It All Depends on What the Data Tell You

"Cheshire..." Alice began rather timidly, "would you tell me please, which way I ought to go from here?" "That all depends a good deal on where you want to get to," said the Cat.

—Lewis Carroll, Alice’s Adventures in Wonderland

Perhaps the Cheshire Cat’s answer to Alice’s question should have been, “That all depends a good deal on what the data tell you.” According to Merriam-Webster’s Eleventh Collegiate Dictionary (2007), the word data is defined as “factual information (as measurement or statistics) used as a basis for reasoning, discussion or calculation” (p. 316). Teachers use different sources of assessment data to “reason” about student learning, to “discuss” what learning has been accomplished, and to “calculate” levels of achievement. In short, they use assessment data to draw inferences about learning to inform educational decisions.

Formative assessment functions as a component of a comprehensive assessment system. Formative assessment provides one source of data to inform educational decisions, specifically those decisions that guide instruction minute by minute, day by day (Leahy, Lyon, Thompson, & Wiliam, 2005). Other educational decisions are informed by different data sources, for example, data from the end-of-unit assessments or the end-of-year state tests. While all the remaining chapters focus exclusively on formative assessment, this chapter steps back to locate formative assessment in the larger assessment context.

We know from the field of educational measurement that “one assessment does not fit all” (National Research Council, 2001, p. 220). Different assessments have different purposes, and therefore they inform different decisions. In this chapter, we’ll consider the various sources of assessment data available
to teachers, what purposes they each serve, and how they inform a range of teacher decisions. When we are considering the array of available assessment data, we’ll also be able to see where formative assessment fits into the big picture of assessment.

THE BIG ASSESSMENT PICTURE

In 2001, the authors of Knowing What Students Know: The Science and Design of Educational Assessment (KWSK) advanced a model for an assessment system to serve multiple, decision-making purposes (National Research Council, 2001). The authors proposed that such systems should be coherent, comprehensive, and continuous (3Cs).

- A coherent assessment system is built on a well-structured conceptual base—an expected learning progression, which serves as the foundation of all assessments.
- A comprehensive assessment system “provides a variety of evidence to support educational decision making.” (p. 259)
- A continuous assessment system provides “indications of student growth over time.” (p. 259)

In Figure 3.1, we can see what a system characterized by the 3Cs might look like: a range of assessments, from minute by minute to the annual state assessments, providing different levels of detail about student learning over time to be used for various decision-making purposes.

While we remain at some distance from a comprehensive system built on the same underlying model of learning advocated by KWSK, nonetheless the framework set forth there helps us think about the different purposes assessments can serve. Taken together, all assessments in the system should provide a continuous picture of student learning, yielding data about the extent to which students have met or are on the way to meeting learning goals. Later in the chapter, we will look at the different assessments shown in Figure 3.1, but first, we need to consider the all-important issue in educational assessment: appropriateness for purpose. Or put another way, which assessments in the system are appropriate for each of the decisions teachers need to make?

**APPROPRIATENESS FOR PURPOSE**

The most important question teachers need to answer when they are using assessment data is “Do these data come from assessments that are appropriate to purpose?” Teachers need to know that the inferences they draw from the data are valid, and inferences will not be valid if the assessments are not appropriate to purpose.

Validity is the key issue in educational measurement. Validity refers to whether an assessment is measuring what it is intended to measure and can well serve the intended purpose. Validity is always related to a specific use of the assessment or the interpretation of the data yielded by the assessment (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

What an assessment measures is termed a **construct**. A construct is the specific characteristic—for instance, the ability, skill, understanding, psychological trait, or personal quality—that the test is intended to measure. For example, reading comprehension, number sense, and scientific inquiry are all constructs. When teachers are using assessment results, they need to be sure that the assessment is measuring the construct they think it is measuring. Otherwise, the inferences they draw from the data may be inaccurate and lead to inappropriate decisions. For example, if the assessment is intended to measure students’ reading comprehension, then it must measure the range of abilities, skills, and understandings comprising the construct of reading comprehension. If it only measures decoding skills or vocabulary, then teachers may draw inferences about students’ decoding skills or vocabulary knowledge. However, they will not be able to draw inferences about students’ reading comprehension because the decoding and vocabulary constructs do not comprise the full range of abilities, skills, and understanding comprising the reading comprehension construct.
Assessment reliability is another important issue in educational measurement. Reliability refers to how consistently an assessment measures what it is intended to measure (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). If a test is reliable, the results should be repeatable. For instance, changes in the time of administration, day and time of scoring, who scores the assessment, and changes in the sample of assessment items should not create inconsistencies in results. Reliability is important because it is necessary for assessment validity. If assessment results are not consistent, then we can conclude that the results are not accurate measurements of what the assessment is intended to measure.

