spring 2002

adventures in assessment

Learner-centered approaches to assessment and evaluation in adult literacy

volume fourteen

Editor: *Marie Cora*

Design: *Marina Blanter*

SABES

SABES is the System for Adult Basic Education Support, a comprehensive training and technical assistance initiative for adult literacy educators and programs. Its goal is to strengthen and enhance literacy services and thus to enable adult learners to attain literacy skills.

SABES accomplishes this goal through staff and program development workshops, consultation, mini-courses, mentoring and peer coaching, and other training activities provided by five Regional Support Centers located at community colleges throughout Massachusetts. SABES also offers a 15-hour Orientation that introduces new staff to adult education theory and practice and enables them to build support networks. Visit us at our website: www.sabes.org

SABES also maintains an adult literacy Clearinghouse to collect, evaluate, and disseminate ABE materials, curricula, methodologies, and program models, and encourages the development and use of practitioner and learner-generated materials. Each of the five SABES Regional Support Centers similarly offers program support and a lending library. SABES maintains an Adult Literacy Hotline, a statewide referral service which responds to calls from new learners and volunteers. The Hotline number is 1-800-447-8844.

The SABES Central Resource Center, a program of World Education, publishes a statewide quarterly newsletter, *"Field Notes"* (formerly "Bright Ideas"), and journals on topics of interest to adult literacy professionals, such as this volume of *"Adventures in Assessment."*

The first three volumes of *"Adventures in Assessment"* present a comprehensive view of the state of practice in Massachusetts through articles written by adult literacy practitioners. Volume 1, Getting Started, includes start-up and intake activities; Volume 2, Ongoing, shares tools for ongoing assessment as part of the learning process; Volume 3, Looking Back, Starting Again, focuses on tools and procedures used at the end of a cycle or term, including self, class, and group evaluation by both teachers and learners. Volume 4 covered a range of interests, and Volume 5, The Tale of the Tools is dedicated to reflecting on Component 3 tools of alternative assessment. Volume 6, Responding to the Dream Conference, is dedicated to responses to Volumes 1-5. Volume 7, The Partnership Project, highlighted writings from a mentoring project for practitioners interested in learning about participatory assessment. Volumes 8-12 cover a range of topics, including education reform, workplace education, and learner involvement in assessment. Volume 13 focuses on issues related to building systems of accountability. Volume 14, *Examining Performance*, presents a range of articles that focus on performance-based assessment.

We'd like to see your contribution. If you would like to submit an article for our next issue, contact Editor Marie Cora.

Opinions expressed in *"Adventures in Assessment"* are those of the authors and not necessarily the opinions of SABES or its funders.

Permission is granted to reproduce portions of this journal; we request, however, appropriate credit be given to the author and to *"Adventures in Assessment."*

"Adventures in Assessment" is free to DOE-funded Massachusetts programs; out-of-state requests will be charged a nominal fee. Please write to, or call:

Marie T. Cora, World Education,
44 Farnsworth St., Boston, MA 02210-1211
617-482-9485 • mcora@worlded.org

Introduction. Volume 14: Examining Performance

BY MARIE CORA

Our last issue of Adventures in Assessment, Volume 13, examined some of the challenges the ABE field faces in meeting state and federal demands of accountability. Writers from that volume consistently noted the difficulty of capturing students' performance while at the same time striving to meet the reporting demands of funders. They noted that instructional purposes and administrative purposes of assessment often require very different approaches, tools, and documentation. In fact, efforts across the country are now focusing on trying to align these purposes.

Volume 14 continues to look at issues of accountability, but through the lens of capturing performance without the use of traditional tests. Practitioners take on questions including:

- How do we determine which assessment tools are appropriate for which purposes?
- Can we utilize performance-based assessment as a system of accountability?
- How can we capture students' knowledge *and* application of skills?
- How can we involve adult students in this journey?

These are only a few of the many questions we must answer together as a field if we are to build a strong and healthy Adult Basic Education system across our nation.

You will learn, as I did while putting together this volume, that there are many places where performance is in fact being examined in non-traditional ways. I also noted that in many of these places adult students are playing central roles of leadership. In this spirit, Volume 14 looks at non-traditional assessment in the classroom (where one might naturally expect to see more performance-based assessment), the program, and across programs, and in several instances, specifically highlights the roles that the adult students play.

The first article by **Linda Suskie**, reprinted from the American Association for Higher Education Bulletin, was written with a different population in mind, but the points she raises around fair assessment practices pertains to the ABE field as well and sets the stage for the ideas touched upon in the rest of this volume. Indeed, Suskie's Seven Steps to Fair Assessment can and should be applied to all stages of learning—from early childhood to adult education.

Two articles describe and critique oral assessments from two different ESOL programs in Massachusetts. **JoAnne Hartel** and **Mina Reddy** write about the OPT (Oral Proficiency Test) that was developed in their program over the past $1^1/_2$ years; and **Betty Stone** and **Vicki Halal** write about the BROVI, which they began working on in the Fall of 2000. These tools have both similarities and differences,

and I expect the reader will find the descriptions thought-provoking.

Inaam Mansoor and **Suzanne Grant** of the Arlington Education and Employment Program (REEP) in Arlington, VA contribute their writing rubric to assess ESL student performance. They also describe the process of developing and field-testing this tool with the guidance of the What Works Literacy Partnership (WWLP).

Tricia Donovan, writing about performance assessment in math, carefully outlines for us that the most important part of examining performance is how it reveals a person's understanding of a problem or task. She discusses some examples of tasks that help us examine both students' knowledge and application of skills.

Mary DuBois describes her work with the Student Action Health Team in Southeastern Massachusetts. This model for student learning and leadership development is highly participatory, and it incorporates a variety of performance-based assessments including conducting needs assessments, carrying out research, and delivering information in accessible ways to other adult students.

Early in 2001, the International Language Institute of Massachusetts (ILI) launched a program-wide effort to develop an approach to assessment which would be consistent with their learner-centered teaching philosophy. **Kermit Dunkelberg's** article describes a process which involved their students in the research and critique of various assessment methods and tools.

Finally, **Pat Mew** and **Paul Hyry** write about the Western Massachusetts Assessment Study Group, which has been meeting since January 2001. This endeavor brings together a collection of adult education programs from that region of the state to examine and critique various assessment methodologies, much like Kermit's group at ILI. Pat and Paul's group, however, is examining processes and tools that could be helpful across *all* the participating programs. Again, both the content and process focus on performance assessment, but in this case, the students are practitioners.

In this age of education reform, adult education might be farther behind than our K12 counterpart, but our approach to education has always considered the ways in which less traditional approaches to learning and assessing might better serve students and teachers. We are innovative and determined. We are in a position to develop new contributions to our field that could forever change the way our work is done. Practitioners and adult students, working side by side, are already improving their classrooms and programs.

Several articles in this volume refer to the Massachusetts Curriculum Frameworks. These can be found at: http://www.doe.mass.edu/acls/frameworks/

Your thoughts and ideas are welcomed and encouraged. If you would like to submit an article or have comments, please feel free to contact me at mcora@worlded.org

Table of Contents

Introduction: Volume 14: Examining Performance ..... **page 1**

Marie Cora, Editor, Adventures in Assessment, SABES Central Resource Center, World Education, Boston, MA

Fair Assessment Practices: Giving Students Equitable Opportunities to Demonstrate Learning ..... **page 5**

Linda Suskie, the American Association of Higher Education.

Assessing Oral Communication at the Community Learning Center Development of the OPT (Oral Proficiency Test) ..... **page 11**

JoAnne Hartel and Mina Reddy, Community Learning Center, Cambridge, MA.

So What IS a BROVI, Anyway? And how can it change your (assessing) life? ..... **page 19**

Betty Stone and Vicki Halal, SCALE, Somerville, MA.

A Writing Rubric to Assess ESL Student Performance ..... **page 33**

Inaam Mansoor and Suzanne Grant, REEP, Arlington, VA.

Illuminating Understanding: Performance Assessment in Mathematics ..... **page 39**

Tricia Donovan, TERC, Cambridge, MA.

Student Health Education Teams in Action ..... **page 43**

Mary Dubois, New Bedford Public Schools/Division of Adult and Continuing Education, New Bedford, MA.

Involving Learners in Assessment Research ..... **page 49**

Kermit Dunkelberg, International Language Institute, Northampton, MA.

WMass Assessment Group–Tackling the Sticky Issues ..... **page 67**

Patricia Mew, SABES West, Holyoke, MA, and Paul Hyry, Community Education Project, Holyoke, MA.

Fair Assessment Practices: Giving Students Equitable Opportunities to Demonstrate Learning

This article is reprinted by permission from the May 2000 AAHE Bulletin (American Association for Higher Education).

BY LINDA SUSKIE

I am a terrible bowler. On a good night, I break 100. (For those of you who have never bowled, the highest possible score is 300 and a score below 100 is plain awful.) This is a source of great frustration for me. I've taken a bowling class, so I know how I'm supposed to stand and move, hold the ball and release it. Yet despite my best efforts to make my arms and legs move the same way everytime, the ball only rarely rolls where it's supposed to. Why, I wonder, can't my mind make my body perform the way I want it to, every time I roll the ball?

If we can't always control our bodily movements, we certainly can't always control what goes on in our heads. Sometimes we write and speak brilliantly; sometimes we're at a loss for words. Sometimes we have great ideas; sometimes we seem in a mental rut. Is it any wonder, then, that assessment—finding out what our students have learned—is such a challenge? Because of fluctuations in what's going on inside our heads, we inconsistently and imperfectly tell our students what we want them to do. Because of similar fluctuations in what's going on in our students' heads, coupled with cultural differences and the challenges of interpersonal communication, they can't always fully interpret what we've told them as we intended them to, and they can't always accurately communicate to us what they know. We receive their work, but because of the same factors, we can't always interpret accurately what they've given us.

A colleague who's a chemist throws up his hands at all this. Having obtained controlled results in a laboratory, he finds assessment so full of imprecision that, he says, we can never have confidence in our findings. But to me this is what makes assessment so fascinating. The answers aren't there in black and white; we have, instead, a puzzle. We gather clues here and there, and from them try to infer an answer to one of the most important questions that educators face: What have our students truly learned?

Seven Steps to Fair Assessment

If we are to draw reasonably good conclusions about what our students have learned, it is imperative that we make our assessments—and our uses of the results—as fair as possible for as many students as possible. A fair assessment is one in which students are given equitable opportunities to demonstrate what they know (Lam, 1995). Does this mean that all students should be treated exactly the same? No! Equitable assessment means that students are assessed using methods and procedures most appropriate to them. These may vary from one student to the next, depending on the student's prior knowledge, cultural experience, and cognitive style. Creating

"Equitable assessment means that students are assessed using methods and procedures most appropriate to them."

custom-tailored assessments for each student is, of course, largely impractical, but nevertheless there are steps we can take to make our assessment methods as fair as possible.

1. *Have clearly stated learning outcomes* and share them with your students, so they know what you expect from them. Help them understand what your most important goals are. Give them a list of the concepts and skills to be covered on the midterm and the rubric you will use to assess their research project.

2. *Match your assessment to what you teach* and vice versa. If you expect your students to demonstrate good writing skills, don't assume that they've entered your course or program with those skills already developed. Explain how you define good writing, and help students develop their skills.

3. *Use many different measures and many different* ***kinds*** *of measures*. One of the most troubling trends in education today is the increased use of a high-stakes assessment—often a standardized multiple-choice test—as the sole or primary factor in a significant decision, such as passing a course, graduating, or becoming certified. Given all we know about the inaccuracies of any assessment, how can we say with confidence that someone scoring, say, a 90 is competent and someone scoring an 89 is not? An assessment score should not dictate decisions to us; we should make them, based on our professional judgement as educators, after taking into consideration information from a broad variety of assessments.

   Using "many different measures" doesn't mean giving your students eight multiple-choice tests instead of just a midterm and final. We know now that students learn and demonstrate their learning in many different ways. Some learn best by reading and writing, others through collaboration with peers, others through listening, creating a schema or design, or hands-on practice. There is evidence that learning styles may vary by culture (McIntyre, 1996), as different ways of thinking are valued in different cultures (Gonzalez, 1996). Because all assessments favor some learning styles over others, it's important to give students a variety of ways to demonstrate what they've learned.

4. *Help students learn how to do the assessment task*. My assignments for student projects can run three single-spaced pages, and I also distribute copies of good projects from past classes. This may seem like overkill, but the quality of my students' work is far higher than when I provided less support.

   Students with poor test-taking skills may need your help in preparing for a high-stakes examination; low achievers and those from disadvantaged backgrounds are particularly likely to benefit (Scruggs & Mastropieri, 1995). Performance-based assessments are not necessarily more equitable than tests; disadvantaged students are like-

ly to have been taught through rote memorization, drill, and practice (Badger, 1999). Computer-based assessments, meanwhile, penalize students from schools without an adequate technology infrastructure (Russell & Haney, 2000). The lesson is clear: No matter what kind of assessment you are planning, at least some of your students will need your help in learning the skills needed to succeed.

5. *Engage and encourage your students.* The performance of "field-dependent" students, those who tend to think more holistically than analytically, is greatly influenced by faculty expressions of confidence in their ability (Anderson, 1988). Positive contact with faculty may help students of non-European cultures, in particular, achieve their full potential (Fleming, 1998).

6. *Interpret assessment results appropriately.* There are several approaches to interpreting assessment results; choose those most appropriate for the decision you will be making. One common approach is to compare students against their peers. While this may be an appropriate frame of reference for choosing students for a football team or an honor society, there's often little justification for, say, denying an A to a student solely because 11 percent of the class did better. Often it's more appropriate to base a judgement on a standard: Did the student present compelling evidence? summarize accurately? make justifiable inferences? This standards-based approach is particularly appropriate when the student must meet certain criteria in order to progress to the next course or be certified.

   If the course or program is for enrichment and not part of a sequence, it may be appropriate to consider growth as well. Does the student who once hated medieval art now love it, even though she can't always remember names and dates? Does another student, once incapable of writing a coherent argument, now do so passably, even if his performance is not yet up to your usual standards?

7. *Evaluate the outcomes of your assessments.* If your students don't do well on a particular assessment, ask them why. Sometimes your question or prompt isn't clear; sometimes you may find that you simply didn't teach a concept well. Revise your assessment tools, your pedagogy, or both, and your assessments are bound to be fairer the next time that you use them.

Spreading the Word

Much of this thinking has been with us for decades, yet it is still not being implemented by many faculty and administrators at many institutions. Our challenge, then, is to make the fair and appropriate use of assessments ubiquitous. What can we do to achieve this end?

- *Help other higher education professionals learn about fair assessment practices.* Some doctoral programs offer future faculty studies in pedagogy and assessment; others do not.

Encourage your institution to offer professional development opportunities to those faculty and administrators who have not had the opportunity to study teaching, learning, and assessment methods.

- *Encourage disciplinary and other professional organizations to adopt fair assessment practice statements.* A number of organizations have already adopted such statements, which can be used as models. Models include statements adopted by the Center for Academic Integrity (McCabe & Pavela, 1997); the Conference on College Composition and Communication (1995); the Joint Committee on Standards for Educational Evaluation (1994); the Joint Committee on Testing Practices (1988); the National Council on Measurement in Education (1995); and the first National Symposium on Equity and Educational Testing and Assessment (Linn, 1999); as well as AAHE (1996). (See *Assessment Policies*, below).

- *Speak out when you see unfair assessment practices.* Call for the validation of assessment tools, particularly those used for high-stakes decisions. Advise sponsors of assessment practices that violate professional standards, and offer to work with them to improve their practices.

