Chapter 8

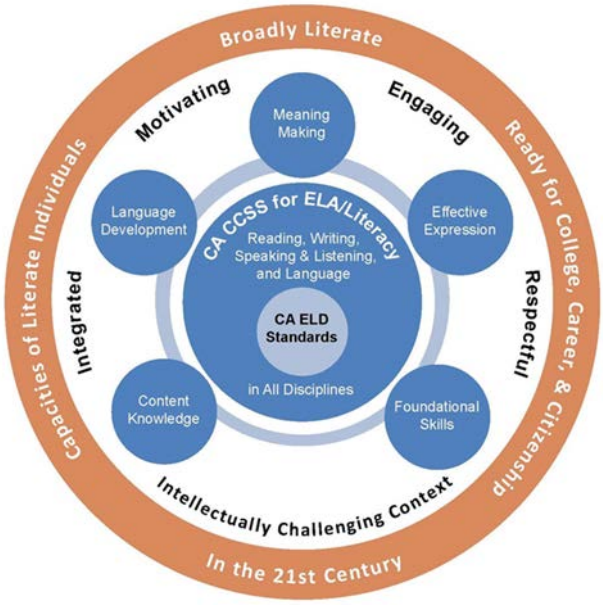
Assessment

Chapter at a Glance

Student achievement of the California Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects (CCSS for ELA/Literacy) and the California English Language Development Standards (CA ELD Standards) depends on educators' skilled use of assessment information. With

these standards, the landscape of assessment and accountability in California has experienced a dramatic shift. Not only do the standards present new goals for California educators as depicted in the outer ring of Figure 8.1 below, but the institution of the California Measurement of Academic Performance and Progress (CalMAPP) and the Smarter Balanced Assessment Systems represent a major shift in the intent of statewide assessment. “It is the intent of the Legislature…to provide a system of assessments of pupils that has the primary purposes of assisting teachers, administrators, and pupils and their parents; improving teaching and learning; and promoting high-quality teaching and learning using a variety of assessment approaches and types” (E60602.5(a)). Redundancies in the state are being reduced, and a sharper focus on the role of classroom assessment has emerged. Although a brief discussion of statewide assessment is included, this chapter will primarily address local, teacher- and school/district-driven assessments, and it is important to note the congruence of the intent of assessment at all levels—to improve teaching and learning.

Figure 8.1. Goals, Themes, and Contexts for Implementation of the CA CCSS for ELA/Literacy and the CA ELD Standards



This chapter describes what is involved in the skilled use of assessment to support student attainment of the CA CCSS for ELA/Literacy and the CA ELD

Standards--and ultimately the overarching goals of development of the capacities of literate individuals, being broadly literate, readiness for college, career, and citizenship, and acquisition of the skills necessary for living and learning in the 21st century. It begins with a discussion of the purposes of different assessments, followed by a description of different assessment cycles, the types and purposes of assessment within each, and the decisions that each assessment type can inform. Snapshots of teacher use of assessment in the cycles are included throughout. The use of primary language assessments and assessment accommodations of ELs are also discussed. In addition, the chapter provides information about the Smarter Balanced Assessment Consortium's (Smarter Balanced) annual statewide assessments, their optional interim assessments and formative assessment tools and practices, and the English Language Proficiency Assessment for California (ELPAC). Also included is a discussion of student involvement in assessment. The chapter concludes with a consideration of the technical quality of assessments to ensure that assessments yield accurate information for their intended purpose.

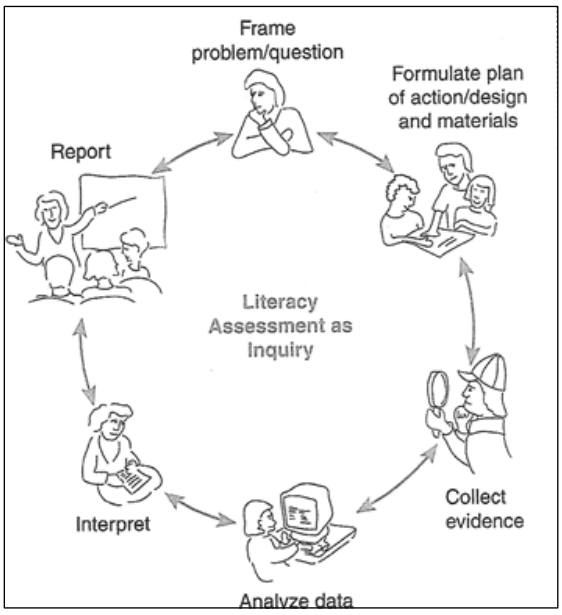
The CA CCSS for ELA/Literacy and the CA ELD Standards, as discussed throughout the framework, constitute shifts that have implications for assessment. First, the organization of the CA CCSS for ELA/Literacy is constant from kindergarten through grade twelve. The standards within each strand (reading, writing, speaking and listening[1], and language) can be backward mapped from the CCR Anchor Standards, meaning that students work on a relatively small number of broad competencies to move from novice to expert, so that teachers' work is simpler and communication is easier. Second, the standards encompass the full spectrum of language and literacy competencies from kindergarten through grade twelve, meaning that students apply and transfer skills from the earliest grades. Third, the standards encourage educators to "work and think big." "…[E]ach standard need not be a separate focus for instruction and assessment. Often, several standards can be addressed by a single, rich task…, [so that] students can develop mutually reinforcing skills and exhibit mastery… across a range of texts [and tasks]" (5). And finally, the standards recommend that language and

[1] As noted throughout this framework, speaking and listening should be broadly interpreted to include signing and viewing for Deaf and hard-of-hearing students whose primary language is American Sign Language.

literacy learning be connected with the academic disciplines from the earliest grades onward. Assessment, then, should enable educators to determine a student's trajectory in developing proficiency within and across the years in the standards. Assessment, also, should enable educators to determine a student's progress in language and literacy in application and in conjunction with larger projects and units, as well as in connection with other academic disciplines.

Formative assessment, discussed in Chapter 3 and below, is especially important in assessing the broad range of language and literacy skills and their application. Classroom teachers, school leaders, and professional learning providers should consider the support that educators will need to understand and implement formative, as well as summative, assessment effectively. Most importantly, educators need to interpret the results of assessment in order to plan and modify instruction. Collaborative professional structures, such as learning communities, should be the nexus of learning and work that teachers do relative to assessment (see Chapter 11) in which assessment is viewed as a cycle of inquiry. See Figure 8.2.

Figure 8.2. Literacy Assessment as Inquiry



Permission to be sought. (Graves, Juel, Graves, Wilson, and Calfee 2004)

Purposes of Assessment

Assessments are designed and used for different purposes. For example, an annual assessment designed to assess how well students have met a specific standard

(for example, CA CCSS for ELA/Literacy RI.4.8: *explain how an author uses reasons and evidence to support particular points in a text*) cannot serve the purpose of diagnosing a particular reading difficulty a fourth grade student is experiencing in achieving the standard. Nor can it provide substantive insights into how a student is beginning to understand what constitutes evidence in a specific text. In the use of any assessment, a central question is, "Am I using this assessment for the purpose for which it is intended?"

Assessment has two fundamental purposes: One is to provide information about student learning minute-by-minute, day-to-day, and week-to-week so teachers can continuously adapt and plan future instruction to meet students' specific needs and secure progress. This type of assessment is intended to assist learning and is often referred to as formative assessment or assessment ***for*** learning. Formative assessment occurs in real time, during instruction while student learning is underway (Allal 2010; Black and Wiliam 1998; Bell and Cowie 2000; Heritage 2010; Shepard 2000 2005). For example, a third grade teacher working with small groups of students on distinguishing their point of view from a particular author's is able to gain insights into students' developing skills through the use of strategic questions and can adjust instruction immediately based on the students' responses.

A second purpose of assessment is to provide information on students' current levels of achievement. Such assessments serve a summative purpose and are sometimes referred to as assessments ***of*** learning. They help determine whether students have attained a certain level of competency after a more or less extended period of instruction, for example, at the end of a unit which may last several weeks, at the end of a quarter, or annually (National Research Council [NRC] 2001). Inferences made by teachers from the results of these assessments can be used to make decisions about student placement, instruction, curriculum, and interventions, and to assign grades. For example, the English Language Proficiency Assessment is an assessment used for summative purposes to make decisions about the classification and placement of students according to English language proficiency levels.

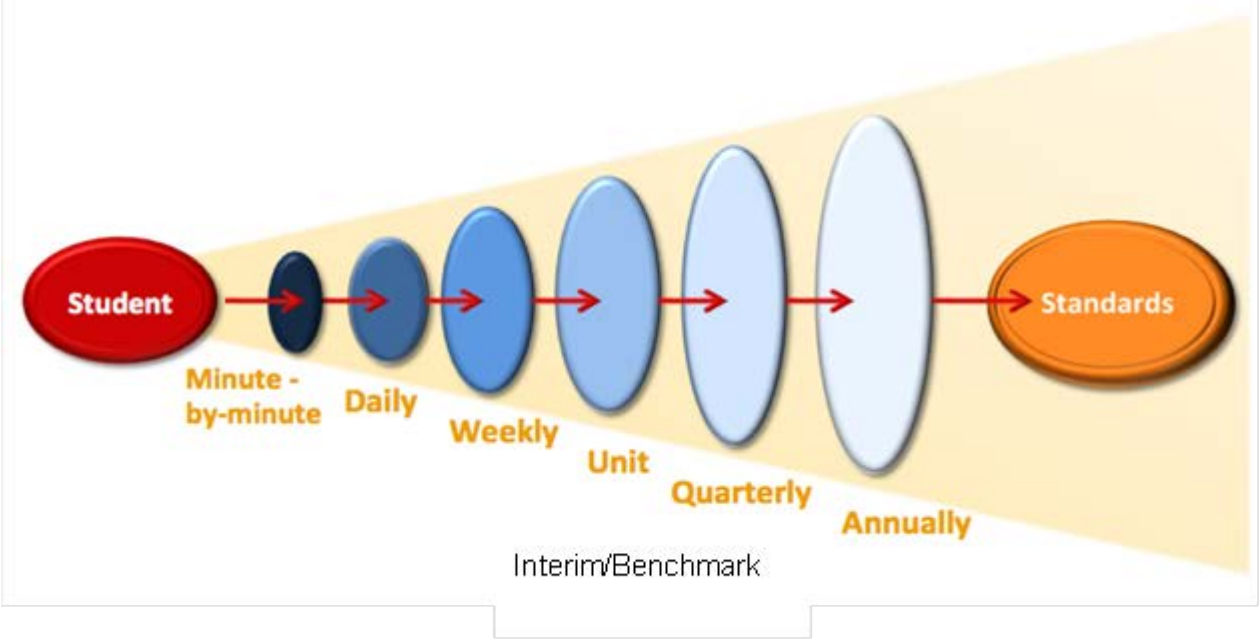
In order to maximize the use of assessment information for decisions related to student achievement of the CA CCSS for ELA/Literacy and the CA ELD Standards,

teachers need to make full use of assessment for both formative and summative purposes.

Assessment Cycles

One way to think about assessment for different purposes is to conceptualize assessment as operating in different cycles: short, medium and long (Wiliam 2006). Figure 8.3 shows a range of assessments within a comprehensive assessment system. Those assessments that are more proximate to student learning (i.e., minute-by-minute, daily, weekly) operate in a short cycle because they address a short period of teaching and learning. Short-cycle assessment serves a formative purpose because its intended use is to inform immediate teaching and learning. Assessments administered at the end of the year are long-cycle because they cover a much longer period of learning. They are primarily used for summative purposes.

Figure 8.3. Assessments in the System (Adapted from Herman and Heritage 2007).



Occupying a middle position between annual and formative assessment is interim or benchmark assessment: “assessments administered periodically throughout the school year, at specified times during a curriculum sequence to evaluate students’ knowledge and skills relative to an explicit set of longer-term learning goals” (Herman,

Osmundson, and Dietel 2010, 1). In Figure 8.3, they are referred to as quarterly assessments. Such assessments operate in a medium cycle because they address longer-term goals than those assessments more proximate to student learning but not as long-term as the annual assessment. Interim/benchmark assessments are generally used for summative purposes—evaluating what has been learned—although they may be used formatively if they inform decisions that teachers make within the school year to improve student learning. However, they are distinct from formative assessment as assessment ***for*** learning because, by their nature, they do not inform immediate teaching and learning. Unit assessments primarily serve a summative function but can be formative if the teacher uses the assessment information to improve learning before moving on to the next unit. Progress-monitoring assessments can be short, medium or long cycle, depending on whether they are administered after a longer or shorter period of instruction and they can serve both a formative and summative function. (For more information on screening, diagnostic assessment, and progress monitoring, see subsequent sections of this chapter).

