

Contents

Acknowledgments

vii

1. Introduction

1

Growing Suspicion About Testing in Schools

3

The Power of Portfolios

6

Arguments Against High-Stakes Standardized Testing

7

Yearning to Know More About Learning

9

Tests Have Some Benefits If Used Wisely

10

Numbers Are Meaningless Without Descriptors

11

All Assessment Requires Judgment

12

About the Seven Assessments

13

2. Purposes of Education in a Democracy

16

Democratic Education and Assessment

20

Local Communities Should Decide What Their Students Study in School

23

3. Assessment 1: Student Self-Assessments

27

Student-Led Family Conferences

30

Class Meetings

31

Rubrics

33

Checklists

35

Written Contracts and Self-Reflections

36

The Recollections Project

38

Counteracting Harmful Hidden Messages to Students	39
4. Assessment 2: Teacher Observations of Students and Their Work	41
Parents as Assessors	42
Teachers as Ethnographers	43
The Limits of Teacher Memory	45
Teacher Checklists	46
Narrative Reports	46
5. Assessment 3: Descriptive Review Processes	49
What Is a Descriptive Review?	50
Putting Theory Into Practice	55
Equity and Multicultural Issues	67
Concluding Thoughts	70
6. Assessment 4: Reading and Math Interviews	71
Reading Interviews at Mission Hill School	74
Math Interviews	78
7. Assessment 5: Portfolios and Public Defense of Student Work	80
Portfolios at Mission Hill School	81
Developing Public Speaking Abilities from an Early Age	86
8. Assessment 6: School Reviews by Outside Experts	88
School Quality Review Process in Boston	89
The Office for Standards in Education, Children's Services, and Skills	92
Accreditation of Educator Preparation	95
9. Assessment 7: School Boards and Town Meetings	97
Reforming and Reinvigorating School Boards	101
New England Town Meetings	104

10. Building a System of Assessment: Examining the New York Performance Standards Consortium	107
<i>Ann Cook and Phyllis Tashlik</i>	
Components	108
Impact on Curriculum and Instruction	110
Multiple Ways to Demonstrate Learning	113
Inquiry and Discussion	114
Collaboration and Professional Development	115
Results and Outcomes	117
11. Conclusion	119
Better Ways of Assessing Knowledge	120
Keeping the Purposes of Education at the Forefront	122
Rethinking Assessment	123
Defending the Public Good	126
References	129
Index	133
About the Authors	141

Introduction

As an anchor pulled up after many years on the bottom of a harbor brings with it long strands of attached material, the subject of educational evaluation necessarily [brings] into the spotlight many topics in education—all of them inextricably bound up with it.

—Brenda Engel in *Holding Values: What We Mean by Progressive Education*

For those of us who have spent our professional lives in schools and have thought about what good education requires, it is not always easy to imagine how policymakers see the world. How can this or that policy possibly support what we are doing in schools? It can be a chasm too far to cross, leading some people to cynicism. How can large schools that feel anonymous meet the needs of troubled children? And yet, schools the size of small towns continue to be built. How can funds for public schools be used for private schools, when those schools are not accountable to a public body and can admit and expel students at will? Still, we see powerful proponents of school vouchers at the highest levels of government.

Perhaps most distressingly, in recent decades we have seen policymakers require standardized tests with high-stakes decisions tied to the scores, including student promotion, teacher evaluation, grades for schools, and the evaluation of educational policies themselves. Now this takes some work to comprehend.

James C. Scott (1998) helps us step into the shoes of policymakers in his landmark book, *Seeing Like a State*. He argues that policymakers simplify the world in order to “see” it. Using this generous view of what are, we believe, harmful policies, standardized tests can be understood as a sort of map. Map makers necessarily simplify complex terrain so they can gain a bird’s-eye view. But though this view may be helpful, it is useful to remember that a simplification process—such as shrinking a 3-D globe-like Earth onto a 2-D surface—inevitably distorts the land and sea masses and prioritizes some parts over others. Scott offers dozens of examples of

how the attempts of policymakers to simplify and micromanage the world around them has caused harm to people being governed, even when the intentions of the policymakers seemed benevolent.

High-stakes standardized test results are the data that policymakers use to evaluate and to rationalize educational policies. They create the data that constitute their simplistic and flawed “map” of schools. However, for those of us who have seen how these policies have affected students and schools we care about, we know this map is distorted and prioritizes flawed data. And it is not inevitable. Assessments of students and schools are important. They are the tools used by educators and school communities to understand the needs of students and schools and to evaluate how they are doing and we must pay careful attention to them. But some assessments are better than others.

