Improving Distance Education for Adult Learners

Working Paper No. 1

Assessment and Accountability Issues in Distance Education for Adult Learners
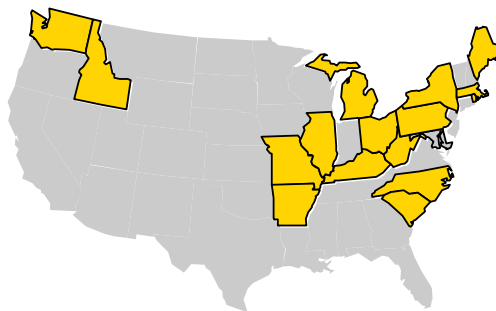
October, 2002

Project IDEAL is a consortium of states working together to develop effective distance education programs for adult learners. The Project IDEAL Support Center at the University of Michigan helps consortium states by developing training materials and web-based tools. The Center provides technical support in the areas of teacher training, research design, data collection, data analysis and reporting. Through collaborative research and practice, we are working to provide quality distance education for adult learners across the country.

Publications include:

- *Handbook of Distance Education for Adult Learners. Second Edition, November, 2003*

- *Assessment and Accountability Issues in Distance Education for Adult Learners. October, 2003.*

- *Accountability in Adult Education for Learners Studying at a Distance, December, 2003.*

- *Using Assessment to Guide Instructional Planning for Distance Learners. January, 2004.*

- *Study Guide for Distance Learning 101: Teaching Adult Learners at a Distance. Second Edition, November 2003. (Available only to member states)*

- *Trainer's Guide for Distance Learning 101: Teaching Adult Learners at a Distance. Second Edition, November 2003. (Available only to member states)*

For further information visit: www.projectideal.org

Shannon J. Young, Jerome Johnston, and Susanna E. Hapgood

Assessment and Accountability Issues in Distance Education for Adult Learners

First Edition, September 2002

First Printing: September, 2002

File: DEAssessment1stEdition.doc

Table of Contents

Acknowledgements

Executive Summary

Assessment and Accountability Issues in
Distance Education for Adult Learners

Shannon J. Young, Jerome Johnston, and Susanna E. Hapgood
Institute for Social Research - University of Michigan
First Edition, September, 2002

Under the banner of Project IDEAL (Improving Distance Learning for Adult Learners), adult educators in 13 states have joined together to explore the possibilities of using distance education to deliver instruction to adult basic learners who cannot easily attend classroom programs. A central issue for these states is how to treat distant learners in the NRS accountability scheme. This working paper frames the issues, poses possible solutions, and recommends several research studies to be undertaken in the coming year to inform policy decisions in this arena. A goal of this project is to convene a panel of experts in early 2003 to consider the issues laid out in the paper and recommend best practices for counting distance learners under the current NRS system. An additional goal of the project is to develop recommendations on ways the NRS system might be revised in the next authorization to better accommodate adults studying at a distance.

The Assessment Issues in Brief

From an accountability perspective, studying at a distance presents four problems.

1. **Measuring Seat Time.** Time spent studying is at the heart of "counting" adult learners as students and justifying expenditures on their behalf. After 12 hours of engaging in educational activities (intake, assessment, orientation) a learner is considered an official student. In many states, additional increments serve to trigger assessment activities aimed at seeing if the learner has made educational progress. Classroom programs use attendance at classes, labs and other sanctioned educational activities of known duration to assign hours to each student's record. This approach to measuring seat time can work for distance students for the time they are engaged in intake or orientation activities at a literacy center. But the bulk of their studying is done in a non-certifiable setting. *Issues:* How should "seat time" be measured for distance learners? Some distance programs that utilize online delivery of the instruction (e.g., *PLATO*, *SkillsTutor*, *GEDIllinois*) have mechanisms for measuring the elapsed time the user is connected to each lesson. Other programs that utilize online delivery have no mechanism for measuring elapsed time (e.g., *Workplace Essential Skills* and *GED Connection*). Many programs are based on print and video materials, and there are no precise ways to measure time spent in these activities. *Solution:* One solution would be to have a teacher examine a student's work and judge whether it meets a standard for completeness. If it does, the learner could be granted credit for having spent a pre-determined amount of time studying the lesson in question.

2. **Measuring Educational Gain Appropriately.** NRS gives states latitude in selecting standardized measures of educational progress. Most states have responded by recommending one or two assessments that apply to all adult learners regardless of the particular courses they take. To assess progress, a student must complete the same (or parallel) form of the test before and after instruction. But what if the recommended test is not a good measure of what a student could learn in a course? The testing would show minimal or no gain when in fact learners may have made gains. Learners (and the literacy centers that claim them as students) would be penalized. Of the curricular materials being selected by Project IDEAL states for teaching at a distance, there is a great deal of variation in the fit between the curriculum and state-mandated measures of educational attainment. These courses include: *Crossroads Café, PLATO, SkillsTutor, GED Connection*, custom online GED prep programs[1], and *Workplace Essential Skills*.

> *Crossroads Café.* There are well-accepted measures of educational progress in the English language skills taught in this series (e.g., the BEST). Data are needed to establish the link between studying *Crossroads Café* at a distance and educational gain on standardized measures.

> *PLATO & SkillsTutor.* These computer-assisted instruction (CAI) programs are designed to improve a learner's skills in the very areas measured by standardized tests of reading and math such as the TABE. While it is not known whether a student *will* improve his/her score by using the programs, the match between instructional content and assessment is generally good.

> *GED Connection & custom GED online courses.* Assessing educational progress while studying for a GED is a special situation in the NRS. To study for a GED, learners must first demonstrate that their reading and math scores are at the high school level.[2] While enrolled in GED-specific courses, the candidate must be tested for educational progress in reading/math using a standardized measure such as the AMES, ABLE, CASAS, or TABE. There is no evidence (and little reason to believe) that studying to pass the GED social studies or science tests would have the unintended side effect of improving general reading or math scores. In other words, the measure of educational progress is not well aligned with instruction. Some leaders in the field are urging that GED practice tests be accepted as measures of educational progress.

> *Workplace Essential Skills (WES).* WES presents unique assessment problems that apply to both classroom and distance learners. Content analyses of WES and the standardized tests recommended by NRS show an imperfect fit. The assessment problem varies by which of the four "strands" (content areas) of WES a learner is studying. Some standardized tests are marginally acceptable measures of progress for learners studying the Workplace Reading and Workplace Math strands, though the best match is not always with the test a state has selected as its primary indicator. This working paper provides a

---

[1] GEDIllinois and GED Online—Missouri.
[2] Students that pre-test at the 11th or 12th grade level in reading and math are exempt from progress testing.

detailed analysis of the fit between WES Reading and the four most commonly used standardized measures. This type of analysis can provide useful general information regarding strengths and weaknesses of various assessments in relation to a specific curriculum.

For learners studying the Employment or Workplace Communication strands there are no standardized, easily administered measures of educational progress. New measures are needed. Research on efforts to teach WES at a distance in Pennsylvania showed that most learners spend less than 30 hours studying over the course of a year. This raises the possibility that learners studying WES at a distance could be classified as Work-based Project Learners in the NRS. Doing so does not eliminate the assessment requirement, but changes it from pre-post testing with a standardized measure to a single end-of-project assessment that validates that learners possess the skills taught.

3. **Educational Gain and Hours of Study.** Various studies of educational gain done with classroom-based learners estimate that a gain of one educational level requires 100-150 hours of study. The Pennsylvania data suggest that, in the course of one year, many distance learners will study less than 50 hours—well less than what is needed to make an educational level gain. If further studies find this to be typical, it may be that distance learners should be treated differently. One state is considering adjusting its negotiated outcomes to permit counting distance learners without the expectation that they will make educational gain.

4. **Certifying Educational Level Gain.** Assessing educational progress—like any high stakes assessment activity—requires that testing be done in a secure environment where someone other than the learner certifies the identity of the test taker and certifies that the appropriate test was used and that time requirements (if any) were met. Typically, testing is done at a literacy center that has procedures in place to ensure these requirements are met. But, experience to date indicates that it is difficult to get distance learners to come to a testing location far from their home, especially for a posttest. Moving the assessment to the Web so that learners can take the test at a distance is unacceptable because the security requirements cannot be met. Two solutions are possible. One is to develop a network of test locations (K-12 schools, libraries) where an educator or librarian provides the necessary security without requiring the learner to go to the host literacy center.[3] Another solution is to maintain a cumulative portfolio of each student's work and use a standardized rubric to certify progress. Both options need to be further explored.

Next Steps

These issues and the related policy recommendations need additional research to help make wise choices. In the coming year a number of research efforts will be incorporated into the distance experiments of the states collaborating in Project IDEAL.

---

[3] In this case, an online test would be helpful. It would solve a thorny problem with decentralized testing: having the student first take a locator test which in turn dictates which form of the test needs to be taken.