Assessments within each cycle function best when they are part of a comprehensive, coherent and continuous system of assessment (as shown in Figure 8.4) that provides ongoing information to teachers throughout the year (NRC 2001). Within such systems, minute-by-minute, daily, and weekly assessment feeds into unit assessment, which, in turn, feeds into quarterly (interim or benchmark) assessments, and multiple interim assessments feed into the annual assessment of the standards. A comprehensive, coherent and continuous system of assessment provides mutually complementary views of student learning, ensures that assessments within each cycle are focused on the same ultimate goal—achievement of standards—and push instruction and learning in a common direction (Herman 2010).

Each assessment cycle provides information at varying levels of detail and inferences drawn from the assessment results are used to address specific questions about student learning and inform a range of decisions and actions. Figure 8.4 summarizes the types and purposes of the assessments within each assessment cycle.

158 Figure 8.4. Types and Uses of Assessments Within Assessment Cycles

Cycle	Grain Size	Methods	Information	Uses/Actions
Short				
Minute-by-minute	An instant or "snippet" of learning	Observation, Questions (teachers and students)	Examples of progress, of misunderstanding, of explanation	"Stoplight" – keep going, slow down, stop and find out more
Daily Lesson	A small chunk of learning; a short-*story*	Organize around beginning, middle, end, quizzes, written products	Story of a learning episode; pieces of a Standard	Follow-up for next lesson; feedback to class or individual students
Week	A big chunk of learning, 3-5 lessons	Introduction of goals for week, responses and products during week, review at end of week	Trends and progress across lessons;	Plans for start of new week; review progress and any revisions
Medium				
End-of-Unit/Project	A really big chunk of learning; a story with events; Progress on a standard	Process and product review. Planned feedback for group and individuals. Rubrics.	Grades Achievement of clusters of standards	For students; for class; for reflection on project
Interim/ Benchmark	Large	Portfolio Test	Status of achievement with respect to intermediate goals toward standards (results aggregated and disaggregated)	Making within-year and year-to-year instructional adjustments: Monitoring, reporting; grading; same-year adjustments to curriculum and instruction; professional learning and resource decisions

159

Cycle	Grain Size	Methods	Information	Uses/Actions
Long				
Annual	Very large	Smarter Balanced Assessment English Learner Proficiency Assessments for California Portfolio District/school created test	Status of student achievement with respect to standards (results aggregated and disaggregated)	Gauging student, school, and district's year-to-year progress: Monitoring, reporting and accountability; classification and placement; certification; adjustments to following year's instruction, curriculum, programs; grading; professional learning and resource decisions

Short-Cycle Formative Assessment

Short-cycle formative assessment is a process used by teachers and students *during instruction* that provides feedback to adjust ongoing teaching and learning to improve student achievement of intended instructional outcomes (McManus 2008, 3). Short-cycle formative assessment occurs when evidence of learning is gathered minute-by-minute, daily, and weekly from a variety of sources during ongoing instruction for the purpose of moving learning forward to meet short-term goals (i.e., lesson goals) (Black and Wiliam 1998; Council of Chief State School Officers Formative Assessment State Collaborative 2006; Heritage 2010; Popham 2010). In the remainder of this chapter, short-cycle formative assessment is referred to as formative assessment.

This type of assessment provides the most detailed information for teachers and their students. The idea of formative assessment, or assessment *for* learning, does not apply to a specific tool or assessment. This is not to say that an assessment cannot be used for formative assessment purposes—it can, but only if it provides information about students' learning status relative to the desired lesson goal and teachers can use it immediately to adjust their instruction. Many assessments marketed under the

formative assessment label do not (Perie, Marion and Gong 2009; Shepard 2005).

The sources of evidence available to teachers in short-cycle formative assessment are what students do, say, make, or write (Griffin 2007). For example, sources of evidence can be teacher-student interactions fuelled by well-designed questions (Bailey and Heritage 2008; Black, Harrison, Lee, Marshall, and Wiliam 2003), structured peer-to-peer discussions where the teacher observes (Harlen 2007), dialogues that embed assessment into an activity already occurring in the classroom (Ruiz-Primo and Furtak 2004 2006 2007), student work from well-designed tasks (Poppers 2011), and web-based reading assessments that provide immediate feedback (Cohen, Hall, Vue, and Ganley 2011).

The report of the FAST/SCASS Project (McManus 2008) emphasizes several features of formative assessment. First, "formative assessment is a *process* rather than a particular kind of assessment…. There is no such thing as a 'formative test'" (3). Second, "the formative assessment process involves both teachers *and* students…, both of whom must be actively involved in the process of improving learning" (3). Third, teachers must be clear about the ultimate goal of a unit and the sub-goals or stepping stones that are important along the way, "…from a learning progression teachers have the big picture of what students need to learn, as well as sufficient detail for planning instruction to meet short-term goals" (4). Fourth, criteria and evidence of learning need to be laid out at the beginning of the project and reviewed along the way. "…teachers must provide the criteria by which learning will be assessed… using language readily understood by students, with realistic examples of what meets and does not meet the criteria" (4). Another definition of formative assessment designed for the standards follows. Formative assessment is a process in which teachers employ an inquiry process at varying levels of grain size and cycle time, allowing them to monitor students' learning paths, to relate these movements to the expected instructional progression, and to decide how to modify instruction accordingly.

Whatever the source of the evidence, the teachers' role is to construct or devise ways to elicit responses from students that reveal where they are in their learning and to use the evidence to move learning forward (Sadler 1989). For effective formative assessment, teachers will need to be clear about the short-term learning goals (for

example, for a lesson) that cumulatively lead to students' attainment of one or more standards. They will also need to be clear about the performance criteria for the lesson goal—how will the students show if they have met, or are on the way to meeting the lesson goal. The evidence-gathering strategy can then be aligned to the performance criteria.

Questions that formative assessment can answer include the following:

- Where are my students in relation to learning goals for this lesson?
- What is the gap[2] between students' current learning and the goal?
- What individual difficulties are my students having?
- Are there any missing building blocks in their learning?
- What do I need to adjust in my teaching to ensure students learn?

Information from short-cycle formative assessment is used to make instructional adjustments in real time, to continue with the planned lesson, or to provide feedback to students that will help them take steps to advance their learning. (Feedback to students is discussed in the Student Involvement section of this chapter.)

An important point about teachers' use of evidence in formative assessment is that their inferences from the evidence and actions focus on individual students. This does not mean that instruction for students is on a one-to-one basis, but rather that individual needs are addressed in the context of a class of students. This orientation to individuals is necessary if students are going to have the opportunity to learn and progress equally (Heritage 2013). To do so, instruction needs to be contingent upon each student's current learning status. In other words, instruction has to be matched to where the students are so that they can be assisted to progress and meet desired goals.

While formative assessment evidence is not aggregated in the same way as medium- and long-cycle assessment information, teachers can categorize individual student responses to look for patterns across the class or for particular students who are outliers. For example, at the end of a lesson after students have completed a

[2] The gap refers to the distance between where the students' learning currently stands at particular points in the lesson (a lesson can be several periods) and the intended learning goal for the lesson. The purpose of short-cycle formative assessment is to close this gap so that all students meet the goal (Sadler 1989).

response to a question about a text, a teacher can quickly categorize them into students who are showing they understand, students who are nearly there, and students who need more work. The next day's instruction can be planned accordingly. Teachers of English learners should take great care in making these formative assessment decisions. Depending on their level of English language proficiency, some EL students may not be able to express their ideas orally about a topic during a class discussion; however, this does not necessarily mean that they do not understand it. In addition, an informal observation that indicates that EL students are not orally proficient in English should not determine how the students should be taught reading. EL students do not have to be proficient in oral English before they can learn to read in English (Bunch, Kibler and Pimental 2012). In addition, the CA ELD Standards clearly demonstrate that all ELs, regardless of their level of English language proficiency, are capable of engaging in intellectually-rich tasks at the same cognitive level. Teachers can use their in-the-moment formative assessment practices to ensure that the appropriate level of scaffolding is provided for EL students to do so. (For more information on scaffolding, see Chapter 3.) Primary language assessments also help to ensure that appropriate instructional decisions are made.

The use of technology that enables students to give immediate responses to teachers (e.g., clickers, mobile devices) can also help teachers with large numbers of students to get an ongoing sense of where students are during the lesson. For example, halfway through a lesson, a tenth grade teacher asks three or four questions related to multiple meanings and word phrases in a literary text the class has been analyzing. The results immediately appear as a pie chart on the smart board. Both teachers and students can quickly see how the class responded and can decide together if more work needs to be done in this area before the lesson progresses.

Snapshot 8.1 Formative Assessment in Grade Five

Fifth graders are working on the following CA CCSS for ELA/Literacy standards: a) applying the reading standard for informational text*: explaining how an author uses reasons and evidence to support particular points in a text, identifying which reasons and evidence support which particular points* (RI.5.8)*;* b) the writing standard: *produce clear and coherent writing in which the development and organization are appropriate to task, purpose, and audience* (W.5.4)*;* and the language standard *vocabulary use* (L.5.4-6),

particularly transition words to help their writing flow logically. They are writing an argument to encourage their readers to take more care of the natural environment. In their reading instruction, they have analyzed a text to identify where the sections with the "arguments," "counterarguments," and "evidence" to support the arguments are located. In their writing, they are learning to apply these ideas, as well as how to organize their arguments effectively.

While the students are involved in the independent writing part of the lesson, the teacher sits with a student to discuss his writing progress. She has a ring binder open to a page with these headings at the top: *Child's Name/Date, Research Compliment, Teaching Point*, and *What's Next for this Child*? Further down the page is a self-adhesive note that lists five students' names. These are the other children she plans to meet with during the session.

The teacher's initial purpose with the student is to follow up from two days ago when she provided him with feedback based on the evidence she had elicited from her interaction with him; in that interaction she determined that the student needed to provide stronger sources of evidence to support his argument. On this occasion, she wants to see how he has used her prior feedback:

T: *You're working on evidence? Tell me about it.*

S: *I found good information in the book of the Environmental Protection Agency and on the Internet.*

T: *And what do you think about what you found so far? Do you think that it supports your argument?*

S: *I guess….*

At this stage, the teacher reminds the student that the purpose of the evidence is to support his argument. She explains what "supporting an argument" is, in a way that is meaningful to a fifth grader, by telling him that you have to prove it with what is in the text or the readers may not believe you. She asks him to read his argument aloud. Having established that the focus of his argument is to "stop dumping in the ocean because all the beautiful animals we see are going to start vanishing," the teacher asks: *So, what evidence did you find to support that claim – that all the animals will die if we don't stop dumping? What evidence did you find that will help you to strengthen that argument, or prove it to your readers?* The teacher then helps her student to recognize which of the information he has located is from a reliable source and will be effective in supporting his argument. Satisfied that the student can move forward on his own to incorporate his evidence, she then asks him to go over the organization of his argument and to let her know where he will place the evidence. When the student does this, it is evident to the teacher that he has some confusion about the overall structure and that his writing needs to be reorganized. This is a moment in the interaction when she targets a teaching point for him. She goes over the organization with him and writes the organizational elements on a self-adhesive note and includes specific support, such as putting the evidence in order to help the flow or adding transitional sentences.

Throughout this interaction, the teacher has made notes in her ring-binder file. Under *Research Compliment* she has written that the student recognized the reliability of his source, in the section labeled *Teaching Point* she wrote that she had discussed how evidence supported his argument, and under the

heading *What's Next for this Child*? she wrote "organization and transitional sentences," noting that the student had problems organizing his writing to effectively convey his argument to the reader. By gathering evidence in the course of this interaction, the teacher was able to match her teaching points to the specific student's needs. Additionally, after several interactions of this kind, she may find that there are common needs among several students and might pull them together for a mini-lesson.

