



ASSESSMENTS TO GUIDE ADOLESCENT LITERACY INSTRUCTION



CENTER ON
INSTRUCTION

ASSESSMENTS TO GUIDE ADOLESCENT LITERACY INSTRUCTION

Joseph K. Torgesen

Debra Houston Miller

Florida Center for Reading Research

Florida State University

2009



CENTER ON
INSTRUCTION

This publication was created for the Center on Instruction by the Florida Center for Reading Research at Florida State University. The Center on Instruction is operated by RMC Research Corporation in partnership with the Florida Center for Reading Research at Florida State University; Instructional Research Group; the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; and the Meadows Center for Preventing Educational Risk at The University of Texas at Austin.

The contents of this document were developed under cooperative agreement S283B050034 with the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Editorial, design, and production services provided by RMC Research Corporation.

Citation: Torgesen, J. K., & Miller, D. H. (2009). *Assessments to guide adolescent literacy instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Extended quotations from *Formative Classroom Assessment* are reprinted by permission of the publisher. From James H. McMillan, Editor, *Formative Classroom Assessment: Theory into Practice*, New York: Teachers College Press © 2007 by Teachers College, Columbia University. All rights reserved.

Excerpt from "Can Diagnostic Reading Assessment Enhance General Educators' Instructional Differentiation and Student Learning?" (p. 329) by L. Fuchs and D. Fuchs in *Preventing and Remediating Reading Difficulties: Bringing Science to Scale*, B. R. Foorman, Ed. (2003). Austin, TX: PRO-ED. Copyright 2003 by PRO-ED, Inc.

To download a copy of this document, visit www.centeroninstruction.org.



CONTENTS

- 1 OVERVIEW**
- 3 EXECUTIVE SUMMARY**
- 5 PART I—USING ASSESSMENT TO IMPROVE LITERACY INSTRUCTION FOR ADOLESCENTS**
- 5 INTRODUCTION**
 - 7 Adolescent academic literacy*
 - 9 Data-based decision making*
- 11 A COMPREHENSIVE ASSESSMENT SYSTEM FOR ADOLESCENT LITERACY**
 - 11 Goals for adolescent literacy instruction*
 - 11 The role of reading assessments in meeting these goals*
 - 13 Literacy assessments in a comprehensive assessment plan*
- 18 FORMATIVE ASSESSMENTS TO GUIDE INSTRUCTION IN ADOLESCENT LITERACY**
 - 19 Current definitions of formative assessment*
 - 21 Examples of classroom-based formative assessment*
 - 24 Further examples from high-achieving secondary schools*
 - 28 Evidence for the impact of classroom-based formative assessments on student achievement*
 - 29 Two caveats*
 - 31 Six characteristics of high-quality classroom-based formative assessment*
 - 38 Curriculum-based measurement (CBM): A special case of classroom-based formative assessment*
 - 42 Summary and conclusions about classroom-based formative assessment*
 - 43 The role of standards-based benchmark assessments in improving student literacy outcomes*
 - 48 The role of screening and diagnostic assessments in a comprehensive assessment plan*
 - 53 Summary and review: Assessment types required in a comprehensive assessment plan*

57	CHALLENGES TO THE SUCCESSFUL USE OF ASSESSMENT FOR INSTRUCTION WITH ADOLESCENTS
57	<i>Challenge #1: Fundamental changes in attitudes and instructional practices</i>
61	<i>Challenge #2: Tension between types of formative assessments useful to administrators and types useful to teachers and students</i>
63	<i>Challenge #3: Lack of teacher “buy in” and divergence of educators’ attitudes and beliefs about indicators of student success</i>
65	CONCLUDING COMMENTS
67	RECOMMENDATIONS FOR FURTHER READING
68	REFERENCES
75	PART II—TEN EXAMPLES OF ASSESSMENTS, OR ASSESSMENT SYSTEMS, IN CURRENT USE OR UNDER DEVELOPMENT TO GUIDE INSTRUCTION IN ADOLESCENT LITERACY



OVERVIEW

This document was prepared to assist literacy specialists and other technical assistance providers in their work with states to improve educational policy and practice in adolescent literacy. It is a companion volume to two other Center on Instruction publications that offer guidance in improving literacy outcomes in adolescents—*Academic Literacy Instruction for Adolescents: A Guidance Document from the Center on Instruction* (Torgesen, Houston, Rissman, et al., 2007) and *Improving Literacy Instruction in Middle and High Schools: A Guide for Principals* (Torgesen, Houston, Rissman, 2007). Both may be downloaded at <http://www.centeroninstruction.org>.

As in the two other documents, adolescent literacy is defined here as covering the range of students from grades four through twelve. The present document has two major parts:

Part I—Using assessment to improve instruction in literacy for adolescents

Part I describes the key elements of a comprehensive assessment plan to improve literacy instruction for adolescents. It details the purposes of assessment and identifies the major types of assessments that can be used to meet each purpose. The focus is on formative assessments, whose purpose is to guide instruction for individual students or groups of students. The information in this part of the document is based on research about the nature of adolescent literacy as well as research on the utility of various types of assessments for learning.

Part II – Ten examples of assessments, or assessment systems, in current use or under development to guide instruction in adolescent literacy

Part II contains short summaries of approaches to assessment for instruction in adolescent literacy currently in use or under development in the United States. These examples come from individual schools, from state-level systems, and from both commercial companies and nonprofit organizations. They are meant

to convey the wide variety of approaches in addressing some of the challenges in adolescent literacy assessment identified in Part I. They are not meant as endorsements of either the procedures themselves or of the companies or organizations using them. In fact, the companies, organizations, or schools associated with the examples are purposely not identified by name. Although the examples are real, they are identified here only by number. They were selected through an Internet search and by polling experts in adolescent literacy around the country. We also made inquiries to a number of organizations not included here but were unable to obtain sufficient information to summarize their procedures. The list is obviously not exhaustive, and there may be other, even better, examples that we were not able to locate.



EXECUTIVE SUMMARY

Assessments to Guide Adolescent Literacy Instruction addresses the growing national concern about adolescent literacy achievement. In a focused examination of literacy assessments and their role in instructional decision-making and growth in student learning, it provides a research-based overview of the types of literacy assessments that can help direct teaching and learning at all levels, with special attention to assessments in the classroom setting. Part I examines formative assessments, or assessments *for* learning (as distinct from assessments *of* learning), placing clear emphasis on formative assessments and their significant direct connection to classroom instruction. Academic literacy is explicitly defined as the kinds of reading skills students need to be successful both in school and on state-level accountability measures.

Torgesen and Miller compare five types of assessments, detailing the specific characteristics of each in terms of methods of implementation, the frequency with which they are employed, and their purpose. A discussion of the importance of classroom-based formative assessments in the larger plan to improve adolescent literacy offers examples of formative assessment practices and outlines the research supporting their effective use. The authors also draw conclusions about the four other types of assessments that schools can include in a comprehensive assessment plan—standards-based benchmark assessments, curriculum-based measures (CBM), screening assessments, and diagnostic assessments.

Their review of standards-based benchmark tests looks at both the increased frequency with which they are used and at concerns about the lack of adequate research to support any positive effect they may have on students' literacy growth. The authors examine curriculum-based measures as a specific kind of formative assessment that can guide instruction and note the differences between and uses of *mastery oriented* CBM versus *general outcome* CBM when monitoring adolescents' grade-level literacy skills. Their discussion of screening distinguishes between *general screening* to identify students who require intervention instruction and *targeted screening* to determine specific types of interventions that at-risk students might require. The authors' discussion of the diagnostic information gained through both *formal diagnostic tests* and *informal diagnostic assessments* proposes minimal,

tactical use of formal measures and highlights the close relationship between informal diagnostic assessments and classroom-based formative assessments.

The authors identify three challenges educators must address in order for formative assessments to have their full effect on improving student outcomes: 1) the need for fundamental change in attitudes and instructional practices, 2) the need to resolve tensions between teachers and administrators about the types of formative assessments each prefer, and 3) the need to change educators' attitudes and beliefs about indicators of student success. All will require new thinking about the purposes of assessment, student abilities, and teaching methods; attention to new instructional resources; and the development of new assessment and pedagogical skills.

Part I concludes by emphasizing that an effective comprehensive assessment plan is crucial for guiding instruction that will improve overall levels of academic achievement and meet the learning needs of all students.

Part II profiles ten examples of assessments and assessment systems currently in use or under development in the U.S. While evidence is not currently available for the effectiveness of these systems in improving literacy outcomes for adolescents, the systems do provide insight into the range of possible approaches to using assessments for improving literacy instruction for students in grades 4 through 12. Each profile contains a succinct description and a commentary that links the types of assessments to information presented in Part I.



PART 1—USING ASSESSMENTS TO IMPROVE LITERACY INSTRUCTION FOR ADOLESCENTS

INTRODUCTION

This document presents research-based information on assessment practices to support more effective instruction in academic literacy for students in grades 4 through 12. It assumes more than a beginning level of knowledge of reading and assessment practices, and is based on research in two areas: 1) the nature of reading and reading comprehension in adolescents and 2) the use of assessments and assessment systems to increase student learning and performance. Although research on literacy assessments for adolescents is actually less well developed than research on instruction, we believe that it provides sufficient basis for several strong recommendations to improve practice in adolescent literacy.

In any discussion of educational assessments, it is important to distinguish between *assessment of learning*, and *assessment for learning*. This distinction centers principally on how the assessment information is used. Assessments *of learning*, frequently referred to as summative assessments, indicate how well students have learned, or how well they can meet performance standards in a subject area such as reading or math. The best example of these types of assessments currently in wide use in the United States are the end-of-year, state-level accountability assessments required by the No Child Left Behind legislation. Assessments *for learning*, in contrast, are designed to help teachers provide more effective instruction so that individual students can improve their learning, or so that more students can reach acceptable performance standards.

The focus of this document is on assessments *for learning* 1) because these assessments are intimately connected to instruction and 2) because they are currently much less well developed and less widely used than are assessments *of learning*. A final reason for our focus on *assessments for learning*, or formative assessments, is that 3) improvement in their use has significant potential to increase the effectiveness of teaching and learning in adolescent literacy (Black & Wiliam, 1998). While classroom-based formative assessments are discussed extensively, this document also addresses the

formative uses of three other types of assessments: benchmark (interim), screening, and diagnostic assessments.

Improving literacy outcomes for adolescents will entail a comprehensive effort that effects changes in state- and district-level policies, shifts instructional practices in all content areas, organizes schools more efficiently, provides more powerful interventions for struggling readers, develops more involved and effective leadership, and offers extensive professional development to all leaders and teachers. Many of these steps have been outlined in recent documents available on the Internet. Four of the most widely disseminated are:

- *From State Policy to Classroom Practice: Improving Literacy Instruction for All Students*. National Association of State Boards of Education (2007), available at www.nasbe.org/index.php/file-repository?func=fileinfo&id=132
- *Reading at Risk: How States Can Respond to the Crisis in Adolescent Literacy*. National Association of State Boards of Education (2006), available at www.centeroninstruction.org/files/Reading_At_Risk_Full_Report.pdf
- *Reading Next: A Vision for Action and Research in Middle and High School Literacy*. Alliance for Excellent Education (2006), available at www.all4ed.org/files/ReadingNext.pdf
- *Reading to Achieve: A Governor's Guide to Adolescent Literacy*. National Governors Association Center for Best Practices (2005), available at www.nga.org/Files/pdf/0510GOVGUIDELITERACY.PDF

All four emphasize improvements in the use of various forms of assessment for learning, or formative assessment, as part of an effective overall strategy for improving adolescent literacy. In light of the consistency of these recommendations, recent comments by Dr. Richard Stiggins, an expert in classroom-based formative assessments, are sobering. In a chapter titled *Conquering the Formative Assessment Frontier*, Stiggins (2007) acknowledges recent accomplishments in developing high-quality summative assessments, but adds, "...behind these considerable accomplishments there is almost complete neglect of assessment where it exerts the greatest influence on pupils' academic lives: day to day in the classroom, where it can be used to help them learn more" (p. 10). He also suggests, "The principal assessment challenge that we face in schools today is to ensure that sound assessment practices permeate every classroom—that assessments are used to benefit pupils....This challenge has remained unmet for decades, and the time has



come to conquer this final assessment frontier: the effective use of formative assessment to support learning” (p. 10).

Adolescent academic literacy

Although both practitioners and researchers use the term *adolescent literacy* to include a very broad range of competencies extending beyond the reading and writing proficiencies required for school success (e.g., Alvermann, 2001; Moje, 2006), this document focuses on assessments to improve instruction in adolescent academic literacy. Most definitions of literacy include writing as well as reading, but our focus is on the assessment of reading skills. Although writing and reading skills are closely allied in both instruction and assessment (as when students make extended written responses in answering questions or discussing text), we discuss writing only as a means of assessing reading skills, not as an object of assessment itself.

The description of adolescent literacy used in an earlier document (Torgesen, Houston, Rissman, et al., 2007) remains relevant to the current document:

Academic literacy is usually defined as the kind of reading proficiency required to construct the meaning of content-area texts and literature encountered in school. It also encompasses the kind of reading proficiencies typically assessed on state-level accountability measures, such as the ability to make inferences from text, to learn new vocabulary from context, to link ideas across texts, and to identify and summarize the most important ideas or content within a text. Notice that the definition of academic literacy includes not only the ability to read text for initial understanding but also the ability to think about its meaning in order to answer questions that may require the student to make inferences or draw conclusions. (p. 3)

This definition is consistent with the definition of reading in the new framework for the National Assessment of Educational Progress (NAEP) proposed by the National Assessment Governing Board (NAGB, 2007): “Reading is an active and complex process that involves understanding written text; developing and interpreting meaning; and using meaning as appropriate to type of text, purpose, and situation” (p. 2).

This definition applies to the assessment of reading achievement on the NAEP and is not intended as an *inclusive* definition of reading. It indicates that the NAEP assesses reading comprehension, and not more basic reading skills (i.e., decoding, reading accuracy, and fluency) that students are supposed to acquire by the end of third grade. Assessments such as the NAEP assume that students have acquired the necessary word-level reading skills to identify the

Academic literacy is embedded in content domains but assessed as a general skill

Reading comprehension in adolescents is strongly influenced by the content of the texts being read (Anderson & Pearson, 1984; Hirsch, 2006). That is, students who have extensive knowledge of the topic they are reading about typically comprehend the material more efficiently than students who have less knowledge in that domain. This is one reason our earlier document on academic literacy instruction (Torgesen, Houston, Rissman, et al., 2007) recommended more powerful teaching of essential content for improving overall levels of adolescents' academic literacy.

However, the NAEP literacy standards, like those of most states, apply across content areas, and the items that assess performance on those literacy standards are frequently written so that correct responses do not depend heavily on prior knowledge except for essential vocabulary. In the current document, we discuss literacy assessment as it relates to relatively content-free literacy standards (e.g., being able to identify the main idea and supporting details, being able to draw accurate conclusions) while at the same time recognizing that a student's ability to meet these standards may vary across content areas. The ability to meet standards within specific content domains is influenced by prior knowledge in each domain and prior experience reading texts in each domain.

In our view, academic literacy in adolescents is strongly embedded in content, and it is the essential responsibility of all content-area teachers to provide instruction that enhances their students' ability to comprehend texts within their content area (Torgesen et al., 2007). As content-area teachers more fully assume these responsibilities, their instruction should enable students to comprehend texts more deeply and think about the content. Students' ability to comprehend and think about texts within the standards-based, year-end outcome tests in literacy should also improve. Although no one may be specifically assigned to help most students meet literacy standards in middle and high school, it is actually the responsibility of all teachers to work toward improving adolescents' literacy skills, not only English/language arts teachers or reading teachers.



words in text accurately and fluently, and they focus on students' ability to understand the meaning and use the information contained in a text. This focus is appropriate for reading assessment in grades four and higher, but it renders them insensitive to the instructional needs of many struggling readers who continue to have difficulties with word-level skills in middle and high school.

In its focus on reading comprehension, our definition of academic literacy for adolescents is very similar to that which guided the development of the NAEP. However, because we are concerned with assessment to guide instruction, we discuss assessments for a considerably broader range of reading skills than those that are the focus of the NAEP and of most state-level accountability measures. For example, we discuss the utility of assessments of both word-level skills and reading comprehension, because some adolescents still require focused instruction on word-level skills in order to achieve the goal of grade-level reading comprehension. We acknowledge that the ultimate outcome of reading instruction should be attainment of well-defined comprehension standards as outlined in the NAEP framework and in various state literacy standards for adolescents, but we also recognize the need for appropriate assessment of more basic reading skills as an aid to instruction for students who continue to struggle at this level.

Data-based decision making

It may be helpful to place our current focus on formative literacy assessment within the larger context of data-based decision making. The systematic use of empirical data in making management and instructional decisions in districts, schools, and classrooms encompasses several types of data. For example, *process data* obtained from classroom observations or interviews with school leaders can provide information about quality of instruction or intervention support systems. *Input data*, such as expenditures per pupil, number of inservice hours provided for instructional and leadership personnel, or characteristics of students when they enter the school, can be useful for understanding school performance or identifying possible areas for improvement. *Satisfaction data* from student or teacher surveys may also provide information helpful in making good classroom or school-level decisions.

Within this broad context, *student outcome*, or *performance data* are among several sources of information teachers and schools can use to guide their

decisions. However, a recent report by the Rand Corporation (Marsh, Pane, & Hamilton, 2006) found that student outcome data are by far the most widely available at present, and thus are most widely used in making decisions. This document examines the types of reading assessments schools currently use to generate student outcome data, but will focus on those formative assessments that have the greatest potential for improving instruction in the classroom.

Another important consideration in data-based decision-making is converting *raw data* into *usable information*. Raw assessment data must be organized and combined with other data to produce usable information, and for most assessments, this requires considerable training and experience. Simply having assessments in place will have little impact on teaching or student learning if no one has the time or expertise available to turn the raw assessment data into usable information.

Finally, another set of skills and knowledge is required to turn useful information from assessments into *effective instructional adjustments*. For example, knowing that a student struggles to identify the main idea in grade-level expository text does not give the teacher direct guidance in how to help the student become more skillful at this task. The instructional expertise required to respond appropriately to detailed assessment data must also be present if assessment information is to produce improvements in student learning. In other documents, we have provided guidance on instructional practices to improve adolescent literacy outcomes (Torgesen, Houston, Rissman, et al., 2007) and initial guidance for principals on school-level procedures required to turn raw assessment data into usable information (Torgesen, Houston, & Rissman, 2007).



A COMPREHENSIVE ASSESSMENT SYSTEM FOR ADOLESCENT LITERACY

Goals for adolescent literacy instruction

In thinking about a comprehensive assessment plan for students in late elementary, middle, and high school, we first identify the three most important goals of academic literacy instruction at these grade levels. They are:

1. *To increase overall levels of reading proficiency in order to better prepare students for increasing demands for high-level literacy skills both in post-secondary education and the workplace.* There is evidence that current average reading levels of students graduating from high school are not sufficient to meet post-secondary literacy demands, both in many work place settings and at community colleges and universities (Biancarosa & Snow, 2006; Williamson, 2004). The practical translation of reaching this goal in the current accountability environment is to have ever higher numbers of students achieve the highest levels of proficiency on state end-of-year reading examinations.
2. *To help students who have achieved grade-level reading standards by the end of third grade continue to meet increasingly difficult standards in middle and high school.* Students must acquire many additional reading skills after third grade in order to be proficient readers in high school. Many students are able to meet grade-level standards in late elementary or middle school, but cannot meet them in high school (Leach, Scarborough, & Rescorla, 2003).
3. *To assist students who are reading below grade-level standards in acquiring the skills and knowledge necessary to meet those standards.* This will require instruction sufficiently powerful to *accelerate* reading development so that students make more than one year's progress during one year of school.

The role of reading assessments in meeting these goals

A well-developed, comprehensive literacy assessment plan is essential if schools are to meet these three goals for improving outcomes in academic

literacy. The need for accurate assessments arises because of the enormous diversity in the rate of learning and level of literacy skills among adolescents. Because of this enormous diversity, a key to meeting all three goals is accurate assessment that allows teachers and schools to differentiate instruction according to individual students' needs.

