http://www.ncsall.net/?id=574

Literacy Education in Adult Basic Education

Volume 3: Chapter Four
John Kruidenier

Adult basic education programs, sometimes called adult basic and secondary education programs, typically serve adults over the age of sixteen who do not have a high school diploma and are no longer eligible for traditional secondary education programs. Although adult basic education (ABE) is situated apart from the elementary, secondary, and college education systems, it does

not exist in a vacuum. This is especially true of literacy assessment in adult basic education now, at the turn of the century. Adult literacy assessment is affected by changing definitions of literacy, changes in the needs of federal and state funders, and changes in assessment tools and practices. These changes will be discussed in detail in this chapter in order to present a broad picture of the current state of literacy assessment in adult basic education. Because none of these changes are completely new, a brief history will be presented to put them in context. Literacy must be defined before it is possible to know how to assess it, and assessment must be defined before it is possible to know how best to implement it. This chapter thus begins with definitions of literacy and then describes assessment in adult basic education within this definitional framework. Implications for practice, research, and policy follow.

DEFINING ADULT LITERACY

ASSESSMENT

This section presents various views of literacy and identifies three dimensions that appear to be especially important for adult literacy: context, practice, and ability. A working definition of literacy assessment is then presented, and important characteristics of both traditional and newer forms of assessment are introduced.

Views of Literacy

A straightforward, though narrow, definition of literacy is the ability to read and understand written text. This definition is roughly doubled in complexity when written expression is added to the way in which literacy is viewed: the ability to write understandable text. Even more complex and expansive views of literacy are possible. There is no single, fixed view of literacy. The existence of multiple viewpoints makes sense given the following statements about literacy, all of which are true: reading ability itself is

a continuum (adults are described as high- and low-literate); reading is both a psychological or cognitive phenomenon and a sociocultural phenomenon (occurring within and outside the individual); writing is a form of literacy that is virtually inseparable from reading; numeracy, or the ability to read, write, and manipulate numbers is considered a form of literacy by many; literacy may develop differently in different types of individuals (native-language learners and those attempting to become literate in a second language, those with and without a specific learning or reading disability, females and males); and oral communication differs from written communication along a continuum from the less formal to the more formal (Harris & Hodges, 1995, p. 140).

THE ROLE OF CONTEXT. Expansive definitions of literacy abound. Most dictionaries define a literate person not only as one who can read and write but also as one who is well-informed,

educated, or cultured. Although this is a relatively old definition, it has led recently to a phenomenon that might be called "literacy with an adjective." More than thirty-five types of literacy are listed in the international reading association's literacy dictionary (Harris & Hodges, 1995, p. 141). Some of these are more directly related to reading and writing, such as family literacy and adult literacy, while many others are more expansive, such as computer literacy, cultural literacy, and media literacy, to name just a few.

Implicit in the literacy with an adjective phenomenon is the view that literacy is more than reading, writing, and computing with efficiency and understanding. It is also the ability to practice reading and writing in specific situations to obtain or communicate specific information (Guthrie & Greaney, 1991; Smith, 1995, 2000; Reder, 1994). Although the number of situations or contexts linked with literacy

may be new, the central role of context in defining literacy is not.

One dimension of the history of the development of reading and writing over the past six thousand years is the expansion in the number of situations in which literacy may be used and the number of people using it (Kaestle, Damon-Moore, Stedman, Tinsley, & Trollinger, 1991; Venezky, 1991). Literacy was originally a craft confined to a select group of clerics and government and business bureaucrats (ecclesiastical, governmental, and business literacy). It was then extended to many societies' elite classes (cultural literacy, perhaps, is added to the mix of literacies). Finally, after the invention of the printing press in the fifteenth century and through the second half of the nineteenth century, literacy was put within reach of most people (Kaestle et al., 1991).

Perhaps the most expansive view of literacy is critical literacy, wherein reaction to a text is considered to be grounded in one's social, political, or economic situation (Brookfield, 1997; Fehring & Green, 2001; Hiebert, 1991). Literacy in this context is "reading the world" (Freire & Macedo, 1987), and its goal is to continue the spread of literacy to adults as a form of empowerment. All that we express about a text we read is bound to our past experience, which is shaped by society (Alvermann, Young, Green, & Wisenbaker, 1999).

LITERACY PRACTICES. Literacy practices are closely related to context. Practices describe how individuals use reading and writing in various situations and include, for example, reading books, newspapers, or magazines, reading job-related texts, writing letters, and so on (Guthrie & Greaney, 1991; Smith, 1995, 2000; Diehl & Mikulecky, 1980; Mikulecky & Drew, 1991; Sticht, 1995; Kirsch, Jungeblut, Jenkins, & Kolstad,

1993).

Practices are sometimes associated with specific contexts. Guthrie and Greaney (1991) found that adults do most of their reading at work while scanning brief documents such as tables, schedules, memos, and bulletins. The next largest amount of time is spent reading books during leisure time, and then newspapers and magazines, also during leisure time. Some practices, however, may occur in several contexts. Reading a newspaper, for example, could take place when looking for a job, buying a house, or learning about a political candidate. Literacy practices, because they are not always linked to one specific context, could be considered a separate dimension in definitions of literacy.

PSYCHOLOGICAL PROCESSES. Context is an important dimension in definitions of literacy. It incorporates all that might be going on around or outside an individual. An equally important

dimension is what goes on within individuals as they read and write, what enables an individual's literacy practices in various situations. Like the issue of context, the study of psychological or cognitive processes involved in reading also has a history, although it stretches over roughly the last one hundred years instead of thousands of years.

As Stahl (1999) notes, the history of reading instruction in the United States over the last century reflects the changing views of the internal mechanisms or cognitive processes involved in reading and writing. Instruction in reading at the turn of the century focused on the ability to decode text. By mid-century, the focus had shifted to an emphasis on meaning, typically the ability to read a passage and answer factual questions about it.

In the 1980s, according to Stahl, the definition of reading shifted again to include an emphasis on meaning

construction, the ability to combine ideas that exist in memory with ideas derived from a text being read (Anderson, 1984; van Dijk & Kintsch, 1983; Lesgold, Roth, & Curtis, 1979). In this view, constructing the meaning or mental representation of a text while reading involves the actions of many processes.

Many different processes are involved in constructing these representations. To mention just a few, there is word identification, where, say, a written word like bank must somehow provide access to what we know about banks, money, or overdrafts. There is a parser that turns phrases like the old men and women into propositions [ideas in memory] . . . There is an inference mechanism that concludes from the phrase The hikers saw the bear that they were scared. There are macro-operators that extract the gist of a passage. There are processes that generate spatial imagery from a verbal description of a place. [Kintsch,

1988/1994, p. 951]

Next, continues Stahl, the whole language movement brought with it a new emphasis on reading as a response to a text, along with issues such as motivation to read and an appreciation of literature (for example, Cramer & Castle, 1994). More recently, "balanced" reading instruction has emerged, in which decoding, meaning construction, and motivation or engagement are all considered important aspects of the reading process (Stahl, 1999; Baker, Dreher, & Guthrie, 2000; Pressley, 1998; Snow, Burns, & Griffin, 1998; National Reading Panel, 2000).

Results from studies of basic reading and writing abilities indicate that within individuals both reading and writing are cognitive processes made up of several components (Perfetti, 1985; Curtis, 1980; Perfetti & Curtis, 1987; Chall & Curtis, 1987; Carr & Levy, 1990; Snow

& Strucker, 2000; Gregg & Steinberg, 1980; Torrance & Jeffery, 1999; Levy & Ransdell, 1997; Kruidenier, 1991). This is an attractive notion for some educators because it suggests that teachers can focus on specific aspects of the reading and writing process during assessment and instruction.

Components or aspects of the reading process that are typically addressed by instruction include word analysis (phonemic awareness and phonics), word recognition, fluency (accuracy, rate, and prosody in the reading of connected text), word meaning, and reading comprehension and metacomprehension (knowledge of comprehension strategies) (Chall, 1994; Chall & Curtis, 1992; Curtis, 1999; Curtis & Chmelka, 1994; Roswell & Natchez, 1979; Strucker, 1997a, 1997b; Kruidenier, 1990). Although components of the writing process are not as well defined through research, they include both general or

global processes as well as lower-level processes (Flower & Hayes, 1981; Hayes, 1996; Torrance & Jeffery, 1999; Levy & Ransdell, 1997; Kruidenier, 1991, 1993). The more general or global processes include planning (generating and organizing ideas), forward production (translating ideas into text), and editing and revising. Lower-level processes include word production (spelling) and sentence production (syntax and morphology). An additional component of both the reading and writing process is motivation or engagement (Beder, 1990; Guthrie & Wigfield, 1997; Baker, Dreher, & Guthrie, 2000).

These aspects or components of reading and writing processes develop over time. Individuals may be described as being at various levels or stages in the development of their literacy abilities (Chall, 1996; Adams, 1990; Collins & Gentner, 1980; Bereiter, 1980). This is

the basis for some assessments that place readers at a developmental level based on ability. It is also the basis for some forms of diagnosis that describe students' strengths and weaknesses. One component or aspect of the reading process may develop at a rate different from that of another. Looking at these different rates across components yields profiles of literacy abilities (Chall, 1994; Strucker, 1992, 1997b; Snow & Strucker, 2000). The notion that component processes are active whenever reading and writing take place and that they develop over time is another important dimension in views of literacy.

The definition of literacy that will be used in this chapter to discuss adult literacy assessment includes the three dimensions described thus far: context, practices, and ability. It might be summarized as follows: Literacy is the ability to read (construct meaning from

text) and write (create text that is meaningful). Reading and writing are processes, consisting of specific subprocesses or components operating in memory within individuals. These processes are expressed through literacy practices in specific contexts among individuals.

As important as describing what will be included in a discussion of adult literacy assessment is a description of what will not be covered. First, although the assessment of numeracy, mathematics, or quantitative literacy could easily be incorporated into this definition, it is left out because it is beyond the scope of this chapter. Also left out of the discussion are literacy contexts that are not fairly directly related to adult literacy or that have not received as much attention in the adult literacy literature. Contexts that will be considered are those especially important to adults, including the workplace (Diehl & Mikulecky, 1980;

Mikulecky & Lloyd, 1997; Sticht, 1995), the home or family (National Center for Family Literacy, 1996), and health and community settings (Davis, Crouch, & Long, 1992; Nurss, Parker, Williams, & Baker, 1995).

The assessment of specific types of adult learners is also beyond the scope of this chapter. Second-language adult learners, adults with learning disabilities, and other subgroups of adult learners will not be considered separately. The purpose of literacy assessment is not to identify a learning disability, although good assessments of literacy should provide adequate information on instructional planning for all adults, including those with a reading disability. One possible exception would be testing that attempts to measure native-language literacy to help determine the global literacy ability of students in English for speakers of other languages (ESOL) classes. Readers interested in adults with learning

disabilities may want to focus on the discussion of assessments that provide the most information about beginning readers (Snow & Strucker, 2000; Corley & Taymans, Chapter Three of this volume).

Views of Educational Assessment
Assessment in education is defined by Harris and Hodges (1995) as "gathering data to understand the strengths and weaknesses of student learning" (p. 12). Using the description of literacy provided in this chapter, literacy assessment might be defined as gathering data to understand the strengths and weaknesses of student reading and writing abilities and practices in various contexts. Adult literacy assessment has been heavily influenced by several developments in the field of educational assessment: standardized testing and more recent innovations in assessment, including criterion-referenced testing and performance or alternative assessment.

STANDARDIZATION: TESTING, VALIDITY, AND RELIABILITY. Educational testing, including tests of literacy ability, has a long history. School examinations in china were administered as early as the twelfth century B.C. (Nitko, 1983). The first recorded reading assessments in England and France occurred in the fourteenth century A.D. or earlier and consisted of oral reading (reading aloud) (Resnick & Resnick, 1977; Venezky, 1991). The history of educational assessment in the United States in the past hundred years, however, is dominated by the development of standardized testing. During the first half of the twentieth century, testing was heavily influenced by theories of mental abilities developed in the field of psychology and by the use of individually administered IQ tests, first in France by Binet and Simon and then in the United States in the early 1900s (Nitko, 1983, p. 445). The first group-administered intelligence test,

which included a silent reading comprehension section, was developed by the U.S. Army during World War I (the Army Alpha) (Sticht, 1995).

Advances in the field of statistics beginning in the mid-1800s also contributed to the development of standardized tests. Statistical analysis of raw scores (usually the total number of correct answers) enabled one person's score to be compared with the scores of all others taking a test in numerically objective, accurate, and precise ways.

With compulsory education in the 1920s and 1930s came the rapid development and increased use of standardized achievement and intelligence tests, as well as their misuse by Social Darwinists and the eugenics movement (Nitko, 1983). These tests were considered to be standardized because administration and scoring procedures were the same for all examinees. Exam questions were

presented in the same way to everyone, and tests were all scored in the same way, using detailed examination guides and trained examiners. (See Exhibit 4.1 for a description of some common assessment terms, such as standardized.)

By referencing one person's score to the scores of a representative group of those for whom the test was developed (a norm group), examiners could compare learners' abilities and use this information in the process of making decisions on, for example, which candidates to admit to an educational program and where to place them. Over the years, several types of norm-referenced scores have been developed that can be used to compare one person's raw score to another's: percentile ranks and stanines (what percentage of students score below a given raw score), scale scores (comparing one person's score to a norm group using a scale that, unlike percentile ranks, is an equal-

interval scale), and grade-equivalent scores (which relate a raw score to the typical or average performance of students at specified grade levels) (Nitko, 1996).

Standardization of testing has also led to relatively specific, agreed-upon methods for evaluating a test using the concepts of validity and reliability. A test is considered valid if it is judged to adequately measure the domain of knowledge that it was designed to measure. A test is judged to be reliable primarily by means of statistical measures that indicate how reliable its scores are, including reliability coefficients that measure how consistent the scores are and a standard error of measurement that suggests how accurate they are. The statistical measures of reliability are tools that are used to address the broader, more qualitative aspect of a test's validity (Nitko, 1996). A test must be reliable to be valid;

reliability is a necessary but not sufficient condition for validity. For adult literacy assessment, these developments in standardized testing culminated in the development of norm-referenced tests for use specifically with ABE students in the 1950s and 1960s, including, for example, the Adult Basic Learning Exam (ABLE) (Karlsen & Gardner, 1986).

Standardized tests have played a significant role in what Linn (2000) has identified as the five prominent "waves of reform" that have swept through education since World War II:

1. The movement toward grouping or tracking in the 1950s to handle the diverse population of elementary and secondary students entering public schools. (Standardized tests were important in placing students.)
2. Large, federal expenditures for compensatory education in the 1960s through the Elementary and Secondary Education Act. (Tests were used to satisfy congressional demands for evaluation and accountability.)
3. Minimum-competency testing in the 1970s and 1980s.

4. High-stakes standardized testing in the 1980s and 1990s. (Teachers and administrators were held accountable for test results.)
5. Current reform efforts. (These include the high-stakes accountability element of earlier reforms along with "ambitious con-tent standards," assessment and accountability based on these performance standards, performance-based assessment, and inclusion.)

INNOVATIONS: CRITERION-REFERENCED AND PERFORMANCE-BASED ASSESSMENT. Several innovations in assessment also occurred during the post-World War II period. Minimum-competency testing, the third reform wave, is a type of criterion-referenced testing originally developed for the military (Sticht, 1995) in the 1960s by Glaser and others as an alternative to norm-referenced testing (Glaser, 1963, cited in Nitko, 1983, p. 445). Instead of comparing a test taker's score to others' scores (a norm group), criterion-referenced tests compare the test taker's performance to the domain of performances being assessed (Nitko, 1996). Assuming that reading ability can

be represented along a continuum from no or very few literacy abilities (competencies) to advanced forms of literacy, for example, a criterion-referenced reading test is used to determine how literate a learner is, or where along the continuum that learner could be placed. Similarly, performance standards specify the domain of instructionally relevant tasks that a learner should have mastered at a given level or point along the continuum (Nitko, 1983). Criterion-referenced measures focus on determining what an individual already knows and therefore what needs to be taught as opposed to an individual's standing relative to a group of peers.

