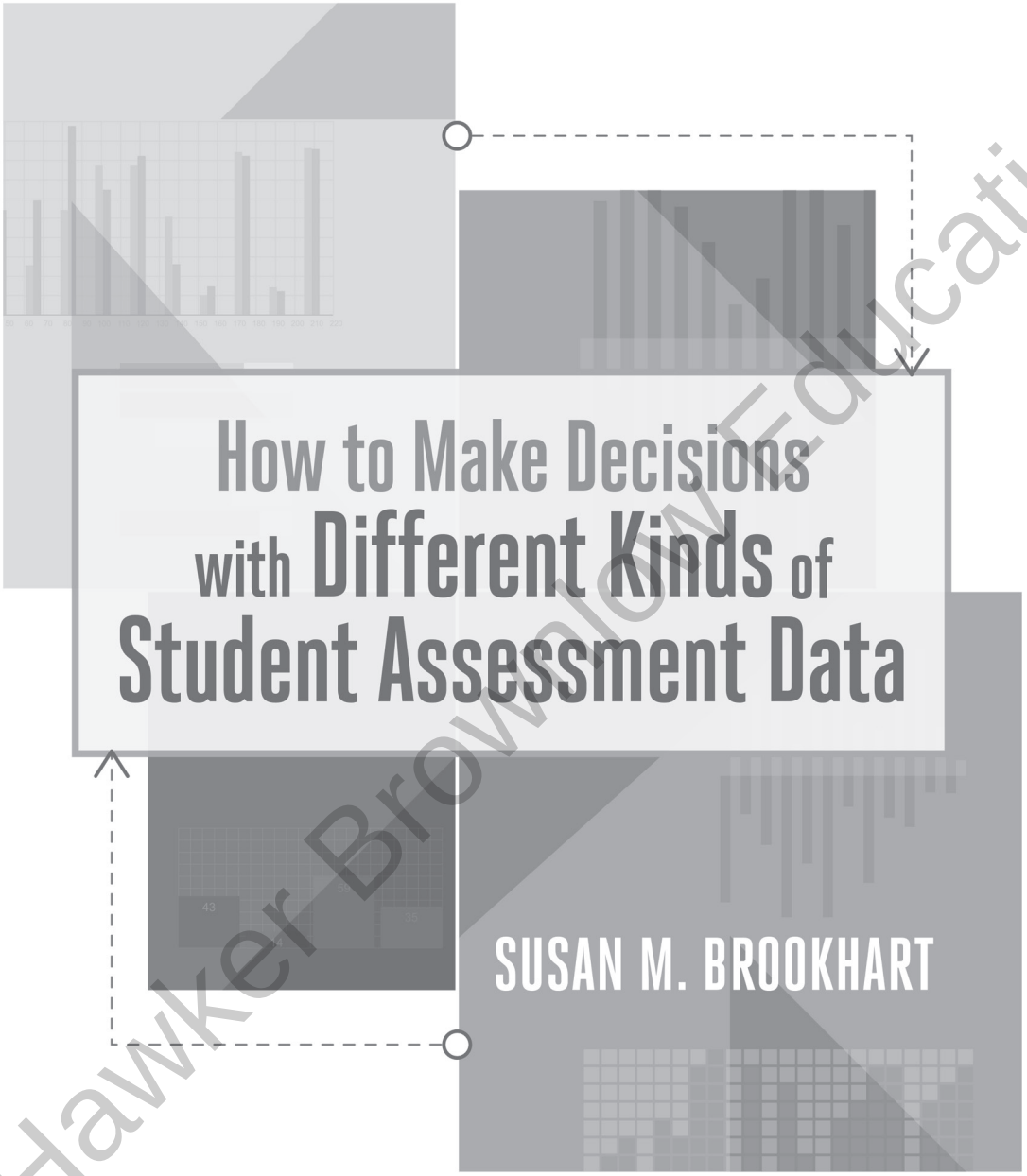


How to Make Decisions with Different Kinds of Student Assessment Data

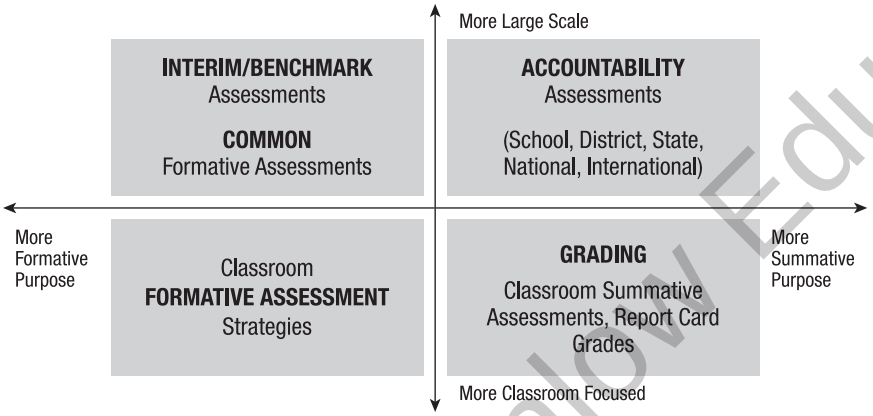
SUSAN M. BROOKHART

Hawker Brownlow EDUCATION

ASCD®

How to Make Decisions with Different Kinds of Student Assessment Data

An Introduction to Different Kinds of Data



Simple logic does not always help us with data interpretation. Here are two examples, both of which are true stories.

Once during a workshop session, a school administrator explained to me that he saw standardized test scores as a sort of barometer. He aimed to make changes in his school that would lead to a rise in test scores, and that would be his indication that his reforms were successful. The analogy he gave me was the population of oysters in the Chesapeake Bay. Environmental reforms were needed in the bay area, changes were made, and the oyster numbers are increasing (see http://www.chesapeakebay.net/issues/issue/oysters for more information). Similarly, he explained, reforms in his school should lead to higher test scores.

The second example occurred at an airport, where I struck up a conversation with a businessman and his young son who were waiting for the same airport shuttle I was. The man said that he was really glad that his state now administered standardized tests, as per No Child Left Behind, because finally he had what he called a "bottom line" that he could watch to know how his child and his child's school were doing. His analogy was to the bottom line in a profit-and-loss statement in a business.

The administrator and the businessman dad were well-meaning people who valued education; they were not nay-sayers. They were both bright and successful individuals who applied logic and common sense to a problem they cared about. And they were both wrong.

Here's the thing. If you are an environmentalist or an oysterman, the oysters *are* the issue—or in the case of the Chesapeake Bay, one of the issues. Increasing the oyster population, to improve the habitat and the economy, *is* the purpose of the scientific reforms and management strategies. More oysters means the program is achieving its goal. Similarly, generating a profit *is* the purpose of being in business. Higher profits mean more money for shareholders, employees, and product development. Making money means your business is achieving its goal.