Validity and reliability are important issues for all assessments, but particularly for those assessments where the consequences of student performance are very high—for example, the annual state tests that can have significant consequences such as student retention. These kinds of high-stakes tests will have (or should have) undergone a rigorous process to establish their technical quality. Establishing technical quality involves accumulating validity and reliability evidence to support the use of the assessment for the intended purpose. All the assessments shown in Figure 3.1 should be valid and reliable. In Chapter 5, we will discuss the issues of validity and reliability as they apply to formative assessment that is integrated into instruction—in particular for those assessments that occur minute by minute or daily—where the consequences for an inaccurate inference are not very great. If teachers make a mistake one day, there is a good chance this can be rectified by the inference they draw the next.

However, at the very minimum, all assessments, formative, benchmark, progress monitoring, or accountability, should be aligned to learning goals, be they long term or short term. Within this array, the purpose of formative assessment is to promote further learning. In this vein, Stobart (2006) has argued that overriding concern in the validity of formative assessment is consequential validity; in other words, does the use of the assessment result in effective learning in subsequent instruction? According to Stobart, absent consequential validity, formative assessment does not and cannot meet validity standards.

Now that we have established the importance of the ideas of validity and reliability in assessment, let's take a look at processes teachers can use to make sense of the data to make them usable in the classroom.

**MAKING SENSE OF THE DATA**

Teachers have to make sense of the data to render them usable in the classroom. Figure 3.2 shows four steps teachers can follow to make sense of the data.
Curriculum, instruction, and feedback are the means teachers use to assist students to reach learning goals. Either during the course of learning or after a period of learning, assessment data are gathered to evaluate the learning that is either developing or has been achieved. Teachers then analyze what the data show and interpret what they mean in terms of student learning. For example, from annual state assessments teachers might see that a large number of students have scored below basic in a strand of mathematics and might infer that they have a weakness in this particular area. However, we should inject a word of caution here about drawing inferences concerning either strengths or weaknesses from one assessment, especially if there are only a few items to measure the construct. Before jumping to any conclusions, teachers must look for corroborating evidence, for example, from other assessments or analysis of student work. With enough corroborating evidence in hand to support their inferences, teachers then come to a conclusion about what kind of decisions they will make to further learning based on their interpretation. Finally, they have to implement their decision and take action.

Of course the time frame for analyzing and interpreting the data, making decisions, and taking action will vary according to the purpose of the assessment. For example, decisions about program or curricular changes might occur from the analysis and interpretation of annual state tests and other kinds of benchmark tests, while decisions about what to do in the next lesson could result from the minute-by-minute or daily assessment of learning.
DIFFERENT ASSESSMENTS FOR DIFFERENT PURPOSES

As we have already discussed, assessments shown in Figure 3.1 serve different purposes. Each one covers different-sized chunks of learning, ranging from learning occurring on a minute-by-minute basis, to the cumulative learning that has taken place over the course of a year, which is sampled by the annual state assessments.

To provide a coherent picture of learning, all the chunks, regardless of size, should be connected together in a clear progression of learning. Essentially, each learning chunk is made up of the subgoals of the next, larger chunk. Subgoals are the building blocks that enable students to meet longer-term goals. For example, small chunks of learning occurring over a lesson or several lessons should be connected to the larger chunk of a unit, the larger chunk of a unit connected to the chunk of learning that takes place over a quarter, and the even larger chunks of quarterly learning need to connect to the ultimate goals, the state standards. Ideally, the standards themselves should be connected so they provide a clear and coherent picture of the knowledge, concepts and skills students need to acquire over the course of their schooling.