The last current wave of reform described by Linn (2000) includes the development of performance-based assessment. Performance assessments are used to evaluate how well students complete tasks that require the

application of knowledge or skills in a realistic, or authentic, situation. A performance assessment designed to assess adult literacy students' reading, for example, might have them use a manual to troubleshoot a specific problem in a workplace setting (Sticht, 1972; Mikulecky & Lloyd, 1997). Or, to assess writing, students might be asked to help construct a portfolio of their best written work generated in a classroom setting (Fingeret, 1993). Generally, performance tasks involve lengthy written (or spoken) responses or participation in group or individual activities (Nitko, 1996). Assessing specific literacy practices, such as how frequently newspapers are read at home, is also a form of performance assessment, although it is often based on retrospective self-reports rather than direct observation by an examiner.

Performance assessments, like standardized tests, can be evaluated for

validity-that is, judged on the basis of how well they measure the literacy task they purport to evaluate and how consistently they are administered and scored. Performance assessment also includes the use of one or more scoring rubrics to increase reliability. Rubrics are sets of rules that can be used as a guide for scoring and administration (Nitko, 1996) and usually include some sort of scale or checklist. Scoring guides for the holistic or analytic scoring of student writing samples are an early example of this type of rubric. Numbered quality scales are established (a scale from 1 to 4, for example, with 4 being the highest), and descriptions of what is expected of an essay at each level are provided. Evaluators read each essay and assign it a score based on which level of quality it most closely matches.

Performance assessment is conceived of by some as an alternative to standardized, norm-referenced, and

criterion-referenced testing (for example, Garcia & Pearson, 1991). As will be shown later in this chapter, performance assessment is an important part of the reforms under way in adult literacy, including the new National Reporting System for adult literacy (DAEL, 2000).

WHY ASSESS?

Recent reports suggest that many adult educators remain unconvinced that assessment is an important part of the teaching process (General Accounting Office, 1995; Kutner, Webb, & Matheson, 1996; Condelli, Padilla, & Angeles, 1999). These are educators who have, in the past, not used any formal assessment tools or procedures when teaching reading and writing, who have used them only for posttesting, not diagnosis (Beder, 1999), or who have been reluctant to use them because of possible negative side-effects (Ehringhaus, 1991). A review of eleven states' assessment systems (Kutner et al., 1996), for example, found that

Administering standardized assessment instruments is not a priority for most programs; pretests are often administered only to participants whose literacy is considered to be at a sufficient level and very few programs have post-test data, even for learners remaining in a program for a substantial number of hours. Furthermore, standardized assessment instruments are often selected for ease of administration rather than because they reflect the content of what is being taught. [p. 2]

Tests are not directly related to the instruction offered by local adult education programs. [p. 12]

Instructors . . . may need assistance in becoming familiar with the relationship between learner competencies, curriculum, and assessment measures. [p. 17]

Many within and outside the field of adult literacy have described the possible negative effects of assessment, particularly when standardized tests are used. Students, for example, may be anxious about testing, and negative results from tests may lead to a loss of self-esteem and motivation (Ehringhaus, 1991). A standardized test may be culturally biased, particularly when normed on groups that are different either culturally or in some other significant way from those taking the test, and this may lead to misdiagnosis (Garcia & Pearson, 1991; Joint Task Force on Assessment, 1994; Askov, Van Horn, & Carman, 1997).

When used professionally and carefully to minimize possible negative side-effects, however, assessment can be beneficial. The most common uses of assessment in adult literacy include

- Screening to place students in appropriate programs

- Diagnosis of individual strengths and weaknesses in literacy to help plan for instruction
- Measurement of individual growth
- Self-evaluation and personal growth
- Program evaluation and accountability (Askov et al., 1997; Askov, 2000)

These uses are generally accepted in areas of education other than adult literacy as well (Joint Task Force on Assessment, 1994; Joint Committee on Standards for Educational and Psychological Testing, 1999).

Given the apparent usefulness of assessment, is there evidence that it really works, that it leads to improved student learning? Linn (2000) examined test score trends over the last several decades following the use of high-stakes accountability testing. He found a pattern of early gains in average achievement test scores followed by a leveling off. The examination of large-scale assessment programs, however, is difficult and controversial because of the large number of uncontrolled variables

that may affect results. Few carefully controlled studies of the direct effects of assessment in education exist, and there may be none in the field of adult literacy. In a comprehensive review, Dochy and colleagues (Dochy, Segers, & Buehl, 1999) found eleven experimental studies of progress assessment in education. In these studies, teachers assessed students at least twice to measure progress. Most of these studies indicate that the assessment of progress for instructional purposes, when compared with no progress assessment, leads to greater student gains. The researchers suggested that progress assessment may give teachers a better understanding of student ability and thus lead to better, more focused instruction, or that frequent testing may provide students with explicit information about what they need to know.

Very few assessment models in adult literacy go beyond the model described

by Askov (Askov et al., 1997): diagnostic pretests to determine strengths and weaknesses, instruction based on these assessment results, informal assessment during instruction, and posttests to determine gains. One model for assessment and instruction that adds the research-based notions of literacy components and developmental levels (discussed earlier) to this general model is described by Chall (1994; see also Curtis, 1999; Curtis & Longo, 1997; Strucker, 1997a; Kruidenier, 1990). This model, originally developed for use in literacy instruction with children (Chall & Curtis, 1987, 1990, 1992), suggests that each aspect or component of the reading process be assessed to determine a learner's developmental level for each one (for example, word analysis, word recognition, fluency or oral reading of connected text, oral vocabulary, silent reading comprehension, and motivation).

This form of assessment results in a

comprehensive profile of relative student strengths and weaknesses in reading (Roswell & Chall, 1994; Strucker, 1992, 1997b; Snow & Strucker, 2000; Chall & Curtis, 1992; Curtis, 1999). The profile is used to design a program of instruction that addresses all aspects of the reading process while taking into account the unique needs of each learner. Instruction is built around each component, ensuring that developmentally appropriate materials and instructional methods are provided for both strengths and weaknesses. Ongoing, informal assessment is used to continually adjust instruction as needed. Addressing all components during instruction ensures that no one aspect of the reading process is overemphasized (Strucker, 1997b).

In the description of this model, Chall (1994) notes that assessment also takes into account adult needs and interests and elicits the adult learner's

collaboration. The unique needs and abilities that adults bring to literacy instruction are an important theme in adult education (Kasworm & Marienau, 1997; Sticht & McDonald, 1992; Curtis, 1990). Kasworm and Marienau (1997) propose five key principles for assessment derived from "commonly held premises about adult learning" (p. 7):

Assessment recognizes that adults come to literacy instruction with a wide variety of experiences and an extensive knowledge base and that what they learn will be applied to specific situations.

- In addition to the need to improve their literacy abilities, adults also have affective needs and should be involved in the assessment process through, for example, self-assessment and the sharing of assessment results.
- Giving adults feedback promotes learning.
- Assessment should take into account, and use, adults' involvement in work, family, and community.
- Adults' prior experienced-based learning gives them the knowledge to participate in the design of assessment programs and to be actively involved

in their own assessment (through the use of procedures such as portfolio assessment).

- The use of assessment for instruction has not been the focus of very much research in the field of adult literacy. The modern history of adult literacy assessment in the United States has instead been dominated by large-scale assessments of adults' functional literacy abilities and federal adult literacy legislation.

Large-Scale Assessments

Aside from intelligence testing during World War I, direct assessment of the literacy abilities of large groups of adults first occurred in the 1930s in the United States and then not again until the 1970s (Kaestle et al., 1991). Before this, from about 1840 to 1930, national assessments of literacy consisted of asking adults if they were able to read and write a simple message. These self-reports of a literacy practice were obtained during each national census. From 1940 onward, literacy was measured by asking how many grade levels in school adults had completed (Kaestle et al., 1991; Ehringhaus, 1990). This last criterion for literacy demonstrates a central problem

with criterion-based approaches generally-their arbitrariness. The grade-level criterion for being considered literate gradually increased from grade 3 to grade 12 over the years as the literacy demands of society apparently increased (Ehringhaus, 1990).

The first direct assessment of adult functional literacy abilities was conducted by Buswell in the 1930s (1937; cited in Kaestle et al., 1991, p. 94). As a test of functional literacy, it measured the ability of adults to locate information in texts encountered in everyday life, such as catalogs and telephone directories.

Although occurring much later, during the minimum-competency wave of reform (Linn, 2000), a series of large-scale assessments of adult literacy conducted in the 1970s also included measures of functional or everyday literacy. The Survival Literacy Study, the

National Reading Difficulty Index, the Functional Literacy: Basic Reading Performance Test, the Adult Functional Reading Study, the Adult Performance Level Study, and the English Language Proficiency Study all asked adults to read and respond to functional reading material. This material included, for example, classified ads, product advertisements, legal documents, schedules, and other texts people may encounter in their daily lives. The Survival Literacy Study and Adult Performance Level Study also included writing tasks or items that assessed writing ability (Kaestle et al., 1991).

All these assessments were standardized and all except one were criterion-referenced tests. The determination of functional literacy for the criterion-referenced tests was based on the percentage of questions answered correctly. The literate/illiterate cut-off varied from a low of 75 percent correct

to a high of 90 percent. Several of these tests included additional percent-correct cut-offs to establish three levels of literacy instead of just one: literate, marginally literate, and illiterate.

Kaestle and colleagues note the two problems related to the validity of these national assessments, which are issues commonly associated with criterion-referenced tests. Functional literacy competency was defined by means of specific test content, which may not apply to certain subgroups of adults. In addition, the percent-correct criteria were arbitrary and not always clearly defined.

Similar problems have been seen in the two most recent national adult literacy assessments: the Young Adult Literacy Survey, or YALS (Kirsch & Jungeblut, 1986), and the NALS (Kirsch et al., 1993). Both defined five levels of functional competencies. Although the YALS and NALS used item-response

theory, a statistical technique that is more sophisticated than a simple determination of an individual's percentage of correct answers, arbitrary cut-off scores were still used. To be placed at level 3 out of a possible five levels of literacy ability, for example, an adult's answers must indicate that the adult has an 80 percent chance of getting items of average difficulty correct at level 3. Level 3 is the functional literacy standard for the National Governors Association. When the arbitrary 80 percent cut-off is reduced to 65 percent, the criterion used for the National Assessment of Educational Progress, the number of adults in the United States classified as literate increases by 15 percent (Sticht, 1998; Kirsch et al., 1993).

The choice of content for the NALS test items defined what was meant by functional literacy. The content was similar to content used in earlier

functional literacy tests, although it was grouped into three categories-prose, document, and quantitative passages from everyday life-suggesting three forms of functional literacy. Item format, which included questions about prose, document, or quantitative texts, also suggested a view of literacy that focused on reading comprehension as opposed to other aspects of the reading process, such as word recognition or word analysis. NALS responses were not limited to multiple-choice selections but included extended written responses as well.

Large-scale assessments of adults' literacy practices began in the early 1900s in the United States. These included self-reports, responses to questions such as "Have you read a book in the last month?" (Kaestle et al., 1991, p. 180). Studies of reading habits or practices have been undertaken regularly since, and surveys of reading practices have recently been used in large-scale

studies of adult literacy (the YALS and NALS). The type and frequency of reading practices are now associated with reading ability and used to measure literacy development (Smith, 1995; Mikulecky & Lloyd, 1997; Sticht, Hofstetter, & Hofstetter, 1996) as well as the literacy demands of various jobs (Sticht, 1995).

National Legislation and Effects on Assessment Practices

As mentioned, the federal government's role in adult education began in the military with the assessment of recruits during World War I and still continues (Sticht, 1995). The federal role in civilian adult literacy programs began in the 1960s, during the compensatory education reform movement (Linn, 2000) and the passage of the Elementary and Secondary Education Amendments (PL 89-750), of which the Adult Education Act of 1966 was a part. This federal role has continued through the Elementary and Secondary School

Improvement Amendments of 1988 (PL 100-297), the National Literacy Act of 1991 (PL 102-73), and the Adult Education and Family Literacy Act of 1998 (Title II of the Workforce Investment Act, PL 105-220).

This legislation has funded states' adult literacy programs based on the number of adults in a state who are over the age of sixteen, are out of school, and do not have a high school diploma. It has also affected assessment activities in adult education. Legislative guidelines and language have generally reflected the waves of education reform described by Linn (2000), bringing the accompanying innovations and changes in assessment practices to adult literacy programs.

THE ADULT EDUCATION ACT (AEA) OF 1966. The AEA did not require the use of assessment for program evaluation, only that programs would enable adults to "acquire skills necessary for literate functioning" (Merrifield, 1998). The

overall program lacked realistic goals, specific criteria, and ways to measure progress toward goals, according to an independent government review (General Accounting Office, 1975). The 1988 amendments to the AEA listed specific topic areas to be addressed for program evaluation and mandated the use of standardized tests (Condelli, 1996), part of a larger reform wave in education in which standardized tests were used for accountability. The amended AEA specified that at least one-third of the adults in each state's AEA-funded programs be assessed using valid and reliable norm-referenced, criterion-referenced, or competency-based tests (Sticht, 1990).

Despite these efforts, a review by Padak and Padak (1994) found assessment practices in adult education to be haphazard. This review was based on three statewide surveys of evaluation and assessment practices in adult literacy

programs, and more detailed descriptions of nineteen programs. The authors found that evaluations were either not being done or were reported in ways that made interpretation difficult and suggested several reasons for this pattern of poor assessment practices. First, many programs had flexible open-entry, open-exit policies for their adult students. Reporting data on progress that takes into account different amounts of time in a program requires a level of sophistication in the analysis of data that local programs did not have. Second, programs relied on volunteers, who may not have the knowledge needed to conduct assessments or to understand the need for assessment. Third, evaluations were often tied to funding and may have tended to overstate successes and obscure weaknesses (Padak & Padak, 1994).

THE NATIONAL LITERACY ACT (NLA) OF 1991. The NLA incorporated elements of the last wave of reform described by

Linn (2000). Accountability requirements were increased by asking states to develop "indicators for program quality" in three areas: recruitment, retention, and improvement of students' literacy skills. These indicators were envisioned as a step toward the development of measurable performance standards. Quality indicators were to be developed first (for example, students remain in the program long enough to meet their educational needs), measures were to be established next (for example, hours of instruction student receives), and then performance standards established (for example, 80 percent of students stay at least fifty hours) (Condelli, 1996, p. 1).

Most states, in fact, voluntarily developed performance standards for these and additional areas following the development of model standards by the U.S. Department of Education (DOE). The additional standards were related to

program planning, curriculum and instruction, staff development, and support services (Condelli, 1996). The NLA also incorporated new literacy assessment techniques, allowing states to report learner gains using standardized tests, teacher reports, learner self-reports, measures of improvement in job or life skills, and portfolio assessment and other alternative performance assessments (General Accounting Office, 1995).

The NLA required that states use their new quality indicators (and, presumably, the associated performance standards) to evaluate the effectiveness of local programs, although it did not provide evaluation guidelines. A review of usage of indicators and standards found that by 1996 they were being used in virtually all states to evaluate local program effectiveness, to determine which programs needed assistance, and to improve the quality of state programs. A little more than one-half of all states

were using them to make funding decisions, reducing or eliminating funding to those programs not meeting specified standards (Condelli, 1996).

