

# Developing and Validating an Automated Test of Spoken MSA

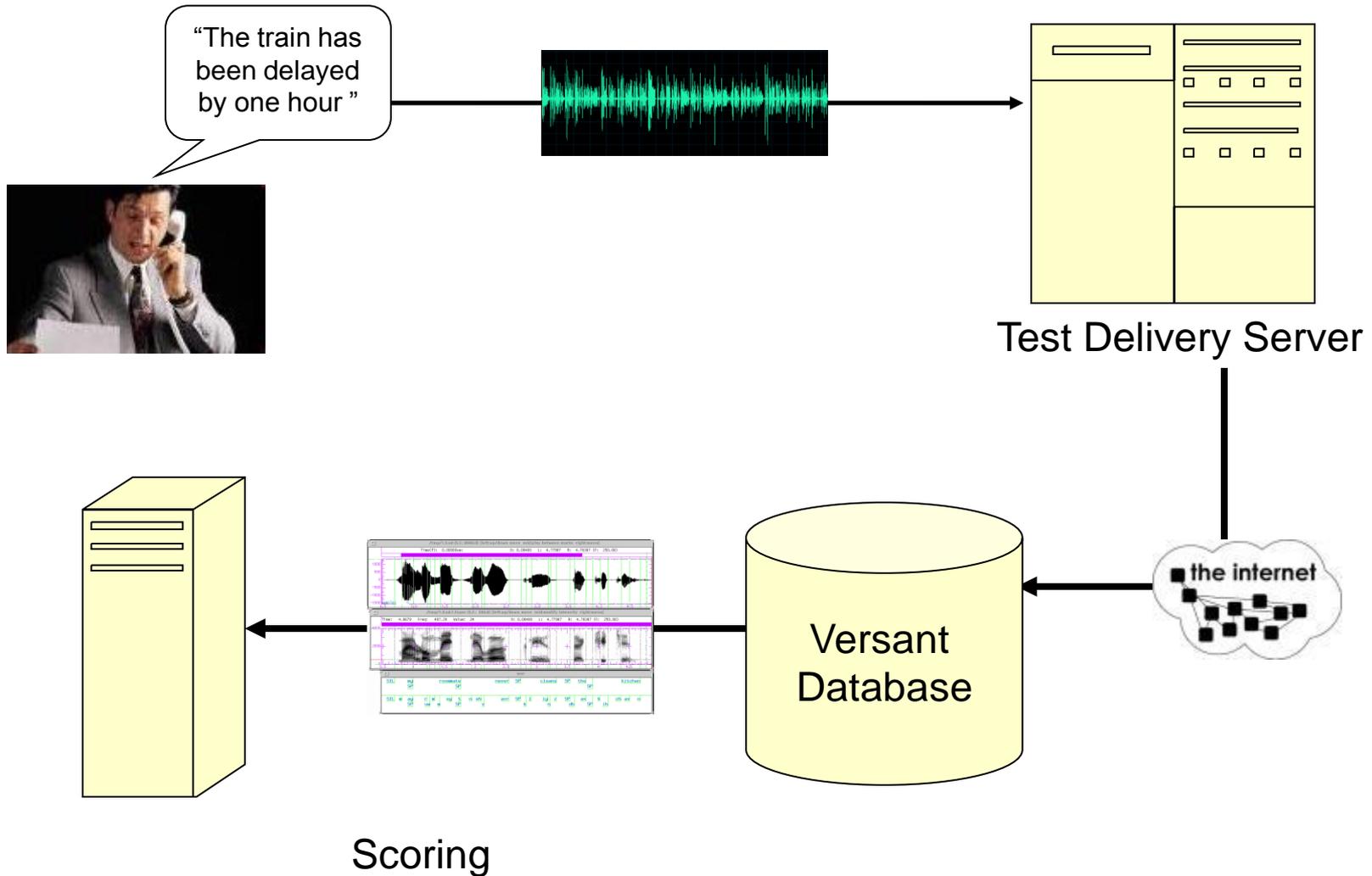
*ECOLT at GWU  
Washington, DC  
8 November 2008*

Jared Bernstein  
Jian Cheng  
Ulrike Pado  
Masa Suzuki  
Mohamed Al-Saffar  
Alistair van Moere  
***Pearson Knowledge Technologies***  
*Palo Alto, California*

# Outline

1. How Versant tests work
2. Versant Arabic Test (development)
3. Validation evidence
4. Predictive accuracy

# How Versant tests work



# Versant Arabic Test

- DLI purpose
  - ~1000 students at DLI need predictive speaking tests
- Requirements
  - Accurate test of Arabic listening & speaking
  - Convenient to use at DLI and worldwide
  - Suitable for repeated formative testing
  - High peak capacity for mass screening