As we stated earlier, the assessments in the system need to align to the learning goals of the different learning chunks to provide teachers with the range of information they need. Because they are aligned to different-sized chunks of learning, they provide different levels of detail. For example, by their very nature annual state tests can only provide a gross measure of learning. In contrast, minute-by-minute or daily assessments provide a much greater level of detail. So how do the differences in the grain size of these different levels of information contribute to differences in the use teachers can make of the various assessments? In the next section, we’ll answer that question.

DIFFERENT GRAIN SIZES

In Figure 3.3, we can see a representation of assessments of different levels of granularity and the kinds of questions that teachers can ask of the each data source to get information about student learning.

The questions guide teachers’ process of “reasoning, discussion, and calculation.” Once teachers have concluded this process—in other words, once they have drawn inferences from the results—they need to decide on and then take various courses of action. Without action, the data serve no purpose, and merely become, in Sadler’s (1989) term, “dangling data” (p. 121)—knowledge
Figure 3.3 - Assessments of Different Grain Sizes for Different Purposes

**Quarterly Assessments**
- What have my students learned?
- Who has and who hasn’t met the benchmark?
- How are students performing on this test on those areas identified as weak on the state tests?
- What are the strengths and weaknesses in individuals’/groups’ learning?
- Does the evidence meet my expectations based on formative assessment data?
- Who are the students most in need?
- What are the strengths and weaknesses in curriculum and instruction?
- Is there evidence our improvement strategies are working?

**Annual State Assessments**
- What have my students learned?
- Have they met the standards?
- What are the strengths and weaknesses in individuals’/groups’ learning?
- What are the strengths and weaknesses in curriculum, instruction, and programs?
- How are subgroups performing currently and over time?
- What are the relative strengths in teaching and learning?

**Minute-by-Minute, Daily, Weekly, Assessments**
- Where are my students in relation to learning goals?
- What is the gap between students’ current learning and the goal?
- What individual misconceptions or difficulties are my students having?
- Are there any missing building blocks in their learning?

**End-of-Unit Assessments**
- What have my students learned?
- Have they met the goals of the unit?
- Does the evidence meet my expectations based on formative assessment data?
- Are there some students who need additional help to meet the goals of the unit?
- What help do they need?
- What improvements do I need to make in my teaching?
of students recorded somewhere in some system, not knowledge that teachers and students can use to inform next steps in learning.

Let’s now look at the actions teachers might take from each source of data.

**TAKING ACTION**

Figure 3.4 shows the action that can result from analyzing and interpreting the data from each type of assessment.

**WHERE DOES FORMATIVE ASSESSMENT FIT IN?**

So far, we have learned about different assessments for different purposes and the decisions about learning each assessment can inform. Now we are going to look specifically at where formative assessment fits in. Recall the definition of formative assessment presented in Chapter 2:

> **Formative assessment is a process that takes place continuously during the course of teaching and learning to provide teachers and students with feedback to close the gap between current learning and desired goals.**

In formative assessment, teachers collect data while learning is taking place. Formative assessment strategies are aligned to the short-term subgoals, which are the focus of the lesson, and data from them provide teachers with a steady stream of information to keep learning moving forward. While teachers get fine-grained information from formative assessment to guide teaching and learning day by day, other assessments linked to longer-term goals provide snapshots of progress on the way to meeting standards. The important point here is that the short-term goals should be subgoals of the longer-term goals so that all the assessments in the system are providing a coherent picture of learning.