Despite the DOE's attempt to provide states with technical assistance related to performance standards, assessment and evaluation, and data collection and reporting systems, several reports were extremely critical of the federal and state adult literacy delivery systems and of assessment practices in particular (General Accounting Office, 1995; Kutner et al., 1996; see also Stein, 1997). Many of the criticisms questioned the validity of the assessment procedures used. As mentioned earlier, validity in assessment refers to how well an assessment measures the domain of knowledge or behaviors it is designed to measure (Nitko, 1996). If the domain is not well defined, measurement may not be adequate. As evidence of this problem, critics pointed to the

inconsistent definitions of learner progress across states, poorly defined objectives, and the use of different standardized tests in different states (General Accounting Office, 1995; Kutner et al., 1996).

Given the very broad definitions of literacy currently used in adult literacy (see "Views of Literacy," earlier in this chapter), it is not surprising that different definitions of literacy exist or that some objectives associated with broader definitions (for example, functional literacy) may be difficult to measure. To a certain degree, questioning the external or domain-related validity of a particular standardized test (or of standardized, norm-referenced, and criterion-referenced tests generally) is simply one way to express disagreement with the aspects of literacy (the skills, contexts, or practices) on which the assessment focuses (see Stein, 1997, and Merrifield, 1998, for examples of this type of

criticism).

In addition, multiple and perhaps conflicting definitions of the literacy domain might be expected in a system that has multiple funders with different interests and views (Merrifield, 1998). Fifty-nine percent of adult literacy programs are funded through local education agencies (primarily local school districts), 15 percent by community colleges, 14 percent by community-based programs, and 12 percent by other agencies (General Accounting Office, 1995; Beder, 1999).

The other major criticism of the validity of assessment practices had to do with serious questions related to the reliability of the data collected for assessment, or how consistent the collection of data was. The General Accounting Office report (1995) found that data collected from local programs often had gaps or was inaccurate (see also Kutner et al.,

1996).

This problem may be related to several factors. First, learner attendance in adult literacy programs has traditionally been poor. Many barriers to attendance exist, such as the need for childcare and transportation, and the demands of work (Merrifield, 1998; Comings, Parrella, & Soricone, 1999), and poor attendance makes it difficult in some cases to give assessments and collect data. Second, adult literacy program staff have traditionally had limited expertise. In 1995, 80 percent of staff were part-time, 60 percent of programs had no full-time staff, and many staff were volunteers (General Accounting Office, 1995; Stein, 1997). Finally, program resources have traditionally been inadequate and may not be capable of supporting the training and monitoring activities necessary for reliable data collection practices (Beder, 1999; Stein, 1997; Merrifield, 1998; Sticht, 1998).

Whatever the causes or reasons, large-scale, independent evaluations of adult literacy assessment over the past ten years have consistently found that assessment practices are frequently haphazard and ineffective. As noted, assessment-particularly the use of standardized approaches to assessment, such as standardized tests-has simply not been a priority. Standardized tests were often chosen not for instructional purposes but for ease of administration (Kutner et al., 1996). When pilot-testing a management information system for adult literacy providers, Condelli and colleagues learned from a user survey that providers liked the system's ability to generate government reports automatically but resisted collecting assessment data and entering it into the system (Condelli et al., 1999).

THE ADULT EDUCATION AND FAMILY LITERACY ACT (AEFLA) OF 1998. In the wake of the evaluations discussed in

preceding sections, the most recent federal adult literacy legislation, implemented in 1998, again attempts to strengthen accountability through the use of more uniform performance standards. In addition, the adult education and family literacy act (Title II of the Workforce Investment Act, 1998, PL 105-220) provides for funding incentives for states based on the performance of their adult literacy programs, a form of high-stakes assessment. The AEFLA expects all states, in turn, to base funding decisions for local programs at least in part on their performance.

Performance measures to be used for accountability include "(i) demonstrated improvements in reading, writing, and speaking the English language, numeracy, problem solving, English language acquisition, and other literacy skills, (ii) placement in, retention in, or completion of, postsecondary education, training, unsubsidized employment or

career advancement, and (iii) receipt of a secondary school diploma or its recognized equivalent." States may add their own measures but are required to use these. Levels of performance must be "expressed in objective, quantifiable, and measurable form" in order to show progress. This information must be reported to the Department of Education and made public; a state-by-state comparison of assessment results must be compiled and disseminated by the DOE (Title II, Chapter 1, Sec. 212 of the Workforce Investment Act). All local programs are required to use these measures.

Specific guidelines for measures to be used in assessing adult learners, recording assessment results, and reporting the results through a computer-based system are provided by the National Reporting System (NRS), implemented in the summer of 2000 (DAEL, 2000; Garner, 1999).

Criticism of the generally poor state of assessment procedures in adult literacy and the need for "uniform valid and reliable data" to evaluate program effectiveness were major factors in developing the NRS (DAEL, 2000, p. 2). The NRS provides states with specific guidance on the types of standards, measures, and collection procedures that must be used by adult literacy providers accepting state and federal funds. It also provides states with technical and training assistance to support data collection and reporting procedures.

Gain in reading or writing ability is a key measure, "probably the most important single measure in the NRS" (DAEL, 2000, p. 38). Every adult entering a program must be pretested to determine a beginning literacy level and posttested before leaving to determine gain. Programs may use either standardized tests (norm- or criterion-referenced) or performance assessments with

standardized scoring rubrics. Although any state-approved assessment may be used for determining beginning and ending levels, each ABE student must be placed in one of six basic education levels defined by the NRS. The first four levels cover, roughly, literacy development through the beginning of secondary education: beginning ABE literacy, beginning basic education, low intermediate basic education, and high intermediate basic education. The last two levels cover adult secondary education: low adult secondary education and high adult secondary education.

Performance standards, or entry-level descriptors, are given for each level. These descriptions of what an adult at each level is expected to be able to do are keyed to scores from common standardized literacy tests. Performance standards for six ESOL levels are also provided. States report to the federal

government the number and percent of learners who advance one or more levels.

In summary, a wide variety of basic reading and writing assessment instruments may be used by states and local programs as long as they are either standardized norm- or criterion-referenced tests or performance assessments with standardized scoring rubrics. Results from these assessments are not reported directly but are translated into the NRS literacy levels, which are then used for reporting purposes.

In addition to basic reading and writing abilities, specific literacy contexts are highlighted in the AEFLA. As in earlier federal legislation, the overall goal of the AEFLA is to increase adults' self-sufficiency and functional literacy. As part of the WIA, however, a greater emphasis is placed on workplace literacy. Along with performance

standards for basic reading and writing, the NRS also describes performance standards for numeracy and for functional and workplace skills in terms of reading and writing. Follow-up measures related to employment, collected after students have left a program, are also required (whether a former learner has entered employment, retained employment, entered postsecondary education, or obtained a General Educational Development [GED] credential or diploma).

Secondary measures, recommended but not required, are also specified for family literacy programs. These measure progress toward the goal of assisting parents in obtaining the skills necessary to be full partners in their children's educational development. These are measures of literacy practices, such as the frequency of helping children with schoolwork and the number of contacts with teachers (measures of involvement

in children's education), and the frequency of reading to children, visits to the library, and book purchases (measures of involvement in children's literacy activities).

The use of uniform measures across states and a uniform, computer-based system for collecting data is designed to increase the validity and usefulness of data collected to evaluate the effectiveness of adult literacy legislation and funding. Additional procedures are recommended by the NRS to improve the validity (and reliability) of the data collected. These include staff development activities for local teachers, volunteers, and other staff; Web-based resources; organized and concrete data-handling procedures; increased resources for data collection; ongoing monitoring of data collection and recording; and formal audits of local program data.

Assessment for Instruction
As discussed, much of the literature

related to assessment in adult literacy has been dominated by large-scale, national assessments of functional literacy. Although federal and state legislation have attempted to shape the use of assessment, it has focused primarily on accountability. The use of assessment in adult literacy instruction, presumably the primary function of assessment, has not been studied in detail.

No national survey or observational studies of local programs' use of assessment for instruction exist. Therefore, it is not possible to determine how closely local practices approximate assessment models described by experts such as Askov (Askov et al., 1997), Chall (1994), and others. State accountability assessment plans, required by national legislation over the past decade, are currently in a state of flux because of new regulations (the AEFLA of 1998 and the NRS). However, past analyses of state plans do provide some

very general information about how reading and writing abilities are assessed in local ABE programs. Some information about the role of the other two major dimensions of literacy, context and practices, is also available. These three dimensions of literacy will frame the discussion of assessment for instruction that follows. Ability and literacy practices will be discussed separately, while literacy context will be discussed in relation to each.

ABILITIES: ASSESSING ASPECTS OF THE READING AND WRITING PROCESS

Until recently, many local adult literacy programs did not pretest incoming learners' reading and writing abilities. According to one survey conducted in the early 1990s (Beder, 1999; Young, Fitzgerald, & Fleischman, 1994), more than one-third did not pretest. Lack of pretesting abrogates the use of the assessment models described earlier (Askov et al., 1997; Chall, 1994):

pretesting to determine strengths and weaknesses in terms of components, use of pretest results to formulate a plan for instruction, ongoing assessment to adjust instruction as needed, and posttesting to measure the effects of instruction.

Most programs, however, have regularly pretested learners, using standardized tests, locally developed measures, or a combination of the two. With implementation of the NRS, all programs are now required to pretest, although not necessarily for instructional purposes. Of the standardized tests that state and local programs report using, the Test of Adult Basic Education (TABE) has consistently been used by more adult literacy programs than any other standardized test (Ehringhaus, 1991; Kutner et al., 1996; Beder, 1999). Reports of its use vary from about 70 percent to 80 percent among programs that use assessment regularly (Kutner et al., 1996; Beder, 1999). The Adult Basic

Learning Examination (ABLE) and Wide Range Achievement Test (WRAT) are the next most frequently used tests (reports of 20 percent), followed by the Comprehensive Adult Student Assessment System (CASAS) (14 percent) and the Slosson Oral Reading Test (SORT) (12 percent) (Beder, 1999).

Assuming these assessments are used for instruction, how well do they measure learner reading and writing processes, and how useful are they in an assessment model that incorporates the concepts of instructional components and developmental levels? To help answer these questions, each component (such as comprehension, vocabulary, fluency, and so on) is discussed in terms of the way it is, and could be, used in adult literacy programs. This part of the chapter is organized as follows:

- For each component, norm-referenced, criterion-referenced, and performance or informal assessments are discussed.
- The most commonly used ABE assessment

instruments are presented first.

- Other ABE assessments and non-ABE assessments are presented when a component is not addressed by one of the common tests, or when another test suggests a significant alternative approach to assessment. Table 4.1 provides a summary of the components measured across tests. All the tests discussed (common ABE assessments, other ABE assessments, and non-ABE assessments) are listed in Table 4.2, categorized by type.

Specific issues related to the literacy context addressed by a test and test reliability and validity are discussed when a test is first mentioned, and then as needed. This includes the types of scores offered by an assessment instrument and the norming group on which the scores are based. The list of tests in Tables 4.1 and 4.2 is not meant to be exhaustive, and many tests that may be just as appropriate to use with adults as those listed are not included. To name just a few, the Test of Applied Literacy Skills (TALS) (Educational Testing Service, 1991), the Peabody Picture Vocabulary Test (PPVT) (Dunn

& Dunn, 1997), the Test of Word Reading Efficiency (TOWRE) (Torgesen, Wagner, & Rashotte, 1999), and the Comprehensive Test of Phonological Processing (CTOPP) (Wagner, Torgesen, & Rashotte, 1999) are not discussed.

Reading

Assessment of the following components of the reading process will be discussed: reading comprehension, vocabulary, fluency (reading accuracy and rate), word recognition, and word analysis.

READING COMPREHENSION. Reading comprehension assessment measures students' ability to understand or generate meaning from a text that is read. This aspect of the reading process is what most people associate with literacy or reading ability.

<u>Norm-Referenced Assessment.</u> The TABE (CTB/McGraw-Hill, 1987, 1994a, 1994b),1 ABLE (Karlsen & Gardner, 1986), and CASAS (CASAS, 1989) all

measure reading comprehension by asking students to answer multiple-choice questions about what they have read. The ABLE and CASAS also include a number of cloze, or fill-in-the blank, items. Although the CASAS reading test includes some word analysis test items, most are comprehension items. All three tests are normed on representative samples of adults from various settings. The TABE and ABLE provide separate norm-referenced scores (percentile ranks and so on) for some of these groups (vocational-technical programs, prisons, ABE programs, and others).

The tests' content reflects the literacy contexts represented. All of these tests contain adult-oriented reading material, a mix of material from educational, daily life, and employment-related contexts. The reading passages in the ABLE seem to contain more academic passages, or what might be expected in a K-12 school

context, such as literature (for example, fiction and poetry) and content-oriented material (for example, science, social studies, or history). The TABE has somewhat more reading related to daily life and employment than the ABLE. In addition to passages from works of fiction and factual passages about topics such as boats, it also has students read and respond to advertisements, letters, and passages of dialogue. Alternate versions of the TABE are also available that focus on any one of four work contexts: health, business, trade, and general occupational (CTB/McGraw-Hill, 1994a, 1994b). These versions, however, are available only for more advanced ABE learners.

At the other end of the continuum, perhaps, is the CASAS system, which includes two CASAS tests, the Beginning Literacy Reading Assessment portion of the Life Skills Assessment and the Reading portion of the Basic Skills

for Assessment in Employability. As their names suggest, they almost exclusively address daily life and employment-related contexts, respectively. The CASAS Life Skills assessment, for example, has students read ads, price tags, restaurant menus, food labels, medical forms, and passages about legal issues and community services. The CASAS assessments also contain what, in the NALS framework, are quantitative items, displays such as graphs and other items that require numerical computations or numeracy skills.

The skills measured on the TABE and ABLE were obtained by examining ABE curriculum guides, texts, instructional programs, and objectives from other achievement tests. The CASAS system is based on more than three hundred competencies related to the Secretary's Commission on Achieving Necessary Skills (SCANS) competencies (1991)

identified by the U.S. Department of Labor to help apply teaching and learning in a real-world context. One competency, for example, is described as the ability to "interpret advertisements, labels, or charts to select goods and services."

These tests actually consist of three or four separate tests, called levels, each at a different level of difficulty. A short "locator" test is given to determine which level a learner takes. All three tests are able to measure gain through the use of scale scores, which provide a single numerical scale that covers all levels of a test. The TABE and ABLE, but not the CASAS, provide reliability data in their manuals.

Criterion-Referenced Assessment. In addition to being norm-referenced, the CASAS could also be considered a criterion-referenced test. Test questions are keyed to its list of SCANS-based competencies, and the competencies are

keyed to suggested instructional material. Instructors may examine a learner's individual responses on a test, note which items are missed, and provide instruction in the corresponding competency. In addition, each scale score is keyed to one of four presecondary ABE literacy levels: Beginning/PreLiteracy, Beginning Basic Skills, Intermediate Basic Skills, and Advanced Basic Skills. Two secondary literacy levels are also described: Adult Secondary and Advanced Adult Secondary. These are all similar to the NRS entry-level performance standards for ABE.

The TABE and ABLE provide criterion referencing in the form of mastery levels for reading comprehension subskills such as Event Interpretation, Main Ideas, and Details. Mastery levels and criteria for establishing mastery levels, however, are poorly defined. The ABLE reports the average number of items correct for

the norm group in each subskill within a level. Each level may span several grade or ability levels. The examiner must decide whether or not this constitutes mastery of a subskill. The TABE provides a three-level index of mastery for each subskill (Not Mastered, Partial Mastery, and Mastery) based on number correct, but it does not provide a rationale for each cutoff score. In addition, the ABLE and TABE mastery levels are based on a relatively small number of test items, too few to be truly useful for placement and instruction. Both should be considered informal, as opposed to criterion, measures.

