

Testing Tales: Discussing the Full Range of Views

Overview:

Participants work in small groups to read and discuss “testing tales.” These “tales” are short vignettes about some aspect of testing. Many are direct quotes from testing experts, others are summaries of research studies or news articles. Each presents a viewpoint usually with evidence, followed by one or two provocative questions designed to help the reader think critically about the viewpoint and evidence that is presented. The testing tales range the full continuum of viewpoints, from pro-testing to anti-testing and many places in between.

The purpose of this activity is to provide participants with an understanding of the wide range of viewpoints there are around testing, and to model the need to ask critical questions and make up one’s own mind. Through this experience, participants gain a more complex understanding of a complex topic.

Use:

Serves as a chance to deepen understanding about testing and accountability issues and provides opportunities to think critically about the barrage of anti- and pro-testing views that assault us all. Is best used after participants have an understanding of what standards-based reform is. Use after an introductory activity, such as the powerpoint presentations, *“What is Meant by Accountability?”* and *“What is Standardized Testing.”*

Nice alternative to straight information delivery. Enables participants to explore the controversy around testing in a comfortable setting.

Key Concepts Addressed from Public Understanding Framework:

Are Tests Accurate?
Standardized Testing

Time Frame: 60-75 minutes

Reading and Discussing Testing Tales (30-45 minutes)
Whole Group Summary (30 minutes)

What You Need:

For each group of 4-6 participants:

- o 1 set of Testing Tales duplicated onto cardstock
- o 1 large ziplock bag (or folder) to hold the set of Testing Tales

For the presenter:

- o overhead transparency of “Testing Tale Summary Points” (master)
- o overhead projector
- o extension cord (optional)

Getting Ready:

Before the Day of the Workshop:

1. **Select which Testing Tales to Use.** Read the two dozen Testing Tales (masters) and **decide which ones you would like to use.** Some people like to use them all, so participants can choose which ones they read. Others prefer to choose a subset so participants are sure to read certain ones. Depending on how much time you allow, even the fastest groups will only get to read and discuss up to ten tales—roughly half of what’s provided. If you choose a subset to use, make sure that it includes the range of viewpoints, not just one side of the issue.
2. **Duplicate Testing Tales.** Duplicate 1 copy of the set of Testing Tales you will be using for each group of 4-6 participants. Copy them on to cardstock. Then cut them up so each Testing Tale is on its own “card.” Put each full set of Testing Tales into a large ziplock bag.
3. **Make Overhead Transparency.** Make an overhead transparency of “Testing Tales Summary Points” (master).

Immediately Before the Workshop:

1. **Set up the room.** Arrange the room so that groups of 4-6 participants can sit at a table together. If you are in a classroom, move desks together to make “tables.” Tables should be oriented so that all of the table groups can join a large group discussion, and see what’s projected on the overhead.
2. **Set up overhead projector.** Set up overhead projector at the front of the room near where you will stand.
3. **Have overhead transparencies on hand.** Place the 7 overhead transparencies (in numbered order) next to the overhead projector.
4. **Have sets of Testing Tales on hand.** Have easily accessible, the sets of Testing Cards.

Reading and Discussing Testing Tales

Note: The challenge in conducting this activity is to make sure that the small groups stay on task, reading and discussing the testing tales, and that each person in the small group is involved. Given the topic of testing and accountability, the situation is ripe for lots of side talk. And because there will naturally be a range of prior knowledge about the topic within the participants in the small group, it's not uncommon for one or two participants to dominate the discussion. We've suggested a structure for dealing with this. Be aware that these problems might arise.

- 1. Set the Scene.** Begin by acknowledging the complexity of the topics of testing and accountability and that there are many different perspectives and viewpoints about the topics. Explain that this activity will enable them to consider many different viewpoints and to think critically about each. The goal is not to convince them that any particular view is correct but rather for them to decide what they themselves think, based on the viewpoints and evidence they read about.
- 2. Introduce the Activity.** Tell them that each group will be given a set of Testing Tales. Each Testing Tale is on a different "card." Testing Tales are short vignettes about some aspect of testing. Many are direct quotes from testing experts, others are summaries of research studies or news articles. Each presents a viewpoint usually with evidence, followed by one or two provocative questions designed to help them think critically about the viewpoint and evidence that is presented. The testing tales range the full continuum of viewpoints, from pro-testing to anti-testing and many places in between. *You may want to add some of your own excerpts, from current articles on testing.*
- 4. Give Instructions.** Tell them that you'd like them to work in groups of 3-5 people. When they begin, each person should choose a Testing Tale and read it silently to themselves. When they have all finished reading, they should go around the table, person by person. The person who read the Tale should explain to the rest of the group what it said. Then the group should briefly discuss it and try to answer the questions in italics at the end of each Tale. When everyone has had a turn, then each person should pick another Tale and the process repeats. Tell them that they have approximately 30 minutes to do this, and that you'd like them to get through three rounds of reading and discussing in that time period.

