

ECOLT, Washington DC, 6 November 2009

NRS Classification Consistency of Two Spoken English Tests

Rebecca Present-Thomas

Alistair Van Moere

Knowledge Technologies, Pearson

Organization of Presentation

- Overview of the National Reporting System (NRS)
- Introduction to the *BEST Plus Oral Proficiency Interview*
- Introduction to the *Versant English Test*
- Methodology
- Data
- Discussion

Overview of the NRS

- The U.S. Department of Education's adult education program enrolls 1.1 million students across the country each year
- Of these, half a million students participate in the English Literacy program
- The National Reporting System (NRS) is the system responsible for tracking the accountability of the adult education program

Overview of the NRS

- All adult education programs are required to report demographic information and NRS-approved test scores for all students.
- In order for a test to be approved by the NRS, it must be established that the test
 - has been designed to measure *educational gain*
 - has multiple forms that can be used for pre-test/post-test comparisons
 - measures content and skills that are consistent with the NRS Educational Functioning Levels

NRS Educational Functioning Levels

1. Beginning ESL Literacy
2. Low Beginning ESL
3. High Beginning ESL
4. Low Intermediate ESL
5. High Intermediate ESL
6. Advanced ESL
7. Exit criteria from the NRS

NRS-Approved Tests for English as a Second Language

- **CASAS**
 - Comprehensive Adult Student Assessment System
 - Provides reading, writing and listening scores
- **TABE CLAS-E**
 - Test of Adult Basic Education Complete Language Assessment System - English
 - Provides reading/writing and speaking/listening scores
- **BEST Plus and BEST Literacy**
 - Basic English Skills Test
 - Versant English Test (spoken)

Research Questions

1. What is the relationship between Versant scores and BEST Plus scores?
2. Can the Versant English Test appropriately classify test-takers according to NRS levels?
3. How consistently does the BEST Plus classify students according to NRS levels?
4. How consistently does Versant classify students according to NRS levels?

BEST Plus Oral Proficiency Interview

- Computer-adaptive
- 3-20 minutes
- Certified raters administer the test and assign scores on:
 - Listening comprehension
 - Language complexity
 - Oral communication
- Tasks & Topics include:
 - Picture description; constrained/open questions
 - Housing, Family, Jobs, Education, Weather
- Scores are immediately calculated by the test scoring software.



BEST Plus - NRS Concordance

NRS Level	BEST Plus Score	Score Range
Beginning ESL Literacy	≤ 400	313 points
Low Beginning ESL	401-417	17 points
High Beginning ESL	418-438	31 points
Low Intermediate ESL	439-472	34 points
High Intermediate ESL	472-506	35 points
Advanced ESL	507-540	34 points
Exit criteria from NRS	≥ 541	459 points

- The published SEM for the BEST Plus ranges from 20-30 points.

((())) Versant English Test



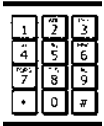
1. Candidate/Proctor downloads test from Versant web site and prints out



2. Candidate reviews test sheet & instructions (5min)



3. Candidate calls Versant Testing System – toll free number on test sheet



4. Candidate keys in unique (8-digit) Test Identification Number



5. Candidate takes test (~15min)



6. Candidate/Proctor retrieves score from Versant web site (within minutes of completing test)

Versant Test Structure

VERSANT ENGLISH TEST (DEMO)

REMINDER: The test begins when you say your name. If you hang up before you complete the test, the test cannot be graded. You cannot reuse the Test Identification Number.

Call: 1-415-738-3800

Thank you for calling the Versant testing system.
Please enter your Test Identification Number on the telephone keypad.
Now, please say your name.
Now, please say the city and country you are calling from.
Now, please follow the instructions for Parts A through F.