Any norm-referenced test, such as the TABE, ABLE, and CASAS, may be used as an informal, criterion-referenced test as an examiner becomes familiar with its content and comes to understand how various norm-referenced scores reflect actual reading behavior (Joint Committee on Standards, 1999). The

NRS and the CASAS system, for example, key scale scores to entry-level performance standards for specific ABE reading levels. Although the NRS does not endorse the TABE or vouch for the validity of its scale scores, it does suggest that a TABE reading scale score between 542 and 679 is associated with the Beginning Basic Education level, which is described as follows: "Individual can read and print numbers and letters, but has a limited understanding of connected prose and may need frequent rereading" (see DAEL, 2000, p. 14, for the full description).

Particularly for the inexperienced examiner, the scores 542 and 679 may seem arbitrary and difficult to interpret. Relating scale scores to performance standards simulates what a TABE examiner may come to know only after extensive experience in using the test and providing a wide range of students

with instruction.

The perceived benefit of Grade Level Equivalent scores, which associate raw scores with the abilities of average students at various grade levels, is that unlike scale scores, they intuitively make sense. Unfortunately, they often make too much sense-the concept of grade-in-school is so familiar that inexperienced examiners may easily misinterpret Grade Equivalents (GEs). GEs, like scale scores, are not derived consistently across tests, so GEs from different tests may have different implications. The TABE derived its GEs by equating TABE scores with California Achievement Test scores, for example. The ABLE GEs were formulated by giving the ABLE to a sample of elementary and secondary school students. The CASAS determined GEs simply by asking students who took the CASAS how many grade levels in school they had completed.

Although norm-referenced GEs are usually reported in terms of years and months (for example, 7.6), average scores for students are actually obtained at only one or two points during the year. Additional points along the GE continuum, as with scale scores, are determined through extrapolation and interpolation. GEs, then, illustrate that the interpretation of norm-referenced tests for criterion-referenced purposes requires a fairly high degree of expertise in both testing and teaching.

In addition to using norm-referenced tests for criterion-based decisions, there are several tests that were constructed as criterion-referenced tests that may be used in adult literacy programs. Two of these are the Reading Evaluation Adult Diagnosis (READ) (Colvin & Root, 1982) and the Diagnostic Assessments of Reading (DAR) (Roswell & Chall, 1992). Both are similar to informal

reading inventories, tests that measure oral reading and reading comprehension by having students read and answer questions about passages written at different levels of difficulty. The levels of difficulty usually correspond to school grade levels or Grade Equivalents.

Both use a simple form of adaptive testing. In adaptive testing, items are first ordered according to difficulty. Both the READ and the DAR, for example, contain reading comprehension test passages beginning at, roughly, GE 3 and continuing through successive grade levels to GE 12. The examiner finds the highest level passage at which a student exhibits mastery (answers a specified number of questions correctly, for example). The level of this passage (somewhere between GE 3 and GE 12) is the student's score. A unique feature of adaptive testing is that the learner need not respond to all of the test items, saving time and perhaps avoiding

frustration.

The READ, developed for Literacy Volunteers of America (LVA), contains passages that represent a more adult-oriented context than the DAR. The DAR, however, is a more reliable test. The level of the DAR content, corresponding to school grade levels, was validated on a large, national sample of students at various grade levels. No validity checks for the READ are reported.

The Test of Functional Health Literacy in Adults (TOFHLA) (Nurss et al., 1995) is an example of a criterion-referenced assessment that focuses on a specific, fairly narrow context. Adults' ability to read and understand health-related texts (X-ray preparation, Medicaid rights, and a consent form) is measured using a multiple-choice cloze passage. The cloze score is combined with the score on a separate numeracy test, and this

combined score is used to place the learner at one of three literacy levels (low, marginal, and adequate functional health literacy). Reliability coefficients are presented in the TOFHLA manual. Performance Assessment. Performance assessment includes the evaluation of student portfolios, demonstrations, projects, oral retellings, and other alternatives to norm-referenced and criterion-referenced test content. Because performance assessments may include tasks normally used for instructional purposes, they have the potential to link instruction directly to assessment (Fingeret, 1993; Leipzig & Afflerbach, 2000; Padak, Davidson, & Padak, 1994).

Although it is conceivable that a performance assessment could focus narrowly on only one aspect of the reading process, most view performance assessments as situated, holistic evaluations, in contrast with tests that

focus on specific parts, aspects, or components of reading and writing processes (Garcia & Pearson, 1991; Paris, Calfee, Filby, Hiebert, Pearson, Valencia, & Wolf, 1999). Most performance assessments, then, measure more global skills, such as reading comprehension.

Performance assessments are considered by many to be a more valid measure of the domain of reading comprehension behaviors (for example, Padak et al., 1994). They are able to measure metacomprehension abilities such as strategy use and comprehension monitoring, and they use an extended, constructed response mode as opposed to multiple-choice or short-answer formats (Martinez, 1999). The fact that they use an extended response mode, however, also makes it more difficult for performance assessments to establish consistent scoring procedures. Perhaps because performance assessment is just

beginning to be used extensively, procedures for establishing and measuring reliability are not well developed (Merrifield, 1998; Leipzig & Afflerbach, 2000).

Performance assessments may be used to assess reading and writing ability to satisfy assessment requirements in the NRS. It is not clear yet, however, how many states and local programs will actually use performance assessments or what form they will take as they incorporate the concepts of ability levels and standardized scoring rubrics that are a part of the NRS. As the NRS has evolved over the past decade, however, at least some states have developed performance assessments. A review of eleven states' ABE literacy assessment systems, based on interviews with state officials and published state plans, showed that at least five were adopting published performance assessments (Kutner et al., 1996). Although very little

detail was provided, these included learner portfolios, writing samples, classroom demonstrations, and reading aloud, as well as documentation of specific practices.

Performance tasks such as project-based learning have been used by local programs for some time, although scoring rubrics and other ways of evaluating student learning gains have lagged behind (for example, see Wrigley, 1998). As performance tasks and related scoring rubrics evolve, more detailed descriptions of existing performance assessment systems, including direct observations, will be needed, as will research related to reliability and validity.

Informal Assessment. The Tests of General Educational Development (GED) are administered by the American Council on Education (American Council on Education, 1993) and adults who pass the GED receive a high-

school-level educational diploma. Many programs for advanced ABE learners base their curriculum on the GED. Although local programs cannot administer the actual GED, GED practice tests are available as informal measures (GED Testing Service, 1997). Test content consists of passages, as well as charts, tables, and other graphics, that cover high school subject matter at the twelfth grade level, including social studies and science, and literature and the arts. Test takers read the passages and answer multiple-choice questions. Although norm tables with percentile ranks and scale scores are provided (based on a national sample of high school seniors) the results are informal because the norms are based on standard administration procedures in official testing centers. As an informal assessment of reading comprehension, the GED practice tests would be suitable for learners who are reading at the high school level or above.

VOCABULARY (ORAL). Vocabulary, or knowledge of word meanings, may be measured either orally or silently. in oral measures, students hear a word and tell what it means, choose the correct illustration of the word, or choose the correct orally presented definition. In silent measures, students must read silently and answer questions about a word. "Silent vocabulary" will be discussed in more detail in the next section.

Norm-Referenced Assessment. Of the commonly used adult literacy assessment instruments (the TABE, ABLE, CASAS, WRAT, and SORT), only the ABLE assesses oral vocabulary, and it does so only for adults who take the lowest level of the test (level 1, for adults with one to four years of formal schooling). Sentences are dictated to students, who must decide which of three alternatives best completes each sentence. A multiple-choice vocabulary item assessing knowledge of a word such as

foot might ask, A foot is made up of . . . 12 inches, 3 inches, 8 inches, with the student marking the correct answer on an answer form. The context represented by test items is roughly the same as in the ABLE reading comprehension assessment; words are drawn from work situations, daily life, and academic texts, with most being from the social, physical, and natural sciences.

Tests not commonly used by adult literacy programs that measure oral vocabulary at all levels are available. The Woodcock-Johnson Diagnostic Reading Battery (Woodcock, 1997) measures both oral and silent vocabulary. Students' oral vocabulary is measured by having them listen to a tape on which they hear a word, then respond with a one-word antonym or synonym.

The Woodcock is normed on children and adults ranging from two to ninety-five years old and provides scale scores,

percentile ranks, age equivalents, grade equivalents, and a mastery score. Extensive data on the test's reliability is provided. GEs are determined by obtaining the average scores for the norm group in a given grade level during each month of the school year, without extrapolating or interpolating. Like the DAR, the Woodcock uses adaptive testing and does not have separate tests for students at different levels of ability (as the TABE, ABLE, and CASAS do). The Woodcock is also different from the TABE and ABLE in that it does not provide an overtly adult context, using more school-like content.

Criterion-Referenced Assessment. As with reading comprehension, the ABLE provides criterion-referenced scores (mastery levels) for vocabulary knowledge and support for item analysis. The same problems related to validity exist for these measures that were discussed above for the ABLE criterion-referenced reading comprehension

measures.

Of the criterion-referenced tests not commonly used in adult literacy, the DAR measures vocabulary by asking learners to define words rather than using a multiple-choice or short answer format. While this makes the test more difficult to score, it may be a more valid measure of vocabulary knowledge. As with reading comprehension, the DAR vocabulary subtest yields a "validated" GE score.

The Woodcock mastery score (Relative Proficiency Index) is based on one of its norm-referenced scale scores. It indicates what percentage of material a test taker would be expected to know when compared with individuals at the same age (or grade) level. Unlike the ABLE, it uses a sufficient number of items overall for reliable indexes of mastery.

VOCABULARY (SILENT). Silent vocabulary assessment, in which

students silently read and answer questions about a word and its meaning, is not as pure a measure of vocabulary as oral assessment. When the learner must read the questions silently, vocabulary knowledge is confounded with other aspects of reading ability, such as decoding. This measure may not be as valid as an oral measure, but it is easier to administer and score.

Norm-Referenced Assessment. The ABLE uses the same item format for silent vocabulary assessment on levels 2 and 3 (for students with five to eight and nine to twelve years of school) as it does for oral vocabulary on level 1, except that the items are read silently by the student, not dictated by a teacher. The CASAS and most recent TABE (CTB/McGraw-Hill, 1994a, 1994b) do not have a separate subtest for vocabulary knowledge. The 1987 TABE does have a separate vocabulary assessment test. It contains multiple-

choice items that are a little more varied than those on the ABLE. In addition to asking students to complete a sentence with the correct word, as on the ABLE, the TABE asks students to find synonyms and antonyms for an underlined word or word part (such as a prefix) in a phrase or sentence. The student reads a phrase (such as Over the mountain) and must select from a list of four options the word with the same meaning as the underlined word (such as above, below, near, through). As with the TABE reading comprehension subtest, the 1987 TABE vocabulary assessment emphasizes functional contexts (life and work-related contexts) somewhat more than the ABLE.

Some tests provide both oral and silent vocabulary measures. The Woodcock, for example, has separate oral and silent vocabulary subtests, and the norm-referenced scores are directly comparable.

Criterion-Referenced and Informal Assessment. The ABLE and the 1987 TABE provide support for item analysis and mastery cutoffs for total vocabulary scores and vocabulary subskill scores. Again, problems with these criterion levels are the same as those discussed earlier. The Woodcock mastery score is more robust because it is based on more items and is referenced to the performance of the norm group.

FLUENCY. Assessments of reading fluency measure learners' ability to read connected text accurately, at a reasonable rate, and with appropriate prosody (intonation and phrasing).

Norm-Referenced Assessment. The TABE, ABLE, and CASAS do not measure oral reading fluency directly, nor do they provide a score for reading fluency. The TABE, because it is timed, penalizes those test takers who read

slowly.

An example of a norm-referenced test that does measure reading fluency is the Gray Oral Reading Test (GORT), now in its third edition (Weiderholt & Bryant, 1992). The GORT is normed on students in grades 2 through 12, not on adults. On this individually administered, adaptive test, students are asked to read aloud from short passages that become progressively more difficult (sentences increase in length and complexity, and vocabulary increases in difficulty). As learners read, their oral reading errors and the time that it takes to read a passage are recorded. Three separate scores-one for rate, one for accuracy, and a total score-may be converted into percentiles, scale scores, and grade equivalents. Student miscues (reading errors such as mispronunciations, omissions, repetitions, and self-corrections) may be analyzed qualitatively to look for patterns in the

errors. A low rate score accompanied by many self-corrections, for example, might be interpreted differently from a low rate score accompanied by many omissions and mispronunciations. The GORT manual suggests using the procedures described by Goodman and Burke (1972) to conduct the error analysis.

Criterion-Referenced Assessment. The DAR is an example of a criterion-referenced assessment that measures fluency in oral reading. Like the GORT, the DAR oral reading subtest is an adaptive test. A student's score indicates the highest level passage that is mastered, with passages spanning GE 1 through GE 11-12. Mastery is defined as pronouncing roughly 95 percent of the words in a passage correctly, as in traditional Informal Reading Inventories (IRIs). As mentioned earlier in the comprehension section, the DAR is similar to IRIs, but differs in the care

taken to establish content validity. The difficulty and grade placement levels of oral reading passages are based on readability measures, experts' judgments, and two research studies in which the passages were given to a wide range of students of different ability levels to verify that student scores on the passages accurately differentiated students at different grade levels and were adequately correlated with a norm-referenced test.

Performance Assessment. Oral reading is a natural performance task that usually involves the analysis of oral reading errors. Miscue analysis is one example of methods used to analyze these errors (Goodman, 1999; Goodman, Watson, & Burke, 1987; Leipzig & Afflerbach, 2000). With this method, miscues are not treated as errors but are evaluated in terms of whether or not they maintain a text's syntactic and semantic integrity. Other informal assessments, such as

IRIs, look at the number and type of errors (mispronunciations, self-corrections, and so on).

WORD RECOGNITION. Word recognition assessments measure students' ability to pronounce individual words presented in isolation. Students may read a word list rather than a passage of text, for example.

Norm-Referenced Assessment. The ABLE, TABE, and CASAS do not have separate word recognition subtests. The WRAT (Wilkinson, 1993) and SORT (Slosson & Nicholson, 1990), although not used as frequently in adult literacy programs, do measure isolated word recognition.

On the WRAT, students are asked to read aloud a list of words that increase in difficulty (cat and red are at the beginning, for example, and disingenuous and inefficacious are at the

end). The test ends when the student either is unable to pronounce ten consecutive words or gets to the end of the list. On the SORT, an adaptive test, students are also asked to read isolated words, but they are asked to pronounce words from lists ranging in difficulty from the primary level through high school. Both are normed on children and adults, and both provide norm-referenced scores derived from raw scores. The WRAT and SORT GE scores are not interpolated or extrapolated scores (there is no need for this because all learners take the same test, there are no separate levels, and an adequate sample is drawn from each grade level). Manuals for both provide reliability data.

Criterion-Referenced Assessment. Although the CASAS does not give a separate score for word recognition, it does include "discrimination among sight words" as one of its reading comprehension objectives, and several

items on the reading comprehension subtest address sight words directly. Item analysis support is given on the CASAS for sight words, although this measure has the same problems as the TABE and ABLE mastery measures discussed earlier.

The DAR and READ also use word lists as a part of their criterion- referenced assessments. The student's score for word recognition is the grade level of the most difficult list on which mastery is exhibited. On these adaptive tests, mastery is determined by the percentage of words pronounced correctly-70 percent of the words on a DAR list, for example. Although the DAR provides information on validity, the READ manual contains none.