1. **Measuring Seat Time.** Research will be done on ways to measure seat time and estimate how much time learners spend studying the same curriculum in classroom and distance settings. In those states using programs such as *SkillsTutor* or state-developed online GED programs where the instruction is contained in the online activities, the computer records will be used. In all states using other curricula, teachers will examine workbook and online assignments and develop standards for determining level of completeness. Comparative research in classrooms will help establish how many hours credit should be granted to learners who submit assignments judged to be complete.

2. **Increasing Seat Time.** Since distance education with adult basic learners is so new, states will try systematic experiments in ways to maximize the amount of time distance learners spend engaged in learning. The goal is to identify best practices in recruitment, orientation, and support.

3. **Appropriate Measures of Attainment.** *WES:* Experiments will be done in at least one state (Pennsylvania) to see if their Workplace Foundation Skills Framework can be adapted to provide a way to measure educational gain for those learners studying the Employment or Communication strands of WES. Other states will be invited to try similar approaches.

4. **Certifying Educational Level Gain.** The feasibility of having libraries and K-12 schools administer tests for adult learners working at a distance will be explored in several states. In other states, the feasibility of collecting and evaluating student portfolio data for educational gain will be explored.

5. **Distance Education Assessment Conference.** When there is sufficient data from these research efforts, a group of testing and policy experts will be convened to consider the issues and the data and make recommendations in each of these areas.

———

Project IDEAL Research and Technical Assistance
Dr. Jerome Johnston, Institute for Social Research, University of Michigan
P. O. Box 1248, Ann Arbor, MI 48106-1248
734/763-3079 – projectideal@umich.edu – www.rcgd.isr.umich.edu/ideal

Project IDEAL Member States (2002)
Arkansas, Idaho, Illinois, Kentucky, Maine, Massachusetts, Missouri, New York,
North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina

Introduction

In the year 2000, a group of state directors of adult education began to explore possibilities for implementing distance learning in their states. After two years of discussion and planning, directors from 13 states joined with researchers from the University of Michigan to form a consortium called Project IDEAL (Improving Distance Education for Adult Learners) dedicated to developing viable distance education programs for ABE, ASE, and ESL learners. Project IDEAL was created to support states as they experiment with various distance learning models and work to develop best practices and sustainable programs. Staff from the University of Michigan are engaged in developing a variety of products and services to facilitate the experimentation and implementation processes. These products include a *Handbook of Distance Education for Adult Learners*, program experiment and planning documents, an online course titled: *Recruiting and Teaching Adult Learners at a Distance*, and this assessment working paper. These products are intended for distance educators, administrators, and policy makers.

This working paper focuses on the role of assessment in distance education. In particular, it examines the requirements of the National Reporting System (NRS) Guidelines implemented nationwide in June, 2000. Developed in response to a mandate set forth in 1998 in the Workforce Investment Act's Adult Education and Family Literacy Act, the NRS defines required practices in assessing and reporting on adult learners as a means of determining program success and funding. As a result of the outcomes-based orientation of the NRS, states must carefully examine a variety of related issues as they select curricula and assessments for use in their distance programs. These issues include how to determine number of instructional hours or seat time equivalents for certain curricula, assessing whether to use existing standardized measures or to develop new assessments or performance-based measures, and deciding how best to categorize learners for reporting purposes (e.g., core outcome of educational gain vs. secondary outcome work-based project learner designation) among others. Our guiding research question for this paper is: Given the current NRS environment, what are assessment issues states face in developing, implementing, and sustaining viable distance education programs? In addition, we provide recommendations and specify research studies that need to be conducted to help guide future policy decisions.

This paper is divided into five sections. We begin with a discussion of assessment in general, considering the many reasons assessment information is needed and used by teachers and administrators. Section 2 provides an overview of the NRS requirements and describes the issues states must face in determining how to count distance learners and establish appropriate standards for educational gain for them. Section 3 examines the curricula that the Project IDEAL states have selected to teach at a distance and considers the measures appropriate to each. Section 4 looks in depth at one curriculum—the Reading Strand of Workplace Essential Skills—and highlights several problems of misalignment between the instruction and the measures of educational gain recommended by the NRS. The final section looks ahead and describes a number of research efforts that will be incorporated into the distance efforts of Project IDEAL states in the upcoming year (October, 2002 – September, 2003).

The Many Uses of Assessment

In general, there are three primary purposes for assessment: (1) for placement into courses and programs, (2) to measure students' progress, and, (3) to determine learning and other gains for accountability purposes. In each arena, assessment measures serve different, sometimes overlapping, functions including gaining individual student information for tailoring teaching and learning, determining curricular selections, assessing program success, planning for professional development activities, and reporting to local, state, and national stakeholders.

Intake/Placement

The first use of assessment occurs during the intake process. While agencies have different protocols, in general, most follow a similar pattern. Students are interviewed by agency staff who gather demographic data and help them identify the primary objective(s) or goal(s) they wish to accomplish. These goals may include earning a GED, finding or improving employment, enrolling in post-secondary education, and/or developing basic literacy skills. Much of the data to be collected for the NRS can be handled during the intake process. NRS data to be collected at intake include: demographics, student status (employment, public assistance, disability, rural residency and primary and secondary goals for attending), and program enrollment type. Optional secondary measures include whether the student is low income, a displaced homemaker, a single parent, a dislocated worker, and/or is learning distabled. (For a complete data collection list, see the NRS Implementation Guidelines, U.S. Department of Education, p. 38.)

During the intake process, students are also tested to determine their incoming literacy skills or "Educational Functioning Level" (NRS Guidelines, p. 6) for placement into appropriate courses. Intake testing using a standardized measure serves two additional purposes. First, it enables agency staff to help students set realistic goals and timelines for study. Second, it serves as the baseline that can be used for accountability purposes as the student progresses. NRS requires that each student who enrolls in a program and has received at least 12 hours of instruction be baseline tested using a standardized assessment to be counted toward a program's total population of students served.

During the intake process, agencies may also choose to use a variety of informal assessments depending upon the type of program the student is interested in and on his/her stated goals. For example, a student interested in gaining employment might be evaluated for general knowledge of the application and interviewing processes. These types of assessments may also be completed at the beginning of a course of study, after students have started classes.

In distance courses, the intake process is fairly similar to that used for placement into on-site courses. Often, distance educators will have only one opportunity to meet face-to-face with their learners to establish a connection, introduce the distance process, and set expectations for teaching and learning. Orientation, however, plays a far more immediate and important role in facilitating students' understanding of the distance education process. Thus, in addition to the baseline testing and interview, the intake process must

also include orientation in how to be a distance learner and how to use technology (where needed) and how to use the specific curricular materials. Of equal importance is helping students determine whether they are well suited for distance education. Some state programs provide an online self-assessment to help learners decide whether distance is a viable option for them. *GEDIllinois* for example, provides such an assessment called OASIS. Students answer a series of questions regarding their life habits and learning preferences. The software analyzes their responses and then offers recommendations regarding whether a student should pursue distance education and if so, what he/she needs to do in preparation for participating. Other adult education providers offer similar self-assessments. Kentucky's Virtual Adult Education site links to its Virtual High School site, which provides an online self-assessment called "Is Online Learning for Me?" Missouri's *Online GED* program website directs students to a "Right Choice Assistant" self-check quiz to help students decide whether online learning is right for them. As with *GEDIllinois*' OASIS assessment, the Right Choice quiz evaluates student responses and provides recommendations. Not every adult learner is a good candidate for studying successfully at a distance. Tools such as the ones described above can help educators and learners determine whether distance education fits for them.

Once students are identified as viable distance learning candidates and have completed the intake and course orientation processes, they must also be properly oriented to the importance of participating in progress and post-testing (where appropriate). Helping students understand the role of assessment in their learning can be an important step toward ensuring their participation in what is often seen by students as a daunting part of their educational experience.

Progress

The second assessment arena involves measuring progress. Progress assessments generally occur both within a course of study on an ongoing basis as students advance through a program of study, and at state-determined intervals to determine gain (e.g., after 40-50 hours of instruction, every four months, upon completing certain coursework, etc.). Progress assessments may or may not include standardized measures. Informal measures include checklists, performance-based simulations, tailored tests, interviews, practice tests, teacher observations, and graded course materials, among others. These various measures serve as the feedback mechanism during program participation. They enable students to see where they are making gains—a potential motivational tool—and they provide teachers and programs with valuable information on what is/isn't working from curricular and delivery standpoints. Measures of progress in distance education play a vital role in determining students' understanding of and engagement with course materials. Frequently, distance teachers have little to go on aside from such measures of progress as completed course assignments, teacher-student email and/or phone exchanges, end of unit self-checks, exams, and practice tests to determine whether a student appears to understand course content.