In contrast, raising test scores is not the purpose of education. The purpose of education has changed with society's needs and values over the years (Sloan, 2012). At this point in time, if you ask people the purpose of education, you will get answers such as these: to create adults who can compete in a global economy, to create informed citizens who can participate in the democratic process, to create critical thinkers and problem solvers, to create lifelong learners, or to create emotionally healthy adults who can engage in meaningful relationships. Obviously, no test score can tell you whether you have achieved these things.

The less obvious problem with our well-meaning administrator and businessman is that *even if you limit your interest to academic learning outcomes, raising test scores is not the purpose of education. The students' learning is.* Test scores are a measure of student learning, but they are not the thing itself. In our analogies, the oysters and the money were themselves the objects of interest. You can count oysters, and you can count money, but you can't "count" learning.

The best you can do to measure learning is to use a *mental measurement* that, if well designed, is a measure of learning in a limited domain. The key is to define clearly what that domain is, use a test or performance assessment that taps this domain in known ways, use a score scale with known properties that maps the student's performance back to the domain, and

interpret that score scale correctly when making inferences about student learning. The purpose of this book is to explain just enough about the properties of data on student learning so that you can make those inferences well. Then—and only then—can you make sound decisions. Another name for this purpose is *developing assessment literacy*. As the examples demonstrate, literacy in educational assessment involves more than counting or ranking. It involves specifying *what specific learning* you are measuring; understanding how the questions or tasks in the measure form a sample of that domain of learning; understanding properties of the scales, numbers, or categories used in the measurement; and being able to reason from all these things to make sound interpretations and decisions.

