



NCSALL Home > Publications > Review of Adult Learning and Literacy
> Volume 1 (1999) > Full Chapters > Perspectives on Assessment in
Adult ESOL Instruction

This page is located at: <http://www.ncsall.net/?id=521>

Perspectives on Assessment in Adult ESOL Instruction

Volume 1: Chapter Six

Carol H. Van Duzer

Robert Berdan

Assessment of outcomes and learner progress is a primary concern in federally funded adult education programs. This concern is not new, but it has gained prominence over the past decade as legislative imperatives, such as the National Literacy Act of 1991 and the Government Performance and Results Act of 1993, have required federally funded programs to be more accountable for what they do. The lack of a consistent assessment system across states and across programs within states has impeded the documentation and reporting of results to state and federal stakeholders. Such a system is needed to demonstrate the difference that adult education makes in the lives of learners, the communities in which they live, and the nation as a whole. Welfare reform and the establishment of one-stop centers for education and training underscore the need for better and more compatible accountability systems across and within states (Short, 1997).

The Workforce Investment Act (WIA) of 1998 called for the establishment of "a comprehensive performance accountability system to assess the effectiveness of eligible agencies in achieving continuous improvement of adult education and literacy activities ... in order to optimize the return on investment of Federal funds in adult education and literacy activities" (WIA, section 212.a). States now award adult education funding to programs that provide adult education services based on twelve criteria, which include the degree to which the program establishes performance measures for learner outcomes, past effectiveness in meeting (or even exceeding) these performance measures, and the maintenance of a high-quality information management system that can report participant outcomes and monitor program performance against the performance measures (WIA, section 231.e). Adult education programs that offer English for speakers of other languages (ESOL) have much at stake in the movement to define and

measure learner outcomes, for adult ESOL instruction is the fastest growing area in federally funded adult education programs in the United States (National Center for Educational Statistics, 1997).

Adult ESOL programs have always grappled with how to measure and report a range of desired outcomes and satisfy the demands of each stakeholder: learners, teachers, program administrators, funding agencies and organizations, policymakers, and the general public. Learners want to know how well they are progressing in learning English. Teachers want feedback on the effectiveness of their instruction. Program administrators want to know how well they are meeting program goals and how they can improve their services. Those funding the programs as well as the general public want to know whether funds spent are yielding results. Policymakers want to know what specific practices are successful so they can establish guidelines for allocating future funds. A single approach to assessment may not provide enough useful information to satisfy each of these demands.

Reports on testing and assessment from the early 1990s (Business Council for Effective Literacy, 1990; Sticht, 1990) show that very few of these concerns about assessment have been resolved. The tests being used in adult education-TABE (Test of Adult Basic Education), ABLE (Adult Basic Learning Examination), and CASAS (Comprehensive Adult Student Assessment System)-and in adult ESOL-BEST (Basic English Skills Test) and CASAS-are largely the same, as are the critiques of their validity and reliability. The call for research to help answer questions about the role and use of standardized tests and other assessments is still appropriate, and even the questions of what should be assessed and for what purpose are still being debated. However, the field has at the least made progress on the two following issues. First, it is generally acknowledged that tests developed for native English speakers are not appropriate for use with English-language learners. Second, certain segments of the field have recognized that assessment is but one component in a larger instructional system that includes standards for content, program design, staff development, and assessment.

This chapter seeks to provide the field of education for adult English speakers of other languages-at local, state, and national levels-with a timely overview of the state of assessment in adult ESOL programs in the United States. It also seeks to provide a brief description of assessment reform initiatives in K-12 education and in adult language education abroad that might serve as models for adult education in the United States. Our intent is to help program staff and state and national policymakers make informed choices about assessment measures and procedures and to foster a collaborative effort to build an accountability system that addresses the needs of each stakeholder in a more effective

adult education instructional system. Many decisions about assessment were being made even as we were writing this chapter. Although legislative requirements demand that states have an accountability system in place by July 1999, the field will be debating assessment issues for a long time to come.

For the purposes of this report, we are using the term assessment in a broad sense: to find out what learners want, know, and can do at the beginning of instruction (needs, placement, and diagnosis), throughout instruction (ongoing progress), and at the end of instruction (achievement and outcomes). The information presented is based on a literature review from the fields of language assessment and government policy, reviews of standardized tests currently being used in adult ESOL programs, and discussions with experienced adult ESOL educators. We explore three key issues and make recommendations based on the findings:

- What implications do legislative requirements for performance measures have for adult ESOL programs?
- What assessment tools and processes are available and how adequately do they meet the needs of all stakeholders?
- What insights can be gained from the assessment reform experience of K-12 education and adult language education abroad?

PERFORMANCE MEASURES IN ADULT EDUCATION

The use of standardized tests to evaluate adult education programs was put into legislation for the first time in 1988, in amendments to the Adult Education Act (Business Council for Effective Literacy, 1990). In that legislation, states were required to evaluate the progress of at least one-third of their grant recipients using standardized tests (Sticht, 1990). The National Literacy Act of 1991, amending the Adult Education Act, required the U.S. Department of Education (ED) to develop indicators of program quality that would assist states and local programs in judging the effectiveness of programs that provide adult education services. The legislation specifically called for indicators in the areas of recruitment, retention, and educational gains.

The Department of Education sought input from the field of adult education by reviewing state and local practices related to program quality, commissioning papers by experts in the field, holding focus groups, and working closely with the state directors of adult education (Office of Vocational and Adult Education, 1992). A quality program indicator was defined as a variable reflecting effective and efficient program performance. It was distinguished from a measure (data used to

determine the level of performance) and a performance standard (the level of acceptable performance in terms of a specific numeric criterion).

Under the area of educational gains, two indicators were identified:

- Learners demonstrate progress toward attainment of basic skills and competencies that support their educational needs.
- Learners advance in the instructional programs or complete program education requirements that allow them to continue their education or training.

Sample measures for the first indicator included standardized test scores, competency-based test scores, teacher reports of improvements in communication competencies, and demonstrated improvement on alternative assessments (such as portfolios, checklists of specific employability or life skills, and student reports of attainment). Sample measures for the second indicator included rate of student advancement to a higher level of skill or competency in the program; attainment of a competency certificate, General Educational Development credential (GED), or high school diploma; and percentage of students referred to or entering other education or training programs.

With the passage of the Government Performance and Results Act (GPRA) in 1993, more emphasis was placed on performance measurement as a requirement of government-funded program evaluations. Now, under the Workforce Investment Act, each state must negotiate acceptable target levels of performance on three core indicators with the ED that encompass the quality indicators identified as the result of the earlier legislation: (1) demonstrated improvement in skill levels in reading, writing, and speaking the English language, numeracy, problem solving, English-language acquisition, and other literacy skills; (2) placement in, retention in, or completion of postsecondary education, training, unsubsidized employment, or career advancement; and (3) receipt of a secondary school diploma or its recognized equivalent (WIA, section 212.b.2.A). The levels of performance for each core indicator are to be expressed in objective, quantifiable, and measurable form and must show the progress of each eligible program toward continual improvement of learner performance.

Within their five-year plans for adult education, each state will establish levels of performance for programs to meet. States are in the process of preparing their plans for ED approval. They became effective July 1, 1999. Each year, states will submit data on the core indicators to the secretary of education, who will issue reports on how each state is doing.

To facilitate the accountability and reporting process, the ED has been

working with the state directors of adult education to establish a National Reporting System (NRS). The ED has granted funding to the American Institutes for Research/Pelavin Research Center (AIR/Pelavin) to help establish the system. The NRS will include a common set of outcome measures; a system for collecting data on these measures; and standard guidelines, definitions, and forms for reporting the data (Office of Vocational and Adult Education, 1997).

The NRS draft outcome measures for adult English-language learners are "educational functioning-level descriptors" that describe what a learner knows and can do in three areas: speaking and listening, reading and writing, and functional and workplace skills. These functioning-level descriptors appear to combine features outlined in the CASAS: level descriptors and the Student Performance Level (SPL) descriptors. States are to use the functioning levels to report educational gains of learners in the programs they fund.