# Versant Arabic Test

## Test Structure

*Part A: Reading*

*Part B: Repeat -1*

*Part C: Short Answers*

*Part D: Sentence Builds*

*Part E: Repeat -2*

*Part F: Passage Retelling*

### VERSANT ARABIC TEST

**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: 1.800.335.6393**

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**

2571 6232

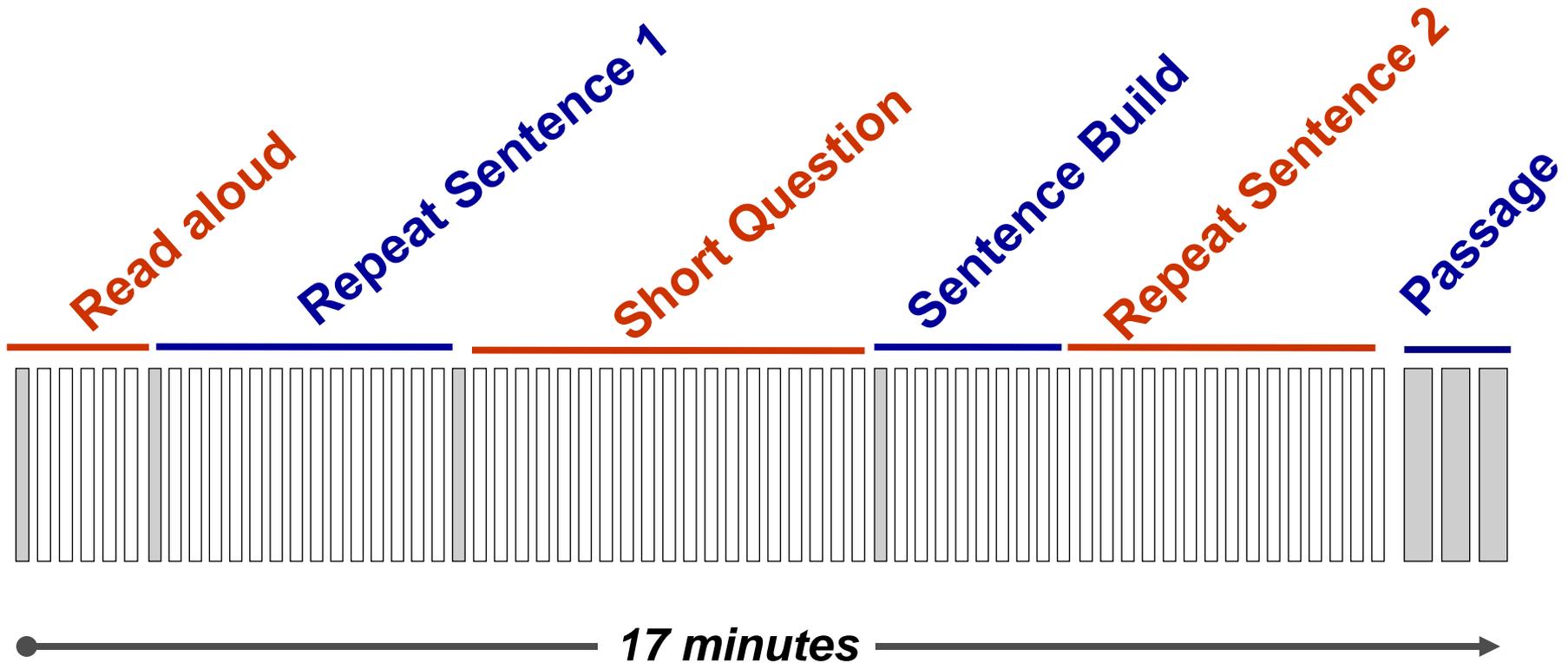
Expires: September 10, 2009

PART	TASK	TEST DETAILS
A	Reading	<p>Please read the sentences as you are instructed.</p> <ol style="list-style-type: none"> <li>1. الرياض مدينة جميلة وجوها دافئ</li> <li>2. هواؤها جاف وشوارعها واسعة</li> <li>3. مبانيها عالية وحدائقها خضراء</li> <li>4. يقصدها الكثير من السياح من مختلف البلدان</li> <li>5. سكان المدينة يحبون الحدائق العامة</li> <li>6. يتمتعون بمناظر الأشجار والمياه</li> <li>7. منهم من يأخذ أولاده ليلعبوا في الحدائق</li> <li>8. ومنهم من يتمشى ويتمتع بجمال الطبيعة</li> </ol>
B	Repeat	<p>Please repeat each sentence that you hear.</p> <p>Example: a voice says, "الأعياد تساعد في التسلية والراحة" and you say, "الأعياد تساعد في التسلية والراحة"</p>
C	Questions	<p>Now, please just give a simple answer to the questions.</p> <p>Example: a voice says, "الطالب يدرس ماذا يفعل الأستاذ؟" and you say, "هو يدرس" or "يدرس"</p>
D	Sentence Builds	<p>Now, please rearrange the word groups into a sentence.</p> <p>Example: a voice says, "الأولاد ... يغضون الأكل ..." and you say, "الأولاد يغضون الأكل خارج البيت"</p>
E	Repeat	<p>Please repeat each sentence that you hear.</p> <p>Example: a voice says, "الأعياد تساعد في التسلية والراحة" and you say, "الأعياد تساعد في التسلية والراحة"</p>
F	Passage Retelling	<p>You will hear several brief passages in Arabic. After each passage, you will hear a beep and then you will have 30 seconds to retell it in Arabic as best you can. Try to retell as much of the passage as you can in Arabic, including the important details.</p>

Thank you for completing the test.

Versant Arabic Test - 71 - 32455 - 1  
© 2008 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).

