Spring 2004

Adventures in Assessment

Learner-centered approaches to assessment and evaluation in adult literacy

Volume Sixteen

Editor: Carey Reid
Design: Marina Blanter
Adventures in Assessment is published by SABES, the System for Adult Basic Education Support. With funding from the Massachusetts Department of Education, Adult and Community Learning Services, SABES promotes high quality adult basic education services through training, support, and resources that improve the skills and knowledge of practitioners and strengthen programs. Our approach to staff development encourages participation, reflection, and innovation. SABES serves Massachusetts adult education staff and programs through five Regional Support Centers and a Central Resource Center.

SABES maintains past issues of Adventures of Assessment online at http://www.sabes.org/resources/adventures/, where readers can download any of the following:

- **Vol. 15** (Spring 2003): *Assessment in Motion*
- **Vol. 14** (Spring 2002): *Examining Performance*
- **Vol. 13** (Spring 2001): *Meeting the Accountability Challenge*
- **Vol. 12** (Winter 2000): *Experiences with Standards-Based Reform*
- **Vol. 11** (Winter 1998): *Aspects, Levels, and Perspectives*
- **Vol. 10** (Winter 1997): *Time to Reflect*
- **Vol. 9** (Winter 1996): *Rethinking Assessment*
- **Vol. 8** (Winter 1995): *Meeting the Challenge of Authentic Assessment*
- **Vol. 7** (December 1994): *The Partnership Project*
- **Vol. 6** (April 1994): *Responding to the Dream Conference*
- **Vol. 5** (October 1993): *The Tale of the Tools*
- **Vol. 4** (April 1993): *Goal-setting, tape journals, math, and other topics*
- **Vol. 3** (November 1992): *Looking Back, Starting Again*: Tools and procedures used at the end of a cycle or term, including self, class, and group evaluation by both teachers and learners.

Opinions expressed in Adventures in Assessment are those of the authors and not necessarily the opinions of SABES or its founders. Permission is granted to reproduce portions of this journal; we request, however, appropriate credit be given to the authors and to Adventures in Assessment.

Adventures in Assessment is free to Massachusetts DOE-funded programs; out-of-state requests will be charged a nominal fee. Please email requests to Carey_Reid@worlded.org.
In the Fall of 2003, I began planning out this issue with Marie Cora, the former Staff Development Specialist for Assessment here at World Education. Marie and I agreed that we wanted this issue to keep very close to adult basic education classrooms.

The articles that we sought from the fine writers who have since contributed reflect that intention. You’ll find articles about integrating student goal setting into classes, using data to improve programs, and using the Arlington Education & Employment Program (REEP) writing assessment to inform instruction. You’ll also find two similar articles that explain basic elements of standardized testing in laypersons’ language. One of these articles has been specifically designed for teachers to use with their students, both to educate them about the many standardized tests that pervade our culture and to help them prepare themselves more effectively for test taking.

We hope you find these articles useful.

My best,

P. Carey Reid
Staff Development Specialist
System for Adult Basic Education Support
How to Do Your Best on Standardized Tests: Some Suggestions for Adult Learners ................................................................. page 5

Ronald K. Hambleton and Stephen Jirka provide an article written for teachers to use with their adult learners. The authors’ intention is to empower learners by providing them with jargon-free explanations of the basics of standardized testing and several proven test-taking strategies.

Using the REEP Assessment for ESOL and ABE Classroom Instruction ........ page 13

Joanne Pinsonneault and Carey Reid are veteran teachers who have found good ways to use the REEP writing assessment for instructional purposes.

Integrating Goal Setting into Instructional Practice ............................. page 25

The staff at The Center for New Americans in Western Massachusetts take the general goals that newly enrolled learners set at intake and use them in rich and constructive ways in the classroom.

A Basic Primer for Understanding Standardized Tests and Using Test Scores ................................................................. page 29

April L. Zenisky, Lisa A. Keller, and Stephen G. Sireci have written a companion piece to the learner-friendly article that opens this issue. The authors’ intentions are to help practitioners understand key concepts underlying standardized testing and to interpret test scores effectively.

Using Data for Program Improvement ............................................ page 35

Luanne Teller and her staff at the Stoughton (Massachusetts) Adult Education Program have been digging deeper and deeper into attendance and other data and using them to strengthen their program.

ACLS, SABES, and UMass: Perfect Together! ................................. page 39

Stephen Sireci offers an upbeat overview of the projects he and his colleagues at UMass Amherst are working on with MassDOE, SABES, and adult basic education practitioners.
**How to Do Your Best on Standardized Tests: Some Suggestions for Adult Learners**

**PART ONE: WHAT ARE “STANDARDIZED TESTS”?**

Educational tests, sometimes called “standardized tests,” seem to be everywhere. In Massachusetts, the Department of Education administers the Massachusetts Comprehensive Assessment System tests (better known as MCAS) in English Language Arts, Mathematics, and Science to learners in the public school system. The Educational Testing Service administers the Scholastic Assessment Test (the SAT) to learners who are considering going to college. The American Council on Education administers the Tests of General Educational Development (GED). You can’t get a high school diploma, go to college, join the military, get a professional license or certificate, or get a job without passing a test. You can’t even get a driver’s license without passing a test. With so many standardized tests around, adult learners would be wise to learn how to do their best on them, and to help their children do well on them, too.

Standardized tests are particular kinds of tests, different from the final examination a high school teacher might design for her math course, or the writing exercise an ESOL teacher might design to see how well his learners are doing. When talking about tests, “standardized” simply means that everyone who takes the test is given the same amount of time and sees the same or very similar test questions. “Standardized” also means scoring is done very carefully so that test scores do not depend upon who happens to be doing the scoring. Why are standardized tests so widely used? Because, by and large, they have been shown to be (1) an efficient way to collect information about what people know and can do, (2) objective in the sense that test scores do not depend to any great extent on who happens to score the answers, (3) valid in that they often provide relevant and useful data for making decisions about mastery of a body of knowledge and skills and potential for success, and (4) convenient and cost-effective because they can be administered to many people at the same time.

Governments, the armed services, industry, universities and colleges, credentialing agencies, and many other groups use standardized tests because they are convinced by the evidence that such tests offer the best basis for making decisions about who has the necessary knowledge and skills for some particular purpose (like going on to college or being hired for a job.) Human beings make tests and human beings administer them, and all human beings have biases. Bias can sometimes creep into standardized tests, but it can usually be spotted and the problem fixed, or, if the problem of bias cannot be solved, the test can be eliminated.

Ronald K. Hambleton and Stephen Jirka provide an article written for teachers to use with their adult learners. The authors’ intention is to empower learners by providing them with jargon-free explanations of the basics of standardized testing and several proven test-taking strategies.
Some people believe that standardized tests are used too often and that there are better ways to measure ability and readiness. For purposes of discussion, let’s consider the task of determining whether adult learners have the same knowledge and skills as high school graduates. This is an important task in the United States, because high school diplomas are an entry to higher education, the military, and lots of jobs. Many adults who did not obtain a high school diploma during their teens later want to demonstrate that they have about the same level of knowledge and skills as high school graduates and thereby gain the same opportunities. Today we have the Tests of General Educational Development (the five tests that make up the GED), which are used around the country as a way for people to demonstrate they have knowledge and skills equivalent to those of high school graduates. One alternative to passing the GED would be for adult learners to return to high school and take regular school tests along with state graduation tests, but with a million persons desiring GED certificates each year, this would surely be impractical. The external diploma programs offered by many adult basic education programs are an excellent alternative, but they require a great deal of individual conferencing.

As a standardized test, then, the GED certainly has its place. It provides many thousands of adult learners in this country with a second chance. Teachers are familiar with the material covered by GED tests, so they can design test preparation instruction effectively. And the GED is widely accepted as a high school equivalent: community colleges, universities, the military, skilled trades, and employers who require a high school diploma welcome those who demonstrate proficiency through the successful passage of the GED tests. Clearly, the GED tests and others like them have an important role to play in this country.

We believe that some of the problems surrounding the standardized tests used in adult basic education programs, such as the GED and the TABE, are not with the tests themselves, but with learners’ test-taking anxiety and lack of test-taking skills. These two factors are interrelated; knowing more about standardized tests and how to take them can boost a learner’s self-confidence and reduce her test-taking anxiety. However, people are not born with test-taking skills, and sometimes learners from other countries have had very little exposure to American-style tests with multiple-choice items and separate answer sheets, or with the computer-administered tests that are becoming popular.

PART TWO: DOING YOUR BEST ON STANDARDIZED TESTS

At this point, we would like to offer six very practical suggestions to help adult learners perform to their best ability on standardized tests.

1. Get positive about taking tests!

Adult learners need to think positively about themselves, the learning they are doing, and the tests they will be taking to assess their learning. While standardized tests can be daunting, they also offer adult learners a way to move up, to provide a role model for their children, to get a better job, or to go to college. All too often, adults without a basic education see themselves as victims. A positive
attitude can boost confidence and improve test performance.