Programs will determine an individual learner's entry-level and subsequent-level gains using a uniform, standardized assessment procedure that has been described in the state plan and approved by the ED (Office of Vocational and Adult Education, 1998). Illustrative examples of test benchmarks for each functioning level in the pilot NRS document include a range of CASAS Life Skills scores and SPLs. (The SPLs, developed under the auspices of the Office of Refugee Resettlement's Mainstream English Language Training Project [MELT], U.S. Department of Health and Human Services, 1985, are also descriptions of adult learners' language abilities. They are correlated to the BEST test; both the BEST and CASAS tests are reviewed below.) Unfortunately, the two test benchmark guidelines provided in the pilot draft do not cover the range of measures previously identified for the quality indicator of educational gains, nor do they allow for the flexibility that local programs may need. More examples need to be identified during the pilot field test and added to the final document. If they are not, there is a danger that in the need to satisfy the demands of policymakers and funding sources, assessment will become too narrowly focused on standardized test scores. This may lead to program designs that do not serve the needs of learners or the communities in which they live or that adequately assess what learners know and can do. The fundamental question of what is to be counted as success-and therefore what skills and proficiencies are assessed-needs to be addressed at the program, state, and national levels before the NRS is finalized.

LANGUAGE AND LITERACY ASSESSMENT IN ADULT ESOL

The Adult Education and Family Literacy Act (Title II of the WIA) defines literacy as "an individual's ability to read, write, and speak in

English, compute, and solve problems at levels of proficiency necessary to function on the job, in the family of the individual, and in society." However, there have been many definitions of literacy over the years, including a school-based view of literacy as basic reading and writing skills, the functional view that is in the legislation, and a view of literacy as social practices. Even this latter definition is acknowledged in the "Family Literacy" designation noted in Title II of the WIA. The field of adult ESOL recognizes that there are many literacies, defined by how individuals use literacy in everyday life to achieve personal, family, job, and community participation goals (Crandall, 1992). Literacy includes the ability to complete a task or solve a problem, such as getting a driver's license, completing the GED, or finding a job; to support the learning of one's children; to comprehend print material (in one's first or second language); and more.

Which literacy is to be assessed: development in reading and writing, speaking, mathematical ability, social practice, or all of these? What constitutes progress in these areas, and how is it assessed? Can a gain in general language proficiency on a given measure be considered sufficient, or is a variety of assessment instruments and processes needed? These are questions that must be answered and agreed on at the local, state, and national levels if we are to establish an accountability system that captures learner progress.

Many adult ESOL programs use a combination of assessment tools to meet their program needs. These include standardized tests such as the CASAS and BEST, materials-based tests such as those accompanying text series, and program-based tools such as teacher-made tests and portfolios. However, the field of adult ESOL lacks a cohesive assessment system that enables comparison of learner achievement and program impact across the wide variety of programs (survival, preemployment, preacademic, workplace, vocational ESOL, ESOL for citizenship, ESOL family literacy) and the wide range of delivery systems (local education agencies, community colleges, libraries, community-based or volunteer organizations, churches, businesses, and unions). The lack of consistent assessment procedures from program to program is problematic for two major reasons. One is that it impedes the documentation and reporting of results to the satisfaction of all stakeholders. The other is that it frequently impedes the movement of learners from ESOL to vocational training and academic programs because learner pathways differ from program to program.

Most programs use placement procedures to match students to the levels or courses offered. This may take the form of a standardized test, a program-developed test or interview, or a combination of these. The kinds of assessments that are used after placement depend to a large

extent on the program's philosophy of language and learning, the roles of teachers and learners, and the measures of success as defined by the various stakeholders (Wrigley, 1992). Program staff need to juggle two important purposes for assessment:

- They need to assess and document the actual progress that learners are making toward English-language development and completion of learner goals.
- They must meet the legislative requirements of the WIA, which requires a standardized assessment procedure and performance measures, and the NRS, which links learner progress to proficiency descriptors.

Therefore, they must select instruments and procedures carefully and, in many cases, use a combination of standardized and alternative assessments.

Standardized Assessment Tools in Adult ESOL

For the purpose of this chapter, a standardized assessment tool is one that has been developed according to explicit specifications, has items that have been tested and selected for item difficulty and discriminating power, is administered and scored according to uniform directions, and has dependable norms for interpreting scores (Ebel, 1979). Standardized tests are used in adult education programs in most states because they are easy to administer to groups, require minimal training for the teacher, and purport to have construct validity and scoring reliability (Solorzano, 1994; Wrigley, 1992).

Standardized tests reviewed for this chapter include the Adult Basic Learning Examination, the Test of Adult Basic Education, the Adult Language Assessment Scales (A-LAS), the Comprehensive Adult Student Assessment System (components appropriate for ESOL), New York State's Placement Test for English as a Second Language Adult Students (NYS Place), and the Basic English Skills Test. Because they were designed for native English speakers, the ABLE and TABE are now regarded as inappropriate for English-language learners. We have included them here because they were used in the 1970s and 1980s in many adult education programs that had ESOL learners in classes with adult basic education (ABE) learners, and they are still used in some of these programs today. The A-LAS parallels the ABLE and TABE in that it has both language and mathematics batteries, but it was designed for nonnative English speakers. The CASAS and BEST are the two most widely used standardized tests in adult ESOL programs. Both were developed for assessing nonnative English speakers. The NYS Place was selected because it is the only oral assessment besides the

BEST that was identified by the California Department of Education Adult ESL Assessment Project (Kahn, Butler, Weigle, & Sato, 1995) as suitable as a placement tool for assessing speaking ability in programs that were implementing the California ESL Model Standards (California Department of Education, 1992), discussed below. Ordering information for these tests is included in the chapter appendix.

ADULT BASIC LEARNING EXAMINATION (ABLE). The ABLE was designed for use with native, English-speaking adults who have limited formal education and is used primarily in ABE and GED programs and in prison education. It is an educational achievement test. It was not designed to be a language development test or even a test of language proficiency, although it does have a language subtest in Levels 2 and 3. The ABLE is available in three levels differentiated by years of formal schooling, and it has six subtests: vocabulary, reading comprehension, spelling, language (Levels 2 and 3 only), number operations, and problem solving. The first four of these subtests relate to language proficiency and are described in Exhibit 6.1. Reviews of the full battery can be found in Fitzpatrick (1992) and Williams (1992).

Although the ABLE is represented by its publisher, the Texas-based Psychological Corporation, to be an indicator of educational achievement, many of the items on it reflect a narrow concept of achievement. For example, there is a heavy preoccupation with the inflectional morphology of auxiliary verbs (for example, We are/was [verb]), which is highly differentiated across social groups. At the same time, there is a complete absence of attention to such language systems as complex nominals (for example, the combination of [noun] with [noun]), which continue to develop in adolescence and later across all groups.

The use of the ABLE with the populations for which it was designed—ABE, GED, and prison education—is problematic, and transporting the test to the adult ESOL population compounds these problems. The test does not reflect what is known about stages or sequencing of English language development. Only Level 1 would be plausible for use in most adult ESOL programs. Furthermore, because the vocabulary test is presented orally and the writing is confined to words in isolation (a spelling test), the ABLE has very few items that actually measure literacy skills.

TEST OF ADULT BASIC EDUCATION (TABE). The developers of the TABE made a conscious attempt to assess the basic skills taught in adult basic education programs. The publisher, CTB/McGraw-Hill in California, reports a systematic effort to limit cultural, gender, and ethnic bias; construct items with content appropriate for adults; and include

items developed through item response theory (IRT) modeling (see Hambleton, 1991), with desirable psychometric properties such as item discrimination and range of difficulty. That effort has resulted in content that is generally accessible to immigrant adults.

The TABE has three basic levels: E (easy, grade equivalents 1.6–3.9), M (medium, grades 3.6–6.9), and D (difficult, grades 6.6–8.9), with an upward extension to A (advanced, grades 8.6–14.9). There are also a downward extension, L (literacy), a Spanish-language version, a computer-based version, a placement test, and several other associated products. Levels E through A consist of two mathematics tests (computation and applied mathematics), a reading test, a language test, and an optional spelling test. There is no writing test. The language tests are reviewed in Exhibit 6.2.