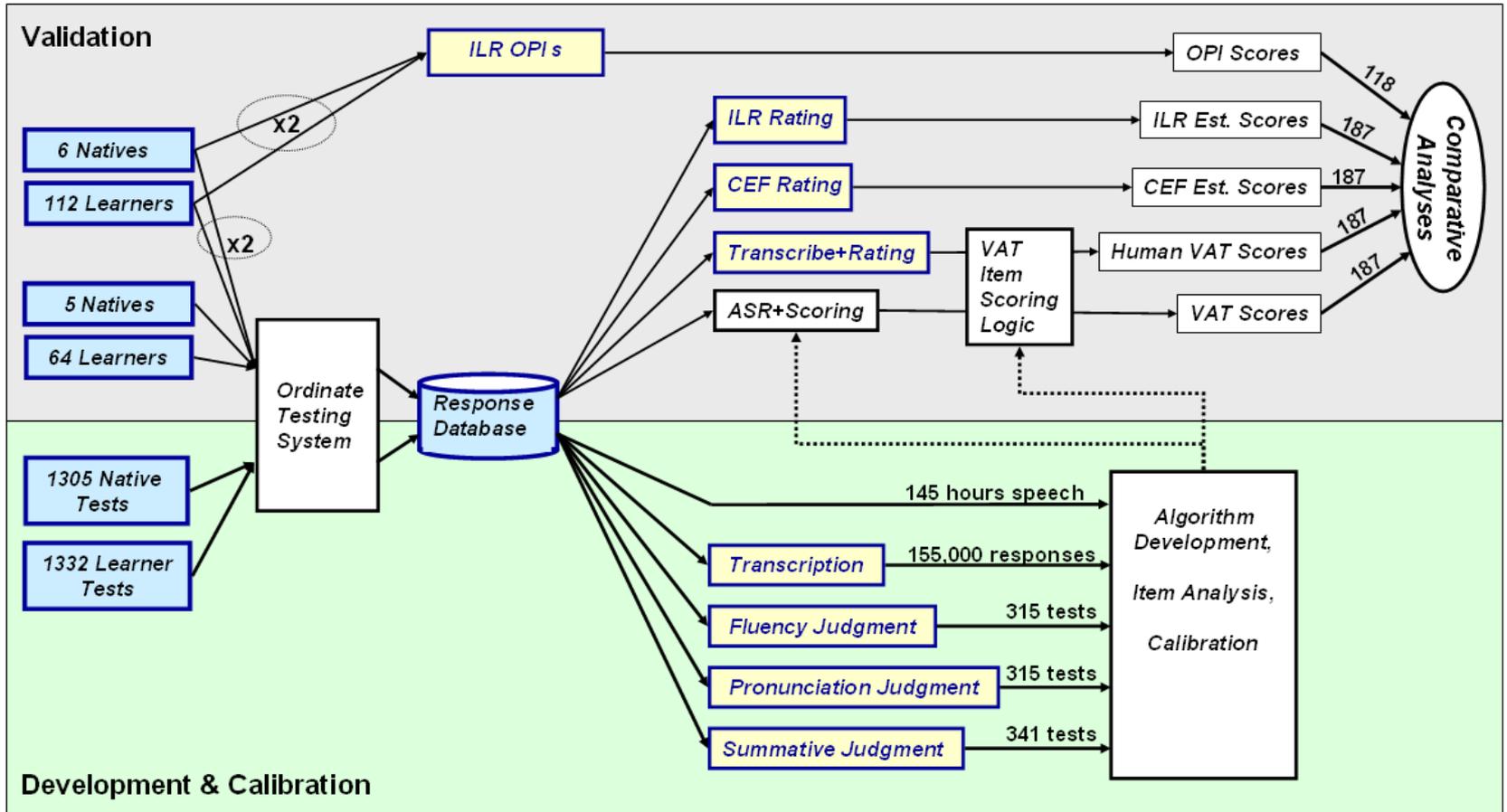
# Schematic test structure



# Prompt recordings

<b>Country</b>	Egypt	Iraq	Jordan	Morocco	Lebanon	Palestine	Syria
<b>Voices</b>	F, M	F, M	M	F	M	F, M	F, M