The Rapid Estimate of Adult Literacy in Medicine (REALM) (Murphy, Davis, Long, Jackson, & Decker, 1994) is an example of a context-specific word

recognition assessment. Modeled after the WRAT and SORT, the REALM consists of three word lists containing medical terms (for example, flu, infection, osteoporosis). Raw scores are converted into GEs that are anchored to descriptions of patients' abilities to read medical-related texts. The GEs were obtained by correlating the REALM with the SORT.

WORD ANALYSIS. Word analysis assessment measures students' ability to recognize, produce, and manipulate individual phonemes or speech sounds (phonemic awareness) in words or syllables that they hear. It also measures their ability to match sounds with letters and letter combinations (their knowledge of letter-sounds correspondences) and to blend letter-sounds into words while reading or spelling (phonics ability). Higher-level word analysis assessment measures students' knowledge of meaningful word-parts, such as

compounds, prefixes, and suffixes.

Norm-Referenced Assessment. None of the norm-referenced tests commonly used in adult literacy give a separate score for word analysis. Several other norm-referenced tests, however, do measure aspects of phonemic awareness and phonics. The Woodcock looks at students' ability to sound out words (in the Word Attack subtest), to supply missing phonemes in words (Incomplete Words), and to blend isolated sounds into words (Sound Blending). The ability to sound out words is measured by having students read phonologically regular nonsense words. The nonsense word stad, for example, can be pronounced (it rhymes with had) even though it is not a real word. Using nonsense words ensures that students are actually sounding out a word as opposed to saying a word that they have memorized (as they might have memorized irregular words such as

enough and though). Scores on the Incomplete Words and Sound Blending tests can be combined to obtain an overall Phonological Awareness score. On these subtests, students listen to a word with one or two missing phonemes (si__ter) and must say the complete word (sister), or they are asked to listen to the individual parts of a word (c-a-t or b-at) and then must say the word (cat and bat).

Criterion-Referenced Assessment. The TABE, ABLE, CASAS, and WRAT do have items that test for knowledge of certain word analysis or phonemic awareness skills when measuring some more inclusive component of the reading process. The WRAT, for example, which measures sight word knowledge, asks beginning readers to read isolated upper- and lowercase letters. The easiest level of the TABE has items for matching and recognizing letters and for identifying beginning, middle, and end sounds in words. The CASAS reading

comprehension subtest includes items that measure the ability to recognize and discriminate upper- and lowercase letters. The 1987 TABE vocabulary subtest tests knowledge of affixes, and both the TABE and ABLE spelling subtests test knowledge of various word parts, such as affixes, vowels, consonants, and vowel digraphs. The TABE, ABLE, and CASAS support item analysis and provide mastery level scores, but, as said before, these scores are not easily interpreted.

The Adult Measure of Essential Skills (AMES) (Steck-Vaughn, 1997) is a newer test of adult literacy that, like the TABE and ABLE, is normed on groups of adults, places questions in an adult context (home, community, workplace, and school), and has forms at different levels for adults who are at different levels of literacy development. Unlike the TABE and ABLE, however, it has a separate auditory discrimination subtest

for nonreaders (for those "who have had from one to two years of schooling"). Students are asked to find words that have the same sound as a stimulus word in beginning, medial, and ending positions. A student might be shown three pictures (of a house, cat, and dog, for example) and then asked by the examiner to locate the one that begins with the same sound as a word pronounced by the examiner (such as hat). Unfortunately, this subtest has not been separately normed, and results are combined with other subtest results to obtain a norm-referenced total reading score.

The Test of Auditory Analysis Skills (TAAS) (Rosner, 1979) is an example of a criterion-referenced word analysis test of phonological awareness. It is a short, thirteen-item test that measures the ability to manipulate phonemes and syllables by asking students to say a word after removing a phoneme or

syllable. The student might be asked to say boat without the /b/ sound, for example (oat). It gives a GE score from kindergarten through grade 3, based on the number of correct items.

The Word Analysis assessment of the DAR measures a student's knowledge of letter-sound correspondences using a series of twelve subtests that correspond roughly to the order in which word analysis skills are introduced to, or learned by, beginning readers: matching words, matching letters, naming lowercase and uppercase letters (pre-reading subtests), consonant sounds, consonant blends, short vowel sounds, rule of silent e, vowel digraphs, diphthongs, vowels with r, and polysyllabic words. These tests of basic word analysis ability are given only to those students who score below the fourth-grade level on the DAR Word Recognition subtest. It is assumed that those scoring above level 3 on Word

Recognition will have mastered basic word analysis skills.

Mastery levels on the twelve DAR Word Analysis subtests are determined individually for each subtest, based on the number of correct responses. GE scores are not provided for these subtests. In addition to simple matching and naming tasks for the pre-reading subtests, students are asked to say the sounds of individual consonants when presented with the letter that represents the sound or to read words containing specific letter-sound correspondences. To assess knowledge of vowel digraphs (two-letter vowel combinations that represent one sound), for example, a student might be asked to read a word list including the word seat to assess knowledge of the sound that the digraph ea makes. Correct answers, indicating that the student knows the ea sound, would include any one-syllable word with a medial /e/ sound, such as beat or

seam, as well as seat.

Several informal tests of basic word analysis are also available, such as Adams's Test of Phonemic Awareness (Adams, 1998).

Writing

Assessment of the following components of the writing process will be discussed: the production of written products, writing vocabulary, sentence production, word production, planning and monitoring, and revising and editing.

WRITTEN PRODUCTS. Essays, reports, stories, and other written products produced in response to a task have increasingly been scored holistically. Readers rate a written product on a scale (usually an ascending four- or five-point scale) using guidelines that describe the characteristics of products at each point along the scale. Analytic scoring is also used, where several specific traits, such

as style or mechanics, are scored separately. This form of performance assessment, consisting of a writing task and scoring rubric, is one of the few that have led to norm-referenced performance assessments, what Nitko (1996) calls structured, on-demand performance assessments.

Norm-Referenced Assessment. None of the most commonly used assessments in adult literacy provide norm-referenced scores for whole written products. There are, however, other norm-referenced writing assessments that do evaluate student essays, stories, and descriptions. The Writing Process Test (WPT) (Warden & Hutchinson, 1992), for example, is a group-administered, structured performance assessment that uses an analytic scoring procedure to evaluate a student's composition. Students are given an academic or school-like writing task, such as writing a personal essay for a school newspaper.

Their response, a written product, is evaluated analytically by giving a score for ten features: purpose and focus, audience, vocabulary, style and tone, support and development, organization and coherence, sentence structure and variety, grammar and usage, capitalization and punctuation, and spelling. These scores are summed to produce a total score, which can then be converted into one of several standardized scores, including percentile ranks and two scale scores.

Although the norming group for this test includes students in grades K-12 and not adults, the manual suggests that the test can be used in ABE settings and can also be used to evaluate any writing assignment. When used with adults, then, it becomes an informal criterion-referenced test because there are no adult norms.

A more varied norm-referenced writing

assessment is the Test of Written Language (TOWL) (Hammill & Larsen, 1996). Its measure of story writing uses an analytic scoring rubric that considers aspects of both sentence-level and story-level production. Like the Writing Process Test, it was normed on a K-12 population and its content reflects an academic context. In addition to measuring whole, written products, it also measures five components of the writing process with a more traditional, short-answer format.

Scoring a test that uses an analytic scoring rubric requires more training than is required for scoring a multiple-choice test. Analytic scoring is more subjective than multiple-choice scoring, and the reliability coefficients reported in the Writing Process Test manuals are generally somewhat lower than those reported for the TABE or ABLE. When judging an entire essay, the evaluator may find many different responses to be

"correct." As the use of a five-point scoring scale suggests, several types of responses may fall between "correct" and "not correct." When judging whether a student has "used techniques to engage a reader," for example, a scorer selects one of five responses (sophisticated, competent, partly competent, not yet competent, and problematic). These are keyed to more detailed criteria. For a piece to be judged sophisticated, for example, it must be one that "uses techniques (e.g., questions, humor, direct address, references to audience) effectively to engage [the] audience throughout the writing." A competent piece, on the other hand, shows "some evidence of techniques to engage the audience, but not all are effective; a clear effort is made once but not carried throughout."

Performance Assessment. Of common adult literacy tests, only the CASAS Functional Writing Test gives scores for

complete written products. The CASAS writing test is a structured, on-demand performance assessment (Nitko, 1996) that measures a learner's ability to produce one or more of three types of adult-oriented texts. The first text type is descriptive and is derived from a picture task. The test taker looks at a drawing of a street scene and writes about what is happening in the picture. The second text type is a form, such as an employment application, and the task for the learner is to complete it. The third type is a description of a common process depicted in a picture, such as obtaining money from an ATM.

The CASAS writing test provides both analytic and holistic writing scores. Analytic scoring is used for two of the tasks, the picture task and the form task. When scoring the description, for example, evaluators give it a score (from 0 to 6) in each of the following five categories: content; organization; word

choice; grammar and sentence structure; and punctuation, spelling, and capitalization. These scores are first weighted and then summed to yield a total that is used to place students at one of six writing ability levels, from a Beginning Literacy level through an Advanced level. A text produced for the Process Task is scored holistically, with evaluators assigning a single score on a scale from 0 to 5, using a scoring rubric that focuses on content, organization, word choice, and mechanics (grammar, spelling, and so on). CASAS will ensure the reliability of scores for the different tasks only if test administrators receive training from CASAS or the essays are sent to CASAS for scoring.

Informal Assessment. The GED practice tests (GED Testing Service, 1997) may be used informally to measure advanced ABE learners' written products (those ready to take the GED high school equivalency exam). Learners are given a

statement (about the effects of watching television, for example) and directions about what to write (a two-hundred-word essay on whether they agree with the statement, for example). The essays are scored holistically. Although scoring guides and sample essays are provided, the results are reliable only for trained examiners.

WRITING VOCABULARY. Although there is considerable overlap, writing vocabulary and reading vocabulary assessment may measure somewhat different abilities. While reading vocabulary assessment typically measures the ability to recognize or state the meaning of a word, writing vocabulary assessment measures how effectively a learner uses or produces a concept while composing.

Norm-Referenced Assessment. None of the tests commonly used in adult literacy provide norm-referenced measures of a

student's ability to use specific vocabulary words while writing. As described earlier, the TABE and ABLE do assess students' knowledge of word meanings, but the assessment requires reading, not writing. The nonadult TOWL is an example of a norm-referenced test for writing vocabulary. Students are asked to write sentences that contain specific vocabulary words. As with the WRAT word recognition reading test, students are given progressively more difficult words until they reach their ceiling, missing a specified number of words in a row (or completing the test). The test begins with words like see, eat, and help and ends with words like evade and inept. Percentiles, scale scores, and grade equivalents can be derived from students' raw scores.

Criterion-Referenced Assessment. The Writing Process Test provides a criterion-referenced score for the use of

words to communicate purpose and style. This vocabulary score is one of the ten analytic scores the test provides to evaluate a whole written product. The criteria range from sophisticated ("uses precise, fresh, vivid words to communicate purpose and style") to problematic (vocabulary is "inadequate, incorrect, or confusing") on a five-point scale.

The CASAS Functional Writing Test also contains an analytic score for vocabulary in which a student's choice of words is evaluated according to specified criteria. Although this is not a direct, controlled measure of writing vocabulary, the scoring rubric might be used as a guide for the assessment of vocabulary used in naturally occurring student writing, such as texts collected for a portfolio.

SENTENCE PRODUCTION: CAPITALIZATION, PUNCTUATION, AND SENTENCE STRUCTURE (SYNTAX AND

USAGE). None of the commonly used adult literacy assessments ask learners to actually write sentences. Sentence production ability is instead evaluated indirectly with measures of capitalization and punctuation knowledge (these are usually measured together) and knowledge of the structure of sentences (both grammar or syntax and conventional usage).

Norm-Referenced Assessment. Capitalization, punctuation, and sentence structure knowledge are all measured by the ABLE and TABE using a multiple-choice format. On the TABE, a Language score is obtained, in part, by asking students to read a sentence or passage and then select the best way to punctuate or capitalize a part of the selection from among four or five choices. For example, if a sentence such as It is I think very hot outside is given, the correct answer among the choices would be It is, I think, very hot outside

as opposed to It is I, think, very hot outside. The Language score on the TABE is also derived from responses to multiple-choice questions related to knowledge of English language usage, or phrase- and sentence-level syntactic structures, and paragraph development, or specific paragraph-level skills. For sentence-level structures, students are tested on their ability to recognize the correct use or form of basic syntactic structures: nouns, verbs, modifiers, and simple sentences. For paragraph-level structures, students are asked to recognize the best topic sentence, the best sequence for sentences in a paragraph, and the best way to combine two simple sentences into one, more complex, sentence.

Unlike the TABE, the ABLE measures only capitalization, punctuation, and sentence-level structures. Neither test measures these skills on the levels of the test designed for beginning readers (for

students with one to four years of schooling on the ABLE and for students at about GE 0-2 on the TABE). Norm-Referenced Performance Assessment. The Writing Process Test and the TOWL (normed on children) provide examples of ways in which extended learner-generated writing may be used to assess sentence production ability. The Writing Process Test contains a norm-referenced score for fluency that is derived from a combination of analytic scores for the following features: sentence structure and variety, grammar and usage, capitalization and punctuation, and spelling. The TOWL scores student stories analytically for both contextual conventions (spelling, capitalization, and punctuation) and contextual language (sentence construction, grammar, and quality of vocabulary). These measures of students' extended writing can be compared with students' scores on style and sentence-combining subtests that

require one-sentence responses to stimuli (dictated sentences or short sentences that are to be combined into one longer sentence).

The Woodcock-Johnson Tests of Achievement (Woodcock & Johnson, 1989) measure a learner's ability to write sentences, evaluating both the quality of the written product and fluency (speed) and then combining these measures into one written expression score. In one task used to generate sentences, the student is shown a picture and three words and then asked to write a sentence containing each of these elements as quickly as possible. The student might be shown, for example, a picture of a house and the words door, man, and knock and be expected to produce the man will knock on the door of the house (students are not penalized for capitalization and punctuation errors). The Woodcock-Johnson Achievement test gives the same types of scores (percentiles, scale

scores, age and grade equivalents, and mastery scores) as the Woodcock-Johnson Diagnostic Reading test discussed above. These tests, part of the Woodcock-Johnson Psycho-Educational Battery, are constructed so that standard scores from one may be compared with standard scores from the other. Reading results, for example, can be compared directly with writing results.

Criterion-Referenced Assessment. The CASAS Functional Writing Test contains a separate analytic score for grammar and sentence structure. The scoring rubric for grammar and sentence structure, like the rubric for writing vocabulary discussed earlier, might be used as a guide for evaluating sentence structure in naturally occurring written work.

WORD PRODUCTION: SPELLING (SOUND-LETTER CORRESPONDENCE) AND MORPHOLOGY (DERIVATION, INFLECTION, AND COMPOUNDS). Word

production ability is most often measured with spelling tests. Spelling tests may be used to test for specific subskill knowledge, such as knowledge of sound-letter correspondences, derivations, inflections, and compounds. Reading teachers sometimes use spelling tests to measure word analysis ability.

Norm-Referenced Assessment. The TABE and ABLE measure spelling by asking students to read a sentence with a missing word and to then choose the correct spelling of the missing word from among a short list of words. Although only one of the words on the list is spelled correctly, all might be confused with the correct spelling. The following is an example of this type of spelling item: The _____ is dry. strem, streme, stream, streem.

This example tests students' knowledge of vowel digraphs (the two-letter vowel combination ea). The TABE and ABLE

also use the spelling subtest to measure knowledge of consonant variants (consonant digraphs like the ph in phone, for example) and structural forms, such as contractions and affixes. The TABE level L, for adults reading at about GE 0-2, does not contain a spelling subtest. The sentences and possible responses for the level 1 on the ABLE (the level for adults with one to four years of schooling) are dictated.