Accountability

Accountability has become the driving force behind most of the laws, funding decisions, and goals established for progress in adult education. Various stakeholders want

evidence that adult learners are making the kinds of gains that will enable them to become more productive contributors to society (Merrifield, 1998). To determine whether a particular program is successful in achieving such outcomes, evaluators need to look across students and programs using the same types of data as points of comparison. As a result, states have had to focus more attention on specific types of student and program data collection as well as on establishing guidelines for selecting and administering standardized assessments. Standardized measures can include tests or performance-based tasks that have a scoring rubric that has been standardized and that is applied uniformly in assessing student performance (NRS Guidelines, p. 13).

Some states are exploring the portfolio process as a means of collecting different types of data on student progress. If the data collection and assessment processes are standardized, student portfolios can also be used to demonstrate educational gain when students are either unavailable or unwilling to be post-tested. Each state is responsible for establishing its own guidelines for what counts as a measure of progress. For purposes of the NRS, however, only standardized assessment results are counted toward educational gain.

The NRS Perspective

Beginning in 2000, states were required to implement the NRS guidelines, which indicate how states must report on the educational achievements of their adult learners. The purpose of the NRS is to enable systematic collection of data that can be aggregated and analyzed at the national level to aid congress and other federal agencies. Federal incentive grants are tied directly to a state's ability to exceed expected learning outcomes for its adult students.

The NRS requires states to collect measurement data in three primary areas: descriptive, participation, and outcome. As mentioned above, descriptive and participation data (with the exception of instructional hours) are generally collected during the intake process.

Determining Instructional Hours

Programs are required to collect data regarding the number of hours each student spends participating in instructional activities. For purposes of the NRS, time data are collected for all students. The NRS recommends that programs include "time in class receiving instruction; time in a learning lab, such as a computer lab; time spent with a program-sponsored tutor; and time on assessment activities, including initial assessment and follow-up tests, such as post-tests" (p. 45). This means that programs may choose to include intake and orientation activities as part of the 12-hour minimum. Note that time is assigned largely in blocks of fixed length; a learner gets credit or not for attending a class of a certain length. In a few cases time is assigned for activities that may contribute unique values for every student, such as working with a tutor.

In distance education, determining instructional hours or what counts as "seat time" is challenging for any activities that do not take place at a literacy center where it can be certified by a teacher or administrator. Intake and orientation activities typically take place at a center, though some distance teachers in Pennsylvania manage to do some of

these activities without seeing their students. In general, once a student is working outside of a center, the task of determining hours students spend studying can be problematic. There are two issues: measuring study time and trusting its validity. CAI programs *(PLATO* and *SkillsTutor)* track instructional hours through the software, which provide the necessary estimates of students' participation.

With other distance education packages such as *Crossroads Café, GED Connection*, and WES, the issue of measuring seat time is more complicated. Unlike software-based skills programs, the online components of WES and GEDC do not have a time tracking mechanism. The nature of the LiteracyLink's Web design does not permit tracking elapsed time while a learner is working on assignments within each lesson. In addition, one part of each lesson requires the learner to leave the LiteracyLink website and search for resources elsewhere on the Web. Two methods could be used to track seat time. One is to have students keep a time log; the other is to have teachers examine students' completed work and give credit for a predetermined number of hours for each assignment that is judged to be complete. Several Project IDEAL states will try out the second strategy in this coming year. In these sites parallel studies will be done in classrooms where students are studying WES and GEDC to secure comparative data and establish a range of times associated with completing various components of the curriculum.

Crossroads Café does not have a Web component, but it presents the same challenge. To study *Crossroads* at a distance students receive a variety of video and print materials. Students may choose to work solely with the video and primary student workbook. Or, they may choose to supplement their learning by using the Partner's guide and/or the Photo Stories in addition to the primary materials. Additionally, many ESL students view videos repeatedly as part of their learning process. In other words, the time spent working with these materials will vary greatly from learner to learner. Distance programs using *Crossroads Café* may ask students to keep a time log of each type of activity. But can learners keep valid logs? In a formative study of *Crossroads Café* researchers found that they had to call learners weekly to obtain what they considered to be valid data on study time. The solution posed for WES and GEDC may be the best one to use with *Crossroads Café* as well. But research should be done using both self-report and teacher rating strategies to compare the validity of each method.

Outcome Measures

Outcome data to be collected include standardized assessment results, total number of instructional hours, and follow-up survey information, through which agencies track employment and educational status. Exhibit 1 of the Guidelines (p. 4) lists the primary and follow-up outcomes as well as secondary outcomes. Reporting on secondary outcomes is optional, though the Guidelines suggest that states may want to track such information for their own analyses.

The primary areas or "core outcomes" in which states must collect data are divided into two categories: (1) measurement of educational gain and, (2) follow-up measures collected at predetermined intervals after students exit programs. These include (a) employment (attainment and retention), (b) receipt of a high school diploma or its equivalent, and (c) enrollment in some form of post-secondary training (e.g., vocational

or technical education, community college, etc.). Data on follow-up outcomes are matched to students' self-identified goals and are collected either via surveys or through data matching with other agencies. For example, students' employment status might be determined either through contacting the student directly or through data matching with the state's workforce department via social security numbers.

Perhaps the most complicated of the core outcomes is educational gain. With the exception of Work-based Project Learner designees, all 12+ hour students are measured for educational gain. Work-based Project Learners are those students who participate in short-term (12-30 hours) work-related programs that include specific achievement objectives. These students are exempt from educational gain measurement requirements. A more in-depth discussion of the Work-based Project Learner classification appears in the WES Case study later in this paper. All other students are pre-tested at intake to determine their educational functioning level. The NRS has identified six educational functioning levels for native English speakers and six for English-as-a-second-language (ESL) students. For native speakers, there are four descriptors for adult basic education (ABE) and two for adult secondary education (ASE). Table 1 lists the different levels.

Table 1: NRS-Defined Educational Functioning Levels

ABE/ASE Levels		ESL Levels	
Literacy Level	**Grade Level Equivalent**	**Literacy Level**	**Student Performance Level**
Beginning ABE	0–1.9	Beginning ESL Literacy	0–1
Beginning Basic	2.0–3.9	Beginning ESL	2–3
Low Intermediate Basic	4.0–5.9	Low Intermediate ESL	4
High Intermediate Basic	6.0–8.9	High Intermediate ESL	5
Low Adult Secondary	9.0–10.9	Low Advanced ESL	6
High Adult Secondary	11.0–12.9	High Advanced ESL	7

Each descriptor is tied to a set of skills students must be able to demonstrate upon entry into that level. To measure students' educational functioning level, states are required to use either a standardized test or a performance-based measure that has been standardized in its administration and scoring. The NRS recommends four standardized tests that have been correlated to the educational functioning level skill descriptors for ABE and ASE. These include the Adult Basic Learning Examination (ABLE), the Adult Measure of Essential Skills (AMES), the Comprehensive Adult Student Assessment System (CASAS), and the Test of Adult Basic Education (TABE). For ESL students the NRS recommends the Oral Basic English Skills Test (BEST) and CASAS Life Skills tests to determine the Student Performance Levels (SPL). States may select from among these tests, choose other standardized tests, or develop their own measures as long as they can demonstrate that the assessment procedures and scoring have been standardized. As Mislevy & Knowles (2002) report, the process of developing a standardized test is both costly and time consuming, which can prove problematic in a field where funding and

human resources are often limited. Currently, few states are in a position to engage in the assessment development process. Two notable exceptions include Ohio, which has developed and is field testing a performance-based portfolio system for its entire adult learner population, and Massachusetts, which is working to develop standardized assessments tailored to its educational standards. Other states such as Kansas and Pennsylvania are working to develop performance standards, checklists, and other frameworks for use in teaching and learning and in all facets of the assessment process.

While accountability is the primary concern of the NRS, meeting such requirements in a manner that both accurately gauges student learning and provides all stakeholders with useful information poses problems. Many states have mandated the use of specific standardized assessments for all measures of educational gain, yet insufficient analyses have been completed to determine the match or fit between these tests and the various curricula being taught in local programs. As Reuys (2001) notes, "standardized tests may do a very poor job of capturing and reflecting the learning that goes on in adult basic education classes" (p. 2). Thus, programs may end up underreporting student learning because the testing used does not measure what is taught in a given curriculum. Reuys further points out an important caveat to accepting testing as a comprehensive means of determining learning: "It may thus prove to be a major risk for adult basic education programs across the country and for the system as a whole to be judged largely on the basis of students' improvements in scores on tests that may be inherently incapable of capturing much of the learning that is taking place for these students at these programs" (p.2).