The TABE is not suitable for learners in beginning-level adult ESOL classes. If given the locator test, many ESOL learners will be assigned to take the Level L test. That test, with its high proportion of literacy readiness items (twenty-three), may be useful for testing adults with little or no previous alphabetic literacy experience, but it is not appropriate for beginning English learners who are literate in their native language. The remaining twenty-seven items of the test, with too few items at the lower range of language development, will not detect the learning that can reasonably be expected in early levels of ESOL instruction. However, for the most advanced levels of adult ESOL instruction, the TABE may prove useful.

ADULT LANGUAGE ASSESSMENT SCALES (A-LAS). The A-LAS, published in New York by McGraw-Hill, is designed to test the English-language skills needed for "entry level functioning in a mainstream academic or employment environment" (Duncan & DeAvila, 1993, p. 1). The A-LAS consists of two test batteries: a set of oral tests and tests of reading, writing, and mathematics, available in two forms. The reading and writing tests are described in Exhibit 6.3.

The A-LAS was constructed for testing the language and literacy of adults learning English. That it was not adapted from another test is apparent in the vocabulary employed (carefully selected for English learners) and the adult level of the item content (for example, concerning employment). None of the items, however, seeks to differentiate stages of English-language development. Because the A-LAS attempts to test the full range of English skills, from no English to entry-level functioning for employment and academic work, it must contain items that range in difficulty from monosyllabic word recognition to essay construction. However, the reading test has relatively few items at any particular level of development. This may be sufficient for a placement

instrument, but if used for assessing achievement, it will be difficult to detect the increment of learning that many adults display in the relatively short time they stay in programs. The A-LAS would be a much stronger tool for assessing achievement if it were available in multiple levels, allowing more items at each difficulty level.

COMPREHENSIVE ADULT STUDENT ASSESSMENT SYSTEM (CASAS). The CASAS Web site (<http://www.casas.org>) lists the availability of more than one hundred standardized assessments and a variety of instructional and supporting materials developed by CASAS. The system was designed for adult basic education, workforce learning, special education, adult ESOL, and various other state and federal programs. CASAS began in the early 1980s as a collaborative effort between adult educators in California and the State Department of Education. Their goal was to develop an assessment model that would help adult education programs to implement competency-based education, as mandated by the 1982 California state plan for adult education (Center for Adult Education, 1983). Over the years, the system developers have identified more than three hundred competencies (that is, statements that describe adult functioning in employment and in society). They then developed and field-tested more than four thousand life skills reading items that assess those competencies. These items are the basis for the array of assessments now available from CASAS. Extensive training is required as part of the CASAS purchase agreement to ensure proper administration and use of the assessments. See Exhibit 6.4 for a descriptive review of the assessments for ESOL populations, particularly the series of tests characterized as life skills.

CASAS tests multiple modalities at multiple levels with multiple forms. Some of the tests are specifically constructed for the adult English-language learner population. In general, the item content is accessible for immigrant populations. All of the items are tied to a list of competencies, and all of the tests are scaled to a single, uniform scale of proficiency. Ranges of the scale are associated with highly generalized statements of language proficiency or skill levels.

The CASAS system provides the instructor with a chart to construct an item-by-item, student-by-student summary, or class profile, for some of the assessments. The student-by-student class profile is keyed to corresponding CASAS competencies rather than directly to test item. Constructing such class snapshots of student performance allows teachers to see where students are performing well and where they need continued instruction. However, one might hesitate to provide instructors with such information because they might limit instruction to the content of the test, and this might pose a problem with the relationship between parallel forms of the tests. At any level, the items are selected not only to

be of comparable difficulty across forms but to have a very high overlap of the competencies they represent. For example, at Level A, Life Skills Listening, twenty-six of the thirty-four items are drawn from competencies shared across the two forms (form 51 and form 52). Many of the competencies are very narrowly defined (for example, "Interpret clothing and pattern size"). When the class profile shows that many students did not get the item for a particular competency correct, there will be a tendency to focus instruction in that area. When that same competency is retested on the parallel posttest form, improvement could be expected, but this does not mean that there would be similar growth in all comparable competency areas.

Another problem could develop if the instruction and testing are drawn into narrow content domains or competencies while the test results are being interpreted in very broad proficiency ranges. Because the posttests do not sample broadly across the competencies but rather concentrate on the taught areas, they will overstate proficiency when it is then interpreted in terms of broad skill levels. (For example, the test results may state that a person is low intermediate ESL, meaning she can satisfy basic survival needs and very routine social demands and can understand simple learned phrases easily. What the results do not state is that she may have this competency only in the area of clothing.)

BASIC ENGLISH SKILLS TEST (BEST). The BEST, published by the Center for Applied Linguistics in Washington, D.C., is designed as an adult English-language proficiency test, focusing primarily on survival and preemployment language skills. It consists of two parts: oral interview and literacy skills.

Like the CASAS assessments, the BEST was an outgrowth of the movement toward a competency-based approach to instruction for adult English-language learners in nonacademic programs. The development of the BEST was funded principally by the Office of Refugee Resettlement (ORR) of the U.S. Department of Health and Human Services (HHS). Teachers and administrators from ORR's Region 1 in Boston and the National Office in Washington, D.C., worked with test developers from the Center for Applied Linguistics (CAL) to develop the original 1982 version (form A). Three additional forms (B, C, and D) were developed and field-tested in 1984 with the help of staff from seven geographically diverse programs that were participating in the ORR Mainstream English Language Training (MELT) Project (National Clearinghouse for ESL Literacy Education, 1989).

The primary goals of the MELT Project were to provide consistency among ORR-funded programs in the United States, continuity between the domestic and overseas training programs (mostly in Southeast Asia),

and guidance for curriculum development, establishment of instructional levels, and assessment. Products other than the BEST included the Student Performance Levels (SPLs) (cited as test benchmarks in the NRS) and a core curriculum document that correlated topics and competencies to the SPLs (U.S. HHS, 1985). Although the BEST was primarily developed for use with English-language learners from Southeast Asia, many of the programs participating in the field test of the BEST also provided other refugee and immigrant populations with services and included those populations in the field test (Allene Grognet, director, BEST Development Project, personal communication, January 14, 1999). However, the test population data reported in the MELT documents includes only refugee populations.

The BEST was made available to ORR-funded programs from the ORR Refugee Materials Center in Kansas City, Missouri, until it closed at the end of 1987. At that time, CAL decided to reprint form B and make it available through CAL. Form C was eventually reprinted as well. The form B oral interview section is described in Exhibit 6.5.

Using only forty-nine items, the BEST oral interview attempts to assess language proficiency from eight topic areas or domains, across eight proficiency levels, using the four response modalities of speaking, listening, reading, and writing. The result is that very few items are actually related to the theoretical model at each proficiency level. The length of the test and the number of items are constrained by the need to administer each interview individually. The consequence of this time constraint and the desire to develop a broadly defined scale of performance levels in a single test is that the test loses stability when used to predict the exact proficiency level of individual students. In fact, the test developers recognized that the BEST discriminates better at the lower SPLs (0ñVI) than the higher ones (VIIñX). In 1992, CAL convened a meeting with potential users of a higher-level BEST to explore the general design and preliminary specifications of a test that would discriminate at the higher levels (CAL, 1992, summary of the higher-level BEST test meeting). Lack of funding prevented the test development from proceeding. The BEST, however, does elicit extended responses for fluency questions so that the proficiency level of the learners can be probed more deeply than is allowed for in the NYS Place.

NEW YORK STATE PLACEMENT TEST FOR ENGLISH AS A SECOND LANGUAGE ADULT STUDENTS (NYS PLACE). The NYS Place is primarily an oral picture description task, cued by brief oral questions from the examiner. The test has an optional initial oral screening component, or oral warm-up, to determine if the examinee has sufficient English to proceed with the test. These seven items are simple greetings and directives. If the examinee fails to answer three items in a

row, testing is suspended. There is also a literacy screen with items asking the test taker to read letters, numbers, and words in isolation as well as one short sentence. Exhibit 6.6 presents a description of Form B, the only form currently available.