5. **Introduce Roles.** To help them achieve their goal, ask each group to assign a timekeeper (to keep the group moving and able to read and share three rounds of Testing Tales), and to assign a facilitator (to make sure that everyone gets a chance to speak).
6. **Begin Reading and Discussing.** Set the groups to work. As they are reading and discussing, circulate among the groups. Help keep groups on task by reminding them periodically that they should be close to reading another set of Tales. Use your judgment about when to do this. While the goal is not to read as many Testing Tales as possible, rather to have meaningful discussion about some number of them, there is a very powerful tendency for inertia to set into these groups, and for one or two people to take over the discussion. The facilitator and timekeeper may be shy about their role in keeping the groups on task and moving, and side discussions may dominate. In our experience, keeping the discussion driven by the Testing Tales is the best way to achieve the goal of the activity.

Whole Group Summary

1. **Focus the Group's Attention.** About 30 minutes from the end of the session, have groups conclude their discussions.
2. **Invite Participants to Share.** Stimulate sharing about the Testing Tales with some questions to the group:
 - Was there any information that surprised you?
 - Did anything you read and discussed change your mind? Deepen your understanding of the issue?
 - What were the key points that kept recurring in different Testing Tales?
3. **Provide Summary Points.** About 15 minutes from the end of the session, present the following summary points using the overheads you have prepared:
 - Standards can be a way to ensure rigorous academic goals for all students and all teachers
 - Large-scale assessment has inherent limitations:
 - lack of timeliness for reporting results
 - multiple-choice question format
 - focus on knowledge of facts and basic skills
 - A single data point, either for a child or a school, cannot provide the whole picture of what learning progress is being made
 - Creating an accountability system that works to improve teaching and student learning is a complex issue.

Testing Tales Summary Points

Standards can be a way to ensure rigorous academic goals for all students and all teachers

Large-scale assessment has inherent limitations:

- lack of timeliness for reporting results
- multiple-choice question format
- focus on knowledge of facts and basic skills

A single data point, either for a child or a school, cannot provide the whole picture of what learning progress is being made

Creating an accountability system that works to improve teaching and student learning is a complex issue.

Statewide testing is seen as a way to raise expectations and help guarantee that all children are held to the same high standards.

Does a focus on testing, narrow student learning to what is tested?

Are tests measuring what's important to learn or what's easiest to measure?

Asked why the annual whale watching trip had been canceled, the superintendent of schools in East Palo Alto, California replied: “Students are not tested on whale watching, so they are not going whale watching.”

Is everything important measured on tests?

How far is too far to take a standards-based approach to learning?

Gerald Bracey, well-known testing expert points out a few of the personal qualities that standardized tests do *not* measure: creativity, critical thinking, resilience, motivation, persistence, humor, reliability, enthusiasm, civic-mindedness, self-awareness, self-discipline, empathy, leadership, and compassion.

Does this mean that the things measured by standardized tests are not important?

A University of California, Berkeley, psychology professor, Rhona Weinstein, has spent years studying how a high or low expectation about a child's academic ability can become a self-fulfilling prophecy. A recent study of hers shows that increased emphasis on testing diminishes the opinions that low-performing youngsters have of themselves.

Is it possible for testing to be done in a way that communicates high expectations to all students?

Is there a way to help parents, students, and teachers to take testing in stride?

A national survey of teenagers conducted by the Public Agenda Foundation found little anxiety among middle school and high school students over taking standardized tests; 28 percent say they do not get nervous at all, and 67 percent say they get a little nervous but that they can handle it. Overall, the Public Agenda survey, “Reality Check, 2001,” found that the students in their survey are not overly concerned about testing.

What about the other 33%?

Is there a different response to testing in elementary school aged children?

Is having the experience of testing in elementary school important to students not being overly concerned in middle and high school?