# How Versants are developed



# Pilot data sources

Native Data					
Egypt	Syria	Iraq	Palestine	Other	Total
484	281	179	187	517	1648

Learner Data		
DLI	Non-DLI	Total
1120	552	1672

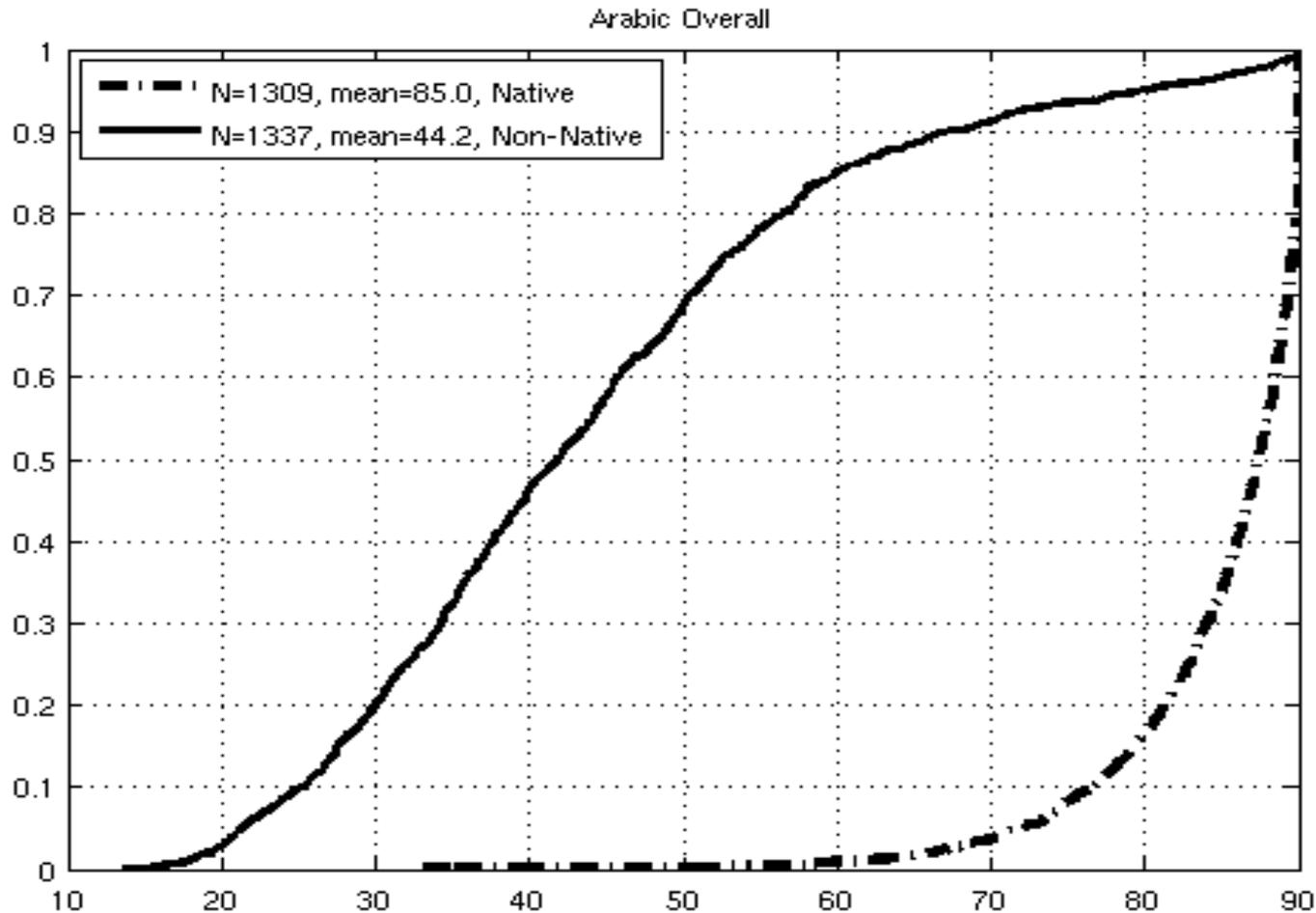
# Reliability

<b>Score</b>	<b>Split-Half Reliability (N = 134)</b>	<b>Test – Retest Reliability (N = 100)</b>
<b>Overall</b>	0.98	0.97
<b>Sentence Mastery</b>	0.97	0.96
<b>Vocabulary</b>	0.89	0.82
<b>Fluency</b>	0.97	0.96
<b>Pronunciation</b>	0.96	0.94