The NYS Place, developed by the State Education Department of New York, is available from the City School District of Albany. It was developed specifically for the initial placement of adults in ESOL programs, and the content is generally accessible to immigrant populations. The NYS Place, along with the BEST, is one of the few tests of spoken English for adults. It is a highly structured test. Unlike the BEST, it does not give the examinee an opportunity to initiate or elaborate at any time. Therefore, the test protocol cannot be represented as eliciting authentic conversation.

GENERAL OBSERVATIONS ON TESTS. There are probably as many definitions of language proficiency as there are programs. Because language has so many facets and so many uses, different tests approach different aspects of language proficiency. Over the years, proficiency testing has reflected changes in our understanding of language theory. It has moved from a structural view (for example, discrete point tests of grammar, phonology, and other components of language), through a sociolinguistic view (integrative tests such as cloze and dictation), to a communicative view (for example, oral proficiency interviews that assess the learner's ability to use language to carry out communicative tasks) (Manidis & Prescott, 1996). Today, given the focus on real-life, practical content in adult ESOL instruction and the goals of the learners, if a test does not in some way look at language as communication, it would seem to be missing much that is important.

Nevertheless, most items in most tests do not relate directly to either theoretically or empirically derived understandings of adult English-language proficiency development. One might assume that if a test is constructed in English and requires responses in English, then higher scores will correspond to higher levels of English proficiency. But this is a very shaky foundation on which to build proficiency assessment. To the extent that the content of the items is known to and within the experience of all examinees, score differentials will more directly reflect language proficiency differentials.

A number of the tests designed for English-language learners reviewed here relate scores to some broadly defined scale of proficiency levels. These proficiency levels tend to be described in very global terms, often corresponding to the complexity of communicative situations (for example, greetings and leave-takings contrasted to explanation and persuasion). The actual items in a test, however, may be particular to a

competency area and may sample very narrowly from the broad ranges of behavior described by the proficiency levels. The generalizability from performance on the test items back to all situations consistent with a particular level description may be difficult to establish. When there is only one form of a test or a very small set of alternate forms, the possibility increases that learners begin to learn the test or that teachers teach to the test.

The use of a single test form to assess the full spectrum of proficiency levels also means that most items will not match any particular test taker's current level of functioning—that is, there are too few items at any one level of proficiency. Tests with multiple levels usually make more accurate assessments of functioning level. This can also be achieved in direct measures of proficiency, such as oral interviews, if the items engage the examinee in extended dialogue so that proficiency is assessed based on learner response and not the difficulty of the question.

Alternative Assessment in Adult ESOL

In the reviews of standardized tests, we pointed out the difficulty of using a single assessment instrument to provide information useful for placement, instructional decisions, and accountability. This fact has led many adult ESOL program staff to develop program-based alternative assessments. Alternative assessment is "any method of finding out what a learner knows or can do, that is intended to show growth and inform instruction, and is not a standardized or traditional test (Valdez-Pierce & O'Malley, 1992, p. 1).

Alternative assessments may take the form of performance assessment, portfolio assessment, or learner self-assessment. In performance assessment, the learner uses prior knowledge and recent learning to accomplish a task related to general language use or relevant to a specific context (Lumley, 1996). The learner response to, or outcome of, performance assessment may be an oral or written report, an individual or group project, an exhibition, or a demonstration (O'Malley & Pierce, 1996). A portfolio is a systematic collection of learner work that represents progress and achievement in more than one area (Fingeret, 1993). In learner self-assessment the learners monitor their own progress and accomplishments in order to select learning tasks and plan their use of time and resources to accomplish those tasks (O'Malley and Pierce, 1996).

Alternative assessments are consistent with emerging models of language acquisition. These models examine how people acquire language competence as they use the language in social interaction to accomplish purposeful tasks, such as to give and receive information and to make requests (August & Hakuta, 1997).

During the past decade, several studies and publications on best practices have guided the development of alternative assessments. The most comprehensive that was written specifically for adult ESOL is *Bringing Literacy to Life*, a document prepared for the U.S. Department of Education by Aguirre International (Wrigley & Guth, 1992). Other publications include the following:

- *Assessing Success in Family Literacy Projects: Alternative Approaches to Assessment and Evaluation* (Holt, 1994)
- *Adventures in Assessment* (McGrail & Simmons, 1991ñ1998)
- *Making Meaning, Making Change: Participatory Curriculum Development for Adult ESL Literacy* (Auerbach, 1992)
- *It Belongs to Me: A Guide for Portfolio Assessment in Adult Education Programs* (Fingeret, 1993)
- *Authentic Assessment for English Language Learners: Practical Approaches for Teachers* (O'Malley & Pierce, 1996)

The last book, though focused on Kñ12 education, can be helpful to adult ESOL programs as well. It devotes considerable attention to assisting teachers in developing assessment tasks and scoring rubrics (that is, rating scales) that ensure the reliability and validity of alternative assessments. Examples of assessment tools and processes advocated in these materials include learner-teacher conferences; questionnaires and surveys; teacher observation forms; checklists of communication skills and behaviors (Crandall & Peyton, 1993); and learner reading, writing, and speaking logs. Learners might prepare narrative writings or keep journals in which they express what they have learned in class; what changes they have made in their language and literacy practices or interactions; and how their goals, needs, and interests have been met or modified (Auerbach, 1992).

Although these alternative assessments are not standardized, they should be consistent with the following principles:

- They are program based, reflecting the program's underlying philosophy of instruction.
- They are learner centered, reflecting the strengths and goals of individual learners.
- They are done with the learner, not to the learner, so that learners are actively involved in setting goals, discussing interests, deciding what to evaluate, and reflecting on their accomplishments.
- They focus on the learning process as well as outcomes, allowing learners to reflect on their progress and make changes in how they are using their time and resources.

- In addition to the linguistic dimension of language and literacy development (for example, vocabulary and grammar), these assessments focus on the metacognitive (for example, developing learning strategies) and affective (for example, increased confidence) dimensions.
- They involve a variety of procedures, not just a single process or tool.

Alternative assessments provide teachers and learners with valuable feedback on learner progress and instructional changes that may need to be made.

Despite the fact that alternative assessment is guided by these principles and affords greater flexibility than standardized tests in gathering a more complete picture of what learners know and can do, its use as a means for accountability has raised serious questions. Without the development of guidelines and rigorous procedures for the collection and evaluation of evidence of learner performance and without the proper training of staff in how to carry out the assessments, alternative assessments do not produce the reliable, hard data that sources of funding require. To dispel the uncertainty of the subjectiveness associated with such assessments, administration procedures and conditions would have to be strictly monitored and a minimum of two raters involved in assessing student performance (Lumley, 1996). Furthermore, the program-specific nature of the assessments, along with difficulty in aggregating the data across programs, make program comparisons difficult.

EDUCATION REFORM INITIATIVES THAT INFORM ADULT ESOL ASSESSMENT

The issues of what and how to assess in language and literacy development are not unique to adult ESOL education. Assessment reform movements in K-12 education and language education abroad face similar challenges, although in some respects they are ahead of adult education in the United States, and the field can learn from their experiences. The next two sections briefly review some initiatives in K-12 education and abroad.

K-12 Education

In K-12 education in the United States, the assessment reform of the early 1990s was tied to the K-12 standards movement, which resulted from the passage of Goals 2000 legislation and the Improving America's Schools Act of 1994. A number of national, state, local, and professional groups have been developing content standards-what students should know and what schools should teach and assess-in a broad array of subjects, including art, foreign language, geography, history, language

arts, mathematics, science, and social studies. In some content areas, performance standards-what students must know and be able to do to demonstrate proficiency in the content area-are also being developed. Performance standards can be used to guide the development of assessment tools and processes. A third type of standard, referred to as opportunity to learn, defines what resources (for example, staffing and materials) are needed to make sure students will be able to meet the content and performance standards (August & Hakuta, 1997).

Early in the standards movement, Teachers of English to Speakers of Other Languages (TESOL) formed a task force to ensure educational equity and opportunity for students learning English as a second or additional language. The task force became concerned that many of the K-12 academic content standards did not take into account the important role of language in academic achievement. The task force developed ESOL standards for pre-K-12 students that specify the language competencies English-language learners need to become fully proficient in English, both socially and academically (TESOL, 1997).