For example, strong students will not attain even higher levels of literacy skills without appropriate challenges. Similarly, struggling readers are unlikely to profit from instruction that assumes they have skills and knowledge they have not yet acquired. In order to target instruction effectively, to re-teach when necessary, to offer needed support, and to provide appropriately challenging assignments, teachers must have an ongoing and accurate understanding of their students' literacy capabilities. District and school leaders also need data on student performance in order to make decisions about allocating school and district resources and to provide additional support where necessary. To help meet the goals for improving adolescent literacy outcomes, assessments of reading growth in grades 4 through 12 must serve three purposes:

1. *To determine the extent to which instructional programs are successful in helping all students meet or exceed grade-level standards by the end of the year.* At the state, district, and school levels, educators need to know at the end of each year how many students at each grade level can meet the state-level literacy standards. They also need to know whether the number of students who can achieve at the highest levels is increasing from year to year, and whether the number of students at the lowest levels is declining. These data should be comparable across a state's schools and districts so that schools can be accountable to their communities and districts, and states can allocate resources and support where needed.
2. *To monitor the progress of all students during the academic year so that appropriate adjustments can be made to ongoing instruction.* Districts, schools, grade-level teams, and individual teachers need to make periodic assessments during the year to tell them which students are making adequate progress toward meeting grade-level standards by the end of the year. This allows teachers to make adjustments and allocate resources while there is still time to help students in need. Classroom teachers also need very frequent assessments to help them and their students understand the necessary "next steps" in learning to improve their literacy



skills. This information needs to be much more detailed than that required by district and school level leaders. It must be sufficiently detailed to help teachers make decisions about the need to re-teach a concept, to explain something in a different way, or to provide additional practice to consolidate a skill. As Stiggins (2007) has explained, “Both pupil and teacher must know where the learner is now, how that compares to ultimate learning success, and how to close the gap between the two” (p.15).

3. *To provide information helpful in identifying students who need intervention to advance reading performance toward grade-level standards.* At the beginning of the year, schools must have data that will help them identify students who may need special instructional interventions because their low literacy skills are likely to interfere seriously with their ability to learn in content-area classes. Additional screening information may also be required in order to identify, from among the larger group of “at risk” students, those who need intensive, comprehensive interventions, versus those who may only require focused instruction in comprehension skills and vocabulary.

Literacy assessments in a comprehensive assessment plan

A key starting place for selecting assessments is a clear understanding of their purpose:

- What decisions will be made from the assessment data?
- Who will make these decisions?
- What information is most essential to making good decisions?

Proposed below is an *ideal* comprehensive system, not necessarily one that can be quickly, practically, or easily implemented. Nevertheless, it should be helpful to have a clear picture of what an *ideal* system might look like, both as a blueprint for long-term plans to develop such a system and as a foundation from which decisions to depart from an ideal system can be made thoughtfully.

An ideal system would contain:

1. *Assessments to measure end-of-year-performance.* The goal here is to provide a reliable and valid assessment of the extent to which all students can meet, or exceed, state-level literacy standards. This goal is satisfied by

the state-level accountability assessments that are currently in place because of NCLB requirements. These summative tests of student learning and school effectiveness typically have minimal implications for improving learning at an individual student level.

2. *Assessments to monitor the progress of all students during instruction.* Here the goal is to provide information for making ongoing adjustments to instruction for individual students or groups of students. A comprehensive system might involve two broad types of progress monitoring assessments: 1) benchmark (interim) assessments and 2) classroom-based formative assessment. Interim assessments may be administered district- or school-wide to all students several times a year to gather objective information about student skills in relation to grade-level standards. They are often used to predict the likelihood that students will be able to “pass” the state-level accountability test at the end of the year. Their popularity has been growing rapidly over the last decade (Marsh et al., 2006).

Ongoing, classroom-based formative assessment is the most relevant to guiding and improving classroom-level instruction. Improved use of this type of assessment may be the most cost-effective way to improve student learning performance in middle and high school (Black & William, 1998; Fuchs & Fuchs, 1986). Classroom-based formative assessment is tailored to each student’s reading proficiency: It is quite different for students in a content-area classroom working on grade-level reading comprehension standards than it is for struggling readers in intervention classrooms working on a broader range of reading skills.

Although both district- or school-level benchmark tests and classroom-based formative assessments provide information that can aid instructional decisions, Stiggins (2007) notes that “teachers and pupils will use continuous day-to-day formative classroom assessments in ways that differ fundamentally from how policy makers and instructional support personnel use their periodic assessments” (p. 15).

3. *Assessments to identify students who need intervention.* At the beginning of each school year, the school should have some data on students’ literacy skills in order to plan literacy instruction and support for the year. This information could come from the previous year’s summative test,



or from a benchmark predictive test administered at the beginning of the year. It would be used to identify initially those students who should be able to meet grade-level reading standards for the current year through the literacy support provided in content-area classes as well as those who are unlikely to meet grade-level standards without extra support or intervention.

Although adolescent struggling readers are very diverse (Buly & Valencia, 2002; Hock, Brasseur, Deshler, Catts, Marques, et al., in press), for assessment and intervention planning purposes, it may be useful initially to think of them as falling into two groups. One group has primary instructional needs in reading comprehension and cannot meet grade-level standards on the end-of-year accountability measure, primarily because of weak vocabulary and comprehension skills. These students can read grade-level text with reasonable fluency and accuracy (not too far below average) but may not be skilled in using reading comprehension strategies or in performing certain kinds of inferential reasoning processes, or they may lack specific vocabulary required for understanding grade-level text.

The second, usually smaller, group contains students with severe and pervasive reading difficulties. Their challenges extend to basic problems with reading accuracy (usually caused by weak phonics/word analysis skills); they are almost always dysfluent readers and frequently have all of the other more complex reading problems involving weaknesses in content knowledge, thinking/reasoning skills, reading strategies, and vocabulary. Of course, there will be substantial variability in the instructional needs of students in each of these groups. However, for purposes of initial assignment to either very intensive interventions encompassing both word-level and comprehension instruction or less intensive interventions focusing primarily on vocabulary and comprehension skills, this two-group division is a good place to start.

A comprehensive assessment plan must contain screening assessments to differentiate between these two groups of students so that interventions can be targeted appropriately. Follow-up screening instruments (given only to students identified as “at risk” on the standards-based accountability measure from the previous year or the first predictive benchmark test) might be brief but reliable tests can quickly determine whether students are sufficiently impaired in word-level reading

skills to require special instruction in this area. A measure of low-level comprehension and reading fluency might also be helpful in identifying appropriate intervention placements for students reading substantially below grade level. Once students are initially placed in an intervention context, informal diagnostic assessments may help further differentiate their instructional needs.

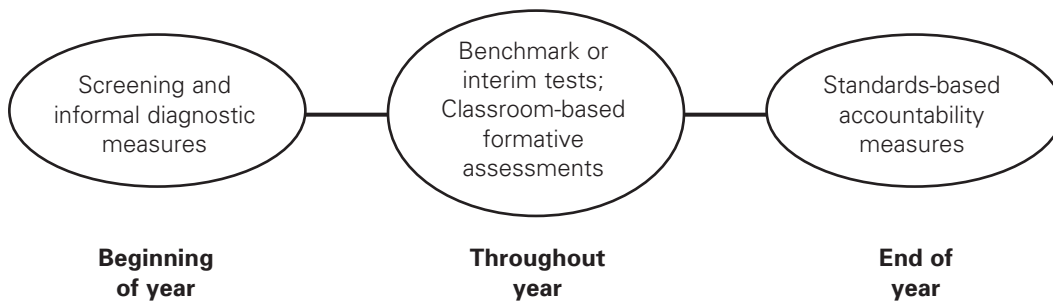
Another way to think about the types of assessments that comprise a comprehensive assessment plan is to categorize them in terms of when they are most likely to be given:

1. *Beginning of the school year.* These may be screening tests to help decide the level of intervention that individual students require or informal diagnostic assessments to help plan and focus the various interventions.
2. *Throughout the year.* These assessments can help teachers and school leaders make necessary adjustments to instruction.
3. *End of the year.* These year-end summative tests are typically given to determine whether students can meet grade-level standards in literacy.

Figure 1 on page 17 represents the nature and timing of these assessments. The middle shape is larger than the other two, not only to indicate the strong potential for these types of assessments to improve instruction in adolescent literacy but also to suggest the size of the professional development investment necessary to help teachers use such assessments more effectively. Note that although benchmark assessments are listed as part of ongoing assessment, the first administration of a benchmark test at the beginning of the year can serve as a general screen to identify students in need of special reading interventions. Subsequent administrations of alternate forms of the benchmark test could be used to monitor growth in grade-level reading skills across the academic year.



Figure 1: Optimal Timing of Assessments, by Type



Many schools and districts are using commercially available intervention programs that include screening (placement) and progress monitoring tests as part of their programs. It would probably be most efficient for them to use the tests associated with their programs, since the program developers have produced guidance for using the assessments to place students in the program and to monitor progress through the lesson elements. However, even when program-specific assessments are available, the district or school should also require general progress monitoring assessments at benchmark periods in order for schools and districts to have uniform information for evaluating student progress toward grade-level standards in specific areas.

As mentioned earlier and discussed elsewhere (McEwan, 2007; Phillips, 2005; Torgesen et al., 2007), having the “right” assessments in place is only one element of an effective literacy assessment plan. In order for assessment data to aid decision making, the data must be easily accessible, in a form that will allow instructional support personnel and teachers to sort, aggregate, or scan it, and in a way that produces timely information. Schools should also create systems that allocate sufficient time for data analysis and collaborative decision making, and leaders must continually support and encourage data-based decision making in the school. Next we consider the role of each type of assessment in a comprehensive assessment plan to improve adolescent literacy outcomes.

FORMATIVE ASSESSMENTS TO GUIDE INSTRUCTION IN ADOLESCENT LITERACY

Assessments given to obtain information useful for guiding instruction are typically referred to as formative assessments, in that they help to form instruction. Many definitions of formative assessment are in circulation, and many different kinds of tests are extolled and marketed as providing information that can enhance instructional decision making. At least part of the reason for current confusion about definitions and examples of formative assessment is that different kinds of tests support different kinds of instructional decision making. Some types of formative assessment provide a lot of detail about what should be taught (i.e., they identify the actual misconceptions, or types of errors students make with regard to certain standards, or they identify inappropriate reading strategies being used). They are sometimes referred to as “classroom-based” or “short-cycle” because they are administered frequently and are embedded within ongoing instruction. Another type of formative assessment, sometimes referred to as a benchmark (interim) assessment, is administered less frequently than classroom-based assessments and provides less detail about the specifics of what needs to be taught. Rather, these assessments typically describe the student’s current level of performance in relation to grade-level standards. Such tests might be used to identify students unlikely to be successful on the state accountability assessment or to provide information about specific standards or skills on which performance is currently inadequate. These assessments are less useful for guiding classroom instruction on a day-to-day basis, but they may provide grade-level teams, school leadership, or district support personnel with information that is helpful in allocating resources and providing additional support where needed.

In current practice, three types of formative assessments are used with adolescents: 1) ongoing classroom-based and instructionally embedded assessments; 2) benchmark assessments, sometimes referred to as interim assessments; and 3) formal and informal screening and diagnostic assessments.



Current definitions of formative assessment

At the most general level, formative assessment can be defined by its purpose, which is to provide information useful for guiding instruction. For example, one group of experts offered this general definition:

An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed. (Perie, Marion, & Gong, 2007, p. 4)

Similarly, Black & William (2007), two prominent researchers in this area, state:

“An assessment is formative to the extent that information from the assessment is fed back within the system and actually used to improve the performance of the system in some way.” (p. 31)

Within these definitions of formative assessment, any of the three types of assessments considered in this section (ongoing classroom-based, benchmark or interim, or screening/diagnostic) could potentially be used for formative purposes. However, the literature also contains definitions of formative assessment that are more restrictive, and that point more directly to ongoing, classroom-based assessments as fulfilling the ideal of formative assessment most completely. For example, according to Perie, Marion, Gong, and Wurtzel (2007):

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes. Thus, it is done by the teacher in the classroom for the explicit purpose of diagnosing where students are in their learning, where gaps in knowledge and understanding exist, and how to help teachers and students improve student learning. The assessments are small-scale (a few seconds, a few minutes, certainly less than a class period) and short-cycle (they are often called “minute-by-minute” assessment or formative instruction). Furthermore, the tasks presented may vary from one student to

another depending on the teacher's judgment about the need for specific information about a student at a given point in time. Providing corrective feedback, modifying instruction to improve the student's understanding, or indicating areas of further instruction are essential aspects of a classroom formative assessment. There is little interest or sense in trying to aggregate formative assessment information beyond the specific classroom. (p. 1)

In their seminal review of the impact of formative assessments on student outcomes, Black and Wiliam (1998) offered another restrictive definition of formative assessment as "encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 7–8). Abrams (2007) adds that, to be most useful for guiding instruction, formative assessments should uncover the "thought process, logic, and understandings students use to arrive at selected answers or narrative responses" (p. 91) rather than simply measuring what students know and can do.

These latter definitions would most likely exclude from the formative assessment family the kind of medium scale (with large scale being state-level accountability measures), district or school level, benchmark tests that are frequently given monthly or quarterly (cf. Perie, Marion, Gong, & Wurtzel, 2007; Stiggins, 2007). They might also exclude screening tests or standardized diagnostic tests which simply indicate whether a student is weak or strong on certain reading or language skills compared with age or grade-level peers. McMillan (2007) defines formative assessment in a way that suggests that differences among formative assessments are not necessarily dichotomous, but rather may exist along a continuum. He suggests (p. 4) that one can tell whether "true" formative assessment is happening by looking for three key things:

1. The extent to which assessments are embedded within instruction;
2. The extent to which additional instructional strategies are employed; and
3. The extent of student engagement and learning.

All three types of assessments may have a place in a comprehensive assessment plan, but classroom-based assessments are most closely linked to ongoing instruction in the classroom. Further, according to such experts as



Black and Wiliam (2007) and Stiggins (2007), they are most in need of further development and broader implementation because of their potential to inform instruction and improve student outcomes significantly.

Examples of classroom-based formative assessment

In this example, Stiggins (2007) captures most of the elements identified with strong formative assessment practices in the classroom:

The assignment was to read three novels by the same author, develop a thesis statement, and defend it in a term paper referring to the literature. To set pupils up for success, the teacher began by providing the pupils with a sample of an outstanding paper to read and analyze to determine what features made it outstanding. They brainstormed and discussed what made it good in class. Then the teacher gave them a sample paper that was of very poor quality. Again they analyzed and evaluated its features in some detail. Comparing the two papers, they listed essential differences, ultimately collaborating in the development of a set of keys to quality, which they transformed into a set of rating scales that depicted the continuum of quality along each dimension, all in pupil-friendly language and accompanied by examples of pupil work to illustrate each key. Then, with this specific guidance in place, they drafted their own papers. They exchanged drafts, analyzing and evaluating each other's work and providing descriptive feedback on how to improve each. If pupils wanted feedback from their teacher on any particular dimension of quality, they could request it and would receive it. The paper was done when the pupil decided it was done. In the end, not every paper was outstanding, but most were of very high quality, and each pupil was confident of that fact before submitting his or her work for final evaluation and grading. (p. 9)

One of the most obvious points to emerge from this example is that strong formative assessment practices are intimately linked to, and associated with, strong instructional practices. Marshall (2007), for example, identified four characteristics of effective classroom-based formative assessment observable in the example above. Effective formative assessment involves:

-
1. Establishing specific and well-explained learning goals;
 2. Setting up instructional contexts that provide lots of opportunities for the teacher to assess how well students are doing;
 3. Providing responsive correction and support to build understanding; and
 4. Engaging students in the establishment of criteria for evaluation—ensuring that they understand them and think they are reasonable.

Commenting on his own example, Stiggins (2007) further explains some of the qualities of effective classroom-based formative assessment that his example illustrates:

The most unique feature of the assessment for learning process is that it acknowledges the critical importance of the instructional decisions made by pupils and their teachers while working as a team—it provides the information they need when they need it. In this context, pupils become consumers of assessment information, too, using evidence of their own progress to understand what comes next for them. (p.17)

Below is another example of a classroom-based formative assessment from the experience of Miller, co-author of this document. The instruction involved a lesson on magnets designed to teach a group of 8th grade students with learning disabilities to locate information in a text in support of an argument, which was one of the standards on the state assessment test in reading:

After discussing the lesson objective and conducting an introductory activity to allow the students to make observations about how magnets react to other magnets and various objects, I led a discussion based on the student observations to introduce key characteristics of magnets. I then asked the students to work in pairs to search the text for “evidence” to support their conclusions about each characteristic. As I circulated among the students, I observed that many were reading sentence by sentence from the beginning of the section rather than using the text structure (sub-headings, bolded words, signal phrases, etc.) to search for their evidence. I interrupted the student searching and asked students to describe how they were searching the text. Building on student comments that most closely



aligned with effective use of text structure, I modeled a search for the evidence from the text using a “think aloud” approach. I explained how each of the key text features helped me make a reasonable guess about whether the information I needed might be located in that part of the text. Next, I guided the students as they used text features to locate the section most likely to have information about one of the characteristics. At the conclusion of the lesson, I mentioned that the students will be asked to gather information from a variety of sources to complete an end of unit writing assignment. I decided that I would model the use of text features in several future lessons and give students opportunities to practice this strategy throughout the remainder of the unit. I also made a note to remind myself to monitor students to see if they could use the strategy without a prompt from me.

In this example, the teacher’s assessment of student strategies for searching text led her to teach an impromptu “mini-lesson” as well as to develop a longer term instructional agenda responsive to these students’ specific needs. One aspect of the teacher’s plan, after identifying an important instructional target based on her initial assessment of student strategy use, was to monitor her students’ growing ability to use the strategy independently. Presumably, if she noticed certain students were lagging in independent use of more efficient strategies, she would plan differentiated instructional and learning experiences for them that would lead to mastery in using effective text-searching strategies.

The example also illustrates a confusion that can sometimes arise in defining formative assessment. An essential idea in most definitions is that formative assessment provides information to guide instruction, or points to the “next steps” for instruction. In one sense, good formative assessments do provide guidance about how to reach instructional goals because they provide information about what needs to be taught. In this case, the teacher observed that the students needed to be taught a more efficient strategy for finding supporting information in text.

However, there is still a lot of information that the assessment cannot provide. For example, the teacher needed to have a very good understanding of the task herself in order to notice that the students were using an inefficient strategy. In other words, she had to have a good theory of the task (Black & Wiliam, 1998) to understand that the students were not approaching the task

efficiently, and that they had more to learn. Moreover, she had to understand how best to teach the new strategy to the students.

Formative assessments provide information about what must be taught, but they cannot provide information about *how* to teach it. They can monitor whether an instructional approach is accomplishing its aim, but they provide no information about what instructional or corrective approach to take initially. This underlines a point we will return to later, that *formative assessments are only effective if they are followed by effective instructional responses or appropriate types of feedback* (Fuchs & Fuchs, 2002; Shute, 2007). Both the capacity to create effective classroom-based formative assessments and the capacity to respond appropriately to the information they provide depend on the teacher's deep understanding of the skill or concept being taught.

Other examples of formative assessment in the classroom are found in a professional development study (William, Lee, Harrison, & Black, 2004) in which a group of high school teachers were given extensive training (7.5 days spread over 18 months) in the principles and goals of classroom-based formative assessment. They were encouraged to implement this training in their classrooms in whatever manner they saw fit. The five most common formative assessment strategies the teachers implemented in their classrooms were:

1. Rich questioning and discussion that provided ample opportunities for discovering student thinking and knowledge;
2. Comment-only marking which focused on providing descriptive, rather than evaluative (as in grading), feedback about the strengths and weaknesses of student work;
3. Sharing scoring and grading criteria with students so that they had a very clear idea of what successful performance on a given task or activity looked like;
4. Providing many opportunities for peer- and self-assessment; and
5. Group review of outcomes from tests.