Unfortunately, “accountability for schools” and “high-stakes standardized tests” are too often confused in the public imagination. We certainly want accountability. We want assurance that our schools are effective and that our nation’s future—our children—are in good hands. But growth on standardized test results alone is not a legitimate way to ensure accountability, despite the claims of presidents, governors, superintendents, scholars, and journalists. Even realtors, who frequently publish the test results of local schools on websites, use this information to value and sell houses.

Over time we have seen standardized test scores become increasingly tied to high-stakes decisions regarding the evaluation of students, teachers, schools, and educational policies, especially since the implementation of the No Child Left Behind law in 2002. But many parents and educators have begun to realize that accountability and standardized test scores are not the same. In a recent Phi Delta Kappa/Gallup poll (Bushaw & Calderon, 2014), for example, 68% of parents surveyed reported that they were skeptical that standardized tests helped teachers know what to teach. And the No Child Left Behind law, which required annual standardized testing for most grades, and specific consequences tied to those test results, became unpopular, receiving more unfavorable than favorable ratings in opinion polls before the law’s name was changed during its reauthorization in 2015 to the Every Student Succeeds Act (ESSA). Yet, standardized tests remain the key tool used by policymakers and others to judge a student and a school’s performance. Under ESSA, schools are required to test students every year in grades 3–8 and once in high school. The consequences of those test results are now under the jurisdiction of states rather than federal law, but most states have attached high-stakes consequences, such as student, teacher, and/or school evaluation, to the test results.

GROWING SUSPICION ABOUT TESTING IN SCHOOLS

We all have had experiences with standardized testing and probably have formed our own opinions about how well they measured our knowledge and abilities. My own (Deborah's) adventures with standardized testing involved a series of "revelations" that I put together of puzzling information I encountered when I became a teacher more than 5 decades ago. I had, personally, taken no standardized tests during my K–12 years. It's possible that I did once take an IQ test at some age, based on what my parents later told me. But otherwise, I was completely ignorant about these kinds of instruments versus tests designed by my teachers and graded by them in ways that spelled out what I had or had not done well. Moreover, in my school students were encouraged to discuss grades with their teachers. (Note: It was an independent school in New York City, and students were largely upper class.)

In 1951, I decided to transfer to the graduate school at the University of Chicago after 2 years at Antioch College. My acceptance was based on taking their standardized tests, including some in subjects I had never studied. I did very well. It puzzled me. The man I married—who hadn't completed high school for very complicated reasons having nothing to do with academic difficulties—also was allowed into the Chicago graduate school on the basis of test scores!

My next experience was when my son Nicky took a test as a favor to a friend whose coursework required her to administer such a test. She noted that he frequently got the easy questions wrong but rarely got the harder ones wrong. His score, in short, was meaningless. The example she gave me of an easy item was about what to do if sent to the store to buy a certain brand of bread, which they didn't have. The choices included going to another store, choosing another bread, or going home. He chose going home. Or course, in our neighborhood the nearest other store involved crossing a busy street and walking about another four blocks. He was 8 years old at the time, and this was out of bounds. Buying another brand—the "correct" response—was not his style of risk taking.

He also took a New York City (NYC) standardized reading test in the fall of that same year—our first in NYC—and the school recommended remedial help, given his poor score. Because he was a fluent and voracious reader of almost any book he could get his hands on, I was puzzled. I declined the remedial help and when he did very well on the spring test, the school dropped the issue. Meanwhile, I got hold of a copy of a recent 3rd-grade test (I was working in a school at the time) and proceeded to have him show me how he worked the test. I discovered that he considered it

cheating if you didn't cover up the reading passages while answering the questions that followed: "Otherwise you could always get the right answer." No one, he said, had told him he should do this. It was his own idea. He thought it was obvious.

When it came to picking answers, I queried him about a few he had gotten wrong. He said he wasn't surprised. He guessed they wanted him to check B, but he actually thought D was better. "But," I protested, "how would they know what your reasoning was?" His answer: "I explained it in the margins."

After my first central Harlem kindergarten class went on to 2nd grade (when NYC at that time started testing kids), I decided to question them similarly about their answers and their strategies. I received a small grant from a foundation for this purpose and taped the sessions. I tried various techniques. For example, I read the passages aloud to see if it affected their answers, and it didn't improve them. When I asked students to explain their logic, I was astounded. Their difficulty wasn't caused by their inability to read the passages accurately, and so my reading them aloud was no help. Their logic when answering items incorrectly seemed excellent, based on the evidence they presented. I proceeded to write a booklet on the topic for City College called *Reading Failure and the Tests* (1973).

I was astounded too by the number of parents who, when asked how their child read, gave me their test score answer but had no idea what it meant and often seemed confused because they thought their child was reading better than they were told the score indicated.