262

Snapshot 8.2 Formative Assessment in Grade Two

In a second-grade classroom comprised of native English speaking children and children who are English learners, the children have been working on retelling folktales they have read together in class and conveying the central message of the tale (CA CCSS Reading Standard for Literature). The EL children, in particular, have been working on using the past tense to indicate the tales happened in the past (ELD.PII.2.3). In this lesson, students are engaged in small group work, and during this time, the teacher selects groups of three students to recount one of the folktales the class has read that week. In this situation, she wants to give each student sustained opportunities to use language while she and the others in the group listen. She asks the first student to begin, then after a while asks the second child to carry on and so on. When the students have finished, the teacher asks them to say what they think the main message of the story is. Each child offers an opinion and there follows a discussion about whether there is agreement about the main message or not. From the recounting activity, the teacher has evidence that one student uses the past tense consistently and mostly with accuracy, while the other two do not. Two of the children are able to convey the message of the text, but another has not really grasped it. After her discussion with the group, she makes quick notes about each student and what is next for them instructionally. She continues this process with one more group before independent reading time is over, and she will find other opportunities during the week to assess other small groups in the same way.

263

Snapshot 8.3 Formative Assessment with EL Newcomers

In a secondary designated English Language Development (ELD) class, with newcomers whose experience in the United States ranges from three months to one year, the ELD teacher has worked collaboratively with the science teacher to create a five-week unit on animal behavior with the purpose of guiding his students through a deep exploration of the content through the language resources used to convey meaning. The two teachers have agreed that during science instruction, the science teacher will provide appropriate and strategic support so to his EL students so that they can fully participate in the science activities he has designed, gain understanding from the science textbook, and engage in collaborative discussions about the text and content. This strategic support includes using graphic organizers, providing increased opportunities for the students to discuss their ideas in small groups or pairs, and primary language support, including drawing attention to cognates.

During designated ELD instruction, the ELD teacher has agreed to analyze the science textbook and the activities the science teacher has designed in order to identify the language demands they

present and then to address the language demands in her class. This is the third class of the first week on the unit. Having formulated questions they would like to explore around the science topic, students have then perused a variety of texts on the topic to identify meanings and have charted language they think is critical for conveying their understanding of the topic. They now work in pairs to collaboratively write a description about what they have learned so far about one aspect of animal behavior, using as much of the language they have charted as they can. Before the lesson is over, the pairs write their description drafts on large sheets of paper to enable a discussion on what they have done and where they may go next to refine or add to their descriptions. The pairs read their descriptions to the class, and time is provided for the other students to ask questions and make comments. When one pair shares their description about animals and language, an animated conversation develops on whether animals have language or not. Julio, explains the thinking that went into the description that caused the lively discussion.

> Julio: *…First of all, I think that language is a way to **inform** others around you, your feelings or just a simple thing that you want to let know people what is the deal. And it can be **expressed** by saying it, watching a picture, or hearing it, you know what I'm saying? I don't know if you have heard about the kangaroo rat that stamps its feet to **communicate** with other rats. It's really funny cause we humans have more **characteristics** to **communicate** to each other, but we still have problems to understand other people. Characteristics like sound, grammar, pitch, and body language are some of them, while the rat only uses the foot* (he stamps the ground).

The teacher, who has been taking notes on the language students are using in the conversation, also notes that Julio is using some of the academic language the class has charted in both his writing and speaking and has, more importantly, done an effective job of conveying his understanding of the information from his research and persuading his peers using evidence. The ELD teacher decides to examine more closely the students' written descriptions, as well as the language they have used in their conversations, in order to make decisions about what language features of the science texts to focus on as she progresses in the unit. She also plans to make a copy of her notes to share with the science teacher when they meet later that week during collaboration time.

Medium-Cycle Assessment

Assessments that teachers develop, or that are included in the curricular materials and are administered at the end of a unit, are medium cycle. As noted previously, medium-cycle assessments occupy a middle ground between short-cycle formative assessments and long-cycle summative assessments. Some are used to inform instruction during the school year; others serve evaluative purposes.

End-of-Unit Assessment

End-of-unit assessments can serve a summative purpose to evaluate student achievement with respect to the goals of the unit. If such assessments are given to students before the end of the unit when there is still time to take some instructional action before moving on to the next unit, then they can also serve a formative purpose. In developing unit assessments, teachers will need to ensure that the goals of the unit are clear and aligned to standards. In other words, what is to be assessed must be well articulated and derived specifically from the standards. When teachers know *what* to assess, they can determine *how* to assess. In other words, they can decide on the most effective way that students can demonstrate the achievement of the goals.

End-of-unit assessments can help teachers answer such question as:

- Have my students met the goals of the unit?
- 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 next time I teach this unit?

Snapshot 8.4 End-of-Unit (Medium-Cycle) Assessment in Grade Seven

In a seventh grade classroom with native English speakers, recently reclassified ELs, and a group of ELs who are at the Expanding and Bridging levels of English language proficiency, the teacher has been taking the students through a five week unit: Persuasion Across Time and Space: Analyzing and Producing Complex Texts (*Understanding Language* 2013). This unit addresses multiple CA CCSS for ELA/Literacy and CA ELD standards simultaneously. The unit has four primary goals: 1) to read and analyze complex texts; 2) to involve students in reading, writing, listening, and speaking activities that are grounded in evidence from informational texts; 3) to engage students in disciplinary practices highlighting language and purpose that are responsive to audience; and 4) to build history/social studies knowledge through content rich non-fiction.

During the course of the unit, with intentional and strategic scaffolding by the teacher and considerable involvement in collaborative groups, the students engaged in close reading, collaborative discussions, and analysis of the text organization, grammatical structures, and vocabulary of persuasive texts on relevant topics. In the final part of the unit, the students analyzed the video, "The Girl Who Silenced the World for Five Minutes," compared and contrasted persuasive techniques in the video to one of the texts they had read, and produced a persuasive text of their own. The students' analysis of the video and written work served as the summative assessment for the unit. Using the students' work, the teacher was able to make a determination about the students' understanding of the purpose,

organization, and structure of persuasive texts and their ability to use various language resources (including vocabulary, complex grammatical structures, connecting words and phrases) to write a coherent and cohesive persuasive piece for a public audience.

After reviewing the students' responses, the teacher concluded that the students had made good progress toward meeting the goals of the unit, especially in regard to their understanding of persuasive techniques in different contexts (i.e., video and text). Examining her EL students' writing more closely, however, the teacher noticed that most of her students' writing was characterized by text that appeared more like spoken, every day language. In other words, their written arguments were not making use of connecting words and phrases (e.g., for example, therefore, consequently) to create cohesion, nor were they using many complex sentences to connect ideas and create relationships between them (e.g., *Even though governments are taking action*, it is not happening fast enough). This analysis of her students' writing helped the teacher to design lessons where she could show them examples of cohesion and complex sentences that connect ideas, model how to "unpack" the meaning in the texts, collaboratively construct similar writing with the students, and provide them with guided practice in writing related to the unit topic. She also planned to draw her students' attention to various examples of persuasive langauge used in arguments and to observe how her students incorporated them into their own writing in the next unit she had planned. In addition, she made a note to address these linguistic features directly when she teaches the unit the following year. (Snapshot adapted from Understanding Language 2013)

Interim or Benchmark Assessments

Interim or benchmark assessments, such as the Smarter Balanced interim assessments, are medium-cycle and address intermediate goals on the way to meeting standards. The Smarter Balanced assessments are aligned to the standards, and any other interim or benchmark assessment used by districts or schools will also need to be aligned to the standards. Typically administered quarterly or every six weeks, interim assessments cover a shorter period of instruction than long-cycle assessments and consequently give more detail about student learning. Results from interim assessments provide periodic snap shots of student learning throughout the year. These snapshots assist teachers to monitor how student learning is progressing and to determine who is on track to meet the standards and who is not. Results from these assessments can help teachers answer the following questions:

- What have my students learned so far?
- Who has and who has not met intermediate goals?
- Who is and who is not on track to meet the standards?

- How are students performing on this test on those areas identified as weak on the California state long-cycle assessments?
- What are the strengths and areas of need in individual's/groups' learning?
- Who are the students most in need? What do they need?
- What are the strengths and areas of need in my curriculum?
- What are the strengths and areas of need in my instruction?
- What improvements do I need to make in my teaching?

Administrators can also use interim assessment to address many of these questions that are relevant to their decision-making needs, for example, programmatic, professional learning, and resource decisions.

If students are not making desired progress, then teachers and administrators are prompted to consider if changes are needed in curriculum and instruction while adjustments can still be made before the end of the year. In this sense, even though they sum up a period of learning (over a few weeks or months) their use is also *formative* if adjustments to curriculum and instruction are made during the school year. Interim assessments also supply individual performance data. These data are useful to identify individual student's strengths and learning needs. In addition, while these results sum up a period of learning, they can also be used formatively if steps are taken to respond to individual student's needs while there is still time within the year. In instances where no action is taken to support student learning, the results from these assessments remain summative only.

Using data systems, including spreadsheets, interim assessment results can be aggregated and displayed in graphs and charts, so teachers can identify patterns in their students' performance, and disaggregated to provide information on the relative performance of individuals and subgroups.

If districts, schools, or individual teachers use commercially produced interim assessments, they must consider technical quality to ensure that the assessments are appropriate for the intended purpose. (See section on Technical Quality in this chapter.)

Snapshot 8.5 Interim (Medium-Cycle) Assessment in Grade One

All incoming first graders in a school are assessed at the beginning of the school year on the foundational skills of the ELA Standards, specifically, print concepts, phonological awareness, phonics and word recognition, and fluency. Results from their end-of-year kindergarten assessment are used to determine which sections of the assessment they receive. For example, if a student's results indicated a complete understanding of print concepts, that part of the assessment would be skipped, although close observations would be made during class to confirm last year's assessments. The teachers find the results from the beginning of the year assessment to be a useful starting point for their instruction, particularly as students may have either lost or made up ground during the summer. In addition, the teachers assess, or obtain help to assess, the primary language foundational literacy skills of their English learners who are new to the school and use this information for instructional decision-making.

After these initial assessments and appropriately designed instruction, students are administered interim foundational skills assessments every six weeks to determine progress. While the teachers are using opportunities during their instructional time on a more regular basis to gather evidence of students' skill development and adjust instruction accordingly, they find the results of the interim assessments are important for monitoring progress of individuals and the class as a whole, and to indicate to them where they need to make improvements in their teaching to ensure better progress. The teachers also think the results are useful as a means to evaluate and support their own judgments about students' skill development in the period between the interim assessments' administration.

Long-Cycle Assessment

Yearly assessments, such as the Smarter Balanced Assessment Consortium's (Smarter Balanced) annual assessments, are long-cycle assessments. (See elsewhere in this chapter for more information on Smarter Balanced.) They cover a year's worth of learning and, by their nature, provide a large grain size of information about student achievement relative to the standards. They sum up achievement after a year of learning and are therefore most appropriately used by schools and districts to monitor their own longitudinal progress and to ensure individual students are on track in academics and, for English learners, English language development. Schools and districts can ensure that students in dual language programs are maintaining steady progress toward biliteracy.

Long-cycle assessments are also useful to teachers and can help teachers answer such questions as:

- What did my outgoing class of students learn? Did they meet the standards I was

teaching them?

- What did my incoming class of students learn from last year to this year? Which standards did they achieve, and which did they not?
- What are the overall strengths and areas of need in my class's learning?
- What are the strengths and areas of need in individual's and groups' learning?
- What are the strengths and areas of need in my curriculum?
- What are the strengths and areas of need in my instruction?
- Have the improvement strategies I/we put in place worked?

With data systems, the assessment results can be aggregated so that schools and individual teachers can look for patterns in their students' performance. They can also be disaggregated to provide information on the relative performance of subgroups and the performance of individual students. School and district administrators can also use these assessment results to address questions relating to which students have and have not met the standards, and the relative strengths and areas of need in curricula and programs. Successful schools discuss long-cycle assessment in proactive ways so they can adjust the way they collaborate and teach.

Long-cycle assessment results are appropriately used for monitoring and accountability, reporting to parents on their individual child's achievement, adjustments to programs, curriculum and instruction for the following school year, teachers' reflection on their instructional practices, and identifying teachers' professional learning needs. The results also provide a starting point for the students' teachers the following school year, in terms of a picture of a class', a subgroup's and an individual's strengths and weaknesses. Below is a snapshot of the use for long-cycle assessment.