Ideally, states would have sufficient resources to develop tailored standardized assessments to ensure measures accurately reflect instructional content. At this point in time, however, many states have limited time and resources and so must rely heavily on standardized assessments for reporting purposes. In lieu of developing tailored measures, states should carefully examine which of the available standardized tests is the best match with their curricula. Given that the standardized assessments recommended by the NRS were not developed specifically for the curricular materials being used by Project IDEAL states, it is important to know both how closely the items on any standardized assessment match instructional materials as well as how much of the curriculum is captured by the assessment. Understanding the strengths, limitations, and purposes of each assessment as well as how and why specific assessments have been constructed can help educators make more informed assessment selection decisions. For example, CASAS is a competency-based measure designed to assess literacy skills development using workplace-based contexts. Items on the reading assessments reflect a variety of work-related documents such as memos, diagrams, pay stubs, safety instructions, etc. The focus on the AMES, on the other hand, is broader in scope, containing examples of poetry, historical narrative, personal communications (among others) in addition to everyday literacy and work-related texts. The differences in genre and how skills are contextualized affect whether the measure accurately assesses what is learned.

Educational Gain and Distance Education

Currently, no provisions or recommendations have been made in the NRS for considering the population of distance students in terms of (1) counting learners and determining instructional hours, and (2) educational gain. Current research suggests students typically need to participate in approximately 150 hours of instruction to have a strong chance for making an educational gain of one level (see Comings, Sum, and Uvin, 2000 cited in Mislevy & Knowles, 2002). And assessment systems like the CASAS set their recommendations for expected learning gains on the basis of 100 hour instructional increments. Yet the majority of adult education students working in classroom-based programs receive under 100 hours of instruction per year (U.S. Dept of Education, 2001). Indeed, in Pennsylvania, the median time spent in classroom-based programs over the course of a year is only 45 hours. What does this mean for the distance student? Little is known about how many hours distance learners will spend on average engaging instructional materials, though early indications suggest participation may be lower than in classrooms.

In thinking about instructional hours, issues of expected gain come into play. So far, no large-scale studies have been conducted to determine whether distance learners make gains comparable to their classroom-based counterparts when they spend similar numbers of hours in instruction and study. Nor has any data been systematically collected to determine the level of student engagement with the materials in an independent learning environment. In a distance setting, learners have fewer opportunities for immediate feedback and support than they would have in a classroom and must rely primarily on themselves for sustained motivation and engagement. Given the higher levels of self-reliance inherent in distance education, the average number of instructional hours needed for educational gain may or may not be higher than those for classroom-based programs. Project IDEAL considers it important to conduct additional research on instructional hour requirements in terms of distance education before solid recommendations can be made for program development.

It is important to account for the potential differences of distance education when setting expectations for educational gain and for what is feasible for learners to achieve in a given timeframe using a particular curriculum. For example, in instances where teacher-mediated products are used (e.g., WES, GEDC, *TV411*, *On Common Ground*, etc.) states should collect comparative data to determine what differences exist between classroom and distance learners with regard to educational gain, time spent engaging learning activities, and persistence levels to determine appropriate expectations for reporting purposes. For programs like *PLATO*, *SkillsTutor*, *GEDIllinois*, and Missouri's online GED program where time tracking is possible, such data should be readily available for analysis and comparison to determine what, if any, differences exist between classroom-based and distance students' time on task requirements for similar educational gain.

Meeting Standardization Requirements in Assessing Distance Learners for Accountability

As part of distance learning programs, accommodations must be made to test learners in an environment that permits the standardization process required by the NRS. Programs

need to establish a protocol for assessments. This may involve students returning to the orientation site for any follow-up testing. Or, states may arrange with other agencies (e.g., libraries, community based organizations, workplace settings, etc.) to train staff to handle assessments. In this instance, students could travel to the nearest site, which may facilitate their willingness to participate in post or progress testing.

Given the high rate of student drop out of classroom programs prior to post-testing, distance programs may also need to consider alternative means of assessing gain over the course of instruction. States like Ohio are developing and standardizing portfolio systems for their adult education population that can serve to show gain in lieu of standardized post-test measures. Alternative approaches like the portfolio system are gaining in popularity as programs are anxious to report on learning gains and to be able to count students as having achieved their goals. The NRS accepts such alternative reporting methods so long as standardization of data collection and rubric development and scoring can be demonstrated. Alternative data collection is a topic that requires considerable discussion and further research and will be addressed in future Project IDEAL research and assessment-related work.

Distance Curricula and Assessment

Among the first steps in setting up distance experiments, Project IDEAL states identified curricular packages to adapt for use in their programs. The different content and delivery mechanisms as well as the level of teacher mediation in each package impacts the types of assessment measures that can be used in each phase of the assessment process: intake, progress, and accountability. For this paper, discussion will focus on four products currently being used by states in their distance programs: *PLATO, SkillsTutor, GED Connection*, and *Workplace Essential Skills. Crossroads Café* has also been selected for use in distance experiments in five Project IDEAL states. However, current analyses and discussion for this paper is focused on products designed for ABE, ASE, and GED students. ESL issues will be examined in-depth in later editions.

PLATO and *SkillsTutor* are software-based instructional packages that include built-in diagnostics and tailored unit and course completion testing. These packages do not require extensive teacher mediation as the software is designed to guide student learning based on diagnostic and progress assessment results. After learning how to use course software, students can work independently. The systems do not preclude teacher involvement, however. Rather, they are designed to enable students to function at their own pace and to provide learners with immediate feedback and direction. The programs' online management systems track assessment results, time on task, and whether students have achieved predetermined mastery levels for course content. Currently, the assessments provided by *PLATO* and *SkillsTutor* can only be used for diagnostic and progress purposes. Accountability requirements must be met through use of standardized assessments. Both *PLATO* and *SkillsTutor* have correlated their instructional content to a wide variety of state and national k-12 and adult standardized assessments as well as to various standards and benchmarks. Product developers should be contacted regarding recommendations for which NRS-designated or other standardized assessments are thought to best capture instructional content provided in the software.

In addition, existing assessment data from classroom-based programs in the three Project IDEAL states using these CAI programs can be used to compare educational gain and time on task similarities between classroom and distance learners as more data become available for distance learners using these products. Other possibilities for data collection and analyses include examining the comparability in educational gains made when different assessments are used to measure progress. These and other experiments will provide useful information (1) in determining the degree to which various units, courses, and programs offered by the software packages result in educational gain and, (2) in identifying which standardized assessments are best suited for particular curricula.

GED Connection (GEDC) is a multimedia product that was developed based on requirements and content of the newly released GED 2002 test.   GEDC provides self-check measures, unit testing, and practice tests in each of the five GED test content areas. These practice tests have not been standardized, but are modeled closely on GED test content and measurement style. Recent debate at the state and national levels has focused on whether such non-standardized measures as the official GED Practice tests and tailored practice tests provided by software and state-based program developers should be counted toward NRS reporting requirements. As yet, consensus has not been reached regarding whether the reauthorization of the NRS will include recommendations for GED progress reporting.

For NRS accountability, GED students must be pre and post-tested using a standardized assessment. Post-testing at various state-determined intervals is required regardless of which GED content area a student is studying at the time of testing. Thus, at some points, post-testing matches instructional content, such as when a student is preparing for the GED math test. At other times, however, a student might be studying content that is unrelated to standardized test content. For example, a student enrolled in a science preparation course would still be required to be post-tested on reading, writing, or math to determine educational gain. This mismatch between instruction and assessments can result in misleading findings regarding student learning and progress toward the goal of earning a GED. The one reporting exception involves those students who pre-test at the High Adult Secondary level (11th-12th grade). Educational level gain credit is given for those students when they earn their GED.

The Case of Workplace Essential Skills

Developed by PBS LiteracyLink, *Workplace Essential Skills* (WES) is a 24-unit multimedia (print, video and online) course designed for adults who want to learn how to apply for a job, increase their knowledge of the workplace, and refine their reading, writing, communication, and/or mathematical skills to meet the demands of common workplace settings. The program is designed for students reading between the 5th and 9th grade levels. The curriculum is divided into four instructional strands: Employment, Communication, Reading, and Math.

The Employment strand introduces students to the job-search process as well as appropriate on-the-job behaviors, employer expectations, workplace safety issues, etc. The Communication strand provides instruction and modeling in appropriate verbal and

non-verbal communication with customers and coworkers, and demonstrates how businesses use different types of documents to convey information inside and outside the organization.   The Math strand covers basic use of mathematical concepts in the workplace, including the topics of estimation, strategies for setting up problems, applications of basic arithmetic, and use of a calculator. Instruction ranges from beginning ABE concepts to advanced adult secondary concepts. The Reading strand helps students understand a variety of purposes for reading, suggests different reading strategies, and presents different types of documents and their purposes. Each strand contains both conceptual information regarding why certain skills are needed on the job and opportunities to practice skills.