The WRAT spelling subtest is also oral, although it is more like a traditional spelling test, in which the teacher dictates a word, uses the word in a sentence, and then directs students to write the correct spelling without the benefit of being able to select from among a list of possible answers. The TOWL spelling subtest, like most of the other TOWL subtests, requires extended writing. In this case, students write dictated sentences.

Criterion-Referenced Assessment. Both

the Writing Process Test and the TOWL have analytic scores that evaluate spelling in context (in a learner's story, for example). The Woodcock provides a mastery score for spelling.

PLANNING AND MONITORING. Planning includes what a student does before or during writing to generate ideas and organize them coherently, based on the writing task and intended audience. Creating a working outline for a written product is an example of a planning behavior. While writing, writers may monitor their composing to ensure that it conforms with their plans, to change plans, and to check spelling and other lower-level processes.

Norm-Referenced Assessment. None of the common adult literacy assessments measure planning ability in writing. The Writing Process Test does provide norm-referenced scores for development, derived from the sum of analytic scores

for purpose, audience, vocabulary, style and tone, support and development, and organization and coherence. Development is described as the ability to handle the broader concerns of topic, audience, and ideas, as opposed to fluency or the ability to handle the more mechanical aspects of writing (sentence structure, grammar, and so on). Criterion-Referenced Assessment. The CASAS Functional Writing Assessment provides two measures that are very roughly related to generating and organizing ideas: a measure of an essay's content, which is an overall assessment of the quality of the ideas in an essay and the degree to which the ideas expressed address the writing task, and a measure of the degree to which a final written product is well organized. These measures address neither the writer's ability to handle demands related to an audience nor the writer's ability to plan before beginning to write or during the writing process.

The Writing Process Test is unique in that it does attempt to measure writers' views of their own writing ability and use of specific planning and revising strategies. Writers rate their writing using the same analytic scoring features that the examiner uses to evaluate the writing. Teacher and writer ratings can then be compared. Writers' evaluations of their own writing are not very reliable, especially among less experienced writers, according to the test publisher's research. However, self-evaluation provides a natural way for adults to be more directly involved in the assessment process.

REVISING AND EDITING. Revising and editing both involve making changes to what has been written. Although the two overlap to some degree, editing is a more local activity, involving changes in sentence-level structures as students write or as they read over what they have

written. Revising is more global and involves adding, deleting, moving, or otherwise changing sentences or paragraphs within a text to better express an idea.

Norm-Referenced, Criterion-Referenced Assessment. Only one assessment was located that provided norm-referenced or criterion-referenced scores for general revising and editing processes. The TOWL has one subtest that measures an aspect of editing. On the logical sentences subtest, students rewrite illogical sentences so that they make better sense. If given the sentence John washed the sky, for example, the student would be expected to rewrite the sentence so that it made sense (John washed his car, for example).

As discussed, the Writing Process Test is unique in that it does attempt to measure writers' view of their own writing ability and use of specific revising strategies.

Both writers and examiners use analytic scoring rubrics to evaluate the revising process. The writers, for example, rate the degree to which they agree with statements such as the following (using a four-point scale): As I rewrote, I thought about the assignment.

Informal Assessment. Although none of the commonly used adult literacy tests evaluate the way in which students edit or revise their own work, both the TABE and ABLE language subtests do ask students to make decisions about secondary texts that are similar to decisions that writers make when editing or revising their own text. A careful item analysis by an examiner can serve as an indirect, informal evaluation of some aspects of these processes. To measure capitalization and punctuation skills, for example, the ABLE asks students to read a sentence that may or may not contain an error and then to select a better version of the sentence or a part of the

sentence if there is an error. A student may be given a sentence like the following: Should I wash the cloths. The student selects the best alternative to the underlined part of the sentence from a list like the following: a. Correct b. Clothes. c. cloth? d. clothes?

The TABE indirectly measures more sophisticated editing and revising abilities as part of its language expression measure: recognizing correct sentence structures, combining sentences, working with topic sentences, and sequencing sentences in a logical manner.

Motivation

Motivation is an important aspect of reading and writing, especially for adult learners, most of whom are not required to attend literacy classes and who must find the time and energy to do so. Motivation, attitude, and engagement in literacy are frequently associated with

time spent reading and reading achievement (Smith, 1990; Guthrie & Wigfield, 1997; Mikulecky & Lloyd, 1997). Motivation has traditionally been assessed in adult literacy during intake interviews, when new learners are asked about their goals and interests (Askov et al., 1997).

Normally, change in motivation to read is not measured, and none of the assessments considered so far contain a measure for motivation. Examples of measures that do exist, in addition to the informal measures mentioned earlier (Askov et al., 1997), are measures developed primarily for research purposes (Beder, 1990; Guthrie & Wigfield, 1997), for statewide performance assessment programs at the K-12 level (Leipzig & Afflerbach, 2000), and in assessments of K-12 literacy curricula (Au, 1997). Among the items in the questionnaire used by Wigfield are, for example, I have favorite subjects

that I like to read about and I like to read about new things. Students indicate their degree of agreement on a four-point scale (Guthrie & Wigfield, 1997, p. 432).

Au's evaluation of a literacy curriculum involved the use of a performance assessment with children. The assessment included grade-level benchmarks to measure ownership of literacy (ownership is considered an aspect of motivation). Teachers used checklists, anecdotal records, collections of student products, and questionnaires to evaluate progress in meeting the benchmarks. Some examples of the benchmarks used are "enjoys writing" (kindergarten) and "makes connections between reading and writing" (grade 3) (Au, 1997, p. 178).

Assessments developed for research and large-scale assessment may provide items that have more validity than those developed by teachers for local programs

(those used during intake interviews, for example). The reliability of motivation questionnaires may be problematic because they are fairly transparent, especially for adults, and the natural tendency is to respond in the way that you think the examiner would want you to respond.

PRACTICES: ASSESSING THE USE OF READING AND WRITING PRACTICES

The frequency of reading practices, such as document, book, newspaper, or magazine reading, is positively associated with literacy ability (Smith, 1995; Sheehan-Holt & Smith, 2000). A goal for many adult literacy programs is to increase both the amount of time adults spend reading and the volume of material they read. Although there are no standardized assessments of literacy practice, it can be assessed informally when a teacher is interested in whether or not a literacy program has positively affected the frequency of specific

reading practices.

Assessment of literacy practices involves self-reports and the use of diaries to record what is read (Alvermann et al., 1999; Kirsch & Jungeblut, 1986; Mikulecky & Lloyd, 1997; Smith, 2000; Sticht, 1995). In a study of after-school "read and talk" clubs, adolescents were expected to keep a daily log in which they answered questions about what they read, where they read, why they read, how much time they spent reading, and how much they used the library as a source for reading (Alvermann et al., 1999). Assessment may be associated with a specific setting or context, such as family literacy practices (National Center for Family Literacy, 1996) or workplace practices (Mikulecky & Lloyd, 1997; Sticht, 1995). Mikulecky and Lloyd, for example, in a study of workplace literacy, asked participants, "Tell me the sorts of things you read and write on the job during a normal

week" (1997, p. 563).

Direct observation and recording of literacy practices can also be used (Sticht, 1995). Direct observation is more reliable than self-reports, although it is more difficult to implement. Interview questions that elicit self-reports must be constructed carefully. Small changes in the phrasing of questions can have a large impact on the information obtained. For example, the question, "Have you completed a book in the past month?" would probably result in fewer positive responses than, "Have you read in a book in the past month?" (Kaestle et al., 1991, p. 189).

Change in literacy practices over time can be assessed by collecting practices data more than once (Mikulecky & Lloyd, 1997), as is required by the NRS. Self-reports can be used to obtain the data specified in the NRS, such as family literacy practices, and to evaluate a

program of instruction. The NRS, for example, suggests that family literacy programs ask adults about practices such as how frequently they read to their children. Unlike more typical forms of performance assessment, results from the assessment of literacy practices are not tied to developmental levels. It is not known, for example, precisely how growth in the frequency or number of reading practices is related to growth in literacy ability.

THE STATE OF ABE LITERACY ASSESSMENT
The most frequently used literacy assessments in adult basic education (the TABE, ABLE, WRAT, CASAS reading tests, and SORT) each provide norm-referenced scores for one or two components of the reading process. The TABE, ABLE, and CASAS measure comprehension, the ABLE has a separate vocabulary measure, and the WRAT and SORT have scores for word recognition. These assessments do not have norm-

referenced scores for fluency, word analysis, or aspects of the writing process other than sentence production and spelling. Some have criterion-referenced measures for word analysis and a few additional components of the writing process, but they generally rely on too few items or are otherwise difficult to interpret.

Norm-referenced, criterion-referenced, and standardized performance assessments for adults that measure other components do exist, including measures of fluency (the GORT and DAR), word analysis (the Woodcock and DAR), and written products (the CASAS Functional Writing Assessment and Woodcock). Two criterion-referenced or performance-based assessments that were developed primarily for the K-12 level might also be used with adults to measure written products and writing vocabulary (the TOWL and the Writing Process Test) and planning and revising

or editing (the Writing Process Test).

Of all the tests mentioned here, only the DAR and the Woodcock (Reading) attempt to measure all aspects of the reading process, and only the Woodcock (Achievement) attempts to measure multiple components of both the reading and writing process. Unfortunately, the DAR has only one form, which makes it difficult to use for both pre- and posttesting, and the Woodcock is available only to those with specified credentials (requiring a fairly high level of expertise).

There are no formal, adult-oriented assessments of the motivational aspect of reading and writing. Assessments of motivation designed for research with adults (Beder, 1990) or at the K-12 level (for example, Guthrie & Wigfield, 1997) might serve as examples. There are also no formal assessments for literacy practices, although research may again

serve as a guide for the creation of questions that help to generate reliable self-reports of adult practices (Purcell-Gates, Degener, Jacobson, & Soler, 2000; Mikulecky & Lloyd, 1997; Kaestle et al., 1991).

Most of the common adult literacy assessments (the TABE, ABLE, and CASAS) use adult-oriented contexts, including functional, life-skills, and workplace content for test items. The ABLE has the most academic content, while the CASAS has the most functional content. Although the WRAT and SORT do not use adult contexts, there are other word recognition tests that focus on specific contexts, such as health and medicine (the TOFHLA and REALM).

Performance assessments have the potential to measure many aspects of reading and writing ability. Although there is no detailed, comprehensive

survey of their use in adult literacy, K-12 and adult education literature indicate that they have traditionally focused primarily on reading comprehension, written products, and oral reading. They are, for example, used to measure aspects of reading comprehension that common assessments do not, such as comprehension monitoring and strategy use. They are also used to gauge the ability to use reading and writing in naturally occurring situations. Methods to use in evaluating the reliability and validity of performance assessments are still evolving (Leipzig & Afflerbach, 2000).

Most of the common adult literacy assessment instruments are group-administered tests (the TABE, ABLE, and CASAS). They provide brief scripts for test administrators to use and so can be administered fairly easily and reliably. The WRAT and SORT are somewhat more difficult to administer.

They are given individually and the tester must be able to interpret and score oral responses as either correct or incorrect, and must know when to end the testing. Less frequently used tests, such as the DAR, Woodcock, and CASAS writing test are more complex to administer. Performance assessments, because they are a newer form of assessment and do not have established procedures for constructing tasks and developing scoring rubrics, are perhaps the most complex assessments to administer. Setting up performance assessment systems is an extended, iterative process even for those who are experts (for example, see Paris, 1999).

The amount of training that adult literacy staff need in order to reliably administer literacy assessments varies along with the complexity of the assessments. Training is necessary, however, when scoring and interpreting even the simplest tests. A task as simple as using

a norms table to convert a raw score into a percentile rank or GE can create problems even for a trained professional (Nitko, 1983, p. 361). Knowing which forms and levels of a test to use is problematic for many adult educators (Kutner et al., 1996). Interpreting the wide variety of derived scores requires training and experience as well.

When administered by properly trained staff, all the assessments mentioned above can be used to satisfy the accountability requirements of the NRS (with the exception, perhaps, of the DAR, which has only one form). With more training and experience in selecting and using the right combination of tests, practitioners can use these tools to inform instruction. Scale scores and GEs can be used to help guide instruction, for example, but it is important to know that different tests construct these scores in different ways, and that the way in which they are constructed can affect

interpretation.

IMPLICATIONS FOR PRACTICE, RESEARCH, AND POLICY

How well do common adult literacy assessments align with views of literacy in adult basic education, particularly along the dimensions of practice, context, and ability? First of all, none of the formal assessments discussed here were designed to assess literacy practices. Second, some of the commonly used adult literacy assessment instruments use content from multiple adult contexts, although none, of course, are able to provide information about all contexts. Third, the most commonly used standardized tests in adult education each measure just one or two components of the reading process and only a few of many aspects of the writing process.

The NRS requires just one assessment of any one aspect of basic literacy ability in virtually any context, however, so any of

the commonly used tests could be used for federal accountability purposes. Adult literacy programs are not required by the NRS to measure literacy practices, but those focusing on family literacy are encouraged to measure literacy practices related to parents' interaction with their children. For this reason, instruments or procedures for measuring practices that have been validated through research or extensive use are needed. Literacy practices have been investigated throughout the history of adult basic education (Kaestle et al., 1991), and some of this research may serve as a starting point (for example, Purcell-Gates, Degener, Jacobson, & Soler, 2000; Mikulecky & Lloyd, 1997; Sticht, 1995).

Literacy assessment should not be used solely to satisfy requirements for accountability but should be fully integrated into instruction (Askov et al., 1997; Askov, 2000; Joint Task Force on

Assessment, 1994; Joint Committee on Standards, 1999). How well do the most common adult literacy assessments support instructional models? For those programs that construct profiles of student strengths and weaknesses to provide guidance in the selection of instructional methods and materials (Chall, 1994; Chall & Curtis, 1992; Curtis, 1999), even a combination of the tests commonly used in adult literacy is insufficient (Strucker, 1997b; Chall, 1994; Snow & Strucker, 2000). Reading specialists have used combinations of other standardized norm-referenced and criterion-referenced tests to construct complete profiles (Chall, 1994; Chall & Curtis, 1992; Strucker, 1997b; Curtis, 1999). Using the ABLE, GORT, and WRAT together during assessment, for example, would provide information about all aspects of the reading process except word analysis. There are also single, standardized assessments that provide measures of many aspects of

reading and writing (for example, the Woodcock and DAR).

Even for adult literacy programs that focus most of their energies on only one aspect of reading, such as reading comprehension, a single norm-referenced or criterion-referenced test may not be adequate. For some, the use of multiple-choice or short-answer formats, as opposed to extended, constructed responses (Martinez, 1999), is seen as a real limitation (Merrifield, 1998; Garcia & Pearson, 1991). These formats do not directly measure some comprehension abilities, such as comprehension monitoring and strategy use. Performance assessments are capable of directly measuring a wider range of comprehension abilities because they do not rely on short-answer formats (Martinez, 1999). These have probably been used by some of the 31 percent of programs that construct their own assessments (Kutner et al., 1996),

although no research on the types of performance assessments actually used in adult literacy programs is available.