Snapshot 8.6 Long-Cycle Assessment in Grade Eight

An eighth grade teacher receives the annual assessment results for her students. Due to last year's results, when they were in seventh grade, she and her eighth-grade colleagues have worked diligently to improve the students' close and analytic reading skills with respect to literature and informational text, and their ability to write arguments effectively. To address weaknesses evident in the seventh grade assessment results, she has paid particular attention the literature standards: 1) *Cite textual evidence that most strongly supports an analysis of what the text says explicitly as well as inferences drawn for the text* (RI.7.1), and 2) *Compare and contrast the structure of two or more texts and*

analyze how the differing structures of each text contribute to its meaning and style (RL.8.5). She has paralleled the focus of the first literature standards in informational text as well. In addition, to address the weaknesses evident in the seventh grade writing results, she has worked with her students extensively on the standard: *Write arguments to support claims with clear reasons and relevant evidence* (W.6-8.1)

The first question she wants to answer when reviewing the annual assessments is: Have my students met the standards? This year, most students in her classes have achieved proficiency with respect to the reading and writing standards, and there is growth from last year. She is satisfied with the overall result and feels that the instructional focus that she and her colleagues identified for the year has yielded positive results. However, there are more students achievement levels 1 and 2 of the state achievement assessment than she would like, so she plans to follow up with her colleagues to look at the overall grade level performance to identify if there are students in other classes that are in this category. She also plans to investigate the scores of individual students who have not met the standard to see where specific areas of need lie and if the results of summative assessment are consistent with what she has observed through formative assessment. At the same time, she will examine the ELPAC for her English learners, some of whom have been in U.S. schools for only a couple of years and others for many years, as well as data about their literacy proficiency in their home/primary language. She wants to make sure that she uses all available information to design appropriately differentiated instruction for each of her students. This information provides evidence to help guide any changes in her instruction for next year's eighth graders. She also knows this will be valuable information for the ninth grade teachers.

Additional Methods of Medium- and Long-Cycle Assessment

Additional methods for evaluating student achievement in medium or long cycles include rubrics and student portfolios.

Rubrics

Performance assessments that require students to demonstrate learning through an oral, written, or multimodal performance task (e.g., a presentation, a report) are usually scored according to a rubric. A commonly accepted definition of a rubric is that of a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality (Andrade, Wany, Du, and Akawi 2009). Criteria should relate to the learning that students are being asked to demonstrate rather than the tasks themselves, and they should provide clear descriptions of performance across a continuum of quality (Brookhart 2013). The criteria should be linked to standards and reflect what is required to meet a specific standard or cluster of standards.

Descriptions of performance are usually presented within score levels, and the number of score levels depends on the extent to which criteria across the levels can distinguish among varying degrees of understanding and skills. The knowledge and skills at one level should differ distinctively from those at other levels (Lane 2013). If schools are using commercially produced performance assessments for high stakes assessment purposes, for example, placement or end-of-year grades, they will need to be assured that the rubrics have undergone a series of studies to provide evidence of their technical quality. (See the section on Technical Quality in this chapter.) Examples of such studies include review by language and literacy experts, review to ensure cultural and language sensitivity, and field tests to provide evidence that the rubric can differentiate performance across levels of the rubric and across grades.

For classroom assessment, in situations where stakes are not so high, teachers can also develop rubrics for their own classroom performance assessments. When creating rubrics, there are a few points to bear in mind. First, rubrics should express as clearly and concisely as possible what the performance at each level entails, so it is important to try to avoid unclear language. Before using the rubric, the language of the rubric will need to be explained to students. Second, in communicating expectations negative language should be avoided. Third, the gradations of quality need to be specifically articulated across levels. Figure 8.5 shows an example of a rubric for scoring an essay. The dimensions of the rubric are listed on the left-hand side and the criteria are clearly described across four levels of performance.

Figure 8.5. Essay Scoring Rubric (Andrade 2013)

	4	3	2	1
Ideas and Content	The essay has a clear thesis and supports it with evidence. Relevant comparisons b/w the paintings are made. Reasons for the similarities and	The essay has a clear thesis. Comparisons b/w the art works are made. The discussion of influences might be thin.	An opinion is given. The support for it tends to be weak or inaccurate. May get off topic.	The thesis and support for it is buried, confused and/or unclear.

	differences are discussed in terms of the influence of one art movement on another.			
Organization	The paper has an interesting beginning, developed middle, and satisfying conclusion in an order that makes sense. Paragraphs are indented, have topic and closing sentences, and main ideas.	The paper has a beginning, middle and end in an order that makes sense. Paragraphs are indented; some have topic and closing sentences.	The paper has an attempt at a beginning &/or ending. Some ideas may seem out of order. Some problems with paragraphs.	There is no real beginning or ending. The ideas seem loosely strung together. Poor paragraph formatting.
Voice & tone	The writing has a clear perspective, sophisticated style, and appropriate tone.	The style and tone are appropriate. The writer's perspective fades in and out.	The writer's perspective is obscure. The paper shows little awareness of audience and purpose.	The writing is flat, lacks a perspective, and uses an inappropriately formal or informal style and tone.
Word choice	The words used are descriptive but natural, varied and vivid.	The words used are correct, with a few attempts at vivid language.	The words used are ordinary. Some may sound forced or clichéd.	The same words are used over and over, some incorrectly.
Sentence fluency	Sentences are clear, complete, begin in different ways, and vary in length.	Mostly well-constructed sentences. Some variety in beginnings and length.	Many poorly constructed sentences. Little variety in beginnings or length.	Incomplete, run-on and awkward sentences make the paper hard to read.
Conventions	Spelling, punctuation, capitalization, and grammar are correct. Only minor edits are needed.	Spelling, punctuation and caps are usually correct. Some problems with grammar.	There are enough errors to make the writing hard to read.	The writing is difficult to understand because of errors.

It is preferable for teachers to design rubrics collegially as a group rather than as individuals. Taking advantage of how school teams already work together, as well as ensuring that the appropriate content expertise is represented in the group is a useful operating procedure for rubric development (Brookhart 2013). There is no rule of thumb for the frequency with which teachers should use rubrics. The use of a rubric depends on the purpose for which is being used (Brookhart 2013). For example, a rubric may be used at regular intervals during a writing assignment or once each week to assess oral reading. Given the time and effort to develop quality rubrics, it will be important to make sure that the learning goal or standard is best assessed by a performance task and a rubric, and that the investment in rubric development is worthwhile (Arter and Chappuis 2006).

Rubrics can improve student performance, as well as monitor it, by making teachers' expectations clear and by showing students how to meet these expectations. When teachers provide an evaluation of student work using a rubric, students should be clear about what they need to do to improve in the future. Rubrics can also help support student self- and peer-assessment (see section on Student Involvement, p. 36 for more information on self- and peer- assessment). Rubrics are particularly useful for assessing oral language development, particularly for English learners. For example, rubrics can focus teachers' attention on particular discourse practices, grammatical structures, and vocabulary as they observe and listen to their students' during collaborative discussions, oral presentations, and informal conversations. These observations can then guide instructional decision-making, including how teachers structure conversations, how they model different uses of English, and how they ensure that students receive ample exposure to rich oral language. The CA CCSS for ELA/Literacy and the CA ELD Standards provide useful information for designing rubrics to gauge students' progress in oral language (including vocabulary and presentations), collaborative discussions, writing, and other areas of the curriculum. Since the two sets of standards are aligned, teachers can work together to create streamlined rubrics using both sets of standards, as well as the standards' companion appendices and documents, so as not to create multiple rubrics for evaluating the same tasks.

Portfolio Assessment

Student portfolios are another useful source of evidence for making judgments of student learning and for making instructional adjustments and refinements. They can be either medium- or long-cycle assessments, depending on the length of the period of learning they cover.

A portfolio is a systematic collection of student work and related materials that tells the story of a student's activities, progress and achievement in a given subject area (Arter and Spandel1992; Venn 2000). Portfolios can provide a progressive record of student growth or they can be used to demonstrate mastery of specific learning goals and contain only samples of a student's highest achievement (Venn 2000). Whatever the purpose of the portfolio there should be sufficient samples related to specific learning goals that enable an evaluation of either growth or achievement (Stiggins, Arter, Chappuis and Chappuis 2006). The specific learning goals should be aligned to the standards and the evidence that is included in the portfolio should reflect either students' progress toward meeting standards or achievement of specific standards.

Portfolios can contain a range of evidence: student learning goals; samples of written work; images of work samples (e.g., digital images of models or other representations); audio samples (e.g., student narratives; oral presentations or read-alouds), video files (student performances; signed presentations); student reflections; teacher observations; teacher-student conference notes; and documentation of any other assessment results. Digital portfolios allow students to both assemble and publicly present their work while they also learn 21st Century skills, such as using technology, creativity, and communication, to name a few.

Assembling a portfolio should involve students in the selection of the content as well as student self-reflection on the contents related to why they were selected, what they represent and what they show about the student's learning (Arter and Spandel 1992; Stiggins et al. 2006).

Some questions teachers should keep in mind when using portfolios are:

- How representative is the work included in the portfolio of what students can really do?
- Do the portfolio pieces represent coached work, independent work, or group

work?

- Do the portfolio pieces represent student language and literacy progress across the content areas?
- How well do the portfolio items match standards?
- Are there clear criteria for judging the work and do the criteria represent the most relevant dimensions of student work products?
- Is there a method for ensuring that evaluation criteria are applied consistently and accurately? (Arter and Spandel 1992)

It is important to ensure that well developed criteria are used to evaluate the evidence assembled in the portfolio in terms of what the portfolio items reveal about student achievement and that a scoring process is in place, for example, if the portfolio is to be scored by one or more raters and when the scoring will be done. It is also important to consider if the portfolio should be rated as a whole (for example, a portfolio of student writing exemplifying achievement relative to specific standards), or if the portfolio is to be rated as individual samples. An additional consideration is the weighting of items in a portfolio. For example, will videos of performances of children's spoken language be weighted more or less than children's written responses?

Portfolios have the added benefit of providing valuable information to parents, particularly the parents of ELs and other language minority students who may not be completely familiar with the way U.S. schools work. When portfolios are designed for "telling the story" of student growth during a particular time frame, this communicates to parents how their children are developing in a variety of important areas. This information can guide parents to support their students to continue developing in these areas at home, thereby creating an opportunity for collaboration between schools and families.

Accommodations for ELs on Medium- and Long- Cycle Assessments

To ensure as accurate a picture as possible of students' learning status, assessment accommodations may be needed. The goal of an accommodation is to make an assessment more accessible for ELs and to produce results that are valid for these students. The intent is not to give them an unfair advantage over those who are not receiving that accommodation (Abedi and Ewers 2013). There are five major

considerations when selecting assessment accommodations for ELs: 1) Effectiveness: an accommodation must be effective in making an assessment more accessible to the recipients; 2) Validity: an accommodation should not alter the focal construct, i.e., the outcomes of accommodated and non-accommodated assessments should be comparable; 3) Differential Impact: an accommodation should be sensitive to student's background characteristics, and their academic standing, i.e., one size may not fit all; 4) Relevance: an accommodation should be appropriate for the recipients; 5) Feasibility: an accommodation must be logistically feasible to implement in the assessment setting (Abedi and Ewers 2013, 4). The Smarter Balanced assessment consortium will offer tools that improve the accessibility for all students while supporting accommodations to meet the particular needs of ELs (Smarter Balanced 2013b). Examples of accommodations, depending on the type of assessment, are bilingual dictionaries or English dictionaries. Because the type of accommodation useful to ELs will vary depending on the student's age, the student's level of English language proficiency, the topic, type of assessment task, and other factors, accommodations should be used strategically and intentionally.

Student Involvement

Whatever the assessment cycle, one goal of assessment is to promote a positive orientation to learning for students. Assessment, particularly when stakes are attached to it, creates a strong reason for learning. Assessment can also impact the learner's willingness, desire and capacity to learn (Harlen and Deakin Crick 2002). For example, if passing the test becomes the reason for learning, then students run the risk of developing a performance orientation, rather than a learning and mastery orientation (Ames and Archer 1988; Dweck 1999). Students with a performance orientation tend to use passive rather than active learning strategies, they avoid learning challenges, and their learning tends to be shallow rather than deep (Crooks 1988; Harlen and James 1997). While teachers can help students learn, only the students can actually do the learning. For this reason, successful achievement of standards will require students to develop a learning orientation evidenced by an interest in learning and meeting challenges, and a belief that effort, engagement in learning, and the development of learning strategies can lead to increased achievement.