Further examples from high-achieving secondary schools

Several years ago, Langer (2001) conducted a study of instruction in middle and high schools that were achieving higher than expected outcomes in reading and writing. Although the study focused on instruction, descriptions of that



instruction contain many comments about assessment strategies, suggesting that teachers in the higher performing schools were using many of the classroom-based formative assessment strategies identified above. For example, “Students were taught to become more reflective about their own reading and writing performance, sometimes using rubrics throughout the school year in order to help them gain insight into their better or less well developed reading and writing performance in response to particular tasks” (p. 861). An example that elaborates on this observation is provided below:

At Hudson Middle School, Cathy Starr taught her students strategies to use to reflect on their progress as they moved through an activity. After a research activity, the students were asked to rate themselves on their research and writing using rubrics they had developed.

- 1. Where do you think you fall for the research (grade yourself)? Did you spend the time trying to find the information? Did you keep going until you had learned enough to write your report?*
- 2. Whether this is a short and informal or longer and more formal piece, you should spend time thinking about the writing. Did you plan what you were going to say? Did you think about it? Did you review it and revise it before putting it in the back?*
- 3. Did you edit? Did you check the spelling and punctuation?*

Most of the teachers in the higher performing schools shared and discussed with students rubrics for evaluating performance; they also incorporated them into their ongoing instructional activities as a way to help their students develop an understanding of the components that contribute to a higher score (more complete, more elaborated, more highly organized response). Use of the rubrics also helped students develop reflection and repair strategies relevant to their reading, writing, and oral presentation activities. (p. 868)

More successful teachers, such as those studied by Langer, provide a clear description of the learning or performance goal, develop scoring rubrics to help students identify critical elements of success, and then provide repeated opportunities for students to apply the rubrics to their work and that of their

peers. This general assessment strategy leads to a core theoretical outcome of high-quality classroom-based formative assessments, which is to help students become more independent learners (Marshall, 2007). As students are involved in developing criteria for evaluating their own work, have repeated experiences in which they receive descriptive feedback about their performance, and gain experience evaluating their own and others' work, they presumably become more able to make needed corrections to their work when they are working independently.

Successful teachers in the high-performing schools also used interactive discussions frequently in their classrooms to provide feedback to students about their text analysis, thinking, and reasoning skills. "Teachers expected their students not merely to work together, but to sharpen their understandings with, against, and from each other," Langer writes (p. 872). One effective teacher used explicit activity guides to help students work together in groups toward common goals, and conducted "debriefing sessions and conferences [which] provided a time for each student to discuss not only the group's work but also to describe her or his own areas of accomplishment and need" (p. 873). The use of explicit guidelines and performance rubrics was widespread in the effective schools Langer observed and is consistent with the frequent emphasis on the involvement of students in their own assessment and that of their peers in discussions of classroom formative assessments (Marshall, 2007).

In light of the current high-stakes (accountability) testing environment, it is notable how Langer found that the effective schools tended to incorporate the state-level standards (as exemplified and assessed on the end-of-year examination) into their curriculum and ongoing assessment activities:

Teachers in the higher performing schools used the tests as an opportunity to revise and reformulate their literacy curriculum. The primary approach to test preparation involved relevant teachers and administrators in a careful deconstruction and analysis of the test items themselves, which led to a deeper understanding of the literacy skills, strategies, and knowledge needed for students to achieve higher levels of literacy performance. (p. 861)



Less successful schools in Langer’s study typically used test preparation strategies such as commercial materials or example tests that were essentially separate from the scope and sequence of the curriculum.

To be maximally effective, both instruction and classroom formative assessments should be aligned with state standards. In fact, at least one author (Abrams, 2007) has proposed that high-quality formative assessment practices that focus on a sound understanding of grade-level academic performance standards can help alleviate some unfortunate consequences of “high-stakes” testing such as use of commercial test preparation materials and emphasis on surface content coverage at the expense of deeper understanding. A good formative assessment program that has “unpacked” the state standards and identified the specific learning goals they contain can help focus classroom activities on real learning rather than test preparation.

A good example of how this kind of activity can improve the quality of instruction related to grade-level performance standards comes from a discussion by Ainsworth and Viegut in their book *Common Formative Assessments: How to Connect Standards-Based Instruction and Assessment* (2006). The example concerns middle school math teachers who were analyzing the results from their state-level accountability measure:

The teachers were confident that certain content had been covered and that students knew that content prior to the test, yet the data did not indicate that this was the case. Upon closer examination, the teachers realized that even though they had taught and assessed these particular concepts and skills, they had taught and assessed them as discrete, isolated elements. However, the large-scale assessment required students to combine multiple concepts into one cohesive response. (p. 5)

Langer’s study of high-performing schools showed that teachers in those schools made careful efforts to align their literacy curriculum with state standards, and used many classroom-based formative assessment practices to ensure that students received the instructional support and feedback they needed to learn to meet those standards. The essential point is that teachers should spend time studying the standards so they can systematically unpack them in order to identify the most important targets for instruction. The

understanding they gain from doing this can help to produce useful formative assessments and help focus classroom instruction properly.

In his discussion of the teacher's role, Stiggins (2007) describes the knowledge and skill required to integrate instruction and classroom-based formative assessments within a standards-based curriculum and illustrates how formative assessments aligned with curriculum standards can help to improve student learning. He suggests that teachers must play five roles:

1. *To become competent masters themselves of each of the standards their pupils are to master;*
2. *To understand how those standards transform into the curriculum that forms the scaffolding that pupils will climb on their journey up to each standard;*
3. *To make classroom-level achievement targets clear to pupils;*
4. *To transform the classroom targets into high-quality classroom assessments that are capable of accurately reflecting pupil achievement of those targets; and*
5. *To use those assessments over time in collaboration with their pupils to inform key decisions and to help motivate pupils to keep learning. (p. 18)*

These examples of classroom-based formative assessment suggest characteristics that help define the types of assessment that may be most effective for improving student performance. As we shall see in the next section, *classroom-based, short-cycle assessment has the strongest research base for improving student achievement.*

Evidence for the impact of classroom-based formative assessments on student achievement

Black and Wiliam (1998) provided the definitive review of the impact of classroom-based formative assessments on student learning. Encompassing approximately 250 studies, their review found that effective use of classroom-based formative assessments had an impact on student achievement of between .40 and .70 standard deviation units, a substantial effect. (For example, an effect size of .70 means that the median score of the group



whose instruction was guided by classroom-based assessments fell at the 76th percentile of students who were taught without the use of formative assessments.) Based on both the work covered in their review and on more recent research, Black and Wiliam (2007) concluded, "The question has been settled: Attention to classroom assessment practices can indeed have a substantial impact on pupil achievement" (p. 30). Other recent reviews (Brookhart, 2007; Gusky, 2007) reached similar conclusions, although two earlier reviews by Natriello (1987) and Crooks (1988) produced less conclusive support. According to Black and Wiliam (2007), the weaker support for classroom assessments in the earlier reviews most likely resulted from including a broad variety of practices, many of them ineffective, in the category.

An early review (Fuchs & Fuchs, 1986) of the impact of frequent classroom assessments on the performance of students with learning disabilities used much more restrictive criteria for including studies, and obtained a mean effect size on student performance of .70 standard deviation units across 21 studies. Students in the studies included in this review ranged in age from Pre-K through grade 12, and assessments were administered from two to five times a week. The impact of these assessment practices did not vary across the age range from early primary grades through high school. However, about half the studies required teachers to follow specific rules for reviewing the data along with prescribed actions to follow. The effect size for these studies was .92, compared with .42 in studies in which follow-up actions were left to teacher judgment. Additionally, studies that trained teachers to produce graphs of student progress as a guide to action obtained an average effect size of .70, compared with an effect of .26 for studies in which teachers simply recorded results.

Two caveats

It seems safe to conclude that high-quality classroom-based formative assessments can produce a significant impact on student performance. However, two important caveats to this conclusion must be introduced at this point.

The first caveat. It is impossible to disentangle the impact of formative assessment from the instruction that follows it. When improved student outcomes are used as the criterion for effective practices, evaluations of classroom-based formative assessment are as much an evaluation of the

instructional adjustments resulting from the assessments as they are of the assessments themselves. Formative assessments have no real utility apart from the feedback they provide to guide further student work or alterations to instruction. If the feedback to students or the instructional adjustments by teachers are not appropriate and skillful, improvements in student performance are unlikely to occur. This intimate, entangled relationship between the assessments themselves and the instructional adjustments they enable introduces an implementation challenge. We will elaborate on this challenge later in this document.

Clearly, increasing student performance through the use of formative assessments involves much more than simply adding assessments to the educational program; it is likely to involve fundamental changes in teacher instructional practice as well. One obvious change is that the teacher must carefully take into account a student's present level of understanding and skill when providing instruction. Another involves the idea that even the performance of students with low scores can be improved if instruction is targeted to their specific needs. The ability to make these instructional adjustments entails a large step forward in teaching skills. Serious efforts to help students acquire a skill or understanding when they do not acquire it "the first time through" raise questions about how instruction can be modified to make it more understandable, or how to engineer and motivate extra practice when initial attempts do not produce the level of knowledge or skill required to meet grade-level standards. These questions all require sophisticated educational decision-making and high levels of instructional skill.

The second caveat. Most of the research on assessment for learning with adolescents has been conducted in math and science or on lower level literacy skills (decoding and fluency) rather than with complex reading comprehension skills. The potential for widespread use of classroom-based formative assessments to improve adolescent student outcomes in academic literacy depends largely on generalizations from their effectiveness in other content areas, or from their utility in promoting lower level literacy skills. Furthermore, some principles of formative assessment may be easier to implement in these other disciplines than they are in literacy and social sciences. Bethan Marshall, an expert in formative assessment in these latter areas, cautions that: "Unlike math and science, which have developed a specific predetermined sequence for the teaching of certain concepts, in English, the humanities, and social



science, progression becomes a more meandering, organic affair” (p.137).

In other words, it may be easier in some disciplines to formulate the “next instructional steps” based on formative information than it is in adolescent-level academic literacy, where the target standards often involve complex thinking skills rather than the acquisition of specific concepts, or the mastery of well-defined algorithms. For example, in the proposed framework for the next National Assessment of Educational Progress in reading (NAGB, 2007), the “cognitive targets”—reading comprehension skills identified as standards for adequate performance—do not change much between 4th and 8th grade, but the passages from which the items are developed increase in sophistication and complexity. At 4th grade, a proficient reader is expected to be able to “identify the author’s implicitly stated purpose, summarize major ideas, find evidence in support of an argument, distinguish between fact and opinion, and draw conclusions from informational text” (p. 45). With more sophisticated text of the same type at 8th grade, proficiency in reading requires that students be able to perform the tasks listed above and “analyze and interpret implicit causal relations” (p. 46). Although we have reasonably strong research support for a variety of instructional techniques to build reading comprehension skills like these (National Reading Panel, 2000; Torgesen, Houston, Rissman, et al., 2007), the “path of growth” or “learning progression” (Popham, 2007) for improvement between 4th and 8th grade reading comprehension ability may be less clear than a comparable path between 4th and 8th grade proficiency in science or math.

Six characteristics of high-quality classroom-based formative assessment

We have previously alluded to some characteristics of effective classroom-based formative assessments. In this section, we systematically discuss the more salient characteristics of this type of assessment as described in the research and practice literature. In so doing, we also discuss the theoretical role each plays in helping to improve student learning outcomes.

1. ***Classroom-based formative assessment involves short-cycle, frequent assessment.*** The most straightforward characteristic of classroom-based formative assessments is that they take place every day in the classroom and are typically given repeatedly during each instructional period. The

short assessment-feedback-instructional adjustment cycle has a number of benefits for student learning. Perhaps the most important is that it allows the teacher to respond more quickly to correct student misconceptions or provide scaffolding to help move students to a higher level of understanding or learning. Rea-Dickens (2001) has suggested that frequent formative assessments “reduce the size of the task for the child,” and help keep the child “in the field” and motivated (p. 454). Stiggins (2007) pointed out that frequent “assessment *for* learning enables pupils by helping them watch themselves grow—by entitling them to retain the belief that, if they keep trying, success is within reach for them” (p. 17).

The idea that assessments should motivate students to continue to try by helping them track their incremental learning toward the ultimate goal is very different from the ideas behind most current assessments. A key to ensuring that formative assessments help maintain student motivation to learn involves the nature and quality of feedback from the assessment provided by the teacher or other students.

2. ***Classroom-based formative assessments can take a variety of forms.***

They are defined by their purpose, not their form. Effective formative assessments in the classroom can be administered in a variety of ways:

- *Rich performance tasks.* In their comprehensive literature review on formative assessments, Black and Wiliam (1998) indicate that “formative assessment which guides learners toward valued learning goals can only be generated with tasks that both work to those goals and that are open in their structure to the generation and display of relevant evidence, both from student to teacher and to students themselves” (p. 31). They cite a qualitative study (Garnet & Tobin, as cited in Black & Wiliam, 1998) of instruction in two very strong science classrooms that found one key to the teachers’ success was the way they continually monitored students’ understanding. Both classrooms sustained a broad diversity of activities with an emphasis on question-asking, with the students asking 60% of the questions. Marshall (2007) has suggested that “on a more practical level, teachers are involved in carefully devising and creating tasks that maximize opportunities for students to think through and develop their ideas as an aid to understanding and writing. The richer the task, the more meaningful the teacher’s feedback. Within this model of teaching and learning, the



most frequent form of feedback will be oral—characteristically...a follow-up question that prompts further thinking” (p.139).

- *Classroom and peer-to-peer discussions.* It is not a coincidence that improving the quality and frequency of discussions about the meaning of text is one of the major instructional improvements supported in current research on adolescent literacy (Applebee, Langer, Nystrand, & Gamoran, 2003; Torgesen et al., 2007). Classroom discussions are a way of determining the level of student understanding, while at the same time deepening it. In fact, the dynamics of a good discussion illustrate in some ways the interaction between formative assessment and instruction. If, in the course of a discussion, the teacher senses a misunderstanding on the part of students, the teacher’s next move is to present information or ask leading questions to help resolve that misunderstanding. A good interpretive discussion of text is an excellent vehicle to uncover the “thought process, logic, and understandings students use to arrive at selected answers or narrative responses” (Abrams, 2007, p. 91), one of the goals of high quality, classroom-based formative assessment.
- *Improving the quality of questions that teachers ask about the meaning of text.* This type of formative assessment is similar in many ways to discussion. Usually, higher-level, complex questions are followed by discussion, as a form of feedback to the student’s response. There is strong evidence from research on reading comprehension that instruction in which teachers ask high-level questions or teach students to generate useful questions is effective in improving reading comprehension (National Reading Panel, 2000). The questions teachers ask about text can be directly aligned with state-level literacy standards (Ainsworth & Viegut, 2006), but extensive research indicates that often they are not (Cizek, 2007). Because the definition of academic literacy includes both the ability to develop a basic construction of a text’s meaning and the ability to draw inferences and conclusions from it in order to answer questions, experience answering such questions in class, with usable feedback from the teacher or peers, helps build the literacy competencies identified in many national and state literacy standards.

-
- *Classroom tests and quizzes.* Black and Wiliam (1998) note that whether tests play a formative role in the classroom really depends on how the information they generate is used. Given frequently as a means for checking student understanding, with feedback and additional instruction following, tests can offer useful formative information. Tests and quizzes that require extended responses are more likely to provide useful formative information than those composed primarily of multiple choice items (Shepard, 2000). Scoring extended responses using rubrics developed in consultation with students is an excellent way to involve students in the formative assessment process. Not only do such rubrics provide a clear vision of the learning or performance target, they can be easily used to facilitate self- and peer-assessment.

3. ***Classroom-based formative assessments have clear, well-defined targets.*** In standards-based instruction, the targets of formative assessment for students working at grade level should be closely aligned to the state grade-level standards. As Stiggins (2007) points out:

Target definitions begin with state standards. While the learning is unfolding, pupils progress through the levels of proficiency that lead up to each standard. To make this possible, each standard must be deconstructed into a scaffolding that pupils must climb on their journey up to that standard. These continuously unfolding classroom targets (foundations of ultimate competence), then, become the focus of day-to-day formative assessments. For this deconstruction to be accomplished, each teacher must be a “confident competent master” of the standards his or her pupils are expected to master. If teachers don’t meet this professional standard, then unclear or inappropriate achievement expectations can result, leaving dependable assessment beyond reach. (pp. 15–16)

Ainsworth & Viegut (2006) provide an extended discussion of the value of aligning formative and summative assessments in the instructional system, and some of these advantages were discussed earlier in this document. Cizek (2007) reports that classroom assessments are



frequently not aligned with state standards; this area needs both further research and much deeper development in practice. It should be recognized that establishing clear, standards-based targets for formative assessment in literacy will be challenging because of the complexity of grade-level standards once students are past the initial stages of learning to read.

The value of aligning formative assessments with state-level standards applies primarily to students whose reading skills are reasonably close to grade-level standards. Other students, such as those receiving intensive intervention because of difficulties with basic reading skills, may have immediate instructional goals and assessment targets for a broader range of reading skills (e.g., reading accuracy and fluency) and for comprehension standards that are closer to their actual level of performance than the grade-level standards. Of course, the ultimate goal of intensive interventions is to enable students to read at grade level, but short-term goals and assessment targets need to be within reach and thus should be closer to the student's current reading level. It is also possible that some students with learning disabilities who have specific difficulties with word-level reading skills might have performance goals and assessment targets for listening comprehension that are quite different than those for reading comprehension. That is, we might expect students with strong broad cognitive abilities but specific weaknesses in their ability to read words in text to work on higher comprehension goals when material is read to them than when they have to read it for themselves.

4. ***Classroom-based formative assessments provide enough detail about student understanding, knowledge, and skill to suggest next instructional steps.*** As with defining clear assessment targets, developing assessments that provide useful information about "next instructional steps" in literacy may be more challenging than in some other content areas. As mentioned earlier, Black and Wiliam (2007) emphasize that both the creation of formative assessments and the interpretation of formative assessment data require a well developed "theory of the task." The teacher must have a deep understanding of the performance standard as well as the steps involved in attaining the standard in order to recognize the specific ways a student's current

performance falls short so that either corrective feedback or corrective instruction can be provided. Other researchers (Masters & Forster, 1997; Popham, 2007; Wilson & Sloane, 2000) have used the term *learning progressions* to describe a similar idea. A learning progression is a “carefully sequenced set of building blocks that students must master en route to a more distant curricular aim. The building blocks consist of subskills and bodies of enabling knowledge” (Popham, 2007, p. 83). Some sense of the learning progression required for attaining grade-level standards must underlie effective classroom-based formative assessments.

Research on reading and reading development over the past 30 years has produced a relatively coherent theoretical understanding of the major factors that influence reading comprehension in adolescents (Chall, 1996; Pressley, 2000). Broad factors such as the ability to read words in text accurately and fluently, vocabulary and background knowledge, use of comprehension strategies, thinking and reasoning skills, and motivation to perform well all influence performance on measures of reading comprehension in adolescents. However, state literacy standards often point to specific competencies within comprehension for which there is a less well developed “theory of the task” to guide formative assessment. We have little theoretical research on the component skills required to meet such standards as being able to describe the relationship between theme, setting, and character, or the ability to judge the logic, coherence, and credibility of an argument, and almost no instructional research focused on such specific standards. Broad instructional recommendations for improving adolescent literacy (Biancarosa & Snow, 2006; Torgesen, Houston, Rissman, et al., 2007) focus on general methods to improve word-level reading skills or reading comprehension, and do not address steps to achieve more specific state- or national-level literacy standards. It remains an empirical question whether specific “learning progressions” need to be developed for each end-of-year literacy standard or whether a more general approach to teaching reading comprehension (such as focusing on the ability to identify the main idea and supporting details, or the ability to make inferences) will suffice for helping students meet the most essential standards.



5. **Effective formative assessments are followed by appropriate types of feedback and instructional adjustments.** Black and William (1998, 2007) have emphasized the intimate connection between effective formative assessments and effective feedback. We also know that the impact of feedback on subsequent performance can be quite variable, depending on a number of features (Butler, 1988). Communication to the student based on formative assessment needs to energize and empower improved performance and not be discouraging. An early review by Crooks (1988) emphasized that feedback should focus on the task, and other research has stressed that feedback from formative assessments should be descriptive rather than evaluative (Kluger & DeNisi, 1996). A recent review of formative feedback by Valerie Shute (2007) produced a set of 31 guidelines for providing effective feedback that take into account learner characteristics and the timing of the feedback.

