

**MAKING A  
MEASURABLE  
DIFFERENCE:** CASE STUDIES FROM  
THE HIGH-QUALITY ASSESSMENT PROJECT

STRENGTHENING  
THE ASSESSMENT  
TRAJECTORY:  
ENGAGING EDUCATORS  
IN THE POLICY PROCESS

States are now using higher quality tests to better measure important student skills such as writing, problem-solving and critical thinking. This case study captures advice from teacher leaders who reviewed the quality of the tests and discusses how the experience affected their teaching. This brief also offers input on how states can engage educators in similar reviews.

PREPARED BY



Dr. Catherine McClellan, Ph.D.  
Dr. Rebecca Snyder, Ed.D.

Dr. Maryann Woods-Murphy, Ed.D.  
Ms. Katherine Bassett, M.Ed.

**INTRODUCTION:** Great teachers recognize great assessments. As policy and education leaders work to make sure state tests are measuring the problem-solving, writing and critical thinking skills students need for success, they should convene and rely on teachers to review test quality and help answer the question: Do the questions on our state test reflect excellent instruction?

Nearly every state has reworked its summative assessment over the past few years to try to better measure skills and knowledge that prepare students for success after high school. Two large groups of states collaborated to create the PARCC and Smarter Balanced assessments, which were designed to be more innovative and of higher quality than states could produce on their own. With so many states using PARCC and Smarter Balanced as their new state assessments, the National Network of State Teachers of the Year (NNSTOY) has been gathering teacher leaders from around the country to put these claims to the test, comparing these two new state tests against older state tests. These teachers—publicly recognized as exceptional classroom instructors—looked carefully at how well test questions reflected skills and knowledge their students should be learning and how the tests could be improved.

**ON THE RIGHT TRAJECTORY** NNSTOY conducted three research studies into the quality of state assessments, including the quality of the new PARCC and Smarter Balanced assessments used by many states. Using methodologies and practices such as Evidence-Centered Design, Webb’s Depth of Knowledge and custom survey instruments, panels examined consortia and state assessments at grades 5 and 11. Each study panel was composed of state and national Teachers of the Year and finalists for State Teacher of the Year.

**THE RIGHT TRAJECTORY 2015:**

Teachers with experience in grade 5 reviewed the consortia assessments and compared them to previous state tests from Delaware, Illinois, New Hampshire and New Jersey.

**STILL ON THE RIGHT TRAJECTORY**

**2016:** Teachers reviewed two more prior state assessments (Nevada and Oregon) and Smarter Balanced grade 5.

The two grade 5 panels found that the new consortia assessments better reflect the range of reading and math knowledge and skills, as well as the full range of cognitive complexity. These assessments better align with strong instructional practices, providing information relevant to a wide range of student performers. And while the new consortia assessments are more rigorous and demanding, they are grade-level appropriate, and even more so than prior state tests. These expert teachers believe that there are challenges still to be overcome, but the new assessments represent movement in the right direction.

**BEGINNING A  
HIGHER TRAJECTORY**

**2016:** Teachers with experience in grade 11 reviewed only the Smarter Balanced assessment at that grade level (prior state tests generally tested at grade 10).

At grade 11, the panel examined only the Smarter Balanced assessment because there were no previous state assessments to use for comparisons. These educators found many things to like about the assessment, and they made recommendations for continuous improvement.

To encourage teachers to reflect on the value of this review and lessons learned, NNSTOY reconvened 24 of the participating teachers (in three groups) in Spring 2017. The participants came from all over the country: Colorado, Washington D.C., Department of Defense, Georgia, Idaho, Illinois, Minnesota, Nevada, New Jersey, New York, Oregon, South Dakota, Utah and Washington State. Some of their states use the consortia tests reviewed; some do not. All of them brought to the table excellence in teaching, passion for education and profound knowledge of and care for their students

All participants were eager to share their thoughts and felt that their earlier study experiences were valuable. Many volunteered to collaborate with policymakers to replicate these studies in their own locales. They told us that when we fail to include the view from the classroom, assessments may trade valuable instructional time for testing of low-level skills and factual recall. Students may be bored or disengaged with questions that are disconnected from their classroom experiences, don't make them think or challenge their understanding. Score reports arrive too late and lack sufficient detail to improve education. Each participant was clear: teachers' voices need to be heard and heeded as student assessments are built and administered.