Test Identification Number (TIN)

8383 1210

Expires: October 30, 2010

PART	TASK	TEST DETAILS
A	Reading	<p><i>Please read the sentences as you are instructed.</i></p> <ol style="list-style-type: none"> 1. Traffic is a huge problem in Southern California. 2. The endless city has no coherent mass transit system. 3. Sharing rides was going to be the solution to rush-hour traffic. 4. Most people still want to drive their own cars, though. 5. Larry's next door neighbors are awful. 6. They play loud music all night when he's trying to sleep. 7. If he tells them to stop, they just turn it up louder. 8. He wants to move out of that neighborhood. 9. My aunt recently rescued a dog that was sick. 10. She brought her home and named her Margaret. 11. They weren't sure she was going to live, but now she's healthy. 12. I just wish she could get along better with their cat.
B	Repeat	<p><i>Please repeat each sentence that you hear.</i></p> <p>Example: a voice says, "Leave town on the next train." and you say, "Leave town on the next train."</p>
C	Questions	<p><i>Now, please just give a simple answer to the questions.</i></p> <p>Example: a voice says, "Would you get water from a bottle or a newspaper?" and you say, "a bottle" or "from a bottle!"</p>
D	Sentence Builds	<p><i>Now, please rearrange the word groups into a sentence.</i></p> <p>Example: a voice says, "was reading" ... "my mother" ... "her favorite magazine" and you say, "My mother was reading her favorite magazine."</p>
E	Story Retelling	<p><i>You will hear three brief stories. Each story will be spoken once, followed by a beep. When you hear the beep, you will have 30 seconds to retell the story in English. Try to retell as much of the story as you can, including the situation, characters, actions, and ending. You will hear another beep at the end of the 30 seconds.</i></p>
F	Open Questions	<p><i>You will hear two questions about family life or personal choices. Each question will be spoken twice, followed by a beep. When you hear the beep, you will have 40 seconds to answer the question. You will hear another beep at the end of the 40 seconds.</i></p>

Thank you for completing the test.

70 - 29366 - 1
© 2009 Pearson Education, Inc. or its affiliate(s). All rights reserved. Ordinate and Versant are trademarks. In the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s).

Tasks:

- Read sentences*
- Repeat sentences*
- Short questions*
- Sentence builds*
- Story Retelling*