If students are involved in the assessment process, they are more likely to develop a learning orientation than if they are solely passive recipients of test scores. They are also more likely to develop the skills of setting goals, managing the pursuit of those goals and self-monitoring, all important 21st Century skills (NRC 2012). Active student involvement in the assessment process is a vital element in the development of student self-direction in learning. Feedback is a crucial key to student involvement in assessment because it is a critical factor in the development of students' insight into their own learning and understanding (NRC 1999; OECD 2005).

Feedback

Feedback provides an indication to students of what they have done well – the degree to which they have met the learning goals – and what they can do next to improve their learning (Bangert-Drowns, Kulik, Kulik, and Morgan 1991). Importantly, feedback from either teachers or peers should avoid focusing on the student rather than the task (Kluger and DeNisi 1996), and on making comparisons with other students (Black and Wiliam 1998; Wiliam 2007). Basically, as Wiliam (2011) suggests, feedback should prompt a cognitive reaction (focused on active steps to achieve mastery) and not an emotional reaction (focused on anxiety and embarrassment).

Long- and medium-cycle assessments usually produce a score indicating the status of achievement. While the scores typically tell students what they have achieved, they do not tell them how or why they achieved what they did. The role of teacher feedback in relation to these types of assessment results is to assist students to understand where they were successful or not, and to set some goals with the students so they know where they need to improve and have some ideas of how to do so. This approach will mean that teachers need to spend time with students discussing assessment results and setting goals and strategies for improvement. Even when teachers have used a rubric and provided an evaluative score, students need feedback about how to improve. Although potentially time consuming, the pay-off for students is that assessment is more transparent, and the students are more oriented to goals and feel more ownership in future learning.

When considering what kind of feedback to give their EL students, teachers should focus first and foremost on communication and meaning-making, rather than

correcting every grammatical error, and they should encourage EL students to take risks when using English. These risks need to be taken in a safe and supportive environment where students are free to make mistakes as they learn English. For example, a student might say, “How fast the lava go?” If a teacher stops to correct every grammatical error, such as this one, the focus on meaning can be lost. Instead, teachers should think carefully about where and when to address this type of error. The teacher may recast the statement “How fast does the lava flow? Let’s read to find out.” In addition, the teacher could take note of the error and make plans to address the grammatical structure, as well as vocabulary, more explicitly during designated ELD. This is not to say that errors should be ignored. Rather, the way in which errors are addressed needs to be carefully considered in order to maximize student learning. Overcorrection, particularly when it feels like ridicule, can take the focus away from content knowledge development and discourage EL students from participating in conversations or inhibit their desire to write their ideas, which impedes their English language development.

Snapshot 8.7 Student Involvement in Assessment in Grade Four

Miss Nieto, a fourth grade teacher, has a discussion with each of her students about their reading scores from the interim assessment. In her meeting with Henry, she notes that the student has done well on the items related to using explicit details about the text and summarizing central ideas and is on track to meet the associated standards. She also discusses with the student that his scores indicate that he is not as strong in using supporting evidence to justify or interpret how information is presented. Miss Nieto and Henry have a conversation about why he thinks he scored lower on those items. He tells her that he thinks he is getting the idea of using evidence for justification but he still thinks it is difficult for him. She suggests that this will be a focus for the student between now and the next interim assessment and gives the student some ideas that can support his learning.

Feedback is particularly salient in the context of short-cycle formative assessment. Students can receive feedback in three ways: from their teachers, from peers, and through their own self-assessment. The purpose of the feedback is to close the gap between the student’s current learning status and the lesson goals (Sadler, 1989). Students need to be given opportunities to use the feedback, otherwise it does not serve the intended purpose.

Teacher Feedback

Three questions provide a frame for feedback in short-cycle formative assessment:

1. Where am I going?
2. Where am I now?
3. Where to next?

To answer the first question, both teachers and students need to be clear about the goal or target of the learning and what a successful performance of learning will be. Answering the second requires teachers and students to elicit and interpret evidence of learning. In other words, they need to decide where the students' learning currently stands in relation to the learning goal. Feedback addresses both the second and the third questions. The teacher provides feedback that indicates to the student where he or she has been successful and provides a hint or cue of what to do next.

Snapshot 8.8 Teacher Feedback in Grade One

Kathleen, a first grader, is preparing to read aloud to her teacher. Before she begins, Mr. Silverstein reminds her to think about the reading strategies they have been working on. The text states: *Fish swim in the river*. Kathleen, reading very slowly, says: *Fish…swim…in…the…water. No. That's not water. It doesn't begin with 'w.' R* (says letter name) *r* (letter sound*)... i...v... River! Fish swim in the river.* Mr. Silverstein provides feedback after the student finishes reading the sentence: *You did a very good job of using your decoding strategies to read the text accurately. Let's keep on reading and while you are reading think about: is what you are reading making sense, and does what you are seeing match with what you are reading? Just like you did when you noticed that* water *could not be the right word. Water made sense, but the letters indicated a different, equally sensible word: river.*

Peer Feedback

Peers are also sources of feedback for learning. Peer feedback has a number of advantages both for those students providing the feedback as well as those receiving it. It involves thinking about learning and can deepen students' understanding of their own learning. Research shows that the people providing the feedback benefit just as much as the recipient, because they are forced to internalize the learning goals and performance criteria in the context of someone else's work, which is less emotionally charged than their own (Wiliam 2006). The same three questions listed above apply to

peer feedback. Without clarity about the goal and the performance criteria peers will find it difficult to provide useful feedback to each other. Peers need to assess the status of classmates' learning against the same success criteria they use to check their own learning. Additionally, providing constructive feedback is a skill students need to learn, so instruction will need to focus on this as well. It is worth remembering that learners who are adept at giving and receiving feedback to complete learning activities are acquiring important 21st Century skills (NRC 2012).

Snapshot 8.9 Peer Feedback in Grade Three

In a third-grade class students are focusing on Speaking and Listening standard 3.4, one of several that focus on *presentation of knowledge and ideas.*Their learning goal is to write an informative speech to present to the class about a topic of interest to them. The criteria they have to bear in mind when writing their speeches include the following:

- Introduce your topic in a way that engages your audience
- Put your ideas in a logical sequence
- Make an impact on your audience with your ending

Once the students have an initial draft, they exchange their papers with a partner. Then the students provide each other with feedback. One student's feedback to her partner is: *I liked how you started your speech with a question…that's a good way of getting your audience's attention. I think your ideas are logical. I think it would be a better impact at the end of your speech if you go back to your question and maybe finish with a sentence that tells how you answered the question.*

Self-Assessment

Teacher and peer feedback are externally provided. When students are involved in self-assessment they are generating internal feedback. Generating and acting on internal feedback is a form of metacognition and self-regulation. Metacognition is basically thinking about one's thinking, and self-regulation refers to the ability of learners to coordinate cognitive resources, emotions and actions in the service of meeting learning goals (Boekaerts 2006). In the realm of 21st Century learning, metacognition and self-regulation are important skills (NRC 2012). The most effective learners are self-regulating (Butler and Winne 1995; Pintrich 2000; Schunk and Zimmerman 2008). Additionally, training students in metacognition raises their performance (e.g., Lodico, Ghatala, Pressley, Levin, and Bell 1983) and helps them generalize what they have

learned to new situations (Hacker, Dunlosky, and Graesser 1998). Because of the importance of metacognition and self-regulation to successful learning, teachers will need to pay attention to ensuring the students develop these skills in the context of language and literacy learning.

Self-assessment can be developed from the early grades onwards (Perry, VandeKamp, Mercer, and Norby 2002; Puckett and Diffily 2004). For example, a first-grade teacher provides her students with a graphic organizer with the headings: *date, book title, my goal today as a reader, pages read, how well did I meet my goals?* She asks her students to set goals for their independent reading time each day, and at the end of the session to think about how well they met the goals. During the week, when she has individual reading conferences with students, she reviews the self-assessment sheets and where students have not met their goal she asks them what the student did or needs to do to improve. Together, they set a strategy for the student to focus on. As well as providing the students with the opportunity for self-assessment, the teacher offers advice on strategies for improvement, which in turn become part of the students' internal repertoire of strategies that they can employ on subsequent occasions. In effect, they are developing the skills of self-regulation.

Self-assessment becomes more sophisticated as students gain more experience with the skill. For example, in a ninth grade science class where the teacher is integrating ELA and science standards, the students are involved in a short research project on distinct regions of the brain. As called for in the ELA writing standards for literacy in science (WHST.9-10.6), they are to display their information "flexibly and dynamically." Students in this class have time toward the end of every session to complete a reflection and planning log where they answer the following questions: *What was successful about your learning today? What difficulties or problems did you encounter? How did you manage those difficulties? Were you successful? If not, what plans do you have for dealing with them in the next lesson?* These logs serve as a means of self-assessment for students and support self-regulation because they have to think about strategies to solve difficulties. The logs are also sources of information for teachers about the progress students are making on their projects.

Assessment for Intervention

Screening, diagnostic, and progress-monitoring assessments are discussed in this section. Screening assessments identify students who may have difficulties, diagnostic assessment give specific information about the difficulties, and progress-monitoring assessment provides feedback on whether planned interventions to address the difficulties are working. These assessments can operate in short or medium cycles.

Universal Screening (Medium Cycle)

Universal screening is a critical first step in identifying students who are at risk of experiencing reading difficulties and who may need more instruction. Universal screening consists of brief assessments focused on target skills (for example, phonological awareness) that are highly predictive of future outcomes (Jenkins 2003).

An expert panel convened by the United States Department of Education's Institute of Education Sciences recommended that screening should take place at the beginning of each school year in kindergarten through grade two, and a second screening mid-year for kindergarten and grade one (Institute of Education Sciences [IES] 2009).

Because of students' development, the panel also recommended target areas for early screening. Kindergarten screening batteries should include measures assessing letter knowledge, phonemic awareness, and expressive and receptive vocabulary. As children move into grade one, screening batteries should include measures assessing phonemic awareness, decoding, word identification, and text reading. By the second semester of grade one the decoding, word identification, and text reading should include speed[3] as an outcome. Grade two batteries should include measures involving word reading and passage reading. For a reasonably accurate identification of students, the panel also recommended the use of two screening measures at each juncture. When schools or districts are selecting screening measures they should carefully examine the technical information available from the publisher's manual (IES 2009).

[3] As noted earlier, fluency rates do not apply to deaf and hard-of-hearing students who use American Sign Language as they are actually translating from one language to another when they storysign.

Diagnostic Assessment (Medium Cycle)

While the purpose of diagnostic assessments is to improve student learning, they should not be confused with short-cycle formative assessment. Formative assessment is used to guide ongoing decisions about student learning, whereas diagnostic assessment is used to provide targeted intervention for students who struggle and may fall well below classroom learning goals (Carnegie Council on Advancing Adolescent Literacy 2010).

Poor performance might reflect any one of a number of problems including, but not limited to, struggles with language and literacy. For example, if students are struggling with reading and understanding grade-level text, they may have short-term memory issues, or are not able to read fluently enough to focus their attention on comprehending the meaning, or cannot process connections across phrases and sentences in the text. Diagnostic assessment is the means to identify the precise source of the student's difficulty so that an appropriate intervention can be planned. Timely identification of students' difficulties is essential to ensuring the right intervention is made so students can progress.

Great care should be taken when approaching diagnostic assessments in English for English learners and Deaf students. For example, an EL student at the emerging level of English language proficiency or a Deaf student may appear to struggle with reading comprehension when reading a complex text in English. However, it could be that the student has not had sufficient opportunity to build up the language resources in English (including vocabulary and grammatical structures) or background knowledge needed to apply reading comprehension strategies. With appropriately adjusted instructional support, the students may demonstrate comprehension. Diagnostic assessments administered to EL and Deaf students in English need to be interpreted carefully. Teachers should consider possible linguistic and cultural biases of assessments. (See section on Technical Quality in this chapter.) use multiple types of assessments (including, where appropriate, primary language assessments) to gain a comprehensive portrait of students' learning needs, and compare the student to their EL or Deaf peers and not just native English speakers.