The effective use of formative assessment depends on the judgments teachers make about the data and the action they take based on those judgments. If the learning goals of each lesson are subgoals of the teacher’s longer-term learning goals, the formative assessment strategies are aligned with the subgoals, and the teacher judgments are accurate, there really should be no surprises in the data from the assessments that provide snapshots of progress toward the larger goals. Either way, in addition to using other assessments to check on progress, teachers can also use them to check on their own judgments in formative assessment, assuming that the calibrating assessments are of high enough technical quality.
Figure 3.4  Action From Assessments of Different Grain Sizes

**Quarterly Assessments**
- Adjustments to curriculum, instruction
- Targeted intervention for students who are weak in specific areas
- Reporting to administrators, parents

**Annual State Tests**
- Adjustments to curriculum, instruction, programs
- Targeted intervention for students who are weak in specific areas
- Reporting to administrators, parents

**Minute-by-Minute, Daily, Weekly Assessments**
- Adjustments to ongoing instruction/learning
- Feedback to students

**End-of-Unit Assessments**
- Targeted intervention for students who have not met unit goals
PUTTING IT ALL TOGETHER

To help illustrate how teachers can use the different assessments available to them, we turn to the fourth-grade teachers at Harrison Elementary School. What is important to note from the outset is that although this scenario is from an elementary school, it is equally applicable to middle and high schools. Grade-level groups or even whole departments could come together and engage in the same practices we will see at Harrison Elementary.

It is the middle of June and the fourth-grade teachers are meeting to prepare for the upcoming school year. The teachers, who are gathered around a table in the teachers' lounge, are examining a report they generated from their districtwide database. They want the answer to the question: "How well did this year's third-grade students perform on the statewide reading assessment?" As they review the report, the teachers notice that about half of their incoming fourth-grade students score in the advanced and proficient categories for reading, and half score in the basic and below basic categories. Already, the teachers have a sense of the work cut out for them in the coming school year, and they are anxious to get more detailed information to help them understand student needs more fully.

The teacher leading the meeting, Ms. Watson, then proposes they look at the student performance by the subscales of the test. The others agree, and she quickly queries the database and generates another report. The teachers examine the report and see that students who scored at the "basic" and "below basic" levels performed well in word analysis but were weaker in vocabulary and reading comprehension.

So far the teachers have used data from the annual state tests, which are summative assessments. These assessments will have undergone a rigorous process to establish their validity and reliability—two central aspects of their technical quality. They examined aggregated data, which is when student data is combined so that individual performance or performance of particular groups cannot be identified. The aggregated data they reviewed showed a

These teachers are using the results of statewide tests, summative data, to begin to answer the questions: "What have my students already learned?" and "Have they met the standards?"

The teachers get more information from statewide tests in answer to their question "What have my students already learned?"
summary of the performance of all the incoming fourth-grade students on the statewide assessment. From this aggregated, summative data, the teachers were able to get an idea of the overall levels of achievement in reading of their incoming students, as measured by the assessment. As well as the aggregated achievement levels, they were also able to look at the subscales of the test, results of the items that test specific areas—in this case, the items testing vocabulary, word analysis, and reading comprehension—to find more information about strengths and weaknesses.

It is worth noting that when the third-grade teachers look back at the results of this test to answer the questions “What are the strengths and weaknesses in curriculum and instruction?” and “How can I improve my teaching?”, they will have the same information about their own students’ performance to give some guidance about improvements they could make to curriculum and instruction for the next year’s third-grade students. Now let’s return to the fourth-grade teacher meeting.

Once the teachers have reviewed the subscales, Ms. Watson says, “I suggest we look at the student scores on the school’s quarterly district reading inventory to see how much progress these students made over the year, and to see if the same patterns show up as we see on the state test subscales.” Reviewing the line graph, showing the four data points, the teachers notice that while student vocabulary and comprehension skills improved over the previous school year, the growth of these skills was still slower than growth in word analysis skills. “So we are seeing the same pattern in these results as we did in the state test results,” says Ms. Chapman. Ms. Watson agrees that they are seeing the same results and adds that they are making a good start on getting a handle on where are students are in reading and deciding the focus for their planning to begin the year.

At the end of their meeting, the teachers decide to repeat the same process to examine the assessment results of the specific students who will be in their classes the next school year. Individually, they examine aggregated class performance on the state test, then performance by subscales, and finally, they look at student performance on the quarterly district assessments. The teachers share their findings at the following grade-level meeting a week later. Toward the end of the meeting Ms. Watson sums up: “It seems that in general the individual class findings reflect those of the grade level. But there are some definite differences among our classes.” Ms. Hamley notes it was really clear from the reading inventory data that some of the individual students in each class have specific needs, particularly in her class and Ms. Chapman’s class with predicting and summarizing and that they are going to need a lot of help.
What have the teachers done so far in their meetings? They have used two sources of summative assessment data: one source to give them information about levels of performance students had reached in reading at the end of the school year; the other source to show achievement at four points throughout the year, and from which they could see students’ growth trajectory in reading. From these data, the teachers were able to answer the questions:

- What have my students learned?
- What are the strengths and weaknesses in individuals’ and groups’ learning?
- What are the differences among groups?