#### **FIVE LESSONS EMERGED:**

1. Seeing assessment through “students’ eyes” changes instruction for the better.
2. Examining state tests helps educators inform the design and use of assessment throughout the education system.
3. Fewer, high quality assessments that align with excellent instructional practices are worth the investment.
4. By contributing first-hand knowledge of students and instruction, teacher perspectives can help strengthen state—and local—assessments.
5. More teachers should have opportunities to participate in assessment development and reviews.

***This report gathers their main messages for administrators, policymakers, advocates and parents about improving quality of student assessments and better engaging educators<sup>1</sup>.***

*“I want policymakers to know that the assessment should measure what is most important for students to get out of school. And, if we think that students should get more than basic skills and knowledge, we need an assessment that measures those things. And then that will drive instruction to develop more capabilities in students.”*

1. Throughout this case study you'll find quotes from some of the teachers that participated in the state test reviews and these subsequent discussions with NNSTOY about lessons learned. These statements have been edited for clarity and brevity.

## LESSON 1: “SEEING ASSESSMENT THROUGH STUDENTS’ EYES” CHANGES INSTRUCTION FOR THE BETTER.

Although the participants already experienced and successful teachers, they explained how the opportunity to look at state assessment questions more carefully and see the complete tests that students take improved their teaching practices back in the classroom. By being part of review panels, participants could put themselves in the students’ shoes and take the same tests the students face—reminding them of what the experience is really like—and what support and teaching students need to succeed at meeting the higher expectations embedded in the new tests. Panelists reflected on how their and other adults’ experiences with assessments differ from the students’. Teachers and administrators are anxious about the outcomes, especially when professional evaluations are affected. Teachers feel self doubt about whether they did the best job preparing students to demonstrate what they know and can do on these particular assessments. Parents and teachers worry that the assessment is too long, doesn’t measure the things most valued nor uses time better spent in instruction. Everyone feels sympathy when students become stressed and anxious about their performance. Teachers taking the tests highlighted the experience of current students taking the test and what instruction can do to improve it. The participating teachers were reminded that the new tests are more rigorous and require more thinking and problem-solving than even recent versions of state tests did—much less the kinds of tests their parents or teachers took.

*“I definitely think it was useful to go through the new state tests... It helps you empathize with your kids about how long the test takes and how difficult these questions are.”*

*“The new tests require you to go deeper in your teaching. You have to. You have to teach problem-solving and critical thinking (which I call real thinking). Otherwise, students are being cheated.”*

Much is asked of students now, and classroom instruction should prepare them to respond confidently. The tests are difficult, often assessing multiple content standards in a single item. Items often have multiple parts. Questions that demand depth of knowledge mean that students must persist, push through, make connections and find insights. Students are not just choosing option C or the longest answer to a multiple-choice question anymore. Most items require a student to construct an answer, not just select one. Teachers must ask complex questions in class, provide experience struggling with problems and use language that prepares students for the challenges they face.

Panelists told us that when they returned to the classrooms with the insights from the study results in mind, they pushed themselves and their students deeper into the material they teach, demanding more engagement and providing increased cognitive challenge.

They reduced or eliminated instructional time spent preparing students for “basic skills” assessments. Because of these insights, teachers who participated in the study recognize that classroom lessons involving critical thinking and problem-solving best prepare students for new forms of assessment, rather than traditional testing tips and strategies. In some classrooms, project-based learning, portfolio assessment and mastery-based assessments have taken the place of traditional classroom tests, which do little to prepare students to succeed on the new tests.

*“I really feel like the new tests match the rigor that we are now putting into our lessons, much more than 20 years ago.”*

*“I think participating in this review of state tests was the best professional development of my career. . . . Going through actual tests changed my practice because of this experience. If other teachers can go through this sort of review, then they can create changes in their classrooms.”*

Many teachers also reported they increased focus on project work and team assignments. They asked questions at deeper levels of knowledge and understanding. Complex and demanding texts were selected as the basis of discussion and analysis. Assignments that spanned multiple standards were given and discussed. Persistence was encouraged and rewarded. The joy of struggling with and solving a really difficult problem was incorporated into everyday learning.

