup> Gewertz, C. (2014, March 21). Field-testing Set to Begin on Common Core Exams. *Education Week*. Retrieved from [http://www.edweek.org/ew/articles/2014/03/21/26fieldtests\\_ep.h33.html](http://www.edweek.org/ew/articles/2014/03/21/26fieldtests_ep.h33.html)
- <sup>57</sup> PARCC. New York. Retrieved from <https://www.parcconline.org/new-york>
- <sup>58</sup> Ed Week. (2014, May 19). The National K–12 Testing Landscape. Retrieved from <http://www.edweek.org/ew/section/multimedia/map-the-national-k-12-testing-landscape.html>
- <sup>59</sup> PARCC. PARCC Task Prototypes and Sample Questions. Retrieved from <http://www.parcconline.org/samples/item-task-prototypes>
- <sup>60</sup> PARCC. Item Development. Retrieved from <http://www.parcconline.org/assessment-development>
- <sup>61</sup> Brown, E. (2014, March 2). D.C. Mulling Over Common Core Test Switch. *The Washington Post*. Retrieved from [http://www.washingtonpost.com/local/education/dc-mulling-over-common-core-test-switch/2014/03/02/29478710-a0b3-11e3-a050-dc3322a94fa7\\_story.html?wprss=rss\\_education](http://www.washingtonpost.com/local/education/dc-mulling-over-common-core-test-switch/2014/03/02/29478710-a0b3-11e3-a050-dc3322a94fa7_story.html?wprss=rss_education)
- <sup>62</sup> PARCC. (2013, August). PARCC Fact Sheet and FAQs. Retrieved from <http://www.parcconline.org/sites/parcc/files/PARCCFactSheetandFAQsBackgroundFINAL.pdf>
- <sup>63</sup> Smarter Balanced Assessment Consortium. Computer Adaptive Testing. Retrieved from <http://www.smarterbalanced.org/smarter-balanced-assessments/computer-adaptive-testing/>
- <sup>64</sup> Alpert et al. (2013)
- <sup>65</sup> PARCC. (2013, August). PARCC Fact Sheet and FAQs. Retrieved from <http://www.parcconline.org/sites/parcc/files/PARCCFactSheetandFAQsBackgroundFINAL.pdf>
- <sup>66</sup> PARCC. Practice Tests. Retrieved from <http://www.parcconline.org/practice-tests>
- <sup>67</sup> Hanushek, E.A., Raymond, M.E. (2005). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management*, 24(2), 297–327. Retrieved from <http://hanushek.stanford.edu/sites/default/files/publications/hanushek%20Raymond.2005%20jpm%2024-2.pdf>
- <sup>68</sup> Chiang, H. (2009, October). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93(9–10), 1045–57. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0047272709000693>
- <sup>69</sup> Rouse, C.E., Hannaway, J., Goldhaber, D., et al. (2013, May). Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure. *American Economic Journal: Economic Policy*, 5(2), 251–81. Retrieved from <http://www.aeaweb.org/articles.php?doi=10.1257/pol.5.2.251>
- <sup>70</sup> All of these studies measure scores based on assessments other than the state exam, so cheating, gaming, or test prep cannot explain these results.
- <sup>71</sup> Rockoff, J.E., Staiger, D.O., Kane, T.J., et al. (2010, July). Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools. *The National Bureau of Economic Research Working Paper No. 16240*. Retrieved from <http://www.nber.org/papers/w16240>
- <sup>72</sup> Deming, D.J., Cohodes, S., Jennings, J., et al. (2013, September). School Accountability, Postsecondary Attainment and Earnings. *The National Bureau of Economic Research Working Paper No. 19444*. Retrieved from <http://www.nber.org/papers/w19444>
- <sup>73</sup> DiCarlo, M. (2010, July 14). Teachers Matter, But So Do Words. (Weblog). Retrieved from <http://shankerblog.org/?p=74>
- <sup>74</sup> Di Carlo, M. (2012, February 2). The Perilous Conflation of Student and School Performance. (Weblog). Retrieved from <http://shankerblog.org/?p=4980>
- <sup>75</sup> Di Carlo, M. (2013, October 3). Are There Low-Performing Schools With High-Performing Students? (Weblog). Retrieved from <http://shankerblog.org/?p=8887>
- <sup>76</sup> Value-Added Modeling 101. (2012, September). *Rand Education*. Retrieved from [www.rand.org/education/projects/measuring-teacher-effectiveness/value-added-modeling.html](http://www.rand.org/education/projects/measuring-teacher-effectiveness/value-added-modeling.html)
- <sup>77</sup> McCaffrey, D. (2012, October 15). Do Value-Added Methods Level the Playing Field for Teachers? *Carnegie Knowledge Network*. Retrieved from <http://www.carnegieknowledgenetwork.org/briefs/value-added/level-playing-field/>