- *Help improve our assessment methods.* Sponsor and participate in research that helps create fairer assessment tools and validate existing ones. Collaborate with assessment sponsors to help them improve their assessment tools and practices. Help develop feasible alternatives to high-stakes tests.

- Help find ways to share what we already know. Through research, we have already discovered a great deal about how to help students learn and how to assess them optimally. With most of us too busy to read all that's out there, our challenge is finding effective ways to disseminate what has been learned and put research into practice.

As we continue our search for fairness in assessment, we may well be embarking on the most exhilarating stage of our journey. New tools such as rubrics, computer simulations, electronic portfolios, and Richard Haswell's minimal marking system (1983) are giving us exciting, feasible alternatives to traditional paper-and-pencil tests. The individually custom-tailored assessments that seem hopelessly impractical now may soon become a reality. In a generation—maybe less—it's possible that we will see a true revolution in how we assess student learning, with assessments that are fairer for all . . . but only if we all work toward making that possible.

When this article was written, Linda Suskie was director of AAHE's Assessment Forum, and assistant to the president for special projects at Millersville University of Pennsylvania.

Assessment Policies

Several organizations have developed statements that include references to fair assessment practices. Some are available online:

Code of Fair Testing Practices in Education by the Joint Committee on Testing Practices, National Council on Measurement in Education
ericae.net/code.txt

Code of Professional Responsibilities in Educational Measurement by the National Council on Measurement in Education
www.natd.org/Code_of_Professional_Responsibilities.html

Leadership Statement of Nine Principles on Equity in Educational Testing and Assessment by the first National Symposium on Equity and Educational Testing, North Central Regional Educational Laboratory
www.ncrel.org/sdrs/areas/issues/content/cntareas/math/ma1newst.htm

Nine Principles of Good Practice for Assessing Student Learning by the American Association for Higher Education
www.aahe.org/principl.htm

Writing Assessment: A Position Statement by the Conference on College Composition and Communication
www.ncte.org/ccc/12/sub/state6.html

References

American Association for Higher Education. (1996, July 25). *Nine principles of good practice for assessing student learning* [Online]. Available: *http://www.aahe.org/assessment/principl.htm*

Anderson, J. A. (1988). Cognitive styles and multicultural populations. *Journal Teacher Education*, 24(1), 2-9.

Badger, E. (1999). Finding one's voice: A model for more equitable assessment. In A. L.

Nettles & M. T. Nettles (Eds.), *Measuring up: Challenges minorities face in educational assessment* (pp. 53-69). Boston: Kluwer.

Conference on College Composition and Communication. (1995). *Writing assessment: A position statement* [Online]. Available: *http://www.ncte.org/ccc/12/sub/state6.html*

Fleming, J. (1998). Correlates of the SAT in minority engineering students: An exploratory study. *Journal of Higher Education*, 69, 89-108.

Gonzalez, V. (1996). Do you believe in intelligence? Sociocultural dimensions of intelligence assessment in majority and minority students. *Educational Horizons*, 75, 45-52.

Haswell, R. H. (1983). Minimal marking. *College English*, 45, 600-604.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.

Joint Committee on Testing Practices. (1988). *Code of fair testing practices in education* [Online]. Available: http://ericae.net/code.txt

Lam, T. C. M. (1995). *Fairness in performance assessment: ERIC digest* [Online]. Available: http://ericae.net/db/edo/ED 391982.htm (ERIC Document Reproduction Service No. ED 391 982)

Linn, R. L. (1999). Validity standards and principles on equity in educational testing and assessment. In A. L. Nettles & M. T. Nettles, (Eds.), *Measuring up: Challenges minorities face in educational assessment* (pp. 13-31). Boston: Kluwer.

McCabe, D. L., & Pavela, G. (1997, December). The principled pursuit of academic integrity. *AAHE Bulletin*, 50(4), 11-12.

McIntyre, T. (1996). Does the way we teach create behavior disorders in culturally different students? *Education and Treatment of Children*, 19, 354-70.

National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement* [Online]. Available: http://www.natd.org/Code_of_Professional_Responsibilities.html

Russell, M., & Haney, T. (2000, March 28). Bridging the gap between testing and technology in schools. *Education Policy Analysis Archives* [Online serial], 8(19). Available: http://epaa.asu.edu/epaa/v8n19.html

Scruggs, T. E., & Mastropieri, M. A. (1995). *Teaching test-taking skills: Helping students show what they know*. Cambridge MA: Brookline Books.

Assessing Oral Communication at the Community Learning Center
Development of the OPT (Oral Proficiency Test)

BY JOANNE HARTEL
AND MINA REDDY

Why Create a New Assessment?

The impetus for developing a new form of oral assessment at the Community Learning Center (CLC), a large adult education center in Cambridge, Massachusetts, came from new federal and state requirements that began in the summer of 2000. Before then, we had been using a standard in-house procedure to assess speaking, listening, reading, and writing on intake. We also had curricula for each level and criteria for moving students up to the next level. We had developed a writing sample administered under standard conditions and scored using a rubric. At the end of each semester teachers held individual conferences with students to discuss their progress. However, there was no program-wide oral assessment. Teachers created their own in-class processes to assess speaking and listening or, more often, based their evaluations entirely on classroom observation. SPL (student performance levels) levels, required by the state for reporting purposes, were assigned based on the classes students were placed in.

Given the increased emphasis on accountability and the need for standardized assessment procedures, we realized that this would no longer be sufficient. We considered using the BEST test, the most common off-the-shelf, standardized oral assessment. We liked the idea of a picture-based test that could be administered in a conversational, informal way. However, the BEST test was not a good match with our ESOL core curriculum, which had recently been revised based on the Massachusetts Curriculum Frameworks. Since we did not find any existing tests that matched our curriculum well, we decided to develop our own assessment of students' oral communication. The assessment needed to match our curriculum, provide information for placement and advancement, yield an SPL level for accountability purposes, and work for beginning, intermediate, and advanced students. We also decided to create alternate forms of the assessment so that it could be given up to three times a year, and to design something that would be easy to administer and score. Finally, we wanted the actual administration of the assessment to take no more than 10 minutes because we planned to administer it individually and we did not have the resources to give a longer test to our entire ESOL population. We realized that satisfying all of these criteria in one assessment would be no easy task.

"We liked the idea of a picture-based test that could be administered in a conversational, informal way."

Description of the Assessment

Each form of the Oral Proficiency Test (OPT) consists of a line drawing with six questions. Three questions involve describing what is in the picture. The last

three questions pertain to the student's own experience. One question is intended to prompt a past tense answer, and another a response with a modal. The test assesses comprehension of the questions, the use of certain grammar forms, vocabulary, syntax, fluency, and pronunciation. (See the sample picture and questions.)

The OPT is administered by a trained tester (a teacher or counselor in the program) who is not the student's own teacher. The tester begins by introducing him or herself and meeting the student. He/she then says, "This is a very short test. It's for listening and speaking. It's only one measure of your progress in learning English. There are many things you and your teacher will talk about. I'm going to show you a picture and I'm going to ask you six questions about the picture. Please give me **big** answers. I'm going to write down the things you tell me so that I can remember what you said." He/she then briefly introduces the picture and proceeds to the questions. The questions can be repeated once if the student wishes, but without changing the wording. As the student answers the questions, the tester writes down what the student says or takes some notes if the response is very fast and long. The tester also makes a symbol to indicate whether the question was repeated. Once the test is over, the tester says, "Thank you very much. It was a pleasure to talk to you," and adds some words of encouragement. The test is scored immediately based on the guidelines (see appendix). There is a range of scores for each answer, depending on the accuracy and completeness of the student's response. There are also holistic scores for pronunciation and fluency. The scores are totaled, and an SPL is assigned and entered on the Department of Education database for accountability purposes. The scored rubric with the tester's notes and/or transcription of the student's answers is given to the teacher to use when conferencing with the student. It is one among several factors to be considered when deciding whether to move a student into the next class. The others include classroom performance, homework, attendance, and the writing sample.

Designing the Assessment

We decided to base the assessment on a conversation about a picture with the aim of making the language as natural as possible. We started with six pictures drawn for us by Joann Wheeler, an artist and former Community Learning Center teacher under the direction of JoAnne Hartel, who also made up the first draft of the questions. Each picture was used for a different form of the test. JoAnne started with the CLC's ESOL curriculum, using topics and vocabulary from the beginning and intermediate levels. The questions were designed to elicit simple sentences.

During the summer of 2000, the new oral assessment was piloted with students in several CLC classes and with new students on intake. At the same time, the BEST test was given to students in two classes for comparison purposes. Beyond the beginning level, the BEST proved to be very unsatisfactory for our students. The scores did not seem to reflect oral ability, particularly with more educated students.

Our pilot worked well enough to reassure us that we were on the right track. We continued administering the test to incoming students in the fall, and the ESOL teachers and counselors gave feedback on it. JoAnne trained and worked with a team of eight teachers to administer the OPT to every student in ESOL levels 1 to 4 in January 2001, at the end of the semester. Training involved discussion of the scoring criteria and practice scoring to make sure the results were as reliable as possible.

After all students were tested in January, the testing team met again to revise the questions and scoring criteria. They chose the three pictures that worked best and asked for some modifications of the pictures (e.g. "Make the woman in the clinic look more pregnant").

JoAnne and Mina sat down with lists of all the students by class and looked at their OPT scores and their class levels based on the judgment of their teachers. We recalibrated the scoring so that these matched more closely and served to discriminate better between students at different levels. The original scoring seemed to work less well at the upper level, so we adjusted it accordingly.

Evaluation of the assessment

We feel confident that the results of the OPT are, in most cases, a true reflection of students' oral communication ability. The new assessment has a number of advantages. It is a standard procedure for all students, administered by a few trained testers, so the results are more comparable than those that would come from individual teachers each using their own methods. In a program with many part-time ESOL teachers who may not have had an opportunity to teach more than one level, as is the case in many ABE programs, making judgments can be difficult. This also helps students feel that there are clear criteria for advancement. The OPT is quick, and it yields a numerical score. Raw scores can indicate improvements within an SPL level. It is based on the grammar and content in the curriculum. Although it is a test, it feels close to a natural conversation and does not cause students to feel intimidated. This is particularly true for those who have been given the test more than once and are familiar with the process. There are three forms of the test available.

According to one CLC ESOL teacher, "It's useful to see how and how much the students can express with someone other than the teacher. Sometimes they can do more. It reminds us that our students need to communicate with other people in a different context. It's more realistic than the classroom."

One of the ESOL counselors said that the OPT is another tool that combined with everything else we use, gives us a clearer picture of the correct ESOL placement level. It gives a better idea of students' grammar skills and sentence structure. And after doing a second round of OPT testing with the same group, the counselor noticed an improvement in the area of conversation. Also, the intake process is more complete now. On some occasions, when it is difficult to make a correct class placement, the OPT has been a key factor in placing students in class.

"Although it is a test, it feels close to a natural conversation and does not cause students to feel intimidated."

However, like any point-in-time assessment, the results can vary depending on how the person is feeling that day. Some ESOL students made mistakes, not because of their English, but because they misunderstood the intent of the line drawing they were looking at. Photographs might help to solve this problem. Although it is short, it is time intensive because it has to be administered individually. It was not scientifically designed. We had some initial discussions about designing procedures for assessing the reliability of the OPT, e.g. administering two forms of the test to the same person and seeing how closely the scores matched (to see how comparable different forms were), taping the test and having two testers score it (to check inter-rater reliability), etc. However, before investing the time needed for these efforts, we have decided to wait for the Massachusetts Department of Education to make some decisions about assessment and accountability. We have also continued to make small revisions in the questions and are collecting data on student scores that will help us to evaluate the effectiveness of the procedure in the future.

The Community Learning Center is a large adult basic education center located in Cambridge, Massachusetts. It serves 1000-1200 students each year, over 60% of them in ESOL classes. Students come from between 60 and 80 countries. Most attend class 5 to 6 hours per week. The majority are working.

Funds for the development and initial administration of the OPT came from the City of Cambridge and the Massachusetts Department of Education.

JoAnne Hartel is a teacher and curriculum and staff development coordinator at the Community Learning Center. Until recently, Mina Reddy was the director of the CLC.

CLC ORAL PICTURE TEST SCORING

ORAL PICTURE TEST SCORE	SPL	ESOL LEVEL
0 – 6	3	ESOL 1
7 – 12	4	ESOL 2
13 – 16	5	ESOL 3
17 – 20	6	ESOL 4
21 – 24	7	ESOL 5

Scoring Criteria

0—Doesn't understand the question, or
Doesn't answer the question in English, or
Answer is not related to the question

1—Answers question, but uses isolated words or very short, simple phrases

2—Answers question in phrases and sentences with little or no control over basic grammar; may display some difficulty expressing ideas

3—Answers question in complete sentences; control over basic grammar is evident but not consistent; may display some hesitation in expressing ideas

4—Answers question completely; good control over basic grammar; can speak creatively, but may display some hesitation in expressing ideas

PICTURE TEST 2 **AT THE CLINIC** **DATE**

	0	1	2	3	4	Comments/Student Response
1. (Explain that this is a picture of a clinic waiting room These people are waiting to see the doctor. Point to the elderly man.) What's the matter with him?						
2. (Point to the receptionist.) What is she doing?						
3. (Point to the other people in the picture.) Tell me about the other people who are at the clinic.						
4. You have an appointment to see a doctor at the clinic. You come in, and you speak to the receptionist. What do you say to the receptionist?						
4. If I have a fever and a cough, what do you think I should do?						
5. In your country what did you do when you got sick?						

Student _____ Teacher _____ Tester _____

Raw Score _____ SPL Level _____

Pronunciation 1 2 3 4 Fluency 1 2 3 4

So What IS a BROVI, Anyway? And how can it change your (assessing) life?

BY BETTY SONE AND VICKI HALAL

Why did we develop our own assessments?

Through the years at SCALE (Somerville Center for Adult Learning Experiences), ESOL assessment had continually been a source of debate and concern. Frankly, the topic drove people crazy. What kinds of assessment were teachers using to support their "feelings" that students were ready to be promoted? Did the assessment results match teachers' intuitions? How similarly did different teachers rate the abilities of the same students? Why couldn't we agree? Staff often raised the issue of assessment in terms of documentation needed to support level change recommendations. We long sought easily administered and appropriate assessment tools that would more consistently measure all language areas (reading, writing, listening, speaking) across all program levels. More recently, this coincided with the state and national movement toward "reliable and valid" assessment, as well as the Massachusetts Department of Education (DOE) requirement to report learner progress according to Student Performance Levels (SPLs).

From instructor to instructor ongoing assessment style and content varied widely. While such a range of assessment strategies would not affect the appropriateness of particular assessments within a class, as a program we lacked consistency and the level recommendation process could sometimes become murky. Counselors were sometimes called upon to mediate lively testimonials between a teacher who wanted to promote "M____" and a second teacher who refused to accept her. Without a program-wide assessment tool, we could not easily come to a consensus on when learners were prepared to move ahead. Since our primary objective is to help students realize their fullest potential, we felt that perhaps we would serve them better by having at least one method of assessment that all staff would utilize. We hoped that would help us more clearly identify students' strengths and weaknesses over time and, therefore, keep better track of their needs as they proceeded through the program. A consistent assessment protocol would also make clear to the students the expectations of the program at each level.