## LESSON 2: EXAMINING STATE TESTS HELPS EDUCATORS INFORM THE DESIGN AND USE OF ASSESSMENT THROUGHOUT THE EDUCATION SYSTEM.

In addition to affecting instructional practices, the experience changed panelists' discussions with other teachers, parents and administrators. They shared their perspective with their colleagues and stakeholders. The experience of evaluating these assessments altered how the participants talked about and advocated for good assessments, as well as helped them construct stronger arguments for students to take the tests and not opt out. Questions about alignment to standards or appropriateness of the assessment for students were answered from direct knowledge.

*“As an educator, actually participating and taking the test helps me feel the rigor that is expected. I know what it’s about.”*

Participants also relayed that they began changing the types of assessments they were using in the classroom to regularly check student progress. For example, one teacher challenged their students to write their own test items and then discussed and critiqued the suggestions. Other teachers shared how they started using assessments with their students that were more similar to the open-ended or constructed-response questions the PARCC and Smarter Balanced tests use. Students were tasked to explain how and why they know something in new ways. As questions with similar format, language and structuring of the tasks to the state assessments make their way into the tests students see frequently throughout their academic career, teachers said students were being encouraged to aim for the same high standards.

*“My message to policymakers is that we need to have high-quality assessments, and we’re moving in the right direction with the newer state tests.”*

Along with insights into the cognitive complexity and difficulty of the newer state tests, teachers reported that participating in these assessment reviews helped them better understand the additional demands on students that should be addressed. Each content area requires multiple days to assess, leading to days or weeks of testing. The emotional, mental and physical stamina required must be developed over time, starting in the earliest years of schooling. Recognizing and preparing for these demands requires collaboration with school administrators and teachers at all grades. Responding to questions on the new tests also requires technology skills, such as using multiple tabs and toolbars, scrolling through and highlighting text, and building graphs and distributions. So students are well prepared to show what they know and can do, these skills must be regularly built in to learning experiences.

Finally, some participants expressed concerns about whether the local, interim assessments and preparation materials used in their districts were as well-aligned with college- and career-ready expectations. The difficulties faced by students who are academically behind or at risk, those with special needs, and those still learning English provoked anxiety, empathy and conversation about how to do better. Participants agreed that richness of students' lived experiences was not always fully represented in the assessments, but should be. Teachers stressed that tests must be culturally responsive, equally measuring the progress of students in both underserved and affluent contexts. And the balance between summative proficiency measures and evaluation of student growth is not yet where they want it to be. These topics will continue to be a focus of effort and progress as assessments evolve and improve.

*“We all came to the NNSTOY review sessions with some biases about what families were saying, what our administrators were saying, what the press was saying about the new state tests. It was at the time when many people were opting out. I felt really well-equipped to return with better information and kind of start engaging some of those people in more conversations about the value.”*

*“Most importantly, I want an assessment to acknowledge the bias that is in design. Are we assessing the skills or some random experience that some random white person had in suburbia?”*

### LESSON 3: FEWER, HIGH-QUALITY ASSESSMENTS THAT ALIGN WITH EXCELLENT INSTRUCTIONAL PRACTICES ARE WORTH THE INVESTMENT.

Another major theme heard from these experienced teachers was that, for assessments to be worth the investment in time and effort, they need to be of high quality. More tests are not necessarily better.