Researchers have found that test performance is, in part, psychological. When learners receive positive messages about their ability to learn and to succeed academically, they are less likely to conform to stereotypes that they believe others have of them, and they perform significantly better on tests. So, adult learners and their teachers must be positive!

Adult learners need to see testing as an opportunity to demonstrate their ability, not evidence that they are victims of a system that cares little about them. Doctors often tell their patients to be positive, because research has shown that patients who remain positive live longer and avoid illnesses better than those who do not. The same is true for adult learners when taking tests—be positive and you’ll perform at a higher level.

2. Clear the brain for learning and testing!

Many adult learners lead stressful lives. Stress comes from family, from the job, from personal health concerns, from the times we live in, and so on. But if adult learners want to improve their lives and those of their family members, they need to find time to concentrate on learning. Adult learners need to have some quiet time each week to study, and regularity and consistency make learning easier. They must see this “learning time” as something they deserve. The study place should be quiet to allow for concentration—perhaps the local library on a Saturday morning, or a quiet place at home in the early morning or late night if necessary, and should be dedicated to studying, with books, paper, and pen readily available. Learners need to stay organized because this time is precious, and they owe it to themselves to make the most of it.

Adult learners also need some quiet time right before taking a test. An hour or two to clear their heads of life’s stresses, away from family, away from the job: time to think about the challenges associated with the upcoming test. An adult learner who arrives late for a test, huffing and puffing, upset about a family- or job-related problem is not emotionally ready for the challenges of a test. If failure follows, the test is often blamed, but the real problem might be that the adult learner was not psychologically ready to perform to her capabilities. If prior test-taking experience resulted in failure, the adult learner should strive to put that behind her and focus on the present test and her efforts to perform well on it.

3. Prepare for the test “strategically”!

We were talking with a colleague the other day who told us about an adult learner who persisted in studying for one section of a GED test that he thought he was weak in. He had failed the test several times previously, yet this one section was only 10% of the test. This learner would have been much wiser to consider the content coverage of the test (which was information readily available to him) and to plan his study time accordingly.

There are two key strategies for preparing to take a standardized test. The first strategy is to become familiar with the format of the test: What sorts of questions are asked, how is information conveyed, and how are answers logged in? This knowledge will reduce the level of surprise and confusion that robs the
test taker of time she could be using to answer questions. The second strategy is to research the content coverage of a test and then to apply the study time the learner has available on the content that will count the most.

With most standardized tests such as the TABE and GED, the format and content information is readily available. Let's take the GED as an example. It is based on a high school curriculum and performance standards that are used throughout the country. The five tests are in a multiple-choice format (except for one essay), and have been developed by experts familiar with secondary and adult education. The Language Arts Test emphasizes organizing text and the mechanics of writing. The Mathematics Test includes computational problems and real world problems and applications. The test will give you any formulas you will need to use. Calculators are used with one of the sections. Some math answers are multiple choice, but many are marked on little “bubble charts.” The Social Studies Test draws content from United States and World History, Government, Economics, and Geography. That test contains at least one excerpt from a major historical document, such as the Declaration of Independence. The Reading Test will have the adult learner read and interpret many different forms and varieties of literature, such as fiction, nonfiction, prose, poetry, and drama from different cultures and time periods, as well as use business-related documents. The Science Test has the test taker interpret and use scientific information in the form of text or graphics, and material from the life sciences or physical sciences. Adult learners might be asked to interpret experimental results or explain how results from a classic study apply to the everyday world. Even more detail on specific GED tests is readily available from the GED testing service and in bookstores.

Adult learners who want to study strategically can use information like that provided above to orient themselves to tests and focus their study time for maximum results. They can easily find out what content is covered by a particular test and how much importance will be given to various topics; for example, geometry makes up a small part of the GED math test. With this kind of information, learners can focus their study time on the most important topics, and when those topics have been mastered, they can move to the less-important ones. In addition, knowing what the most important content areas are can help learners find the right study aids.

4. Become familiar with test-taking techniques!

Going into a test with a good knowledge of basic test-taking techniques will help a learner to do his best. Much has been written on good techniques; here is a sampling of the most often repeated advice:

- Listen carefully to directions.

One of the most critical rules for adult learners is to listen carefully to the test directions: How much time is available? How will the test be scored? What advice, if any, is given about when to randomly guess on multiple-choice test questions? Does the test administrator have any special instructions? Knowing available time allows adult learners to apportion their time
so that they don't need to rush to finish at the end. Knowing about scoring also helps with time use: if 50% of the score will be assigned to essays, then test takers should devote 50% of their test time to writing the essays. And as for whether to guess on multiple-choice test questions, the answer depends on how the test items are scored. If there is no penalty for wrong answers, learners would be smart to answer all questions, so when time is about to run out, they should randomly guess at any remaining answers prior to handing in their answer sheets. On the other hand, if there is a small penalty for wrong answers, learners should be encouraged to answer if they can eliminate at least one of the answer choices. Otherwise, guessing has no particular advantage. Concerning special instructions, adult learners must remember to listen carefully: the instructions might include information about the most important questions on the test, whether or not calculators can be used, the desirable length of essay questions, and so on.

- **Scan the test before starting to answer questions.**
  Adult learners must remember to scan the test first to get an idea of length and difficulty. If the test is made up of multiple-choice questions, they should work on the questions in order and not spend too much time on any one question. Skipping around the test and doing a question here and there is not a good strategy because valuable time is wasted and might lead to errors in marking the answer sheet. If essay questions are part of the test, however, it makes sense to scan these questions and do the easier ones first.

- **Understand a question before answering it.**
  With multiple-choice questions, adult learners must read the questions carefully prior to answering. One of the most common mistakes is not answering the question that is actually being asked. Negative words in the “question stem” can be especially confusing. Sometimes words are highlighted in the question stem and these too are important clues. When in doubt, adult learners should eliminate choices that they know to be wrong, and then choose an answer, at random if necessary, from the remaining choices. Their partial knowledge will be rewarded with such a test-taking strategy.

- **Review the choices.**
  Here are a few additional tips for multiple-choice questions: (1) Read the question stem, try to think of an answer, and then look for it among the available answer choices. If that doesn’t work, at least eliminate the choices that appear to be wrong prior to guessing an answer. (2) If the answer choices are numbers or dates, middle choices are often correct. Note also that longer answers and/or more general answers among the answer choices are more likely to be correct. (3) Sometimes test takers are given a choice among essay questions. Adult learners should be encouraged to watch for this option. Sadly, many test takers fail to heed directions such
as, “Answer one of the three questions below” and try to answer all three instead, thus scoring lower than they could have.

- **Be flexible in approaching essay questions.**
  With short answer and essay questions, adult learners should be encouraged to try to write at least something, even if it’s just a few sentences. Often partial marks are assigned, so even a partial answer will generate some points. Before starting to write their essay, adult learners should try to prepare an outline. Paraphrasing the question itself is often a great way to start an essay. Clear writing, along with good grammar and spelling, are typically important in the way essays are scored. Adult learners should therefore remember to review their written answers for the use of good sentence structure, grammar, and spelling.

- **Review your work.**
  It’s important to remember to review your answers and essays. We all tend to breathe a sigh of relief when the last question has been completed, but adult learners who leave a test with time still available are missing an opportunity to improve their scores. The test is not over until the time is up, or at least until every answer has been checked and essays have been reviewed for grammar and spelling.

- **Stay as calm as you can.**
  Above all, adult learners should stay calm and simply do the best job they can with the time available. Staying calm will make you more efficient while you are answering.

5. **Take a practice test—or even better, take several practice tests!**
   No one learns to fly a plane, drive a car, swim, or play golf just by reading how-to books. Practice makes perfect, as the saying goes, and testing is no exception. There are lots of practice tests available for the GED; in fact, bookstores are full of books containing practice tests for most national standardized tests. However, adult learners need to take these tests under test-like conditions, and that means with the time limit that will be in place when the test counts. They need to be exposed to some of the natural anxiety that arises when seeing firsthand the test and test question formats. They need to practice their pacing, practice reading the questions and answering them carefully, practice making judgments about when and how to guess, and so on. Of course, these practice tests can be scored, so both weak and strong knowledge and skill areas can be identified. In a sense, every test, whether it is intended for practice or not, provides experience that can help one perform better on future tests. Adult learners can mull over their performance and how they might do better the next time—by being better rested, being more prepared on the content area, making improved use of available time, and so on.

6. **Read, read, read!**
   Studies have shown that vocabulary is one of the most important factors in doing well on standardized tests. Every time a test taker encounters a word he doesn’t know, he is less likely to understand a reading passage or a question. It sounds overly simple, but the fact is that vocabulary development is critical to success in all subject areas. The best way to
build vocabulary is by reading, reading, and then more reading. Reading shows words in context—that is, how they are really used in sentences to make meaning—and that’s the best way to learn them. Adult learners should read in their spare time, read on the bus to work, and read before going to bed...and should try to read for understanding.