The framework for the standards is based on three overarching goals: developing competence in English in social language, academic language, and sociocultural knowledge. These goals are supported by nine standards, organized by grade-level clusters (pre-K-3, 4-8, and 9-12), that can be used to guide curriculum development at the state or local level. Descriptors of representative behaviors that demonstrate a standard is being met, along with sample progress indicators, are included in the document to assist in the development of curricular objectives and benchmarks for reporting student performance. Examples are included of how the sample progress indicators can be used across grade-level clusters to account for different English-language proficiency levels (beginning, intermediate, advanced, and limited formal schooling).

Exhibit 6.7 shows how tasks can be constructed across proficiency levels to demonstrate the same standard. It can be used to guide the development of assessment tasks to monitor the progress not only of K-12 English-language learners but also of adults. The goal in the example, "To use English to achieve academically in all content areas," is admittedly geared toward academics in pre-K-12 education. But the standard, "to obtain, process, construct, and provide subject matter information in spoken and written form," describes uses of language that adult learners identified as important to their own literacy development in the adult-focused Equipped for the Future Project (Stein, 1997) and that employers defined as important in a high-performing workforce (Secretary's Commission on Achieving Necessary Skills, 1993). The TESOL pre-K-12 standards document, along with its accompanying Scenarios for ESL Standards-Based Assessment (TESOL, 1999), can

serve as a model for creating performance assessment tasks that probe the functioning proficiency level of the adult learner.

Adult Language Education Abroad

Developments in assessing language learners in Australia, Canada, and Europe can inform efforts in adult education in the United States. Stakeholders in each country have expressed the need for a common way to measure and describe learner outcomes across a wide variety of programs. In Australia and Canada, there are strong links between immigration policy and economic and labor policy. As Europe moves toward closer ties under the European Union, educators face the daunting task of linking assessments in various member languages to a common framework in order to compare language certification across member countries.

AUSTRALIA. Australia traditionally has welcomed immigrants. Before World War II, most immigrants to Australia came from Ireland and the United Kingdom. After the war, Australia's immigration policies looked to non-English-speaking countries to meet the demands for new settlers. As a consequence, the Adult Migrant Education (now English) Program (AMEP) was established to provide these new settlers with English-language instruction and settlement information.

The AMEP has become the largest government-funded English-language training program in the world. At first a centralized curriculum was designed by the government, but in the 1970s curriculum development was decentralized so that teachers at individual programs became the primary developers of curriculum, learner placement, needs assessment, and procedures for monitoring progress (Burns, 1994). In the 1990s, a collaborative initiative was undertaken by the Australian government, industry, and unions to develop a curriculum framework linking industry needs with training in workplace competence (Hager, 1996). This occurred in response to changing world economic conditions, a need for increased continuity between language-training programs and vocational and job-training programs, and a demand for measurable outcomes. The result has been the development of a number of competency-based curriculum documents that are nationally or state accredited. Among the most well known is the Certificate in Spoken and Written English (CSWE), developed by the New South Wales Adult Migrant English Service and the National Center for English Language Training and Research (NCELTR).

The aim of the CSWE is to enable adult English-language learners to develop the language and literacy skills that will enable them to participate in further education or training, seek and maintain employment, and become contributing members of the community.

Competencies that describe what learners can do at the end of study are identified at three stages of proficiency (beginning, postbeginning, and intermediate), based on the Australian Second Language Proficiency Ratings (ASLPR). Within each stage, learners can be grouped by a slow, standard, or fast learning pace to accommodate for differences in educational experiences and native language literacy abilities.

The CSWE document is considered a curriculum framework rather than a syllabus, from which individual programs can create their own courses of study. However, the document specifies the explicit criteria under which the competencies must be assessed. Within these criteria, the competencies may be assessed through teacher observations, interviews, role plays, learner self-assessments, and other means, following the guidelines. Possible exit points occur at the end of each stage, but the certificate is issued only after Stage 3 competencies have been achieved. It takes a learner approximately 250 to 300 hours to pass through a stage. (For a detailed description of the CSWE, see New South Wales Adult Migrant English Service, 1993.)

Initially teachers were concerned about the appropriateness of formal assessment with their learners, the time it takes to assess performance, and validity and reliability issues among teachers and across programs. After an initial adjustment period, teachers found that the new system enabled them to focus their teaching in a way that provided clearer direction and more explicit feedback to learners on progress (Burns & Hood, 1995). Issues of validity and reliability continue to concern Australian adult educators as proficiency assessments based on rating scales are increasingly used for accountability purposes (Manidis & Prescott, 1996; Brindley, 1999).

Another concern is the apparent inadequacy of the system for assessing the progress of learners who have limited experience with formal learning, low literacy levels in both their native language and English, and difficulty adjusting to new cultural expectations (Jackson, 1994). These same concerns for the low-level learner have surfaced in the United States as well. These learners tend to require longer periods of instruction and make smaller language gains than more literate learners. The gains they make tend to be nonlinguistic outcomes in the affective, cognitive, and sociocultural domains.

A research project undertaken by the NCELTR identified eight major categories of nonlinguistic outcomes: confidence; social, psychological, and emotional support in one's living and learning environments; knowledge of social institutions; cultural awareness; learning skills; goal clarification; motivation; and access and entry into further study, employment, and community life (Jackson, 1994). Despite the

incorporation of nonlanguage outcomes in the knowledge and learning competencies of the CSWE (New South Wales AMES, 1993), teachers lack adequate tools to assess and document these outcomes. Research has shown that these outcomes reflect characteristics of good language learners and appropriate language teaching methodologies (Jackson, 1994). They also reflect characteristics and skills favored in the workplace (Hager, 1996; Secretary's Commission on Achieving Necessary Skills, 1993). They should be considered in every assessment system, and teachers need help in finding ways to assess and record them.

CANADA. In response to the field's expressed need for a common set of standards to measure and describe language development of learners in Canadian adult ESOL programs, Employment and Immigration Canada (now Citizenship and Immigration Canada [CIC]) developed national English-language benchmarks with the assistance of a working group of ESOL learners, teachers, administrators, immigrant service providers, and government officials (CIC, 1996; Pierce & Stewart, 1997).

The Canadian Language Benchmarks (CLB) project, launched in spring 1996, is a task-based descriptive scale of ESOL proficiency. It identifies twelve benchmarks across three stages of proficiency (basic, intermediate, and advanced) in three skill areas: speaking/listening, reading, and writing. Each benchmark describes a person's ability to use English to accomplish a task and includes information on the abilities of the person (for example, "Can copy information, describe personal situations"), performance conditions (for example, "Copies words and numbers clearly and accurately"), situational conditions (for example, time limit, length of text, number of mistakes allowed), background knowledge helpful to completing a task (for example, cultural expectations, availability of community services, note-taking conventions), and sample tasks. As compared with the American SPL descriptors, the level of detail in the Canadian benchmarks makes the document more useful for creating assessment tasks. The CLB does not purport to be a proficiency test or a syllabus, but curriculum writers, materials developers, and practitioners can use it as a guide for syllabus development, instruction, and monitoring of learner progress. (For a detailed description of the benchmarks see CIC, 1996.)

Concurrent with the field testing of the draft CLB document, CIC contracted with the Peel Board of Education in Ontario to develop the Canadian Language Benchmarks Assessment (CLBA) to assist in placing learners in programs and assessing learner progress. It was to be a task-based assessment instrument that addressed Benchmarks 1 through 8 (Stages 1 and 2) in each of the three language skill areas. The developers faced many challenges: the instrument had to consist of tasks

representative of those identified in the CLB document, reconcile the assessment of each separate skill area with the holistic approach implicit in task-based assessment, balance cultural bias and authenticity of task, and be user friendly-for both the test giver and the test taker-across a wide range of program types and settings (Pierce & Stewart, 1997).