We had at times considered adopting published assessment materials as well as possibly instituting a formalized portfolio system. The popular ESOL assessment tools were ill-suited to our population. Maintaining an elaborate portfolio assessment system for over 300 learners at two sites was not realistic for a primarily part-time staff. A mini-grant from the Adult Literacy Resource Institute (ALRI) gave us an opportunity to create our own assessment package. As the project flourished, we realized that the tools not only

"Without a program-wide assessment tool, we could not easily come to a consensus on when learners were prepared to move ahead."

"A consistent assessment protocol would also make clear to the students the expectations of the program at each level."

served SCALE's program well, but could be replicated in other programs, and might even provide MA DOE with an example of effective alternative assessment for ESOL.

Who was involved in developing the assessments?

SCALE was fortunate to have received an ALRI mini-grant for two of its part-time ESOL instructors, Laura Brooks and Vicki Halal, to coordinate research, development, and implementation of the initial assessment package. Program Administrators, Betty Stone and Ngaio Schiff, also contributed time and expertise to the project. Additionally, this small team involved the entire staff by soliciting their ideas and feedback through surveys, staff meetings, a pilot round of assessment, and trainings. Some members of the staff chose to dedicate a portion of their staff development hours to the project as well. The team also surveyed groups of students (one from each class) at the outset about their ideas on assessment. This comprehensive collaboration created a great deal of intellectual and practical momentum during the process, and insured that everyone was invested in the project and all voices were heard from the very beginning.

What kind of assessment did we develop?

SCALE's current assessment package consists of two tools, the Writing Sample that measures writing skills, and the BROVI, which has two forms to measure oral/aural skills, an individual speech and a role-play. What, you ask, does BROVI mean? It is, as you may have suspected, an acronym made from the names of the developers that takes the place of "the listening/speaking assessment," a phrase that tripped up all our efforts to exchange ideas about that developing tool.

Assessments are routinely administered program-wide three times per academic year to track students' progress through SCALE's internal levels as well as to satisfy the reporting requirements of the Massachusetts (DOE). The program designates two-week periods in October, January, and May for assessment, and accommodates students who enter SCALE classes at other times with special assessment arrangements.

Our Writing Sample consists of the following:

- "To the Student" Instructions
- Administration Instructions to teachers
- Master Writing Sample sheets for each topic: lined sheets headed by the topic or a picture (Teacher selects a single topic from the Master List for the class.)
- A scoring rubric for teachers

Typically, teachers set aside approximately 40 minutes of class time: 10 minutes for explanation of the purpose and instructions, and 30 minutes for the class to complete it. Once the Samples are collected, instructors score them outside of class according to the rubric and are compensated based on the number of samples per class.

The BROVI incorporates the option of an individual Speech or a paired Role Play activity. The teacher selects one of the options and administers it to the entire class on the designated assessment day(s). The components of the BROVI are

- To the Student' Instructions (see appendix)
- Administration Instructions to teachers
- Speech Topics (Master List includes choice of three per level; teacher selects one for all) (see appendix)
- Role Plays for literacy to high beginner levels (laminated photo cards with scenario descriptions printed on the reverse side)
- Role Plays for intermediate to advanced levels (laminated scenario cards)
- Audience listening activity worksheet (see appendix)
- A scoring rubric for teachers (see appendix)

The amount of class time needed to complete the BROVI will vary according to the option chosen as well as the class size. In both cases, the teacher reviews with the students the purpose of assessment and the instructions. The speech topic or role-play scenarios are distributed to learners who work for 10-15 minutes to prepare (speech topics are discussed in groups of 3-4; role-plays are prepared in pairs). As each person or pair bravely performs the speech or role-play without notes, the rest of the class are filling in their "Audience Listening Activity," preparing questions for those giving speeches or answering questions about the role plays, and the instructor is completing the rubric. All BROVI scoring is done during class time.

Both the Writing Sample and the BROVI are given raw scores from 0-76, which correlate to SPLs 0-8, the range of ability among SCALE ESOL students. The components of each rubric and their weights (Note the x2, x3, x4, x5) represent the relative importance of those aspects of language within our program.

What was the hardest part of the development process?

From the beginning, we were aware that we would face a number of challenges. First, we needed a set of *user-friendly tools for students and instructors to use during class time*. Most of our staff is part-time and limited financial resources for extra paid staff time mandated that the bulk of the assessment work take place within the framework of class hours. We succeeded in raising supplementary grant funds to provide the necessary training for all staff, but the core assessment responsibilities and ongoing feedback on our model fit within expected expenditures for teaching, meeting, and staff and program development time.

Next, we wanted to ensure that our *scoring system would reflect meaningful progress through internal program levels and be correlated to the SPL system*. As previously stated, we wanted to avoid the standardization that might limit the possibilities of student performance. We chose to develop our performance-based assess-

"All through the development process, we tried to keep our staff involved. With each step, we asked for feedback and suggestions for change."

ments so that we could offer students opportunities to use their English to their fullest abilities. The challenge in scoring was to be able to give credit (and points) for the complete range of proficiency levels that exist in our ESOL levels. In this way, we wanted the instruments to reflect the patterns of progress within our entire program. By creating a weighted system of scoring the various components of writing, speaking, and listening, the raw score range covers 0-76 and correlates with SPLs 0-8. Following the second full round of assessment (May 2001) we had a large enough number of raw scores to re-adjust the raw score/SPL correlation based on how actual students scored at each internal SCALE level.

All through the development process, we tried to keep our staff involved. With each step, we asked for feedback and suggestions for change. We needed their input to improve content and administration of the Writing Sample/BROVI. By being involved during the development process, we hoped the staff would feel more confident in utilizing the resultant tools. *Without the participation of the entire staff in both development and implementation, this assessment package would be compromised*. Of course, the more we asked, the more feedback we needed to incorporate. Initially, the pages of notes seemed daunting; however, once we began to sort through them and incorporate their suggestions, we found we appreciated the input even more. Staff involvement in the entire process made us feel more confident in the final product and helped avoid the feeling that the final tools would be an imposition on teachers or students.

Once the Writing Sample/BROVI was ready to be used, the issue arose of training and compensating staff fairly. Even though staff were familiar with the package through its development, the implementation approaches still varied from teacher to teacher. Additionally, scoring could be rather subjective, so there needed to be consensus in order to have "reliable" and "valid" assessment. SCALE offered four sessions of paid program development dedicated to staff training that allowed instructors, counselors, and administrators to discuss and fine tune the administration and scoring procedures involved in the assessment package. After the initial pilot of the Writing Sample/BROVI, we were able to verify the number of hours generally needed to score the Writing Sample, and, we developed a pay scale accordingly. Instructors are allotted a certain number of hours based on their class size and are paid for them at their regular hourly rate.

Facing and overcoming the many hurdles inherent in this project led us to develop what we feel is a user-friendly, meaningful, and fair assessment package.

Self-assessment of our assessment

We have been pleased and encouraged that both the BROVI and Writing Sample assessments have gotten high marks from the ESOL teaching staff at SCALE, as well as from the ESOL program administrators. Practitioners particularly like the following features:

Strengths:

The assessment tasks are performance-based and learner-centered. They are related to the learners' goals of communi-

cating more effectively in English and/or improving writing skills. Teachers report that students have fun preparing and performing the role-plays and learners enjoy hearing each others' "speeches" and asking follow-up questions. The BROVI and writing sample topic selections offer a reasonable degree of choice and allow students to display their language ability, though we continue to refine the master lists in response to teacher feedback. While the exact topics for the "official" assessments are considered "secure," teachers are encouraged to practice role-plays and sustained speaking activities as part of their usual classroom routine. The bottom line is that the assessment tasks themselves are representative of activities in an interactive ESOL classroom. These are not strange, threatening, or irrelevant tasks that suddenly invade the classroom; rather they are natural language learning activities that are easily integrated into curriculum design. The "To the Student" handouts keep the "test stress level" among students in check. Teachers are listening or reading for what students know, not what they don't.

Materials are well "packaged" and easy to use. Administration guidelines and directions are standardized, clear, and easily accessible. Assessment protocol, pay for related work (scoring writing samples), and timelines are unambiguous. This is particularly significant at SCALE where the ESOL teaching staff is primarily part-time. Special student handouts make an effort to demystify the assessment process to the students. We want the learners to know what we are asking them to do, why we do it several times each year, and what we expect of them. Assessment drawers contain classroom packets of Writing Sample master sheets, BROVI and Writing Sample rubrics, laminated beginning level role-play photo scenario cards, intermediate/advanced level role-play scenario cards, BROVI Speech Topics, and "To the Student" handouts to assist learners with understanding the purpose and expectations for each of the three assessments (BROVI speech, BROVI role-plays, and Writing Sample). January 2002 marked the fourth and the smoothest administration round of these assessments at SCALE. Teachers and students are beginning to take the process in stride.

We have achieved a uniformity and consistency of assessment conditions with these tools that had never before existed across the range of classes in our program. Though we needed to invest in a second round of intensive training in January 2002, to orient new teachers and reinforce scoring practices of veteran teachers, consistent scoring of BROVIs and writing samples is improving. Interrater reliability among assessors is key in a program such as ours, where 19 instructors teach and assess five core ESOL levels and three ESOL literacy levels, representing the range from SPL – to SPL 8.

As a program, we are beginning to witness predictable patterns of progress as we track learners through various classes. Assessment results for a sample student who has repeated ESOL 1 two times and then is promoted to ESOL 2, come from three different teachers in the three distinct classes. (ESOL 1, ESOL 1, ESOL 2) Raw scores and correlated SPLs over the student's career at SCALE show little improvement or sometimes some slide-back initially. Ultimately, however,

sufficient raw score (and SPL) increases indicate the student's readiness to advance to the next SCALE class level.

The rubrics are clear, specific, and easy to use. They have seen numerous iterations, always in response to teacher feedback, and always with the goal of facilitating the process of capturing learner performance in a fair, accurate, and streamlined fashion. Teachers have commented that using the rubrics has been helpful in sharpening their diagnostic skills in general. They are regularly reminded of the *objective* criteria the program uses to rate a learner's competence. Good attendance, cheerful attitude, and social connection to the class are not on the rubrics. While those may be characteristics of many of our successful learners, they are not the components the BROVI and the Writing Sample are designed to assess. On the reverse side of the rubrics, teachers have the opportunity to add anecdotal comments on an atypically poor or outstanding BROVI due to extenuating circumstances. Each rubric entry stands as documentation of a learner's performance on a specific task at a given moment. The rubrics are designed to capture the initial, ongoing and final assessment history for a student on a single page. Because each student has a BROVI and a Writing Sample rubric for the year, it is convenient to see, at a glance, how she is progressing over time.

The BROVI and the Writing Sample are significant, but not the only criteria for promotion. As the time for level change recommendations approaches, teachers consider BROVI and Writing Sample assessment results, classroom participation, homework, attendance, and other informal assessment they have made of each student, as they weigh a student's readiness for the next level. The official assessment record is just one bit of data, one piece of the puzzle to consider in the level recommendation process. It is an aid, not a straightjacket. Clear agreement on this point has freed staff up to complete the assessments as honestly and consistently as possible, and to continue to consider how to improve our assessment tools.

Limitations:

No matter what, there is always some subjectivity in evaluating language proficiency. Efforts to quantify the components of effective oral and written communication are elusive. Describing fluency, richness of expression, and grammatical accuracy in a speaking activity with a numerical score will always be part art. Satisfactory levels of inter-rater reliability can be achieved, but intensive training, which is costly in time and money, is still required.

Role-plays are dependent on the strength of one's partner. As in sports, if you "play" with a partner who has equal or better skills than you, you will "play up." It is often difficult, though, to maintain your own level of skill when you "play down" with someone who is less skilled. For this reason, the BROVI guidelines require that initial assessments always use the speech option. By the time a class gets to the ongoing assessment, students know each other well enough and have improved their skills sufficiently to manage role-plays. Furthermore, the teacher is better able to pair students effectively for role-plays.

The listening comprehension aspect in the Speech format is limited. It is weighted less because it is dependent only on the 1-3 questions that the audience asks the speaker. The concept of the listening activity handouts was included thanks to the persistent enthusiasm of Tim Laux, a part-time member of the original assessment team.

The Writing Sample is useable, but not ideal for low literacy students. It tends to show what they cannot do, rather than what they can do.

Feedback to students is limited. The protocol now calls for teachers to encourage the class in general terms following assessment, but to avoid "reviewing the assessment with individual students." The rationale is based in preserving the official assessment as an assessment, not an instructional activity. This is an area where we are divided on how much more time we might give to one-on-one feedback.

Each teacher assesses her own class. Each task is assessed only once. Ideally, teams of teachers would assess a class batch of writing samples to guarantee accurate scoring. Teachers might also be more objective if they assessed another's class on the BROVI (though the students would likely perform worse for an unfamiliar teacher.) The cost of multiple assessors is prohibitive and the logistics of swapping classes for BROVIs is too unruly. We acknowledge, however, the energetic team spirit that surfaces during group trainings and the benefit of sharing scoring tips and rationales.

What's next?

One of the advantages of alternative assessment tools such as the BROVI and SCALE Writing Sample is that the program has full control to reflect on their usefulness, identify priority points for modification, and incorporate improvements. Not only do teachers regularly ask "what if..."questions about administration, but they also suggest new topics for both the oral/aural and written tools. We have refined the rubrics several times in minor ways to make them easier to use and clearer to read. We have tweaked the scoring correlation and likely will make one last adjustment at the end of the current year. We continue to train together to share strategies on how to listen to or read the same samples and hear or see similar strengths and weaknesses. Our aim is to become reliable within a few (raw score) points, such that the SPL correlation will generally be the same. We anticipate that training will be an annual event, to sharpen the skills of veteran teachers and to orient new staff to the fine points of our tools.

We would love to revisit our student focus groups, especially to collect ideas from learners who have been through several rounds of assessment. Follow-up focus groups would give us valuable information: Have we succeeded in making the assessment process "friendly" for the students? Do they see the connection between assessment, curriculum, and class activities? Does regular assessment help the learners understand their own learning curves? Are there changes the students would recommend? Are there essential elements in assessment from the student perspective that we have overlooked?

"We would love to revisit our student focus groups, especially to collect ideas from learners who have been through several rounds of assessment."

Like many adult education programs, we are struggling to abide by the regulations our funders have required, while we design and implement an assessment system that is integrated with our program. Teachers, counselors, administrators, and students are learning to understand the role of this type of assessment and to assess fairly and honestly, without fearing for the program if there are occasional backslides in learner scores, or outrageously low scores for students with performance anxiety on the day of assessment. Assessment contributes to level promotion criteria, informs curriculum design, and represents part of the picture of the success of our programs and of our field. We need to remember, however, that assessment is still only a snapshot of how a learner is doing at a particular moment on a particular day. We load the film, prepare the subjects, focus as best we can, shoot, and hope for the best.

Betty Stone earned her M.A.T. in French and ESL from the School for International Training in Vermont. She has been ESOL Program Administrator at the Somerville Center for Adult Learning Experiences, SCALE, since 1978, and keeps her hand in teaching through subbing, occasional guest teacher appearances, and sharing teaching, learning, and assessment ideas with the great staff at SCALE and around the Boston area.

Vicki Halal has been an instructor in the adult education field since1993, after receiving an MA in ESL Studies from UMass/Boston. She has worked mainly at UMass and SCALE teaching an array of ESOL learners and levels. Currently, she is working at SCALE in the ESOL and GED departments.

TO THE STUDENT: The Speech

- **Why?** You do a speaking/listening activity 2 or 3 times a year to see what you know and how much you have learned. This activity shows us one way you can use the English you are learning.