# Notes

- <sup>78</sup> Baker, E., Barton, P., et al. (2010, August 27). Problems with the use of student test scores to evaluate teachers. Economic Policy Institute. Retrieved from <http://www.epi.org/publication/bp278/>
- <sup>79</sup> Glazerman, S., Loeb, S., et al. (2010, November 17). Evaluating Teachers: The Important Role of Value-Added. Brown Center on Education Policy at Brookings. Retrieved from [http://www.brookings.edu/~media/research/files/reports/2010/11/17%20evaluating%20teachers/1117\\_evaluating\\_teachers.pdf](http://www.brookings.edu/~media/research/files/reports/2010/11/17%20evaluating%20teachers/1117_evaluating_teachers.pdf)
- <sup>80</sup> Di Carlo, M. (2010, December 7). The War on Error. (Weblog) Retrieved from <http://shankerblog.org/?p=1383>
- <sup>81</sup> Chetty, R., Friedman, J., Rockoff, J. (2011, December). The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w17699>
- <sup>82</sup> Stern, G. (2013, October 15). N.Y.'s Teacher Evaluation Faulted in Study. The Journal News. Retrieved from <http://archive.lohud.com/article/20131015/NEWS/310150042/N-Y-s-teacher-evaluations-faulted-study>
- <sup>83</sup> Lower Hudson Council of School Superintendents. (2013, October). Review and Analysis of the New York State Growth Model. Retrieved from <http://www.lhcss.org/positionpapers/nysgrowthmodel.pdf>
- <sup>84</sup> Decker, G. (2013, June 18). State to Use Value-Added Growth Model without Calling it That. Chalkbeat. Retrieved from <http://ny.chalkbeat.org/2013/06/18/state-to-use-a-value-added-growth-model-without-calling-it-that/#.U61JpVNyjec>
- <sup>85</sup> Ehlers, M., Koedel, C., Parsons, E., et al. Selecting Growth Measures for School and Teacher Evaluations: Should Proportionality Matter? National Center for Analysis of Longitudinal Data in Education Research. Retrieved from <http://www.caldercenter.org/publications/upload/wp-80-updated-v3.pdf>
- <sup>86</sup> Lankford, H., Loeb, S., Wyckoff, J. (2002, March). Teacher Sorting and the Plight of Urban Schools. Education Evaluation and Policy Analysis. Retrieved from <http://epa.sagepub.com/content/24/1/37.short>
- <sup>87</sup> Koedel, C. (2014, May 27). The Proportionality Principle in Teacher Evaluation. Shanker Blog. Retrieved from <http://shankerblog.org/?p=9924>
- <sup>88</sup> Cramer, P., Decker, G. (2013, September 16). Instead of Telling Teachers Apart, New Eval Lumps Some Together. Chalkbeat. Retrieved from <http://ny.chalkbeat.org/2013/09/16/instead-of-telling-teachers-apart-new-evals-lump-some-together/#.U4-VpFNyjec>
- <sup>89</sup> Di Carlo, M. (2012, May 29). We Should Only Hold Schools Accountable for Outcomes They Can Control. (Weblog). Retrieved from <http://shankerblog.org/?p=5959>
- <sup>90</sup> We distinguish this practice from evaluation systems that have school-wide rating components, meaning that all teachers in a school are judged by a school's overall components. This practice has several pros and cons; in this paper, we do not take a position on it.
- <sup>91</sup> Decker, G. (2014, May 14). Appeal Process in New Evaluation Plan Shifts Weight from Student Scores for Some. Chalkbeat. Retrieved from <http://ny.chalkbeat.org/2014/05/14/appeal-process-in-new-evaluation-plan-shifts-weight-from-student-scores-for-some/#.U4-ViVNyjec>
- <sup>92</sup> Goldstein, D. (2012, June 13). No More Ditching Gym Class. Slate. Retrieved from [http://www.slate.com/articles/double\\_x/doublex/2012/06/standardized\\_tests\\_for\\_the\\_arts\\_is\\_that\\_a\\_good\\_idea\\_.html](http://www.slate.com/articles/double_x/doublex/2012/06/standardized_tests_for_the_arts_is_that_a_good_idea_.html)