**These educators believe that worthwhile assessments have some common characteristics:**

- **Assessments and instruction must be aligned.** However, assessments should not drive instruction alone, nor should instruction solely determine assessment. Rather, the assessment and instruction should be aligned with the same standards and goals and use the same language and approaches. Additionally, the content, depth of knowledge and cognitive rigor of both should be similar so students recognize their classrooms in the assessment and the assessment in their classrooms.
- **The purpose of assessments should be clear.** Assessments are pressured to do many things for many audiences, which dilutes the focus of the assessment and adds required time or extra content. The purpose of the assessment should be explicit and be clear to those administering and taking it. The content should be meaningful to students and to teachers. The format of the assessment should suit its purpose. Consideration must be given to the multiple experiences and lives of the students, as well as the actual experience of interacting with the test. It is not necessarily cheap to assess in this way, but it is valuable and worth the investment.
- **Data matters.** This means that reports have to be useful and provide sufficient information to feed back into teaching practices. A scale score with no context, detail, or profile of mastered and emerging skills is not valuable, nor is it an appropriate return on the investment made by all participants. Teachers understand the need to maintain test security for analytic purposes, but that must be balanced with the needs of teachers, students and parents to truly understand and be able to utilize the results.
- **Timing matters too.** Even excellent reports that are returned after the students have moved on to the next grade are of limited value. Though such data may reveal information about past learning, they do not offer the educator an important opportunity to diagnose student strengths and weaknesses in order to design precise interventions that improve instruction. Assessment results need to be returned while the student achievement data are still actionable for teachers, school leaders and parents— not well into the following school year.

*“Once we know the purpose of a test and how it’s going to be used and what we’re going to do with it in schools, districts and the state, that makes it better for all educators. We then know how to address these expectations and how to work with our kids.”*

*“Assessment results should be given back in a way that a teacher or system can use them to make decision. I don’t want an autopsy; that’s a waste of time. I want a test that is actionable.”*

## LESSON 4: BY CONTRIBUTING FIRST-HAND KNOWLEDGE OF STUDENTS AND INSTRUCTION, TEACHER PERSPECTIVES CAN HELP STRENGTHEN STATE—AND LOCAL—ASSESSMENTS.

A powerful theme running throughout all of NNSTOY's "Right Trajectory" reports has been the importance of including teacher experiences and perspectives in major education policy decisions, including assessment design. The teachers in these focus group follow-up conversations again emphasized that teachers are the ones working directly with the students every day—and thus have a powerful and unique vantage point to gauge the quality of and expectations embedded in different assessments. It is not possible to define goals for assessments, much less build good ones, without understanding the instructional experience. Teachers structure and provide the learning experiences that the tests are intended to measure. And it isn't just the teachers in the target content areas, either: For example, career and technical teachers use applied mathematics in their work, and history teachers work with document-based questions frequently. There are opportunities for teachers across the school and across subject areas to get others involved.

*“The teachers are the ones with the boots on the ground. You can't shoot at a target you can't see. Teachers need to see what students have to do. If we don't have teachers participating in this process, we won't have a quality assessment.”*

Excellent teachers are not going to teach directly to any test, and no test is going to be perfectly aligned to any single classroom. But the teachers who participated in the NNSTOY reviews articulated that the more teachers have input, the closer to each other the assessment and the learning become; this should be an outcome that everyone desires.

One of the insights shared during these conversations was that test developers are thoughtful about design and construction. Teachers participating in test construction and review are able to align instruction and activities to the same outcomes and standards as the assessment. Teachers can and do learn how to write items that assess deeper knowledge and that parallel the style and approaches taken on the assessment because they know what that is. And test developers benefit from the experiences of the teachers who understand how students of various ages and interests think; how they

NNSTOY makes its focus group protocols and all tools used in conducting research available in our study appendices. You can find the protocols used in all the Right Trajectory studies, including this one, in our reports referenced on page one above.

For more information, please contact Katherine Bassett at [katherine@nnstoy.org](mailto:katherine@nnstoy.org).

learn, and how they respond in very different ways depending on what seem like small things. For example, an adult may consider a “central” idea the same as a “main” idea in a text. But on the assessment, specific differences in language may alter a student’s understanding of the question.

To the extent that they do not get a chance to share their perspectives, test developers and teachers are “working in the dark” or “shooting at an unseen target,” to borrow two analogies used by the experienced teachers in these focus groups. The focus group educators reflected on how working together brings benefits to both parties. Teachers learn skills to align their practice, thus better equipping students to be successful on assessments designed with them and their experiences in mind.

## LESSON 5: MORE TEACHERS SHOULD HAVE OPPORTUNITIES TO PARTICIPATE IN ASSESSMENT DEVELOPMENT AND REVIEWS.

*“Something that is teacher-led has a greater potential for success. I think it has a lot of potential to change teacher practice and give them the tools to be successful.”*

Participants saw tremendous value in the experience of reviewing the tests, and believe that the more teachers share this opportunity, the richer the conversation becomes. And although these studies evaluated states’ or the consortia of states’ assessments, they believe that the benefits would accrue if this process were followed with assessments developed and administered at a local, district or group-of-districts level. We heard that teachers trust other teachers, so this effort can and should be led by those within their own community who have the expertise and experience to do so.