Given the complexity and depth of the cognitive and intellectual factors that can influence students' ability to achieve such literacy standards as making inferences that describe a problem and solution or analyzing and interpreting implicit causal relations, providing appropriate feedback and implementing next instructional steps should be at least as challenging as implementing other features of high-quality formative assessments in literacy. These concerns about the complexity of the instructional targets in grade-level literacy standards for adolescents underline the need for continuing study and discussion of the standards among teachers and concerted professional development opportunities to build both assessment and instructional capacity.

6. **Classroom-based formative assessments involve students in self- and peer-assessment.** An important idea in the literature on classroom-based formative assessments is that the frequent assessment-feedback-new performance cycle can actually help students become more independent, self-regulated learners (Marshall, 2007). As students are involved in developing scoring rubrics for complex performance tasks, as they participate in open discussions of test results that provide feedback for improvement, and are asked to evaluate their own and others' work, they gradually learn to be more reflective about their own performance. This aspect of classroom-based formative assessments is supported by the

literature on goal-setting for and self-assessment of performance goals, which consistently finds a positive impact on student performance (Paris & Paris, 2001; Schunk, 2003).

Curriculum-based measurement (CBM): A special case of classroom-based formative assessment

We have discussed the value of short-cycle, classroom-based formative assessments that are closely linked to ongoing instruction and provide guidance to both teachers and students to improve teaching and learning. These assessments can take a variety of forms, can assess immediate instructional objectives, and can vary across time and across classrooms.

Curriculum-based measurement (CBM) is a specific form of classroom-based formative assessment that shares a number of important features with the type of assessments we have been discussing. Curriculum-based measures are similar to the type of assessments Black and Wiliam (1998) and Stiggins (2007) describe in that they can be given very frequently, are designed to provide information that can lead to effective instructional adjustments, and focus on important curriculum goals (Deno, 2003; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). However, the most common form of CBM, *general outcome measurement*, has some important differences that distinguish it from the type of classroom-based formative assessment that we have been discussing. As used widely to guide instruction in early literacy (Good, Simmons, & Kame'enui, 2001) and in special education, general outcome CBM has three important characteristics (Fuchs & Fuchs, 2003):

First, CBM is standardized: the behaviors to be measured and the procedures for measuring those behaviors are specified. Second, CBM's focus is long term: testing methods and content remain constant, with equivalent weekly tests spanning the school year. Third, each week's test content reflects the performance desired at the end of the year and therefore samples the many dimensions of the year's curriculum. So, standardized procedures, long-range consistency, and an integrated focus on the many skills cumulatively addressed in the annual curriculum are CBM's distinguishing features.
(p. 329)



One important value of general outcome CBM is that it can be used to set precise intermediate goals for student performance related to end-of-year grade-level expectations.

For example, one widely used set of general outcome measures for early literacy is the Dynamic Indicators of Basic Early Literacy Skills (Good, Simmons, & Kame'enui, 2001). This suite of assessments allows teachers to track the development of phonemic awareness, letter knowledge, phonemic decoding, and oral reading fluency using brief, one-minute probes that can be administered up to 20 times a year if desired. Development work on these probes established desirable performance levels, or "cut points," at three times during the year (fall, winter, spring) so that teachers can determine which students are making adequate progress in acquiring critical pre-reading and reading skills at several points during the year (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002).

General outcome measures do not provide detailed information about *why* students might be performing below expectations, but rather indicate *whether* students are making acceptable progress on general indicators of important reading outcomes. The measure of oral reading fluency, for example, provides a single outcome indicating the number of words read correctly in one minute. Although close teacher observation while students perform the task can indicate more detailed features of student capability, such as whether the student self-corrects or has a serious problem with reading accuracy, the test outcome coded in most scoring systems is the single number describing overall performance. The task is not designed to provide information about *why* the student is struggling with oral reading fluency; it provides a reliable indicator of whether or not growth in reading fluency is occurring at an adequate rate.

General outcome CBM has been studied extensively as a tool for guiding instruction with students in special education, using a procedure referred to as *data-based program modification* (Deno, 2003; Fuchs, Deno, & Mirkin, 1984). This approach to using general outcome CBM formatively is also sometimes referred to as *experimental teaching*. The meta-analysis mentioned earlier (Fuchs & Fuchs, 1986) reviewed a large number of studies showing the positive impact of data-based program modification on student outcomes in basic reading skills as well as other academic areas.

In the experimental teaching approach to formative assessment, the teacher implements an intervention program and evaluates its effectiveness

using general outcome CBM. If the intervention is ineffective (i.e., the slope of improvement indexed by the general outcome measure is too shallow), the teacher modifies the intervention in a way that is expected to make it stronger. If the modification proves effective (i.e., slope of improvement increases), the teacher continues to implement the more successful instruction. The teacher might also introduce other modifications to the intervention with the hope of further increasing the rate of learning. Experimental teaching requires at least weekly data collection; otherwise, the effectiveness of various program components cannot be evaluated in a timely way. There is a large literature involving randomized control trials demonstrating that using general outcome CBM (formative assessment) to modify and build individualized programs for students produces positive learning outcomes on widely accepted, standardized measures of reading skill (cf. Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 2003).

Although curriculum-based measures that assess general outcomes have proven successful in helping teachers monitor the progress of early reading growth in young children and in special education students receiving instruction to improve basic reading skills, they are more difficult to use with adolescents reading at grade level. The problem with using quick and simple measures to chart literacy growth in adolescents working on grade-level skills is that the ultimate target of instruction, the ability to comprehend and reason about the content of text, is so multiply and complexly determined that no simple and brief measure can adequately chart its developmental course. For example, in a study examining the relationships of various reading and language skills to performance on the Florida Comprehensive Assessment Test (FCAT) in reading, Schatschneider et al. (2004) showed that individual differences in oral reading fluency accounted for 55% of the variability on the FCAT in 3rd grade but only 32% of the variability in 10th grade. Measures of the other, more complex dimensions of reading comprehension performance, such as vocabulary, reasoning skills, or background knowledge, all take considerable assessment time to produce reliable estimates of ability as does complex reading comprehension itself. It is also true that these “part” measures of the component skills involved in reading comprehension would not adequately reflect the context-dependent strategies for reading comprehension that are so important in being able to respond to the thinking and reasoning requirements embedded in most state literacy standards in middle and high school. Thus, the



best general progress monitoring test for grade-level reading skills in late elementary, middle, and high school is likely to be the type of benchmark assessment that mimics as closely as possible the actual information-processing and test-taking skills assessed on end-of-year summative tests. Such tests take substantial time to administer but can be administered via computers or in classroom-sized groups.

Although general outcome CBM, as currently developed, may not be useful for charting the development of grade-level skills in adolescents' reading comprehension, some measures of this type may be useful indicators of progress for students receiving intensive interventions in word-level skills or more basic aspects of reading comprehension. For example, measures of oral reading fluency might be useful as part of a progress monitoring assessment for 7th grade students reading at the 2nd or 3rd grade level who are receiving interventions to improve their word-level reading skills as well as their comprehension skills. If one goal of the instruction they are receiving is to strengthen their ability to read accurately and fluently, then it makes sense that a reliable indicator of progress in this area should be used to monitor the effectiveness of the instruction being provided. If very little progress is being made, this indicates that the instruction needs to become more intensive or that the instructional program needs to be changed.

Oral reading fluency measures may help document improvements in struggling readers, but Wayman et al. (2007) suggest that maze selection tests may be an even better measure of improvements in adolescents' basic reading skills. These tests can be administered in a group or by computer and assess both fluency and a basic level of comprehension. They ask students to read passages and supply (or select from among several choices) words to fill in blanks that are spread throughout the text, requiring the student to maintain at least gist-level comprehension of the passage in order to select words that best fit the context. The score on such a test is typically the number of maze items that are correctly completed within some time limit (often three minutes), so silent reading fluency does have an effect on the score.

General outcome measures are only one of two broad types of curriculum-based measures. The other type is referred to as *mastery measurement*, in which the objectives targeted for assessment change as students move through the curriculum. That is, alternative forms of a criterion-referenced assessment are used to assess progress until an instructional objective is

achieved, then the assessment changes to a new set of targets. Mastery measurement is closer to the type of classroom-based formative assessment studied and advocated by Black and Wiliam (1998) than general outcome CBM. Mastery measures can target specific instructional objectives and can also be developed to assess component skills of more complex behaviors that are the ultimate targets of instruction. These types of measures could be developed to assess aspects of the complex skills required for grade-level reading comprehension, such as the ability to make inferences, identify the main idea and supporting details, or make good summaries.

Examples of curriculum-based mastery measurement assessments in support of literacy instruction for adolescents can be found in Part II. To illustrate, example #4 involves a curriculum to help students make better inferences from text, and the instructional materials include progress assessment tests that evaluate how well students can perform steps involved in making inferences. Frequent administration of these formative assessments helps teachers identify aspects of the strategy that need additional instruction and helps students chart their progress in mastering the steps of the strategy.

Summary and conclusions about classroom-based formative assessment

Better use of classroom-based formative assessments appears to be one important way to improve student outcomes in academic literacy for students in grades 4 through 12. When done well, these assessments help teachers provide instruction, feedback, and support tailored specifically to student needs. Because they provide detailed information about what students know, how they think, and what they are able to do, classroom-based formative assessments can direct teachers to the appropriate next steps in instruction. They also help students stay engaged in the sometimes lengthy process of acquiring complex literacy skills by helping them see their improvement across relatively small learning steps. As Stiggins (2007) has suggested, "Both pupil and teacher must know where the learner is now, how that compares to ultimate learning success, and how to close the gap between the two. Pupils must not wonder whether they will succeed, only when. Incremental success must always be within reach in their minds" (p.15).



These types of frequent, instructionally embedded assessments have received strong support in the research literature as a means of improving student learning in many areas. Because of the complexity of instructional goals for adolescent literacy, however, using classroom-based formative assessments to support literacy instruction may present unique challenges. These challenges, outlined on pages 57–64, will need to be addressed in future research and through extensive professional development for teachers. Part II describes several examples of current attempts to incorporate classroom-based formative assessments into middle and high school literacy instruction.

It seems clear that we will not make major improvements in developing more powerful literacy instruction that meets the needs of *all* adolescents unless we also make significant improvements in the classroom-based formative assessments needed to guide that instruction. We turn now to another type of formative assessment increasingly used in middle and high schools. These *benchmark*, or *interim*, assessments meet some, but not all, of the goals of formative assessment.

The role of standards-based benchmark assessments in improving student literacy outcomes

Standards-based benchmark assessments typically assess student performance on state literacy standards—in a manner similar to the end-of-year accountability measure—so that gaps in student performance can be identified in time to provide extra instructional support for students who may need it. These assessments are frequently referred to as interim assessments, and those who conduct research on formative assessment consider their function and utility as falling somewhere between “true” formative assessments (described in the previous section) and end-of-year summative assessments. In *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief*, Perie, Marion, Gong, and Wurtzel (2007) offer this definition:

Interim assessment is the term we suggest for the assessments that fall between formative and summative assessment, including the medium scale, medium-cycle assessments currently in wide use. Interim assessments: 1) evaluate students’ knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and 2) are designed to inform decisions at both the

classroom and beyond the classroom level, such as the school or district level. Thus, they may be given at the classroom level to provide information for the teacher, but unlike true formative assessments, the results of interim assessments can be meaningfully aggregated and reported at a broader level. As such, the timing of the administration is likely to be controlled by the school or district rather than by the teacher, which therefore makes these assessments less instructionally relevant than formative assessments. These assessments may serve a variety of purposes, including predicting a student's ability to succeed on a large-scale summative assessment, evaluating a particular educational program or pedagogy, or diagnosing gaps in a student's learning. Many of the assessments currently in use that are labeled "benchmark," "formative," "diagnostic," or "predictive" fall within our definition of interim assessments. (pp.1–2)

A recent report commissioned by the Rand Corporation (Marsh et al., 2006) noted that 80% of school superintendents in California, Georgia, and Pennsylvania found results from their local benchmark assessments more useful for decision-making than state test results. In another district that was studied intensively, 80% of the principals indicated that their standards-based progress tests were moderately to very useful for guiding decisions about instruction. Furthermore, one organization that manages high schools across the United States (see Example #7 in Part II) uses computer-based, standards-aligned progress tests extensively, and interviews indicate that the data are considered very useful by teachers and principals, as well as corporate staff who use them for monitoring schools (Marsh et al., 2006). The same study, however, noted that teachers typically find classroom-based tests more helpful for guiding instruction; the benchmark tests often provide information already known to them and are frequently seen as taking time away from instruction.

Interim progress tests aligned to state literacy standards are widely used by districts and schools as part of their adolescent literacy assessment plans. However, numerous experts in formative assessment (Black & Wiliam, 2007; Cizek, 2007; Stiggins, 2007) question whether these tests, which are marketed as formative assessments, can actually serve the most important purposes of formative assessment. In a paper prepared for the Council of Chief State School Officers, Perie, Marion, and Gong (2007) state:



If the purpose of these assessments is to enrich the curriculum, challenge the students to self-diagnose their own learning, provide insights into any misconceptions the students have, or provide additional professional development for the teachers, many of these types of assessment systems are woefully inadequate. (p. 34)

In another policy guidance paper, the same authors (Perie et al., 2007) concur with many other researchers in this area that the marketers of interim assessments improperly cite research on the positive impact of classroom-based formative assessments, which are very different from the products they are marketing to schools and districts.

The research supporting the efficacy of assessment to improve teaching and learning is based on formative assessment—the minute-by-minute, classroom assessments....There is simply no research base to support the claim that interim assessments improve student learning. While interim assessment has considerable intuitive appeal, experience shows that there is often a considerable gap between intuition and research evidence. (p. 21)

These are harsh criticisms of a form of assessment currently in wide use in middle and high schools across the country. However, these same authors also acknowledge that there may, indeed, be an important place for this type of assessment within a comprehensive assessment plan, as long as policymakers clearly understand what they are getting for their investment. A number of questions must all be carefully considered before investing the considerable resources that these types of assessments require. Questions include the purpose of the assessment, who will use the information it generates, what kind of actions will be based on its results, what kinds of professional development and support are required to execute the actions it prescribes, and how, exactly, it will affect student learning.

Some discussions of interim or benchmark progress monitoring assessments have outlined ways that these assessments might be made more relevant for instructional decision-making in the classroom. For example, Perie et al. (2007) listed 11 criteria of interim assessments that would make them more useful in guiding a broader range of instructional decisions about individual students:

-
1. *Questions are not all multiple choice items;*
 2. *Provision is made for qualitative insights about understandings and misconceptions and not just a numeric score;*
 3. *Immediate implications for what to do besides re-teaching every missed item;*
 4. *Rich representation of the content standards students are expected to master;*
 5. *High-quality test items that are directly linked to the content standards and specific teaching units;*
 6. *A good fit within the curriculum so that the test is an extension of the learning rather than a time-out from learning;*
 7. *A good fit with curriculum pacing so that students are not tested on content not yet taught;*
 8. *Clear reporting that provides actionable guidance on how to use the results;*
 9. *Validation of the uses of and information provided by the assessment;*
 10. *Administration features (speed, availability of normative information, customization, timing flexibility, adaptability) match the assessment purposes; and*
 11. *Professional development for teachers. (pp. 7–8)*

Some features that make benchmark assessments more useful for guiding classroom instruction may make them less useful for school-level decision making. For example, too much focus on qualitative insights about understandings and misconceptions may make it difficult to aggregate this information to the school or district level. There may be tension between the type of formative information that is most useful for principals and district staff and what is most useful for teachers in planning instruction (Ingram, Louis, & Schroeder, 2004; Marsh et al., 2006). If an assessment is given district-wide at a specific time, it may be difficult to satisfy the 7th criterion that students not be tested on standards they have not yet been taught. This would require relatively uniform instructional pacing across all schools in a district, or all teachers within schools, which might place too many constraints on teachers' ability to adjust instructional pacing according to their students' needs.



An example of standards-based periodic assessment

A number of companies throughout the United States are now producing benchmark assessments tailored to the needs of individual districts or states. One of these organizations produces assessment software and items that can be used to administer a computer-adaptive test of reading comprehension that is customized to individual state literacy standards. The developers indicate that the test provides a valid and reliable assessment of students' ability to meet grade-level standards in reading that can satisfy four purposes: 1) help to assign students to classes, 2) help form instructional groups within classes, 3) help teachers and parents understand student progress, and 4) identify, or diagnose, areas of skill within the standards that are in need of extra support. The first three purposes are accomplished by providing an accurate assessment of the students' overall reading level at different times during the year. If a school wanted to group students for literacy instruction according to their general instructional needs (how well they can meet state literacy standards), data from this test could be used to do that, at least at a beginning level of differentiation. The test can also provide a reliable estimate of the extent to which a given student is making expected yearly progress in academic literacy.

It is in the area of "diagnostic" assessment of students' differential abilities to meet various standards that this test, as well as many other benchmark tests, is of limited utility. The developers acknowledge this difficulty, which results from the limited number of items on a test of reasonable length that can be devoted to assessing student performance on each standard. Since only a few items on the test are used to assess each standard, performance on individual standards (or groups of standards) is assessed less reliably than is overall performance. Cizek (2007) has discussed this problem in some detail and points out that the reliabilities for assessment of individual standards on many benchmark tests is often so low that, "when you look at the reliability of the difference scores across areas, it can easily approach zero" (p. 104). This problem, of course, makes it difficult to use these assessments reliably to identify areas in which students are in need of special instructional support.

Because these standards-based benchmark tests can be developed to predict with reasonable accuracy which students are likely to be successful on the end-of-year summative test, they are frequently used to identify "bubble students"—students who are close to being able to meet grade-level standards, and who might be helped to meet standards with some extra instructional support. The Rand study (Marsh et al., 2006) reported that three fourths of the principals in three states encourage teachers to focus on these students, and about a third of teachers reported that they do focus extra teaching effort on them. Although this is clearly one way data can guide instructional decisions, it raises concerns about instruction for students who are "outside the bubble," performing either above or seriously below standards.

Teachers, schools, and districts have much to learn about how to use benchmark, or interim assessments, in a way that maximizes their instructional utility. Part II presents several examples of ways schools are currently using benchmark assessments.

The role of screening and diagnostic assessments in a comprehensive assessment plan

Focused on assessment for learning, this document rests on strong assumptions about the utility of guiding instruction by using information about students' current knowledge and skills related to academic literacy. Literacy instruction is improved when it is guided by sound *diagnostic information* regarding students' instructional needs. Without this information, instruction might be provided where it is not needed, or at a level beyond the student's understanding, or in a way that is insufficiently challenging to move students to higher levels of proficiency.

It is important to distinguish between *diagnostic tests* and *diagnostic information*. *Diagnostic information* is any knowledge about a student's skills, abilities, knowledge, or attitudes that is useful in planning instruction to help the student learn to read more proficiently. It can be generated by classroom-based formative assessments, by benchmark or interim assessments, by curriculum-based measures, by beginning-of-the-year screening tests, or by formal or informal diagnostic tests or procedures. Information from these assessments can help a teacher understand why a student is having difficulty with a certain aspect of reading.

In educational circles, the term diagnostic test is often restricted to formal (standardized) or informal tests that are given at one point in time and assess a variety of reading components or other abilities related to reading. In the context of a fully implemented, comprehensive assessment plan, formal diagnostic tests have a much smaller role to play than in plans where no other reliable and valid assessment information is available.