I was also amused at reading a news story complaining that after several years of increased public funds for public schools exactly the same percentage were reading above and below grade level. This education reporter did not know that the test scores were simply a report of scores that rested on setting grade level at the median—thus ensuring that the percentages remained static. Were we expecting poor kids to surpass middle-class and wealthy kids in rank order? Scores indicating that rich kids outperformed poor kids should not be used as evidence to decrease funding for those most in need.

I gradually became aware, as well, of the amount of cheating going on, especially during the years when it was not hard to know exactly what would be on the test. I was startled when reporters and superintendents took seriously reports of big jumps in a school's test scores, or one teacher's scores versus another's. I was embarrassed to discover that a teacher who had informed me about her cheating and was afraid it would be discovered still bragged the next year about her students' scores.

In one case there was a media story celebrating a Lower East Side school whose scores had risen amazingly. I discovered that the school had, during the year involved, become the district's gifted-and-talented site. Hmm?

I began to try to explain more and more to youngsters about the test—including some hints about how to go about improving their guesses and never to leave a question unanswered. I had them make up tests for one another so they might get a better idea of what the test maker was up to. I showed them that because I knew them well, I could design a test that favored Jackie as well as one that favored others. It helped a little—above all, in terms of how seriously they took the results. They were less anxious. But I also knew it could lead to negative results if students stopped taking the tests seriously. I did feel, however, that explaining to them that the items got harder as they went along would be somewhat reassuring, and that in fact the last few items were ones the test makers were sure very few kids would get right.

I realized as I learned more that my skill at taking standardized tests was related to my skill at understanding how the test maker saw the world and what he and other adults I knew would want me to answer. The problem was that my students' past experiences had led them into a trap. They too gave the answers they thought adults would approve of—but experience led them to the wrong conclusions. My students and their families and community had different experiences than the test makers and the kids they had in mind when deciding what answers were “right.” My students had sufficient vocabularies in terms of size, but they weren't the words that had been selected for the test. In short, their own sophisticated and intuitive intelligence was a disadvantage, not an advantage.

Clearly my knowledge base—talking and reading with and to children—had given me a far greater basis for assessing their skill and knowledge. Alas, their parents and the children themselves had far too much trust in the tests, and it led them into feelings of inadequacy and despair: “I'm still below grade level no matter how hard I try and how much I think I've improved.” It was for this reason that at Central Park East Elementary School (CPE) we developed a different reading scale, along with a different way to present the information to the kids and their families. We also explained all the reasons that their standardized test scores misrepresented their skill—including standard measurement error, which was quite considerable. So we audiotaped children's reading and developed a scale that we could reliably use among ourselves and for families. Kids and parents enjoyed listening to their recordings over the years and seeing their improvement acknowledged. It included reading, talking about assigned reading, and conversations about aspects of reading. And we continued to rely on a growing number of assessments other than standardized tests.

THE POWER OF PORTFOLIOS

When I became the principal of Central Park East Secondary School, I was amazed at how seriously and respectfully both students and families took the portfolios and their presentations (see Chapter Seven). We had explained to them that they were doing something similar to what PhD candidates do in presenting their theses to their committees, as well as being much closer to what employers do when assessing candidates for jobs that require real skill and understanding. They—professors and employers—if possible, seek to assess candidates on the basis of actual work they do and their explanations about how they went about the work. It was also closer to the professional system developed by the National Board for Professional Teaching Standards (on whose founding board I served). Their system included portfolios and videos of actual teaching followed by discussion and explanations regarding why they had done x or y.

The authenticity of the process was impressive to the kids. They rarely questioned its accuracy, and if teacher or student or family had doubts about the results, we could always redo the assessment with their input. They got a chance to question our conclusions and sometimes even to change our opinions. They offered us, as teachers, a chance to better understand one another's work and how our students understood their own work. Involving external reviewers, parents, and younger students also gave us information useful to improving our work.

It was so impressive that, in retrospect, clearly we should have used something similar for moving up from 6th grade to 7th and from 8th to 9th and 10th to 11th. We used a much less rigorous or formal approach that was somewhat based on the same idea, but rested much more on our individual intuitions and less on open dialogue between observing adults and the students themselves.

Also useful at the end of 12th grade was the required reflection piece about their own self-assessment, their plans for the next phase of their life, and ideas about what practices perhaps needed to be improved or changed (see Chapter Three). This was not scored in any way, but was an opportunity for families, students, and faculty to reflect together on ways in which each might be helpful. We should have started earlier.

The fact that in all these processes adult judgment and expertise was respected and displayed openly to students was an important aspect of the design. We believed it would help them in the future in viewing adulthood and expertise as highly valued—although not perfect! We were thus not surprised when college faculty and admissions officers commented about these qualities in our graduates, as well as their ease in discussions with adults.