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) 1999), any test that uses language is a test of language. Therefore, for EL and Deaf students, every test written in English – despite the content area – is partially a test of their English language proficiency and may not adequately assess EL students' knowledge and skills (Abedi 2002). For this reason, it may be beneficial to assess EL students in their primary language in order to gain a more complete picture of their strengths and needs. However, it is important to bear in mind that it may not be appropriate to use primary language assessments with every EL student. For example, students who are literate or received formal education in their native language are likely to benefit from primary language assessments more than those who have not (Bowles and Stansfield 2008; Stansfield and Bowles 2006). In the same vein, evaluating emerging bilinguals' writing by looking at their Spanish writing side by side with their English writing can help teachers see the how the languages reinforce each other, and provide a bigger picture view of the students' developing biliteracy (Soltero-Gonzalez, Escamilla and Hopewell 2006).

A range of assessments is available for diagnosing the source of a student's difficulties and it will be important to ensure the appropriateness of these assessments for diagnostic purposes. (See the section on Technical Quality in this chapter). Administering and interpreting some diagnostic assessments requires special training and licensure so when selecting diagnostic assessments it will be important to determine if the school has access to the relevant professionals who can administer them. Teachers can benefit from working closely with reading specialists who have the necessary specialized knowledge to interpret diagnostic data and provide guidance regarding specific interventions (International Reading Association 2000). It is advantageous for the professionals available (for example, teacher, reading specialist and school psychologist) to work together in diagnosing a student's problem and planning appropriate interventions (Joseph 2002).

Progress Monitoring (Short or Medium Cycle)

Progress monitoring (sometimes referred to as curriculum-based measurement or curriculum-based assessment) is the practice of assessing students' academic performance on a regular basis for three purposes: 1) to determine whether students are profiting appropriately from the instructional program, including the curriculum; 2) to create more effective programs for those students who are not benefitting; and 3) to estimate rates of student improvement (National Research Center on Learning Disabilities 2006). To implement progress monitoring, a student's current level of performance is determined and goals are established for learning that will take place over a specific period. The student's academic performance is assessed on a regular basis (see IES recommendations above) and progress toward meeting the goal is determined by comparing the actual and expected rates of learning.

In addition to the general screening measures described above, a system of progress monitoring is recommended in response to intervention (RTI) programs (IES 2009). Based on available evidence, the panel convened by IES recommended that progress-monitoring assessments be administered to Tier 2 students at least once each month. For those students who are not making sufficient progress, a Tier 3 intensive intervention will need to be planned. Progress-monitoring assessments should be used in Tier 3 to determine the effectiveness of the intervention (IES 2009).

The National Association of State Directors of Special Education (NASDSE) identified nine essential characteristics for progress monitoring to be useful in an RTI context. Recommendations include that progress monitoring should assess marker variables that have been demonstrated to lead to the ultimate instructional target, be sensitive to small increments of growth over time, be administered repeatedly using multiple forms, be administered efficiently over short periods, and result in data that can be summarized in teacher-friendly data displays (NASDSE 2005, pp. 25-26).

If teachers, schools, or districts wish to adopt progress-monitoring assessments, careful attention will need to be paid to the technical quality of any proposed assessments to ensure they are appropriate for the intended purpose. (See the section on Technical Quality in this chapter.)

Mandated California Assessments

On October 2, 2013, AB 484 established the California Measurement of Academic Performance and Progress (CalMAPP) assessment system, which replaces the Standardized Testing and Reporting (STAR) program. The primary purpose of the CalMAPP system is to assist teachers, administrators, and students and their parents by promoting high-quality teaching and learning through the use of a variety of assessment approaches and item types.

Beginning in the 2014-2015 school year, student performance in grades three through eight and in grade eleven will be assessed by annual assessments developed by the Smarter Balanced Assessment Consortium and administered in the last 12 weeks of the school year. The eleventh grade assessment provides evidence of students' college and career readiness.

To ensure the assessments address the full range and depth of the CA CCSS for ELA/Literacy, and the breadth of achievement levels, Smarter Balanced assessments combine item types, including selected response (multiple-choice items with one or multiple correct responses and two-part items) and constructed response (students write a short text or long essay in response to a prompt). For example, for the third grade reading standard, *determine the main idea of a text; recount the key details and explain how they support the main idea* (RI.3.2), selected-response items could be used to assess *determine the main idea of a text; recount the key details,* while a constructed-response item could be used to assess *explain how they support the main idea.* A computer-based assessment, item response types will also include matching tables, fill-in tables, select or order text or graphics, and drag and drop.

For results to be timely and useful, achievement of students in kindergarten through second grade, and reading standards for foundational skills for kindergarten through grade five that are critical to every student's success in reading, are typically assessed through locally determined assessment. It is recommended that the foundational skills are assessed intensively at kindergarten through grade two and then systematically at grade levels above grade two. In selecting appropriate assessments for the purpose of assessing kindergarten through second grade students' achievement relative to standards, and assessing foundational skills, it will be important to refer to the

section on the technical quality of assessments in this chapter to ensure that the assessments are appropriate for the intended purpose.

Optional interim assessments are also available to be administered at locally determined intervals. The interim assessments are reported on the same scale as the year-end assessments and permit teachers to assess either clusters of standards or the full range of the CA CCSS ELA and Literacy Standards. In addition, Smarter Balanced has a digital library of formative practices and tools for teachers' use. These tools include model units and lessons with embedded formative assessment strategies for teacher use.

The Smarter Balanced end-of-year and interim assessments comprise computer adaptive tests and performance tasks, which are described in more detail below.

Computer Adaptive Tests

Computer-adaptive tests (CAT) tailor an assessment to individual students by presenting items based on a student's performance or responses to previous items in the test (Smarter Balanced 2013a). The Smarter Balanced summative assessments are being developed for use with CAT technology known as computer adaptive testing. The CAT assessment "engine" begins by delivering a short series of moderately difficult grade-level test items to the student, and then, depending on the student's initial performance, delivers items that are either more or less difficult. This process continues until the student's level of proficiency is determined (Smarter Balanced 2013a). For example, if a student has performed well on prior items, then more difficult items will be given thereafter, but if a student has performed poorly on prior items, then easier items are presented to the student. By matching the difficulties of new items more closely with a student's demonstrated level of performance, fewer items are needed. Some of the competencies assessed by CAT items include students' ability to use evidence to support their analyses (i.e., claims, conclusions, inferences) from reading different levels of text and their ability to edit and revise writing samples of different levels of complexity.

Because the test is taken on the computer, it is critical that students have developed the necessary technology skills, such as keyboarding, manipulating a mouse, and using pull down menus.

Performance Tasks

Performance tasks provide opportunities for students to demonstrate learning in ways that "emulate the context or conditions in which the intended knowledge and skills are actually applied" (AERA, APA, and NCME 1999, p. 137). They can take the form of demonstrations, oral performances[4], investigations and written products (Lane 2013). Performance assessments provide better possibilities to measure complex skills and communication, important competencies and disciplinary knowledge needed in today's society (Palm 2008) and important learning goals that cannot be easily assessed with other formats (Resnick and Resnick 1992).

The Smarter Balanced performance tasks, some of which are lengthy and will take considerable time to complete, emphasize deep knowledge of core concepts and ideas, analysis, synthesis, communication and critical thinking. For example, to assess the writing standards across all grade levels, full compositions, involving planning and revision are assessed with performance tasks. Similarly, performance tasks are used to assess grade 6-12 reading and writing standards for literacy in history/social studies, science and technical subjects. For instance, short research projects that involve applying research and inquiry as well as a demonstration of many 21st Century skills to produce a range of products (e.g., script for a presentation, PowerPoint, public service announcement) are assessed with end-of-year performance tasks. Other constructed-response tasks include asking students to respond to a question about a passage they have read and use details from the text to support their answer, to write an ending to story by adding details to tell what happens next, revising a paragraph by adding details to support an argument, and highlighting parts of a text that provide evidence to support a core idea of the text.

Smarter Balanced assessments also include multiple-choice items. For example, students watch a video and select a response that assesses comprehension, they read a text and are asked to select the most precise meaning of a word based on the context, and they select a sentence that best identifies an idea from the text.

[4] The term "oral language" refers to signed language for Deaf and hard-of-hearing students who use ASL as their primary language.

Assessments for Students with Significant Cognitive Disabilities

The Common Core State Standards are for every student, including students with significant cognitive disabilities. All students with disabilities will take the new assessments, with the exception of students who cannot achieve at or near grade level as identified by the members of the IEP team. These students are students with the most significant cognitive disabilities and make up approximately one percent of the population. They will require substantial supports and accommodations. These supports will allow them to have meaningful access to certain standards and assessments that are appropriate to the students' academic and functional needs. On October 1, 2012, California joined the National Center and State Collaborative (NCSC) Consortia as a Tier II state. The NCSC is committed to developing professional development modules and curriculum/instruction resources, creating alternate achievement standards and to developing a multi-state comprehensive assessment system for students with significant cognitive disabilities. The long-term goal is to ensure that students with significant cognitive disabilities achieve increasingly higher academic outcomes and leave high school ready for post-secondary options. The curriculum, instructional materials, and assessments targeted for students with significant cognitive disabilities are currently being developed. The NCSC is a standards-aligned assessment that is targeted to replace the previous alternate performance-based assessment known as the California Alternate Performance Assessment (CAPA). The alternate assessments based on modified achievement standards, known as the California Modified Assessment (CMA) targeted previously for towards two percent of students receiving special education services in California, will no longer be necessary with the transition to the Common Core State Standards. For more information, contact the California Department of Education Common Core Resources for Special Education website http://www.cde.ca.gov/sp/se/cc/.

Biliteracy Assessment

When instruction is provided in English and in a language other than English in bilingual or dual language programs, assessment for academic and language development progress in both languages should be implemented. This assessment should be designed according to the same principles and recommendations articulated

throughout this framework and throughout this chapter for both ELs and for students whose primary language is English. Frequently and closely monitoring students' progress, assessing in both languages used for instruction, and interpreting assessment results in accordance with the research on effective bilingual education ensures that students make steady and consistent progress toward full biliteracy.

English Language Proficiency Assessment

The English Language Proficiency Assessment for California (ELPAC), based on the CA ELD Standards adopted in 2012, will replace the California English Language Development Test (CELDT) in 2016-17. During the transition period, an item alignment study, test blueprint development, professional learning, and new item development will occur. The CELDT will be administered as usual until the ELPAC is fully operational.

Technical Quality of Assessments

When considering the use of Smarter Balanced, the ELPAC or other assessments to support student achievement of the CA CCSS for ELA/Literacy and the CA ELD Standards, it is important to keep in mind the purpose for which the assessment is intended. If an assessment does not provide accurate information for the decision-making purpose, its use may constitute misuse (Herman, Aschbacher, and Winters 1992).

This section elaborates the idea of the intended purpose of assessment. It will be particularly important to refer to this section when selecting assessments other than the Smarter Balanced assessments or the CELDT whose technical quality has already been established through rigorous studies.

Elements of Technical Quality

The idea of the "technical quality" of assessment refers the accuracy of information yielded by assessments and the appropriateness of the assessments for their intended purposes. There are three key elements related to the technical quality of assessments: validity, reliability, and freedom from bias (AERA, APA, and NCME, 1999). Each element is described here, and Figure 8.6, summarizing the key points for each, is included at the end of this section.

Validity

Validity is the overarching concept that defines quality in educational measurement. It is the extent to which an assessment provides accurate information for making decisions about student learning and the adequacy and the appropriateness of the use of assessment results for specific decision-making purposes (Herman, Heritage, and Goldschmidt 2011). No assessment is valid for all purposes. While people often refer to the validity of a test, it is more correct to refer to the validity of the *interpretations* that can be made from the results of a test. Validity is basically a matter of degree; based on its purpose, an assessment can have high, moderate or low validity. For example, a diagnostic reading test might have a high degree of validity for identifying the type of decoding problems a student is having, a moderate degree for diagnosing comprehension problems, a low degree for identifying vocabulary knowledge difficulties and no validity for diagnosing writing conventions difficulties. Similarly, the annual end of sixth grade assessments will have a high degree of validity for assessing achievement of standards for those students, but no validity for assessing the incoming group of sixth graders' achievement.

For an assessment to be valid for the intended purpose, there should be evidence that it does, in fact, assess what it intends to assess. Test publisher manuals should include information about the types of validity evidence that have been collected to support the use of the assessment.

Reliability

Reliability refers how consistently an assessment measures what it is intended to measure (Linn and Miller 2005). If an assessment is reliable, the results should be replicable. For instance, a change in the time of administration, day and time of scoring, who scores the assessment, and any changes in the sample of assessment items should not create inconsistencies in results.