There are partners who can facilitate this process and offer support, resources and training. NNSTOY is an obvious choice, since it organized and led this series of studies. Other professional development organizations can also provide guidance. NNSTOY’s methodology is freely available, and participating teachers generally perceived it as worthwhile. Many panelists shared that they would willingly volunteer their skills to become the trainers of the next group of teachers to experience what they did.

Such an effort requires trust and investment from all parties of the process. Test developers and states have invested substantial resources in creating assessments, and have reasonable fears about security. This is not only due to fears of content exposure, but because they use analytic processes that rely on item re-use. Educators are professionals and understand that the offer of access to live test forms comes with non-disclosure requirements. Teachers don't want to cheat—they want to understand expectations. Teachers have understandable anxiety about the uses to which test scores are put. Everyone involved has to understand validity, reliability and consequences of test use, going beyond test design and content. Test developers must accept input, even when it is critical, in order to make the end product better. Participants appreciated the high quality of tests, but their interest focused on ways that the tests could form an integral part of the best teaching and learning. They were skeptical, however, about the use of tests for purposes like teacher evaluation, while remaining enthusiastic about ways that tests could capture students thinking and progress. Furthermore, focus group educators shared concerns about ways that tests could be more suitably adapted to English Language Learners and students with disabilities.

*“When teachers have time to work with each other and create curriculum, together, that’s going to have the biggest impact on teacher practices. But every teacher also needs to be exposed to what is on these new assessments.”*

## CONCLUSION

The result of collaboration such as this between policymakers, teachers and test developers would result in a better product: one that is aligned to the experiences occurring in the classroom; one that better reflects real student achievement; and one worth the time, money, stress and effort put into them by everyone.

APPENDIX A: DEMOGRAPHIC DATA OF TEACHERS WHO REVIEWED STATE TESTS & PARTICIPATED IN SPRING 2017 REFLECTION STUDY

**TEACHER GENDER**



FEMALE: **17**  
MALE: **7**

**TEACHER ETHNICITY**

African American: **2**  
Pacific Islander: **1**  
Caucasian: **21**

**TESTS TEACHERS TOOK FOR STUDY**

Smarter Balanced: **15**  
PARCC: **9**

**TEACHER EDUCATION**



B.A.: **1**  
M.A.: **23**  
Ed.D.: **1**

**NUMBER OF YEARS TEACHING**

1 – 9 years: **3**  
12-15 years: **7**  
20-25 years: **4**  
27-29 years: **7**  
30 – 35 years: **3**

**EXPERIENCE WITH POLICY MAKERS**

Yes: **14**  
No: **10**

**TEACHING ENVIRONMENTS**



**DISTRICT TYPES**

RURAL: **7**  
SUBURBAN: **13**  
URBAN: **4**

**GRADE LEVELS**

ELEMENTARY: **7**  
MIDDLE: **10**  
SECONDARY: **7**

**TEACHER CONTENT AREA**

- ENGLISH LANGUAGE ARTS/ READING
- MATHEMATICS
- MUSIC
- SCIENCE
- SOCIAL STUDIES
- TECHNOLOGY
- VOCATIONAL & TECHNICAL

**TEST STUDENTS IN THEIR STATES TAKE**

- 9:** Smarter Balanced Assessment (SBAC)
- 1:** SAGE
- 9:** PARCC
- 1:** End of Course exam
- 2:** Georgia Milestones Assessment
- 1:** Minnesota Comprehensive
- 1:** New York State

## APPENDIX B: THE RIGHT TRAJECTORY FOCUS GROUP: PROTOCOL & SCRIPT

### ***Consent:***

Thank you for agreeing to be part of this focus group. We appreciate your willingness to participate and we want to learn your perspectives.

