

Accountability and Assessment in Adult ESL Education

Dorry M. Kenyon

Director, Language Testing Division, CAL

Invited Colloquium: *Critical issues at the interface between
assessment and U.S. language education policy*

AAAL Conference

March 22, 2009

Denver, CO



Deriving Meaning from Context

- Second/Foreign Language Learner
- EFL/ESL Teacher
- Psychometrician
- Professional Language Tester
 - at the Center for Applied Linguistics, which houses the *Center for Adult English Language Acquisition (CAELA)* and developed and makes available two adult ESL assessments used for accountability purposes: *BEST Plus* and *BEST Literacy*

Interface with John's Introduction

Assessments are...

not values-free technical procedures for measuring constructs in systematic and maximally valid ways, accountable only to standards of psychometric quality...

but USED...SOCIAL...POWER

From **my** perspective:

Assessments must, however, **at least** be measuring constructs appropriate for their use and provide evidence that they are up to the professionally recognized standards of psychometric quality

Thesis

In addition to asking questions about the USE of assessments, investigating their SOCIAL dimensions, and recognizing their POWER, it is appropriate and imperative for applied linguists to:

- (a) investigate and contribute to the definitions of the constructs the assessments claim to measure, and
- (b) demand that such language assessments are up to recognized standards of psychometric quality (e.g., *Standards for Educational and Psychological Testing* of APA/AERA/NCME)

Outline

1. Background to the National Reporting System (NRS)
2. NRS and standards of psychometric quality for adult ESL assessments
3. NRS and the construct being measured by adult ESL assessments
4. Wrap-up

Outline

1. Background to the National Reporting System (NRS)
2. NRS and standards of psychometric quality for adult ESL assessments
3. NRS and the construct being measured by adult ESL assessments
4. Wrap-up

(1a) What is the National Reporting System (NRS)?

A cooperative data gathering venture between states and the Division of Adult Education and Literacy (DAEL) of the U.S. Department of Education, which publishes an annual report summarizing the data collected to Congress

For complete information: www.nrsweb.org

(1b) History of the NRS

- 1993 Government Performance and Review Act (GEPRA)
- 1995 Concerns of Division of Adult Education and Literacy (DAEL) in U.S. Dept. of Education
- 1998 Workforce Investment Act (WIA)
- 2000 First year of NRS operations

(1c) 5 Core Outcome Measures of the NRS

Topic	Measure	Categories or Definition
Educational gains	Educational gains	Educational functioning levels in reading, writing, speaking, and listening and functional areas
Follow-up measures	Entered employment	Learners who obtain a job by the end of the first quarter after the exit quarter
	Retained employment	Learners who obtain a job and remain employed in the third quarter after program exit
	Receipt of secondary school diploma or GED	Learners who obtain a GED, secondary school diploma, or recognized equivalent after exit.
	Placement in post-secondary education or training	Learners enrolling after exit in a postsecondary educational or occupational skills program building on prior services or training received

Exhibit 1.

National Performance: Adult Completion of Educational Levels and Core Outcome Measures PY 2000–01 Through PY 2004–05

	Percentage Achieving Outcome					Number Achieving Outcome (5-Year Total) (PY 2000–01 to PY 2004–05)
	2000–01	2001–02	2002–03	2003–04	2004–05	
Educational Gain ABE/ASE*	36	37	38	38	40	2,510,582
Educational Gain English Literacy*	32	34	36	36	37	2,006,175
High School Completion **	33	42	44	45	51	889,531
Entered Postsecondary Education**	25	29	30	30	34	241,520
Entered Employment**	31	39	37	36	37	668,376
Retained Employment**	62	63	69	63	64	903,046

Outline

1. Background to the National Reporting System (NRS)
2. NRS and standards of psychometric quality for adult ESL assessments
3. NRS and the construct being measured by adult ESL assessments
4. Wrap-up

(2a) Four USED Initiatives to Professionalize Adult ESL Assessments

1. Funding for development of *BEST Plus*
2. Workshop of the National Research Council
3. Adult ESL Assessment Project at CAL
4. Review to Approve Assessments

(2b) 1. *BEST Plus*

Programs developing survival level listening and speaking skills for adult ESL learners were widespread and needed an NRS-compatible assessment

Basic English Skills Test (BEST) 1982-84

Oral Section: Two static forms of scripted questions

Need for questions that couldn't be memorized (for pre- and post-testing) and measured a wider range of ability levels => computer-assisted, computer-adaptive instrument