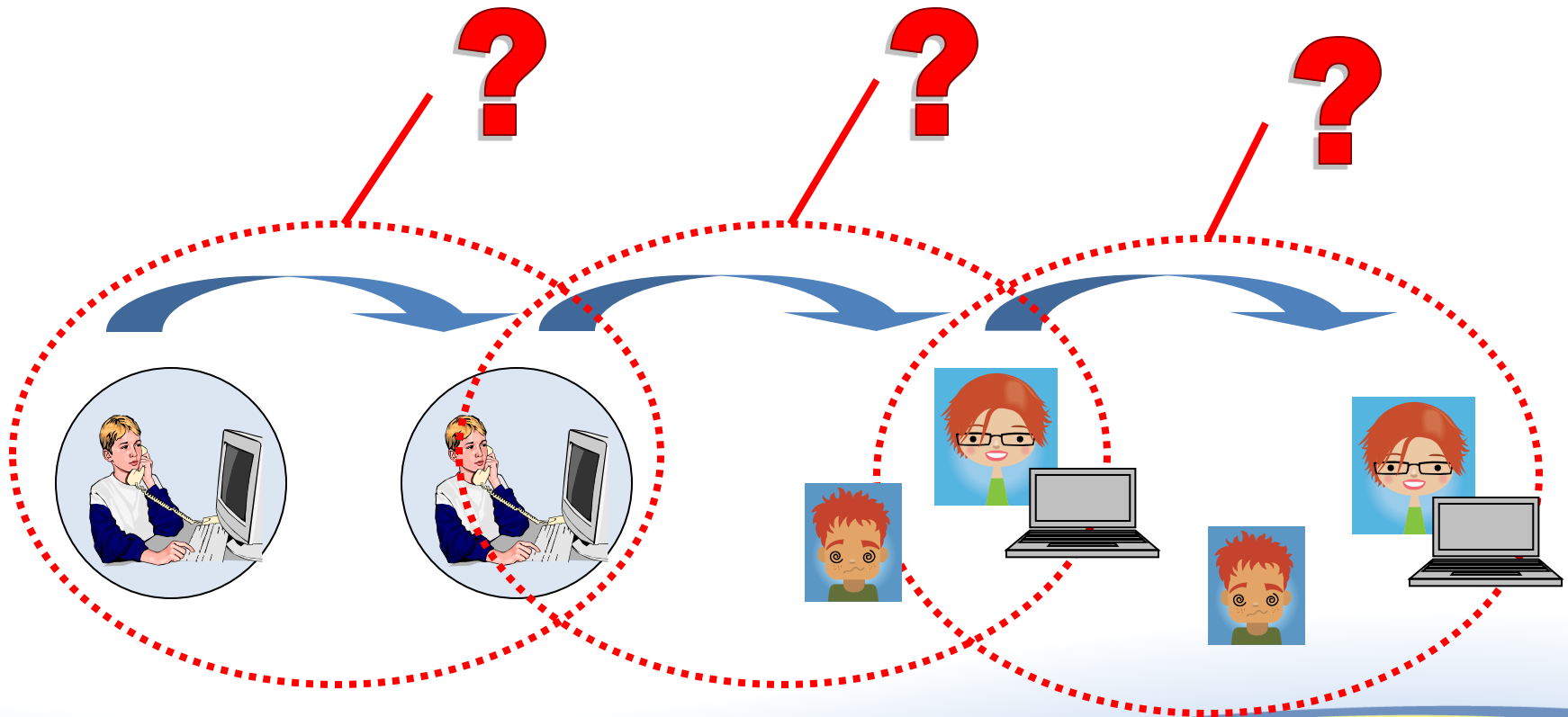
Scores:

- Sentence mastery*
- Vocabulary*
- Pronunciation*
- Fluency*
- Overall Speaking (20-80)*

Methodology

2 automatically scored
Versant tests

2 computer-adaptive
BEST Plus interview tests



Participants

Of the 151 participants:

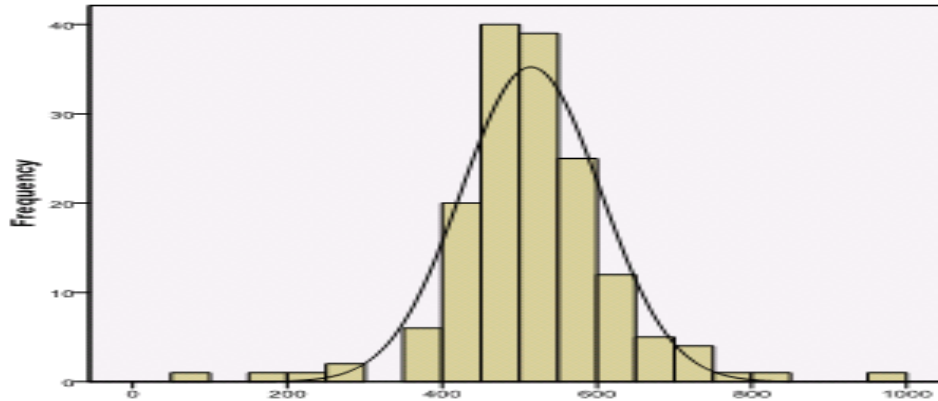
- 51% were enrolled in an adult education program in New York
- 49% were enrolled in or eligible for adult education programs in California
- More than 25 language backgrounds represented

BEST Plus Raters

Rater	Number of Tests	Education	Experience
1	101	M.A. TESOL	5 years teaching ESL in a variety of settings
2	93	M.A. Linguistics	2 years teaching ESL
3	67	B.A. Language and Literature	Over 20 years experience teaching ESL in adult school settings.
4	39	B.A. Psychology	2 years language assessment experience
5	2	M.A. Applied Linguistics	5 years teaching ESL in a variety of settings