- <sup>93</sup> EngageNY. Overview of Student Learning Objectives. Retrieved from [http://www.engageny.org/sites/default/files/resource/attachments/overview\\_of\\_student\\_learning\\_objectives.pdf](http://www.engageny.org/sites/default/files/resource/attachments/overview_of_student_learning_objectives.pdf)
- <sup>94</sup> For one example, among many others, of this argument, see: <https://www.aft.org/pdfs/teachers/devmultiplemeasures.pdf>
- <sup>95</sup> Worrell, C. (2013, October 25). In Teacher Evaluations, Student Data and Multiple Measures Show Progress. Data Quality Campaign. Retrieved from <http://www.dataqualitycampaign.org/blog/2013/10/in-teacher-evaluations-student-data-show-progress/>
- <sup>96</sup> New York City Department of Education. NY State Policy Context: Education Law 3012-c. Retrieved from <http://schools.nyc.gov/Offices/advance/Background/Policy+Context/default.htm>
- <sup>97</sup> New York City Department of Education. (2014, April 9). Chancellor Fariña Announces New Promotion Policy for Students in Grades 3-8. Retrieved from <http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2013-2014Chancellor+Fari%C3%B1a+Announces+New+Promotion+Policy+for+Students+in+Grades+3-8.htm>
- <sup>98</sup> New York State Department of Education. (2013, June). Diploma/Credential Requirements. Retrieved from <http://www.p12.nysed.gov/ciai/gradreq/diploma-credential-summary.pdf>
- <sup>99</sup> Wall, P. (2013, November 14). Tougher Diploma Rules Leave Some Students in Graduation Limbo. Chalkbeat. Retrieved from <http://ny.chalkbeat.org/2013/11/14/tougher-diploma-rules-leave-some-students-in-graduation-limbo/#.U44Jm1Nyjec>
- <sup>100</sup> Jacob, B. (2001, June). Getting Tough: The Impact of High School Graduation Exam. Educational Evaluation and Policy Analysis. Retrieved from <http://epa.sagepub.com/content/23/2/99.short>
- <sup>101</sup> Marchant, G., Paulson, S., (2005, January). The Relationship of High School Graduation Exams to Graduation Rates and SAT Scores. Education Policy Analysis Archives. Retrieved from <http://files.eric.ed.gov/fulltext/EJ846516.pdf>
- <sup>102</sup> Grodsky, E., Warren, J., Kalogrides, D. (2009, May). State High School Exit Examinations and NAEP Long-Term Trends in Reading and Mathematics, 1971-2004. Education Policy. Retrieved from <http://epx.sagepub.com/content/early/2008/06/13/0895904808320678.abstract>
- <sup>103</sup> Reardon, S., Arshan, N., Atteberry, A., Kurlaender, M. (2010, December). Effects of Failing a High School Exit Exam on Course Taking, Achievement, Persistence, and Graduation. Education Evaluation and Policy Analysis. Retrieved from <http://epa.sagepub.com/content/32/4/498>
- <sup>104</sup> Baker, O., Lang, K. (2013, June). The Effect of High School Exit Exams on Graduation, Employment, Wages, and Incarceration. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w19182>
- <sup>105</sup> Darling-Hammond, L., Rustique-Forrester, E., Pecheone, R. (2005). Multiple Measures Approaches to High School Graduation. The School Redesign Network at Stanford University. Retrieved from <https://edpolicy.stanford.edu/sites/default/files/publications/multiple-measures-approaches-high-school-graduation.pdf>
- <sup>106</sup> New York State Education Department. (2014). Graduation Rate Data. Retrieved from <http://www.p12.nysed.gov/irs/pressRelease/20140623/home.html>