1999 – proof of concept project

2000-03 – *BEST Plus* developed and operationalized

(2c) 2. National Research Council Workshop

December 12-13, 2001

“Performance Assessments for Adult Education: Exploring the Measurement Issues” (Mislevy & Knowles, 2002)

Brought together needs in adult assessment and experts with years of experience researching psychometric issues in K-12 assessment (National Council of Educational Measurement [NCME] types):

Bob Mislevy, Pamela Moss, Stephen Dunbar, Mark Reckase, Henry Braun, Ronald Hambleton, David Thissen, Stephen Sireci

(2d) 3. Adult ESL Assessment Project at CAL

2004-2006

1. Gap Analysis (19 adult ESL assessments)
2. Design Plan

Lyle Bachman, Kathleen Bailey (Applied Linguists/Language Testers)

Ronald Hambleton, Steven Sirici (Psychometricians)

3. Feasibility Study

(2e) Findings of the Gap Analysis

- (1) lack of well-defined test purposes, test uses, and language constructs being tested;
- (2) lack of reference to theories underlying second language acquisition in defining or describing the test construct;
- (3) lack of evidence of psychometric rigor in the test development process;
- (4) lack of evidence supporting equivalence between alternate test forms;
- (5) limited consideration of logistical factors that may affect the implementation of the assessment within the adult education context; and,
- (6) limited consideration of the potential use of technology to administer and score assessments.

(2f) 4. Review of Approved Assessments

First submission summer 2008, no final word yet

Evaluation Criteria:

1. What is the intended purpose of the instrument?
2. What procedures were used to develop and maintain the instrument?
3. Does the assessment match the content of the NRS educational functioning level descriptors?
4. Can the scores on the assessment match the NRS educational functioning levels?
5. Is there evidence of reliability and classification consistency?
6. Has construct validity of the assessment been demonstrated?

Outline

1. Background to the National Reporting System (NRS)
2. NRS and standards of psychometric quality for adult ESL assessments
3. NRS and the construct being measured by adult ESL assessments
4. Wrap-up

(3a) What are NRS Educational Functioning Level Descriptors? (1/2)

6 levels are defined (with the ability to exit out)

For adult ESL (three sets):

Listening and Speaking

Basic Reading and Writing

Workplace Skills

*Development/Origin UNDOCUMENTED

(3b) What are NRS Educational Functioning Level Descriptors? (2/2)

“The descriptors are entry-level descriptors and are illustrative of what a typical student functioning at that level should be able to do. They are not a full description of skills for a particular level. When a student has skills at one or more levels above the placement level, he or she has completed that level and can advance to the next level.” (www.nrsweb.org)

= NRS Level Gain

(3c) Sample Adult ESL Listening and Speaking

High Beginning ESL Individual can understand common words, simple phrases, and sentences containing familiar vocabulary, spoken slowly with some repetition. Individual can respond to simple questions about personal everyday activities, and can express immediate needs, using simple learned phrases or short sentences. Shows limited control of grammar.

Low Intermediate ESL Individual can understand simple learned phrases and limited new phrases containing familiar vocabulary spoken slowly with frequent repetition; can ask and respond to questions using such phrases; can express basic survival needs and participate in some routine social conversations, although with some difficulty; and has some control of basic grammar.

(3d) Original Testing Benchmarks: Adult ESL Language Skills

CASAS Life Skills (scale scores)	Basic English Skills Test (BEST) (raw scores)	Student Performance Levels (SPLs)
Beginning ESL Literacy		
180 and below	Oral BEST 0 – 15 Literacy BEST: 0-7	Speaking: SPL 1 Reading/Writing: SPL 1
Beginning ESL		
181-200	Oral BEST 16-41 Literacy BEST: 8-46	Speaking: SPL 2-3 Reading/Writing: SPL 2
Low Intermediate ESL		
201-210	Oral BEST: 42-50 Literacy BEST: 47-53	Speaking: SPL 4 Reading/Writing: SPL 5
High Intermediate ESL		
211-220	Oral BEST 51-57 Literacy BEST: 54-65	Speaking: SPL 5 Reading/Writing: SPL 6
Low Advanced ESL		
221-235	Oral BEST 58-64 Literacy BEST: 65 and above	Speaking: SPL 6 Reading/Writing: SPL 7
High Advanced ESL		
236-245	Oral BEST 65 and above	Speaking: SPL 7 Reading/Writing: SPL 8