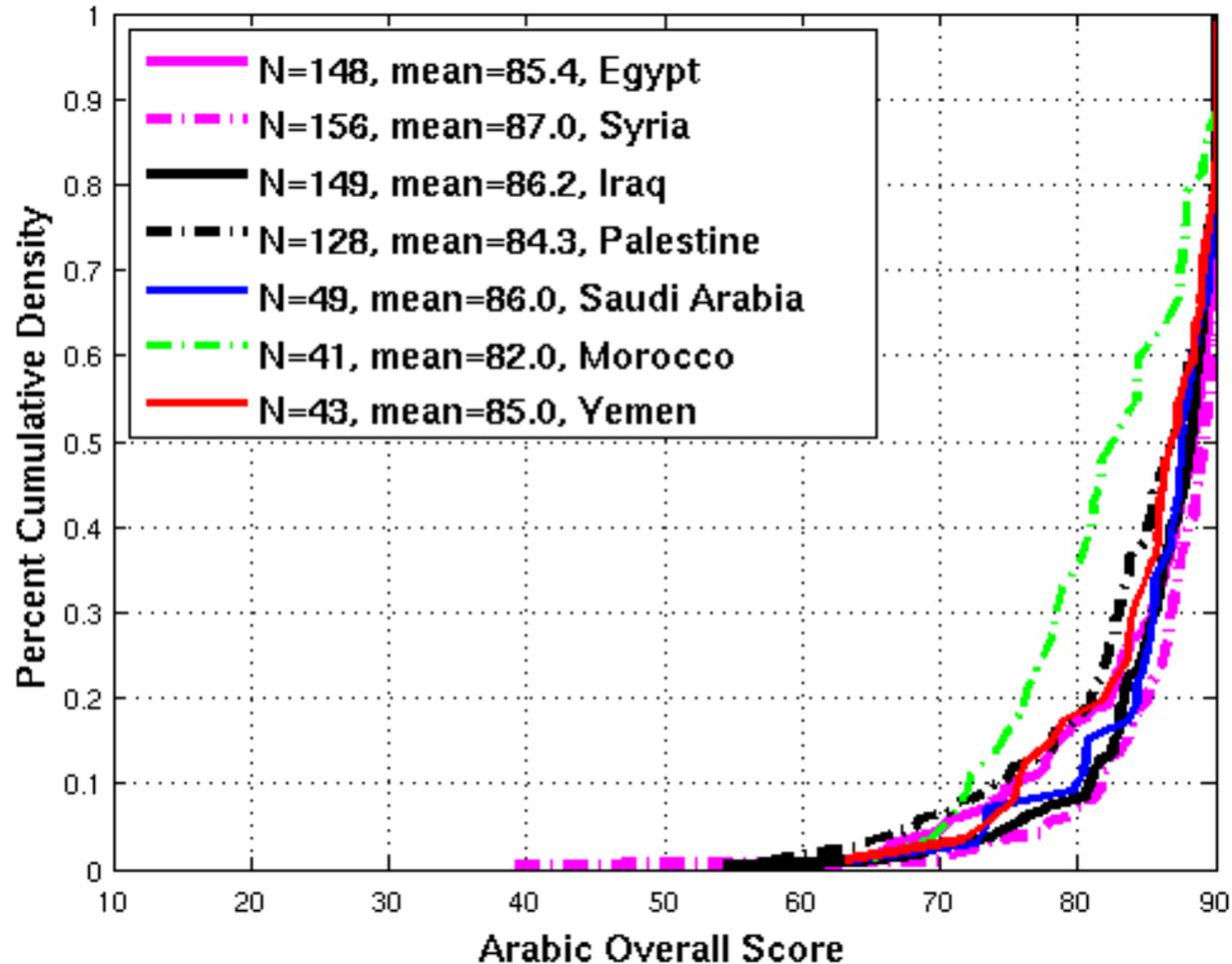
# Machine – Human Comparison

<b>Score</b>	<b>Correlation (N = 134)</b>
<b>Overall</b>	0.97
<b>Sentence Mastery</b>	0.97
<b>Vocabulary</b>	0.96
<b>Fluency</b>	0.84
<b>Pronunciation</b>	0.83