Reliability is important because it is a necessary adjunct of assessment validity (Linn and Miller 2005). If assessment results are not consistent, then it is reasonable to conclude that the results do not accurately measure what the assessment is intended to measure. A general rule of thumb for reliability is that the more items on an assessment the higher the reliability. Reliability is assessed primarily with statistical indices.

Publishers' manuals should provide information about the reliability evidence for an assessment and the relevant statistical indices.

A variety of factors can influence the reliability of an assessment. For example, if a test is administered in an extremely hot or noisy room, students may not be able to complete the test to the best of their ability. If students are asked to provide an oral presentation when the instructions or expectations have not been made clear, this affects the reliability of the performance assessment. A number of other factors, including students' health, level of stress, and motivation can affect the reliability of an assessment. Teachers should use their judgment in interpreting assessment results when they suspect students are not able, for whatever reason, to perform to the best of their abilities. It is equally important for teachers to understand that a test or performance assessment may be reliable but not valid. For example, a student may consistently do well on an assessment, but the assessment may not be measuring what it claims to measure.

Freedom from Bias

Bias is the presence of information in an assessment or a condition of the assessment that unfairly disadvantages a student or group of students so that the student(s) are unable to accurately show what he or she knows and can do with respect to the content of the assessment. As a result, the assessment results may underestimate the student's achievement or reflect abilities that are not related to the assessment's content (Abedi and Lord 2001). Bias arises from tests that favor students of a particular gender, ethnicity, cultural background, geographic location, disability and primary language. An assessment that is free from bias will produce the same scores for students of the same attainment level, irrespective of their demographic subgroup.

Popham (1995) identifies two forms of bias, offensiveness and unfair penalization. Offensiveness occurs when the content of an assessment offends, upsets, or distresses particular subgroups, thus negatively influencing the test performance of these students. Items that present stereotypes of girls, boys, or particular cultures, or that portray certain groups as inferior, could adversely affect certain students' performance.

Unfair penalization occurs when the test content makes the test more difficult for some students than for others. Bias may occur, for example, if a test includes vocabulary that is unfamiliar to students because of their culture or geographic location. Bias may also occur if the test contains images that are more familiar to one group than another, or demands language skills beyond those of the targeted students. For example, if a reading assessment contains vocabulary related to rural life, then inner city students will potentially be more disadvantaged than rural students. In addition, bias occurs when assessments that are based on letter-sound principles are used with students who do not have access to the sounds of language (i.e., students who are deaf or hard-of-hearing).

Assessment developers typically go to great lengths to make sure assessment items are not biased. Examine the publishers' manual for evidence that item reviews to guard against bias have been conducted.

Validity, reliability and freedom from bias are all necessary conditions for all assessment. They are not interchangeable (Linn and Miller 2005). For example, an assessment may offer consistent results (high reliability) without measuring what was aimed at (low validity); and conversely a measurement with all the hallmarks of validity may not have high reliability.

Figure 8.6. Key Points in Technical Quality of Assessments: Long- and Medium-Cycle Assessments

Technical Quality	Key Points
Validity	• Assessments need to be valid for the intended purpose • The extent to which the information the assessment provides is accurate, adequate, and appropriate for a specific decision-making purpose • While people often refer to the "validity of a test," it is more correct to refer to the validity of the *interpretations* that can be made from the results of a test • No test is valid for all purposes
Reliability	• Consistency of the test results, repeatedly and over time • Results of a test are reliable if they are replicable (despite changes in test administration and scoring, e.g., time of administration or who

	scores a test) • Reliability is important because it is a necessary, but not sufficient condition for validity. If assessment results are not consistent, then it is reasonable to conclude that the scores do not accurately measure what the test is intended to measure
Freedom from Bias	• Information or condition in an assessment that unfairly disadvantages a student or groups in showing knowledge in the content • An assessment free from bias produces same scores for students at the same attainment level, despite students' demographics (e.g., gender, ethnicity, primary language) • Two forms of bias: (1) offensiveness – content offends or upsets particular subgroups, (2) unfair penalization – content more difficult for some students than others

In the next section, the ideas of validity, reliability and bias are considered in the context of formative assessment practice.

Technical Quality and Formative Assessment

In formative assessment, the evidence generated by a variety of means is intended to provide information about the students' learning progress in relation to the specific learning goals (i.e., for a lesson) and to be used to inform immediate decisions about next steps in teaching and learning. As alignment to goals is important for annual and interim assessment, so it is for formative assessment. Teachers will need to be clear about the specific learning goals (what students will learn, not what they will do) and what a successful performance entails. For example, learning goals for third grade readers might be to 1) understand that the main idea is the author's message about a topic, minus all the details; and 2) determine the main idea of a text. The performances of understanding and skills for these goals would be for the students to 1) explain the main idea of a text; 2) locate where the author directly expresses the main idea (message) in text; and 3) explain how the important details describe the main idea. The teacher can align her evidence gathering strategies with the goals and performance criteria.

For assessment to be formative it must be both timely and produce information that can inform teaching practice during its ongoing course (Erickson 2007). For this

reason the immediate or proximate timing of evidence is a key component of formative assessment validity. In addition, for formative assessment to be valid the resulting information must also yield substantive insights into students' current learning status that can be used in subsequent pedagogical action (Heritage 2013).

An important point about validity in formative assessment concerns the consequences of the assessment use. Because action resulting from the use of formative assessment evidence is intended to produce benefits to student learning, consequences represent an important component of the validity of such assessment. Even if assessments are formative in intention they may not be so in practice if they do not generate further learning (Stobart 2006; Wiliam and Black 1996).

Reliability for classroom formative assessment takes a very different form because errors in instructional decisions can be rectified quickly through gathering more evidence of learning (Shepard 2001). Reliability in relation to instructional decisions can be thought of as "sufficiency of information" (Smith 2003, p. 30). In other words, teachers have to be confident that they have enough information about the student's learning to make a reasonable judgment about the current status of that learning. This idea of sufficiency of information for reliability argues for multiple sources of evidence before a teacher makes an instructional decision. The wider the range of information, and the more frequently the information is collected the more accurately learning can be inferred (Griffin, Murray, Care, Thomas, and Perri 2010). In practical terms, this might mean that before making a judgment about student learning on specific features of language, a teacher has evidence from students' oral language production, from a quick-write and from a text that has been underlined by the students to identify the specific language feature in question. The more this kind of evidence can be gathered in the context of everyday learning tasks, and so not take time away from instruction, the more the number of learning events as assessment tasks can be increased to improve the reliability of the information gathered (Linn and Baker 1996).

Because reading, writing, speaking and listening skills do not develop in lockstep across all students, formative assessment is inevitably personalized and teachers will need to employ strategies that tap into individual's knowledge and skills. Whatever evidence sources a teacher selects, they should account for the range of students

present in the class so that all students have the opportunity to show where they are in their learning and have the prospect of moving forward from their current status. For example, well-designed questions and tasks that are sufficiently open-ended can give all students the opportunity to reveal their learning. Similarly, formative assessment should not include any elements that would prevent some students from showing where they are relative to goals.

Figure 8.7. Key Points in Technical Quality of Assessments: Short-Cycle Formative Assessments

- Evidence gathered by the teacher is in alignment to specific student learning goals derived from standards
- Evidence gathered needs to be timely and contain information that can inform teaching
- Validity of formative assessment mainly lies in the use of evidence: information gathered must yield substantive insights to students' current learning status that will be used for pedagogical action in order to move students toward achieving learning goals
- Reliability pertains to gathering enough information (e.g., multiple sources) about student learning in order to make a reasonable, accurate judgment for subsequent instructional decisions
- To ensure freedom from bias, evidence gathering should be personalized to students so all students have the opportunity to show where they are in their learning

Conclusion

The use of assessment by teachers is a critical component of students' achievement of the CA CCSS for ELA/Literacy and the CA ELD Standards. Only when teachers have a range of accurate information about student learning can they be in a position to make decisions that will advance learning. Key to informing the decisions teachers need to make is a system of coherent assessment that provides different levels of detail for different decision-making purposes. Within such an assessment system, districts and school personnel will need to strike the right balance in terms of the range of available assessments to teachers from state or district mandated, to those adopted by individual schools, to assessments embedded in curriculum materials, to ongoing day-by-day assessment that teachers conduct during instruction.

Assessment operates in the service of learning and striking this balance involves careful consideration of the decisions that teachers need to make, when in the school

year they need to make them to ensure student progress, and the assessment tools they need to inform their decision-making. In combination with the right assessments for the right purposes, teachers' skillful use of assessment to support learning will go a long way to ensuring that students in California meet the ambitious language and literacy standards that have been set forth.

Works Cited

Abedi, Jamal. 2002. “Assessing and Accommodations of English Language Learners: Issues, Concerns and Recommendations.” *Journal of School Improvement* 3 (1): 83–89.

Abedi, Jamal, and N. Ewers. 2013. *Smarter Balanced Assessment Consortium: Accommodations for English Language Learners and students with disabilities*: A research-based decision algorithm. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-represented-students.pdf

Abedi, Jamal, and C. Lord. 2001. *Assessment and Accommodations for English Language Learners: Issues and Recommendations.* (CRESST Policy Brief No. 4). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Allal, L. 2010. “Assessment and the regulation of learning.” In P. Peterson, E. Baker, and B. McGraw (eds.), *International Encyclopedia of Education Vol. 3,* 348–352. Oxford: Elsevier.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 1999. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Ames, C., and J. Archer. 1988. “Achievement Goals in the Classroom: Students’ Learning Strategies and Motivation Processes.” *Journal of Educational Psychology* 80: 260–267.

Andrade, H. G. 2013. Essay scoring rubric. Unpublished document.

Andrade, H., X. Wang, Y. Du, and R. Akawi. 2009. “Rubric-Referenced Self-Assessment and Self-Efficacy for Writing.” *The Journal of Educational Research* 102 (4): 287–302.

Arter, J. A., and J. Chappuis. 2006. *Creating and Recognizing Quality Rubrics*. Portland, OR: Educational Testing Service.

Arter, J., and Vicki Spandel. 1992. “Using Portfolios of Student Work in Instruction and Assessment.” *Educational Measurement: Issues and Practice* 11 (1): 36–44.

Bailey, Alison L., and Margaret Heritage. 2008. *Formative Assessment for Literacy, Grades K-6: Building Reading and Academic Language Skills Across the Curriculum*. Thousand Oaks, CA: Corwin/Sage Press.

Bangert-Drowns, Robert L., Chen-Lin C. Kulik, James A. Kulik, and Mary Teresa Morgan. 1991. "The Instructional Effect of Feedback in Test-like Events." *Review of Educational Research* 61: 213-238.

Bell, B., and Cowie, B. 2000. "The Characteristics of Formative Assessment in Science Education." *Science Education* 85: 536–553.

Black, Paul J., and Dylan Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles Policy and Practice* 5: 7–73.

Black, P., C. Harrison, C. Lee, B. Marshall, and Dylan Wiliam. 2003. *Assessment for Learning: Putting It into Practice*. New York, NY: Open University Press.

Boekaerts, M. 2006. Self-Regulation and Effort Investment. In K. A. Renninger and I.E. Siegel (eds.). *Handbook of Child Psychology: Vol. 4. Child Psychology in Practice* (6th ed.)*,* 345–377*.* New York: Wiley.

Bowles, M., and C. Stansfield. 2008. *A Practical Guide to Standards-Based Assessment in the Native Language* (NLA—LEP Partnership). http://www.ncela.gwu.edu/files/uploads/11/bowles_stansfield.pdf (accessed June 10, 2013).

Brookhart, S. M. 2013. *How to Create and Use Rubrics for Formative Assessment and Grading.* Alexandria, VA: ASCD.

Bunch, George C., A. Kibler, and A. Pimental. 2012. "Realizing Opportunities for English Learners in the Common Core English Language Arts and Disciplinary Literacy Standards." In Kenji Hakuta and M. Santos (eds.), *Understanding Language: Commissioned Papers on Language and Literacy Issues in the Common Core State Standards and Next Generation Science Standards,* 1–16. Palo Alto, CA: Stanford University.

Butler, D., and P. Winne. 1995. "Feedback and Self-Regulated Learning: A Theoretical Synthesis." *Review of Educational Research* 65: 245–281.

Carnegie Council on Advancing Adolescent Literacy. 2010. *Time to Act: An Agenda for Advancing Adolescent Literacy for College and Career Success.* New York: Carnegie Corporation of New York.