How are tests helpful to students, parents and teachers? Quite simply, tests provide a clear measure of how well students are learning. As we set high standards for what we want our students to know and be able to do, good tests are tools that help us gauge student progress. While they are not the only measure of achievement, test results can help parents understand better where their children are succeeding and where they need extra help. Tests can help teachers assess how their class as a whole is doing and how individual students are doing.

How do these results help teachers when typically, standardized test results are received at the end of the school year or after students have moved on to their next year’s teachers?

Could test results be provided to the following year’s teacher and be a basis for early conferencing with parents?

Do not misunderstand me. I am an outspoken advocate of high standards and rigorous testing. For example, I favor an absolute standard for reading achievement that every third grader must meet, and I would require testing to assess students' progress. Yet the crucial issue is, how are we going to use these tests? Do we use them as a stethoscope, to diagnose strengths and weaknesses so we can give every child the assistance he or she needs to succeed? Or do we use tests as a sledgehammer, as high-stakes instruments for determining winners and losers?

If there’s no consequence to schools and teachers for their students’ poor performance, will those schools change their practices?

Is it possible to use large-scale assessment results constructively to improve educational practices?

Politicians have focused on an issue that they think appeals to voters. It's seen as good to be "tough on schools." Public schools are in great danger. The notion that test scores equal quality means that money will flow to schools that have the greatest increase in test scores. Money will flow out of those schools that don't do well. This accountability system establishes an underlying mechanism for market forces to take over education. Our nation seems to be applying the same regulatory model to schools that we apply to under performing savings and loan institutions. In the future, will regulators go in and seize non-performing schools, dissolve their assets, and sell them to the highest bidder?

Is there something intrinsically wrong with tying funding primarily to performance versus the needs of schools sites

Just as medical tests help diagnose and treat patients, rigorous and meaningful education assessments can help ensure the academic health of all students. Just as in medicine, assessments in education are a means to an end. Assessments provide information on where students and schools need to improve, and they may provide incentives for students and schools to make the necessary improvements. But tests alone cannot create improvement. Educators, parents, and students must do the work of raising student achievement. To make this possible: Schools must get test results in a timely and useful manner. Waiting six months to see how students scored is of little value in helping those students improve. States must provide the test data in a form that educators can understand and use, with a finer degree of specificity than just a number on a scale. For example, a particular score in phonemic awareness conveys more to a teacher than an overall score in reading and certainly more than a score in English language arts. Specific results that identify students' particular strengths and weaknesses enable teachers to target instruction to meet the needs of each student. Meaningful data turns a diagnosis into action, thereby enabling educators to respond to individual student needs. They also make assessments a helpful tool for educators rather than simply an accountability hammer.

Is it practical to think that teachers can use assessment information to create individualized learning plans for each student?

Are these kinds of diagnostic assessments widely available?

Tests should be used now and improved later—rather than resisted until they are perfect—because they provide the best means of charting our progress toward the goal of improved academic achievement.

How do testing companies create tests?

Who is involved in revising them?

Do testing companies want more sophisticated tests?

What will motivate an ongoing evolution of tests?

Critics are concerned that decisions about students' graduation or promotion from grade to grade are now being made based solely on performance on one multiple-choice test. They also question whether tests should be used to hold individual students accountable when it is not clear whether schools are providing students with the tools they need—such as high-quality teachers, strong curricula, and extra time to master what's expected on the tests. For example, only half of the states that now require students to pass a test to graduate also provide funds to give students additional assistance if they fail.

Will the education system change without pressure being applied to this one aspect—student performance?

Is there any evidence that using high school exit exams, can improve teaching and learning?

Rigorous testing that decides whether students graduate, teachers win bonuses and schools are shuttered, an approach already in place in more than half the nation, does little to improve achievement and may actually worsen academic performance and dropout rates, according to the largest study ever on the issue. The study, performed by researchers at Arizona State University, found that while students show consistent improvement on these state exams, the opposite is typically true of their performance on other, independent measures of academic achievement.

"Teachers are focusing so intently on the high-stakes tests that they are neglecting other things that are ultimately more important," said Audrey Amrein, the study's lead author, who says she supported high-stakes tests before conducting her research. "In theory, high-stakes tests should work, because they advance the notions of high standards and accountability. But students are being trained so narrowly because of it, they are having a hard time branching out and understanding general problem-solving."

The study was commissioned by the Great Lakes Center for Education Research and Practice, a Midwestern group of six state affiliates of the National Education Association, which has opposed using any one test to determine when students graduate, schools get more money and teachers are replaced. Perhaps most controversial, the study found that once states tie standardized tests to graduation, fewer students tend to get diplomas.