- **What are we looking for?**
  We are looking for strong speaking:
  - speak for a few minutes about 1 subject
  - use good grammar
  - speak clearly with good pronunciation

Read the Speech Instructions below. Then, turn this paper over and look at the "remember" hints.

SPEECH INSTRUCTIONS

- Your teacher will give you a topic to speak about for 1 or 2 minutes.
- In a small group, share information about the topic. Talk about what you know about it.
- After you prepare, tell the class about the topic.
- The other students are listening and writing on the paper ("Audience Listening Activity").
- Answer questions that your classmates ask.

- **Remember, to speak well, you:**

1. Speak for enough time to give the information.
2. Stay on the topic.
3. Use good grammar structure.
4. Use all the vocabulary you know for that subject.
5. Speak clearly.

- **Remember, when you are listening, you:**

1. Do not talk.
2. Try to understand the people who are talking.
3. Pay attention to the people who are talking.
4. Write anything you want to remember.

BROVI - ESOL SPEECH TOPICS

ESOL 1/A/1Lit

1. Describe your job. Do you work in your home or outside? Do you work alone? What kind of work do you do?

2. Describe a special place in your country. Why is it important to you? Why do you go there?

3. Describe your favorite (living) relative. Who is she/he? What do you do together? Why do you like her/him?

ESOL 2/Basic Skills

1. Compare the weather in your country and in the United States. How are they the same or different? Which weather do you like better (prefer)?

2. Describe an ideal job you would like to have or a great job you had in the past. Describe the job and the working conditions. What is/was your favorite part of the job? Why?

3. Describe the life of older people in your country. Where do they live? Are older people in your country happy?

ESOL 3/B/ESOL Intermediate R/W

1. Describe your first trip away from your home in your birth country.

2. Describe when and why you miss a typical food from your country so much.

3. Describe a valuable lesson you learned in life when you were younger.

ESOL 4/C

1. Explain how you get news about your country (now that you live in the United States.)

2. Describe the reaction you had the first time you ever used a computer.

3. Begin your speech with the words: "Let me tell a story about an accomplishment that makes me feel very proud."

ESOL 5/D/ESOL R/W

1. Describe a piece of excellent advice you once gave a friend or family member.

2. Describe how you will continue to learn after you leave SCALE.

3. Describe the advantages and disadvantages of being an immigrant in the Massachusetts.

AUDIENCE LISTENING ACTIVITY FORM-CONVERSATIONS

Name: _____ **Class:** _____ **Date:** _____

WHO IS SPEAKING?	WHAT ARE THEY TALKING ABOUT?
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	

WHO IS SPEAKING?	WHAT ARE THEY TALKING ABOUT?
10.	

SCORING RUBRIC FOR BROVI ASSESSMENT

Student: _____ **Program Year: 2001-2002**

Circle **R** (role-play) ***or*** **S** (speech) **below for each assessment. S *only* for INITIAL.**

COMPONENT	**Initial** - Circle form: **S**					**Ongoing** - Circle form: **S R**					Final - Circle form: **S R**				
	0	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
1. Fluency x 5 • Fluidity • Length • Elaboration • Focus															
2. Listening Comp. x 2 • Basic understanding • Responsiveness to others															
3. Grammar & Sentence Structure • Subject verb agreement **x 5** • BE • Present tenses • Past tenses • Perfect tenses • Complexity • Variety															
1. Word Choice x 3 • Appropriate use • Richness of expression															
5. Pronunciation x 4															

Put the TOTAL raw score and circle the SPL in the space to the right. (there is a space for the initial, ongoing and final assessment scores)	Initial:	Date:	Ongoing:	Date:	Final:	Date:
	0 1 2 3 4 5 6 7 8		0 1 2 3 4 5 6 7 8		0 1 2 3 4 5 6 7 8	
	Tchr:	Class:	Tchr:	Class:	Tchr:	Class:

Scoring Key

0=not evident	Component is demonstrated	0 - 10% of the time
1=emerging		10 - 35%
2=evident		35 - 60%
3=established		60 - 85%
4=consistent		85 - 100%

COMMENTS Teachers, please initial and date comments

RAW TOTAL	SPL
0 - 3	0
4 - 12	1
13 - 21	2
22 - 30	3
31 - 41	4
42 - 53	5
54 - 64	6
65 - 72	7
73 - 76	8

A Writing Rubric to Assess ESL Student Performance

BY INAAM MANSOOR
AND SUZANNE GRANT

The Challenge

Performance-based assessments are popular because they are often program-based and learner-centered; however, funders tend to question their credibility. We challenged ourselves to address this issue by finding a way to satisfy technical quality issues, such as validity and reliability, while also keeping in mind how assessment influences learning. We believed that this approach would facilitate reporting student achievement both fairly and credibly.

Who We Are

The Arlington Education and Employment Program (REEP) is an adult English as a Second Language (ESL) program administered through the Arlington Public Schools in Arlington, Virginia. Because of its close proximity to our nation's capitol, the area draws large numbers of immigrants attracted by job opportunities in the service industry and a large number of national and international organizations. Nine levels of ESL instruction are offered, including workplace literacy and computer-assisted instruction. There are some 6,000 enrollment slots at 8-10 locations throughout Arlington County. There are 55 trained and experienced ESL teachers, who are supported by 5 coordinators. In addition, more than 100 volunteers support instruction.

Our Story

In 1995, REEP staff developed a writing rubric. A rubric is a scoring device that specifies performance expectations and the various levels at which learners can perform a particular skill. By articulating what our adult ESL learners could do at various proficiency levels, we hoped to fine-tune placement of learners into appropriate class levels and monitor their progress. Our rubric was developed by collecting writing samples from each class level and analyzing them. We found that although we had nine instructional levels, our students' writing fell into six distinct writing performance levels. The differences in these levels could be articulated using five characteristics (learning targets) of our learners' writing: ***content and vocabulary, organization and development, structure, mechanics, and voice*** (See REEP Writing Rubric attached). As part of our work with the What Works Literacy Partnership (WWLP: a group of adult basic education programs from across the country building their capacity to effectively use data for program improvement and decision-making. For more on WWLP, please go to www.wwlp.org), we designed and implemented a study to determine

"A rubric is a scoring device that specifies performance expectations and the various levels at which learners can perform a particular skill."

the effectiveness of using the REEP Writing Rubric to measure progress. With support from WWLP, we developed pre- and post-test writing tasks to assess writing gains.

Developing writing tasks that could be used for program-wide testing of beginning through advanced level students was challenging. To be fair, the tasks needed to generate a wide variety of responses and enable students at different levels to demonstrate their abilities and life experiences. We decided that the performance task of writing a letter of advice based on their own experiences would meet the above criteria and be consistent with skills that students were practicing in class. Moreover, we structured the testing process to mirror instructional practice by engaging students in warm-up activities prior to the actual writing test.

What Works

Reliability of test data is extremely important in the context of program-wide assessment, especially when the assessments are reported to funders. To maximize the reliability of our results, WWLP researchers provided extensive guidance on field-testing, test administration procedures, scoring, performance task development, and rater training. As a result, we implemented the following:

- **Field-testing.**

Before administering the pre- and post- writing tests to hundreds of students, we conducted field-testing to answer the following questions:

1. Can we expect measurable progress within the specified test interval, that is, 120-180 hours of instruction?

2. Can beginning through advanced level students demonstrate their writing skills in response to our writing tasks?

3. Are the pre- and post-test tasks equivalent, that is, do they represent the same level of difficulty?

To answer questions 1 and 2, a small group of experienced teachers administered the pre-test to five students from each class level at the beginning of an instructional cycle. At the end of the cycle, the teachers administered the post-test to the same group. Students were asked for feedback and they said they felt that they were able to demonstrate their writing skills with these tests. Teachers also thought that the tests demonstrated the students' writing abilities. Experienced readers scored the tests, and then a WWLP researcher analyzed the results. The analysis showed that significant gains could be measured and that reliable results could be achieved using the scoring procedures we had implemented. We were ready for large-scale testing.

To answer question 3, the same group of students representing all class levels was given the pre-test followed by the post-test within a three-day period. A WWLP researcher analyzed the results and found no difference between students' pre- and post-test scores, which demonstrated that the two tasks represented the same level of difficulty.

One of the key elements in achieving equivalence was the use of the letter genre and parallel warm-up activities for both the pre- and post-tests.

- **Test Administration.**

Prior to each test administration, testers participated in trainings on ground rules and how to administer the test, for example, time limits, no dictionaries, and how to conduct warm-up activities developed for the particular writing task. This ensured that all students completed the pre-writing activities and the test in a uniform way.

- **Scoring Procedures.**

Each of the five writing characteristics receives a score between 0 and 6, with 6 the highest. The total score is determined by adding each characteristic score and dividing by 5. A sample scoring grid follows.

	Content & Vocabulary	Organization & Development	Structure	Mechanics	Voice	Total (5 subsections)
Pretest Score	3	4	3	4	3	3.4 (17/5)
Post-Test Score	4	4	4	4	3	3.8 (19/5)

- **Building scoring consensus.**

REEP staff were trained to use the writing rubric to score the two (pre- and post-) performance tasks. developed Readers scored a range of essays. Scores for each writing characteristic were charted out as shown above, and the scoring rationale was discussed. This enabled the trainers to see how consistently the rubric was being interpreted, to pinpoint areas of discrepancy, and build scoring consensus.

A shortened version of this process was repeated prior to each scoring session to ensure continued consistency in rubric interpretation and scoring. Consistency among the readers was tracked to determine how many tests needed a third reader.

Each test was scored by two readers, and a third reader was used if the total score was more than one point different. The second reader did not know how the first reader had scored the test. In this way, the first reader's score did not influence the second reader. Similarly, students' class levels were not indicated on the test paper.

Scoring of the tests occured in group sessions of no longer than two hours each. This seemed to be the point

"(Teachers) used their students' test results to inform their instruction so that they could better meet the needs of their students. "

"Students at all levels started paying more attention to their writing as a result of the more formalized writing test."

at which readers began to "burn out."

The training and scoring procedures described above resulted in an inter-rater reliability of 98%. Only 2% of the tests needed a third reader.

Lessons Learned

REEP teachers were involved in every step: developing writing tasks and warm-up activities, administering tests, developing scoring procedures, scoring tests, and analyzing data. Through this involvement, teachers developed a deeper appreciation of testing. They used their students' test results to inform their instruction so that they could better meet the needs of their students. Scoring tests written by beginning to advanced level students gave them a broader picture of writing levels within the program and informed their decisions about subsequent class placements.

Teachers shared the writing rubric with their students, giving them a better sense of how they were being evaluated. Students at all levels started paying more attention to their writing as a result of the more formalized writing test. Many began to embrace writing instruction in the classroom. Learning English now meant more than learning to "speak" English.

We have all gained a greater understanding of the testing process and its need to be both fair and credible to all stakeholders. By participating in the test development process, teachers have developed skills and knowledge that will enable them to develop performance-based classroom assessments which meet this criteria as well. These skills enable us to feel more confident about accepting and reporting gains derived by performance based assessments.

A Word to the Wise

Developing and using a performance-based assessment requires tremendous time and financial commitment as well as access to the expertise of researchers. This commitment must be weighed against the outcomes, and in our case, the results for the program were significant and extremely positive.

We had hoped to demonstrate that a performance-based assessment could be a potentially superior instrument for measuring learner gains and thereby gain credibility with funders. Indeed, our work with WWLP gave us access to researchers who both guided us through the testing process and provided feedback on quality issues. At this writing, we are pleased to report that our WWLP researcher has concluded that "the REEP Writing Rubric is a carefully designed and validated instrument with sufficiently high reliability." We were fortunate in having access to the WWLP project and the professional support it provided. Practitioners need opportunities like this in the future if performance-based assessments are to become accepted measurement instruments.

Contact

Inaam Mansoor, Director
Arlington Education and Employment
Program (REEP)
2801 Clarendon Boulevard, #218
Arlington, Virginia 22201
Tel.: (703) 228-4200
Fax: (703) 527-6966
Imansoor@arlington.k12.va.us
http://www.arilngton.k12.va.us/depart-ments/adulted/REEP

R	Content/Vocabulary	Organization and Development	Structure	Mechanics	Voice
0	• no writing • no comprehensible information				
1	• little comprehensible information • may not address question • limited word choice, repetitious	• weak, incoherent	• frequent grammatical errors • mostly fragments • 2-3 phrases/simple patterened sentences	• lack of mechanics • handwriting and/or spelling obscure meaning	• not evident
2	• addresses part of the task (some but little substance) or copies from the model • irrelevant information • frequent vocabulary errors of function, choice, & usage with meaning obscured	• thought pattern can be difficult to follow, ideas not connected, not logical	• serious and frequent grammatical errors • meaning obscured • sentence structure repetitive (or copies from model)	• frequent errors • inconsistent use of punctuation • spelling may distract from meaning • invented spelling	• not evident
3	• addresses at least part of the task with some substance • limited vocabulary choice • occasional vocabulary errors but meaning not obscured	• limited in appropriate details-insufficient amount of detail or irrelevant information • trouble sequencing • may indicate paragraphing	• restricted to basic structural patterns (simple present, subject-verb), has some errors • correct usage of adverbials (*because clause*) and conjunctions (*and/or/but*) • goes outside of model	• some punctuation and capitalization though frequent errors that distract from meaning	• emerging voice • some engagement • some personalization
4	• addresses the task at some length • begins to vary vocabulary choice • occasional vocabulary errors but meaning not obscured	• uses details for support or illustration (reasons, contrasts),but development of ideas is inconsistent. Some ideas may be well developed while others are weak. • indicates paragraphs	• has some control of basic structures (simple present/simple past) • attempts compound sentences (e.g. *with and, or, but, so*) • some complex sentences (e.g. *with when, after, before*, *while, because, if*) • errors occasionally distract from meaning	• uses periods and capitals with some errors • may use commas with compound and complex sentences • mostly conventional spelling	• shows some some sense of purpose • some engagement • more personalized, may provide opinions and explanations
5	• effectively addresses the task • extensive amount of information • varied vocabulary choice and usage although may have some errors	• can write a paragraph with main idea and supporting details • attempts more than one paragraph and may exhibit rudimentary essay structure structure (into, body, conclusion)	• attempts a variety of structural patterns • some errors • uses correct verb tenses • makes errors in complex structures (*passive, conditional, present perfect*)	• uses periods, commas, and capitals • most conventional spelling	• authoritative, persuasive, interesting • emerging personal style
6	• effectively addresses the task • substantive amount of information • varied and effective vocabulary choice and usage	• multi-paragraph with clear introduction, development of ideas, and conclusions • ideas are connected (sequentially & logically) • appropriate supporting details	• syntactic variety • well-formed sentences • few or no grammatical errors (*verb tense markers, comparative and/or superlative*)	• appropriate mechanical and spelling conventions	• authoritative • strongly reflects the writer's intellectual involvement • personal style is evident

Illuminating Understanding: Performance Assessment in Mathematics

BY TRICIA DONOVAN

"Observation of a student's actual performance a on task has been a fundamental tool of assessment throughout history...But mathematics students fill in a bubble or a blank to indicate that they can understand somebody else's solution to a problem." (Mathematics Assessment: Myths, Models, Good Questions and Practical Suggestions, NCTM, p. 13) Too often this is the case in mathematics classrooms. You either have the answer right or wrong, and who cares how you figured it out. Using a performance assessment means, to the contrary, that you do care how a person arrived at an answer.