- All 5 raters participated in a 6-hour training workshop with a *BEST Plus* certified trainer from the Center for Applied Linguistics

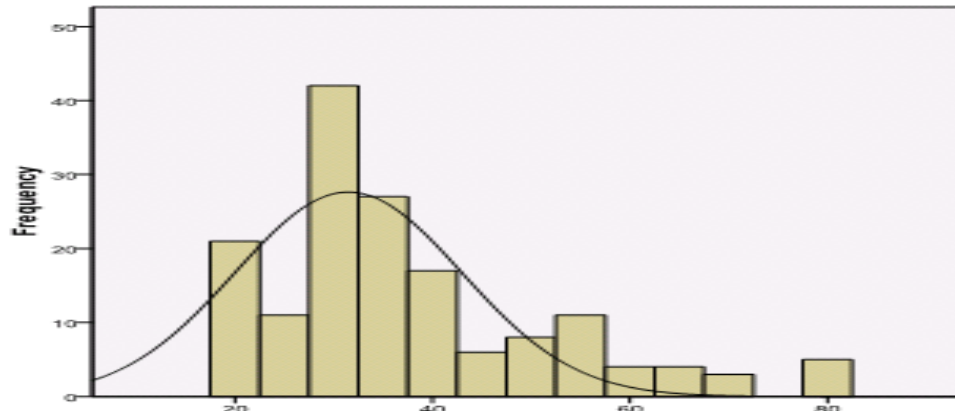
Distribution of Scores (n = 151)