Cohen, N., Hall, T. E., Vue, G., and Ganley, P. 2011. "Becoming Strategic Readers: Three Cases Using Formative Assessment, UDL, and Technology to Support Struggling Middle School Readers." In P. Noyce and D. T. Hickey (eds.), *New Frontiers in Formative Assessment,* 129–140. Cambridge, MA: Harvard Education Press.

Council of Chief State School Officers (CCSSO) Formative Assessment State Collaborative. 2006. *Definition of Formative Assessment.* Washington, DC: CCSSO.

Crooks, T. J. 1988. "The Impact of Classroom Evaluation on Students." *Review of Educational Research* 58: 438–481.

Dweck, Carol S. 1999. *Self-Theories: Their Role in Motivation, Personality and Development.* Philadelphia, PA: Psychology Press.

Erickson, F. 2007. "Some Thoughts on "Proximal" Formative Assessment of Student Learning." *Yearbook of the National Society for the Study of Education* 106: 186–216.

Gersten, Russell, Donald Compton, Carol M. Connor, Joseph Dimino, Lana Santoro, Sylvia Linan-Thompson, and W. David Tilly. 2008. *Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention for Reading in the Primary Grades. A Practice Guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Graves, Michael F., Connie Juel, Bonnie B. Graves, Kathleen M. Wilson, and Robert C. Calfee. 2004. *Assessment and Lesson Plans for Teaching Reading in the 21st Century* 3rd ed. Upper Saddle River, NJ: Pearson.

Griffin, P. 2007. The Comfort of Competence and the Uncertainty of Assessment. *Studies in Educational Evaluation* 33: 87–99.

Griffin, P., L. Murray, E. Care, A. Thomas, and P. Perri. 2010. "Developmental Assessment: Lifting Literacy Through Professional Learning Teams." *Assessment in Education: Principles, Policy and Practice* 17: 383–397.

Hacker, D. J., J. Dunlosky, and A. C. Graesser (eds). 1998. *Metacognition in Educational Theory and Practice.* Mahwah, NJ: Lawrence Erlbaum.

Harlen, W. and R. Deakin Crick. 2002. *A Systematic Review of the Impact of Summative Assessment and Tests on Student Motivation for Learning.* London: EPPI-Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=Pbyl1CdsDJU%3D&tabid=106&mid=1967 (accessed February 2013).

Harlen, W., and M. James. 1997. "Assessment and Learning: Differences and Relationships between Formative and Summative Assessment." *Assessment in Education: Principles, Policy and Practice* 4 (3): 365–379.

Harlen, W. 2007. "Formative Classroom Assessment in Science and Mathematics." In J. H. McMillan (ed.), *Formative Classroom Assessment: Theory into Practice,* 116–135. New York: Teachers College Press.

Heritage, Margaret. 2010. *Formative Assessment: Making It Happen in the Classroom*. Thousand Oaks, CA: Corwin Press.

Heritage, Margaret. 2013. *Formative Assessment in Practice: A Process of Inquiry and Action*. Cambridge, MA: Harvard Education Press.

Herman, J. L. 2010. *Coherence: Key to Next Generation Assessment Success*. Los Angeles, CA: CRESST. http://www.cse.ucla.edu/products/policy/coherence_v6.pdf

Herman, J. L., P. R. Aschbacher, and L. Winters. 1992. *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Herman, Joan L., and Margaret Heritage. 2007. *Moving from Piecemeal to Effective Formative Assessment Practice: Moving Pictures on the Road to Student Learning.* Paper presented at the Council of Chief State School Officers Assessment Conference, Nashville, TN.

Herman, Joan L., Margaret Heritage, and P. Goldschmidt. 2011. *Developing and Selecting Assessments of Student Growth for Use in Teacher Evaluation Systems.* Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, Joan L., E. Osmundson, and R. Dietel. 2010. *Benchmark Assessment for Improved Learning* (AACC Report). Los Angeles, CA: University of California.

International Reading Association. 2000. *Teaching All Children to Read: The Roles of the Reading Specialist: A Position Statement of the International Reading Association.* http://www.reading.org/Libraries/position-statements-and-resolutions/ps1040_specialist.pdf

Jenkins, J. R. 2003. *Candidate Measures for Screening At-Risk Students.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention symposium, Kansas City, MO. http://www.nrcld.org/symposium2003/jenkins/index.html (accessed May 15, 2008).

Joseph, L. M. 2002. "Best Practices in Planning Intervention for Students with Reading Problems. In A. Thomas, and J. Grimes (eds.), *Best Practices in School Psychology IV*, 803–816. Bethesda, MD: National Association of School Psychologists.

Kluger, A. N., and A. DeNisi. 1996. "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin* 119: 254–284.

Lane, S. 2013. "Performance Assessment." In J. McMillan. (ed). *Sage Handbook of Research on Classroom Assessment*, 313–331. Thousand Oaks, CA: SAGE Publications.

Linn, R. L., and E. L. Baker. 1996. "Can Performance-Based Student Assessment Be Psychometrically Sound?" In J. B. Baron and D. P. Wolf (eds.), *Performance-Based Assessment – Challenges and Possibilities: 95th Yearbook of the National Society for the Study of Education Part 1* 95 (1): 84–103). Chicago, IL: National Society for the Study of Education.

Linn, R. L., and M. D. Miller. 2005. *Measurement and Assessment in Teaching*. Upper Saddle River, NJ: Pearson Education.

Lodico, M. G., E. S. Ghatala, J. R. Levin, M. Pressley, and J. A. Bell. 1983. "The Effects of Strategy-Monitoring Training on Children's Selection of Effective Memory Strategies." *Journal of Experimental Child Psychology* 35 (2): 263–277.

National Association of State Directors of Special Education (NASDSE). 2005. *Response to Intervention: Policy Considerations and Implementation.* Alexandria, VA: Author.

National Research Council. 1999. *Improving Student Learning: A Strategic Plan for Education Research and its Utilization*. Washington, DC: National Academy Press.

National Research Council. 2001. *Knowing What Students Know: The Science of Design and Educational Assessment*. Washington, DC: National Academy Press.

National Research Council. 2012. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academy Press.

National Research Center on Learning Disabilities (NRCLD). 2006. *Progress Monitoring.* http://www.nrcld.org/rti_manual/pages/RTIManualSection2.pdf

Organisation for Economic Co-operation and Development (OECD). 2005. *Formative Assessment: Improving Student Learning in Secondary Classrooms*. Centre for Educational Research and Innovation. http://www.oecd.org/dataoecd/19/31/35661078.pdf (accessed May 9, 2008).

Palm, T. 2008. "Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature." *Practical Assessment, Research, and Evaluation* 13 (4): 1–11.

Perie, M., S. Marion, and B. Gong. 2009. "Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments." *Educational Measurement: Issues and Practice* 28 (3): 5–13.

Perry, N., K. VandeKamp, L. Mercer, and C. Nordby. 2002. "Investigating Teacher-Student Interactions that Foster Self-Regulated Learning." *Educational Psychologist* 37 (1): 5–15.

Pintrich, P. 2000. “The Role of Goal Orientation in Self-Regulated Learning.” In M. Boekaerts, P. Pintrich, and M. Zeidner (eds.), *Handbook of Self-Regulation*, 451–502. San Diego, CA: Academic Press.

Popham, W. J. 1995. *Classroom Assessment: What Teachers Need to Know.* Boston: Allyn and Bacon.

Popham, W.J. 2010. *Transformative Assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Poppers, A. E. 2011. “Identifying Craft Moves: Close Observation of Elementary Students’ Writing.” In P. Noyce and D. T. Hickey (eds.), *New Frontiers in Formative Assessment,* 89–107. Cambridge, MA: Harvard Education Press.

Puckett, M. B., and D. Diffily. 2004. *Teaching Young Children: An Introduction to the Early Childhood Profession* (2nd ed.). Clifton Park, NY: Delmar Learning.

Resnick, Lauren B., and D. P. Resnick. 1992. “Assessing the Thinking Curriculum: New Tools for Educational Reform.” In B. R. Gifford and M. C. O’Connor (eds.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*. Boston, MA: Kluwer.

Ruiz-Primo, Maria A., and Erin M. Furtak. 2004. *Informal Assessment of Students’ Understanding of Scientific Inquiry*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ruiz-Primo, Maria A., and Erin M. Furtak. 2006. “Informal Formative Assessment and Scientific Inquiry: Exploring Teachers’ Practices and Student Learning.” *Educational Assessment* 11 (3 and 4): 237–263.

Ruiz-Primo, Maria A., and Erin M. Furtak. 2007. “Exploring Teachers’ Informal Formative Assessment Practices and Students’ Understanding in the Context of Scientific Inquiry.’ *Journal of Educational Research in Science Teaching* 44 (1): 57–84.

Sadler, D. R. 1989. Formative assessment and the design of instructional strategies. *Instructional Science, 18,* 119-144.

Schunk, D. H., and B. J. Zimmerman. 2008. *Motivation and Self-Regulated Learning: Theory, Research, and Applications*. Mahwah, NJ: Lawrence Erlbaum.

Smarter Balanced Accessibility Consortium (SBAC). 2013a. http://www.smarterbalanced.org/

Smarter Balanced Accessibility Consortium (SBAC). 2013b. *Smarter Balanced Accessibility and Accommodations: Meeting the Needs of All Students.* http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/07/SmarterBalanced_Accessibility_Factsheet.pdf

Shepard, L. A. 2000. "The Role of Assessment in a Learning Culture." *Educational Researcher* 29 (7): 4–14.

Shepard, L. A. 2001. The Role of Classroom Assessment in Teaching and Learning. In V. Richardson (ed.), *Handbook of Research on Teaching* (4th ed.), 1066–1101. Washington, DC: American Educational Research Association.

Shepard, L. A. 2005. "Linking Formative Assessment to Scaffolding." *Educational Leadership* 63 (3): 66–71.

Shepard, L.A. 2005. *Will Commercialization Enable or Destroy Formative Assessment?* (Original title: Formative assessment: Caveat emptor). Paper presented at the ETS Invitational Conference 2005: The Future of Assessment, Shaping Teaching and Learning, New York.

Smith, J. K. 2003. "Reconsidering Reliability in Classroom Assessment and Grading." *Educational Measurement: Issues and Practices, 22*(4), 26-34.

Soltero-Gonzalez, L., Kathy Escamilla, and S. Hopewell. 2012. "Changing Teachers' Perceptions About the Writing Abilities of Emerging Bilingual Students: Towards a Holistic Bilingual Perspective on Writing Assessment." *International Journal of Bilingual Education and Bilingualism* 15 (1): 71–94.

Stiggins, R. J., J. A. Arter, J. Chappuis, and S. Chappuis. 2006. Classroom Assessment *for* Student Learning: Doing It Right – Using It Well. Portland, OR: Educational Testing Service.

Stansfield, C. W., and M. Bowles. 2006. Study 2: "Test Translation and State Assessment Policies for English Language Learners." In C. Rivera and E. Collum (eds.), *State Assessment Policy and Practice for English Language Learners: A National Perspective*, 1–173. Mahwah, NJ: Lawrence Erlbaum.

Stobart, G. 2006. “The Validity of Formative Assessment.” In J. Gardner (ed.), *Assessment and Learning,* 133–146. London, UK: Sage.

Understanding Language. 2013*. Instructional Unit: Persuasion Across Time and Space*. Palo Alto, CA: Stanford University. http://ell.stanford.edu/sites/default/files/ela_archives/understanding_language_materials_Jan2013.pdf

Venn, J. J. 2000. *Assessing students with Special Needs* (2nd ed.). Upper Saddle River, NJ: Merrill.

Wiliam, Dylan, and Paul J. Black. 1996. “Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment?” *British Educational Research Journal, 22*(5), 537-548.

Wiliam, Dylan. 2006. *Does Assessment Hinder Learning?* Paper presented at the ETS Invitational Seminar at the Institute of Civil Engineers, London, UK.

Wiliam, Dylan. 2007. “Keeping Learning on Track: Classroom Assessment and the Regulation of Learning.” In F.K. Lester Jr. (ed.), *Second Handbook of Mathematics Teaching and Learning*,1053–1098. Greenwich, CT: Information Age Publishing.

Wiliam, Dylan. 2011. “What is Assessment for Learning?” *Studies in Educational Evaluation* 37: 3–14.