# Native ~ Non-Native Scores

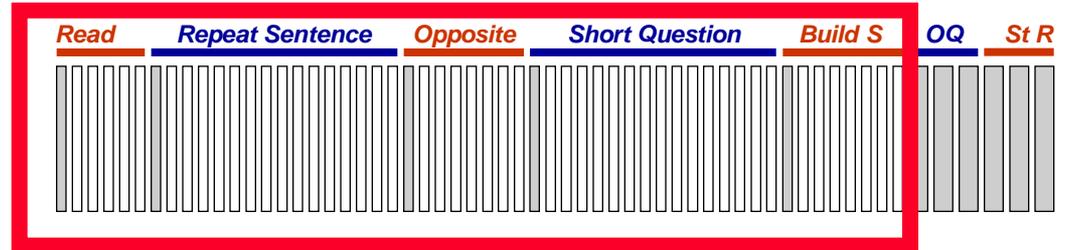


# Natives by Countries

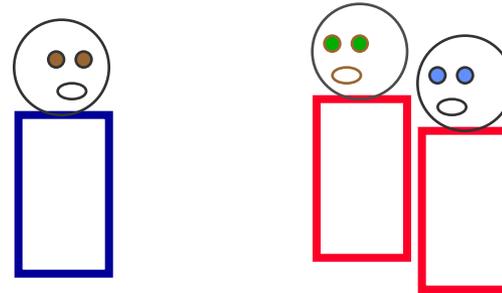


# Concurrent Validation

## 1. Versant Machine Scores



## 2. ILR-OPI Human Interview Scores

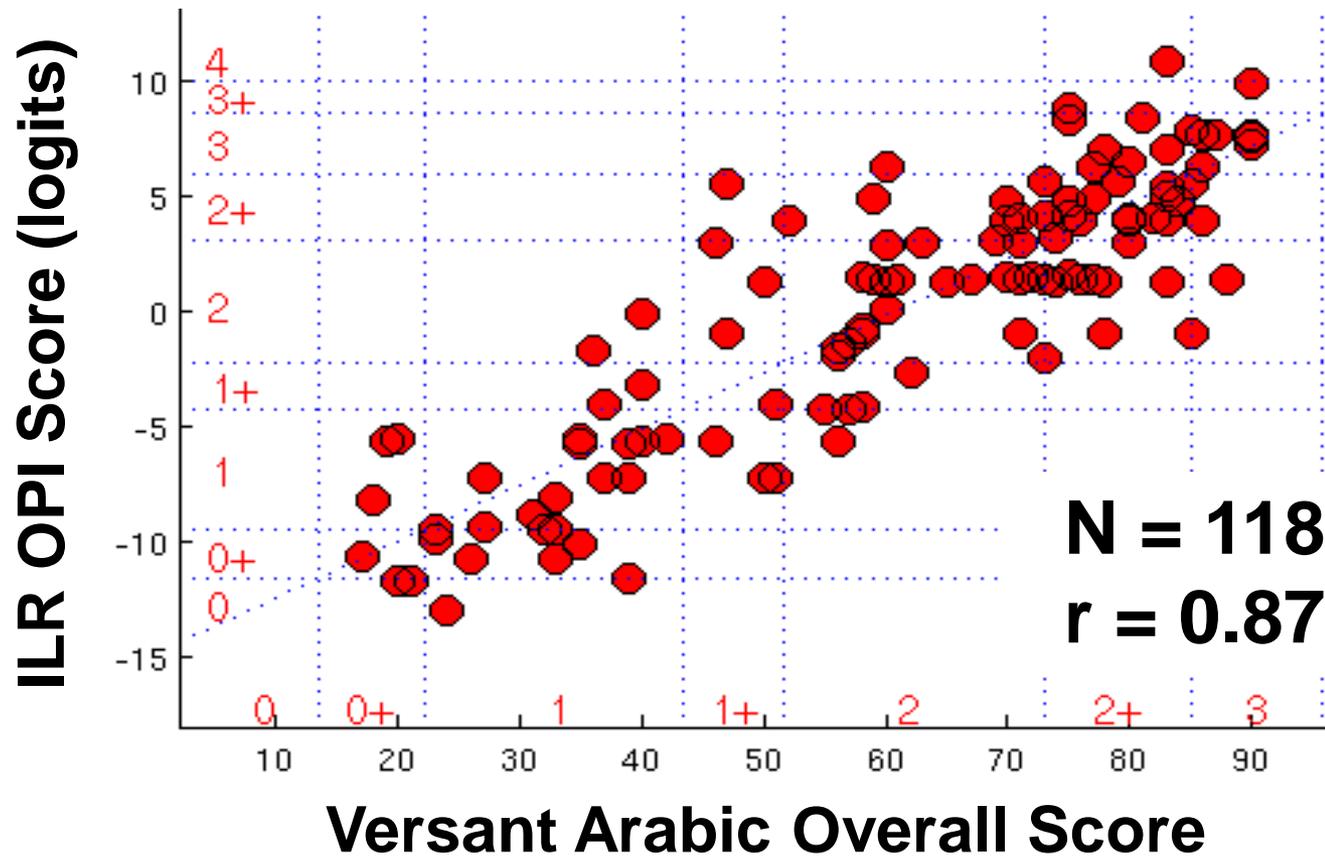


# Rater-rater & rater-OPI correlations

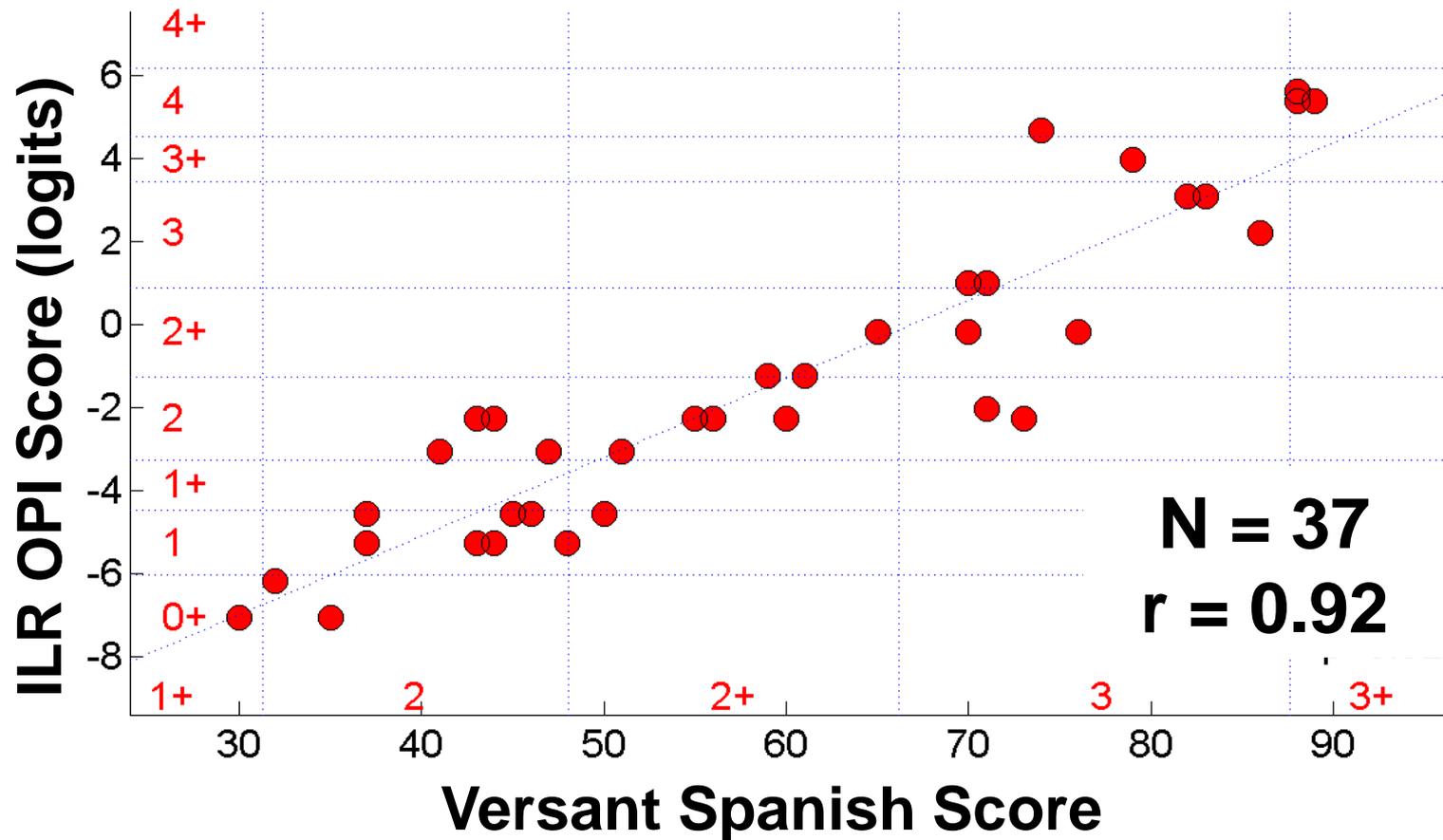
Interview/Rater	N	Rater-rater Correlation	Rater-OPI Correlation
A	85	0.88	0.89
B	84	0.86	0.86
C	92	0.88	0.90
D	76	0.89	0.91
E	49	0.92	0.93
F	53	0.90	0.92
G	33	0.84	0.86
Average	472	0.88	0.90