Summary

In this article we have tried to give a good overview of standardized testing and provide practical suggestions for helping adult learners demonstrate their knowledge and skills on these tests. Our hope is that when learners are equipped with basic knowledge about these tests and proven test-taking approaches, they will be able to demonstrate what they are truly capable of.

Ronald K. Hambleton holds the title of Distinguished University Professor and is Chairperson of the Research and Evaluation Methods Program and Co-Director of the Center for Educational Assessment at the University of Massachusetts Amherst. Professor Hambleton has been teaching graduate-level courses at UMass since 1969. Stephen Jirka is a doctoral candidate in the Research and Evaluation Methods Program at UMass Amherst. His current research includes external validation of test scores and standard setting methods.
Using the REEP for ESOL Instruction: Joanne Pinsonneault

The REEP is the mandatory assessment in Massachusetts for measuring adult ESOL learner gains in the area of writing. After administering the REEP twice to my class of mid-level ESOL learners, I discovered the test—with its warm-up activities, familiar essay prompts, and simple scoring rubric—to be very learner friendly. As long as I did not use official testing prompts in my classroom, there was no reason why I could not share useful aspects of it with my learners. After beginning where I usually do (by talking to my colleagues), I eventually wrote a lesson plan. My initial objectives were to familiarize the learners with the REEP Writing Assessment’s pre-writing activities, essay prompts, and scoring rubric as one way to help me determine the best course for teaching writing. I decided to walk them through the testing procedure, using my own pre-writing questions and writing prompts and discussing the content and purpose of each step as they actually completed it. I began with the rubric because I wanted them to understand how their writing was being judged.

Step 1: Helping learners to understand the REEP testing process

I gave each learner a copy of the REEP rubric. I then explained the essay-scoring process: Two teachers who do not know the learner read and score the essay by using the rubric. If the scores are within a one-point range, they are averaged. If the range is greater than one point, a third reader provides a score that must be within one point of either of the first readers. Those two scores are then averaged. I demonstrated how the averaging system works.

Next, I wrote the five scoring categories and my explanations of them on the board:

1. Content and Vocabulary (addresses task, answers questions, varies vocabulary)
2. Organization and Structure (paragraph formation, details, essay structure)
3. Structure (sentence structure, grammar)
4. Mechanics (punctuation, capitalization, spelling)
5. Voice (addresses audience, engages/persuades reader, embellishes if necessary to make interesting)

We discussed each category in general, and I described the expectations for each category at every level on the point scale.

Step 2: Administering a practice test

I asked the learners to recall the testing procedure: introduction, brainstorming, conversation, prompt reading, and writing. I explained standardization and the rules for administering this test. Now the learners were ready to take a practice test.
**Step 3: Brainstorming**

We discussed the purpose of brainstorming (to get ideas flowing), and, keeping within the five-minute time limit that the REEP allows, practiced brainstorm listing with a prompt that I created: *What kinds of challenges do new immigrants encounter when they first come to America?* Afterward, I asked the learners what they had learned from the process. At first, many had a hard time differentiating between the items on the board and the process that led to the discovery of those items. Then they began thinking about personalizing the list, discarding irrelevant information, and grabbing onto others’ ideas. I modeled how they could create their own lists in their minds by adopting or discarding other learners’ ideas.

**Step 4: Conversation practice**

I began to wonder if the conversational “warm-up” activity used as part of the official assessment could also be used for instructional purposes. Back in the classroom, I explained to my learners that the conversation piece is another pre-writing activity that can help them to generate ideas for writing. As an intervention, I pointed out that they should ask themselves, “What did I learn from my partner?” not “What did I learn about my partner?”

Keeping within the REEP’s ten-minute time limit for this activity, they practiced with the following questions keyed to the new prompt:

1. What is your name?
2. Where are you from?
3. When did you immigrate to America?
4. What was one challenge that you confronted when you first arrived in America?
5. How did you meet this challenge?

Considering the time limit, I suggested that they focus on the last two questions, because these will provide them with ideas for writing. After ten minutes, I asked: “What did you learn from your partner? Did she or he give you any good ideas for writing?” I reminded the learners that this exercise does not require that they tell “the truth” about their experience; they can create circumstances and embellish to make their essays more complex and interesting. Therefore, if they listen carefully as the teacher gathers comments from other learners, they might just get a few ideas to use in their essays.

**Step 5: Understanding the prompt**

To successfully address any writing task, the learners must decide who their audience is, what the topic is, and what questions have been asked. I divided them into small groups and asked them to consider three variations on the same theme (see Figure 1).

In their groups they discussed the differences among the prompts and answered the following questions: “Whom am I writing this for?”; “Whom/what am I writing about?”; and “How many questions do I have to answer?”

Afterwards, they returned to the whole group for a discussion of the differences among the prompts. I asked the learners how they could use the pre-writing activities to their advantage in responding to each prompt. All of sudden it made sense to them! The comments were overwhelm-
ing. "My transition was so easy, but I could use what Mayra said about her life to help me answer the third prompt." By referring back to the rubric, we could point to how these pre-writing activities might lead to higher levels of writing quality.

Step 6: Administering the prompt

Finally, the learners took the practice REEP, under much the same circumstances as an official program assessment. All in all, these first six steps took about three hours, or one class session. But the work had just begun!

Step 7: Scoring

I scored the essays myself and, using a simple form, highlighted the strengths of each and offered concrete suggestions for improvement based on skills that I knew the learner had been working on and had demonstrated some ability with. (See Figures 2 & 3.) For example, if the learner had previously demonstrated the ability to write complex sentences but didn't do so throughout the essay, I suggested places where simple sentences could have easily and effectively been combined. During learner-teacher writing conferences, we discussed the scores and comments. I also gave the learners the opportunity to rewrite their essays, a standard option in our class.

Conclusions

I felt that I had achieved my primary objectives for the lesson: The learners gain increased awareness of the standards by which their writing is judged, and they had some initial ideas about how they can improve their writing. But how could I use this experience to inform my writing instruction? I'm still working on this, but I did spend some time looking at the scores of the class overall, and found a few surprises. Some of my work had been effective. For example, the majority of learners demonstrated good control over the verb tenses we had been studying! However, many learners were having difficulty with paragraph formation and needed a review of the punctuation rules. (Try again, Joanne!) I felt like I had a new place to begin, both with each learner individually and with the class in general.

I have also noted a few unexpected results, including reduced test anxiety and improved REEP scores. But best of all, my learners feel like they have more control over meeting their mandatory ‘improve writing skills’ goal. Teaching from the REEP has been very empowering.

Using the REEP for ABE Instruction: Carey Reid

I have been co-teaching a pre-GED writing class at the Jamaica Plain (Boston) Adult Learning Program one evening a week with the class's primary teacher, David Stearns. David and I have been experimenting with an authentic materials curriculum, using Boston Globe articles, Net-based research, and even research reports from Focus on Basics as starting points for activities that require our learners to summarize, analyze, and respond in writing to these sorts of challenging materials. We’ve been delighted at how willing our learners are to plunge into this stuff and make sense of it, even when a good understanding will most often require at least three readings.

When I learned that Joanne
Pinsonneault was using the REEP in her mid-level ESOL class, it occurred to me that the learners in David’s class might be able to handle it. David was game to try it, and so were the learners. We decided to use the REEP materials as instruments for periodic formative assessment. Here’s how we did it.

**Stage One: Unpacking the rubric**

We gave each learner a copy of the scoring rubric and spent about 90 minutes going through it in a whole group discussion. David and I explained that the rubric taken as a whole represented most of the elements that make up writing. We explained further that the rubric as an assessment tool provided good indicators for levels of quality and sophistication. At that point, we started to just cruise through it, pointing out indicators at random and giving oral samples of how these indicators might align with a particular piece of writing (e.g., “If the writer provides a lot of supportive detail and can use complex sentences, then she’s writing at such-and-such a level.”) After this exploratory phase, we asked the learners if they were willing to look at some sample essays and try their hand at scoring them. I am not exaggerating when I report that the learners were very intrigued by the rubric and equally eager to try applying it to actual writing.

**Stage Two: Scoring the anchor essays in small groups**

David and I really kept our fingers crossed with the next stage. Could the learners apply the rubric to actual writing with some accuracy, or would they become frustrated by such a task? We reminded ourselves that one of the reasons we decided to use challenging materials and exercises in our class was our determination to put into actual practice ideas and theories we professed to believe in—e.g., that adults learn best by doing, that they know what they want, and that they have acquired tons of knowledge through their life experiences. We did not want to appear to be the harbingers of special knowledge (“We know all the grammar rules, but you don’t”), or superior in position (“The REEP is for teachers, but not for learners”), or unconsciously condescending by coddling or “protecting” our learners (“Concepts and materials from standardized testing are much too complicated for you to understand”). In fact, in our class-planning sessions, David and I would call each other on perceived instances of these kinds of presumptions.