WES provides informal student self-checks (skills previews and reviews) in each content area workbook and unit-level testing. No other assessment measures are provided. WES was under development during the time when the concept of a national reporting system was being debated, but it was too early for the developers to give serious consideration to creating a separate standardized group-administered assessment. Besides, the goal of WES was to prepare learners for employment, not to raise scores on standardized tests. Success for learners studying WES could be measured by whether they obtained employment and could handle workplace reading, math and communication tasks on the job. With the introduction of the NRS, however, the definition of WES learner success needed to be expanded to include demonstration of educational gain. The NRS requires educational gain be measured in one of three areas: "Basic Reading and Writing," "Numeracy Skills," and "Functional and Workplace Skills." While standardized assessments have been developed that are designed to measure change in students reading and math skills—two of the four content areas taught in WES—as yet much work needs to be done to develop standardized measures that capture gain in the Functional and Workplace Skills category. Currently, the field is not adequately prepared to account for the types of workplace skills that are taught in the Employment and Communication strands in ways that meet NRS requirements. Determining appropriate means of assessing gain in workplace programs is considered a state-level issue. To address the need for measuring educational gain of workplace skills, many states use or are developing performance-based and other alternative assessments with standardized administrative procedures and scoring rubrics that will meet federal measurement and reporting requirements.

To address the immediate concerns of IDEAL states using WES at a distance and wanting to explore educational gain possibilities, we turned attention to analyzing the WES Reading and Math curricula and comparing those strands with NRS-recommended standardized assessments. For purposes of this case study, our discussion focuses on work completed around the Reading strand content. The content analysis and comparisons of WES Reading curriculum with the standardized assessments are meant to serve as an illustration of the types of issues associated with measuring a specific curriculum using a nationally standardized test not developed specifically for that curriculum. These analyses are not intended to be exhaustive nor do they account for the host of psychometric considerations that are essential for good test design and evaluation. Instead, the analyses are intended to highlight ramifications of using certain types of assessments as the primary indicators of change. They are also designed to help states

make better-informed decisions when selecting from among available measures. In the high stakes, outcomes-based arena of the NRS, it behooves states to understand the curriculum-assessment matching process to ensure they understand the strengths and limitations of using standardized tests as the primary or sole means of demonstrating educational gain.

The WES Workplace Reading Strand

The first step in examining fit between curricula and standardized assessments involves completing a detailed content analysis of the curriculum. As mentioned above, the WES Reading strand is intended to introduce learners to the various ways in which reading is used on the job, to useful reading strategies, and to the different types of documents learners might encounter in their work environments. Unit by unit analyses were completed to identify each topic presented in the curriculum. Topic areas were then grouped into three categories: purposes for reading, reading strategies, and document types. Ten key concepts regarding purposes for reading were identified. These concepts are designed to help learners understand the connection between improving their reading skills and accessing information that enables them to function more successfully on the job. The Reading strand also provides 16 strategies for reading that help learners understand the various elements and structure of texts as well as approaches to successfully locating and remembering information. Finally, a variety of texts are used as instructional examples. Table 2 lists the specific contents of each category.

Table 2. Content of the WES Workplace Reading Strand

Purposes for Reading	Reading Strategies	Document Types
10 Concepts	16 Concepts	17 Concepts
Understanding the purpose of a text	Re-reading	Flowcharts
Gaining familiarity with different types of & organizations of texts	Using reference books & other resources (e.g. dictionary)	Tables & charts
	Asking for clarification	Employee handbooks
	Varying reading speed	Policies
Inferring from texts	Scanning & skimming	Forms (e.g., purchase order, application, change of address, etc.)
Problem solving using texts	Taking notes & highlighting	
Understanding instructions	Restating what you've read	Pay stubs
Using (reading/completing) forms & charts	Using context clues	Training manuals
Reading for specific information	Monitoring understanding	Product specifications
	Picturing & visualizing texts	Material data safety sheets
Understanding specific words & phrases	Thinking of an example of a key point	Tables of contents
		Diagrams
Finding information	Paying attention to details	Pictorial instructions
Identifying main ideas	Figuring out abbreviations	Schedules (bus, work)
	Using multiple approaches to finding information	Memos
	Using structure of diagrams & written directions (e.g. titles, headings & legends)	Step-by-step instructions
		Labels
		Guidelines
	Reading thoroughly	

The NRS-Recommended Tests

As noted above, the NRS recommends four standardized assessments for measuring gains made by ABE and ASE learners: the AMES, ABLE, CASAS, and TABE. The four measures share some common features. All are group administered, multiple-choice batteries that include two parallel versions for pre- and post-testing in each content area. The ABLE, AMES, and TABE have been norm-referenced, which involves interpreting students' scores in relation to how well a comparison group "normally" performs on a test (Sticht, 1999). The CASAS is a criterion-referenced test. Criterion-referenced tests are those that "set a required level of performance and interpret scores in relation to that standard or criterion" (Forlizzi, 1998, p. 22). All four tests include a locator instrument for use as a screening or placement measure. Each includes reading and math subtests for different literacy levels ranging from pre- or beginning literacy to advanced secondary education and all claim to have items that were developed from "real life" or everyday contexts that adults routinely encounter. Table 3 lists general features of the tests followed by a brief description of the content of Reading and Math tests from each battery.

Table 3: Features of Standardized Assessments

	ABLE	AMES	CASAS ECS	TABE
Screening instrument	*Select*ABLE	Locator	Appraisal Form 130	Locator
Sub-tests used for NRS*	Reading Comp Number Operations Problem Solving	Reading Communication Computation Applied Prob Solving	Reading Math	Reading Math Computation Applied Math Language
Ed'l Range	Grades 1-12	Grades 0-12+	Grades 1-13+	Grades 0-12.9
Test Levels	1: Gr 1-4 2: Gr 5-8 3: Gr 9-12	A: 0-2 (LEP/ESL) B: 3-4 (Low literacy) C: 5-6 (medium lit) D: 7-8 (high literacy) E: 9-12+ (advanced lit)	A: Beg literacy/pre beg B: Beg basic/intermed C: Advanced basic D: Adult secondary	L: (literacy) 0-1.9 E: (easy) 1.6-3.9 M: (medium) 3.6-6.9 D: (difficult) 6.6-8.9 A: (advanced) 8.6-12.9
Testing Time in minutes**	Reading (35) Number ops (35) Prob solving (35)	Reading (35) Comm. (35) Computation (15) Applied prob sol (35)	No time limit	Reading (15) Math Comp (50) Applied Math (50) Language (55)

*ABLE also offers subtests in spelling, vocabulary, and language; TABE includes a spelling subtest that is scored separately from the rest of the battery.
**Note. Time limits for the ABLE and AMES are suggested, but not required.

Adult Basic Learning Examination (ABLE). The ABLE is a basic skills test battery that has three levels of exams and six sub-test content areas. The Reading Comprehension tests for Levels 2 and 3 each contain 48 items, which are divided equally between inferential and literal questions derived from passages of academic content and from materials that might be found in daily life (e.g., forms, letters, warranties, reports, directions, etc.). ABLE has two mathematics sub tests. The Number Operations tests are

comprised of computations using basic operands in calculations with fractions, whole and mixed numbers, decimals, percents, powers, dollar amounts, basic algebraic equations and negative numbers. The Problem Solving test includes word problems that measure a student's ability to read charts, graphs, tables, etc., to extract information for use in setting up and solving equations.

Adult Measure of Essential Skills (AMES). The AMES consists of five testing levels developed from the definitions of adult literacy provided by the National Adult Literacy Survey. Three content-area subtests are used for NRS accountability: Reading, Communication, and Math. Reading test items are designed to reflect workplace or everyday literacy activities and information. Reading passage questions measure students' vocabulary knowledge, literal and inferential skills, and ability to read for meaning. The tests present questions regarding passages written in a variety of genres including fiction, poetry, prose, and expository writing. The Communications tests measure writing processes including creating, revising, and editing different types of texts (primarily business-related), and knowledge of grammar and mechanics. The Math tests were developed using standards provided by the National Council of Teachers of Mathematics and are divided into two sections. Computation items include questions requiring the use of basic operands in problems with whole and mixed numbers, decimals, fractions, percents, and algebraic equations. Applied problem solving items focus on mathematical reasoning, estimation, applying geometrical concepts, interpreting statistical information, mathematical reasoning, and computing functions among others.

Comprehensive Adult Student Assessment System (CASAS). The CASAS is a competency-based system containing 80+ standardized and alternative assessment instruments that have been developed from a list of over 300 core competencies for adult basic education. Test items are designed to demonstrate student mastery of basic skills at different ability levels. Unlike the other NRS recommended assessments for ABE/ASE, the CASAS does not assign grade levels to test results, instead converting its scaled scores to skill level descriptors. Rough grade-level equivalents are provided in the test administration manual for those programs required to report grade levels, but CASAS warns against using such equivalents for anything else. The Employability Competency System (ECS) tests are designed to measure students' reading and math abilities "in relation to the skills necessary to get and to keep a job" (Test Administration Manual, p. 3). The ECS Reading tests assess functional skills using work-related texts, signs, charts, reading passages, and other related materials to measure students' literal and interpretive skills. The Math tests require knowledge and application of basic operands, whole and mixed numbers, fractions, percents, decimals, basic geometry and algebra, and word problems set in work-related contexts.