At this point, the teachers have a sense of the strengths and weaknesses of their incoming students and are able to begin some program planning for the start of the next school year. However, once the students begin school, they plan to use additional fine-grained assessment to have more detailed information about individual student’s needs.

We know from cognitive research the importance of prior knowledge to learning. Effective teachers identify student prior knowledge and build on it to make connections to new learning (National Research Council, 2000). At the start of the school year, the teachers continue their investigation about student prior knowledge and are still focused on the question, “What have my students learned?” The answers will provide the basis for understanding the needs of individuals and groups more fully, and enable teachers to plan instruction accordingly. Let’s return to the teachers and see what they do to gather more information.

Each of the fourth-grade teachers has a plan for how they are going to get more fine-grained information about their students’ knowledge and skills in reading so they can match their instruction to the learning needs. Among the strategies they use to elicit the information are:

- Visualization—drawing pictures and diagrams of their understanding of their ideas in text with explanations
- Write questions that will test their understanding of a passage
- Explain and show how to clarify meaning, for example, using commas or conjunctions as guides to where the author provides definitions of words that may be unfamiliar to the reader
- Summarize—telling the most important ideas in two or three sentences
It All Depends on What the Data Tell You • 33

• Write or tell what they already know from the text and writing or telling some questions that still need to be answered in the subsequent text
• Make predictions in response to questions based on the questions
• Read a passage and infer two things that were not explicitly stated in the text: Explain the reasons for their inferences

The teachers use the data from these strategies to corroborate the information about the students' skills they received from their analysis of assessment data. They do this because they know the value of data from multiple measures for decision making. Multiple measures increase the validity and reliability of the teachers' interpretations. The teachers are also aware that the students' reading status may have changed somewhat over the summer—students may have increased their skills through practicing reading or other interventions, but they may have also lost some of their skills. Either way, to ensure valid and reliable inferences of students' current reading levels, it is important to check exactly where the students are.

Not only do the teachers use these strategies to elicit detailed prior knowledge, but they continue to use the same strategies to give them fine-grained data to guide ongoing teaching and learning.

At their grade-level meetings they share strategies to elicit ongoing evidence. Ms. Hamley tells the group that today she tried something she hadn't done before which gave her valuable information. She asked a group of students to tell her what they already knew from part of a chapter they had read and then to write some questions that still needed to be answered. She reported that she could really see Ryan's problem with prediction because he wasn't able to clearly state what was already known and relate it to what he still needed to know. Ms. Ross wondered if Hailey had the same problem and told the group that she would try the same strategy with her.

Throughout the first quarter, the teachers use ongoing, frequent formative assessment. With this evidence, they are able to determine what the next steps in learning should be, anticipating that adjustments to teaching and learning made on the basis of the evidence should result in improved learning. And, of course, if it does not, they will use other formative assessment strategies to figure out why no improvement has occurred and what alternative action can be taken. In their classrooms, the process of using formative assessment is continuous: assess, make adjustments, assess again, make adjustments, and so on.
Toward the end of the quarter, the teachers administer the schoolwide reading inventory. The results from this assessment provide answers to these questions:

- What have my students learned?
- Who has and who hasn’t met the benchmark?
- How are students performing on those areas identified as weak on the state tests?
- What are the strengths and weaknesses in individual’s/groups’ learning?
- Does the evidence meet my expectations based on formative assessment data?
- Who are the students most in need?
- What are strengths and weaknesses in curriculum and instruction?
- Is there evidence our teaching strategies are working?

Returning to the teachers we’ll see how they use this information at their grade-level meeting to review the results from the assessment.