For this stage, we decided to use the essays currently used in the training of Massachusetts teachers for REEP scoring. As in that training, we decided to use the six “anchor essays,” each of which represents the same score across all levels—for example, a “3” all the way across. Specifically, the six anchor essays represent scores from 1 through 6. We told the learners that they would be looking for that consistent score level for each essay, explaining that this would simplify the exercise a bit, adding that this was the same process used to train Massachusetts teachers to understand the REEP.

David and I then organized the class into three workgroups of five learners each. We gave each learner a packet of anchor essays and a blank scoring sheet and asked that they read a particular essay together, score it together using the rubric, and try to reach agreement on a score. For some time, we’d been using collaborative learning approaches, so the
learners stepped right into this task. One by one, the workgroups read and scored a particular essay, and then when all three groups were finished David and I asked the facilitators to report the scores. Then we discussed the scores and the reasons behind them.

Frankly, David and I were surprised by the results of this exercise. In nearly every case, the workgroups gave the same score to a given essay as the REEP “experts” assigned to it. In those cases where their and the experts’ scores were different, the variance was only a single point. In addition, the three workgroups nearly always agreed on the same score, and when they varied it was only by a single point. In full class discussions, the reasons for giving scores were discussed, and where scores differed the workgroups defended their choices. Occasionally, an individual learner would disagree with the other members of her group; in the full class discussions, she would get a chance to defend her choice. (A great anecdote to share: One time a very shy learner revealed that she did not agree with the score the other members of her group had given. At that point, the other members asked her to speak up and defend her score. As it happened, her score matched that of the experts. Needless to say, the shy member’s standing in her group went up a notch as a result of that episode.)

Stage Three: Using the rubric to score learner essays

The class explorations of the REEP really intrigued the learners. When David and I said we hoped to apply the rubric to their own writing, they were completely supportive of the idea. At this point, David and I began to realize that because the REEP is not used as an official learning gains assessment in Massachusetts for non-ESOL learners, such as those in our class, we were free to adjust the materials and approaches as we felt we needed to. For example, we could alter the rubric if we wished, to make it more appropriate for our class. We were also free to design prompts that we felt suited our class better than those that are more appropriate for ESOL learners, which are based, for example, on personal letters or narratives. It struck us that it might be better for our learners to use prompts that resembled those used for the GED Writing Test.

The first prompt we developed basically centered on the question, What is the most important profession in modern society and why? To get the ball rolling, we started with an open discussion and some brainstorming, as Joanne Pinsonneault had done. Then we asked learners to write original essays for 30-45 minutes in response to a prompt based on the brainstorm question. (See Figure 5.)

David and I then read over these essays and, again as Joanne had done, scored them and noted those scores on a single-learner score sheet. We decided to list for each learner only a few Notable Strengths and one Area for Improvement. Our idea was to provide the learners with a Next Step for improving the draft, based on ideas about process writing we’d been applying all along. At the next class, David or I sat down with each learner and went over the scoring and the notes we’d made together. (We’ve provided an example of an essay and score sheet as Figures 4 & 5.)

Stage Four: Using the essays as “authentic materials”

At this point, everything we were doing with the REEP and with our overall
authentic materials approach came together in a nice neat package: David and I decided to use the learners’ essays as source materials, as the stuff that future lessons could be based on. For example, we gleaned fifteen sentences with errors from the various essays, listed them on a single sheet, and then asked the learners to discuss and edit them together in workgroups. The learners were much more motivated to edit samples of their own work than materials from workbooks.

Next steps: Using the rubric more widely

David and I are just beginning to experiment with using the rubric as a self-assessment and peer-assessment tool. For example, the class is now knee-deep in a several-week project in which we’re applying the NCSALL research on learner persistence to the learners’ own lives. We began by reading and discussing together the research findings (Focus on Basics, Volume 4, Issue A). The learners are now writing essays composed of an initial summary of the research, a commentary on how the findings apply or do not apply to their own lives, and a final section on ways in which our own writing class might be altered to better support their persistence(). Now that most of the learners have a first draft, we will be asking them to use the rubric to self-assess the piece and, perhaps in conference with one of us, plan out the second draft.

David and I are also just beginning to try to integrate the rubric with a peer-editing process that the class is developing together. We’ve come up with rules and guidelines around constructive criticism; now we’re wondering if the rubric would be a good tool for promoting objective and constructive critiquing. We shall see!

A final note...

David, Joanne, and I have presented these ideas in conferences and workshops over the past year. A common concern among participants is that the REEP scoring rubric might be set too high for new readers and beginning ESOL learners. Gradually, however, we’ve all come to realize that teachers are free to adapt these materials for their own learners. They can simplify the indicators, make them more positive sounding, reduce the number of levels, and so forth.

We would like to offer one caveat, however. Using the REEP rubric has, in our opinion, substantiated the claim among many theorists and experienced practitioners that adult learners know a lot more about writing than might be readily discernible. The learners often attach, constructivist-fashion, their working knowledge of writing elements to rubric terminology: in a sense, they are enabled to give voice to what they already know. So, before simplifying or adapting the rubric, you might first want to find out if you really need to.

Joanne Pinsonneault is an ESOL teacher with the UMass Dartmouth Workers Education Program in New Bedford, Massachusetts. Carey Reid is Staff Development Specialist for Licensure and Assessment with the System for Adult Basic Education Support in Massachusetts.
Joanne's Three Practice Prompts

[In an actual writing situation, each would be printed on a separate piece of paper with lines provided for the learner's writing.]

**REEP Practice: Immigrating to America**

**Question Number 1**

Your friend writes, “I am moving to America and I need your help. What kinds of challenges and problems can I expect to encounter when I first arrive? Where can I go to get help? Do you have any advice or suggestions for me? Thank you for your help!”

Instructions: Write a letter to your friend. Be sure to answer his/her questions. You have 30 minutes. You may NOT use a dictionary or talk to other learners during the test. Do not write on the back of the paper. You may use extra paper. Check your letter before you hand it in.

**Question Number 2**

Think about how difficult it is to immigrate to a new country. What challenges confront an immigrant family after moving to America? What resources are available to immigrants once they arrive here? Who can immigrants turn to when they need assistance?

Instructions: Write an essay about the challenges confronting new immigrants to America. Be sure to answer the questions. You have 30 minutes. You may NOT use a dictionary or talk to other learners during the test. Do not write on the back of the paper. You may use extra paper. Check your essay before you hand it in.

**Question Number 3**

When did you immigrate to America? Where did you come from? What challenges did your family confront after moving to America? What resources did you use to help you confront these challenges? How long did it take for your family to adjust to life in America?

Instructions: Write an essay about the challenges you confronted when you immigrated to America. Be sure to answer the questions. You have 30 minutes. You may NOT use a dictionary or talk to other learners during the test. Do not write on the back of the paper. You may use extra paper. Check your essay before you hand it in.
FIGURE 2

An Example of a Learner's Response to a Practice Prompt

NAME: Zega

REEP Practice: Immigrating to America
Question Number 1

Your friend writes, "I am moving to America and I need your help. What kinds of challenges and problems can I expect to encounter when I first arrive? Where can I go to get help? Who can I ask for help? Do you have any advice or suggestions for me? Thank you for your help!"

Instructions: Write a letter to your friend. Be sure to answer his/her questions. You have 30 minutes. You may NOT use a dictionary or talk to other students during the test. Do not write on the back of the paper. You can have extra paper. Check your paper before you hand it in.

Dear Anna, how are you.

I received your letter, and I am so happy for you. You are a good person. I never forget those all days when we played together in the High School. We had a fun time with a bigger person... and jokes, those times we never forget.

I am more than glad to help you. In America, life isn't easy. The biggest problems you can expect will be the language. English is a difficult to learn and to find a job. Without understand English, it is very difficult.
FIGURE 2 (PAGE 2)

An Example of a Learner’s Response to a Practice Prompt

Name: Zega

But you are a hard worker, I don’t think you have any problems. We have a few once programs that can help you to find a job and some school programs to help in English. I believe you will have a better life in America. You can count on me for everything do you need, remember that it’s friends are for. I hope see you soon. Zega

5-103

Sincerely, ____________________
FIGURE 3

Joanne’s Scoring and Commentary Sheet for the Learner’s Response

REEP ESSAY SCORING SHEET

Scored by: Joanne
Date: 5-6-03

<table>
<thead>
<tr>
<th>Student Name</th>
<th>Content &amp; Vocabulary</th>
<th>Organization Development</th>
<th>Structure</th>
<th>Mechanics</th>
<th>Voice</th>
<th>TOTAL (sum?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beza</td>
<td>[Actual scores have been omitted to avoid “debates”]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Beza
Wow! This is a very successful letter. You were able to put my general suggestions to good and immediate use.