The resulting CBLA consists of three separate instruments-one for each of the skill areas. The listening/speaking assessment is administered one-on-one and is scored by the interviewer as it is being administered. It takes ten to thirty minutes to complete. The reading and writing assessments can be administered individually or in a group setting and take at least another one to two hours to administer. There are parallel forms of the reading and writing assessments-two to be used for placement and two for assessing outcomes.

The developers consider the CBLA a work in progress. Assessors are being trained, and an interrater reliability study is underway. They caution, however, that it remains a low-stakes instrument that would need to have its validity and reliability enhanced before it could be used for a high-stakes assessment such as job entry, academic opportunity, or immigration (Pierce & Stewart, 1997).

EUROPE. The assessment reform effort in Europe is perhaps more complex than the efforts in Australia, Canada, and the United States because the Europeans are attempting to correlate assessments in many languages to a common set of standards. Two projects in particular, the Council of Europe's Common European Framework and the Association of Language Testers in Europe's (ALTE) Framework Project, are worth noting. The first establishes a comprehensive framework for the description of language proficiency and its relationship to content. The second then applies this framework to the various examinations and certificates offered in member countries to promote the recognition of language certification across Europe.

Common European Framework. The Council of Europe assists its forty-four member states in encouraging all citizens to learn nonnative languages to promote mutual understanding, personal mobility, and access to information in a multilingual, multicultural Europe (van Ek & Trim, 1996). The second draft of the Modern Languages: Learning, Teaching, Assessment-A Common European Framework of Reference (Council for Cultural Cooperation Education Committee, 1996) represents the council's latest effort in a collaborative process that began in 1971. The document identifies four contexts for language use (personal, public, occupational, and educational) and specifies ranges of language knowledge and skills. Language programs can use the document to develop curricula, select instructional approaches and

techniques, and establish assessment procedures. It includes information on the purposes of communication (for example, completing insurance forms, taking public transportation, reporting an accident); communication activities (productive, receptive, interactive) and processes (for example, plan, organize, execute); texts (for example, spoken: public speeches, news broadcasts; written instructions, letters, signs); learning strategies; the processes of language learning and teaching; establishment of levels and scales; and assessment purposes and types. It is not prescriptive in the sense of telling practitioners what to do or how to do it. Rather it provides a range of elements for programs to choose from and establishes principles for language teaching, learning, and assessment to facilitate program design. It thus provides a common basis for discussing these choices among programs.

ALTE Framework Project. The ALTE, established in 1989, is an association of European institutions that offers learners language training and certification in the language of the institution's country or region. Fifteen languages are represented among the association's eighteen members. As the European workforce becomes more mobile within the European Union, both employees and employers need to know what language qualifications mean across countries. To meet this need, ALTE members are establishing common levels of proficiency and common standards for the language testing process (ALTE, 1998). They have drawn from previous work of the Council of Europe that specified the descriptions of language ability (van Ek & Trim, 1996) as well as from their collaboration with the council on the Common European Framework project.

The ALTE framework identifies five main levels of proficiency, each defined by statements of what a user can do at that level across the productive (speaking and writing) and receptive (reading and listening) skill areas. Examinations given by ALTE members have been charted along this continuum by comparing the content and the demands each examination makes on the examinee (ALTE, 1994); this allows for comparisons of skill level attainment across language tests. For example, a learner of French as a foreign language achieving the Diplôme de Langue Française issued by Alliance Française would have the same proficiency in French that a learner of Italian would have in Italian upon achieving the Certificato di Conoscenza della Lingua Italia, Livello 3. The "can do" statements for each level are currently being validated so that they can be used to describe what a language test score actually means in the real world (for example, "Can offer help to a client or customer: 'I'll give you our new catalogue'") (Jones, 1999).

Works in Progress. Both the Council of Europe and the ALTE consider their frameworks to be works in progress-to be used, commented on, and

further developed in response to experience gained from using the documents, to developments in research in language acquisition and learning, and to new needs that may arise.

Lessons from These Assessment Initiatives

Assessment initiatives in adult education in the United States can apply the lessons learned from K-12 efforts and adult-language education efforts abroad. The primary lesson to be drawn from these efforts is that in the overarching structure, assessment is but one component in a larger instructional system. The K-12 system is well established in the United States. Australia's Adult Migrant English Program was strongly backed by government funding initiatives and linked to a major government-funded research institute, the National Center for English Language Training and Research (NCELTR), which works with practitioners to produce research and provide training that reflect the practical classroom issues language teachers face (for example, assessing oral proficiency [Manidis & Prescott, 1996]). The Common European Framework project is producing a comprehensive compendium of elements of a language instructional system so that programs can base their instructional designs on solid knowledge about language acquisition and principles of language teaching and assessment.

These initiatives also show that assessment reform does not begin with assessment procedures but with content standards-what learners need to know and be able to do to function successfully in the communities in which they live. Once the content is identified, curriculum and instructional approaches are chosen and performance standards and assessment procedures developed. This does not mean that every program that uses the pre-K-12 TESOL standards document, the Canadian Language Benchmarks, or the Common European Framework will be teaching and assessing the same things in the same way. However, these documents enable individual programs to select elements that match program and learner purposes and goals, while at the same time providing a common frame of reference among the programs.

Although the requirements of the Workforce Investment Act and the development of the National Reporting System represent strong moves toward a national accountability system in the United States, by and large, there is no national comprehensive learning system for adult education (Merrifield, 1998) with which accountability and assessment fit coherently. What to assess, how to assess, how to report the data, what they mean, and how to use them are questions that still need to be agreed on. This does not mean that traces of such a system do not exist. In fact, there are initiatives at local, state, and federal levels that indicate the field is lurching toward building a system. For example, many programs have developed instructional designs with articulated mission statements and

goals, approaches to curriculum and instruction, and procedures for assessment. The way these components are defined varies, depending on the type of program and the goals of the program and its learners. What is lacking is a common framework, as in the European model, to enable programs to compare what they are doing across their varying types and missions.

Several states, including Massachusetts and California, have elements of statewide systems in place. California was the first state to identify standards for adult ESOL. The English-as-a-Second-Language Model Standards for Adult ESL (California Department of Education, 1992) provides standards for program design, curriculum, instruction, and student evaluation. The document also identifies proficiency levels and describes course content and sample lessons for each level. The Massachusetts effort to establish standards is part of a larger effort across the state to implement standards in all of adult education. The first step in the process is to establish a curriculum framework that identifies what learners need to know and be able to do. From this framework, assessment and implementation phases will follow (Massachusetts Department of Education, 1998).

On the national level, the National Institute for Literacy's Equipped for the Future (EFF) project is attempting to answer the question, "What is it that adults need to know and be able to do in order to be literate, to compete in the global economy, and to exercise the rights and responsibilities of citizenship?" (Stein, 1997, p. 2). The goal is to establish content standards that articulate what adults need to know and be able to do to be effective family members, community members, and workers. These standards will cut across the four purposes of literacy identified in phase 1 of the project: gain access to information, give voice to ideas, act independently, and build a bridge to the future by learning how to learn. ESOL programs in several states have been involved in the development and pilot testing of the content standards. These standards can provide the basis for establishing performance standards from which programs and states can develop assessment systems. In fact, the EFF team is already beginning to conceptualize what an assessment system for the standards might look like (National Institute for Literacy, 1999).

Another project at the national level is the "What Works" Study for Adult ESL Students Evaluation, a five-year study sponsored by the U.S. Department of Education and being conducted by the Pelavin Research Center and Aguirre International. Its purpose is to evaluate the effectiveness of instructional approaches for adult English-language learners with limited literacy skills and then make recommendations to the field. A small part of the study will examine the types of assessment and outcome measures that are being used and identify those that are

most appropriate for these learners.

Finally, TESOL has appointed a task force to develop adult education program standards and sample performance measures that will address program goals, structure, implementation, curriculum, instruction, and assessment (Bitterlin, 1997). This document will be a useful guide for program staff and policymakers in establishing performance measures to ensure high-quality educational services for adult learners of English.