## THE 2014 EDUCATORS 4 EXCELLENCE NEW YORK TEACHER POLICY TEAM ON TESTING AND ASSESSMENT

Trevor Baisden

*Founding Fifth-Grade ELA and History Lead Teacher  
Success Academy Bronx 2 Middle School*

Elizabeth Barrett-Zahn

*Kindergarten to Fifth-Grade Science Facilitator  
Columbus Elementary School, New Rochelle*

Rachael Beseda

*First-Grade Special Education Teacher  
Global Community Charter School*

Ezekiel Cruz

*Ninth- to 12th-Grade Social Studies Teacher  
Manhattan Bridges High School*

Suraj Gopal

*Ninth-Grade STEM Special Education Teacher  
Hudson High School of Learning Technologies*

Vivett Hemans

*English and Language Arts Teacher  
Eagle Academy for Young Men of Southeast Queens*

Maura N. Henry

*Sixth- to 12th-Grade ESL Teacher  
The Young Women's Leadership School of Astoria*

Michelle Kniffin

*Ninth- to 12th-Grade Math Teacher  
High School of Telecommunication Arts and Technology*

Jason Koo

*Math Teacher  
Albert Einstein Junior High School I.S. 131*

Christine Montera

*Social Studies Teacher  
East Bronx Academy for the Future*

Liliana Ruiz

*Sixth- to Eighth-Grade Special Education Teacher  
Bea Fuller Rodgers School I.S. 528*

Charlotte Steel

*Seventh-Grade Math Teacher  
Booker T. Washington M.S. 54*

Blackfoot U-Ahk

*Fourth- and Fifth-Grade Teacher of Students  
with Severe Emotional Disabilities  
Coy L. Cox School P.369k*

Iris Won

*Ninth- to 12th-Grade Mathematics and Technology Teacher  
Renaissance High School for Musical Theater & Technology*

---

*This report, graphics, and figures were designed by Kristin Girvin Redman and Tracy Harris at Cricket Design Works in Madison, Wisconsin.*

*The text face is Bembo Regular, designed by Stanley Morison in 1929. The typefaces used for headers, subheaders, and pull quotes are Futura Bold, designed by Paul Renner, and Museo Slab, designed by Jos Buivenga. Figure labels are set in Futura Regular, and figure callouts are set in Museo Slab.*



**For far too long,** education policy has been created without a critical voice at the table—the voice of classroom teachers.

Educators 4 Excellence (E4E), a teacher-led organization, is changing this dynamic by placing the voices of teachers at the forefront of the conversations that shape our classrooms and careers.

E4E has a quickly growing national network of educators united by our Declaration of Teachers' Principles and Beliefs. E4E members can learn about education policy and research, network with like-minded peers and policymakers, and take action by advocating for teacher-created policies that lift student achievement and the teaching profession.

Learn more at [Educators4Excellence.org](https://Educators4Excellence.org).