Formal, standardized diagnostic batteries assess a range of reading skills such as word identification, phonemic decoding, vocabulary, and sentence or passage comprehension. They typically compare the student's performance to a large representative sample of students (the norming sample) at the same age or grade level and indicate a student's relative strengths and weaknesses in



the areas assessed on the test. They typically provide percentile ranks or standard scores to indicate how the student compares in each area with other students of the same age or grade. The profiles of students unable to meet grade-level reading standards might be quite different on such tests. For example,

1. Student A might perform very poorly on all of the subtests, and would be identified as a student with pervasive and serious reading difficulties;
2. Student B might perform within the average range on word identification and phonemic decoding, but relatively poorly on vocabulary and reading comprehension. This student would not need intervention for word-level skills, but would need extra support for the growth of vocabulary and comprehension;
3. Student C might perform poorly on the word identification and phonemic decoding tests, but better (although perhaps not in the average range) on vocabulary and comprehension. This pattern is typical of older dyslexic students who often have strong general cognitive ability, but need specific and intensive support in acquiring accurate and fluent word reading skills (Shaywitz, 2003).

Formal diagnostic tests identify broad areas of strength and weakness in students' literacy skills but do not offer specific information about what kind of phonemic decoding skills are deficient or what type of comprehension strategies the student is not yet using. Such diagnostic information can be useful for identifying a broad focus of instruction but does not really help in planning details of the first day's lesson.

The second type of diagnostic test widely used to guide instruction is often referred to as an *informal reading inventory* or *informal diagnostic inventory*. These tests ask students to perform a number of different reading tasks such as reading lists of words, reading passages orally, or answering questions about the meaning of words. In the hands of a skilled teacher, they provide a richer context than formal diagnostic tests for observing what the student knows and can do with regard to a range of reading skills and are often used to identify a starting place for instruction in areas such as word analysis, vocabulary, or comprehension strategies.

At this point, it may be useful to consider an example of a 9th grade student who has performed below grade-level standards on the state

accountability measure in literacy. Possible reasons for this student's difficulties include:

1. Difficulties reading the words on the test accurately and fluently;
2. Limited knowledge of the vocabulary on the test passages;
3. Limited knowledge of the content domains represented in the text, or lack of schema-related knowledge that is relevant;
4. Fundamental language comprehension difficulties;
5. Inefficient use, or failure to use active thinking and comprehension strategies while reading in order to maximize comprehension of deep meanings;
6. Limited knowledge of various terms (author's purpose, theme, characters, motivation, etc.) that might appear in the questions;
7. Limited ability to do the kind of thinking required by the questions—making connections, drawing inferences, forming conclusions;
8. Poor test-taking strategies, which can be caused by lack of understanding the nature of the questions or lack of familiarity with and understanding of testing format; and
9. Weak motivation to expend the effort required by the test.

Rather than resulting from only one of these factors, the student's poor performance on this state-level examination is likely to be associated with weaknesses in a number of these areas. If this student were attending a school that had a range of instructional/intervention options available, how could information be obtained that would allow instruction to focus on the areas most responsible for his or her poor performance? We consider each difficulty in turn.

1. *Problems at the word-level of reading.* These could be diagnosed either through a brief screening measure administered to all below-grade level students at the beginning of the year or assessed using a formal or informal diagnostic assessment.
2. *Problems with vocabulary.* These could be assessed through a formal diagnostic instrument or a vocabulary screening instrument, or the school could simply assume (and be right most of the time) that the student needed extra vocabulary support, because almost all struggling readers



in middle and high school are relatively weak in vocabulary (Cunningham & Stanovich, 1998). If this assumption were incorrect, the student's intervention teacher would quickly discover this through ongoing classroom-based formative assessments.

3. *Limited knowledge of content domains.* This is not assessed on formal diagnostic inventories in reading and would be very difficult to determine in any event, because it would depend on the context of specific passages on the test. Again, the school could simply assume (and, again be right almost all the time) that struggling adolescent readers will have missed out on significant learning opportunities and thus will be as relatively deficient in conceptual knowledge as they are in vocabulary (Cunningham & Stanovich, 1998).
4. *Fundamental language comprehension difficulties.* This relatively rare problem would require diagnosis using a formal test of linguistic competence. However, such a test should not be given until the student shows lack of response to high-quality, intensive interventions focused on reading difficulties identified through other means.
5. *Inefficient use of comprehension strategies.* No currently available formal diagnostic tests examine the extent to which students use effective, grade-appropriate reading comprehension strategies. However, "think aloud" procedures can be used to obtain diagnostic information in this area and these procedures have the additional advantage of being very closely linked to instruction (Kucan & Beck, 1997; Laing & Kamhi, 2002).
6. *Limited knowledge of terms in the questions.* There is no formal diagnostic assessment for this; it would be best determined through classroom-based formative assessments.
7. *Limited ability to do the kind of thinking required by the questions.* Again, no adequate formal diagnostic test for this is currently available, and it would be best determined through classroom-based formative assessments.
8. *Poor test-taking strategies.* These are not assessed on formal or informal diagnostic inventories and would be best determined through classroom-based formative assessments.

-
9. *Weak motivation to do the work required by the test.* Although surveys of reading motivation are available (Wigfield & Guthrie, 1997), this student characteristic is probably best assessed by teacher observations during classroom work.

A formal diagnostic test would not be needed to determine that this student had a problem with reading comprehension, because that had already been demonstrated on the end-of-year, standards-based accountability measure. However, since 9th grade literacy standards typically require students to answer relatively complex, high-level questions that involve making inferences, analyzing arguments, and drawing conclusions, it might be useful to determine whether the student also has difficulty with lower level comprehension skills, such as those assessed on a three-minute maze selection task or on a comprehension test involving short passages and questions that do not require complex inferences. Such measures are available on formal diagnostic batteries and can be administered in groups or on computers.

Full, formal diagnostic batteries will be administered rarely, if the purpose is to obtain information that can be directly used to improve instruction. Such batteries may be required to identify students for special education, but that is a different purpose from planning instruction. Formal diagnostic assessments must meet high standards for reliability and validity because they are often used to make “high-stakes” placement decisions involving the allocation of special education resources. However, if the purpose is to obtain diagnostic information at the beginning of the year (or really, anytime during the year) to place students in appropriate intervention groups, it would be much more efficient and cost-effective for schools to use screening measures that are either very quick to administer or can be given to groups of students.

Once students have been placed in interventions broadly targeted to their major educational needs, it may be helpful for the intervention teacher to administer an informal reading inventory or an informal diagnostic assessment to gain more detailed information about what the student knows and can do. Information from these more detailed assessments should help teachers identify subgroups of students within their intervention classes who have similar instructional needs.

There is one final point to make about the utility of diagnostic tests in planning instruction in academic literacy. Many formal diagnostic tests are

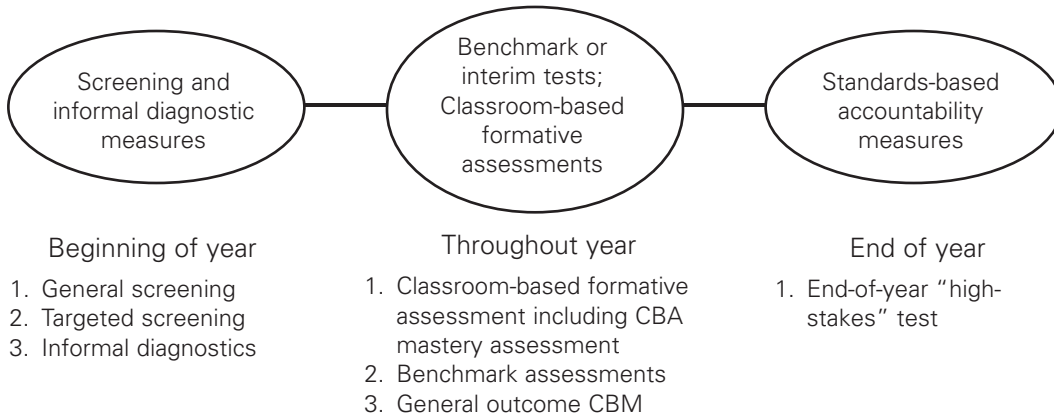


purported to diagnose specific cognitive or language skills that are important for reading growth. These tests measure such constructs as verbal short-term memory, visual processing ability, auditory processing ability, rapid automatic naming skill, spatial or visual memory, etc. Although some of these constructs may have strong or moderate predictive relationships with reading growth, there is no compelling evidence that knowing a child's score on any of these non-reading tests can help teachers provide more effective instruction in reading. These constructs are sometimes assessed to determine whether a child has a learning disability in reading, but, according to the preponderance of evidence from research, they do not help plan more effective instruction for students with learning disabilities (Torgesen, 2002). For example, children who perform poorly on measures of rapid automatic naming of digits and letters frequently have difficulties acquiring fluent reading skills. In fact, very low performance on this measure can indicate the presence of a learning disability in reading (Wolf, 1991). However, no interventions with demonstrated effectiveness are available to directly improve adolescents' performance on this construct. Rather, low performance on this measure indicates the need for careful attention to the acquisition of fluency at all stages of learning to read. If fluency on phonemic awareness tasks, letter knowledge tasks, phonemic decoding tasks, and text reading is regularly monitored during reading instruction, then teachers will be alerted in a timely fashion to students who require more support for the development of reading fluency.

Summary and review: Assessment types required in a comprehensive assessment plan

We have discussed the assessments that are logically part of a comprehensive assessment plan for adolescent literacy. We have focused primarily on assessments to guide instruction, while also recognizing the importance of end-of-year summative measures to determine how well overall goals for literacy instruction are being achieved. In Figure 2 below, we have added some text to Figure 1 (see p. 17) to identify all the types of assessments we have discussed thus far. This figure shows where each of the assessments we have discussed fits into a comprehensive assessment plan for adolescent literacy.

Figure 2: Details of Assessment Types and Timing



Formal and informal diagnostic measures can really be given anytime during the year, whenever the need arises.

CBA = Curriculum-based assessment

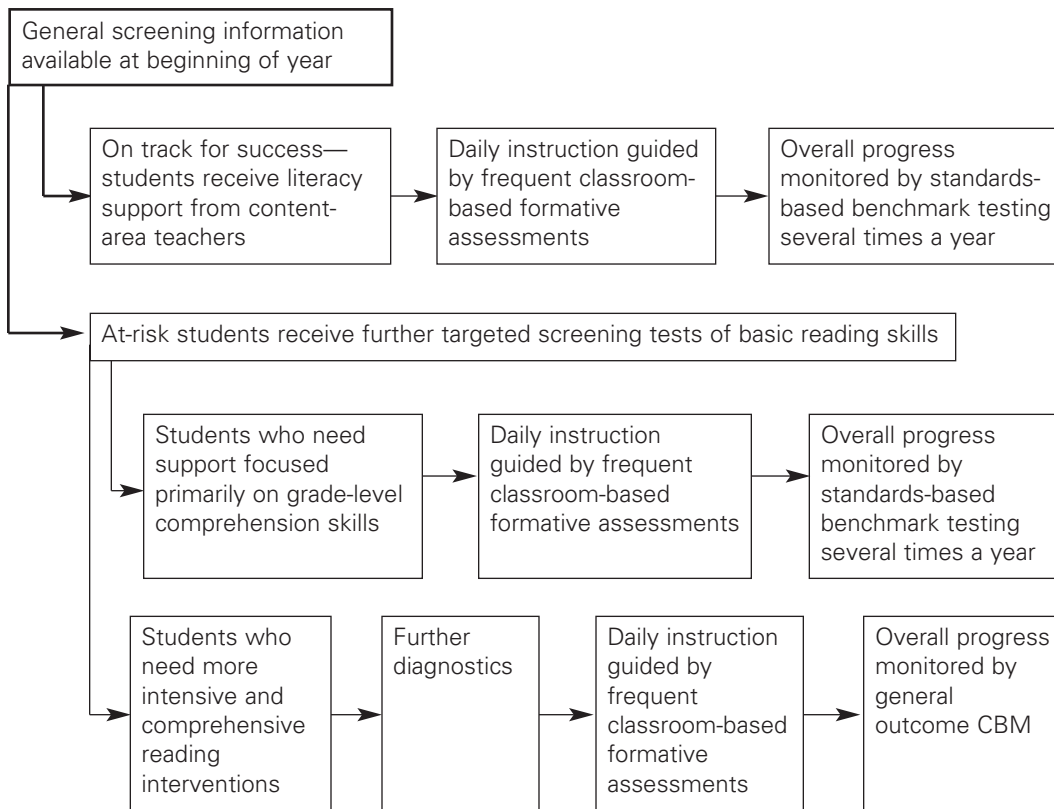
CBM = Curriculum-based measurement

Figure 3 provides a more detailed flowchart showing how the formative assessments summarized in Figure 2 might be used at different points in the instructional year with different students.

The box on the top left in Figure 3 labeled "General screening information available at beginning of year" identifies the need for data about every student's literacy skills that will be helpful in planning literacy instruction and support for the year. This information could come from the previous year's summative test or from a benchmark predictive assessment administered at the beginning of the year. These data would be used to make an initial classification of students into those who should be able to meet grade-level standards in reading for the current year (on track for success) through the literacy support provided in their content-area classes, and those who are unlikely to meet grade-level standards without extra support or intervention (at risk).



Figure 3: Application of a Comprehensive Assessment System



Content-area literacy instruction for students classified as *on track* would be guided by *classroom-based formative assessments*, and their progress toward grade-level standards for the current year could also be assessed through formal *benchmark*, or interim assessments administered several times during the year.

Students identified as *at risk* for failure to meet current grade-level standards would receive one or more *targeted screening tests of basic reading skills*. Students who did *reasonably well* on these screening tests could receive intervention support focused primarily on helping them acquire the complex comprehension skills described in the state literacy standards for their grade. Instruction in these intervention settings would be guided by *classroom-based formative assessments*. The students might also take the grade-level *benchmark assessments* to help measure their progress toward meeting grade-level literacy standards.

Students who were found to have *serious difficulties with basic reading skills* would receive a more intensive and comprehensive intervention where they might be administered a variety of *informal diagnostic assessments* to help identify their most critical instructional needs. Following this initial assessment, their instruction would be guided by *classroom-based formative assessments*. Their progress in acquiring basic reading skills might be assessed using *general outcome CBM*. If the intervention teacher were following procedures for data-based program modification (Fuchs, Deno, & Mirkin, 1984), these general outcome measures might be administered weekly. At present, no research-based information indicates whether students with serious reading difficulties should take grade-level benchmark tests alongside students reading close to grade level. Most likely, these benchmark tests would not be very sensitive to reading growth made by students with reading skills substantially below grade level.

It is unlikely that many schools currently implement such a wide range of assessments. Many schools may be using one type of test, hoping that it can be useful for several different ends. As Perie, Marion, Gong, and Wurtzel (1997) observe: “Educational leaders are often trying to squeeze as many purposes as possible out of a single system. However, few assessments or assessment systems can serve more than two or three purposes well and they tend to work best when the various purposes have been prioritized explicitly” (p. 4). Educational leaders and teachers need to be very careful about trying to make one assessment, or assessment system, serve too many ends. As we have seen, each broad type of assessment in Figure 3 is useful for a relatively narrow range of goals, and serious distortions can result if a single assessment is required to serve too many purposes. Creating a comprehensive system such as the one described in this document may extend over several years, with many obstacles to overcome along the way. Nonetheless, such fully developed systems may be critical to support large improvements in student outcomes over time.



CHALLENGES TO THE SUCCESSFUL USE OF ASSESSMENT FOR INSTRUCTION WITH ADOLESCENTS

Teachers and educational leaders face many challenges in making assessment for learning an important part of their instructional work with adolescents, ranging from individual changes in attitudes and practices to alterations in school organization and resource allocation. After indicating that classroom-based formative assessments have the potential to improve student learning outcomes significantly, Black and Wiliam (2007) also said, “What is less clear is...how the gains in pupil achievement that the research shows are possible can be achieved at scale” (p. 30). In this section, we identify several important obstacles school leaders must confront and overcome if formative assessment is to realize its potential for improving student outcomes. We also offer suggestions for meeting these challenges.

Challenge #1: Fundamental changes in attitudes and instructional practices

Effective assessment for learning will involve much more than simply adding new assessments. Three aspects to this challenge seem most daunting. Schools’ effective use of formative assessments to improve student learning requires: 1) reorienting to the purposes of assessment and methods of teaching; 2) acquiring many new assessment and teaching skills; and, 3) identifying new instructional resources or reallocating instructional resources.

A central aspect of standards-based education is setting basic achievement goals that *all* students are expected to achieve. Students are free (and many are expected) to achieve far above a state’s basic standards for grade-level proficiency, but, at a minimum, schools are expected to increase substantially the number of students who can meet those basic grade-level standards. This goal is completely consistent with assessment for learning; in fact, one can argue that the goal *cannot* be achieved until we become reoriented to the purposes of assessment and begin to incorporate assessment *for* learning as an integral part of all classrooms.

A number of studies (Abrams, 2007) have shown that standards-based educational reform has effectively changed the content of *what* is being taught

in many middle and high schools, but there is little evidence that it has changed *how* instruction is being delivered. One aspect of teaching that must change if more students are to receive the instruction and support they need to achieve grade-level standards is that it must become more *differentiated*. This concept is well understood and quite broadly practiced in the early elementary grades (Kosanovich, Ladinsky, Nelson, & Torgesen, 2006) but is less a part of instructional practices in middle and high school.

The concept of differentiated instruction rests on the fact that students come to any learning environment with different learning abilities and different kinds of preparation for learning. Some students will require more instruction and support to reach important learning goals, and others may require different teaching approaches with the same content. Where there is little variability in teaching students in diverse classrooms and schools, there is usually great variability in student outcomes. If schools and classrooms are not organized to provide some students with more intensive instruction, many students will not learn at rates sufficient to meet grade-level standards. Strictly evaluative assessment practices, such as those used to assign classroom grades, frequently do "...little more than show the pupils for whom the initial instruction was and was not appropriate" (Gusky, 2007, p. 65).

Rather than being strictly evaluative, assessments should serve as a signal that re-teaching must occur, or that concepts must be presented in a different way, or that the intensity of instruction must increase so that all students can meet learning expectations. Formative assessments should lead to second, third, or fourth chances for learning—they are a means by which teachers form partnerships with students to ensure that all students receive the instruction they require. As Stiggins (2007) has suggested, "Changing schools from places that merely sort pupils based on achievement into places that assure that all pupils will meet standards brings with it the challenge of rethinking the dynamics of assessment" (p. 11).

Not only does the effective use of formative assessment require such a rethinking, it will also require many new skills for teachers. Stiggins (2007) has noted that most teachers lack formal training in formative assessment, and Black and Wiliam (1998) have underlined the difficulties practicing teachers face in making the kinds of pervasive changes in assessment and teaching we have discussed:



Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles which leave entirely to them the task of translating them into everyday practice—their classroom lives are too busy and too fragile for this to be possible for all but an outstanding few. What they need is a variety of living examples of implementation, by teachers with whom they can both derive conviction and confidence that they can do better, and see concrete examples of what doing better means in practice.
(pp. 15–16)

In addition to reorienting to the purposes of assessment and acquiring new assessment and instructional skills, the use of formative assessment data will also require some school-level organizational changes. Not only does time need to be set aside for discussion, interpretation, and planning based on assessment information, but instructional resources need to be identified to respond to assessment data. It may be of little use to know why some students cannot comprehend grade-level material if resources are not available for appropriately focused and sufficiently intensive interventions to accelerate their growth toward grade-level standards.

One school district that has had significant success in teaching all students to read at grade level by third grade (Fielding, Kerr, & Rosier, 2007) attempts to increase instructional time in *proportion to how far struggling readers are behind* their peers. If the resources to provide such increased time have not been identified, or if scheduling inflexibility precludes extended instructional time in literacy, then knowing precisely how far a student is behind may be of little value. For example, school-level data showing that large numbers of students are struggling with specific 7th grade standards will only be useful if the school can offer teachers additional support or professional development to become more effective in addressing those standards.

Meeting Challenge #1

The most obvious response to this challenge is to provide ongoing professional development in classroom-based formative assessment that includes discussion of strategies for providing feedback and additional instruction when needed. States, districts, and schools should think of this professional development as extending over a relatively long period of time because it involves growth in both assessment and instructional skills (Shepard, 2000).