Test of Adult Basic Education (TABE). The TABE provides basic skills tests for five literacy levels. Each TABE test overlaps other tests at the low and high end of the grade-level range for that test (e.g., TABE 7M is designed for up to grade 6.9, while the TABE 7D starts at the 6.6 grade level). The Reading tests measure vocabulary knowledge, critical thinking skills, and finding and using information from reference passages. Unlike the other measures, the TABE has a more academic orientation in its Reading tests. For example, included in the tests are items asking about uses for reference

materials as well as about citation styles. As with the AMES, the TABE Math tests are divided into two sections. The Computation subtests focus on basic operands used in conjunction with whole and mixed numbers, fractions, and decimals, among others. The Applied Math tests focus on solving word problems presented in a variety of everyday contexts using graphs, tables, and other pictorial representations.

Matching WES Workplace Reading with the Tests

We turn in this section to examining the fit of each of the four NRS recommended standardized assessments with the WES Workplace Reading Curriculum. Before delving into detailed analyses, it is useful to step back and look at the task in general terms. Two questions are being asked. Question 1: Given a particular test, does WES teach all of the concepts being assessed? Question 2: Given the content of WES, how much is measured by the test? Figure 1 illustrates these questions graphically as two circles. For question one the circle represents all the skills assessed on a particular test. The issue is how many of these skills are taught in the WES curriculum? A line divides the circle into two parts corresponding to those skills that are taught and not taught. The research question asks: where should the line be drawn that divides these two parts? For question two the circle represents all the skills taught in the WES Workplace Reading strand. The issue is how many of these skills are assessed by the test? Again, a line divides the circle into two parts, this time corresponding to those WES reading skills that are assessed by the test and those that are not. The research question asks where to draw the line that divides the circle into two parts.
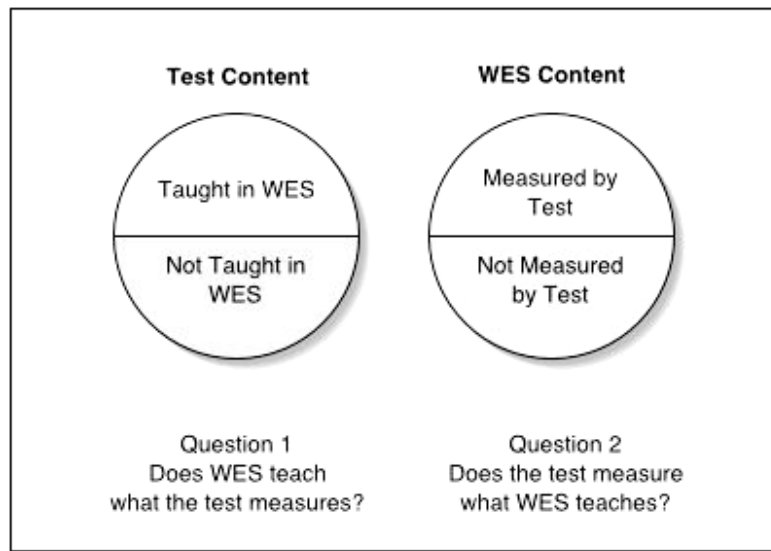

Figure 1. The Curriculum-Test Match Questions

A perfect test would be one for which WES teaches everything that is measured by the test, and, conversely, the test measures everything that is taught in WES. In reality, no standardized test not developed specifically for WES will ever be a perfect match. The more relevant issue involves whether the degree of fit between a curriculum and test contents is sufficient to recommend use of the test as a good indicator of learning gains.

To answer this question, we analyzed the content of each of the appropriate forms of the above four tests and then assessed the fit using a process described below.

Selecting Forms of Each Test Battery

Not all forms of the four recommended tests are appropriate, so some forms were excluded from the analysis. Forms intended for students functioning below the 5th grade level were not analyzed as WES requires a minimum of 5th grade reading level. In recent studies, distance teachers using WES strongly recommended learners possess a minimum of 7th grade reading level to work on WES materials (Petty & Johnston, 2002). Thus, ABLE Level 1, AMES Levels A & B, CASAS ECS Level A, and TABE Levels L and E were not considered for this analysis. Content analyses were completed for the following subtests and item-level comparisons were completed with the WES Reading strand:

> ABLE Reading Levels 2 & 3;
>
> AMES Reading & Communication, Levels C-E;[4]
>
> CASAS ECS Reading, Levels B-D; and,
>
> TABE Reading, Levels 7D and A.

Coding the Test Items

Each item in each of these tests was scored using the scheme shown below.

1. Direct match (D)—between instructional content and test item.

2. Transfer (T)—requires transfer of skill or information learned in curriculum (e.g., WES Reading strand provides instruction in reading simple charts and forms. Students who have mastered basic chart reading should be able to successfully apply this knowledge to test items requiring reading basic charts or forms containing information on a variety of subjects).

3. Prior Knowledge (P) required to complete item (e.g., a student would be disadvantaged if they did not have some experience with a reference book such as the Reader's Guide).

4. Not covered (N) in WES curriculum.

For this analysis items with codes of 1-3 were considered to be a match, leading to a liberal estimate of the quality of the match. (If the item requires too much prior knowledge, only students with broad knowledge could answer correctly—even if they learned the WES material very well. Subsequent studies could eliminate items coded 2 or 3 and this would result in a more conservative estimate of match.)

---

[4] Calculations for the AMES Communication tests indicated that these tests were not a good match with WES Reading content and therefore, should not be used to measure educational gain.

Table 4 shows the match between the instructional areas presented in WES and the standardized assessments recommended by NRS. It is organized to address the two questions raised earlier: does WES teach what the tests measure and do the tests measure all that WES teaches? Column 1 addresses the first question. As an example, consider the TABE 7A. Analyses revealed that 36 of the 50 items in the TABE 7A test of reading tested content taught in WES. This is a 72% match. Looking across all of the tests, WES covers quite well the concepts tested in the ABLE, AMES, and the CASAS Levels C and D. It teaches only about three quarters of the concepts measured by the TABE tests. In other words, students who study WES Reading, but are assessed using the TABE, are less likely to show growth because they would not have been taught one-quarter of the concepts measured by the TABE.

Table 4. The Match Between WES Content and Various Standardized Tests

| | Question 1 | Question 2: Do the tests measure all that WES teaches? * | | | |
| | (1) | (2) | (3) | (4) | (5) |
	Does WES teach what the tests measure? †	All WES Content (n=43)	Purposes for Reading (n=10)	Reading Strategies (n=16)	Document Types (n=17)
ABLE 2E	100%	42%	50%	56%	24%
ABLE 3E	100%	40%	60%	50%	18%
AMES C2	100%	49%	**80%**	44%	35%
AMES D2	94%	42%	60%	50%	24%
AMES E2	100%	47%	**80%**	50%	24%
TABE 7D	76%	51%	**100%**	50%	24%
TABE 7A	72%	53%	**80%**	**69%**	24%
CASAS 14B	82%	47%	**80%**	38%	35%
CASAS 16C	94%	**58%**	**80%**	44%	**59%**
CASAS 18D	100%	**65%**	**90%**	56%	**59%**
AVERAGE	*92%*	*49%*	*75%*	*51%*	*33%*

† Cells show percent of test items that are covered by WES content
* Cells show the percent of the WES content covered by the test

Columns 2-5 address the second question. Look at the TABE 7A again. Column 2 shows that only 53% of the content taught in WES (23 out of 43 identified instructional topics) were measured by this test. Columns 3-5 show the match for the three content subcategories. The TABE 7A is the best match of all the tests for covering the purposes of reading (80%) and again the best for measuring the WES instruction on reading strategies, though the figure (69%) is perhaps unacceptably low. The TABE 7A is

weakest as a measure of the concepts taught in WES regarding document types that are found in the workplace.

Examining column 2, none of the tests appears to be a good measure of what the Reading strand of WES teaches. Not surprisingly, the best coverage is with the CASAS 16C and 18D, tests that us an employment-based context for the test questions. However, these tests cover only three-fifths of the skills taught in WES. Breaking down WES into three subsets of instruction reveals more variation in the measurement fit. In terms of the Purposes of Reading (column 3), seven of the 10 tests cover 80% or more of the WES instruction. Yet when it comes to Reading Strategies (column 4), nine of the tests are quite weak. The best coverage comes with the TABE 7A, and it measures only 69% of the strategies taught in WES Reading. In terms of Document Types (column 5) only the two higher-level CASAS ECS tests cover a reasonable amount of WES content.

Overall, none of the measures is an ideal choice, with only the TABE 7D and 7A and the CASAS 16C and 18D covering over half the WES content. If states choose to use one of the four NRS-recommended standardized measures, then they should be aware that much of what students learn in WES will not be tapped.

There is one additional caveat. The matching methodology used in this study is crude; it looks for a simple match between the topics in the test and in WES, ignoring the psychometric qualities of the test. For example, to be sensitive to change in a learner's knowledge or skills, each concept area must be represented by multiple items with varying degrees of difficulty. Applying principles of good test design would further reduce the goodness of fit.