Are there any alternative studies that have shown an increase in academic achievement and graduation rate due to high stakes testing?

The El Centro (California) School District implemented a district wide focus on inquiry-based science instruction five years ago. In this traditionally low-performing district with high levels of poverty and minority enrollment, teachers and principals received extensive training in inquiry-based science that continued over the four years of the study, and they were encouraged to focus their efforts on inquiry-based science instruction. The results show that standardized science achievement scores improved, with improvements increasing the longer students were taught using the inquiry and hands-on methods. Interestingly, grades 4 and 6 also showed impressive improvements in their SAT 9 mathematics and reading scores, as well as on the district's writing proficiency exam. The district focus developed out of "a belief that the skills of reading and mathematics are strengthened when taught using the engaging, high-interest content" of inquiry-based science.

How do they know that the improvements in math and reading are connected to the science instruction? Has it been shown to be connected?

If the students and teachers are stable over 4 years, does this select for a specific population that is not comparable to all students/teachers?

Non-instructional factors explain most of the variance among test scores when schools or districts are compared. A study of math results on the 1992 National Assessment of Educational Progress found that the combination of four such variables (number of parents living at home, parents' educational background, type of community, and poverty rate) accounted for a whopping 89 percent of the differences in state scores.

Should the presence of natural variation be a reason to settle for lower standards for some students?

How can NAEP be designed to better measure factors we want to compare: differences in reading programs used?

Norm-referenced tests were never intended to measure the quality of learning or teaching. The Stanford, Metropolitan and California Achievement Tests (SAT, MAT, and CAT), as well as the Iowa and Comprehensive Tests of Basic Skills (ITBS and CTBS), are designed so that only about half the test-takers will respond correctly to most items. The main objective of these tests is to rank, not to rate; to spread out the scores, not to gauge the quality of a given student or school.

What is the value of making these kinds of comparisons?

What does that information tell you about different populations of students?

It is notable that the majority of the public and most politicians support the need for high stakes testing, while the majority of educators and testing experts do not.

What does each group's stance say about their responsibility?

Is there a common ground for these groups? Do they have the same goals for students?

What is the process for designing these kinds of solutions to educational problems?

George Madaus supports testing, but he's well aware of the weaknesses of multiple-choice questions. "The adults who write the questions sometimes lose sight of the way kids will read those questions. There's a standardized test question that shows a cactus in a pot, a rose in a pot, and a cabbage, and the question is which needs the least amount of water. To the item-writer 'cactus' was the right answer, but some kids pick the cabbage. And the reason they gave was that the cabbage had been picked and so it didn't need water anymore. That's a perfectly good answer, but the machine had been set to score it as wrong." Students who get the "right" answer have demonstrated, perhaps, that they think like an adult -- or like the test-maker. The kid who thinks differently, or whose frame of reference is different, is marked down, and perhaps eliminated from the competition.

What % of tests are flawed?

Why are so many tests in multiple-choice format?

Is it possible to write multiple choice questions that are fair and accurately measure student learning?

Polls by Public Agenda and other organizations continue to show that 72 percent of Americans -- and 79 percent of parents -- support tougher academic standards and oppose social promotion "even if [the outcome is] that significantly more students would be held back." Those numbers seem to reinforce the argument of Diane Ravitch, an education historian, an education official in the Bush Administration, and a strong supporter of standards, who has described the protesters as "crickets" -- few in number, but making a disproportionate amount of noise. "There's tremendous support" for tests, Ravitch says, "among elected officials and in the business community." She may also be correct when she says that a great many of those who profess to oppose the high-stakes tests oppose all testing and all but the fuzziest standards. They are the same people, Ravitch argues, who in the end cheat kids by demanding too little and forever blaming children's inability to read or to do elementary math on the shortcomings of parents, neighborhoods, and the culture. Scrap the tests and we're back to the same neglect and indifference, particularly toward poor, marginal students, that we had before. Letting students who can't read, write, or do basic math graduate is doing no one a favor.

Why might the "business community" support testing? Does it include people who make money through tests and test-driven materials?

Are there any studies indicating that more standardized tests help "poor, marginal" students to succeed?

Is it true that opponents to high stakes tests also oppose all tests and standards in general?

As Ted Sizer (a well-known education reformer) notes, the real world is not "a series of set, pre-digested answers" but a set of questions. "Take the issue of cloning, an issue so difficult that very few teachers want to talk about it, or know how to talk about it. Cloning raises all sorts of difficult questions. What are the right answers? That can't be put on a multiple choice test."