Chappuis, Stiggins, Arter, and Chappuis (2005) suggest ways that district-level policies and administrative practices can support more effective classroom-based formative assessment. They range from outlining clear achievement targets for each grade level and ensuring that intra-district assessments complement one another, to supporting effective data communications systems in the district. In the same book, the authors also describe a set of leadership competencies that must be in place to support the development of useful formative assessment practices in schools.

It may also be possible to accelerate the development and use of effective classroom-based assessments by state- or district-level work to produce “assessment exemplars.” These could, for example, take the form of discussion questions linked to content-area texts that reflect reading comprehension standards for a given grade. They could also include examples of assignments, with appropriate scoring rubrics, that reflect literacy competencies required at each grade, as well as extended response questions (with associated scoring rubrics) that could be used as formative assessment quizzes by classroom teachers. Langer’s (2001) study of high performing schools demonstrates that teachers, or teacher teams, are capable of deconstructing state-level literacy standards for classroom-based assessments. This process might be accelerated, or at least facilitated, if state- or district-level efforts also developed practical examples of classroom assessment strategies linked to the state’s literacy standards at each grade level. Part II describes several examples of assessments developed by programs or school districts that can help to serve this purpose.

Various organizations have also begun to provide professional development and support for developing classroom-based, instructionally embedded assessments. For example, the Berkeley Evaluation and Assessment Research Center (BEAR) at the University of California, Berkeley has a formal system, supported by test development software, that can be used to design classroom-based assessments (Kennedy, 2005; Kennedy & Wilson, 2007). The Council of Chief State School Officers has launched a state cooperative to develop materials in support of classroom-based formative assessment. The Council’s Formative Assessment for Students and Teachers (FAST) initiative, a project within the State Collaborative on Assessment and Student Standards (SCASS), is developing materials to support state-level initiatives to increase the use of classroom-based formative assessments in all subject areas. See



http://www.ccsso.org/projects/SCASS/Projects/Formative_Assessment_for_Students_and_Teachers/ for more information. There are also several for-profit organizations in the United States that provide training for districts or schools in developing more effective classroom-level formative assessments.

Given the complexity of simultaneously developing assessment and instructional capacity, it will be essential to follow a long-term development plan that includes a sequence of professional development and other supports and that leads eventually to a well-coordinated system of assessment for learning. For example, although benchmark or interim assessments lack many characteristics of effective classroom-based assessments, they may be a helpful place to begin the process of developing an effective formative assessment system. In Part II, Example #9 shows that good benchmark assessments can catalyze increased teacher-level understanding of state literacy standards. Data from these assessments can also focus collaborative work by teacher teams to develop improved instructional strategies based on the standards, and to monitor the effectiveness of the strategies that are developed. Experience in discussing data from benchmark assessments and planning additional instruction in response to these data might help teachers more easily develop their own, instructionally-embedded assessments to monitor student progress and provide more details about next instructional steps. This is much more likely to happen with support and encouragement at the policy level in schools and districts.

Challenge #2: Tension between types of formative assessments useful to administrators and types useful to teachers and students

The large-scale investment in benchmark assessments at least partially reflects the needs of school- and district-level administrators for more frequent data about student progress toward grade-level standards. Data from the Rand study (Marsh et al., 2006) indicate that superintendents and principals consider these data valuable for the kinds of decisions they need to make at the district and school level. Although there is a paucity of research to demonstrate the impact of these benchmark assessments on student outcomes, there is apparently wide agreement among educational leaders about their perceived usefulness. Whether this perceived usefulness is associated with real improvements in

student performance has not yet been demonstrated in rigorous research (Henderson, Petrosino, Guckenburg, & Hamilton, 2007). The key to improved use of these assessments will involve moving beyond what Perie and others (Perie, Marion, & Gong, 2007) have referred to as the “zone of wishful thinking” that is often associated with the implementation of benchmark assessments. “Policymakers often hope that data will automatically lead to improved practice. However, experience shows that data must be accompanied by the reporting systems, professional development, support structures, and management practices that will impact teacher and student beliefs and behaviors” (p. 21).

From a practical standpoint, it does appear that data from these kinds of assessments could be useful for a variety of purposes in a comprehensive assessment plan. For example, they could be used to identify students in need of more intensive instruction, to stimulate discussions among teachers about instructional strategies related to standards, or to monitor the impact of instructional innovations. However, districts or schools whose investment in these types of assessments precludes support for effective classroom-based formative assessments are likely to fall short of their ultimate instructional goals. Given that the preponderance of research support for formative assessments to improve student performance comes from studies of classroom-based assessments, long-term development plans that do not include increasing teacher skills in classroom-based assessments seem short-sighted.

Meeting challenge #2

The most direct means of meeting this challenge is for educational leaders to become educated about the differences between benchmark (interim) and classroom-based assessments. This would include developing an understanding of the limitations and uses of both types of assessment as well as learning how they might complement one another in an overall assessment system. It is particularly critical for district- and school-level leaders to receive professional development in this area, since they are the ones likely to make decisions about long-term assessment plans, including investments in professional development for teachers. If they choose to invest only in benchmark assessments, without understanding their limitations for guiding daily classroom instruction, they are unlikely to obtain the significant changes to classroom instructional practices that are most directly supported by short-cycle classroom-based formative assessment.



Challenge #3: Lack of teacher “buy in” and divergence of educators’ attitudes and beliefs about indicators of student success

A key issue in developing an effective comprehensive assessment plan is reaching substantial agreement within a school about important indicators of student success. If the principal places primary value on one set of outcomes and teachers on another, it will be difficult to align assessments of student progress at the classroom and school level. In a study of nine high schools that had all been working for at least four years to implement a quality improvement model of school change that emphasized the systematic use of data to guide instruction, Ingram, Louis, and Schroeder (2004) found wide divergence of opinion among teachers and principals about the most important indicators of student success. For example, teachers placed more trust in locally developed measures, while principals depended more on standardized tests. Asked about how they evaluated student success, teachers were more likely to cite such outcomes as student behavior and affect, student feedback on their courses, and college success rather than student performance on standardized tests. Standardized test data were cited much more frequently in evaluating school success than teacher success, but about half of interview respondents did not mention any achievement indicators at all. The authors commented, “Teachers are articulate about the lack of congruence between the goal assumptions of their state’s accountability legislation and their own and their community’s expectations for a much less well-defined product of ‘effective adults’” (p. 1279).

Meeting Challenge #3

These findings about teacher attitudes suggest that this factor must be addressed in developing any assessment system to improve student performance on grade-level standards for literacy. Teachers must “buy in” to the idea that the state standards represent important and valid literacy outcomes so they can have confidence that their instruction is focused on “real learning” and is not simply preparation for taking an end-of-year accountability test. Although research shows that teachers do give greater attention to areas of performance assessed by state-level tests, the extent to which this change in focus reflects change in real instructional values is unclear. Thus, part of the

preparation for improving both assessment and instructional practices in literacy is to work toward consensus on valued instructional goals. This may be particularly challenging in the area of literacy, because it is critical that teachers whose primary focus has been teaching specific content in science, social studies, English, or history also “buy in” to general literacy standards related to their disciplines. Unless science teachers accept the idea that part of their responsibility is to help students make inferences from texts, or combine information across textual sources to form conclusions, they will not engage in formative assessments to guide their instruction on these skills.

Given the complexity of changing attitudes and beliefs, it is likely that a variety of approaches to aligning instructional and assessment goals at the classroom and school level will be needed. A sound beginning step would be to support sufficient study and discussion of each grade level’s literacy standards so that all teachers understand the important literacy skills they embody. Professional development and cooperative activities designed to help teachers turn their understanding of the standards into classroom-based formative assessments might be helpful because they illustrate how the standards translate into meaningful, everyday literacy competencies in the classroom and across disciplines. In addition, as teachers receive professional development on instructional strategies related to desired literacy outcomes, they will see the link between instruction and improved student competence. They should not only grow in their understanding of the standards (their theory of the task will improve), but they should also gain a better appreciation of the links between instruction and student outcomes.



CONCLUDING COMMENTS

We have outlined and discussed the major types of assessments that could contribute useful information to guide instruction in adolescent literacy. These assessments are a *vital* part of a comprehensive plan for improving literacy outcomes, because they help intervention teachers, classroom teachers, principals, and district staff respond more effectively to the wide diversity of literacy and literacy learning abilities among students. These seven key observations summarize our most important points:

1. Assessments can guide instruction in a number of ways at the classroom, school, district, and state levels. Different types of assessments are required to meet these different purposes.
2. Classroom-based, instructionally embedded assessments currently have the strongest research support as a means for improving student learning outcomes. Assessment initiatives that do not support the improvement of this type of assessment are likely to fall short of their long-term goals of helping all students become proficient readers.
3. Standards-based benchmark tests, given monthly or quarterly, are increasingly used to monitor student progress toward literacy standards assessed on year-end accountability measures. Although educational leaders report that these tests are useful for instructional decision-making, currently no rigorous research supports the premise that they have a positive impact on student literacy growth. Schools and districts need to monitor the value of their investment in this type of assessment as they advance their literacy plans.
4. Curriculum-based measures that have been widely used to guide instruction in early literacy and special education are a special case of classroom-based formative assessment. The kind of brief, *general outcome* measures that have been effectively used to monitor growth in basic reading skills have little utility for monitoring growth in grade-level literacy skills for middle and high schools students because they do not adequately assess complex comprehension processes. However, *mastery oriented* CBM may be very helpful in guiding instruction in comprehension skills, and brief general outcome measures may be very useful for

monitoring the effectiveness of instruction for students who are still struggling to acquire more basic reading skills.

5. Two types of screening information seem necessary to identify the type of intervention support needed by students who are reading substantially below grade level. *General screening information* (from the previous year's summative test or from a benchmark test at the beginning of the year) can be used to identify students whose reading ability is so weak they are unlikely to profit from grade-level literacy instruction from content-area teachers. Among these at-risk students, *targeted screening tests* of word-level skills, vocabulary, or lower-level comprehension can be used to identify students for specific intervention placements within a school.
6. In the context of a comprehensive assessment plan, *formal diagnostic instruments* should be used sparingly and strategically. *Informal diagnostic assessments*, which are sometimes very similar to classroom-based formative assessments, may be used more frequently by intervention teachers to establish an instructional plan. In addition, program-specific placement and progress tests can help guide instruction for struggling readers when a published intervention program is used to guide instruction.
7. Effectively using assessment for instruction to improve student literacy outcomes involves a great deal more than adding different types of assessments to the school program. Assessment without the capability and will to respond to the information it generates is essentially a waste of time. Taking full advantage of the information generated by a variety of assessments will require important changes in both attitude and capability at every level in the educational hierarchy, from classroom teachers and teacher preparation programs to state-level policy-makers.

Although there are many obstacles to overcome in the creation of an effective comprehensive assessment plan to guide instruction in adolescent literacy, it is an essential key to meeting our goal to leave no student behind. Effective assessments provide the guidance to help each student stay “on course” toward important literacy goals. When they are followed by appropriate instructional adjustments, formative assessments can provide a powerful tool to improve learning for all students.



RECOMMENDATIONS FOR FURTHER READING

In preparing this document, several publications were immensely helpful. For those wishing to deepen their knowledge in this area, we recommend the following as a place to start:

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.

Black, P., & William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principals, Policy and Practice*, 5, 7–74.

Chappuis, S., Stiggins, R., Arter, J., & Chappuis, J. (2005). *Assessment FOR learning: An action guide for school leaders* (2nd ed.). Portland, OR: Assessment Training Institute.

McMillan, J. H. (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College Press.

Perie, M., Marion, S., & Gong, B. (2007). *A framework for considering interim assessments*. Dover, NH: National Center for the Improvement of Educational Assessment.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief*. Washington, DC: The Aspen Institute.

REFERENCES

- Abrams, L. M. (2007). Implications of high stakes testing for the use of formative classroom assessments. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice*. (pp. 79–98). New York: Teachers College Press.
- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- Alvermann, D. E. (2001). *Effective literacy instruction for adolescents: Executive summary and paper commissioned by the National Reading Conference*. Chicago: National Reading Conference.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). White Plains, NY: Longman.
- Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, 40, 685–730.
- Berman, I., & Biancarosa, G. (2005). *Reading to achieve: A governor's guide to adolescent literacy*. Washington, DC: National Governors Association Center for Best Practices.
- Biancarosa, G., & Snow, C. (2006). *Reading next: A vision for action and research in middle and high school literacy: A report to the Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment: Principals, Policy and Practice*, 5, 7–74.
- Black, P. & Wiliam, D. (2007). A theoretical foundation for formative assessment. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice*. (pp. 29–42). New York: Teachers College Press.



- Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43–62). New York: Teachers College Press.
- Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis, 24*(3), 219–239.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*, 1–14.
- Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Fort Worth, TX: Harcourt Brace.
- Chappuis, S., Stiggins, R., Arter, J., & Chappuis, J. (2005). *Assessment FOR learning: An action guide for school leaders* (2nd ed.). Portland, OR: Assessment Training Institute.
- Cizek, G. J. (2007). Formative classroom and large-scale assessment: Implications for future research and development. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 99–115). New York: Teachers College Press.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438–481.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*, 8–15.
- Deno, S. L. (2003). Developments in curriculum based assessment. *Journal of Special Education, 37*, 184–192.
- Fielding, L., Kerr, N., & Rosier, P. (2007). *Annual growth for all students, catch-up growth for those who are behind*. Kennewick, WA: The New Foundation Press.
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449–460.

-
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199–208.
- Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education, 77*, 64–84.
- Fuchs, L. S., & Fuchs, D. (2003). Can diagnostic assessment enhance general educators' instructional differentiation and student learning? In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 325–351). Baltimore: York Press.
- Good, R. H., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257–288.
- Good, R. H., Simmons, D., Kame'enui, E., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene: University of Oregon.
- Gusky, T. R. (2007). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and practice. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 63–78). New York: Teachers College Press
- Guthrie, J. T., Wigfield, A., and Perencevich, K. C. (2004). *Motivating reading comprehension: Concept-oriented reading instruction*. Mahwah, NJ: Lawrence Erlbaum.
- Haynes, M. (2007). *From state policy to classroom practice: Improving literacy instruction for all students*. Washington, DC: National Association of State Boards of Education.
- Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007–No. 039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved January 7, 2008 from <http://ies.ed.gov/ncee/edlabs>



- Hirsch, E. D. (2006). *The knowledge deficit: Closing the shocking education gap*. New York: Houghton Mifflin.
- Hock, M. F., Brasseur, I. F., Deshler, D. D., Catts, H. W., Marques, J., Mark, C. A., & Wu Stribling, J. (in press). What is the nature of struggling adolescent readers in urban high schools? *Learning Disability Quarterly*.
- Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*, 1258–1287.
- Kennedy, C. A. (2005). *The BEAR assessment system: A brief summary for the classroom context*. Berkeley, CA: Berkeley Evaluation & Assessment Research Center.
- Kennedy, C. A., & Wilson, M. (2007). *Using progress variables to map intellectual development*. Berkeley, CA: Berkeley Evaluation & Assessment Research Center.
- Kluger, A. N., & DeNisi, A. (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.
- Kosanovich, M., Ladinsky, K., Nelson, L., & Torgesen, J. K. (2006). *Differentiated reading instruction: Small group lesson structures for all students*. Florida Center for Reading Research. Retrieved July 12, 2007 from www.fcrr.org
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research, 67*(3), 271–299.
- Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities, 35*(5), 436–447.
- Langer, J. A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal, 38*, 837–880.
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211–224.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: Rand Corporation.

-
- Marshall, B. (2007). Formative classroom assessment in English, the humanities, and social sciences. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 136–152). New York: Teachers College Press.
- Masters, G., & Forster, M. (1996). *Progress maps*. (Part of the Assessment Resource Kit). Melbourne: The Australian Council for Educational Research.
- McEwan, E. K. (2007). *Raising reading achievement in middle and high schools: Five simple-to-follow strategies* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- McMillan, J. H. (2007). *Formative classroom assessment: The key to improving student achievement*. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 1–7). New York: Teachers College Press.
- Moje, E. B. (2006). Motivating texts, motivating contexts, motivating adolescents: An examination of the role of motivation in adolescent literacy practices and development. *Perspectives*, 32, 10–14.
- National Association of State Boards of Education. (2006). *Reading at risk: How states can respond to the crisis in adolescent literacy*. Washington, DC: Author.
- National Assessment Governing Board. (2007). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: American Institutes for Research.
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read. Report of the subgroups*. Washington, DC: National Institute of Child Health and Human Development.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155–175.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36(2), 89–101.
- Perie, M., Marion, S., & Gong, B. (2007). A framework for considering interim assessments. Dover, NH: National Center for the Improvement of Educational Assessment.
- Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief*. Washington, DC: The Aspen Institute.



- Phillips, M. (2005). *Creating a culture of literacy: A guide for middle and high school principals*. Reston, VA: National Association of Secondary School Principals.
- Popham, J. (2007). The lowdown on learning progressions. *Educational Leadership, 64*, 83–84.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 545–561). Mahwah, NJ: Lawrence Erlbaum.
- Rasinski, T. V. (2003). *The Fluent Reader*. New York: Scholastic.
- Rea-Dickens, P. (2001). Mirror, mirror on the wall: Identifying the processes of classroom assessment. *Language Testing, 18*, 429–462.
- Schatschneider, C., Buck, J., Torgesen, J. K., Wagner, R. K., Hassler, L., Hecht, S., & Powell-Smith, K. (2004). *A multivariate study of factors that contribute to individual differences in performance on the Florida Comprehensive Reading Assessment Test*. Technical Report No. 5. Tallahassee, FL: Florida Center for Reading Research.
- Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly, 19*(2), 159–172.
- Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York: Knopf.
- Shepard, L. (2000). *The role of classroom assessment in teaching and learning*. Technical Report No. 517. Boulder, CO: National Center for Research on Evaluation, Standards, and Student Testing.
- Shute, V. J. (2007). *Focus on formative feedback*. ETS Research Report. Retrieved from <http://www.ets.org/research/contact.html>.
- Stiggins, R. J. (2007). Conquering the formative assessment frontier. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 8–28). New York: Teachers College Press.
- Torgesen, J. K. (2002). Empirical and theoretical support for direct diagnosis of learning disabilities by assessment of intrinsic processing weaknesses. In R. Bradley, L. Danielson, & D. Hallahan, (Eds.), *Identification of learning disabilities: Research to practice* (pp. 565–613). Mahwah, NJ: Lawrence Erlbaum.

-
- Torgesen, J. K., Houston, D., & Rissman, L. (2007). *Improving literacy instruction in middle and high schools: A guide for principals*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., Wexler, J., Francis, D. J., & Rivera, M. O. (2007). *Academic literacy instruction for adolescents: A guidance document from the Center on Instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*, 85–120.
- Wigfield, A. & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420–432.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on pupil achievement. *Assessment in Education, 11*, 49–65.
- Williamson, G. L. (2004). *Student readiness for postsecondary options*. Metametrics Technical Report. Retrieved October 2005 from [http://www.lexile.com/lexilearticles/Student%20Readiness%20for%20Postsecondary%20Options%20\(v4_1\).pdf](http://www.lexile.com/lexilearticles/Student%20Readiness%20for%20Postsecondary%20Options%20(v4_1).pdf)
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*, 181–208.
- Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26*, 123–141.



PART II—TEN EXAMPLES OF ASSESSMENTS, OR ASSESSMENT SYSTEMS, IN CURRENT USE OR UNDER DEVELOPMENT TO GUIDE INSTRUCTION IN ADOLESCENT LITERACY

In Part II, we offer 10 examples of assessments to guide instruction in adolescent literacy that are in current use or under development. They were not selected because empirical research indicates that they have demonstrated a positive impact on student achievement. To our knowledge, other than the research on assessment types described and reviewed in Part I, research evidence of the impact of specific instantiations of the types of assessments is not yet available. In any case, because these assessments are all embedded in complex educational contexts, it would be difficult to identify their separate impacts on student achievement.

This set of examples is not exhaustive. These assessments and assessment systems were selected primarily because they exemplify a range of current approaches to assessments of adolescents' literacy learning. They are not meant as an endorsement of either the procedures themselves or of the companies or organizations using them. Because we do not want to appear to endorse specific organizations or companies, we have not identified by name the organizations associated with the examples, but rather we have simply identified them by number. These examples were selected through an Internet search and by polling adolescent literacy experts around the country. After a brief description of each example, we comment on its characteristics in terms of the types of assessments discussed in Part I.