Performance assessments are designed to reveal a learner's understanding of a problem/task and *her/his* mathematical approach to it. The task can be a problem or a project; the task might even be a performance — demonstrate the balancing of equations (and, therefore, the essential nature of the equal sign). It can be an individual, group or classwide exercise.

What the task purports to measure should be clear. Furthermore, it should emerge from classroom curriculum, for the object of any performance assessment assignment is to determine what learners know and how they use what they know. Performance tasks are not about good guessing, and usually not about single right answers. When we teach measurement for instance, we might want to know if learners are able to convert smaller units to larger units and visa versa, so we create an assessment task that requires finding the lengths of various objects and reporting those lengths in several units. Learners demonstrate what they know and their method of solution as they undertake the task. As teachers, we use this information to set the academic agenda (and, in some cases, the social agenda— working together more effectively as a group, etc.) for the individual, group and/or class. Therefore, any task not related to the anticipated or implemented curriculum is inappropriate for our purposes.

Finding a task that illuminates a person's knowledge and application of skills is no easy search: Is it the summative assessment task, or an emergent one embedded in the instruction that we seek? Are we creating a pre-assessment to determine prior knowledge of a subject? Whatever our intent, we should communicate it clearly to the learners. Whatever our intent, we need a task that is valid; that is, one that reveals levels of understanding regarding particular learning objectives addressed in the classroom, and for which criteria regarding what constitutes performance from entry to excellence have been articulated.

"...the object of any performance assessment assignment is to determine what learners know and how they use what they know."

"A good performance task provides a lens through which to view student understanding. However, it's important to have a clear vision of what's being assessed, and the criteria should be transparent to all, including the learners."

A good performance task usually has eight characteristics (outlined by Steve Leinwand and Grant Wiggins and printed in the NCTM Mathematics Assessment book). Good tasks are: essential, authentic, rich, engaging, active, feasible, equitable and open. In adult education, we might add that they should connect to participants' goals.

An essential task represents a 'big idea' and aligns with the core of the curriculum. To be authentic, a task must use processes appropriate to mathematics practice and learners should value the outcome of the work. A rich task is one that has many possibilities, raises other questions, and can lead to other problems. An engaging task is one that challenges the learner to think, yet encourages persistence. Active tasks allow the learner to be the worker and decision-maker, and allow students to interact as they construct meaning and deepen understanding. Feasible tasks are safe, developmentally appropriate, and able to be completed during class time and as homework. Equitable tasks promote positive attitudes and develop thinking in a variety of styles, while open tasks have more than one right answer and offer multiple entry points and solution paths. Of course, to have all these qualities, a task must be near perfect. Good tasks hit most of the characteristics.

Examples of performance tasks follow. In a class working on fractions, for instance, the teacher might assign a task that asks groups or individuals to design an activity that will help the class understand how small 1/10 is. S/he might seek a broader task, too, by asking learners to list everything they have learned about fractions so far. If the class has been studying averages, s/he might ask learners to write an explanation that proves the statement "median is always the middle number" is either true or false. In addition, s/he might ask learners to look at some real estate listings in which a median house price is listed and discuss, given the range of houses listed, why the realtor chose to look at the median as opposed to the mean. The task might be extended by asking learners, "Who might want to know the mean in this case and why?" If studying geometry, learners might be presented with a diagram of a right triangle with a 45° angle and one leg that measures 5cm and be asked to list out everything they can tell about this triangle. For a class in which percents are the focus of study, a teacher might present an 'eating out' situation and ask how to figure out how much to leave, tip included.

A good performance task provides a lens through which to view student understanding. However, it's important to have a clear vision of what's being assessed, and the criteria should be transparent to all, including the learners. Not sharing the criteria for assessment has been compared by some to asking someone to take a driver's license test without telling them what's being tested. How do you prepare for such a test? How do you know if you're doing what's expected?

Most performance tasks are scored using a 'rubric'. A rubric can be divided into sections such as: understanding the problem; planning a solution; getting an answer. Points are then awarded for various levels of performance, such as "no attempt to plan a solution" or "completely inappropriate solution" or "partially cor-

rect plan" or "workable plan" that could result in correct answer.

A sample rubric set from the fall 1996 edition of The Problem Solver (Problem Solver Special Edition: Assessment of Mathematics Understanding, vol.4, No.1, Western Mass. SABES) was devised for a performance task that involved investigating rents in town, graphing them and finding averages. There was a 'Skills Assessment' rubric and a 'Habits of Mind' rubric. They looked something like this:

SKILLS ASSESSMENT:

3-mastery 2-demonstrated use 1-unused or misused

Skills Assessed in Task	Competency Level
1. Computation (adding and subtracting whole numbers, dividing)	
2. Finding the average	
3. Comparing averages	
4. Following directions for setting up graphs	
5. Using a graph to answer questions about information contained in graphs	
6. Recording data	
Comments:	

HABITS OF MIND:

3-highly visible 2-evident 1-not evident 0-N/A

Affective Domains Assessed in Task	Expression Level
1. **Persistence** (sticks with problem)	
2. **Curiosity** (engages in problem)	
3. **Flexibility** (attempts alternative solution methods)	
4. **Thoroughness** (checks answers, responds to all questions, compiles sufficient data)	
5. **Creativity** (unique approaches, responses or presentations)	
6. **Cooperation** (shares ideas and materials, listens, etc.)	
7. **Communication** (states ideas clearly, asks appropriate questions)	
8. **Reasoning** (shows logical and/or intuitive reasoning; inductive and/or deductive reasoning; proportional reasoning; generates hypotheses)	
9. **Problem Solving** (uses a variety of strategies and/or appropriate strategy; poses interesting, sensible problems...)	

"It's easier to be a good teacher if you know what's understood and what isn't."

Obviously, the nature of the performance task being used to make assessments as well as the purpose of the assessment determine the rubric form. We may want to assess only one competency — simplifying fractions, for instance – and in that case we can look at a problem that learners are working on, one that requires adding fractional amounts, and choose to look only at the work done regarding simplifying fractions. In such a case, we might look to see if the computations are done mentally or with pencil and paper, and if done with paper and pencil, we could then ask if the fractions are being simplified by the largest factors possible or by 2's, etc.

Perhaps the most difficult work with performance assessments, as with any assessment, lies in the final act. What recommendations do we make based on what's been illuminated? At least with a performance assessment, there is a clearer idea as to where the problems in understanding or skill exist. We can tell if careless computation or total lack of place value understanding is at play; we can tell if the concepts of perimeter and area are clear, but a person is using counting or adding as opposed to formulas to determine each. It's easier to be a good teacher if you know what's understood and what isn't. Performance assessments in mathematics make teachers wise in the ways of their learners.

Tricia Donovan taught GED classes in Western Mass for 12 years before joining TERC in Cambridge as a curriculum developer/writer on the EMPower math project. She worked on the original ABE Math Standards and on the current ABE Math Frameworks. In addition, she is editor of The Problem Solver*, an ABE math newsletter funded by DOE and SABES West, and a doctoral candidate in the Teacher Education Curriculum Studies Department at the University of Massachusetts, Amherst.*

Student Health Education Teams in Action

BY MARY DUBOIS

The following excerpt was taken from the Spring 2001 issue of Field Notes, written by Marcia Hohn, Regional Coordinator of the Northeast SABES Regional Support Center.

The Massachusetts Health Education team is a group of health educators and adult literacy practitioners in Massachusetts. Its goal is to promote the health of learners and teachers through health education activities in adult literacy programs. We understand health to include physical, emotional, and spiritual aspects, and that a healthy life is defined individually and culturally. We intend to support individuals, family, and community/environmental well-being through facilitating ways for learners and teachers to evaluate current health choices, make informed decisions, and create new options. This includes decisions concerning prevention and access to health care. We understand that our goal may also involve advocacy for policy and other social changes at the state, regional, and local levels.

At first glance, it looks like any other adult education class, multiage, multicultural, and multilevel. Upon closer inspection, however, questions arise. Who is the teacher? Is it the middle-aged woman at the board? Could it be the young man who appears to be taking notes? Perhaps it is the older woman who is speaking? When entering the classroom, the casual observer would notice a group of adults ranging in age from mid 20's to early 70's. During the discussion everyone takes part. The material they are talking about comes from the ABE Health Curriculum Framework, and the students are members of the Health Team at a weekly meeting. The woman at the board is this week's leader, Amelia. She is a student is an ABE GLE (Grade Level Equivalent) 4-5.9 class, and she moderates the discussion and writes the brainstorming results. The young man is Jose from an ESOL SPL (Student Performance Level) 2-3 class and is he taking notes of the meeting. The older woman voicing a suggestion, Martina, just moved from an ABE GLE 0-3.9 class to a GLE 4-5.9. She is making a point during the discussion.

The teacher/facilitator has just distributed a copy of the ABE Health Framework to each of the eight team members. While looking over the framework, one of the team members notices a diagram illustrating the importance of good health in every aspect of life. The team decides to make a poster of the diagram to display on the Health Bulletin Board. They discuss making a new heading which would be easier for all students at the center to understand. "I need a com-

"The Health Team concept is a participatory model."

pass," says David, an ESOL 2-3 student from Portugal. He wants to draw and enlarge the diagram, which includes a sizeable circle. The teacher/facilitator goes to the GED Classroom to look for one. Jose, David's classmate from Cape Verde, suggests making a compass. Jose's wife, Helena, and Winnie from China, both students in ESOL SPL 6-8, take two pencils and some string and make a compass. Amelia, the leader, needs a ruler to measure out the lines for the heading. All of this is accomplished without any input from the facilitator. The students drive the process.

After the poster is completed, the team practices their cancer presentation. They are researching cancer on the Internet in response to a health issues survey they developed for all students at the center to complete. Initially, the survey was posted on newsprint and students voted during class break time. When checking the results, the team felt that the numbers were low and perhaps invalid. They decided to ask students to complete a paper copy of the survey during class (see Initial Assessment in appendix). This produced more accurate results. The cancer presentation includes a two-sided paper with a drawing done by David illustrating "What is Cancer?" The team reviews the language to assure understanding of all classes at the center, which range from ESOL 0-1 to GED. On the reverse is a typed list of prevention tips augmented with clip art drawings (see appendix). After practicing, the team decided to pair up for the class visits. They talk about which team members would be most effective in each class in terms of translation requirements of ESOL. Next they discuss the best time and day to visit the classes.

The last item on the agenda is the Health Team T-shirts. A member of the Center's Advisory Board had referred the team to the adjacent public junior high for possible printing of the shirts. The Advisory Board at the program is composed of representatives from community agencies, business people from the community, the program's director, counselor, and volunteer coordinator, a state representative, and students from the Adult Learning Center. After several phone conversations and visits regarding colors and graphics, the school declined the job. One of the team members now proposes making their own shirts using the computer and iron-on material for the graphics. This is enthusiastically received and will be the plan for the next meeting.

The Health Team concept is a participatory model. A former health team member describes the process. Sandra, a dentist in her native country of Colombia writes, "I became involved in the health team in a voluntary way. One of my teachers gave me an application. I filled it out and after that we had a meeting to know each other. We started to plan how to do good things." All students at the center are offered the opportunity to join the Health Team. The application is modeled on an employment form (see appendix). After the teacher/facilitator reviews the applications, interested students gather for an "interview." Potential members meet with current or past team members and the teacher/facilitators. The former team members explain the duties of the team and offer examples of the work previous teams have done. Candidates answer questions such as: Why do you want to join the team? How do you think

you can help the team? Do you have any health-related experience? Do you understand the way the team works? Do you have any questions for us? Based on both the application and the interview, the facilitator and former team members vote on candidates for the current team. Although the members receive a stipend, this is not mentioned until after the members have been chosen. At the first meeting, the new members are told about the stipend as well as the fact that it is based on attendance and participation. Members receive the stipend at the conclusion of the school year.

Health topics are identified by surveying all the students at the center. Team members research the health issues indicated by the survey results. Sandra reports, "We did research into depression, asthma, high blood pressure, nutrition, diabetes and quit smoking. We made brochures, contests, newsletters, bulletin boards, presentations, and we explained the topics in an easy way for everyone. We went to other schools and performed a skit. We shared coloring books (teaching about asthma) and crayons with children." The team also arranged for Public Health nurses to come to the center to check blood pressure. To follow-up they arranged for a visit by the mobile health van for additional screenings. Team members called and scheduled both visits. They notified all students and staff and also developed a timetable for classes to be checked. The team escorted each class to the screenings and wrote thank you notes to the health practitioners. All of this work was accomplished during weekly Health Team Meetings.

At first the new team members are a little nervous about taking charge. Their previous academic experience, especially for ESOL students, is to sit quietly without much participation while the teacher runs the class. In addition, some students are not confident in their ability to speak English. However, they recognize right away that the experience of speaking and writing English will improve their skills in both areas. Once they grasp the concept that they are in charge, and the teacher is a facilitator who stays in the background and is used as a resource, a transformation takes place. The team members work together in a cooperative fashion or independently on an aspect of a project they feel strongly about. They come up with ideas, they decide how to address the health issues, i.e. developing a brochure, video, pamphlet, newsletter, or bulletin board, arranging for guest speakers, contacting community health practitioners, etc. In addition, their research is the basis of a curriculum, which is distributed to all staff members. The curriculum contains an initial assessment as well as a post assessment. The initial assessment can be a survey, and true-false "quiz", a brainstorming discussion, or a K-W-L process where students list what they Know (K) about a topic as well as what they would like to know (W). At the conclusion, they will note what they have learned (L). The post assessment may be the initial assessment, a survey, or a product. Last year's team was concerned mid-year as to whether they were reaching the learners at the center. They decided to develop a survey asking for additional learner input and suggestions.

Overall, the participatory concept meets the needs and requirement of adult learners. It touches upon the Freirean approach in that learners identify prob-

"The team members work together in a cooperative fashion or independently on an aspect of a project they feel strongly about."

lems and issues from real-life experiences and seek solutions. The theories of Malcolm Knowles, which suggest that adults move from dependency to self-directedness, draw upon their experience for learning, and want to solve problems, also support the health team model. The strongest evidence lies in the words of the student members of the health team, "I have a good time being part of health team. I learned about health, I improved my relationship with classmates, teachers, and students. I met people from different schools. Had been part of the health team was an unforgettable and nice experience."

Mary Dubois has been associated with the New Bedford Public Schools/Division of Adult and Continuing Education's Health Team since its inception. She is currently the Curriculum Facilitator and has taught ESOL, ABE, and Adult Diploma Classes during her seven years with the program.

NEW BEDFORD ADULT LEARNING CENTER
STUDENT ACTION HEALTH TEAM INITIAL ASSESSMENT

This will follow the process of the Student Action Health Team members.

Brainstorm Ideas
The class will have an oral discussion on "What I Know About Cancer." The teacher will list this information on the board as students offer input

Suggested vocabulary:

symptom	surgery
treatment	diagnosis
radiation	cell
chemotherapy	tumor
risk	benign
screening	malignant
prevention	carcinoma

Writing
Students will write what they know about symptoms and prevention.

Survey
The class will complete a survey listing the different types of cancer and the number of people they know with each type. One class at the center will compile the surveys for the school.

Don't Smoke

Use Sunscreen

Eat a Healthy Diet

FIND THE ANSWERS TO THE FOLLOWING QUESTIONS IN THE NEXT ISSUE.

1. What is the most common type of cancer for both men and women?
2. What is the most common cancer in women?
3. What is the most common cancer in men?