"At the worst," Sizer adds, "these standardized tests provoke a kind of drilling mentality. It's a game. And so students learn the game. What they learn is to hire people to teach you how to figure the test out. Not the substance, but the test."

Can't large-scale tests be designed that measure more than "set pre-digested" answers?

We should hold schools accountable for meaningful outcomes that have a connection to the American economy. That economy demands citizens with creative, intelligent minds. Not everyone fits into the same box. Standardized tests should be used on an informative basis--to gauge what's going on in a district, not to make decisions about individuals. You don't need to test everybody to do this. You just need a sampling.

Is business really looking for creative, intelligent minds?

Are these qualities that can be easily assessed?

The Texas “Miracle”

Perhaps no state testing program has aroused the ire of testing critics more than the Texas Assessment of Academic Skills (TAAS), for over a decade the backbone of the Texas accountability program. Texas students' average state test scores have shown achievement gains year after year, and Texas parents are overwhelmingly positive about the accountability system, with a majority in favor of the high stakes high school exit exam. Critics say the Texas tests rely almost entirely on multiple-choice items and they are opposed to the high-stakes graduation test. Still, these critics have to “concede that the improvements are impressive,” reports *Education Week*, “but [many] claim that an enriched curriculum, not test preparation, is behind the shifts.” Others are quick to point out that Texas standards are low, some say too low to be standards of excellence.

Unintended effects of the Texas accountability system are also considered to be serious by some. One research study has shown that tests in Texas have led to enormous numbers of students being held back in ninth grade, before the test, then dropping out later. The Texas program has also been the subject of two racial bias lawsuits by the NAACP and the Mexican-American Legal Defense Fund (MALDEF). In one year, of the 10,000 Texas seniors who failed the graduation exam, about 7,000 were Latino or African-American, even though they make up only 40 percent of state students. Some also argue the test is unfair because blacks and Latinos generally go to worse schools than whites and don't therefore have the same opportunities to learn. But others say while that may be true, it has only been the state's commitment to rigorous assessment that has improved minority schools and minority school performance in the last 10 years.

As the standards requirements take effect, and more parents face the possibility their children will not graduate, pressure to lower the bar or eliminate it will increase. Conversely, as more people come to understand that the “Texas miracle” and other celebrated successes are based on embarrassingly low benchmarks, those, too, will come under attack.

Is setting low standards okay or does that defeat the purpose of standards-based reform?

How can we set rigorous standards that reflect reasonable goals for all of our students?

A research report by the Chicago Annenberg Research Project showed that students who received assignments requiring more challenging intellectual work also achieved greater than average gains on the Iowa Tests of Basic Skills in reading and mathematics, and demonstrated higher performance in reading, mathematics, and writing on the Illinois Goals Assessment Program. This flies in the face of the increasingly common practice of focusing more class time and class assignments on review and teaching to the test. Contrary to some expectations, the authors find that high quality assignments were found in some very disadvantaged Chicago classrooms and that all students in these classes benefited from exposure to such instruction, as measured by higher test scores. The results suggest that if teachers, administrators, policymakers, and the public-at-large place more emphasis on authentic intellectual work in classrooms, rather than teaching to the test, yearly gains on standardized tests in Chicago could surpass national norms.

What further evidence would be needed to conclude that “teaching to the test” is less helpful to student scores than presenting “authentic intellectual work?”

The news media and other popular forums continue to tell us that all students are not reaching the standards that would help the U.S. become first in the world in mathematics and science or to make all students scientifically and mathematically literate. Can standards make a difference? Advocates of standards in mathematics and science believe that standards offer teachers a guide for setting clear and coherent learning goals for all students. Similarly, Kati Haycock, of the Education Trust, declares standards as a key to setting benchmarks for what students should know at each grade level. For the past six years, Haycock and colleagues have been observing in high poverty classrooms, to see what happens there—what kinds of assignments teachers give, for example—compared to what happens in other classrooms. “We have come away stunned. Stunned, first, by how little is expected of students in high-poverty schools—how few assignments they get in a given school week or month. Stunned, second, by the low level of the few assignments that they do get. Clear and public standards for what students should learn at benchmark grade levels are a crucial part of solving the problem.”

Can one set of standards work for all students, without being impossibly high for some or impossibly low for others?

What’s the process for developing high quality standards that can be achieved in all classrooms with all students?