The findings of this analysis are discouraging if the goal is to insure that learners are assessed in a way that samples what they learn. It is not to say that learners studying WES Workplace Reading will not show any progress if they are assessed using one of the four measures. In a field test of WES Workplace Reading, 43 adults studied the four chapters for six hours each. They were tested before an after using the CASAS ECS Reading test. Twenty-five percent showed an increase of one educational functioning level; another 30 percent showed movement within their level. The remaining 45 percent made no progress at all (Johnston, Young & Petty, 2001).

The analyses in Table 4 point to the need for alternative assessments for measuring what WES teaches. States are increasingly aware of the need to develop or modify existing work-related, performance-based assessments. The findings in this study support the conclusion reached by Sticht (1999): "The nationally standardized and normed tests are not sensitive enough to the specifics of what is being taught in the [workplace] program. Among other reasons, this is why many programs are searching for alternatives to such standardized tests. There is a desire for more curriculum-based assessment so that learners' 'true' gains can be detected" (p. 65).

Classifying WES Distance Learners

Many states that are trying out distance education start with the assumption that distance learners should be treated just like classroom students, with the same expectations for making measurable educational gains in the course of a reporting period. But a large experiment in Pennsylvania teaching WES at a distance found that only a small fraction of distance learners fit the profile of a typical classroom student. A total 1,248 adults were recruited to study *Workplace Essential Skills* (WES) at a distance over a 5-8 month period. Learners elected to work on one of four strands of WES. Although "hours of study" data were not collected in this experiment, retrospective reports by teachers lead to the estimate that less than half of the distance students met the 12-hour criterion and only one quarter studied as many as 30 hours.[5] Classroom based research on WES suggests that the time required to master the content of any one strand is at least 24 hours (Johnston, et al., 2001). What do these preliminary findings suggest regarding classifying distance learners for NRS? Three possibilities should be explored in this coming year: (1) treat them the same as any other learners with regard to expecting educational growth, (2) expect that most will not show growth and adjust the negotiated targets for the state to reflect this, and—in the special case of WES—(3) classify distance learners as Work-based Project Learners. Two of these need further discussion.

Expect No Measurable Gain

In the first years of experimentation with WES at a distance, states should consider adjusting their expected gains percentages to account for the distance learners. Massachusetts, for example, is considering adjusting its expected outcomes to allow for the counting of distance learners without the expectation that they will make educational gain. Massachusetts recognizes they are reaching a new population of learners—ones they might not otherwise reach in classroom-based programs. Expanding the learner population is considered an important step in developing distance-based programs in Massachusetts. By adjusting expected educational gain outcomes for the adult population as a whole to account for an increase in learners drafted through their distance programs, the state can demonstrate that more learners are being served without penalizing programs participating in the Project IDEAL experimental process if learners do not make as rapid progress working at a distance as they do in classrooms.[6] As distance efforts mature and increasing numbers of learners invest the time needed to make measurable progress, this decision can be revisited

Classify WES Students as Work-Based Project Learners

As participants in Pennsylvania's WES at a distance program become more experienced, they will likely discover additional ways to help encourage distance learners to engage in additional hours of study, but it is likely that there will still be large numbers studying for less than 30 hours. Add to this finding some of the challenges to measuring educational

---

[5] See Petty and Johnston (2002) for more detailed data. Determining "seat time" for each distance learner is a key addition to the Pennsylvania research plan for the 2002-2003 school year.

[6] Initially, distance learners will comprise only a small fraction of the total number of adults being served. If the target this year or next was a generous 1,000 learners, this would be only 4% of the Massachusetts program (1,000/25,000) and less than 1% of the Pennsylvania program (1,000/62,500).

gain discussed earlier, some consideration should be given to classifying WES distance learners who meet the 12-hour rule as Work Based Project Learners—especially those who study the Employment and Workplace Communication strands. Learners are placed in this NRS category if they study a workplace curriculum for between 12 and 30 hours. The classification is intended to enable programs to track and be credited for serving learners in short-term educational courses that are designed to facilitate their entry into employment. Distance learning programs using the WES materials may want to consider this as a viable option for counting those students who will likely be studying the materials for under 30 hours. To count students as work-based project learners, programs must designate students in this category prior to the beginning of instruction. As with measuring educational gain, determining student achievement as work-based project learners requires students be assessed with a standardized instrument, though pre-testing is not required. Prior to designating a distance learner in the work-based program category, programs must first identify a set of achievements or expectations for student learning as a result of studying the course. In Pennsylvania, for example, educators are examining whether their state-developed Workplace Foundation Skills measures might be tailored for use in measuring gain for WES learning. The Workplace Foundation Skills measures are currently used for purposes of measuring the degree to which tailored programs are meeting the needs of employers. Employers identify specific skills and knowledge they want their employees to learn. The program designs an instructional curriculum that is based both on the employer's requisites and that draws from the list of competencies identified in the Workplace Foundation Skills content. Tailored checklists are then constructed and administered to determine student achievement.

This approach to using competency-based checklists as measures of performance has potential for success with WES. States with similar workplace checklist materials may wish to correlate their assessment materials with the WES content. The process of assessment of WES distance learners could be integrated into the existing workplace programs of assessment using trained evaluators. Alternatively, Project IDEAL collaborators may opt to pool resources to develop checklists and performance-based assessment approaches similar to those used in Pennsylvania for use in all consortium states using the WES materials.

Recommendations for New Research
to Inform Policy Choices

This paper has identified the major issues surrounding assessment and accountability for adult distance education students. It has examined the various roles for assessment, discussed accountability and the NRS, and presented a case study of the assessment issues related to one curricular product. Throughout, the focus has been on describing the challenges and opportunities for states as they begin to incorporate distance education into their programs' offerings. As is true for any evolving field, there are more questions than answers. Adult educators, policy makers and researchers need to address the following issues:

Measuring Seat Time. How can "seat time" be counted for learners working at a distance? Important administrative decisions are made about counting and testing students based on cuts of 12, 30, and 50 hours. There are two distinct issues here:

1. How can we most accurately estimate seat time for distance learners? Some online curricula *(PLATO, SkillsTutor)* track students' study time; others do not. Learners could be asked to keep a log, but it places an undue burden on learners to keep accurate records of their activities. Project IDEAL has proposed a strategy that uses teachers. For each student they would count actual hours for activities that take place in a literacy center. For distance activities they would judge each learner's completed homework for completeness and then assign a standard number of hours based on that judgment. Tests will be done in IDEAL states this coming year to assess the validity of the data and the ease of data collection. In some states both teacher rating and student self-report data will be used with the same learners in an attempt to cross-validate both forms of data.

2. How does seat time for distance learners compare with seat time for classroom learners? Many standardized assessments assume that students will study for a specified number of hours before they show an educational level increase. If seat time differs significantly for distance students and classroom students, it will have implications for the realistic expectations of educational gains for distance study.

Measures of Curriculum-Based Progress. How can a learner and teacher know if the learner is making progress mastering the material presented in curricula such as WES? It is important for both learners and teachers to have ongoing feedback. It helps learners know if they understand the material and helps the teacher design interventions to help a student who is having difficulty. Some of the curricular programs being used by the IDEAL states have materials designed for these purposes; others do not. Additional work is needed to identify the curricular products that lack materials for student progress assessment and to propose (or develop) ways in which those needs can be met. An initial effort will be made by the Project IDEAL technical support staff to put online the Skills Preview and Skills Review quizzes from WES as a model.

Classifying Distance Learners. How should distance learners be classified for accountability purposes in the NRS system? Should all learners who meet the 12-hour standard be expected to show educational gains or should some students be considered informal, exploratory learners and not subject to meeting an explicit goal? What are the criteria for assigning students to one of these categories or to the Work-based Project Learner category, and when should that decision made?

1. *Should they count at all?* One possibility is that states could negotiate their expected NRS outcomes to allow for counting distance learners without the expectation that they will make measurable educational gains on a standardized assessment. This appears to have the potential to encourage

states to reach a new population and refine their distance learning delivery systems without focusing immediately on accountability issues.

2. *A special category.* If students are considered formal learners in the NRS system and are studying WES Employment or Workplace Communication, should they be treated as Work-based Project Learners?

Measuring Educational Gain. If distance learners are considered regular students, there are additional questions regarding which assessments should be used to ensure that there is a fit between the curriculum being taught and the assessment. Several problems have been identified with the existing gain assessment recommended for learners studying for their GED. Other problems have been identified with assessment of learners studying WES. All of these problems could be solved with the development of new standardized assessments. Is this a reasonable step to take?

Finally, the emerging interest in adult distance education calls for the establishment of mechanisms for sharing and disseminating information and using newly acquired knowledge as a basis for policy decisions. We strongly recommend an annual gathering in which state and national policymakers, as well as researchers and adult education practitioners, meet to consider the issues presented here and make recommendations based upon their experience and knowledge.