RECOMMENDATIONS FOR POLICY, PRACTICE, AND RESEARCH

The field of adult education in the United States is in the midst of implementing accountability systems, as is the field of K-12 education. However, in adult education this is being done without the benefit of the infrastructure of the K-12 system. Programs and states need support from both researchers and those providing funds to develop accountability systems that are part of a larger adult education instructional system, effectively capture what learners know and can do, and can be efficiently and realistically carried out at the program level. The following recommendations focus on the policies, practice, and research needed to make this possible.

Policy

The Workforce Investment Act (with its emphasis on performance measures) and the National Reporting System (with its emphasis on functioning levels) are driving the types of accountability systems that states are formulating. The NRS gives only limited examples of test benchmarks—ranges on CASAS tests and SPLs—to assess learner functioning levels. The SPLs are descriptors of learner abilities and not a test. Determining the SPL of an individual learner can be accomplished by correlating the local program level to which the learner is assigned to the SPL or assigning the SPL according to teacher judgment and verifying with the BEST test (U.S. HHS, 1985). Also, although not as detailed as the Canadian Language Benchmarks, the SPLs can guide the development of informal performance-based assessments (Grognet, 1998). However, it is doubtful that these two benchmarks (CASAS ranges and SPLs) are adequate for assessing the functioning level gains of learners for accountability purposes in all types of adult ESOL programs. The BEST test or formal performance-based assessments (valid, reliable, consistently administered, and perhaps using multiple raters) correlated to the SPLs would need to be administered to ensure that a standardized assessment procedure was used to assess learner gains using the SPLs. Furthermore, the assumptions about curricular content and instructional approaches that underlie the CASAS, and to some extent the BEST, are not necessarily the assumptions that

undergird all adult ESOL programs. In this chapter we have also pointed out the dangers of assuming that a higher test score translates into increased language proficiency when the tests contain too few items at any single proficiency level or assess too few contexts to ensure an accurate proficiency level assessment.

First, agreement needs to be reached as to what constitutes literacy development among adult ESOL learners and how it can be measured for accountability purposes as learners progress. Assessment measures should reflect one's view of literacy and be comprehensive enough to show the full range of learner achievement. If the measures are to be made into tests, the resources are needed to improve the instruments now being used. The SPLs should be revised and revalidated in view of increased knowledge of language development and changing practices since they were first formulated in the 1980s (Grognet, 1998). Sample tasks for each proficiency level, such as those given in the TESOL standards document and the CLB, need to accompany the descriptors. If new standardized instruments are to be developed, they should be able to accommodate learners at every stage of language and literacy development. If alternative measures are acceptable, programs will need clear guidelines on how to select or develop those measures that are most appropriate for their learners and meet the legislative requirements for a standardized and rigorous process.

In addition, consideration must be given to the long time adults need to become fluent in another language. SPL studies indicate that a low-level learner (SPL 1) needs between 110 and 235 hours to advance to the next proficiency level, depending on the characteristics of the program (for example, intensity of instruction and class size) and the learner (for example, educational experience, health, and motivation) (U.S. HHS, 1985). Australian studies support the need for many instructional hours, especially for learners with low literacy skills, to show gains in proficiency level. The National Evaluation of Adult Education Programs (NEAEP) study, conducted between 1990 and 1994, reported that the median amount of time that ESOL learners stay in programs is 113 hours (Fitzgerald, 1995). Furthermore, the initial gains that low-level learners make tend to be nonlinguistic. As mentioned earlier, these are the skills and knowledge that may make it possible to continue learning and may be important in the workplace, but they do not show up on existing tests. Any test or proficiency scale needs to describe sufficiently the benchmarks on the way to achieving the next level, so that even learners with limited literacy skills or little time available to devote to formal instruction can show progress.

Programs need additional resources to implement assessment requirements-whether in the development of rigorous performance

assessments, teacher training on using the assessment tools and procedures, or the ability to compensate teachers for time to assess learners and document results. Policymakers at the federal, state, and local levels must work with program administrators to determine how resources can be allocated to provide this support.

The first few years of WIA implementation will yield rich data for analysis of what is working, for whom, and what adjustments may have to be made to make the system work for everyone. The development of databases about what works best for various types of learners can identify elements of program design that will meet the needs of each stakeholder. Program staff need this information in clear and understandable language and in a timely fashion if they are to make well-informed decisions. Researchers need this information to guide their research and development efforts.

The development of an accountability system should be just one component in an educational system that also encompasses guidelines for curricular content and instructional practice. A comprehensive system such as the Common European Framework presents a model for policymakers, program administrators, practitioners, and researchers to consult as results from EFF, the "What Works" Study, the NRS, and various research studies are reported.

Practice

Practitioners need to be aware of efforts such as EFF, the "What Works" Study, and TESOL's Task Force on Program Standards. EFF can provide program staff with information for evaluating curriculum content, thereby helping to increase their awareness of what counts as success to learners across a spectrum of program types and, therefore, what should be taught and assessed. The "What Works" Study can shed light on best practices for learners with limited literacy skills in both instruction and assessment. TESOL's program standards can be used to guide the development of assessment procedures within adult education programs. These three efforts can work together to provide the types of programs that meet the needs of each stakeholder.

As practitioners consider what counts as progress and how it can best be measured for the learners they serve, they must be able to present and support their decisions before administrators and funding agencies and organizations. Selection of assessment measures should be made with solid knowledge about the proficiencies of learners, the educational context of the program, and the adequacy of the assessment measures being considered to carry out the purpose of the assessment.

Research

To inform both policy and practice, research efforts should address issues that the field itself has identified as critical in the area of assessment and outcomes. These issues are compiled in the Research Agenda for Adult ESL (CAL, 1998, p. 11), in which the following questions are raised:

- What immediate and long-term impact can be expected from the various types of adult ESOL programs? What impact does learner participation in such programs have on the learners and their communities?
- How can adult ESOL programs best capture what learners know and what they have learned?
- What is the cost in time, staffing, and funds to assess and document learning outcomes effectively?
- How can each of the stakeholders in a program participate in determining what counts as progress?
- How do measures of program impact, such as an increase in reading to one's children or a job promotion, correlate with increases in English-language proficiency?
- How might a national proficiency scale facilitate the reporting of learner progress and program impact? How effective is the NRS scale in reporting learner gains?
- Which assessment instruments can reliably document changes in learner performance at what levels? Can these instruments serve all types of adult ESOL programs?
- What changes in program design and staff development are needed to ensure that current and new assessment tools are reliably used?
- How can technology facilitate the implementation of a system for documenting learner outcomes and program impact?
- How do local, state, and national policies affect assessment tools and practices, and what policies need to be created?

CONCLUSION

Whatever the decisions that will be made in the near future about assessment in adult ESOL, it is clear that policymakers, practitioners, and researchers must engage in a collaborative effort to produce substantive, purposeful, and effective change in the field. The U.S. Department of Education's (ED) Office of Vocational and Adult Education continues to make efforts to listen to stakeholders through such venues as the biannual meetings of the State Directors of Adult Education and the studies and the clearinghouse it funds. The ED also sponsors a National Forum on Adult Education and Literacy, which in 1997 brought in learners from across the United States and, in 1998, teachers, to discuss adult education issues. However, more substantive efforts are needed to create an

infrastructure that can sustain a comprehensive adult education system. At local, state, and national levels, representatives from each stakeholder group (policymakers, practitioners, researchers, and learners) should develop a plan of action that supports the implementation of the new legislative imperatives and program designs that effectively and efficiently serve the needs of the community and the learners.