STUDENT ACTION HEALTH TEAM

HEALTH TEAM APPLICATION

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP _____

TELEPHONE __(___)_____

DOB _____

NATIVE LANGUAGE _____

CLASS _____

TEACHER _____

PLEASE COMPLETE THE FOLLOWING:

I would like to be part of the Student Action Health Team because _____

_____

_____

- -

OFFICIAL USE ONLY

RECOMMENDATION: _____

Involving Learners in Assessment Research

BY KERMIT DUNKELBERG

"It is very difficult to begin learning foreign language, coming to an unknown country, especially if you are not as young as you used to be.

When I entered into ILI, I was enrolled into the intermediate level. I didn't understand everything at first, but could catch a gist of speech. We had a very nice teacher who spoke to us coherently. We read the articles from a local newspaper and discussed them, wrote small essays, so that way learning new words and improving our spoken language.

Of course, we had practice after class, by watching TV, reading books, etc. In our class, we had students speaking different languages, it was an original practice too, speaking with them, and understanding them.

One day our teacher offered me and my pal to participate in a project. She briefly told us about it, and we agreed to participate in it. In this project, several students also joined in. The main idea was to help teachers to define the knowledge level of students. Each of us told our own opinions that focused on what could improve our spoke language and grammar. To express your feelings in an unfamiliar language is very hard, especially if you have lived here for only six month. Nevertheless, our teachers listened to us very patiently, and we made a little reference for each level of English study as a second language. Participating in this project, I improved my English, that was very helpful when I took exam for college"

—Dina Bakousseva, student participant on ILI's Curriculum Frameworks Assessment Team

From January to July, 2001, International Language Institute of Massachusetts' Curriculum Frameworks Assessment Team worked to develop an approach to assessment which would be consistent with our teaching philosophy as well as the emerging assessment criteria of the MA Department of Education. Since one of ILI's core values is learner-centered instruction, we were interested from the beginning in learner-centered methods of assessment. After some deliberation, we decided to involve learners directly in our process of researching viable assessment methods.

What follows is an account of our six-month process involving learners in assessment. I have built this account on the minutes I took of our meetings,

in order to give a picture of how the project unfolded over time. As often as possible, I have included the comments of students and other teachers as recorded in the minutes. While my account will inevitably be colored by my own perceptions, I want to emphasize that this was a team effort — involving students, teachers, and administrators. Our Curriculum Frameworks Assessment Team (CFAT, for short) included five teachers (Yvonne Telep, Jennifer Rafferty, Sarah Miller, Kermit Dunkelberg, and ILI's Director of Programs, Caroline Gear), and a diverse group of six ESOL students. The learner participants were: Dina Bakousseva and Elena Sidorova from Russia, Gulsen Kosem from Turkey, Eric Lin Qu from China, Yumiko Millette from Japan, and Amos Esekwen from Congo. They ranged in Level from ESOL 5-8. I was particularly pleased that students outnumbered teachers on this project.

Laying the Groundwork

Several months before our project formally began, teachers at ILI began to meet around issues of assessment. On September 7, 2000, three teachers (Cindy Mahoney, Yvonne Telep, and Kermit Dunkelberg) discussed the pros and cons of standardized vs. alternative assessments:

We discussed the pros and cons of using a standardized test, such as the STEL [Standard Test of English — a grammar test used by ILI's Intensive English Program as part of intake assessment] for ongoing assessment. On the one hand, teachers felt that grammar is a good indicator of level. On the other hand, we questioned the validity of measuring via a standardized test. Some students have expressed a desire for more testing, others have strong opposition to traditional tests. Standardized testing seems, to us, to run counter to the philosophy of both ILI and the Curriculum Frameworks. But DOE's increased demands for accountability, and the fact that they have provided us with a long list of possible standardized tests (in SMARTT, Massachusetts' data collection system) suggests that there may be a place for standardized testing [in future assessment systems].

October 4-5, I attended the DOE Director's Meeting in Falmouth, and wrote the following notes to ILI's teaching staff: *I attended sessions on Performance Accountability and NRS (National Reporting Systems). Of all big topics, this is the biggest:*

— "What motivates students to come to us is what defines our accountability."
(Bob Bickerton, MA State Director of Adult Education, 10/04/2000)

I believe DOE is really committed to learner-centered-ness, and does not want to mandate, from the top down, what we teach. But at every level (us to DOE, DOE to Feds or State Legislature), there are the twin burdens of 1) determining what students need; 2) verifying that through data collection. We could cynically try to give DOE "what they want," or we can really listen to our students.

In the rest of my notes from the Director's Meeting, I raised some of the questions our Assessment Team would grapple with for the next months: Is a standardized procedure the same as a standardized test? What is a rubric? How can assessments be valid, reliable, and learner-centered?

Laying Out the Issues

On January 26, 2001, four teachers (Caroline Gear, Yvonne Telep, Jennifer Rafferty, and Kermit Dunkelberg) met to discuss these issues. From the beginning, we chose to view our work positively:

As we move forward, we will want to think of how to establish a time line and realistic benchmarks for what we can accomplish. We will also want to figure out practical ways to ***divide up the work****, and develop mutually productive scenarios for* ***learner involvement****.*

We move forward in the conviction that we have the potential to positively impact the ways in which we will be "asked" (required) to assess learners and program effectiveness in the future. We also have the chance with this project to strengthen our programs...

We began with each person present brainstorming (on whiteboard) on what s/he associated with the "universe of assessment" we were being asked to address. Among the issues raised were:

- *What is the best timing of assessments?*
- *Will assessment be different in our classroom and Distance Learning programs? Is what is "valid" in one instructional setting equally valid in the other?*
- *How can competencies beyond the four language skill areas be assessed? (For instance, computer literacy, Intercultural Knowledge and Skills, or Navigating Systems)*
- *How can our assessments match what we're doing in the classroom?*
- ***Learners need to be able to interpret the tools!***
- *Let's not reinvent the wheel! Look to models that are out there!*
- *How can our assessment procedures get around teachers' inevitably subjective responses to individual students?*
- *Will standardized tests have a place in our assessment procedure?*
- *What kinds of assessment tools are our learners most comfortable with?*
- *Accountability:*
  - *To our Learner's needs and goals (our primary accountability)*
  - *To Federal and State guidelines (WIA, NRS, DOE): our imperative to stay funded!*
  - *Accrediting organization (ACCET)*
  - *ILI's institutional culture: (Learner-centered, flexible, authentic)*
- *How can learners be meaningfully involved in this process?*

"We move forward in the conviction that we have the potential to positively impact the ways in which we will be "asked" (required) to assess learners and program effectiveness in the future."

"One value of involving learners in our meetings is to offer them a meta-perspective on the assessment issues, allowing them to advocate for what makes sense to them, and to report back to other students from a student perspective."

Beginning the Research

Over the next two months, we explored a wide variety of existing assessment tools, from standardized tests to performance-based prompts and rubrics. We began to review a range of writing rubrics from various sources (including K-12 and English Language Arts, as well as ESOL), and to use them to score sample student writings to see if we agreed on the level the writing represented. (We didn't quite know it yet, but we were testing our "inter-rater reliability" with various rubrics!). We discussed the pros and cons of these rubrics, and began to work on developing our own. At the same time, we worked toward a common understanding of key vocabulary ("valid," "reliable," "authentic," "standardized," "holistic," "analytical"), and key debates in the field. Members of the Team participated in a workshop on standardized testing and another on EFF (Equipped for the Future), both at SABES West (the western region of the state). We also participated in the Western MA SABES Assessment Work Group, in which we joined colleagues from other programs in investigating various models of assessment. We read a host of articles on assessment, which we found in *Field Notes*, *Focus on Basics*, or on the web. In EFF terms, we were rapidly expanding our "knowledge base."

Still, we had not found a comfortable way to involve learners in the process. The learning curve was steep enough for us. How could we expect learners to master these issues? On the other hand, we reasoned, perhaps students would bring their own knowledge base to the project, offering a different kind of expertise? After all, they know far more about what it's like to be a student in our program than we do!

First Meeting with Learners

On April 26, 2001, we had our first meeting involving learners. We were still unsure how to involve them in a meaningful way, but we figured, "let's invite them and decide that together!" We noted that:

One value of involving learners in our meetings is to offer them a meta-perspective on the assessment issues, allowing them to advocate for what makes sense to them, and to report back to other students from a student perspective.

The six learner participants had been chosen from our 5-6 and 7-8 SPL level classes on the basis of their interest, communication skills, and ability to work closely with teachers. Prior to the meeting, student participants had been given a brief orientation, and were given a packet which included:

- *A Position Description (See Appendix B)*
- *"Words and Phrases You Will Hear A Lot" (See Appendix C)*
- *CRESST Assessment Glossary*
- *NRS Guidelines*
- *ESOL Curriculum Frameworks Chart*
- *Timesheet*

The Student Participant Position Description emphasized that:

Our main goal in involving student participants is to be sure that we are ***listening to student voices*** *as we build assessment policy. We would like you to share in our discussions, to offer your opinions, to do some reading and research with us, and to try out some of the assessment methods we are working on. Your experience and perspective as learners in the classroom are important to us.*

We hoped to develop assessment procedures which were consistent with our classroom culture. I decided that our process should reflect our classroom culture, too. From now on our *meeting style* would draw on our *classroom style*, emphasizing pair and small-group discussion. As in the classroom, this would give everyone a chance to speak, and make sure student voices were constantly heard. The first activity of our joint meeting was to pair up teachers and students to discuss:

Why did you say 'yes' to this project? What is interesting about it to you?

Teachers answered that they wanted:

. . . *to know more about different types of assessment.*
. . . *to know how to be able to tell students if they are making progress.*
. . . *to help students learn how to assess their own progress.*
. . . ***a system*** *of assessment.*
. . . *to give the students confidence.*
. . . *to learn about state and federal guidelines.*
. . . *to keep funding for our program.*

Students answered that they wanted:

. . . *more conversation/listening practice.*
. . . *to be part of decision-making.*
. . . *a chance to be involved in how to know. . . am making progress.*
. . . *to help us make the time in class as useful and important as possible.*
. . . *I am against the MCAS—I don't want to see adults take a test.*
. . . *Assessment should be useful.*

We agreed that:

Both students' and teachers' motivations for participation in the project should be respected.

Student/teacher pairs then discussed an area of competence (not language learning), and answered the question:

"How do I know when this is being done well/successfully?"

Examples included: Waiting Tables, Teaching Dance, Movies, Finding and Hiring New Employees, and Dee Jaying in a Club. In groups, we discussed the correspondences between "assessing" these activities and assessing language. We discussed what is subjective, or objective, about these assessment processes. This raised the question of the difference between "assessment" and "evaluation." Evaluation, we felt, was more informal and subjective. For instance, the teacher who had been a Dee Jay stated that she continuously modified the music she played based on her subjective feeling of what was the "right music" for the "right time." Evaluation often employed a "rule of thumb" rather than a standard measure. Assessment relies on

"We hoped to develop assessment procedures which were consistent with our classroom culture."

quantifiable data, and has at least the appearance of being more objective.

Would an "objective" assessment system even be desirable in some of these cases? (Are movies which have been vetted by numerous "focus groups" more satisfying than those which have not? Often the opposite is true). What is easy to count? Does it tell us what we want to know? For instance, when the student who had been a waiter was asked if a waiter's ability could be assessed by counting tips at the end of the night, he emphatically said "no." Ultimately, he felt the best measure of a waiter's ability was evaluation by an informed observer (the manager). And not on one night, but *over time*.

Evaluation and Assessment

Finally, we asked: How does this exercise inform our process of developing assessment strategies at ILI? Students and teachers agreed that evaluation is a strong part of our day-to-day teaching at ILI. Oral and written feedback are incorporated into all of our classes, so that instruction is in response to student's goals and needs. Talking about the non-language learning situations above clarified some of the strengths of constant evaluation. Somewhat like the Dee Jay in the club, the classroom teacher modifies instruction in response to constant feedback and observation. Like the restaurant manager, a classroom teacher observes students *over time*, and has a good sense of students' day-to-day (as opposed to one-time) performance.

While we agreed that ILI had long been strong in the area of evaluation, we also agreed that there was room for improvement in the area of assessment. Specifically, teachers and students both wanted to be able to say more clearly *to what degree* an individual student had improved. Students were not so much interested in an abstract number as they were in knowing *what they needed to do next* to improve. Similarly, while teachers recognized that reporting requirements (to the Feds and DOE) are a "fact of life," and arguably can lead to program improvement, our primary allegiance was to the growth of our students. Consequently, our most pressing concern was our ability to document learner progress *to the learner*: to be able to say "you used to be able to do these things, now you can do these more advanced things, and this is how I know." Again, we were less interested in a number (a BEST or TOEFL type score) than in being able to articulate to students *what skills areas they had improved in, how we noted that change, and what they might need to focus on next.*

Level Descriptors

This discussion brought us back to the subject of level descriptors.[1] At our next meeting (May 3, 2001), we looked closely at the NRS SPL descriptors. Both students and teachers found that the NRS guidelines were general and inconsistent. Moreover, students couldn't understand them easily. *We felt the NRS descriptors were not adequate to serve as the basis for a dialogue between a student and a teacher about what the student's level was, and what that meant for the student's progress*. Why not? We came to two conclusions:

[1] *We began by examining some of the historical background to the NRS. (Our information came from the March 2001 Implementation Guidelines of the "Measures and Methods for the National Reporting System for Adult Education," found online at http://www.air-dc.org/nrs). We were surprised to find that the NRS had arisen, in part, out of an effort to save adult education from being subsumed by "a general system of workforce development." In 1995, Congress had demanded "strong and convincing data" to demonstrate adult education's "effectiveness as a separate program." State directors of adult education had asked for "a national system for collecting information on adult education outcomes." So, while some of us in the group had regarded the NRS guidelines as narrowly workforce-centered, they were in part an attempt to stake a claim for a broader conception of adult education. We were also surprised to learn that:*

- *Among the sources used to develop the NRS were the National Institute for Literacy's EFF (Equipped for the Future) and the CASAS (California Adult Student Assessment System).*
- *The two types of assessment specifically mentioned in the NRS are "standardized test, or a performance-based assessment with a standardized scoring rubric."*

- The NRS guidelines define entry level requirements, but our students want a continuum. *Students want to know all the steps for moving **through** a level to the next one*. They want to be able to see how the levels "connect" to each other.

- The NRS guidelines are difficult for students to interpret. *We want our students to be able to look at the descriptors and say, "Ah! That's my level!"*

Yvonne Telep brought in level descriptors from Washington State.[2] They provided a model for us of a *graded* SPL descriptor, in which each level was divided into three or four sub-levels, enabling students to see more precisely where they stand in relation to a level.

We determined we would write our own level descriptors, consistent with NRS yet more detailed, in accordance with the MA DOE Curriculum Frameworks and our classroom system of instruction. In order to make these descriptors accessible to students, we decided to write them in student/teacher teams. (One of the greatest days for me, as Project Coordinator, was the day our Director of Programs sat down with a Level 5 student to begin writing a draft level descriptor for Level 5 Reading).

This was only the second meeting involving students, but already they were taking their place beside us as equal partners in the work. We were amazed at how interested they were in the issues surrounding assessment (clearly, a lot was at stake for them), how *quickly* they were mastering complicated material, and how *effectively* they were working with us as team members. Students were getting a lot out of it, too. We did short written feedback at the end of each meeting (part of the ILI culture of continuous feedback and observation). Student responses from the April 26 meeting included these comments:

- *Today, we focused on the level description. And we found something that we really like and don't like and even some questions after we deeply touched every sentence what/how they separate the different level. I thought it's very useful and make me more get the topic what we will do and discuss*

- *I really like being in this project.*

- *Today I have get a lot of things. We had conversation about students level and everybody from us decide that a grammar and vocabulary are important for studying English. It was very useful personally for me because I had a big practice speaking with native citizens. I had a fun!*

Student-Centered Assessments

Even as we moved forward with drafting more useable level descriptors, we continued research into what kinds of assessment procedures would be used to *determine* the level. Students had told us clearly that their time was valuable, and that the assessment procedure should be a learning opportunity.