BEST Plus Histogram

Mean = 514.9

Std. Dev. = 105.3

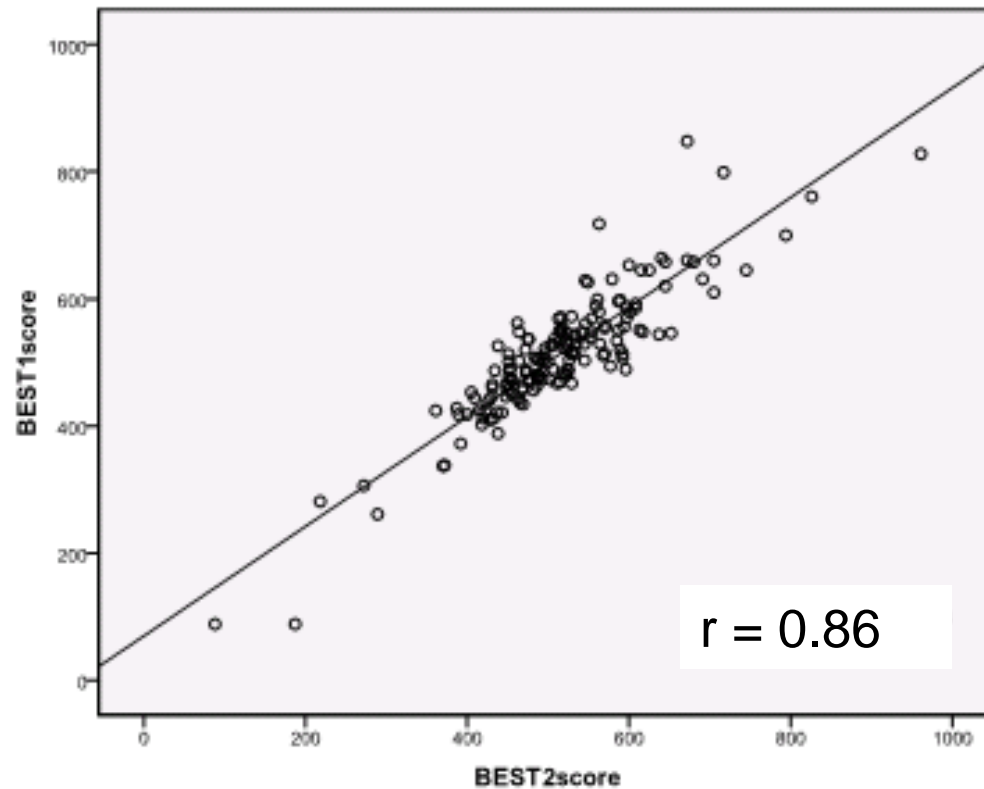


Versant Histogram

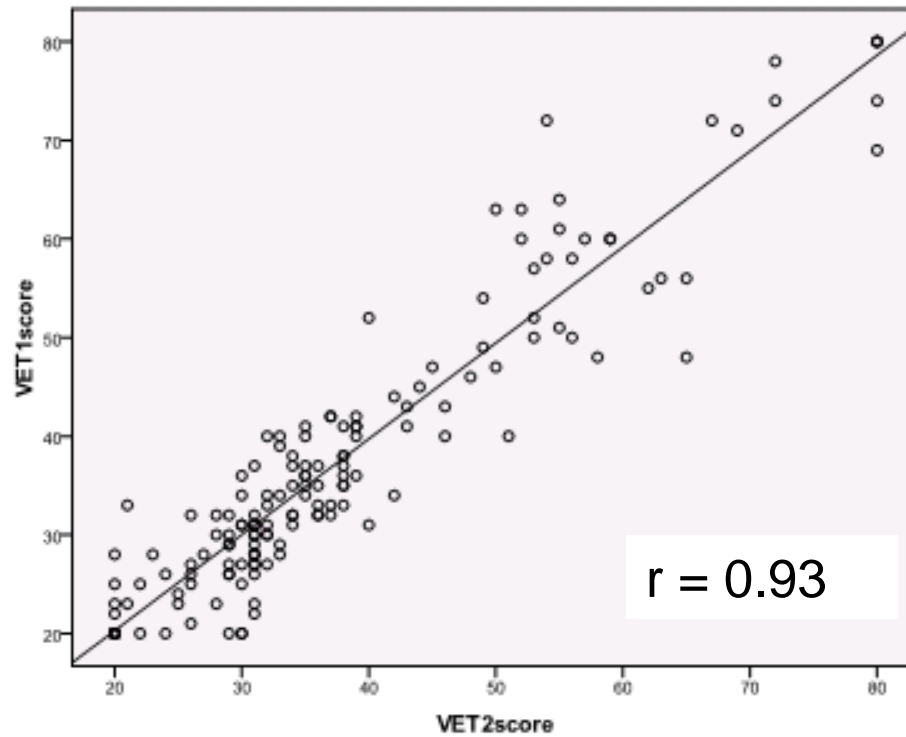
Mean = 37.5

Std. Dev. = 14.4

BEST Plus Test-Retest Scatterplot



Versant Test-Retest Scatterplot



Test-Retest Reliability

Pearson coefficients					Spearman coefficients			
	BEST 1	BEST 2	VET 1	VET 2	BEST 1	BEST 2	VET 1	VET 2
BEST 1	1				1			
BEST 2	.90	1			.86	1		
VET 1	.78	.78	1		.85	.83	1	
VET 2	.74	.76	.94	1	.80	.81	.93	1

BEST-Versant Concordance

- Using average BEST and average Versant scores, a regression analysis was used to predict Versant scores from BEST scores.

$$\text{Versant} = (\text{BEST} * 0.15) - 40.754$$

- This produced a score concordance table

<i>BEST Plus</i>	((())) VERSANT™
380	16.2
400	19.2
420	22.2
440	25.2
460	28.2
480	31.2
500	34.2
520	37.2
540	40.2
560	43.2
580	46.2