1. Identify paragraphs
2. Use question mark after question
3. Look to find words written in Portuguese (problema, programas)
FIGURE 4

An Example of a Learner’s Response to One of David and Carey’s Prompts

REEP WRITING ASSESSMENT: #1

Your name: Evelyn S

Instructions: You have 30 minutes to write a response to the prompt in italics below. If you need extra paper, please ask the teacher for it. Please do not use a dictionary. Look over your writing before handing it in to the teacher. Your teacher will sit down with you next week and discuss your writing.

Here is the prompt:

What, in your opinion, is the most important profession in modern-day society? Give at least three reasons why you believe these professionals are the most important and explain why.

I believe teacher is important because if she or he don’t teach you what you need to know, you cannot become a professional professional in life. A teacher is like a friend that wants you to learn because you could improve in what you need. Teacher is good because they help kids to read and write to the picture. They could help you to accomplish what you need in life. They could help you in the computer also help you to get your G.E.D. also helps with strict in life. They can keep you encourage you to study and help with your goals. They can keep you encourage you to study and help with your goals. If you need extra help they are there for you. In my opinion everybody needs a teacher to learn.
FIGURE 5

David and Carey's Score and Commentary Sheet for that Learner's Response

REEP ESSAY SCORING SHEET

Scored by:  David J. Carey
Date:  9/23/2003

<table>
<thead>
<tr>
<th>Student Name</th>
<th>Content &amp; Vocabulary</th>
<th>Organization Development</th>
<th>Structure</th>
<th>Mechanics</th>
<th>Voice</th>
<th>TOTAL (sum/5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evelyn S.</td>
<td>[Actual scores have been omitted to avoid &quot;debates&quot; on accuracy.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notable Strengths: This writer pushes herself to use new words and write complex sentences.

Next thing to work on: Organizing sentences into paragraphs.
Integrating Goal Setting into Instructional Practice

Our program offers ESOL classes from the beginning to advanced levels to adult immigrants and refugees in Franklin and Hampshire counties in Western Massachusetts. In this article we'd like to share our approach to learner goal setting in the hope that other programs will find it useful.

Goal setting at intake

During the intake interview, Massachusetts adult education programs ask learners about their reasons for wanting to study. Among other things, this activity supports a Department of Education reporting requirement. At this stage, students usually have only a broad idea of what they want to accomplish. Examples of goals frequently mentioned at intake interviews are: “to speak English,” “get a job,” and “learn about the U.S.” However, if students are to feel successful and motivated in class, and to persist in adult basic education classes in the face of the many obstacles in their lives, we believe it is necessary to make these general goals clearer and more specific.

At The Center for New Americans, we have developed a successful classroom-based approach that helps students explore the goals they reported at the intake interview so that they become more specific, measurable, achievable, and realistic. Teachers then link these goals directly to classroom activities. Connecting learners’ goals and class instruction helps teachers meet students’ needs and allows students to experience success in meeting their goals, which contributes to their increased motivation and persistence.

Goal-setting activities in the classroom

The general goal that a student reports at intake is recorded on the Student Goals Form, which is passed on to the teacher. During the first two weeks of a new class session or tutorial, the teacher presents activities that help the student break the general goal into smaller steps, or mini-goals. In accomplishing the mini-goals, students experience success and personal satisfaction. Mini-goals also help students to realize the amount of time needed to achieve their larger goal. These activities are the heart of a successful class and must be viewed as part of instructional time, not separate from the curriculum. In addition, they are crucial to the development of our learners as co-negotiators of the curriculum.

By the end of the second week, students are able to outline several smaller, more specific steps toward their larger goal. In class, each student thinks about what he or she wants to study the following week and reports to the group. These student requests become the basis of the curriculum for the week ahead. The
teacher plans lessons, activities, and materials that respond to these requests. The teacher might include other elements that are needed based on learning assessments. At the end of each week, students reflect on their learning in their logs, and again make requests for the following week—in essence, setting new mini-goals. The reflection time allows students to self-assess what they have learned and how well they have learned it. Both reflection and planning take time and are considered part of instruction as well. The teacher always responds to the new requests the following week, and assesses past lessons by observation, evaluation of performance tasks, quizzes, and other formal or informal modalities. This process progresses in a spiral, with the goals directly informing instruction, followed by assessment by both students and teachers and the setting of new mini-goals as the class continues. (See Figure 1.)

How does it really work?

Let’s think about a beginning ESOL class of ten learners. The primary goals set by these students at intake were to communicate more effectively in English, get a job, learn about U.S. culture, and become a U.S. citizen. A goal-setting activity might begin by using pictures to teach the names of several places in town, including town offices and schools and other places used by the learners. Students could also draw pictures, and a list could be put up on the wall in the classroom. Once these places were identified and could be recognized, the teacher could ask the students if they needed or wanted to use English in these different places. The teacher could ask each student to prioritize which three or four he or she wants to focus on during the class cycle. A calendar is useful here to emphasize the finite amount of time available for a given topic. Students would then write the selected places that interested them in their individual logs.

Among other introductory activities, the curriculum for the initial week of class might include teaching students how to name places where they need or want to communicate in English. The following week might include one specific place—for example, the doctor’s office—where a dialog about calling for an appointment would be studied, practiced, and role-played; a TPR (Total Physical Response) activity might be performed to help students learn what doctors and nurses might say; students might study a vocabulary lesson about the people and things at the doctor’s office—ailments and symptoms, for example. The curriculum depends on what needs the students have expressed.

At week’s end, each student would reflect on their learning and discuss what they want to study the following week. The teacher might need to narrow the focus of the requests by asking questions. These ten student requests would become the basis for the teacher’s lessons in the upcoming week.

In our next example, students of an advanced ESOL class have set similar goals at intake: to communicate more effectively in English, get a better job, and learn about U.S. culture. An activity that helps these advanced students develop mini-goals involves their breaking into small groups and making lists of what they can do in English. A follow-up
activity is to have them develop a list of what they want to be able to do in English. These lists can be put up on the classroom wall to help the students remember what has been discussed.

One item on the second list might be “to speak fluently”—a very broad and long-term goal. Students can work together to explain their reasons for selecting this goal, and through this process, their individual needs will become clearer. For example, one student’s goal of speaking fluently might actually mean being understood by Americans. For another, this same goal might mean being able to express feelings in English. By probing, the teacher might discover that for the first student pronunciation of specific sounds is difficult and for the other a lack of specific vocabulary or cultural appropriateness is the area of concern. Calendars or timelines work well with these students to break the goals down into manageable steps. These mini-goals, different for each individual, should be recorded in the students’ learning logs.

The teacher again collects the requests and plans the lessons for the next week based on the mini-goals developed. At the end of the following week, students reflect individually on the activities and on their learning and set new mini-goals, which become the basis for activities once again. In each class, the challenge is to marry all the requests or mini-goals in one week’s time, to respond to all of them to a reasonable extent. Once the students learn to reflect and are able to recognize what they still don’t know or don’t know well enough, they become adept at making more specific requests. By the end of the semester, another round of formal assessment is conducted and reported. Intake goals that have been met are recorded at this time.

**In conclusion**

We believe this process of week-to-week curriculum design is feasible for all ABE classes, even those with more defined curricula such as GED. If the subject is essay writing, steps that need to be mastered can be identified and a timely plan set in motion. Students can choose to write about topics that interest them.

Elsa Auerbach writes that “The essence of a participatory approach is centering instruction around content that is engaging to students.” As a staff, we have found that responding to specific requests from students for activities makes lesson planning easier and increases students’ motivation and retention, and that students are more likely to be engaged and active learners if the material is relevant to their lives.

---

This article was written by Nicole B. Graves, ESOL teacher (beginners and high intermediate levels) and ESOL Program Coordinator, and Peg Cahill, ESOL teacher (high beginner/low intermediate level) and Support Services Aide, with input from other members of the staff. The Center for New Americans teaches English to immigrants and refugees in Amherst, Greenfield, and Northampton, MA.

---

FIGURE 1
CENTER FOR NEW AMERICANS
How Goal Setting Informs Classroom Practice

Initial Assessment

Reporting

Initial goal setting at intake

Goal setting for the session during class

Initial curriculum development
• Lesson planning
• Teaching/activities
• Students’ reflection

Setting specific mini-goals for the next week

Assessment
• Formal
• Informal
• On-going

On-going curriculum development
Introduction

It’s nearly impossible to live in American society today without having to take some kind of standardized test. You have to pass a test to get a driver’s license, get American citizenship, receive a General Educational Development (GED) certificate, get into college, and be considered for certain kinds of jobs. Here in Massachusetts, our children’s teachers have to pass state tests to be licensed, and the children themselves have to pass the Grade 10 version of the Massachusetts Comprehensive Assessment System test, or MCAS, to graduate from high school.