Example #1

The assessments discussed in this example have been developed as part of a project supported by a Striving Reader's grant funded through the U.S. Department of Education. The assessments were developed to assess progress within an instructional program based on a scope and sequence of reading strategies designed for content-area classrooms in middle and high schools. The schools contracted with a research organization that specializes in developing instructionally embedded classroom-based assessments across a variety of curricular topics. The method used to develop the assessments involves four components or steps: 1) identifying progress variables; 2) designing test items; 3) determining outcome or response space; and 4) creating a measurement model. The assessments are designed to allow teachers to make meaningful interpretations of their students' work based on the progression of their content literacy skills.

Assessment development begins with defining the knowledge and skills that form the core learning goals of the strategy-based literacy curriculum. Known as "progress variables," they represent a developmental progression of skills taught in the curriculum. This step is followed by "items design," the construction of assessment tasks to gather evidence about student knowledge and movement on the progress variables. Items design is grounded in the principle that assessments should be embedded in normal classroom activity and based on authentic instructional tasks. These tasks may use a variety of formats, ranging from activities incorporated into classroom instruction to full-length assessments. Student responses to assessment tasks are used to develop the "outcome space," a rubric that describes more or less proficient applications of skills and knowledge identified in the progress variables.

The rubric is developed jointly by project staff and school literacy coaches and teachers and is extensively modified as teachers use it to analyze student responses in pilot assessments. The process is facilitated by project staff and, while initially time-consuming, allows teachers to develop a deep understanding of the expectations for students within the curriculum. The result is an item-specific scoring guide that gives teachers information about how students are progressing on specific content in the curriculum. The final step in developing the assessment is applying a measurement model to calibrate the assessment items to the progress variables in order to produce reports that quantify



student and class progress on the rubrics designed for progress variables. These reports allow the teacher to track progress along the curriculum's developmental sequence and should also be useful to school leaders for tracking student-, classroom-, and school-level progress because they are common across all classrooms using the same curriculum.

Commentary

The assessments being developed and used in this project represent an attempt to produce a set of classroom-based formative assessments following rigorous design and calibration procedures. The assessments are embedded within the scope and sequence of a well-defined curriculum and are common across classrooms, which would also allow data to be aggregated at the classroom and school level.

The progress variables and rubrics are designed to be “open” so that teachers can share them with students as a teaching tool. Because this is a set of common assessments, computer technology can provide reports to help teachers analyze data at both the student and classroom levels. Individual student reports indicate where each student falls on the overall continuum and reflect growth over time, as well as how the student performs on specific items. Class reports offer a visual display of where all students are performing along the literacy continuum and how they have performed on specific skills. This feedback allows the teacher to make adjustments for individuals or small groups of students who need additional support (e.g., re-teaching a skill or providing additional practice). It also allows the teacher to identify which students are ready to progress to more sophisticated levels on the continuum.

Professional development is an important component of this assessment's utility. Teachers receive expert problem-solving support as they develop their skills in implementing the curriculum and using student data to guide instruction. The nature of the assessment requires teachers to understand the development of skills in the curriculum as well as the characteristics and range of student responses at each level of skill.

Example #2:

This example involves a set of assessments developed by a large metropolitan school district to assess specific aspects of reading comprehension to help teachers in grades 2 through 12 incorporate data about student progress as they deliver literacy instruction. The assessments are intended to promote differentiated instruction based on student needs and to support a collaborative culture in which educators learn to use data to improve student learning.

The assessment system comprises both assessments and extensive support for using the data they generate in guiding instruction. The core of the system is a set of more than 50 passages (six or more at each grade level) written in several different genres. Each has 10 multiple choice questions to assess students' abilities to both find evidence in text and draw various kinds of inferences. Teacher guides help instructors analyze the intellectual demands of the questions and understand some reasons students may choose incorrect answers. Incorrect answers are categorized as:

- Near miss answers—true statements from the text that don't really answer the question. Students who make a lot of these kinds of mistakes may comprehend the text but may not be reading it closely enough or referring to it carefully in answering the question.
- Mis-read answers—answers that are based in the text, but reflect a significant misunderstanding of it or of the question. Students who make these kinds of errors may be reading the questions too quickly or not monitoring their understanding while they read.
- Out of bounds—answers that are not based in the text at all, but may reflect a personal experience, or a "common sense" approach to the question. This kind of answer reflects overreliance on personal experience and failure to read the text closely.

Teachers can have students respond to the multiple choice questions on machine-scored sheets so that within a week they can receive reports that identify types of answers for individual students as well as class summaries. They are encouraged to use information about individual student responses or response patterns as the basis for follow-up discussions and feedback both individually and in groups.



The system is designed to be very flexible; teachers may choose when to use which passages for assessment. For example, a teacher might select a high-interest passage to assess the effect of students' motivation on their comprehension or select another passage with a high proportion of questions requiring a certain type of inference recently discussed in class. The passages vary in complexity within grade level and in the proportion of the different kinds of questions asked. Teachers also have several options for administering the assessments in class: They can ask students to read and annotate the text before receiving the questions, or require students to give reasons for the answers they choose. The goal is to provide highly adaptable assessments that can help teachers answer specific questions about the performance of individual students or the class as a whole.

The testing material offers teachers significant assistance in deepening their understanding of the thinking processes and skills required to answer each type of question correctly. For example, two types of questions require students to find evidence in the text; one type requires students to identify evidence explicitly stated in one location in the text, and the other requires them to gather evidence from several locations. Furthermore, the types of inferences different questions require are broken down into five categories, and the types of thinking required to make each inference are explained in the teacher material. The teacher guide for each passage contains a graphic organizer to help teachers remember the type of thinking for each question.

The training encourages teachers to think of the specific assessments in this system as among many types of assessments they can use to understand their students' learning patterns and needs. For example, in one training exercise, teachers are encouraged to use other data to develop questions about their students' current skills that might be answered with one of the assessments provided by the district. The ongoing teacher support also includes "data coaches" who are responsible for helping teachers learn how to interpret the data from the assessments and develop appropriate instructional responses from them. The system is actually designed to be used as part of a team effort, with the assessment data coaches complementing the work of instructional leadership and reading leadership teams in collaborating to formulate more effective instruction in response to the data.

The district contracted with an outside evaluator for feedback on the continued development of the system, and the most recent evaluation report

found teachers, principals, and literacy coaches very enthusiastic about both the assessments and the way they are being supported by the district. The report indicated that educators had come to see the assessments as a valuable set of tools that are beginning to be part of instructional conversations that extend well beyond the formal data meetings supported by the district. Students interviewed felt that their experience with the assessments helped their test-taking and reading strategies and improved their confidence about taking the year-end “high-stakes” accountability test in reading. The evaluators also identified a number of implementation challenges, perhaps the most important being the problem of linking the assessments more closely to the curriculum.

Commentary

A clear strength of this assessment system is its extensive professional development and teacher support. Part of the professional development covers using the assessments to make real-time adjustments in instructional strategies, and part covers helping the teachers better understand the thought processes and reading strategies the assessments measure. There is a clear focus on differentiating instruction based on the assessment data.

The assessments in this system fall somewhere between short-cycle, classroom-based formative assessments and benchmark, or interim assessments. They are designed to be administered periodically, not daily, and are not embedded within ongoing instruction. However, teachers are allowed to administer them whenever they desire (they are not necessarily all administered at the same time within a school) and can customize their form in a number of important ways. The materials are designed to help teachers understand more about the thought processes or reading strategies that students use in responding to questions about text, so that instruction can be targeted to improve performance. The test materials, the teacher guides, and the professional development should all help teachers become more skilled at analyzing student performance on literacy tasks so that the next instructional steps can be targeted with greater accuracy. Thus, these tests and the support provided for them might be regarded as one of the professional development steps required for more effective use of classroom-based formative assessments to guide instruction.



Example #3

The assessments described in this example were developed to support an explicit and systematic reading instruction/intervention program for students in middle school. The preferred plan for implementation of the program provides reading instruction to all students in the school at the same hour of the day, with students participating in instructional groups matched on reading skill. Strong readers are in larger advanced classes with instruction focused on metacognitive strategy use, recognizing text structures, etc., while the weakest readers are in smaller classes focusing specifically on their needs, which often include instruction in basic reading skills. When the program is used only as an intervention for struggling readers, and more than one class is involved, all intervention classes are also scheduled during the same period. With all reading classes scheduled simultaneously, students may move from group to group as their performance warrants.

Beginning of the year screening

The program typically relies on whatever assessment system the school has in place to determine students' instructional needs at the beginning of the year. This is often the previous year's standards-based outcome test, along with teacher recommendations. If students do not have an outcome test from the previous year, schools are asked to give a standardized test of reading comprehension. The program authors regard initial screening and placement as important to help the school get started, but feel that student misplacement is not an irreversible error. A teacher can quickly determine if a student is working at the right level and move him or her to a more appropriate group if needed. For very poor readers, teachers typically follow up the screening with individually administered tests of word-level reading to assess phonics skills and conduct oral reading fluency checks.

Progress Monitoring

The core of the progress monitoring system is a small booklet, resembling a teacher grade book, for marking progress, with different record booklets for different levels of the curriculum. The booklet contains curriculum goals that students are expected to meet as they move through the instructional materials at their level. This is a formal way to monitor growth on curriculum activities

and set performance targets; the booklets can also be used for curriculum-specific progress monitoring and as a record of how many students have mastered a specific skill.

The progress monitoring booklets focus on the major elements of reading growth and are tied to the lesson structure in the curriculum, but they can also be customized for individual students. For example, a teacher can monitor the growth of a specific skill such as the ability to generate appropriate questions, or the ability to make reasonable predictions about content. Teachers might also use the booklets to monitor fluency. Each student uses a record-keeping tool called a learning guide, a simplified but analogous version of the teacher's record booklet. Student teams also keep records of weekly team growth or improvement and are recognized both for improvement and for behaviors that facilitate it, such as helping and encouraging one another, making sure everyone participates, and active listening. Thus, students keep a continuous record of their short- and long-term goals and see their progress in specific areas; an important aspect of the system is that the instructional goals, and progress toward meeting them, are transparent to the student.

Periodically, the coach and assessment team review the assessment record booklets to see if a student is able to, or needs to, change instructional groups. Along with the marks and notes in the booklet, a student's performance on periodic administrations of a standardized measure of reading comprehension is also considered.

Teacher grade-level teams also meet weekly with their literacy coach to discuss what is working and what needs to be adjusted based on their students' progress. Thus, the progress monitoring assessments inform not only periodic decisions about moving students among instructional groups, but also provide ongoing information about the effectiveness of instruction, and lead to adjustments of instruction.

To summarize, the assessment questions in this system are of two kinds: 1) How is the student progressing in terms of movement through the highly specified curriculum? and 2) How is the student progressing on his or her individually specified reading goals or target areas? This type of assessment guides instruction in at least three ways: 1) It informs the teacher and/or the assessment group about whether the student is in the right instructional group; 2) It helps the teacher monitor the group's progress as a whole to identify needs for re-teaching or additional practice; and 3) It helps the teacher monitor



individual students' progress toward individual goals on important reading subskills. The system actively works to inform students about their own progress toward their goals, and they are encouraged to monitor that progress.

Teachers may initially find this level of progress monitoring overwhelming. What makes it manageable for most teachers is that the assessments are both tied to and defined by a tightly organized curriculum. Professional developers who provide support to these teachers note that if teachers had to develop their own curricula, they would then have to develop their own assessments, which would be particularly difficult.

Diagnostic Testing

Teacher observations and notes in the progress monitoring books provide a substantial amount of diagnostic information. However, if a teacher wants more information, typical district procedures for additional testing would be followed. Referral rates to special education can actually go up because teachers are much more aware of student progress than in more typical instructional settings. In addition to the systematic assessments described above, the cooperative learning model and lesson structure allow the teacher time each day to circulate and listen to students read and discuss their work. If a referral produces a diagnosis, the Individual Education Plan (IEP) determines the instructional response.

Commentary

This program's frequent progress assessments can be classified as classroom-based formative assessments or as mastery-oriented curriculum-based assessment. The assessments are tied specifically to the knowledge and skills being taught but involve more than simply monitoring movement through the lessons. Teachers can identify authentic reading goals appropriate to each student's level of performance and monitor progress in acquiring those skills over time. The assessments are of sufficient detail and frequency to provide rich information about next steps for instruction.

Another important feature of the system is that teachers make observations not only about the development of reading skills but also about student attitudes toward reading and engagement in the instructional tasks. Discussions of these observations are used to help make adjustments that improve student

growth in these critical social and behavioral areas. All of the assessment information is shared with students transparently so that they can actively involve themselves in monitoring their progress toward their literacy goals.



Example #4

The assessments described in this example are part of a set of instructional modules designed to enhance learning, reading, and writing skills in adolescents. Although the initial focus of the modules was on improving school performance of students with learning disabilities, the research and development effort has evolved to encompass learning support for all students. The model includes methods that can be used by all teachers (specifically content-area teachers in middle and high school) to help students learn content more effectively, as well as a set of modules focused on strategies that can be taught directly to students to help them learn more effectively. Some of these strategies, such as those focusing on visual imagery, self-questioning, paraphrasing, and inference-making, are specifically designed to enhance reading comprehension. Assessments to guide instruction in these strategies are included in the support materials for each strategy. Within each instructional module, assessments are tailored to assess student progress in learning to use the various components of the strategy.

For example, the module to teach the inference-making strategy is designed to help students improve their ability to make inferences from text, a skill assessed on most grade-level tests of reading comprehension in late elementary, middle, and high school. In learning to make accurate inferences from text, students are taught a number of steps: 1) identifying the type of question that is being asked (e.g., one that has an answer in the text, or one that requires their own knowledge of the topic); 2) identifying “clue words” that indicate the kind of thinking required to answer the question; and 3) exploring supporting details. Teacher materials include a pre-test to help teachers identify students who may have difficulty using various elements of the strategy as well as a post-test for evaluating student improvement in using the elements and drawing inferences. There are progress charts that can be used to monitor progress over time both in accuracy of inferences and skill in using the strategy steps. An inference answer sheet for use with the progress charts includes space for recording the type of question, clue words used in answering it, and the answer itself. In this way, the teacher can monitor not only student progress in the instructional target (the ability to make correct inferences) but also in student use of specific aspects of the inference strategy. Presumably, this latter information helps identify components that merit additional instruction or practice for some students.

Another set of lessons was developed to help students learn to identify the main idea and supporting details in text. Many struggling readers have trouble with these prerequisite skills when they are learning to use paraphrasing strategies to improve comprehension. Over a series of 14 lessons, students learn what it means to paraphrase and eventually become able to restate ideas in their own words and identify details in relatively complex passages. As they work with learning sheets to practice the skills in each lesson, they receive feedback on their use of elements of the overall strategy. The instructor's manual emphasizes the importance of individual feedback after each lesson in order for a student to correct mistakes and receive encouragement for continued progress. The teacher's manual also discusses four characteristics of effective feedback: It should be 1) positive, 2) corrective (descriptive), 3) provided individually, and given as soon as possible after the student performance, with reminders immediately before the next performance. The manual also contains scoring rubrics for all of the practice exercises, and a progress chart for setting student goals.

Commentary

The assessments in this set of instructional modules are a good example of classroom-based formative assessments within well-designed curriculum materials. These assessments focus on learning strategies and behaviors that are directly taught in the curriculum, with sufficient detail in the assessments to provide useful information about next instructional steps. The materials also allow students to chart their own progress so that they can see how their skills are improving through instruction and practice. Teachers who are trained to use the assessment components provided with the instructional modules have the opportunity to learn many of the important components of classroom-based formative assessment that might help them apply this type of assessment more broadly in their instruction.



Example #5

Various aspects of this assessment model for elementary, middle, and high schools are currently being implemented at school sites around the country. The model includes a set of test development and assessment practices that can be applied to monitor student progress toward achievement standards in any subject area.

This model starts with the assumption that the goal of any literacy or content-area instruction is to help students meet state standards, which increase in difficulty each year students are in school. It has four levels of assessment, each carefully aligned to state standards.

Level 1: Classroom-based assessments implemented in a variety of ways

The assessment format mentioned most often at this level involves performance tasks accompanied by rubrics or scoring guides. These performance tasks provide an opportunity for students to apply the concepts and skills they are learning and to produce a product that can be evaluated with a rubric or scoring guide. As students produce, receive feedback on, and revise their work, these tasks provide rich formative information about students' emerging knowledge and skill related to specific performance targets derived from state grade-level standards. Each teacher develops these classroom-based assessments according to his or her preferences and instructional plan, but they are informed by collaborative work with other teachers to develop the next higher level of assessments in the system. These are referred to as common formative assessments.

Level 2: School-based common formative assessments administered at the beginning and end of instructional cycles

Developed collaboratively by grade-level and subject-area teams, these assessments are based on the idea that all teachers are working toward common instructional goals defined by state-level standards. Their recommended frequency is from once a month to six times a year, but they are planned around important cycles of instruction at each grade level. This level of assessment is unique to this particular comprehensive assessment model. Two test forms of each assessment are developed, to be administered at the beginning and end of the unit of instruction. Developing these assessments involves several steps:

-
1. *Identifying critical standards to be assessed.* Teachers identify the most important standards pertaining to a given unit of study or instructional cycle. The authors of this model suggest that if time constraints make it impractical to develop common formative assessments for all grade-level standards, teachers should work toward identifying “power standards” and focus assessment on those key standards they feel are the most vital in any given unit of instruction.
 2. *“Unwrapping,” or deconstructing, the power standards to be addressed during the instructional cycle.* Unwrapping involves identifying instructional targets, described in terms of tasks students should be able to do (e.g., select, compare, identify, contrast) with specific kinds of knowledge or content. Done collaboratively in subject-area or grade-level teams, this helps ensure that instruction and assessment are consistent within a school. This activity has an important side effect in helping teachers to understand more completely what the standards actually require students to learn.
 3. *Developing the test items to assess student mastery of performance targets derived from the unwrapped standards.* The items should address each skill or type of knowledge identified, and should match the rigor and kind of performance the standard requires.
 4. *Scoring common formative assessments.* This can be done in a variety of ways, but, particularly if the assessments involve constructed responses scored by rubrics, the authors of this assessment model recommend that teachers meet to score responses collaboratively. The authors’ experience suggests that teachers learn an enormous amount from this collaborative scoring process, which helps them to think about their curriculum and design more powerful and targeted instructional strategies.
 5. *Collaborative analysis of results from common formative assessments.* The meaning of the formative assessment outcomes is discussed and instructional or action plans are devised. Pre-test results are used to help focus instruction where it may be needed most, or to identify students who may need special support to meet the standards taught during that unit. Post-test results provide information about the overall success of the instructional plan and identify standards that may need to be reinforced in



subsequent units. In this model, teachers' work in data analysis teams is a critical part of the assessment process and leads to productive use of results to guide instruction across classrooms.

Level 3: District level benchmark assessments

These might be administered quarterly in order to provide information about particular students and their progress toward meeting the end-of-year standards. Instructional support personnel (principals, curriculum leaders, etc.) can use these results in allocating instructional resources and making school-level decisions about instructional focus.

Level 4: Annual state-level assessments

These formal assessments are administered annually to determine the percentage of students who can meet state standards.

Commentary

A powerful feature of this assessment model, if implemented correctly, is its help focusing instruction on important grade-level performance standards. It can provide specific information to differentiate instruction, and information about how students are likely to perform at each successive level of assessment. The authors correctly point out that when assessments are aligned at all levels (classroom, grade level, department, school, district, and state) the information provided by each assessment will be more likely to help guide and focus instruction so that student performance is improved at successively higher levels.

Most of the detailed information about this model focuses on the level of "common formative assessments," which fall somewhere between short-cycle, classroom-based assessments and benchmark, or interim assessments. The model's developers suggest that work on these common formative assessments could be an effective professional development tool to help teachers develop the skills required to create their own classroom-based, instructionally embedded assessments. This seems a reasonable possibility as long as teachers are encouraged to move to that level of assessment and not be content simply with the common formative assessments administered every six weeks.