Related to the use of assessment for instruction is the issue of the use of standardized scores from norm-referenced tests to gauge learner strengths and weaknesses in literacy (for example, Chall & Curtis, 1992; Strucker, 1997b). The NRS uses scale scores and grade equivalent scores (GEs) from common adult literacy tests to help describe levels in the development of adults' literacy abilities (DAEL, 2000, p. 14). Norm-referenced scores are used primarily to compare the performance of a learner with that of a norm group. Using them to describe literacy development requires extensive experience in teaching and assessing literacy ability. An experienced diagnostician can presumably interpret a GE on a test, for example, because the diagnostician is familiar with the test,

what it measures, and the psychometric use of GEs and also knows that even though different tests may use these same terms, GEs and scale scores may be derived from raw scores in different ways. Many recommend that GEs and scale scores be interpreted cautiously by those without this knowledge. The meaning of scale scores is not intuitive, and GEs may be overinterpreted because everyone is familiar with the concept of grade levels.

The use of standardized norm- and criterion-referenced scores for virtually any purpose has been questioned, usually in comparison with performance assessments. Questions about these tests come from within the field of adult literacy (for example, Beder, 1999; Merrifield, 1998; Padak & Padak, 1994) and among educators generally (for example, Pelligrino, Baxter, & Glaser, 1999). Common complaints include the following: standardized tests do not

measure what has been learned, they focus on isolated skills, and they often fail to measure more complex reasoning and problem-solving abilities. Performance assessments can potentially do all of this because they are extremely flexible and can be designed by a particular program's practitioners to fit specific program needs.

As Merrifield (1998) states, the dilemma is that standardized tests do not adequately measure what is learned, while performance assessments, because of their ad hoc, informal nature, are not reliable enough for the comparisons across individuals and programs that policymakers require. As noted, however, some performance assessments, such as writing assessments, are becoming more standardized while some standardized assessments are becoming more flexible. The development of performance assessments seems to be a continuation

of a series of innovations in assessment, such as those that brought criterion-referenced testing in the 1960s, that will add to the tools that can be used rather than supplant all others. Data derived from the NRS, which encourages the use of both performance and norm- and criterion-referenced tests, may help spur the development of reliable performance assessments and help to determine whether or not they will provide information that is sufficiently valid for policymakers' decisions.

Another, more intransigent dilemma in ABE is related to the issue of teacher training. Lack of resources, reliance on part-time staff, and the extensive use of volunteers means that adult literacy teachers on average have less experience and training than teachers at the K-12 level. Greater accountability, however, through the use of formal assessments, means that adult literacy teachers will be expected to do more (Merrifield, 1998;

Beder, 1999). The use of assessment for accountability and instruction requires a greater degree of sophistication in the teaching of reading than recent evaluations of adult literacy programs suggest current staff have (Kutner et al., 1996; Calfee & Hiebert, 1991).

Although this dilemma is not one that will be easy to remedy, focusing on assessment during the training of adult literacy staff may actually have direct beneficial effects. If an adult literacy assessment instrument or system truly represents the domain of behaviors to be addressed during instruction, learning about the assessment will provide teachers with knowledge about adult literacy. Learning about a word analysis or reading comprehension assessment, for example, should provide information about what is expected of adults in these two domains. For instructional models that rely on assessment, assessment is a natural place to begin focusing training.

Adult literacy instructors, and volunteers in particular, need to know about what reading is and how it develops (Wasik, 1998). This knowledge may be presented naturally as practitioners learn about and practice effective assessment procedures.

Practice

In the current environment, with its increased demands for accountability and the new National Reporting System, adult literacy programs cannot avoid formal assessment, as some in the past have managed to do. Assuming also that assessment should be integrated with instruction, the model described by Askov (Askov et al., 1997) and many others should be used: assess student needs, provide instruction based on assessment results, and assess students periodically to adjust instruction and determine whether or not instruction is leading to gains in literacy ability. For those programs that focus on providing direct, explicit instruction in all aspects

or components of the reading process (for example, Chall, 1994; Curtis & Longo, 1997), assessment should include profiling adults' strengths and needs across components, and the assessment instruments chosen should be capable of doing this. Other models are possible, of course. For those programs that focus on one particular aspect of reading, or that view reading as a unitary process, for example, the instrument chosen may assess only this one aspect of reading, such as reading comprehension. Other programs may focus narrowly on one literacy context, such as health, the family, or the workplace, and assessments in these programs may rely more on instruments that have appropriate content.

Training in assessment is key at this point for adult literacy practitioners in local programs. As Calfee and others have noted (Calfee & Hiebert, 1991), teachers must have extensive knowledge

of and practice with assessment to integrate teacher-based assessment effectively and reliably. How training is delivered as well as the content of any training are both important considerations. Training methods need to take into account the high turnover among adult literacy staff, many of whom are part-time or volunteer tutors. One-shot training workshops, for example, will not be effective. Ongoing and on-demand training programs that can be offered as new staff enter would seem to be a more appropriate model. Training program content will need to include instruction in administering assessments and interpreting their results, and it will need to be presented in a way that is understandable to those with the least amount of experience in a program, including volunteers.

Reliable and valid measures should be used by practitioners. This is an NRS requirement for accountability, but it is

also important for instruction. Reliable measures provide better support for instruction. Guidelines provided by professional organizations for the selection and use of assessments should be used (such as Joint Committee on Standards, 1999; Joint Task Force on Assessment, 1994). The NRS requires states to audit local program assessment procedures to help insure reliability. Local programs should also attempt to assess or monitor instructors' assessment and instruction abilities. Assessing teacher knowledge should be just as important as assessing student knowledge.

Research
Research that evaluates whether and how various approaches to assessment in ABE lead to gains in literacy ability is needed. While the recommendation that assessment be used to guide instruction and to evaluate program effectiveness seems to be sound policy (for example,

DAEL, 2000; Joint Committee on Standards, 1999; Joint Task Force on Assessment, 1994), research that links assessment to ABE students' gains in literacy ability is missing. Closely related to this is research that will support the training of ABE staff in the best approaches to assessment. This includes research on effective training methods and research on the abilities and needs of adult literacy staff. What do they know about what literacy is and how it develops? How reliably do they use assessment instruments?

Research is also needed on the most neglected aspects of adult literacy assessment. Formal measures for motivation and for specific literacy practices need to be developed. More formal measures and procedures for performance assessment are needed, as is research that will establish and measure their reliability. This could include broader, comparative research that looks

at validity across various types of adult literacy assessment instruments.

More research is needed on the effects of context on literacy ability. Does the content or context of a literacy program-the degree to which it is functional, for example-affect gain in literacy ability (for example, Sticht & McDonald, 1992)? Do profiles change as content reflecting different contexts changes?

Finally, more research is needed on the best ways to measure various aspects of reading and writing processes to obtain useful profiles of adult literacy learners' strengths and needs. Research is being conducted by NCSALL, for example, that is identifying specific types of learner profiles (Strucker, 1997b; Snow & Strucker, 2000). How to best integrate profiles that result from the assessment of specific abilities into instruction is another area in which research is needed.

Policy

Policymakers need to provide adequate resources for the research described here as well as for the development, purchase, and use of assessments, including training. Although adult education has been, essentially, level-funded (or worse) since its inception in the 1960s (Sticht, 1998), demands for program accountability have steadily increased.

Ways in which to evaluate the reliability of data being collected for the NRS should be specified. The NRS currently relies on states to collect reliability information through program audits. At a minimum, common guidelines or standards for auditing programs should be provided. Assessment data from the NRS will be used to measure the effectiveness of ABE programs. Because states and individual programs may use different criteria to determine adults' beginning and ending literacy levels, results will be open to the criticism that

they are not reliable. A truly reliable assessment of effectiveness can come only from the consistent administration of a common assessment. This might be accomplished best through stratified random sampling of a large number of adults by a third party.

With this in mind, it is important to anticipate and guard against the NRS becoming exclusively a high-stakes system. High-stakes assessment for an instructional program occurs when the results of a single test are used as a basis for delivering consequences, such as funding incentives, or when test results are released publicly so that comparisons can be made across programs (Joint Task Force on Assessment, 1994; International Reading Association, 1999). Although the NRS does not rely on a single measure or test to evaluate program performance, it does provide states with performance incentives, requires them to publish assessment

results, and requires them to evaluate and provide incentives for local programs (DAEL, 2000; PL-105-220, Workforce Investment Act, Title II, Chapter 1, Section 212).

Though the NRS collects data from many measures as opposed to just one, the way in which the system is structured will probably lead at least some states to use a single measure to evaluate many local programs, unless specific evaluation guidelines are provided that encourage the use of multiple measures. The central measure in the NRS system is gain in literacy ability, and this measure may be obtained by administering a standardized test at the beginning and end of an instructional cycle. Although this is not the only way in which gain may be measured, many states will select it because it is efficient and cost-effective.

Potential problems associated with high-

stakes testing include, among others, a narrowing of the curriculum through teaching to the test and focusing attention on those students most likely to show gain on the test being used. To take an extreme example of curriculum narrowing, if the WRAT, a simple measure of word recognition, were the test selected to measure gain, teachers might be tempted to focus on word recognition and neglect other aspects of the reading process during instruction. High-stakes testing can also tempt a program to focus on a specific subset of students most likely to succeed-a practice called creaming-which has occurred in at least one federal program using performance standards (Condelli & Kutner, 1997). This is a potential problem for ABE programs, where so many students may have a reading disability (Snow & Strucker, 2000), and where programs may not assess extremely poor readers until they are "ready" (that is, they read at a higher

level) (Kutner et al., 1996). Although Condelli and Kutner mention several ways to minimize the negative effects of high-stakes testing, such as setting reasonable, obtainable objectives, matching performance measures with program goals, and training and monitoring staff, the most effective approach is probably to require that funding decisions be based on evaluations that use multiple measures.

There is an inherent tension between high-stakes testing and established procedures for assessment within a program. High-stakes tests may be viewed as time-consuming add-ons or as replacements for existing assessment procedures. When a program lacks clear goals and adequate assessment practices, however, even strong opponents of externally mandated testing state that it may "fill a vacuum" and serve as a catalyst for needed change (Calfee & Hiebert, 1991). As the evaluations of

adult literacy programs discussed in this chapter indicate, this seems to be the case for many adult literacy programs. Assuming that the training provided for states through the AEFLA is adequate, and that the states in turn provide adequate training for local programs, a high-stakes assessment implemented through the NRS may in some cases be beneficial. Whatever the outcomes, effective research is needed to describe and understand them. Discussion of any lessons learned should be based on a solid foundation that includes reliable research data.

Note

1. There are two editions of the TABE, the TABE Forms 5 & 6, published in 1987, and the TABE Forms 7 & 8, published in 1994. To distinguish between the two, the most recent TABE will be referred to simply as "the TABE" and the earlier edition will be referred to as "the 1987 TABE." The major difference between them is that the 1987 TABE provides separate reading comprehension and vocabulary scores while the most recent TABE provides only a reading comprehension score (vocabulary is measured as

a part of reading comprehension).

References

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Adams, M. J. (1998). Phonemic awareness in young children. Baltimore: Brookes Publishing.

Alvermann, D., Young, J., Green, C., & Wisenbaker, J. (1999). Adolescents' perceptions and negotiations of literacy practices in after-school read and talk clubs. American Educational Research Journal, 36(2), 221-264.

American Council on Education. (1993). The Tests of General Educational Development: Technical manual. Washington, DC: American Council on Education.

Anderson, R. C. (1984). Role of the reader's schema in comprehension,

learning, and memory. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading (3rd ed., pp. 372-384). Newark, DE: International Reading Association.

Askov, E. N. (2000). Adult literacy. In A. L. Wilson & E. R. Hayes (Eds.), Handbook of adult and continuing education (pp. 247-262). San Francisco: Jossey-Bass.

Askov, E., Van Horn, B., & Carman, P. (1997). Assessment in adult basic education programs. In A. Rose & M. Leahy (Eds.), Assessing adult learning in diverse settings: Current issues and approaches (Fall, pp. 65-74). San Francisco: Jossey-Bass.

Au, K. H. (1997). Ownership, literacy achievement, and students of diverse cultural backgrounds. In J. T. Guthrie & A. Wigfield (Eds.), Reading engagement: Motivating readers through

integrated instruction (pp. 168-182). Newark, DE: International Reading Association.

Baker, L., Dreher, M. J., & Guthrie, J. T. (Eds.). (2000). Engaging young readers: Promoting achievement and motivation. New York: Guilford Press.

Beder, H. (1990). Motivational profiles of adult basic education students. Adult Education Quarterly, 40(2), 78-94.

Beder, H. (1999). The outcomes and impacts of adult literacy education in the United States (NCSALL Reports #6). Cambridge, MA: The National Center for the Study of Adult Learning and Literacy.

Bereiter, C. (1980). A framework for a cognitive theory of writing. In . Gregg & E. R. Steinberg (Eds.), Cognitive processes in writing (pp. 73-93). Hillsdale, NJ: Erlbaum.

Brookfield, S. (1997, Fall). Assessing critical thinking. In A. Rose & M. Leahy (Eds.), Assessing adult learning in diverse settings: Current issues and approaches (pp. 17-30). San Francisco: Jossey-Bass.

Calfee, R., & Hiebert, E. (1991). Classroom assessment of reading. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 2, pp. 281-309). New York: Longman.

Carr, T. H., & Levy, B. A. (1990). Reading and its development: Component skills approaches. San Diego: Academic Press.

CASAS. (1989). CASAS: Comprehensive Adult Student Assessment System. San Diego: Author.

Chall, J. S. (1994). Patterns of adult

reading. Learning Disabilities, 5(1), 29-33.

Chall, J. S. (1996). Stages of reading development. New York: Harcourt.

Chall, J. S., & Curtis, M. E. (1987). What clinical diagnosis tells us about children's reading. Reading Teacher, 40, 784-788.

Chall, J. S., & Curtis, M. E. (1990). Diagnostic achievement testing in reading. In C. Reynolds & R. Kamphaus (Eds.), Handbook of psychological and educational assessment of children. New York: Guilford Press.

Chall, J. S., & Curtis, M. E. (1992). Teaching the disabled or below-average reader. In S. J. Samuels & A. E. Farstrup (Eds.), What research has to say about reading instruction (pp. 253-276). Newark, DE: International Reading Association.

Collins, A., & Gentner, D. (1980). A framework for a cognitive theory of writing. In L. Gregg & E. R. Steinberg (Eds.), Cognitive processes in writing (pp. 51-72). Hillsdale, NJ: Erlbaum.

Colvin, R., & Root, J. (1982). READ-Reading Evaluation Adult Diagnosis: A test for assessing adult student reading needs and progress (Rev. ed.). Syracuse, NY: Literacy Volunteers of America.

Comings, J. P., Parrella, A., & Soricone, L. (1999). Persistence among adult basic education students in pre-GED classes (NCSALL Reports #12). Cambridge, MA: National Center for the Study of Adult Learning and Literacy, Harvard Graduate School of Education.

Condelli, L. (1996). Evaluation systems in the adult education program: The role of quality indicators. Washington, DC: Pelavin Research

Institute.

Condelli, L., & Kutner, M. (1997). Developing a national outcome reporting system for the adult education program. Washington, DC: Pelavin Research Institute.

Condelli, L., Padilla, V., & Angeles, J. (1999). Report on the pilot test for the National Reporting System. Washington, DC: Office of Vocational and Adult Education, Division of Adult Education and Literacy.

Cramer, E. H., & Castle, M. (1994). Fostering the love of reading: The affective domain in reading education. Newark, DE: International Reading Association.

CTB/McGraw-Hill. (1987). TABE Forms 5 & 6: Tests of Adult Basic Education. Monterey, CA: Author.

CTB/McGraw-Hill. (1994a). TABE Forms 7 & 8: Tests of Adult Basic Education. Monterey, CA: Author.

CTB/McGraw-Hill. (1994b). TABE Work-Related Foundation Skills. Monterey, CA: Author.

Curtis, M. E. (1980). Development of components of reading skill. Journal of Educational Psychology, 72(5), 656-669.