*In all cases, rating different OPI performances*  
Rater-based prediction range: 0.84-0.93

# How Versants Compare to OPIs (3)



# Spanish Data: Versant ~ OPI

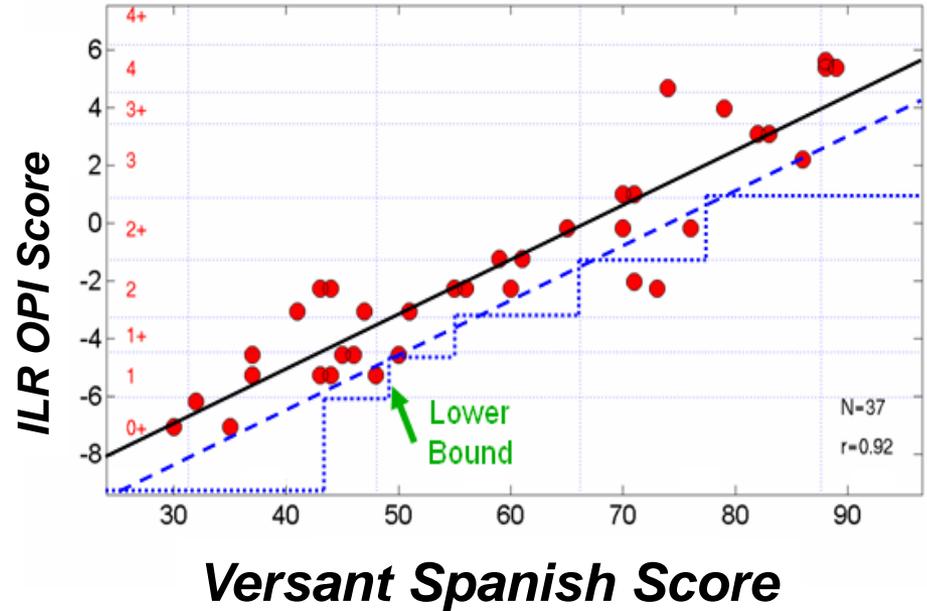
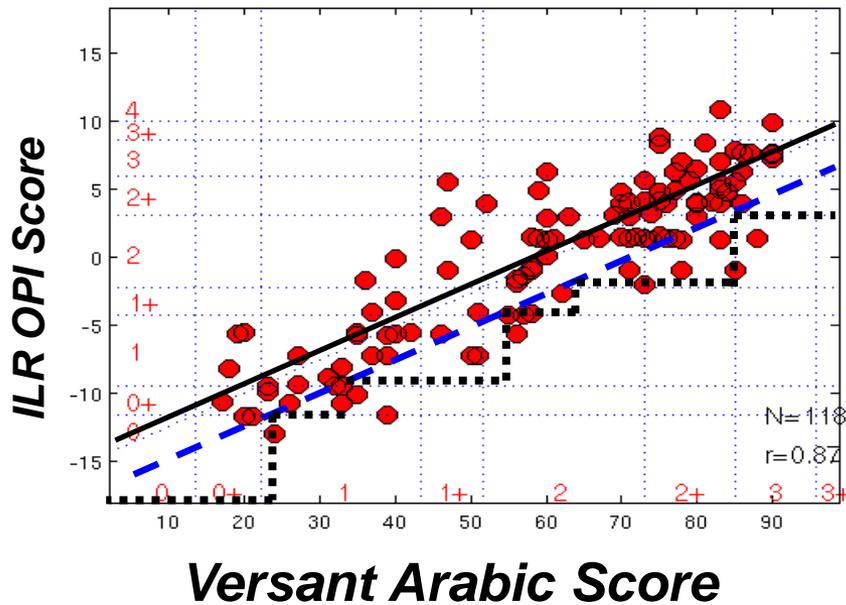


# Concordance Table

SST Overall Score	Best Estimate of ILR Score	$\geq$ ILR Score with 80% Confidence
20	0	0
21- 35	0+	At least 0+
36 - 43	1	At least 0+
44 - 49	1+	At least 1
50 - 55	2	At least 1+
56 - 60	2	At least 2
61 - 66	2+	At least 2
67 - 71	2+	At least 2+
72 - 77	3	At least 2+
78 - 80	3	At least 3