At the same time, teachers were wary of "teaching to the test," if a standardized test were implemented for state-wide assessment. Instead of "teaching to the

"This was only the second meeting involving students, but already they were taking their place beside us as equal partners in the work."

[2] *The Washington state level descriptors are available, with some diligent searching, from http://www.sbct.ctc.edu/Board/Educ/ABE/assess.htm.*

test," could we "test to the teaching"? Could the dog wag the tail, instead of the tail wagging the dog? Could assessments model classroom instruction?

Examples from Colorado State

Caroline Gear brought in materials from the Colorado State Department of Education. The Colorado DOE materials provided examples of clearly-defined performance-based assessments modeled on familiar classroom pair and small-group activities like role plays, interviews, information gap, and "Listen, Repeat, Do."[3]

Both teachers and students were very excited by the Colorado model, because the performance tasks fulfilled some of our emerging criteria for learner-centered assessment:

- The assessments presented authentic speaking and listening situations. Students were communicating with each other with a real need to communicate.
- The authentic situation provided an opportunity for including Standards from ESOL Curriculum Frameworks Strand 5, *Developing Strategies* and *Resources for Learning*, as part of the assessment.
- The tasks were modeled on familiar classroom activities which students had done before.
- Assessment time was also learning time (the assessment procedure included a volunteer to assist pairs/groups who were not being assessed).

On the other hand:

- The assessments were complex to administer, requiring the assistance of at least one trained volunteer in addition to the teacher.
- The pair and group model was very appropriate for our classroom program, but not as appropriate for our Distance Learning Program.

Moreover, despite repeated requests, we were unable to get copies of the scoring rubrics for these assessments from the Colorado Department of Education.

Equipped For the Future (EFF) Standards

Our interest in learner-centered, authentic assessments also led us to examine the National Institute for Literacy's Equipped for the Future Standards. We were attracted to EFF in part because, in determining "What Adults Need to Know and Be Able to Do in the 21st Century," NIFL had begun by *asking adult learners* what they felt they needed to know.[4] Also, as a federally-funded program, EFF had "clout," and seemed to us to offer a different perspective on learning from the NRS. We were therefore astounded and pleased to learn that EFF is developing a comprehensive, performance-based assessment system which will be linked to the NRS![5] Unfortunately, the EFF assessment project had just begun in 2001, so none of their assessment materials were available to us.

Both students and teachers found the EFF Standards to be consistent with the

[3] *The materials were obtained at a presentation by Kathleen Satopietro Weddel (Northern Colorado State Literacy Resource Center) at the 2001 TESOL conference in St. Louis. Colorado assessment materials are available from: Marie Willoughby, Colorado Department of Education, CARE/Family Literacy and Adult Education, 201 East Colfax Avenue, Room 100, Denver, CO 80203.*

[4] The Equipped for the Future Content Standards *are available from the National Institute for Literacy (NIFL), 1775 I Street NW, Suite 730, Washington, DC 20006.*

[5] *See* EFF Voice *2:1, Winter 2001, also available from NIFL, and on-line at http://www.nifl.gov.*

ESOL Curriculum Frameworks. EFF seemed to us to provide new ways of thinking about aspects of the Frameworks which were important to us, but which we weren't sure how to incorporate into assessment. These included Strand 5 (Developing Strategies and Resources for Learning) and the Seven Guiding Principles. We saw a strong correspondence between these aspects of the ESOL Curriculum Frameworks, the EFF Standards for "Life-Long Learning Skills" and "Interpersonal Skills," and learner observations such as *"What is hard for me is not pronunciation, but the fear of (mis)pronunciation!"* and *"Our level can change everyday."*[6]

Finally, we responded warmly to EFF's emphasis on communication as a shared, negotiated act (something not reflected in standardized pencil and paper assessments, but vital to our students' motivations for coming to our classes). This was reflected even in the titles of such EFF Standards as "Listen Actively" and "Speak So Others Can Understand."

Level Descriptors, Again

At our May 10 meeting, we returned to the task of writing better SPL descriptors. We asked ourselves what sources we should consider in writing level descriptors, and agreed on the following short list:

- *NRS Guidelines*
- *Massachusetts Curriculum Frameworks*
- *ILI in-house materials*
- *EFF*
- *Washington State rubrics*

I would add that an additional, implicit resource was our students' own statements about their level. We asked students in the group to describe their "level" to us. What could they do easily, often, with confidence? What was difficult for them? In some cases, student members of the team moderated discussions with their classmates, eliciting responses to these same types of questions. Student's own descriptions of their "level" tended to be experiential rather than academic, with rich detail about such aspects of communication as confidence and cultural context, as well as "learning gains" nearly impossible to assess (i.e., "Now I can *think* in English!") Student descriptions of their challenges and successes informed our understanding of the "levels."

We divided into small groups to consider one skill area, at one level (Speaking Level 5), from each of these perspectives. We then met in large group to identify barriers to our process by discussing:

"What was hard about what we just did?"

We identified the following difficulties, among others:

- *Neither MA Curriculum Frameworks nor EFF are divided by level*
- *The NRS progression is not clear*
- *The levels in NRS and Washington State are not organized the way DOE classes in MA are organized.*
- *Our in-house materials are not clear or accurate enough.*
- *Washington State describes only 6 levels.*
- *It is difficult to make the descriptors precise without being prescriptive*

[6] *Compare these student remarks with the language of the Frameworks: "Language learning requires risk-taking." (Guiding Principle 5). "Teachers and learners need to understand that progress may be inconsistent from day to day." (Guiding Principle 4).*

"If we believe, as Merrifield suggests, that literacy is "rooted in particular social contexts," then considerations such as social conventions, context, and register need to be taken account of in our assessment procedures."

We brainstormed about what "aspects" we should look for when assessing Speaking. Some that we agreed on were:

- *Pronunciation/Intelligibility*
- *Grammar*
- *Usage*
- *Risk-taking, for instance use of new vocabulary*
- *Fluency, Rate*
- *Confidence*

Another group of aspects was more controversial. These tended to reflect socio-cultural dimensions of communication. We all agreed they were important, but could/should they be part of assessment of levels?

- *Social conventions of oral communication*
- *Context*
- *Register*

I was reminded of Juliet Merrifield's comments on shifting definitions of literacy in "Contested Ground: Performance Accountability in Adult Basic Education":

"Over time, views of what literacy means have shifted from academic skills... to functional skills...Literacy is now described as multiple 'literacies' rooted in particular social contexts...When literacy meant what is taught in schools, performance was testable... The research on literacy in its social context has been carried out through careful observations of literacy events and activities which shed light on prevailing literacy practices....While it shifts the focus to performance in life, not in test situations, this new research has not yet been incorporated into practice, assessment, or policy."[7]

If we believe, as Merrifield suggests, that literacy is "rooted in particular social contexts," then considerations such as social conventions, context, and register need to be taken account of in our assessment procedures.

Responses to the May 10 meeting included these comments, from students and teachers:

- *Yesterday I had no idea how we would write a level description, today I have some ideas.*
- *What we did felt like progress, like we are swimming forward instead of treading water.*
- *The last 2 meetings we only talked in general, today we made ourselves more aware and organized to focus on one part.*
- *The most thrilling part was to hear (my student) use the word "concise."*
- *I learned that communication is a big deal! I like the way EFF says "Speak So Others Can Understand."*

After this meeting, we abandoned our "parallel track" of drafting better SPL descriptors, while, at the same time, trying to create performance tasks and scoring rubrics. We focused our remaining time on creating SPL descriptors for as many skills and levels as we had time for, working in student/teacher teams.

As a first step, we each (students *and* teachers) wrote a descriptor for ESOL Speaking Level 5. The richness, complexity, and sheer difficulty of our task is reflected in this list of "highlights" of the various descriptors from our May 17, 2001 meeting:

[7] *Juliet Merrifield, "Contested Ground: Performance Accountability in Adult Basic Education," NCSALL Report #1, July 1998. Available on-line at http://gseweb.harvard.edu/~ncsall/research/report1.htm*

- *All the teachers agree that the thoughtfulness and quality of the students' contributions is extremely high.*

- *Jennifer's is admirably concise and seems to fit with NRS.*

- *Eric's is easy to understand and includes a sense of communication as negotiation.*

- *Dina's suggests good things for students to keep in mind, including confidence.*

- *Caroline's identifies competencies to be addressed throughout the level.*

- *Elena, Yumiko, and Kermit based theirs on NRS mandates. Elena included non-verbal language (eye contact and body movement), and stressed the EFF "Speak So Others Can Understand."*

- *Kermit's had a list of functions, levels of performance, good summing up of grammar and vocabulary, and had Intelligibility as a category.*

- *Amos emphasized active, rather than passive role in Level 5 ("initiates conversation"). Had a real learner's perspective.*

- *Yvonne's charted a clear progression. Elena felt it would help students see where they were on the continuum.*

We all agreed that Yvonne Telep's proposal—based on the Washington State example, with the addition of a range of "competencies" which could serve as the basis for performance-based assessment tasks, was a major breakthrough (*see Appendix D*). We agreed to follow the format she proposed. In the remaining weeks of the project, we developed SPL descriptors for Speaking 3-8, Reading 5, and Writing 5-6.

Curriculum Frameworks Showcase and ALRI Workshop

Students in our project conducted research, debated the conceptual framework of what we were doing, and participated in drafting rubrics and level descriptors. Students also became involved in sharing the results of our work with the field.

Student participants undertook the design of our display poster for the 2001 DOE Curriculum Frameworks Showcase on June 8. Elena Sidorova, a Level 6 ESOL student, designed a computer graphic showing "Our Influences and Inspirations," which was the centerpiece of the poster. The chart included *Curriculum Frameworks, NRS, EFF, MELT, SABES West Assessment Group, Colorado Performance Assessments, Washington State Rubrics, "Field Notes", "Focus on Basics", ILI curriculum, ILI crosswalk, Students' Voices, and Teachers' Experience.*

On June 14, ILI conducted a day-long workshop on Learner-Centered Assessment at the Adult Literacy Research Institute (ALRI) in Boston. Several of our learners participated in presenting the workshop. Like our assessment meetings, much of the workshop was conducted in small groups, in which our learners mixed with the teachers and SABES support staff who were present. The unanimous

feedback of the group was that having learners at the table, helping to conduct the workshop and offering their perspectives throughout, was the greatest value of all.

For our learners, too, this opportunity to synthesize the lessons learned from our project, and to convey them to a group of professionals in the field, was a turning point and a fitting culmination to our work. One student told me several months later that she had accepted a new job which would require her to give occasional presentations. She felt prepared to accept the job because of her work on our project, in particular the experience of presenting the workshop at ALRI.

Moving On

Of the six students who participated in the project, only one is still a student at ILI. Three have gone on to college, and two more moved on to better jobs. Their success certainly says a lot for the individual qualities which led us to select them for the project in the first place. But I believe the project also served as a catalyst for each of them to take the next steps toward their dreams.

Fortunately, they have left a lasting legacy. As we at ILI continue to move forward in developing assessment procedures which will serve both government reporting requirements *and* our students' needs, the lessons we have learned from this project stay with us. The voices of these students still ring in our ears, and so it is only fitting to close with a few of their comments about the project:

- *It was a lot of toil. But I would do it again. All students should participate in this project.*
- *Everything was OK! What was done and what will be do it's very interesting and useful personaly for me...*
- *As a student, I'm very pleased to be part of this working group. It's wonderful to know what teachers think about learners.*
- *We found out that everybody have some good points and very useful. After deeply talking, make more impression in our mind. That's good.*

Kermit Dunkelberg is an ESOL teacher and Program Coordinator at the International Language Institute of MA in Northampton, Massachusetts.

APPENDIX A

Questions which came up in an SPL level 7/8 class

when it was proposed that we try one of the CASAS writing assessments

- You mean, like a “test”?
- What’s wrong with tests?
- Do we want to spend half an hour of class time just writing?
- How often would we do this? (If not often, how can one measure accurately reflect what a learner is capable of? What if I"m tired that day? If too often, see previous question!)
- Is it fair to have a time limit? (Some students write faster than others, although not necessarily better).
- Couldn’t we take it home to do? (If we did, would it still be valid?)
- Could we use a dictionary?
- Should we be assessed on a draft, or a revised piece of writing? (This class had been emphasizing rewriting).
- Could we write on any subject we wanted to?
- Would we write about something we had done in class?
- Why not just use some writing from our journals?
- You mean, like a “test”?

APPENDIX B

Student Position Description

International Language Institute of MA Curriculum Frameworks Assessment Team

Purpose: The main purpose of the Assessment Project is to improve our methods of evaluating student progress in language learning. In other words, how do teachers and students know when, and how much, a student has improved in the "four basic skill areas" of *speaking, listening, reading,* and *writing*? Closely connected to this is the question of which class a student should be in, and when and how students progress to another class. In addressing these questions, our program has to consider ILI's teaching philosophy as well as state and national guidelines for Department of Education programs.

Student Participant's Role: Our main goal in involving student participants is to be sure that we are *listening to student voices* as we build assessment policy. We would like you to share in our discussions, to offer your opinions, to do some reading and research with us, and to try out some of the assessment methods we are working on. Your experience and perspective as learners in the classroom are important to us.

As a student participant, you are expected to:

- Attend a Thursday afternoon meeting once a week or once every two weeks.
- Learn a little about the "big picture" of state and national guidelines which our school must fit in with.
- Learn some special words and concepts we use to talk about assessment.
- Participate in discussions with the teachers about how to improve this part of our teaching.
- Write an evaluation of your participation in the project to include in our final report.

Project Dates: Week of April 23 through June 30

Stipends ($): All student participants will receive a weekly *stipend* of $32 per week, or $8 per hour for four hours of work a week. This will include both meeting time and reading and research time. You will probably find that some weeks you work more than four hours, and other weeks less, but the weekly stipend will be the same. The stipend is to honor the value of your work on the project. It is not a wage.

Thank you for your interest!

APPENDIX C

"Words and Phrases You Will Hear A Lot"

Student Participants — ILI Curriculum Frameworks Assessment Team

ABE: Adult Basic Education

Assessment: A way of measuring progress. Assessments should be "valid" and "reliable." Assessments are also "countable." We must report our assessment results to the Department of Education (DOE) to show that our program is successful. In this project, we are focusing on assessment of the four "skill areas" of language learning.

Curriculum: Plan of study. What is taught, and how it is taught.

Curriculum Frameworks: The state guidelines for DOE Adult Basic Education (ABE) classes in English for Speakers of Other Languages (ESOL). The Frameworks are supposed to be a guide for how to teach, not what to teach.

DOE: Department of Education. The money for our program comes from the Massachusetts DOE.

EFF: Equipped for the Future. A federal study of "what adults need to know and be able to do in the 21st century," as workers, learners, family members and citizens. EFF provides Content Standards with useful descriptions.

NRS: National Reporting System. The NRS describes national SPL's (Student Performance Levels). All programs should fit with NRS descriptions (but the descriptions are not clear).

Performance-based Assessment: Measuring how well a student does at a particular task (speaking, writing, etc.).