Why do we have to take all these tests? Basically, because there is widespread agreement (but not complete agreement) that tests can tell if a person has the knowledge or skills needed for a diploma, a certificate, a school class level, or a job. But it’s not just any test we’re talking about here – it’s standardized tests. Standardized tests are used because people feel that if you’re going to judge someone’s abilities, you’d better use a means that’s reliable and fair, and standardized tests are designed to be reliable and fair – though people might disagree about whether they succeed in those goals.

We will not debate that issue here; the purpose of this article is to equip readers with a basic understanding of what goes into standardized test making and what test scores purport to show about learners’ skills and abilities. We welcome any constructive use of this knowledge, whether it be better instruction or better policies, but all constructive uses start with accurate knowledge.

Meeting the reliability and validity criteria

Federal policies now require that states prove that ABE funds result in learner gains in reading, writing, language acquisition, and math. In addition, they require that states measure these gains with valid and reliable tests. After months of reviewing many standardized assessments and their respective alignment with the Curriculum Frameworks, Massachusetts policymakers and education professionals have agreed to use the TABE for ABE Reading, Writing, and Math; the BEST for ESOL Speaking and Listening; and the REEP for ESOL Writing. Scores in each of these tests are meant to represent what students know or can do in those areas. What does it mean when we say these tests are reliable and valid? Let’s take up each of these concepts in turn.

Reliability

The consistency of scores across different administrations or scorers is known as reliability. It is crucial that test scores effectively reflect the student’s capabilities without influence from extraneous factors.

April L. Zenisky, Lisa A. Keller, and Stephen G. Sireci have written a companion piece to the learner-friendly article that opens this issue. The authors’ intentions are to help practitioners understand key concepts underlying standardized testing and to interpret test scores effectively.
scores be adequately reliable in representing a person's knowledge and skills. Some level of error is always a factor in testing (more on this later) and test scores. If a person takes the same test on different days, we expect the results to be slightly different, but the more error there is in the test's make-up, the more different the two test scores are likely to be. If the two test scores are very different, it is reasonable to conclude that the difference is due to test error and that the scores do not really reflect what the test taker knows and is able to do.

Inconsistencies in scoring tests might also undercut reliability. Some tests are composed of multiple-choice questions, while others require that the test taker construct a response, such as an essay. Scoring a multiple-choice question is straightforward, because there is one right answer; the answer provided is either correct or incorrect. Therefore, regardless of who scores the test, the score on that question will be the same. Essay-type questions, however, require human judgment and are therefore more difficult to score. If two people read the same essay, it's likely that each person will give the essay a slightly different score. However, if the two scores given by the two scorers, or "raters," are very different, then the score on that essay is not very consistent, or reliable.

The measure of consistency between scorers is called inter-rater reliability. The closer the scores assigned to an essay by different raters, the higher the inter-rater reliability of that test. While it might seem impossible to get different raters to assign exactly the same score, it is possible to train raters so that they all score in a very similar way. If this goal is accomplished, there can be more confidence that the score assigned to the essay reflects the ability of the student.

Validity
How do we know whether a test measures the ability we are interested in? Even if a test is perfectly reliable and virtually error-free, how do we know if it is measuring the abilities we want it to and not something else? This is the central concern of validity, and ultimately involves the kinds of judgments that can be based on test scores.

Let's consider a math test consisting only of word problems. The test score could appropriately be used to indicate the student's ability to solve math problems that require reading; that would be a valid use of the test score. However, using the test score as a representation of the student's math ability in general would not be valid.

People who develop tests analyze them in several ways to determine the appropriate (i.e., valid) use of test scores. Let's review some of the issues considered in determining the valid use of test scores:

- Do the questions on the test represent the entire subject matter about which conclusions are to be drawn? For instance, if a test is designed to measure general arithmetic ability, there should be questions about addition, subtraction, multiplication, and division. If there are no questions about division, the test does not measure the entire content of arithmetic, so the test score cannot be said to reflect general arithmetic ability.
• Is the student required to demonstrate the skill that the test is intended to measure? Tests should be directly targeted to the skills measured and that skill should affect test performance. For example, a test designed to measure writing proficiency should ask test takers to write something, and better writers should be shown to receive higher scores.

• Are the test scores consistent with other indicators of the same knowledge and skills? Suppose a student takes a test designed to measure writing ability. If the student does well on writing assignments in class, then he or she should also do well on the writing test, so long as the type of writing on the test is consistent with that done in class. On the other hand, students who do not perform well on writing assignments in class should not do as well on the test. The validity of using that test score as an indication of the person's ability is questionable if there is inconsistency between the score and classroom performance.

Using test scores

By itself, a test score is just a number. Elsewhere in this issue, you'll discover how teachers are finding ways to apply elements of goal setting and assessment to classroom practice; our purpose here, however, is to provide readers with a basic understanding of standardized tests and scoring. When teachers, students, and others who use test scores are looking at a test score for a particular student, there are additional few pieces of information they can use to make that number mean something. In the next few pages, some of these pieces of information are explained to help you understand what test scores do and do not mean.

Test score scales

A score scale is the range of possible scores on the test. Score scales come in all shapes and sizes. On the TABE, for example, different students might get scores as divergent as 212 and 657. In contrast, on the REEP, scores range only from 0 to 6. A student who takes the BEST, depending on his or her ability to comprehend and speak English, will score from 0 to 65 or higher. Is a 212 on the TABE a “better” score than a 5.4 on the REEP? Even though 212 is a bigger number, these two scores come from tests that are very different and are designed to test very different things. For this reason, comparing scores across different tests is generally not a good idea.

Because scores from the REEP, the TABE, and the BEST are all on different score scales, the number a person gets as a score on one of those tests has meaning for that test only. It might be confusing to have different score scales, but the people who develop tests do this on purpose to make sure that users do not interpret scores on a particular test according to some other standard or yardstick.

For example, in the United States the score scale of 0 to 100 is commonly used in many classrooms, but people who make standardized tests often avoid that score scale because many people would assume that such scores mean the same thing they do in the classroom. Sometimes test developers work really hard to create a unique score scale: e.g., on one test used in the United States for admission to medical school, scores are graded from J (the lowest score) to T (the highest score)!
Interpreting Test Scores Rule 3:

Find out what the range of possible scores is for the test you are using. Knowing how high and how low scores can be for a particular test is important to understanding students’ scores.

Error in test scores
As we explained at the beginning of this article, some error is always a factor in test score interpretation. In fact, tests simply cannot provide information that is 100% accurate. This might sound surprising, but this is true for many reasons; for example:

- The extent to which a student has learned the breadth and depth of a subject will influence how she or he performs on a test. On a reading test, for example, a student might do well with questions about word meaning and finding the main idea of a passage but have had less practice distinguishing fact from opinion. The experience (or lack thereof) that a test-taker brings to the test represents a source of error in terms of using the test score to generalize about the student’s reading ability.

- Sometimes a student taking a test is just plain unlucky. If a student is tired, hungry, nervous, or too warm, he or she might do worse on the test than if the circumstances were different.

- A test might have questions that seem tricky or confusing. If a student is not clear about the meaning of a question, he or she will have trouble finding the correct answer.

- As we mentioned earlier, mistakes may be made in scoring a test. When students are not given credit for correct answers or are given credit for incorrect answers, score accuracy suffers.

Standard error of measurement
The score a person gets on the test is meant to indicate how well that person knows the information being tested. One way of looking at a test score is to think of it as consisting of two parts. One part represents the real but unknowable true ability of a person. This part is unknowable because it is never possible to get inside someone’s head and have a perfect measure of their ability in the area of interest. The other part of a test score represents the error, all the things that make the test a less-than-perfect snapshot of someone’s knowledge at one moment in time. Unlike the way we can manufacture a yardstick that is exactly three feet long to measure length, even the best tests can provide scores that are only approximations of the true ability.

Unfortunately, it is impossible to break these two pieces of a test score (the true ability and error) apart. But it is important to understand that any test score contains a certain amount of error, and as we’ve illustrated the error might be due to things that are going on with the test-taker or things that involve how the test is created or scored. Errors in test scores cannot be completely eliminated, but fortunately there are techniques that can be used to provide some idea about how much the score is affected by error.

For example, testing specialists can calculate the standard error of measurement, which can be thought of as the range of scores obtained by the same person taking the same test at different times. The standard error of measurement
is a “best guess” about how close the test is to measuring a person’s knowledge or skill with 100% accuracy. The standard error of measurement is a statistical estimate of how far off the true score the test score is likely to be.

Let’s take the TABE as an example. Suppose a student takes the TABE Reading Test, Level 7E and gets a score of 447. First of all, that score isn’t very low or very high. The next piece of information that will be helpful in understanding this TABE score is the standard error of measurement. The statistics of test development have shown that the standard error of measurement associated with 447 is 17 points, which means that the student’s true score is probably between 430 and 464. This score range was calculated by adding and subtracting 17, the standard error of measurement, from the score of 447.