# ILR triage steps

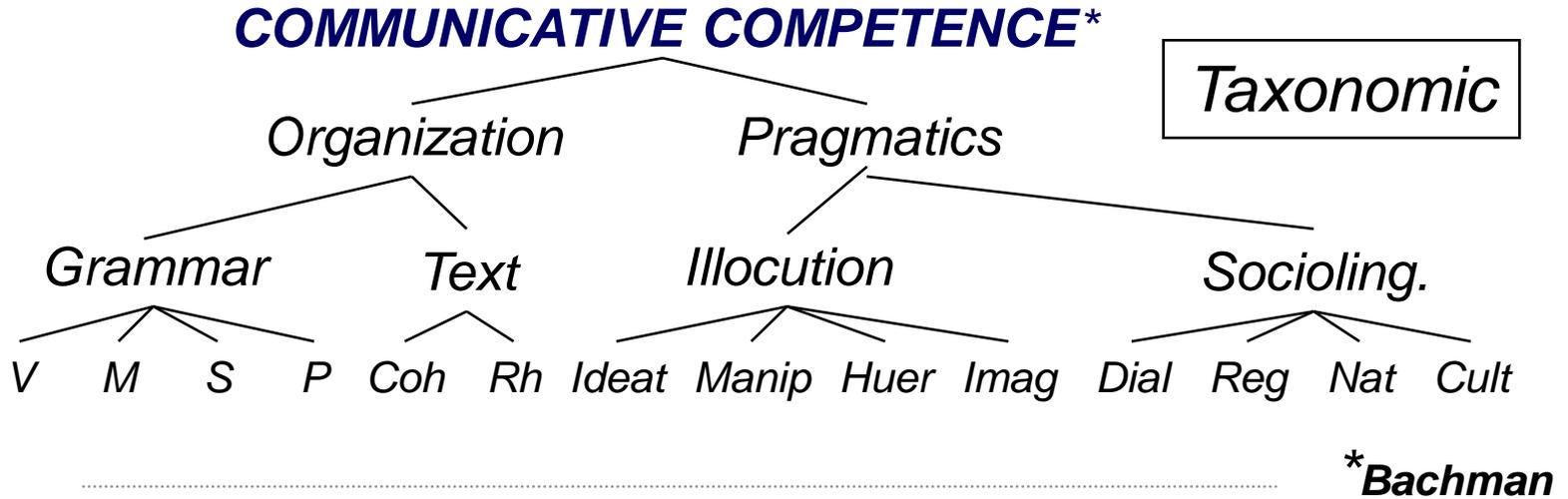
- Versant scores provide ILR level floors (twice) with 92% accuracy



$$\text{LowerBound(ILR)} = \text{ILR} - (t\text{-score})(\text{standard error of the estimate})$$

For 80% confidence, 36 df:  $t = 0.85$  (one tailed)

# Construct Comparison



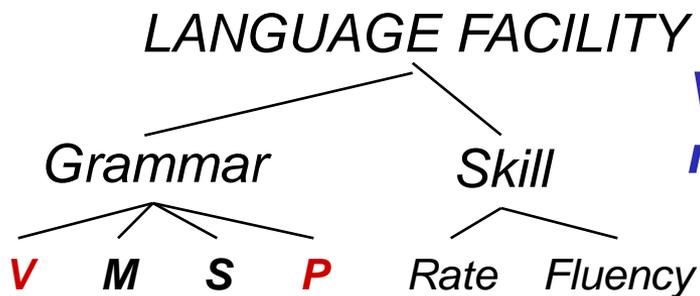
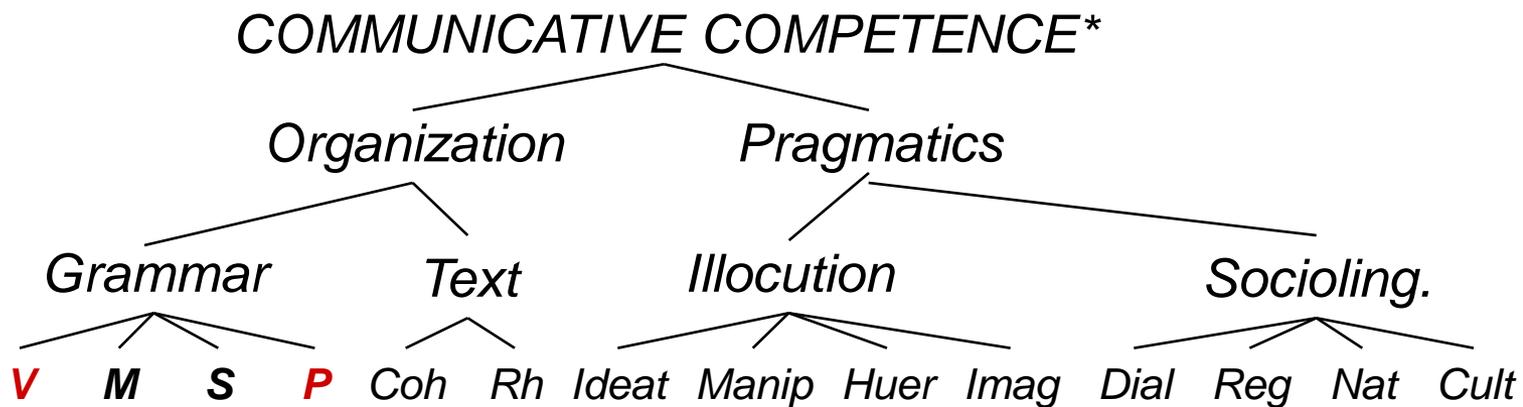
## Construct Comparison

**OPI Construct:** Oral Proficiency as manifest in an Oral Proficiency Interview, is compatible with *communicative competence* as reflected in the functional level and/or complexity of content accurately produced.

**Versant Construct:** *facility in spoken language* – the ability to understand spoken language and speak appropriately in response at a conversational pace on everyday topics.

# Shared variance

**0.87**  
**prediction**



**Versant tests contain sufficient material for predictive scoring**

# Summary

- Versant Arabic Test (VAT) is in operation
- Based on a large and wide body of transcribed spoken material
- VAT is available on demand
- VAT can triage or screen for OPI tests
  
- Useful score correspondence may be provided even when constructs differ

# النهاية