The purpose of this study is to gather information from outstanding educators on ways in which policymakers can effectively engage educators in making decisions about assessment. We wish to provide a policy brief to policymakers that will give them solid ideas about how and when to engage educators in the policy process regarding assessment. We will collect this data via a series of three focus groups. Your responses are critical to the success of this effort. Your participation is voluntary and you have the right to refuse to answer any question. Your decision to participate or not to participate will have no impact on your organization's relationship with NNSTOY. The purpose of this group is not to evaluate your particular school and your name will not be used in any reports.

Although the focus group will be recorded, your responses will remain anonymous.

### ***Ground Rules:***

There are no right or wrong answers.

It's important to speak one at a time and to state your first name each time before speaking.

It's important that everyone participate.

It's okay to have differing points of view.

### ***Warm up and Introduction:***

Please introduce yourselves. Tell us your name, the state in which you teach, the grades and subjects taught, and your years of experience teaching.

***(8-9 minutes per question):***

1. What lessons or reflections do you have from being involved in NNSTOY's research project of student assessments?
  - **Probe:** Have you taken any lessons or insights with you back to the classroom/district?
2. What did you learn about PARCC or Smarter Balanced as part of this research project?
3. Was it important for you to be able to see live test forms? Why/why not?
4. Why should teachers be engaged in helping to validate the quality of student assessments?
  - **Probe:** What can other agencies (state DOEs, LEAs) do to replicate this kind of teacher-led review of state or local assessments?  
[Note that ESSA has money for some local testing audits.]
  - **Probe:** Would local assessment quality reviews be valuable also? What would the focus or most valuable outcome be of this work?
  - **Probe:** Would a review of local assessments be structured pretty similarly to the ones they did, and consider similar issues, or would they need to be different from the reviews NNSTOY conducted of state tests? What changes would they recommend and why?
  - **Probe:** How could a local or state agency or governor's office get started (call the state NNSTOY chapter? The teacher's union?)?
  - **Probe:** What support would the agency need to provide to make this effective?
  - **Probe:** What support and training do teachers need to roll up their sleeves and meaningfully engage (A lot? A little? What type?)?
5. What would you most want policymakers to know about what you learned?
  - **Probe:** In what other ways can policymakers best engage teachers in the design and review of student assessments?
  - **Probe:** If an agency (local school district, state) wanted to engage teachers in work like this, what advice would you give them?
6. What would you most want policymakers to know about what you believe makes a good student assessment?

## ABOUT THE HQAP : HIGH-QUALITY ASSESSMENT PROJECT

Between 2013-2017, the High-Quality Assessment Project—a pooled grantmaking fund created by the Bill & Melinda Gates, Hewlett, Lumina and Schusterman foundations as well as the Helmsley Trust—provided resources to policymakers and advocates around the country to support them in making the transition to higher quality state tests. Rockefeller Philanthropy Advisors and Education First managed the fund.

Using grants and technical assistance from HQAP, these state and national leaders engaged in reviews of state test quality, developed materials for better engaging parents and teachers, and commissioned research into public opinion about the importance of testing. Across the country and more deeply in 17 states over HQAP’s four years, grantees educated state boards, legislators, governors and the media, and engaged parents as new tests were adopted and administered and score reports issued.

With the sunset of HQAP, Education First has been working to take stock of lessons learned from this initiative and to make sure the field continues to give access to the varied and helpful tools and information grantees created about test quality.

As part of this work, HQAP commissioned case studies from three grantees—the National Network of State Teachers of the Year, Partnership for Learning/Ready Washington and UnidosUS—to share their work influencing the debates about states tests and their advice for advocates and policymakers. Each of these organizations engaged their constituencies—parents, teachers and students—in different and innovative ways, and it’s worth thinking about how these efforts could be replicated in more states and communities.

Complementing these case studies, Education First has prepared a report summarizing the work of HQAP and lessons learned for continued advocacy on the quality and usability of state tests.

Finally, the Education First website now includes a permanent “best of” library of resources from many HQAP grantees, including rubrics for reviewing the quality of tests, communications tools and infographics to discourage “opting out” and model materials for engaging both parents and policymakers. The case studies and the Education First final can be found at this online resource as well.



Education First is a strategy and policy organization that helps policymakers design and accelerate policies and plans that help all young people succeed in college, careers and life.



CLICK HERE TO  
LEARN MORE > [www.education-first.com/tests](http://www.education-first.com/tests)