(3e) Revisions (2006) to NRS Levels for Adult ESL Learners

Original Levels (1998)	Revised Levels (2006)
<i>Beginning ESL Literacy</i>	<i>Beginning ESL Literacy</i>
<i>Beginning</i>	<i>Low Beginning</i>
	<i>High Beginning</i>
<i>Low Intermediate</i>	<i>Low Intermediate</i>
<i>High Intermediate</i>	<i>High Intermediate</i>
<i>Low Advanced</i>	<i>Advanced</i>
<i>High Advanced</i>	

(3f) Evaluation of NRS Reading/Writing Descriptors from a Standard Setting Study (1 low, 4 high)

Judges' (n=10) Evaluation of NRS Portion of the Standard Setting Study

NRS Levels	Clarity of the NRS level <i>descriptors</i>	<i>Training</i> on setting the cut scores for NRS levels	<i>Materials</i> used in setting the cut scores for NRS levels	<i>Process</i> used in setting the cut scores for NRS levels	<i>Time</i> allotted for setting the cut scores for NRS levels
Mean	1.80	3.60	4.00	4.00	3.40
St. Dev	.789	.516	.000	.000	.843

Judges' (n=10) Evaluation of SPL Portion of the Standard Setting Study

SPLs	Clarity of the SPL <i>descriptors</i>	<i>Training</i> on setting the cut scores for SPLs	<i>Materials</i> used in setting the cut scores for SPLs	<i>Process</i> used in setting the cut scores for SPLs	<i>Time</i> allotted for setting the cut scores for SPLs
Mean	3.50	3.90	4.00	4.00	3.70
St. Dev	.527	.316	.000	.000	.675

(3f) Evaluation of NRS Reading/Writing Descriptors from a Standard Setting Study (1 low, 4 high)

Judges' (n=10) Evaluation of NRS Portion of the Standard Setting Study						
NRS Levels	Clarity of the NRS level <i>descriptors</i>	<i>Training</i> on setting the cut scores for NRS levels	<i>Materials</i> used in setting the cut scores for NRS levels	<i>Process</i> used in setting the cut scores for NRS levels	<i>Time</i> allotted for setting the cut scores for NRS levels	
Mean	1.80	3.60	4.00	4.00	3.40	
St. Dev	.789	.516	.000	.000	.843	
Judges' (n=10) Evaluation of SPL Portion of the Standard Setting Study						
SPLs	Clarity of the SPL <i>descriptors</i>	<i>Training</i> on setting the cut scores for SPLs	<i>Materials</i> used in setting the cut scores for SPLs	<i>Process</i> used in setting the cut scores for SPLs	<i>Time</i> allotted for setting the cut scores for SPLs	
Mean	3.50	3.90	4.00	4.00	3.70	
St. Dev	.527	.316	.000	.000	.675	

Outline

1. Background to the National Reporting System (NRS)
2. NRS and standards of psychometric quality for adult ESL assessments
3. NRS and the construct being measured by adult ESL assessments
4. **Wrap-up**

(4a) The Government, Adult ESL Assessments, and Standards of Psychometric Quality

The U.S. Dept. of Education **has** provided resources and taken actions to professionalize adult ESL assessment and now requires them to meet high technical standards.

However, to do so requires financial and technical resources that an under-resourced field with a low volume of examinees like adult ESL education (compared to K-12 education) may not be able to provide without additional support.

(4b) The Government, Adult ESL Assessments, and the Underlying Language Construct

All adult ESL assessments must be linked to the six NRS educational functioning level descriptors (for listening and speaking, or for reading and writing).

However, the current descriptors are very problematic. They need to be revised with the input of experts in applied linguistics (and other complimentary fields), or new ones developed that can be formally “linked” to these and meet government approval.

(4c) Summary (1/3)

While not neglecting the importance of how assessments are used, their social consequences, and the power that they represent, ...

(4d) Summary (2/3)

(1) applied linguists should expect and require that the adult ESL assessments used for accountability adhere to standards of psychometric quality, and need to investigate whether they do or not, and

(4e) Summary (3/3)

(2) since adult ESL tests claim to assess second language learning, applied linguists should contribute to the understanding of the constructs claimed to be underlying the assessments.

(Remember the gap analysis of 19 adult ESL assessments, where, with the exception of a few tests, there was no reference to second language acquisition or learning theory underlying the test constructs!)

Thank You!