From the perspective of literacy instruction, this model is missing assessments to determine instructional needs for students with serious reading difficulties. For example, an 8th grade student may fail to meet grade-level standards on a district benchmark assessment, a common formative assessment, or even a classroom performance assessment, for a variety of reasons, including difficulty with basic word-level reading skills. Because state literacy standards in middle and high school focus primarily on dimensions of reading comprehension, standards-based assessments are not likely to provide adequate descriptions of word-level reading difficulties such as problems with decoding multi-syllable words, or dysfluent reading. As noted in Part I, students who have serious problems at the word level are likely to require more intensive instruction across a broader range of skills than students who primarily struggle to meet the reading comprehension standards for their grade level.



Example #6

This assessment system is currently under development in one state and includes reading and pre-reading assessments from Pre-K through 12th grade. This description will focus on assessments for students in 3rd through 12th grades.

Similar to other states, this state already has an end-of-year summative test of reading comprehension used for accountability purposes. The test is a standards-based measure of reading comprehension that requires students to read relatively long passages (i.e., 350 words in third grade, 1000 words in 10th grade) and answer questions requiring thoughtful responses to their content. The questions are mostly multiple-choice but there are extended response questions at 4th and 8th grades. The assessment-for-learning system currently under development will be offered to schools as an aid in guiding reading instruction to improve outcomes on the state literacy standards for students in late elementary, middle, and high school (grades 3 through 12).

The most widely administered tests will be computer-based for ease of administration and data management. Informal assessment procedures will also be developed for individual teachers. At present, three computer-based tests are under development:

1. *Broad screening and progress measure focused on grade-level skills in reading comprehension.* In this computer-adaptive test of reading comprehension (items vary depending on student responses), passages and items mimic as closely as possible those which a student is likely to experience on the end-of-year summative measure. It will provide three types of information about students who take it. The first type will be about a student's current reading comprehension ability, both absolutely (an ability score), and in relation to a representative sample of other students who took the test at the same time. The second type will concern the probability that the student will be able to meet grade-level standards on the summative test in the spring, and the third type will be a Lexile score, which provides information about the level of text the student can be expected to read with good comprehension.

For students in the system without previous year-end summative scores, this test can be used at the beginning of the school year as a reliable estimate of current reading ability relative to grade-level standards in the state or to provide an updated estimate (in the context of standards for the student's

current grade level) for students whose performance on the previous year's summative test was marginal, or below standards. The test will be engineered to be taken in the fall, winter, and spring, in order to gauge individual student progress toward meeting grade-level standards. As currently conceptualized, the most common use of the test would be to identify students in need of extra support at the beginning of the year and to evaluate progress on grade-level standards from those supports during the year.

2. *Targeted diagnostic inventory.* This diagnostic battery will comprise two tests whose purpose is to differentiate between students needing intensive, 90-minute daily intervention periods focused on a broad range of reading skills and students who can profit from less intensive (60-minute daily) interventions delivered by specially trained teachers within content areas. The focus of this latter type of intervention is on the comprehension and vocabulary skills required for success on the year-end summative test at the current grade level, rather than on basic reading skills such as word analysis and reading fluency.

One test in the targeted diagnostic inventory requires students to read two grade-level passages on the computer and respond to embedded cloze items within the text. A timed reading (three minutes per passage) provides a measure of silent reading fluency as well as low-level, or gist comprehension. This measure corresponds closely to the kind of maze test recommended in the review of curriculum-based measures by Wayman and colleagues (2007) as appropriate for monitoring progress in basic reading skills in middle and high school. The outcome of this test will be a statement of absolute performance on the test, as well as how that performance corresponds to a representative sample of other students in the same grade who took the test at the same time.

The other test in the targeted diagnostic inventory is conceptualized as assessing a student's phonological and orthographic knowledge of words in print. It is a computer-adaptive spelling test that will produce an absolute ability score and a statement of a student's relative performance compared with a representative sample of other students at the same grade level. It is expected that experience with the outcomes from both of these tests will eventually provide instructional leaders with the information they need to discriminate between at-risk students who are struggling with basic reading skills (low-level comprehension, reading accuracy, and reading fluency) and those whose projected difficulties on the state summative measure are primarily the result



of difficulties with the complex reasoning and vocabulary demands of the test at their grade level.

3. *Informal diagnostic/progress assessment tool box.* These tests and assessment procedures will be developed for dissemination to both intervention and content-area teachers to aid in day-to-day classroom decision-making. Current development plans include:

- *A phonics screening inventory:* This assesses the specific knowledge required to decode words phonemically, such as letter-sound correspondences, blending ability, and skills with complex syllables or different syllable types. It is intended for intensive intervention teachers.
- *A sight word inventory:* This list of sight words that occur with increasing frequency at each grade level provides the teacher with an estimate of students' context-free word reading ability. It is intended for intensive intervention teachers.
- *Graded passages with linking word lists:* Word lists will be linked to passages in a way that will allow teachers to identify the level of passage that a student can read with sufficient accuracy so that word-level skills are not a hindrance to comprehension. These passages will be used to provide information about a student's comprehension skills on passages that can be read with reasonable accuracy and fluency.
- *Comprehension strategy inventory:* This is an interview protocol that both intervention and content-area teachers could use to make initial assessments of student knowledge of comprehension strategies skilled readers often use to enhance comprehension.
- *Assessment examples for grade-level literacy standards:* These are classroom-based literacy assessment strategies that content-area or literacy support teachers could use to monitor student literacy skills relative to grade-level standards. Examples might include appropriate discussion questions, performance tasks with associated scoring rubrics, extended response questions for quizzes with scoring rubrics, etc. The goal is not to provide a comprehensive set of classroom-based formative assessments to be used at every grade level in every class, but rather to provide examples tied to specific grade-level literacy standards to help teachers develop better classroom-based formative assessments themselves.

Data from the formal, computer-based assessments will automatically enter the state's data management system, which will be able to provide reports of student progress at the individual student, classroom, school, district, and state levels. This information will be immediately and continually available to instructional personnel at all levels to inform decisions about individual students, classrooms that need support, or school- and district-level progress in literacy instruction.

Commentary

The computer-based reading comprehension assessment in this system is most like benchmark, or interim assessments, in that its primary use will be to identify students at risk for failing to meet grade-level standards in reading and to monitor progress toward those standards. At least in its initial version, there will be no attempt to measure student ability on individual state literacy standards. The test is not designed to provide reliable diagnostic information about relative strengths and weaknesses on specific standards. Rather, it will serve primarily to estimate a student's readiness to attain grade-level performance on the grade-level summative test administered at the end of the current year or end of the succeeding year (spring administration). The two tests in the targeted diagnostic inventory can be used for diagnostic screening at the beginning of the year and for progress monitoring during the year, in which case they will function more like general outcome curriculum-based measurements than classroom-based formative assessments.

Extensive professional development supports for teachers and school leaders are anticipated as these assessments are implemented throughout the state. As teachers and school leaders become more familiar with the data the formal assessments generate, and more skilled in using them to plan instruction, the informal assessments are expected to help teachers develop skills in designing and using classroom-based formative assessments to guide their instruction.



Example #7

The organization featured in this example supports schools in several states and acknowledges the importance of the daily assessments that take place in the class through observations, work samples, and unit assessments. However, it also has found it useful to have a regular gauge of how well students are progressing on literacy standards and preparation for “high-stakes” summative tests. Thus it created monthly benchmark assessments to measure progress on state standards and provide data to guide instruction. State-specific question banks were created to match both the content of each state’s standards and how they are assessed at each grade level. Text difficulty is determined using a Lexile scale and passages selected to match the types of texts and genres used in the state assessment. To guide instruction, these benchmark assessments include questions matched to specific state standards and to the emphasis a specific standard receives on the state test.

Students in grades 2 through 11 take electronic reading and math benchmark assessments monthly, with expected student performance levels aligned to state test cut scores. Monthly student scores are used to check whether students are on track and to determine whether an instructional response plan is necessary. Information about the data is provided to students, teachers, and administrators. Students see the results of their monthly benchmark assessment immediately and can review their progress charts. To facilitate students’ ownership of their learning, they are encouraged to know their strengths and weaknesses and to discuss them in the language of the standards.

Teachers receive a wide variety of electronic reports that can be manipulated in various ways to assist with instructional grouping or to monitor class and individual student growth. Teachers also try to use student scores and the percentage correct for individual test items and strands within standards to monitor student learning and guide instructional planning. Teachers review reading passages and questions to monitor possible answers for each question and receive information about instructional implications for wrong responses in order to better understand the nature of instruction a student needs to master a skill or understand a concept. Teachers also use the data to compare student performance to the amount of time and the depth to which they have taught a skill each month, using this information to adjust instructional plans for the coming month. Students can be regrouped or selected for targeted small-group

instruction. Teachers have ready access to large amounts of data, and receive coaching and support in learning to use the data and reports to make effective instructional decisions.

Principals can use the reports to generate individual student-, teacher-, and school-level information, and they receive support to build their expertise in using data for school leadership and instructional decisions. Reports are typically used to determine when professional development or action strategies are needed to improve student learning.

Commentary

This is a benchmark, or interim assessment system, although the assessments are given more frequently than many interim assessments. Feedback on student performance on the monthly assessment is quickly available to students, teachers, and principals in a user-friendly and flexible format. The assessments are aligned to state literacy standards, and efforts are made to provide information to teachers about the meaning of incorrect answers and their instructional implications. The professional development and support to build teacher knowledge about instructional decision-making based upon data are critical to maximizing this model's potential.



Example #8

This example involves an Internet-based system designed to provide reading selections matched to students' reading abilities. This system's innovative feature is that it continuously assesses comprehension performance and rate of silent reading, so that as students become more proficient readers, it makes corresponding adjustments to the difficulty level of the passages that it offers for selection. The system is based on the assumption that reading practice will be most productive when passages are not too easy or too difficult. The system identifies selections for students to read from a large corpus of books and articles that have been rated according to their difficulty level.

The system is based on the Lexile (L) framework, which is widely known among educators. It provides a way to place both students' reading levels and the difficulty levels of individual texts on the same quantitative scale. Theoretically, a student reading at a Lexile level of 910 should be able to attain 80% comprehension of texts that are close to the 910L level of difficulty.

The system in this example has access to more than 25,000 topical articles on a wide range of subject areas. Students may search, by content area or by keywords, for selections that fall within ± 100 Lexiles of their reader ability. If a student's Lexile level is unknown to begin with, he or she may select grade-level passages (depending on the teacher's knowledge of the student's reading skills), and the system will then adjust the difficulty of subsequent passages up or down, depending on how the student performed.

The formative assessment used in this system is a type of comprehension assessment employing embedded cloze items that are created by the program. Items are created within the text by periodically removing key words that fall roughly at the readability of the selection being read. Students complete the cloze by selecting the best response from four choices. Word choices available to complete the cloze are at the same difficulty level as that of the text. For example, in an article titled *Listening to the Wildlife in the Everglades*, which falls at a Lexile level of 1230, a selection of words with Lexile measures within 100L of the text are identified by the computer and deleted from the text. Students are presented with the word from the text and three other syntactically correct words, and are asked to identify the best word to complete the sentence, given the passage's context. If a student's performance falls much below or above the expected level over several passages, the Lexile level of selections is adjusted downward or upward respectively.

This system has been piloted with groups of students in grades 1 through 12, and the data are very promising in terms of the system’s ability to predict how well students will be able to comprehend the passages it provides.

Commentary

This system incorporates short-cycle, classroom-based assessments that continuously adjust the difficulty level of passages that students are assigned to read. Although the formative assessments in this system do not guide instructional changes, they do allow the system to “fit” a student’s reading ability even as that ability changes over time. If reading growth or engagement are facilitated by opportunities to practice reading passages that are “just right” in terms of difficulty level, practice with this system should facilitate reading growth. This system is closer to classroom-based formative assessment than to the other types of assessment discussed in Part I, although it also bears some important similarities to general outcome curriculum-based measurement.

As the corpus of articles in the system expands, it could be used to provide reading assignments in specific content areas aligned with the curriculum at each student’s appropriate level of difficulty. Although currently there is no rigorous research evidence that the algorithms used by this system actually select passages at an optimal level of difficulty for effective reading practice, there is widespread agreement that fluency is enhanced by reading material of appropriate difficulty (Rasinski, 2003). Furthermore, *Concept Oriented Reading Instruction* (Guthrie, Wigfield, & Perencevich, 2004), a new and relatively powerful approach to comprehension instruction for intermediate students, depends on having topical selections available at different reading levels so that students of all reading abilities in a class can participate in content-oriented discussion and cooperative learning projects.



Example #9

In the 1999–2000 academic year, a large state identified the high school in this example as in need of improvement. In response, the principal and faculty adopted a collaborative strategy to improve student outcomes. Their first focus was to improve the curriculum and instructional program that all students received, heavily emphasizing content mastery in all academic programs. They also designed varying levels of academic intervention programs to match their students' needs. They systematically aligned their curriculum and instructional materials to state standards. Any standard that appeared on both the state test and the test from the curriculum publisher was deemed essential. These essential standards were then used to develop pacing calendars in every subject to guide the weekly instructional content for each six-week period.

The second focus involved using data collaboratively to guide decisions, drawing annually on all available national, state, and local data to help ascertain the school's overall progress. However, teachers needed an effective method to review student progress more frequently in order to make daily instructional adjustments. In response, teams of teachers developed benchmark assessments to measure student mastery of the standards covered in each six-week period, and they also developed a process for analyzing the data from these assessments. The results of the benchmark assessments were available to teachers within an hour and to students the day after assessments were given (every six weeks). Teacher coaches facilitated faculty meetings to look at subject-area data for both classrooms and individual students. As they analyzed the data, teachers shared information about the approaches and strategies used in the high-performing classrooms. They also developed a re-teaching plan to address standards and questions that students missed. The school leadership team regularly reviewed data for student subgroups, instructional programs, and classes or instructors. These reviews invited questions by and for the teachers about identifying what was working and which strategies might be duplicated, including ways to share best practices.

As a result of these changes implemented since the 1999–2000 school year, performance for all groups of students has improved at the high school. The school made a very significant gain on the state's evaluative criteria over the last three years. Their "similar schools ranking" moved from 3 to 10 on a 10-point scale.

A closer look at the benchmark assessments in this example

The faculty at this school had national, state, and local data about broad student performance, but also needed frequent reports to assess how students were progressing toward meeting the state standards. They organized an assessment process that allowed regular collection of evidence on individual student academic progress and on student progress across common courses and grade levels. The faculty determined what to measure, who would administer the assessments, and the frequency of assessments. Once they identified essential standards and established the pacing calendar, teams of teachers worked in subject-area groups to write benchmark assessments consisting of five questions that assessed student mastery of the big ideas for each of the five to six standards taught in each six-week period. As teams began writing the benchmark assessment items, they realized the need to check the validity of these assessments, and instituted a three-step plan. First, they examined test questions to determine if a problem arose with either the vocabulary or the question's intent, and repaired items as needed. Second, they verified that the instruction provided was aligned with the assessment and that all topics were covered to mastery. If misaligned, they implemented plans to adjust the instruction. Third, they identified gaps in student understanding. If gaps existed, the content was re-taught. While the faculty teams recognize that the development of high-quality assessments is an ongoing process, they express confidence that the assessments they have already developed provide a significant improvement in their ability to track student progress over methods they were using previously.

Administering the benchmark assessments has become a school-wide event every six weeks. Anecdotal reports indicate that students look forward to learning about their progress and that teachers are equally eager to receive results that enable them to make meaningful adjustments in instruction and keep their students on track academically.



Commentary

The assessments used in this example appear to be most similar to the common formative assessments described by Ainsworth & Viegut (2006) in that they were developed by the teachers at the school and were closely linked to the curriculum sequence and focused on essential state standards. Because they are not administered daily and are common across all teachers of a given content area and grade level, they would not be classified as classroom-based formative assessments. However, they are more closely tied to the instructional sequence, and more focused on specific curriculum outcomes, than is frequently the case for district-level benchmark assessments.

Example #10

When the announcement was made that new graduation requirements included a specific performance level on their state's end-of-year summative test, the principal and staff of this high school determined to increase the percentage of students who scored at that level. However, faculty lacked an accurate process to determine which students required additional help to meet that goal. The district had already worked with a test development company to develop computer-adaptive progress monitoring assessments aligned to state standards in math and reading for grades 3 through 8. These tests were administered individually to students three times each year. The school district agreed to expand the progress monitoring assessments to include grades 9 and 10, a decision that allowed the high school staff to use multiple data points to identify students in need of additional reading instruction. They administered the progress assessment at the beginning of 9th grade, reviewed the 8th grade scores on the state accountability measure, checked attendance data, and verified with middle school staff which incoming 9th graders should be considered for targeted reading instruction.

Language Arts teachers volunteered to teach one section of a reading intervention class for students who were judged to be in need of extra reading support (these students were also scheduled in their regular Language Arts classes). These "intervention teachers" received training and ongoing coaching support for implementing the intervention program, and the schedule was organized to facilitate collaboration among teachers who had the same reading intervention students in content-area classes. The intervention classes concentrated on reading strategy instruction as applied to content-area texts.

The Assessment Process

This high school is part of a school district that provides, along with the annual summative test results, a computer-based progress assessment to periodically measure student progress toward meeting state standards. Online test administration allows efficient determination of individual student knowledge levels by adjusting the question difficulty after the student responds to each question. The student can view his or her score immediately. The teaching staff and school administrators receive individual and class reports and guidance about subskills that need to be taught based on student answers, although this information is less reliable than the total score.



The high school staff uses these data to monitor student progress toward receiving a passing score on the end-of-year summative test. The three administrations of the benchmark progress assessment occur at the beginning of the year, mid-point, and spring. The results provide guidance about student progress, and identify students who have obtained a high enough level of skills to exit the intervention class. Even with the new progress assessment, teachers found that they needed additional data to guide individual student instruction. They also determined that to keep the students engaged, more frequent data collection showing them continual progress was necessary.

Teachers and staff then developed a procedure, derived from intervention classroom instruction, to collect additional data. They designed instruction in the intervention class around the reading to be done in students' content-area classrooms, and content-area teachers then worked with the intervention/ Language Arts teachers to identify reading assignments, vocabulary, and concepts students were supposed to learn. The intervention class teacher taught students strategies to improve their reading fluency and comprehension of content-area texts. Each week, intervention students take a one-minute timed oral reading fluency check on a student-selected passage. Students graph their word counts to track their progress. A literal comprehension check and discussion follow. If a student answers a literal question incorrectly, the teacher and student re-read the text together and the teacher models a comprehension strategy the student could use. The teacher and student also discuss the text selection and the student's overall performance (e.g., "Your graph went up. Why? Yes, you already knew a lot about this content." Or "Your graph went down. Why? That is right. The selection had a lot of technical vocabulary and numbers. These slowed you down because you don't read them very frequently.") The teacher uses these data to guide individual instruction and practice for the following week, and alerts content-area teachers to reading strategies that need reinforcement and content that students find difficult.

Commentary

This assessment system combines benchmark assessments to monitor progress in acquiring grade-level reading skills and general outcome curriculum-based assessments to monitor progress and guide instruction for students in intervention classes. The oral reading fluency probes actually go considerably beyond the information that is typically provided by general outcome CBM measures, because they incorporate follow-up questions and discussion of student answers. They are also used to probe student awareness of factors that influence reading performance. Because this type of questioning and discussion is individually tailored to each student, these fluency probes have acquired some of the characteristics and utility of classroom-based formative assessments. Decisions made from the benchmark assessments focus primarily on determining which students need extra reading support in order to be able to meet grade-level reading standards. The school is also experimenting with using information from the standards-based reading comprehension measure to help focus instruction in grade-level language arts and content-area classes. However, that work is not as developed as work with the fluency probes and follow-up discussion to guide instruction in intervention classes. Instruction for both intervention and grade-level students would likely profit from further work to increase teacher skills in administering and using a broader range of classroom-based formative assessments.



CENTER ON
INSTRUCTION