Ms. Watson begins, “You know, we always look at the results, but really because we continuously assess our students, we are usually not surprised by the results. We already have a good idea of where are students are.” “That’s right,” responds Ms. Chapman. “But I always feel like it is a good check. I see it as another source of evidence that is useful—it gives me a sense of whether my instruction is on the right track.” First, the teachers review the overall grade-level results, and they are pleased to see that there has been improvement in areas that showed up as weak when they analyzed the state assessment data before the start of the school year. Then together they review the results from each class. They notice that Ms. Hamley’s students have made considerable improvement in summarizing and predicting. All the teachers are impressed by the improvement, and a discussion ensues where Ms. Hamley shares some of the assessment and instructional strategies she has used.

The teachers’ use of frequent, fine-grained information from formative assessment goes on throughout the next quarter. At the end of the quarter, teachers examine the schoolwide reading assessment again. They continue the same process until the end of year. By the time they have the results of the last quarter’s reading assessment, they feel confident about the progress the students have made. Indeed, when they examine the results they are very pleased with the outcome. “I feel really good about what we have accomplished this year with these students. Across the board, they have made progress,” comments Ms. Ross. “Yes, I think that’s true,” says Ms. Chapman. “The program planning we did at the start of the year has paid off.”
"It's not just that, though," adds Ms. Hamley. "We've really done a good job of keeping track of our students throughout the quarter—that's also made a big contribution." The teachers note the gains that have been made and also discuss the rate of progress. This leads to a discussion about the program and whether certain elements should have been emphasized more. Ms. Watson makes a suggestion: "Before we go much further in thinking about the year and the program, I'd like to wait until we get the state test results. I want to find out if we are seeing the same things in the state test results as we are here—we'll have more to go on." There is general agreement about Ms. Watson's idea. When the teachers get the results of the state tests they engage in a review of their own program and instruction—what worked, what didn't, and what they could do to improve.

The teachers have now come full circle. Not only do they use the fourth-grade assessment results to evaluate their instruction, but they also use the third-grade results to engage in planning for the next year, just as they did the previous year.

What have we learned from these fourth-grade teachers' use of data?

✓ Data use matters:

The teachers worked together, analyzing data with the goal of taking action to improve student learning. They collaborated to plan the curriculum and instructional practices, and they learned from each other about what works.

✓ One size does not fit all:

The fourth-grade teachers used a variety of assessments to gauge student learning. Each of the assessments had a different purpose, ranging from assessments to determine proficiency levels (have my students met the standards?), to indicate progress (are my students making progress?), and to guide day-to-day instruction (how do I keep learning moving forward?)

✓ Data use is ongoing:

For these teachers, data use is not a single event. Rather, it is a coordinated and systematic approach for analyzing different sources of data to improve learning. Particularly important is the use of fine-grained formative data to guide ongoing teaching and learning.

Formative assessment should be part of a coordinated system of assessment that provides teachers with data to inform the different decisions they make.
need to make to support learning. As we saw at Harrison Elementary, the teachers used all the assessments in the system and moved back and forth from the large grain size—what achievement for students looked like over time (e.g., quarterly and annual assessments)—to the finer grain size—what achievement needed to look like for the students the very next minute or next day.

In the next chapter, we are going to look closely at how teachers determine the small chunks of learning that they assess through the process of formative assessment.

**SUMMING UP**

- Different assessments have different purposes, and they inform different educational decisions.
- Validity is the key issue in educational measurement.
- Reliability is a necessary but insufficient condition for validity.
- When teachers are making decisions based on assessment results, they need to be sure that the assessment is measuring the construct they think it is measuring.
- Teachers should not jump to conclusions about student learning based on the results of one assessment. Multiple measures are needed.
- To provide a coherent picture of learning, all the subgoals or chunks of learning, regardless of size, should be connected together in a clear progression of learning.
- Assessments must be aligned with learning goals.

**REFLECTION QUESTIONS**

1. How does what you now do in your classroom to assess student learning compare with what has been presented in the chapter?

2. How do you know that the assessments you use are valid and reliable?

3. What are your strengths in assessing student learning, and which areas that you have read about in this chapter would you like to develop further?