To complicate matters a bit, as the title of this book indicates, there are different kinds of data. For most educational decisions, you will want to mix the different kinds of data to broaden and deepen the pool of information about student learning that you use to monitor and improve that learning. You will want to know which kinds of data to watch, and when, in order to evaluate the effectiveness of your decisions. This book will help you do that in two ways. First, it offers a framework for thinking about assessment systems that categorizes different measures of student learning. Understanding how information differs from one category to another will help you interpret data. Second, this book offers some insights into different types of scores. Understanding different types of scores and their meanings will help you interpret data properly, as well. Equipped with an understanding of these two big ideas, your data interpretation and subsequent decisions will be more sound, more valid, and more useful.

The Purposes and Uses of Data

The phrase "data-based decision making" is used often and has many meanings. Teachers use data to answer questions about students. Groups of teachers and building administrators use data to answer questions about students, classes, programs, and their school. Central office administrators use data to make decisions about teachers, as well as students, classes, programs, and

schools. An Internet search on "data-based decision making" will bring up dozens of PowerPoint presentations, PDFs, images, and plans. Many books also address this theme.

It may seem like an obvious point, but the data you choose should be related to the intended purpose. If I want to make a tablecloth, I need to measure the length and width of the table; measuring its height won't help me much. The same principle operates in making decisions about student learning, but it's less easy to see. For example, if I want to make a decision about which reading skills to emphasize in my reading class, shouldn't I just look up students' scores on the state reading test? No. What the state reading test measures is general, overall "reading achievement," as defined by a whole set of reading standards. State test results will give you a sense of how your students do at "reading in general," as defined by whatever reading standards your state says its test measures, taken all together.

For example, the Smarter Balanced Assessment Consortium says that, regarding reading, its assessments can support this claim: "Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts" (Smarter Balanced, 2014). If I'm a teacher with a student whose assessment results suggest he can't do that very well, how do I design instruction for him? Low test results relative to this claim suggest a general decision—more or different reading instruction—but don't provide any clues about what aspects of reading to emphasize, remediate, or build on. To design reading instruction for that student, I'll need different data, assessment data that give a more fine-grained description of what the student can and cannot do as he reads.

In this simple example, reasoning from data is a two-step process. Data from the standardized accountability assessment help me identify a problem (Arland doesn't read proficiently) and lead me to another question: Why? To answer that second question, I need different data, because the reading accountability test doesn't give me information that is specific enough. Using complementary kinds of data for educational problem solving requires understanding different kinds of data.

116003 • 9781760019549 • © 2016 Hawker Brownlow Education

This focus on a deeper understanding of data about student learning is what sets this book apart from other data books. I will, of course, also talk about how to use the data to inform instructional improvement. Two other excellent books that talk about using data to inform instructional improvement are *Data Wise* by Boudett, City, and Murnane (2013) and *Using Data to Focus Instructional Improvement* by James-Ward, Fisher, Frey, and Lapp (2013).

This book complements those and other books about data by focusing on developing a clearer understanding of exactly what test scores and other data about student learning *are* and what they *mean*. As an analogy, think about reading a Shakespeare play in a high school English class. If you read the play with a basic understanding of the English language, you will understand the plot. If you take the time to learn some Shakespearean vocabulary, you will understand the plot and the word-play nuances that unlock some of the humor and characterization in the play. In other words, you will understand the play better. Similarly, if you do data-based decision making with a basic understanding of assessment and of numbers, you will be able to make general decisions. If you learn some concepts about how educational assessments are constructed and some nuances about what their results mean, you will understand better what the data are telling you about your students' achievement.

Data About Student Learning

One way to organize and describe the different kinds of data about student learning is to use a four-quadrant framework (Brookhart, 2013). This framework allows us to group different kinds of data according to general type and purpose and to examine how they complement each other. It gives us some vocabulary to describe "assessment" in more specific terms. Figure 1.1 shows a four-quadrant framework for describing different kinds of assessments of student learning. The framework does not attempt to cover other data of interest to educators (e.g., student attendance, the number of books in the library, the ratio of students to teachers), just assessments of student learning.