The standard error of measurement gives us a good idea of score accuracy. In the last example the true score was described as probably falling within 17 points of the score the student got on the test; for a score of 630 on the same test, the standard error of measurement is a much bigger number: 64. In this case, the student’s true score falls between 566 and 694. There is probably a very big difference in TABE reading knowledge between a 566 score and a 694, so it would be harder to interpret a student’s knowledge within such a large range. The size of the standard error of measurement is in large part dependent on the reliability of the test, which was explained previously.

Conclusions

Concepts like reliability, validity, test score scales, and standard error of measurement scales give meaning to numbers that on their own might not mean much. Of course, the score that someone gets on a test is just one piece of information that tells what he or she knows and is able to do in one very specific and carefully defined subject area. While tests and test scores are important, and it is important to try your best on any test you take, it is also important to remember that any one test score is just that: one test score. The sidebar rules for interpreting test scores given in this article might help you use test scores in meaningful ways.

Are all tests as good as they should be? Do all tests provide useful information? Unfortunately the answer is “no,” but researchers at UMass, working in collaboration with the Massachusetts Department of Education, Adult and Community Learning Services, are striving to create tests for ABE students that produce scores that are reliable and can help us make valid decisions about students and programs. Our efforts are focused on making sure the numbers that are test scores – whether from the REEP, the BEST, the TABE, or any new tests that will be developed – are as meaningful and dependable as possible.

Interpreting Test Scores Rule 4:

For standardized tests, look in the technical manual and find out the standard error of measurement. If it seems like a small number relative to the test score scale, you can be more confident in the accuracy of the test score than if it is a big number relative to the test score scale.
April L. Zenisky is Senior Research Fellow and Project Manager at the Center for Educational Assessment at UMass Amherst. Her research interests include computer-based testing methods and designs, applications of item response theory, and innovations in test item formats. Lisa A. Keller is Assistant Professor in the Research and Evaluation Methods Program and Assistant Director of the Center for Educational Assessment at UMass Amherst. Her research interests include Bayesian statistics, computerized adaptive testing and item response theory.

Stephen G. Sireci is Associate Professor in the Research and Evaluation Methods Program and Co-Director of the Center for Educational Assessment in the School of Education at UMass Amherst. He is known for his research in evaluating test fairness, particularly issues related to content validity, test bias, cross-lingual assessment, standard setting, and sensitivity review.
Using Data for Program Improvement

BY LUANNE TELLER

Adult literacy practitioners now collect and report various levels of data to meet state and federal accountability requirements. There is no denying that these tasks are burdensome; however, many programs such as ours are moving beyond simple collection and reporting and using the data to strengthen our programs. In this article I hope to show that measuring and understanding student performance help programs, partnerships, and government to demonstrate and promote our true impact on our communities. Given the increased competition for funds, our field’s best hope for maintaining our funding might be to provide hard evidence of our students’ success.

Our program staff have come to believe that data analysis has the potential to:

- help us identify areas of strength and areas for improvement,
- provide information that can help to deliver services in the most effective and efficient manner,
- provide us and our community partners with information about the value of our programs,
- help us to make informed policy decisions,
- enable us to focus on results,
- help us operate in a way that attracts professional staff,
- enable us and our community partners to take pride in our accomplishments and to enhance our roles in the community,
- ensure that formally reported information is accurate, and
- help us to retain and increase funding.

We all face limited time and resources. At our program we have decided to incorporate data into our decision-making process to insure that the strategies we adopt will be effective. At this point, I would like to illustrate how our staff used data analysis to support a program improvement goal. Here are the major phases of that effort, with special mention of where and how we used data to support it.

**Our goal: To improve student attendance**

Just over a year ago, I met with our staff, students, and members of our community partnership to understand the issues that affect student attendance. Because research shows a direct correlation between the hours of instruction and learning gains, we decided to **increase the intensity of learning**. As a new, small ESOL program, we offered only six hours of instruction a week. More than 80% of our students work full-time, so they are unable to commit to additional structured hours of classroom instruction. We asked ourselves what strategies we could devise to support learning in and out of the classroom.
We checked the attendance data

The attendance issue was initially raised by instructors who expressed concerns about the disruption to class continuity caused by students who arrived late, left early, or had sporadic attendance. Discussions with students echoed these concerns. A review of our SMARTT data revealed some troubling trends; for one thing, many students were not using the six hours of class time we were providing for them. Students also reported frustration with empty seats, especially knowing how long it took for friends and family to get off the waiting list and into the program.

Further data analysis revealed that attendance problems fell into two specific categories. First, we had students who attended every night, but consistently arrived 20-30 minutes late. While 30 minutes might not seem like much, at the end of the year it is the equivalent of missing almost six full weeks of classes. Second, we had students with excellent attendance who suddenly left for extended periods of time, often due to illness or the need to return to their native countries for family emergencies.

In our staff meetings, we decided to develop an attendance policy (like I said, we’re a new program). At this point, our community partners offered a wealth of experience and a range of policy options to consider. From all this information we developed a draft policy and distributed it to all the stakeholders for feedback, which resulted in our adding a provision for leaves of absence.

We then met with each class to review the new policy. Students in our program were already accustomed to signing a Student Learning Agreement, stating that they agree to follow the policies and procedures in the Student Handbook. When we added the new Attendance Policy to the handbook, we were careful to allow ample opportunity for student discussion and questions before asking them to sign the agreement.

We developed several other means of providing more “intensity” without adding class time:

- We implemented a sign-in/sign-out procedure for when students arrive late or leave early. (The number of students arriving late and leaving early immediately decreased).
- We developed a lending library so students could take books home for additional practice.
- The Stoughton Public Library, a partner, began offering ESOL Book Discussion and Conversation Groups. They also offered to house the lending library during the summer when classes are not in session, so students could continue to have access to materials.
- The LVA-Stoughton began to provide tutors for some of our lower level students. (Some of our advanced students volunteer as tutors for the LVA as well, giving them additional English practice.)

We looked at the data again

Another analysis of our attendance data revealed a huge falloff in attendance during the December-January holidays. In the face of this reality, we revised our program schedule to include a longer holiday break in December, which also provides more time for our staff to plan classes and regroup.
We monitored our results over time

After we implemented all the attendance policy changes, we began to monitor the results on a monthly basis. We used the SMARTT attendance reports to identify students whose attendance was not satisfactory. To address that problem, we developed a Monthly Attendance Report for the counselor, who now meets with these students to provide strategies and support where possible. Most students who receive a verbal warning improve their attendance, which is easily tracked by comparing monthly attendance reports.

After the first year, the data revealed that our attendance rate had increased by 12%! We also began to wonder if, in fact, the increased attendance had generated a proportionate increase in learning gains. Thanks to SMARTT, COGNOS, and other data reports, we were able to demonstrate that there is indeed a direct correlation between increased hours and increased student gains. It's much more rewarding to know that relationship exists than to suspect that it exists.

We were also hit with a big surprise. While our attendance had increased, our retention had decreased. Although the numbers were not significant enough to cause panic, they revealed a need to address student retention. Once again, we gathered exit interview data from students to understand what prevented them from staying in the program. We looked at trends to understand under what conditions students tended to leave. As a result, this year we implemented a new intake and orientation procedure. While our year-to-date retention is higher, the true impact can only be measured at the end of the year when we look at the “big picture” relationship among attendance, retention, learning gains, and student progress towards goals.

We are also beginning to improve strategies for transitioning students out of the program. Sometimes, student retention is not a good thing! Some students never feel ready to leave, but now when we arrive at a point where we can no longer serve their needs, we have constructive ways to encourage them to take the next step.

We've gotten into the data habit

Data monitoring is now a regular part of our work life and informs virtually all of program decisions. For example, at staff meetings we distribute class attendance rates to our instructors, with comparisons to the prior fiscal year and to state averages. When we discovered that one of our instructors consistently maintains attendance at around 90%, we started to look at what we all can learn from her!

Basically, we look at our data in two ways: we compare our averages to state averages, and we compare our own data across fiscal years. While state averages are interesting to see, we tend to be more focused on continuous improvement. Looking at data from year to year helps us better understand the impact of our current plan and discover new areas to consider for improvement.

When gathering data, we often find it helpful to substitute the word “data” with the word “information.” The data available in SMARTT and COGNOS is invaluable, but sometimes it is overwhelming anecdotal information that informs our planning. A recent example of this is related to our process for intaking students from the waiting list.
I asked my staff why it was taking so long to fill slots in our beginner ESOL class, since we have over 150 students at this level on our waiting list. I learned that when these students were called, they typically hung up because they didn’t understand us or thought we were telemarketers! To buy some time, we asked current students to translate for us while we looked for a more permanent solution to this long-term problem.