Curtis, M. E. (1990). Developing literacy in children and adults: Are there differences? Paper presented at the Annual Meeting of the International Reading Association, Atlanta, GA.

Curtis, M. E. (1999). When adolescents can't read: Methods and materials that work. Cambridge, MA: Brookline Books.

Curtis, M. E., & Chmelka, M. B. (1994). Modifying the "Laubach Way to

Reading" program for use with adolescents with learning disabilities. Learning Disabilities Research and Practice, 9(1), 38-43.

Curtis, M. E., & Longo, A. M. (1997). Reversing reading failure in young adults. Focus On Basics, 1(B), 18-22.

Davis, T. C., Crouch, M. A., & Long, S. (1992). REALM: Rapid Estimate of Adult Literacy in Medicine. Shreveport, LA: School of Medicine, Louisiana State University.

Diehl, W., & Mikulecky, L. (1980). The nature of reading at work. Journal of Reading, 24, 221-228.

Division of Adult Education and Literacy (DAEL). (2000). Measures and methods for the National Reporting System for adult education: Implementation guidelines. Washington, DC: Office of Vocational and Adult

Education, U.S. Department of Education.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. Review of Educational Research, 69(2), 145-186.

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test-Third Edition (PPVT-III). Circle Pines, MN: American Guidance Service.

Educational Testing Service. (1991). Tests of Applied Literacy Skills. New York: Simon & Schuster.

Ehringhaus, C. (1990). Functional literacy assessment: Issues of interpretation. Adult Education Quarterly, 40(4), 187-196.

Ehringhaus, C. (1991). Teachers'

perceptions of testing in adult basic education. Adult Basic Education, 1(3), 138-154.

Fehring, H., & Green, P. (2001). Critical literacy: A collection of articles from the Australian Literacy Educators' Association. Newark, DE: International Reading Association.

Fingeret, H. A. (1993). It belongs to me: A guide to portfolio assessment in adult education programs. Durham, NC: Literacy South.

Flower, L., & Hayes, J. (1981, December). A cognitive process theory of writing. College Composition and Communication, pp. 365-387.

Freire, P., & Macedo, D. (1987). Literacy: Reading the word and the world. South Hadley, MA: Bergin & Garvey.

Garcia, G. E., & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. H. Hiebert (Ed.), Literacy for a diverse society: Perspectives, practices, and policies (pp. 253-278). New York: Teachers College Press.

Garner, B. (1999). Nationwide accountability: The National Reporting System. Focus on Basics, 3(B).

GED Testing Service. (1997). Tests of General Educational Development: Official practice tests. Washington, DC: American Council on Education.

General Accounting Office. (1975). The adult basic education program: Progress in reducing illiteracy and improvements needed. Washington, DC: U.S. Office of Education.

General Accounting Office. (1995). Adult education: Measuring program results has been challenging

(GAO/HEHS-95-153). Washington, DC: U.S. General Accounting Office.

Goodman, Y. M. (1999). Revaluing readers while readers revalue themselves: Retrospective miscue analysis. In S. Barrentine (Ed.), Reading assessment: Principles and practices for elementary teachers (pp. 140-151). Newark, DE: International Reading Association.

Goodman, Y. M., & Burke, C. L. (1972). Reading Miscue Inventory: Manual and procedures for diagnosis and evaluation. New York: Macmillan.

Goodman, Y. M., Watson, D., & Burke, C. L. (1987). Reading Miscue Inventory: Alternative procedures. Katonah, NY: Owen.

Gregg, L. W., & Steinberg, E. R. (Eds.). (1980). Cognitive processes in writing. Hillsdale, NJ: Erlbaum.

Guthrie, J. T., & Greaney, V. (1991). Literacy acts. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 2, pp. 68-96). New York: Longman.

Guthrie, J. T., & Wigfield, A. (Eds.). (1997). Reading engagement: Motivating readers through integrated instruction. Newark, DE: International Reading Association.

Hammill, D., & Larsen, S. (1996). TOWL: Tests of Written Language (3rd ed.). Austin, TX: PRO-ED.

Harris, T., & Hodges, R. (1995). The literacy dictionary. Newark, DE: International Reading Association.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences, and

applications (pp. 1-27). Mahwah, NJ: Erlbaum.

Hiebert, E. H. (1991). Literacy for a diverse society. New York: Teachers College Press.

International Reading Association. (1999). High-stakes assessments in reading: A position statement of the International Reading Association. Journal of Adolescent and Adult Literacy, 43(3), 305-312.

Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association and American Psychological Association and National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Joint Task Force on Assessment. (1994). Standards for the assessment of reading and writing. Newark, DE: International Reading Association and National Council of Teachers of English.

Kaestle, C. F., Damon-Moore, H., Stedman, L. C., Tinsley, K., & Trollinger, W. V., Jr. (1991). Literacy in the United States: Readers and reading since 1880. New Haven, CT: Yale University Press.

Karlsen, B., & Gardner, E. (1986). ABLE: Adult Basic Learning Examination. New York: The Psychological Corporation, Harcourt.

Kasworm, C., & Marienau, C. (1997, Fall). Principles for assessment of adult learning. In A. Rose & M. Leahy (Eds.), Assessing adult learning in diverse settings: Current issues and approaches (pp. 5-16). San Francisco: Jossey-Bass.

Kintsch, W. (1994). The role of knowledge in discourse comprehension: A construction-integration model. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed.). Newark, DE: International Reading Association. (Original work published in 1988)

Kirsch, I., & Jungeblut, A. (1986). Literacy: Profiles of America's young adults (Report No. 16-PL-02). Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). Adult literacy in America: A first look at the findings of the National Adult Literacy Survey. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Kruidenier, J. R. (1990). Objectives and content of a course for professional and

volunteer teachers of adults. Paper presented at the Annual Conference of the International Reading Association, Atlanta, GA.

Kruidenier, J. R. (1991). Planning and production processes in the written language of skilled and less-skilled writers. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Kruidenier, J. R. (1993). Sentence planning processes in a writing-after-reading task. Paper presented at the American Education Research Association Annual Meeting, Atlanta, GA.

Kutner, M., Webb, L., & Matheson, N. (1996). A review of statewide learner competency and assessment systems. Washington, DC: Pelavin Research Institute.

Leipzig, D. H., & Afflerbach, P. (2000).

Determining the suitability of assessments: Using the CURRV framework. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), Engaging young readers: Promoting achievement and motivation. New York: Guilford Press.

Lesgold, A. M., Roth, S. F., & Curtis, M. E. (1979). Foregrounding effects in discourse comprehension. Journal of Verbal Learning & Verbal Behavior, 18 (3), 291-308.

Levy, M. C., & Ransdell, S. (Eds.). (1997). The science of writing: Theories, methods, individual differences, and applications. Hillsdale, NJ: Erlbaum.

Linn, R. L. (2000). Assessments and accountability. Educational Researcher, 29(2), 4-16.

Martinez, M. E. (1999). Cognition and the question of test item format. Educational Psychologist, 34(4), 207-

218.

Merrifield, J. (1998). Contested ground: Performance and accountability in adult basic education (NCSALL Reports 1). Cambridge, MA: National Center for the Study of Adult Learning and Literacy.

Mikulecky, L., & Drew, R. (1991). Basic literacy skills in the workplace. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 2). New York: Longman.

Mikulecky, L., & Lloyd, P. (1997). Evaluation of workplace literacy programs: A profile of effective instructional practices. Journal of Literacy Research, 29(29), 555-585.

Murphy, P., Davis, T. C., Long, S. W., Jackson, R. H., & Decker, B. C. (1994). Rapid Estimate of Adult Literacy in Medicine (REALM): A quick reading test for patients. In M. C. Radencich

(Ed.), Adult literacy: A compendium of articles from the Journal of Reading (pp. 79-86). Newark, DE: International Reading Association.

National Center for Family Literacy. (1996). Outcomes and measures in family literacy programs. Louisville, KY: Author.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child Health and Human Development.

Nitko, A. J. (1983). Educational tests and measurement: An introduction. New York: Harcourt.

Nitko, A. J. (1996). Educational assessment of students. Upper Saddle River, NJ: Prentice Hall.

Nurss, J., Parker, R., Williams, M., & Baker, D. (1995). TOFHLA: Test of Functional Health Literacy. Atlanta: Emory University.

Padak, N. D., Davidson, J. L., & Padak, G. M. (1994). Exploring reading with adult beginning readers. In M. C. Radencich (Ed.), Adult literacy: A compendium of articles from the Journal of Reading (pp. 56-60). Newark, DE: International Reading Association.

Padak, N. D., & Padak, G. M. (1994). What works: Adult literacy program evaluation. In M. C. Radencich (Ed.), Adult literacy: A compendium of articles from the Journal of Reading (pp. 86-93). Newark, DE: International Reading Association.

Paris, S. G. (1999). Portfolio assessment for young readers. In S. Barrentine (Ed.), Reading assessment: Principles and

practices for elementary teachers (pp. 131-134). Newark, DE: International Reading Association.

Paris, S. G., Calfee, R. C., Filby, N., Hiebert, E. H., Pearson, P. D., Valencia, S. W., & Wolf, K. P. (1999). A framework for authentic literacy assessment. In S. Barrentine (Ed.), Reading assessment: Principles and practices for elementary teachers (pp. 30-43). Newark, DE: International Reading Association.

Pelligrino, J., Baxter, G. P., & Glaser, R. (1999). Addressing the "Two Disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. In A. Iran-Nejad & P. D. Pearson (Eds.), Review of research in education (Vol. 24, pp. 307-354). Washington, DC: American Educational Research Association.

Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.

Perfetti, C., & Curtis, M. E. (1987). Reading. In R. F. Dillon & R. J. Sternberg (Eds.), Cognition and instruction (pp. 13-57). New York: Academic Press.

Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. New York: Guilford Press.

Purcell-Gates, V., Degener, S., Jacobson, E., & Soler, M. (2000). Affecting change in literacy practices of adult learners: Impact of two dimensions of instruction (Report #17). Cambridge, MA: The National Center for the Study of Adult Learning and Literacy, Harvard Graduate School of Education.

Reder, S. (1994). Practice engagement theory: A sociocultural approach to literacy across languages and cultures.

In B. M. Ferdman,

R. M. Weber, & A. G. RamÌrez (Eds.),
Literacy across languages
and cultures. Albany, NY: State
University of New York Press.

Resnick, D. P., & Resnick, L. B. (1977).
The nature of literacy: An historical
explanation. Harvard Educational
Review, 47(3).

Rosner, J. (1979). TAAS: Test of
Auditory Analysis Skills. In J. Rosner,
Helping children overcome learning
disabilities. Novato, CA: Academic
Therapy Publications.

Roswell, F., & Chall, J. S. (1992).
DARTTS: Diagnostic Assessments of
Reading and Trial Teaching Strategies.
Chicago: Riverside.

Roswell, F. G., & Chall, J. S. (1994).
Creating successful readers: A practical

guide to testing and teaching at all age levels. Chicago: Riverside.

Roswell, F., & Natchez, G. (1979). Reading disability: A human approach to evaluation and treatment of reading and writing difficulties (4th ed.). New York: Basic Books.

Secretary's Commission on Achieving Necessary Skills (SCANS). (1991). What work requires of schools: A SCANS report for America 2000. Washington, DC: U.S. Department of Labor.

Sheehan-Holt, J. K., & Smith, M. C. (2000). Does basic skills education affect adults' literacy proficiencies and reading practices? Reading Research Quarterly, 35(2), 226-243.

Slosson, R. L., & Nicholson, C. L. (1990). Slosson Oral Reading Test, Revised. East Aurora, NY: Slosson

Educational Publications.

Smith, M. C. (1990). The development and use of an instrument for assessing adults' attitudes toward reading. Journal of Research and Development in Education, 23(3), 156-161.

Smith, M. C. (1995). Differences in adults' reading practices and literacy proficiencies. Reading Research Quarterly, 31(2), 196-219.

Smith, M. C. (2000). The real-world reading practices of adults. Journal of Literacy Research, 32(1), 25-52.

Snow, C., Burns, S. M., & Griffin, P. (1998). Preventing reading difficulties in young children: A report of the National Research Panel. Washington, DC: National Academy Press.

Snow, C. E., & Strucker, J. (2000). Lessons from Preventing reading

difficulties in young children for adult learning and literacy. In J. Comings, B. Garner, & C. Smith (Eds.), Annual review of adult learning and literacy: A project of the National Center for the Study of Adult Learning and Literacy (Vol. 1, pp. 25-73). San Francisco: Jossey-Bass.

Stahl, S. (1999). Why innovations come and go (and mostly go): The case of whole language. Educational Researcher, 28(8), 13-22.

Steck-Vaughn Company. (1997). AMES: Adult Measure of Essential Skills. Austin, TX: Author.

Stein, S. G. (1997). Equipped for the future: A reform agenda for adult literacy and lifelong learning. Washington, DC: National Institute for Literacy.

Sticht, T. (1972). Determination of adult

functional literacy skill levels. Reading Research Quarterly, 7(3), 424-465.

Sticht, T. (1990). Testing and assessment in adult basic education and English as a second language programs. San Diego: Applied Behavioral & Cognitive Sciences.

Sticht, T. (1995). The military experience and workplace literacy: A review and synthesis for policy and practice. Philadelphia: National Center on Adult Literacy.

Sticht, T. (1998). Beyond 2000: Future directions for adult education. Washington, DC: Office of Vocational and Adult Education.

Sticht, T. G., Hofstetter, C. R., & Hofstetter, C. H. (1996). Assessing adult literacy by telephone. Journal of Literacy Research, 28(4), 525-559.

Sticht, T. G., & McDonald, B. A. (1992). Teaching adults to read. In S. J. Samuels & A. E. Farstrup (Eds.), What research has to say about reading instruction (pp. 314-334). Newark, DE: International Reading Association.

Strucker, J. (1992). Patterns of reading in Adult Basic Education. Unpublished doctoral dissertation, Harvard University Graduate School of Education, Cambridge, MA.

Strucker, J. (1997a). The reading components approach. Cambridge, MA: National Center for the Study of Adult Literacy and Learning, Harvard University Graduate School of Education.

Strucker, J. (1997b). What silent reading tests alone can't tell you: Two case studies in adult reading differences. Focus on Basics, 40(B), 13-17.

Torgesen, J., Wagner, R., & Rashotte, C. (1999). Test of Word Reading Efficiency. Austin, TX: PRO-ED.

Torrance, M., & Jeffery, G. C. (Eds.). (1999). The cognitive demands of writing: Processing capacity and working memory in text production. Amsterdam: Amsterdam University Press.

van Dijk, T., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Venezky, R. L. (1991). The development of literacy in the industrialized nations of the West. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 2, pp. 46-67). New York: Longman.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). Comprehensive Test of Phonological Awareness. Austin, TX:

PRO-ED.

Warden, M. R., & Hutchinson, T. A. (1992). Writing Process Test. Chicago: Riverside.

Wasik, B. (1998). Volunteer tutoring programs in reading. Reading Research Quarterly, 33(3).

Weiderholt, J. L., & Bryant, B. R. (1992). GORT-3: Gray Oral Reading Test: Third Edition. Austin, TX: PRO-ED.

Wilkinson, G. S. (1993). WRAT3: The Wide Range Achievement Test, 1993 Edition. Wilmington, DE: Wide Range.

Woodcock, R. W. (1997). Woodcock-Johnson Diagnostic Reading Battery. Itasca, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). Woodcock Johnson Tests of

Achievement. Itasca, IL: Riverside.

Wrigley, H. S. (1998). Knowledge in action: The promise of project-based learning. Focus on Basics, 2(D), 13-18.

Young, M. B., Fitzgerald, N., & Fleischman, H. (1994). National evaluation of adult education programs: Draft final report. Arlington, VA: Development Associates.