Reliable: Able to give the same results each time, no matter who is testing.

Rubric: A scoring grid to measure student performance.

Skill Areas: We identify four "skill areas" for language learning: speaking, listening, reading, and writing. Students may be at different levels for different skills.

Student Performance Level of performance expected (0-10). At ILI, we have classes at the 3-4, 5-6, and 7-8 levels. These are multi-level classes (more than one level).

Valid: Measures what you want to measure. "Valid" also means that an assessment "fits" with a program's curriculum and philosophy of learning.

APPENDIX D

Level Descriptor
Speaking, SPL Level 5
Draft/ Yvonne Telep
International Language Institute of MA

Functions: (Also referred to as competencies.)

Student will participate in the following activities or exchanges:

- Reporting an event in the past, such as an accident, a previous job or any specific experience.
- Explaining the steps in a process.
- Using polite language/ accepted conventions to request clarification and repetition when needed.
- Making polite requests.
- Conversing in limited social situations in English.
- Expressing agreement or disagreement.
- Stating reasons or giving excuses.
- Expressing future plans or needs.

APPENDIX D

Level Descriptor
Speaking, SPL Level 5
Draft/ Yvonne Telep
International Language Institute of MA

Exceptional (Exit Standard)	Individual demonstrates control of the simple, progressive, and perfect present and past tenses. Errors are infrequent and don't interfere with meaning. Syntax and pronunciation are usually intelligible, although errors still occur. Syntax and pronunciation are not perfect, but competent enough so that meaning is clear. Participation in the exchanges listed above is clear and intelligible with only occasional hesitation or errors. Individual is aware of errors and is able to self-correct, or to restate and clarify when asked to do so.
Competent	Individual demonstrates control of the simple and progressive present and past tenses and errors with infrequent errors which don't interfere with meaning. Individual shows awareness and understanding of the perfect present and past tenses but errors and omissions are frequent. Syntax and pronunciation may be problematic but meaning is generally clear. Individual can participate in the exchanges listed above with some hesitation, but with increasing confidence. Individual can sometimes self-correct, or is able to restate or correct when asked to repeat or clarify.
Developing	Individual uses the simple present and past tenses with occasional errors that don't interfere with meaning. Individual also shows awareness and understanding of the progressive present and past, but makes frequent errors. Grammar and syntax errors are less frequent, but may still interfere with meaning. Pronunciation problems occasionally interfere with meaning. Individual can participate in the exchanges listed above with hesitation and some correction. Individual shows a willingness to participate in the exchanges listed and understands mistakes when made aware of them.
Beginning	Individual has understanding of the simple present and past tenses. Errors of grammar and syntax are frequent and sometimes interfere with meaning. Pronunciation problems occasionally interfere with understanding. Individual can participate in the exchanges listed above with hesitation and assistance.

WMass Assessment Group–Tackling the Sticky Issues

BY PATRICIA MEW
AND PAUL HYRY

- WIA and NRS are facts of our lives. We need to learn how to live with them and to develop inventive ways of handling the accountability and assessment issues they create.

- The Massachusetts ABE system affords us the opportunity to be proactive in the assessment dilemma, to develop some assessment practices that satisfy our funders, our practitioners and our students.

- Both standardized and alternative forms of assessment are useful and necessary.

- We can approach assessment as a way to link SPLs (Student Performance Levels) or GLEs (Grade Level Equivalents), classroom curriculum and the standards in the Massachusetts Curriculum Frameworks documents we are addressing.

- We should try to educate our students in the issues and challenges of assessment to make them partners in the assessment journey.

These are the assumptions with which the Western Mass. Assessment Study Group began its venture in January 2001. The idea was to bring together practitioners from our region to learn about a variety of assessment methodologies, decide what we wanted to investigate further, and develop processes and tools that would be useful for our programs. We also thought the time was ripe for better understanding, and perhaps ultimately influencing, the direction of assessment and accountability in Massachusetts. The PAWG (Performance Accountability Working Group), a group convened by the state director of Adult Basic Education to decide how Massachusetts would meet its accountability obligations, was just beginning its own process. We hoped our parallel process would help us keep abreast of the work the PAWG was doing at the same time as we educated ourselves about assessment based on our own interests.

The programs represented in the study group, as well as Western Massachusetts as a region, have their own challenges. We quickly learned upon advertising the group that both ESOL and ABE practitioners were interested in the assessment problem. This may be due in part to the fact that all programs involved had received state Curriculum Frameworks (CF) grants requiring them to address assessment in their curriculum development work. Regardless, these practitioners brought a diversity of needs and skills to the group. Some had been grappling with assessment and curriculum development using the Massachusetts Frameworks for years. Others were looking at that connection for the first time. Some were

"The idea was to bring together practitioners from our region to learn about a variety of assessment methodologies, decide what we wanted to investigate further, and develop processes and tools that would be useful for our programs."

"How do we create valid and reliable Performance-Based Assessment strategies and tools that link student tasks to Curriculum Frameworks standards... and to ...Federal SPL and GLE levels required by the NRS?"

experts in the Curriculum Frameworks grant process; others were just beginning that process. In addition, Western Massachusetts has many different types of programs — rural, urban, multi-sited, correctional, community-based, large, small — all of which have their issues to overcome regarding the relationship between assessment and program design.

Our work took us down many paths over six months. We began by learning more about the National Reporting System and the Workforce Investment Act upon which the NRS is based. During the course of our process, we also hosted guest presentations on standardized testing issues from both ESOL and the ABE perspectives. But our most important and fruitful work was in the area of Performance-Based Assessment (PBA), and more specifically in the development of rubrics. We did this for several reasons. First, many of the practitioners involved were critical of the limitations of the standardized tests used most in our state. Second, we also all understood that the PAWG, whose work was to last about 18 months, might possibly head in the direction of Performance-Based Assessment, and the group wanted to travel the same route. Finally, we hoped that some of our findings might, in fact, be useful to the PAWG in the context of its lengthier, more intensive process.

As mentioned above, rubric development was a central and practical task in which we engaged ourselves. Our interest was in developing rubrics that would help make a crucial, and heretofore undefined, link between federally defined skills and levels (SPLs and GLEs) and the learning concepts, strands, and standards elaborated in the Massachusetts ABE Curriculum Frameworks documents. This work proved to be incredibly complex more complex, in fact, than we had originally imagined. First, there was the task of understanding rubrics and where they fit into the wider picture of Performance-Based Assessment. Then we examined a variety of rubrics created by other practitioners, programs, and states. Finally, since we found none that explicitly met our needs around the MA Curriculum Frameworks or similar standards, we decided to create our own rubrics, trying in the process to build links between these documents and the qualitative assessment work to which we were so committed. At the same time, we tried to address the NRS performance levels in the categories we included.

This work resulted, first of all, in group products, including several flow charts about the overall process of Performance-Based Assessment and two rubrics – one for ESOL and one for ABE — that link MA Curriculum Frameworks standards to a specific student learning task (development and delivery of an oral presentation). A second result was the development of increased capacity to share assessment theory and develop locally customized rubrics and related tools. Finally, we developed a series of questions and recommendations for our state Department of Education and the PAWG. The most crucial of these are:

How do we create valid and reliable Performance-Based Assessment strategies and tools that link student tasks to Curriculum Frameworks standards (which are defined qualitatively) and, simultaneously, to the quantitatively-defined Federal SPL and GLE levels required by the NRS?

How can a classroom teacher assess in the way the NRS dictates we assess? (Even with standardized testing, we can only assess a single performance at a given time.)

Assessment should be done in concert with curriculum development, but at this moment in the development of ABE in Massachusetts, these two things (assessment and curriculum development) are quite far apart. Once the PAWG reaches its conclusions about how ABE/ESOL assessment should be carried out in Massachusetts, there will need to be an intense training process (supported by funding streams) for teachers to help them learn how to assess students accurately using PAWG-generated processes. This training should include understanding how to move between the CF documents, the assessment process, and the assignment of federal levels to student performances.

When the Western Massachusetts Assessment Study Group finished its first year of work last June, members were appreciative of the chance to learn a great deal together. At the same time, we were clear that we had only just begun the process. We were discovering the possible relationships between assessment practices that are valid, reliable, and pragmatically feasible in light of multiple demands on learners' and teachers' time. We were also discovering the linkage of these practices to state and federal policies to which these processes need to be accountable. As our work progresses this year, we are excited about continuing together toward the development of solutions to the assessment puzzle.

Patricia Mew first worked as an ABE/GED and creative writing teacher at the Hampshire County House of Correction in Massachusetts. She is currently the Curriculum Development and Assessment Coordinator for SABES/West, where she has worked providing program and staff development since 1994. Pat co-facilitates the WMass Assessment Study Group with Paul Hyry.

Paul Hyry has worked in adult education in Holyoke, MA since 1994 as an ESOL teacher, curriculum developer, program director, and local ABE collaborative coordinator. Paul co-facilitates the WMass Assessment Study Group with Patricia Mew.

ESOL ORAL PRESENTATION RUBRIC BLANK (PRODUCED BY NICOLE GRAVES, PAUL HYRY, DIANNE SHEWCRAFT) — WESTERN MASS ASSESSMENT STUDY GROUP 5-01

Class/Level: **Intermediate ESOL(SPL 5-6)** ***Curriculum Area:*** **Oral Presentation (Topic — A famous building)**

_____ESOL CF Strands/ Standards Addressed	**Strand:** Intercultural Knowledge and Skill **Standards:** Recognize and understand the significance of cultural IMAGES and SYMBOLS—American and their own.	**Strand:** Language Structure & Mechanics **Standards:** Acquire/Reinforce basic English literacy skills	**Strand:** Oral/Written Communication **Standards:** 1,5,9 **Strand:** Language Structure and Mechanics. **Standards:** 2, 4, 5, 6, 8
____ Performance Area	**Content of Presentation**	**Organization of Presentation**	**Delivery of Presentation**
Levels			
Exceptional	• Clearly addresses and expands upon all all content areas identified in advance. • Information presented with multiple props (visual aids, etc.) and lots of variety.	• Clear development of beginning, middle, end. • Clear transitions and logical sequence of elements. • Fits time frame with ample time for discussion/Q&A upon completion.	• Establishes & maintains consistent eye contact with entire audience; easily heard by all. • Consistent use of grammar and syntax makes meaning clear at all times. • Pronunciation never impedes listener's ability to understand content.
Competent	• Clearly addresses most content areas identified in advance. • Information presented with a number props of at least one type.	• Clear development of beginning, middle, end. • Transitions clear but not smooth. • Fits time frame with limited time for discussion/Q&A.	• Sporadic eye contact with some audience members; generally easy to hear. • Grammar and syntax help to make meanings clear most of the time. • Pronunciation only occasionally impedes listener's ability to understand content.
Developing	• Addresses some of the content areas identified in advance. • Information presented primarily orally with minimal additional props to aid understanding.	• Develops some but not all parts of the presentation. • Some disorganization in sequencing and transitions. • Body of presentation fills most or all of the allotted time (or goes slightly long).	• Dependent upon paper with limited eye contact; usually can be heard by most. • General meaning usually clear, but grammar/syntax make understanding of specific points difficult. • Pronunciation sometimes impedes listener's ability to understand content.
Beginning	• Addresses few of the content areas identified in advance. Presentation of basic facts in basic oral format.	• No clear organizational strategy. • Time managed poorly (significantly short or long).	• Reads from paper without eye contact. • Hard to hear. • Grammar/syntax errors impede listener ability to understand. • Pronunciation impedes listener ability to understand.

A ORAL PRESENTATION RUBRIC — WMASS ASSESSMENT STUDY GROUP—5-01 (GRAMAROSSA, GREENBLATT, DUVAL, MEW)

Teacher: _____ ***Class/Level:*** 6/7 ABE 2 ***Curriculum Area*** ELA—Oral presentation on famous building

Curriculum Framework Strands/ Standards Addressed	**Strand:** Writing – (Content) Organization **Standard(s):** 1. Express thoughts in writing/speaking 2. Acquire more organizational strategies 3. Write/speak at greater length in response to topic	**Strand:** Writing - Content, Mechanics & Structure **Standard(s):** 1. Revise to include more details and information (6-9) 2. Recognize and use appropriate format & genres 3. Use complete sentences, aware of grammar, mech.	**Strand:** Oral Communication **Standard(s):** 1. Speak so others can understand 2. Communicate complex ideas clearly 3. Respond appropriately to other's questions and statements 4. Summarizes events; restates ideas to clarify
Performance Area/Task Description	**Organization of presentation**	**Content of presentation**	**Delivery of presentation**
Performance Levels			
Advanced	Descriptor: 1. Relevant ideas and information are developed logically, clearly and fully 2. Uses many interesting details, examples, anecdotes to explain, clarify information fully 3. Maintains purpose for speaking all times 4. Includes eff. intro and conclu. to engage listeners 5. Presentation flows well, is sequenced well at all times	Descriptor: 1. All information presented shows signs of accurate research and well-developed detail 2. Relates information well to listeners and their experience 3. Presentation shows a great deal of variety and creativity in content 4. Maintains listener interest in subject throughout 5. All grammar, mech. are correct	Descriptor: 1. Uses correct grammar, well-constructed sentences at all times 2. Able to restate ideas fluently w/out text 3. Seems relaxed and anxiety free, confident 4. Makes frequent eye contact; enunciation, volume and pace engage all listeners
Competent	Descriptor: 1. Relevant ideas and information are well-developed a majority of times 2. Uses several details, examples, or anecdotes to explain and clarify points 3. Purpose for speaking is clear most of the time 4. Introduces and concludes speech adequately 5. Sequences presentation adequately	Descriptor: 1. Most information presented shows signs of research and detail 2. Relates information well to listeners and their experience 3. Shows less than 5 errors in oral/written errors language conventions, self-corrects 4. Maintains listener interest in subject most of time 5. Shows some variety in content	Descriptor: 1. Uses correct grammar, well-constructed sentences, self-corrects when appropriate 2. Enunciates, paces speech adequately 3. Shows minimal anxiety; mostly seems confident 4. Makes eye contact occasionally, attempts to engage listeners 5. Restates main ideas, relies on text minimally

Developing	Descriptor: 1. Sticks to topic most of the time 2. Uses a few details, examples, or anecdotes to explain and clarify points 3. Purpose for speaking is somewhat clear 4. Introduction or conclusion may need work 5. Some ideas may be out of order	Descriptor: 1. Information presented shows little evidence of research or detail 2. Little attempt to relate information to listeners 3. Shows more than 5 errors in oral/written language conventions 4. Maintains some listener interest in subject 5. More variety in content is needed	Descriptor: 1. Uses incorrect grammar, poorly constructed sentences more than 5 times 2. Listeners have trouble understanding some things due to inappropriate pace 3. Shows anxiety 4. Seldom makes eye contact, little attempt to engage listeners; relies on text primarily
Beginning	Descriptor: 1. May speak off topic 2. Uses little or no detail or example to explain and clarify points; information presented is minimal 3. Purpose for speaking may be unclear 4. Little attempt to introduce or conclude presentation 5. Ideas may be out of order	Descriptor: 1. Information presented may be inaccurate, no evidence of research 2. No attempt to relate information to listeners 3. Shows more than 10 errors in oral/written language conventions 4. Little attempt made to interest listeners 5. Little or no variety in content	Descriptor: 1. Uses incorrect grammar, poorly constructed sentences more than 10 times 2. Pace of speech makes speech difficult to understand most of the time 3. Is anxious, nervous a majority of time 4. Unable to summarize or restate ideas 5. Makes no eye contact, may read speech