After several meetings, we decided to create a “We would like you to begin class” post card with our logo and phone number on it. During initial registration/assessment, students are asked to fill in their mailing address on the post card. We explain that when we have an opening, we will mail them the post card with the date for them to begin. When the post card arrives, the student immediately recognizes it and makes the connection with our program. This has resulted in a much more efficient, equitable way to enroll students from the waiting list. While anecdotal data drove this process, the measure of its success will come from hard data. We will document the length of time it takes to enroll new students and determine the number of students enrolled per contact this year compared with last year.

We’ve learned that to use data consistently and effectively we’ve had to “institutionalize” its use. To do that we’ve had to put the following steps into practice:

- **Plan for data analysis.** We’ve learned that good data analysis cannot just happen episodically. We’ve had to set up meeting schedules, choose participants and include them in the process, find meeting space, and prepare copies of data reports.

- **Identify data leaders.** For most of us, understanding data is an acquired skill. We’ve identified people in our program who are skilled at putting data into context and understanding what it is trying to tell us—and let them take the lead!

- **Celebrate success.** If the data points to success, don’t forget to take pride in it!

---

*Luanne Teller is the Director of the Stoughton Adult Basic Education Program, a collaborative partnership among Massasoit Community College, the Stoughton Public Schools, and the Town of Stoughton, funded by a grant through the Massachusetts Department of Education.*
Eight years ago, I was the Senior Psychometrician for the GED Testing Service. My job was to ensure the technical quality of the Tests of General Educational Development (GED), which included making sure the score conversion tables were accurate, making sure the test items were clear and that they were testing what they were supposed to, and conducting research on the psychometric quality of the tests (i.e., studies of score reliability and validity; see Zenisky et al. in this issue). In working with educators throughout the United States to develop and review GED Test items, I fell in love with the adult education community. I became well aware of the dedication of ABE instructors and staff, as well as the amazing success stories of the millions of ABE students across the country.

When I left the GED Testing Service in Washington, DC to come to the University of Massachusetts at Boston in 1995, I wondered whether I would again have the opportunity to work with the ABE community. Thanks to Bob Bickerton and his staff at Adult and Community Learning Services (ACLS) of the Massachusetts Department of Education, I am proud to say that ABE is once again a major part of my life. And this time I have an army of psychometric professionals and students to help out, among the most talented testing professionals with whom I have ever worked. In my new role collaborating with ACLS, I have also discovered a new set of colleagues who go by the strange acronym SABES. These SABES folk are also a pleasure to work with, and are dedicated to improving education and assessment in ABE. In the remainder of this article, I will describe exciting events happening right now that stem from collaborations among UMass, ACLS, and SABES.

In January 2003, ACLS contracted with the UMass Center for Educational Assessment to help improve the assessment of ABE students in Massachusetts and to assist with their ongoing refinements of the processes used to evaluate and monitor all ABE programs in Massachusetts. Since that time we have written more than a dozen reports for ACLS, worked with teams of educators across the state to design new assessments in math and reading, developed and validated new prompts for the REEP writing assessment, and provided a comprehensive set of recommendations to ACLS for enhancing their program monitoring processes. Brief descriptions of three of our major activities follow.

Stephen Sireci offers an upbeat overview of the projects he and his colleagues at UMass Amherst are working on with MassDOE, SABES, and adult basic education practitioners.
Take out your Number 2 pencils: new ABE assessments are coming!

ACLS and SABES have worked hard over the past several years to come up with ways to meet the federal government’s requirements for the demonstration of the effectiveness of ABE programs. Presently, the U.S. Department of Education requires ABE programs to use test scores as one means of demonstrating that students are learning. ACLS and SABES convened a group of Massachusetts ABE educators called the Performance Accountability Working Group (PAWG) to review currently available tests that were suitable for adult learners in Massachusetts. The final report produced by the PAWG is available at www.sabes.org/resources/pawgfinal.pdf. In that report, the PAWG concluded that currently available tests were insufficient for the various needs of ABE students and programs in Massachusetts. They recommended a set of tests, including the TABE, BEST, and REEP, to be used on an interim basis until ACLS could develop new assessments targeted to the recently developed Curriculum Frameworks.

The development of new assessments for ABE students in Massachusetts is one of the key activities we are working on. Our vision is to mobilize ABE instructors and staff across Massachusetts to help us develop these tests. Our initial test development efforts are in the areas of math and reading, and recently we worked with two groups of ABE educators to decide what these tests should look like. One group helped us develop specifications for the math tests; the other group helped us develop specifications for the reading tests. Our next steps are to hold several item-writing workshops for ABE instructors across the state and ask them to write items for us. Thus, ABE instructors in Massachusetts will be the ones who develop the forthcoming tests.

Will our collaborative efforts produce tests that ABE students love to take? Well, probably not love to take; however, we are confident that the tests we are developing will be similar to what students are learning in their classes and will be appropriate for measuring their knowledge and skills. We are also confident that these item-writing workshops will provide valuable professional development for ABE educators. We plan to hold 5-10 workshops over the next year. Check the SABES Website at www.sabes.org periodically for announcements.

Making the REEP deep

Many students throughout Massachusetts strive to improve their writing. Many of these students write in languages other than English, but are taking classes to improve their writing in English. ACLS uses the REEP writing test, developed by the Arlington Education & Employment Program, to measure how much students’ writing improves after receiving instruction in ABE classes.

A key feature of the REEP (or any writing test) is the prompt, which is the topic to which students are asked to respond. An example of a prompt is, “Write a letter to someone about your most recent vacation.” The prompt on a writing test gives the students something to write about and allows a plan to be developed for scoring the essays written in response to that prompt. A year ago, there were only two prompts associated with the REEP.
Thus, with respect to prompts, the REEP was not very “deep.” Students who needed to take the REEP more than twice had to respond to the same prompts over and over again. ACLS asked us to develop new prompts for this test. Because these new prompts would also be used to measure students’ improvement in writing, they needed to be equivalent to the two prompts that were currently in use. After all, if we developed a new prompt that was harder to respond to, students’ newer essays might appear to be worse than their earlier essays.

I’m pleased to report that last spring we pilot-tested four prompts and one was selected for the pool of REEP prompts, expanding it by 50%. During the fall of 2003, we pilot-tested nine new prompts and four of them were approved for addition to the REEP prompt pool. In just one year, the number of REEP prompts expanded from two to seven. We were able to accomplish this goal by calling upon Massachusetts ABE teachers to send us ideas for prompts and administering experimental prompts to their students. ABE students also helped us by writing essays to the experimental prompts. Finally, we used SABES’s network of certified REEP scorers to score the experimental essays. The new prompts were selected after a comprehensive set of statistical and qualitative analyses that led us to conclude they are comparable to the two original prompts with respect to difficulty and scorability. The technical details regarding the prompt tryout and selection procedures are available in two reports we prepared for ACLS.

Monitoring program monitoring

A third project we are working on with ACLS is improving the monitoring of ABE programs throughout the state. ACLS is required to monitor all ABE programs to see whether they are doing a good job in accomplishing their goals and to report program evaluation information back to the federal government as part of the National Reporting System. Over the past year, we followed ACLS staff on several occasions when they gathered information on program quality. We also conducted a survey of ACLS staff members who perform program monitoring and surveyed programs that had recently been monitored. Finally, we took a close look at the instrument used to record program-monitoring data. Using the information we gathered from our observations of program monitoring and the survey data, we made several suggestions for revising the Program Monitoring Instrument. Presently, we are working with ACLS on revising the instrument to make it more efficient.

Introducing UMass

The above descriptions are just brief glimpses of the activities we are working on with ACLS and SABES. At the beginning of this article, I wrote a lot about myself. Before closing, I would like to write a few words about my terrific colleagues at UMass who are also working to improve assessment and evaluation in ABE programs. There are two senior staff members associated with this project: April Zenisky and Mercedes Valle. Both April and Mercedes are experienced in test development and statistics and are working tirelessly...
on the project. There are also several graduate students who are working on the project, including Peter Baldwin, Rob Keller, Drey Martone, and Shuhong Li. In addition, Professors Ronald Hambleton, Lisa Keller, and James Royer are contributing to the project. So, when I mentioned an army of psychometric professionals and students, I was not that far off. We all hope to meet and interact with many of you over the coming months. If you would like to learn more about us, please visit our web site at www.umass.edu/remp.

Stephen G. Sireci is Associate Professor in the Research and Evaluation Methods Program and Co-Director of the Center for Educational Assessment in the School of Education at the University of Massachusetts Amherst. Before UMass, he was Senior Psychometrician at the GED Testing Service, Psychometrician for the Uniform CPA Exam, and Research Supervisor of Testing for the Newark, NJ Board of Education. He is known for his research in evaluating test fairness, particularly issues related to content validity, test bias, cross-lingual assessment, standard setting, and sensitivity review.