NRS Concordances

NRS	Levels	BEST Plus	((())) VERSANT™
1	Beginning	<= 400	<= 19
2	Low Beginning	401 - 417	20 - 21
3	High Beginning	418 - 438	22 - 25
4	Low Intermediate	439 - 472	26 - 30
5	High Intermediate	473 - 506	31 - 35
6	Advanced	507 - 540	36 - 40
7	Above NRS	541 +	41 +

- The SEM for the Versant English Test is 2.9 points

NRS Classification Consistency

NRS	<i>BEST</i>		Agreement Raw-Percent		<i>VERSANT</i>		Agreement Raw-Percent	
	<i>1</i>	<i>2</i>			<i>1</i>	<i>2</i>		
1	3	5	2	67%	7	7	5	71%
2	4	2	0	0%	3	3	1	33%
3	9	13	2	22%	7	12	4	33%
4	20	22	9	45%	25	29	12	41%
5	31	25	15	48%	42	32	14	44%
6	29	29	10	34%	24	20	7	35%
7	55	55	43	78%	43	48	40	83%

NRS Classification Consistency

NRS	BEST 1 & BEST 2	BEST 1 & Versant 1	BEST 2 & Versant 2
1	67%	0%	29%
2	0%	0%	33%
3	22%	22%	14%
4	45%	35%	44%
5	48%	39%	26%
6	34%	31%	33%
7	78%	75%	88%

BEST Plus

Classification Consistency

		<i>BEST 1</i>								
<i>BEST 2</i>	NRS	1	2	3	4	5	6	7		
	1	67		33						100%
	2			100						100%
	3	33	11	22	33					100%
	4		5	20	45	15	15			100%
	5			3	19	48	19	10		100%
	6			3	7	24	34	31		100%
	7				4		18	78		100%

Versant

Classification Consistency

		((())) VERSANT™ 1								
		NRS	1	2	3	4	5	6	7	
((())) VERSANT™ 2	1	71				29				100%
	2		33	33	33					100%
	3	17	8	33	25	17				100%
	4			7	41	52				100%
	5		3		19	44	31	3		100%
	6				5	50	35	10		100%
	7					2	15	83		100%

The “Classification Consistency” Approach

- Claims for linkage are usually conducted on the basis of collective expert opinion
- However, linkage is only valid if scores classify consistently.
- We submit that “classification consistency” is an under-utilized methodology for independently validating claims for linkage to scales of reference.

Implications for NRS

- Based on the concordance approach given here, the National Reporting System (NRS) would be served equally as well by using Versant and/or BEST Plus
- Why did we not use expert judgment for our mapping?
 - Emphasis placed on classification consistency

Next steps

- Documentation of a validation of the NRS scales
- Documentation of the mapping of BEST Plus to NRS

NRS Descriptors

NRS	Listening and Speaking Descriptors	CEFR
1.	Individual cannot speak or understand English, or understands only isolated words or phrases.	<A1 25
2.	Individual can understand basic greetings, simple phrases and commands. ... Demonstrates little or no control over grammar.	
3.	Individual can understand common words, simple phrases, and sentences containing familiar vocabulary, spoken slowly with some repetition. ... Shows limited control of grammar.	26 A1 35
4.	Individual can understand simple learned phrases and limited new phrases containing familiar vocabulary spoken slowly with frequent repetition; ... has some control of basic grammar.	
5.	Individual can understand learned phrases and short new phrases containing familiar vocabulary spoken slowly and with some repetition; ... inconsistent control of more complex grammar.	36 A2 41
6.	Individual can understand and communicate in a variety of contexts related to daily life and work. Shows control of basic grammar but has difficulty using more complex structures. Has some basic fluency of speech.	

Implications

- We do expect measurable gains to take place in adult education courses ...
 - But inspection of the NRS bands reveals they are so narrow as to be difficult to map using standardized proficiency test scores

Alternatives?

- Specific syllabus-based achievement testing
- Gain scores on approved tests (e.g. 50 points on the BEST Plus, 3 points on the Versant, etc)

Thank you