References

CASAS employability competency system: Test administration manual (2000). San Diego, CA: CASAS, Foundation for Educational Achievement, Inc.

Forlizzi, L. (1998). An overview of learner assessment. *The Pennsylvania Able Staff Handbook*. Harrisburg, PA: Pennsylvania Department of Education, Bureau of Adult Basic and Literacy Education.

Johnston, J., & Young, S. J. (1999). *Pilot test of Workplace Essential Skills and LitTeacher professional development courses*. Ann Arbor: University of Michigan Institute for Social Research.

Johnston, J., Young, S. J., & Petty, L. I. (2001). *National field test of Workplace Essential Skills*. Ann Arbor: University of Michigan Institute for Social Research.

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

Mislevy, R. J., & Knoles, K. T. (n.d.) (Eds.) *Performance assessments for adult education: Exploring the measurement issues*. Washington, DC: National Academy Press.

Petty, L. I. & Johnston, J. (2002). *Adult Education in Non-Classroom Settings: A Pilot Test in Pennsylvania*, Phase II: October 2001 - June 2002. Ann Arbor: University of Michigan, Institute for Social Research.

Reuys, S. (2001). New accountability rules pose dilemma for programs. *Adventures in Assessment, 13*. Boston, MA: SABES/World Education.

Sticht, T. G. (1999). *Testing and accountability in adult literacy education: Focus on workplace literacy resources for program design, assessment, testing, & evaluation*. El Cajon, CA: Applied Behavioral & Cognitive Sciences, Inc.

U. S. Department of Education. (2001). *Measures and methods for the National Reporting System for Adult Education: Implementation guidelines*. Washington, DC: Author, Division of Adult Education and Literacy, Office of Vocational and Adult Education.

U.S. Department of Education. (2001). *State reported hours of attendance in adult education programs (1997-2000)*. Washington, DC: Author, Division of Adult Education and Literacy.

Appendix: Detailed Content of WES Reading Strand

WES Unit	Workbook
Skills Preview	*Identifying main purpose of text (pp. 2, 7)* - Identifying main idea in a text (p. 2) - Understanding specific vocabulary (p. 2) - Generating concrete example of meaning in a text (p. 3) - Interpreting simple chart (p. 3, 5) - Interpreting simple work order (p. 4) - Gleaning specific information from text (p. 7) - Scanning text for information (p. 8) - Interpreting instructions (pp. 7, 9) - Problem solving (pp. 9)
Program Lesson 16 Reading for a Purpose	*Reading for a purpose* - Various reasons for reading (pp. 17, 18, 21, 26, 29) - Identifying the main idea of a text (pp. 19, 21, 29, 31, ref. to pages 120-122) - Applying information in job situations (pp. 21, 29, 32) - Interpreting a simple chart (pp. 22-23, 24-25) - Understanding specific vocabulary (pp. 27, 29, 32) *Strategic Reading* - Re-reading (pp. 17, 26) - Use reference books (other "resources") (e.g. dictionary) (pp. 17, 27) - Ask for clarification (pp. 17, 27) - Varying reading speed (pp. 17, 22, 31) - Scanning and Skimming (pp. 16-17, 18, 22-23, 26, 31, 32) - Taking notes & highlighting (pp. 17, 26) - Restating what you've read (pp. 17, 26) - Using context clues (pp. 17, 27) - Monitoring understanding (p. 26) - Picturing and visualizing texts (p. 26) - Thinking of an example of a key point (p. 26) *Types of documents* - Wait-staff job task chart (p. 14) - Excerpt from employee handbook (p. 18) - Excerpt from hospital safety manual (p. 20) - Chart, patient diet requirements (p. 22) - Example of a paycheck stub (p. 24-25) - Example of training manual (p. 26) - Directions from a sewing manual (p. 27) - Posting of training opportunities (pp. 30-31) *Write It* - Email message (p. 21) - Keeping own personal dictionary (pp. 29, ref. to 126) *Tech Tip* - Skimming and scanning with computer displays (menus/icons…) (p. 25)

Program Lesson 17 Finding What You Need: Forms and Charts	*Understanding forms and charts* - Identifying purpose (pp.33, 37, 38-39, 51, 52) - How to read forms and charts (pp. 33, 37, 43) - Completing forms and charts (pp. 33, 37, 41, 46-47, 49, 51) - Gleaning specific information (pp. 35, 41, 42-43, 44-45, 49, 52) - Paying attention to details (pp. 37, 46-47, 49, 51) - Problem-solving when needing clarification (p. 37) - Skimming and scanning (pp. 43, 51) - Using written information in job situations (pp. 43, 51, 52) - Figuring out abbreviations (p. 45) *Types of Documents* - Employment application form (pp. 34) - Change of Address form (p. 38) - Time Request form (p. 40) - Bus schedule (p. 42) - Employee work schedule (p. 42-43 & ref. to page 129) - Hospital patient dietary chart (pp. 44) - Tenant complaint form (p. 47) - Purchase order (p. 48) - Database display (p. 49) - Shipping label (p. 50) *Tech Tip* - Using computerized forms (p. 41) - Getting information from a database (p. 49) *Math Matters* - Abbreviations for measuring amounts (p. 45)
Program Lesson 18 Following Directions	*Following Directions* - Understanding purpose of directions (pp. 54, 58, 60, 62, 68) - Interpreting written directions (pp. 53, 59, 66, 68-69, 71) - Interpreting pictorial directions (pp. 53, 55, 64-65, 72) - Verbal communication and following directions (pp. 53, 70) - Using information for job tasks (pp. 65, 68-69) - Understanding consequences (pp. 66-67, 68-69, 71) *Strategies for Understanding Directions* - Restating/ summarizing directions (pp. 57, 59, 68) - Taking notes (pp. 57, 61, 70) - Explaining them to someone else (pp. 57, 59) - Using structure of diagrams and written directions (pp. 57, 63) - Reading thoroughly (p. 66) - Visualize instructions (p. 66, 69) - Get clarification (p. 66, 69, 70) *Types of Documents* - Flowchart (procedure for directing invoices p. 54, for shipping packages p. 62, ref. to p. 131) - Numbered-step instructions (computer labels p. 58, potato chip weight inspection p. 68)

	▪ Guidelines for a task (preparing garden soil p. 60, troubleshooting printer problems p. 70) ▪ Diagram (floor plan for furniture store p. 64, evacuation procedure p. 66) ▪ Employee handbook (evacuation procedure p. 66) *Math Matters* ▪ Decimal addition/subtraction, fraction – decimal equivalency (p. 69) *Write It/Communicate* ▪ Numbered instructions for a task (p. 61) ▪ Drawing diagram of household tool or appliance (p. 65)
Program Lesson 19 Reading Reports and Manuals	*Reading Reports and Manuals (workplace references)* ▪ Understanding purpose (pp. 73, 77, 78, 80) ▪ Reasons for reading materials (pp. 73) ▪ Types of reports and manuals (pp. 73, 77, 78-79) (e.g. catalog, directory, work manual, memo) ▪ Understanding organization (pp. 73, 77, 78, 85) ▪ Using information to complete job tasks (p. 73, 83, 85, 86-87, 90-91, 92) ▪ Interpreting information (p. 74-75, 80, 82) *Strategies for reading reports and manuals* ▪ Scanning and skimming (p. 73, 77, 82-83) ▪ Using additional resources (p. 73, 77, 81, 87) ▪ Thinking of multiple ways to find information (pp. 79) *Types of Documents* ▪ Product information (p. 74, 83, 88, 91) ▪ Memo (ref. to on p. 79, example of p. 80, 92, 119) ▪ Table of contents (ref. to p. 79, example of p. 81, 82, 84, 90) ▪ Ref. to index (p. 79) ▪ Ref. to glossary (p. 79) ▪ Activity request form (p. 86) ▪ Material Safety Data Sheet (p. 89) ▪ Tabbed binder (p. 90) *Tech Tip* ▪ Using computer program "help" functions (p. 81) *Write It* ▪ Memo to head teacher (p. 85) *Communicate* ▪ Discuss importance of safety (p. 89)
Skills Review	*Reading for Understanding* ▪ Understanding types of and purposes for workplace reading (pp. 94-95) ▪ Identifying main idea (p. 95) ▪ Getting/understanding specific information (p. 95-96, 99, 100-101) ▪ Completing documentation (p. 96) ▪ Understanding consequences, solving problems (p. 97, 99, 101) ▪ Visualizing information (p. 99)

	Types of documents
	▪ Employee manual (p. 92)
	▪ Material certification form (p. 95)
	▪ Order form (p. 96)
	▪ Shipping fee chart (p. 97)
	▪ Numbered instructions (p. 98)
	▪ Table of contents (phone book p. 100)
	▪ Reference information (phone book p. 100-101)