Appendix: Ordering Information for Selected Standardized Tests

ABLE (Adult Basic Learning Examination)
Psychological Corporation
Order Service Center
P.O. Box 839954
San Antonio, TX 78283
(800) 211-8378

A-LAS (Adult Language Assessment Scales)
McGraw-Hill
1221 Avenue of the Americas
New York, NY 10020
(800) 624-7294
<http://www.mhcollege.com>

BEST (Basic English Skills Test)
Center for Applied Linguistics
4646 Fortieth Street, N.W.
Washington, DC 20016ñ1859
(202) 362-0700
<http://www.cal.org>

CASAS (Comprehensive Adult Student Assessment System)
8910 Clairemont Mesa Boulevard
San Diego, CA 92123
(619) 292-2900
<http://www.casas.org>

NYS Place Test (New York State Placement Test for ESOL
Adult Students)
City School District of Albany
Albany Educational TV
27 Western Avenue
Albany, NY 12203
(518) 462-7292, ext. 30

TABE (Test of Adult Basic Education)
CTB/McGraw-Hill

20 Ryan Ranch
Monterey, CA 93940
(800) 538-9547
<http://www.ctb.com>

References

- ALTE. (1994). Using the ALTE framework. *ALTE News*, 3(1), 1ñ3.
- ALTE. (1998). ALTE document 1: European language examinations and examination systems. Great Britain: Author.
- Auerbach, E. (1992). Making meaning, making change: Participatory curriculum development for adult ESL literacy. Washington, DC, and McHenry, IL: Center for Applied Linguistics and Delta Systems.
- August, D., & Hakuta, K. (Eds.). (1997). Improving schooling for language-minority children: A research agenda. Washington, DC: National Academy Press, 1997.
- Bitterlin, G. (1997). Program standards for adult-level ESL programs. *TESOL Matters*, 7(5), 9.
- Brindley, G. (1999, March). Task effects in second language assessment. Paper presented at Teachers of English to Speakers of Other Languages, Thirty-third Annual Convention and Exposition, New York.
- Burns, A. (1994). Adult ESL in Australia. *Digest of Australian Languages and Literacy*, 7, 1ñ4.
- Burns, A., & Hood, S. (Eds.). (1995). *Teachers' voices: Exploring course design in a changing curriculum*. Sydney, Australia: National Center for English Language Teaching and Research.
- Business Council for Effective Literacy. (1990). Standardized tests: Their use and misuse. *BCEL Newsletter*, 22, 1, 6ñ8.
- California Department of Education. (1992). English-as-a-second-language model standards for adult education programs. Sacramento, CA: Bureau of Publications.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *American heritage word frequency book*. Boston: Houghton Mifflin.
- Center for Adult Education. (1983). *Handbook for CBAE staff development*. San Francisco: San Francisco State University.

Center for Applied Linguistics. (1998). Research agenda for adult ESL. Washington, DC: Center for Applied Linguistics.

Citizenship and Immigration Canada. (1996). Canadian language benchmarks: English as a second language for adults and English as a second language for literacy students. Ontario: Ministry of Supply and Services Canada.

Council for Cultural Cooperation Education Committee. (1996). Modern languages: Learning, teaching assessment: A common European framework or reference. Strasbourg, France: Council of Europe.

Crandall, J. (1992). Literacy, language, and multiculturalism. In J. E. Alatis (Ed.), *Linguistics and language pedagogy: State of the art*. Washington, DC: Georgetown University Press.

Crandall, J., & Peyton, J. (Eds.). (1993). *Approaches to adult ESL literacy instruction*. Washington, DC, and McHenry, IL: Center for Applied Linguistics and Delta Systems.

Duncan, S. E., & DeAvila, E. A. (1993). *Adult language assessment scales: Administration and scoring manual*. Monterey, CA: CTB/Macmillan/McGraw-Hill.

Ebel, R. I. (1979). *Essentials of educational measurement*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

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

Fitzgerald, N. B. (1995). *ESL instruction in adult education: Findings from a national evaluation*. Washington, DC: Center for Applied Linguistics.

Fitzpatrick, A. R. (1992). Review of the adult basic learning examination, second edition. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Grognet, A. (1998). *Performance-based curricula and outcomes: The mainstream English language training (MELT) project updated for the 1990s and beyond*. Denver: Spring Institute for International Studies.

Hager, P. (1996). The development of competency-based training: Government, industry, and union pressures. *Prospect: A Journal of*

Australian TESOL, 11(1), 71-78.

Hambleton, R. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA: Sage.

Holt, D. (Ed.). (1994). *Assessing success in family literacy projects: Alternative approaches to assessment and evaluation*. Washington, DC, and McHenry, IL: Center for Applied Linguistics and Delta Systems. (ERIC Document Reproduction Service No. ED 375 688)

Jackson, E. (1994). *Non-language outcomes in the adult migrant English program*. Sydney, Australia: National Center for English Language Teaching and Research.

Jones, N. (1999, March). *Validating can do scales*. Paper presented at Teachers of English to Speakers of Other Languages, Thirty-third Annual Convention and Exposition, New York.

Kahn, A. B., Butler, F. A., Weigle, S. C., & Sato, E. (1995). *Adult English-as-a-second-language assessment project: Final report: Year 3*. Los Angeles: UCLA Center for the Study of Evaluation.

Kenyon, D., & Stansfield, C. (1989). *Basic English skills test manual*. Washington, DC: Center for Applied Linguistics.

Lumley, T. (1996, June). *Assessment of second language performance*. Digest of Australian Languages and Literacy Issues.

Manidis, M., & Prescott, P. (1996). *Assessing oral language proficiency: A handbook for teachers in the adult migrant English program*. Sydney, Australia: National Centre for English Language Teaching and Research.

Massachusetts Department of Education. (1998). *Framework for adult ESOL in the Commonwealth of Massachusetts*. Rev. draft. Boston: Author.

McGrail, L., & Simmons, A. (Eds.). (1991-1998). *Adventures in assessment*. Boston: SABES.

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

National Center for Educational Statistics. (1997). *Participation of adults in English as a second language classes: 1994-1995*. Washington, DC:

U.S. Department of Education.

National Clearinghouse for ESL Literacy Education. (1989). Basic English skills test manual. Washington, DC: Author.

National Institute for Literacy. (1999, January 6ñ7). Developing national qualifications framework for adults. Paper provided for Equipped for the Future Expert Review Session, Washington, DC.

New South Wales Adult Migrant English Service, and National Center for English Language Training and Research. (1993). Certificate in spoken and written English. Sydney, Australia: Authors.

Office of Vocational and Adult Education. (1992). Model indicators of program quality for adult education programs. Washington, DC: U.S. Department of Education.

Office of Vocational and Adult Education. (1995). Adult education for limited English proficient adults. Washington, DC: U.S. Department of Education.

Office of Vocational and Adult Education. (1997, October 9). Thursday notes: A weekly fact sheet on adult education issues. Washington, DC: U.S. Department of Education.

Office of Vocational and Adult Education. (1998, August). Measure definitions for the national reporting system for adult education. Washington, DC: U.S. Department of Education.

O'Malley, M., & Pierce, L. (1996). Authentic assessment for English language learners: Practical approaches for teachers. Reading, MA: Addison-Wesley.

Pierce, B. N., & Stewart, G. (1997). The development of the Canadian language benchmarks assessment. *TESL Canada*, 14(2), 17ñ31.

Secretary's Commission on Achieving Necessary Skills. (1993). *Teaching the SCANS Skills*. Washington, DC: U.S. Department of Labor.

Short, E. (1997). Why a statewide assessment system? *Developments*, 1(2), 2ñ3.

Solorzano, R. (1994). Instruction and assessment for limited-English-proficient adult learners. Philadelphia: National Center on Adult Literacy. (ERIC Document Reproduction Service No. ED 375 686)

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

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

Teachers of English to Speakers of Other Languages. (1997). *ESL standards for pre-K-12 students*. Alexandria, VA: Author.

Teachers of English to Speakers of Other Languages. (1999). *Scenarios for ESL standards-based assessment*. Working draft.

U.S. Department of Health and Human Services. (1985). *Mainstream English language training project (MELT) resource package*. Washington, DC: Author.

Valdez-Pierce, L., & O'Malley, M. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.

Van Ek, J., & Trim, J. (1996). *Vantage level*. Strasbourg, France: Council of Europe.

Williams, R. T. (1992). *Review of the adult basic learning examination, second edition*. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Workforce Investment Act of 1998, Public Law No. 105-220, sec. 212, U.S. Government Printing Office.

Wrigley, H. S. (1992). *Learner assessment in adult ESL literacy*. Washington, DC: Center for Applied Linguistics.

Wrigley, H. S., & Guth, G. (1992). *Bringing literacy to life: Issues and options in adult ESL literacy*. San Mateo, CA: Aguirre International. (ERIC Document Reproduction Service No. ED 348 896)

[Chapter 7](#) ➔