

# Psycholinguistics in Computerized Language Testing

*AAAL/CAAL in Montreal  
June 20, 2006*

---

*Masa Suzuki, Jennifer Balogh, and Jared Bernstein  
Ordinate Corporation/Harcourt Assessment*

## Two Main Objectives

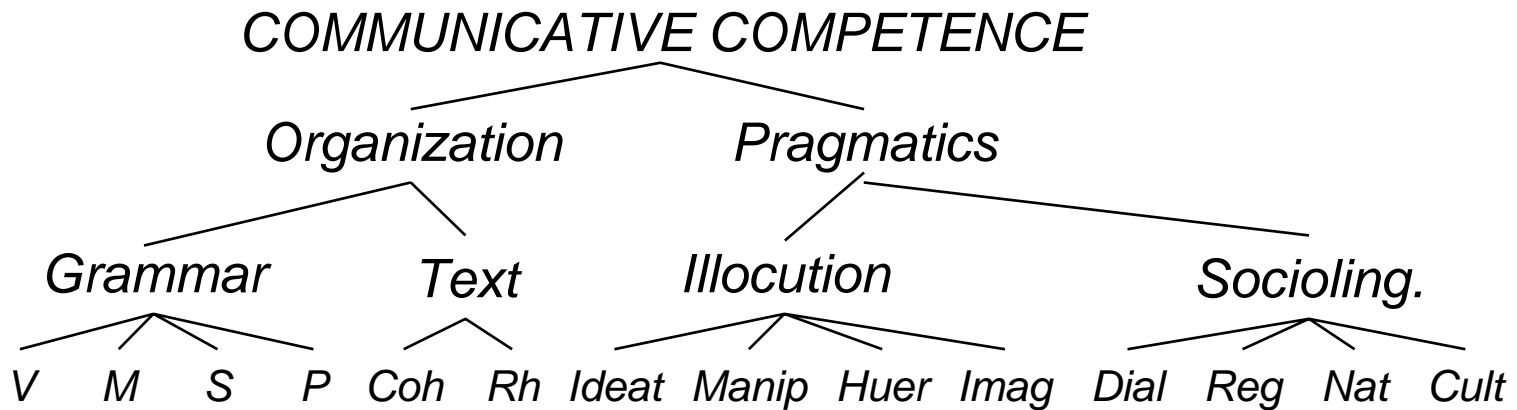
- Psycholinguistic-based Construct
- Automated Spoken Language Tests

## Dr. Jan Hulstijn's Plenary Speech

- Reliable (precise) measurement is hard to obtain in quasi-authentic tasks but is easy to obtain in non-authentic (discrete) task
- Important to assess the core oral language proficiency (language-specific abilities)
- Assess both “what” and “how well”

# L2 Language Proficiency Model

*OPI: Taxonomic*



*Bachman (1990)*

## Some Characteristics of Spoken Language

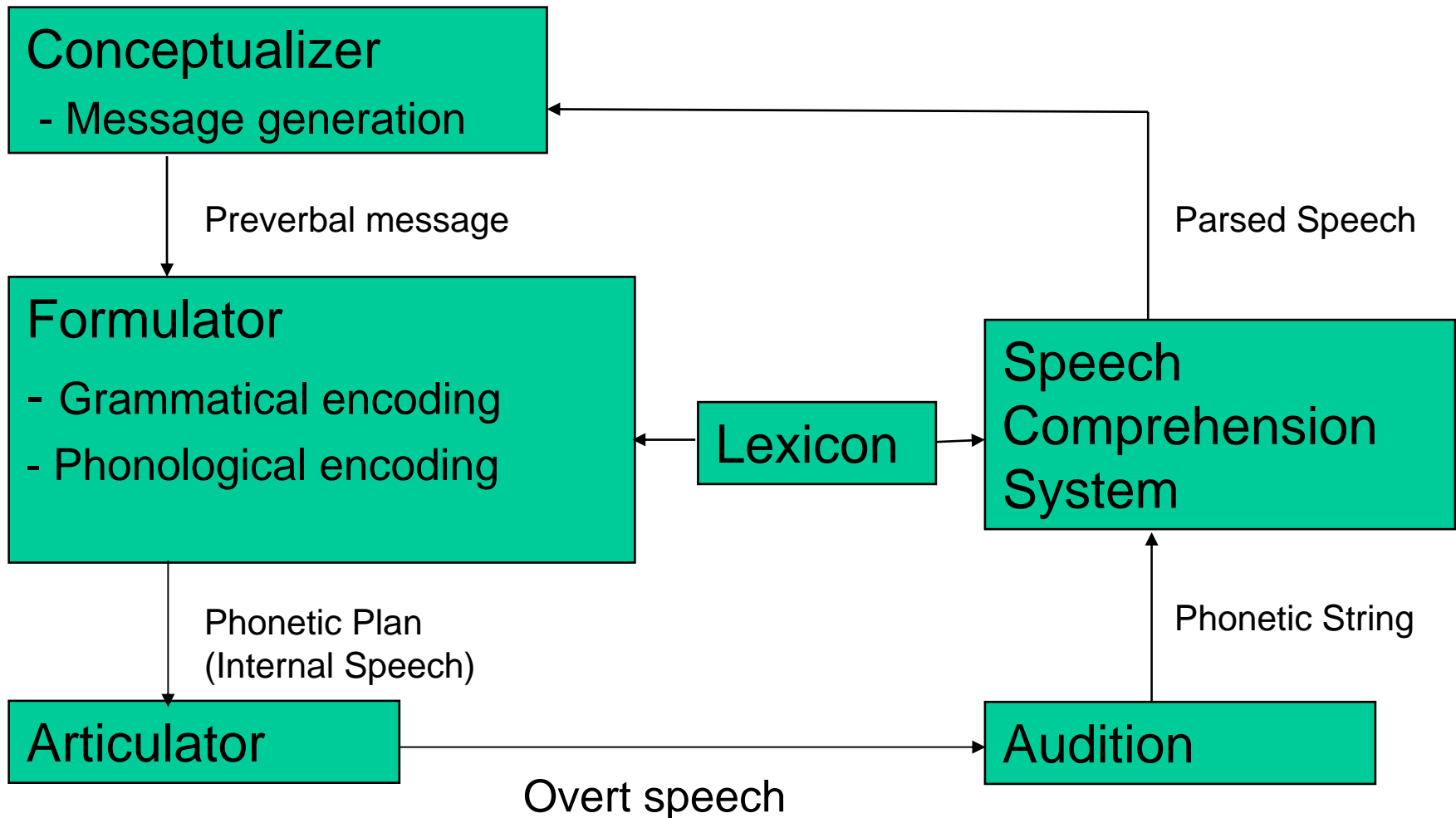
- Real-Time like ping-pong and unlike email
- Both “listening” and “speaking”
- Immediate responses required
- Spontaneous

→ If one cannot do these efficiently, he/she will mostly likely be labeled as “non-fluent”.

## What drives spoken performance

- No matter how good L2 speakers are at e.g. sociolinguistic competence and/or strategic competence, if they do not possess the ability to understand, process, and/or to produce the spoken language efficiently, the other competences are of limited use.
- Psycholinguistic aspect (=language comprehension/production skills) is the **core**.

# Levitt's Model (1989)



# Other Psycholinguistic Theories

- Chunking
- Automaticity (Schneider & Shiffrin, 1977; McLaughlin, 1987)

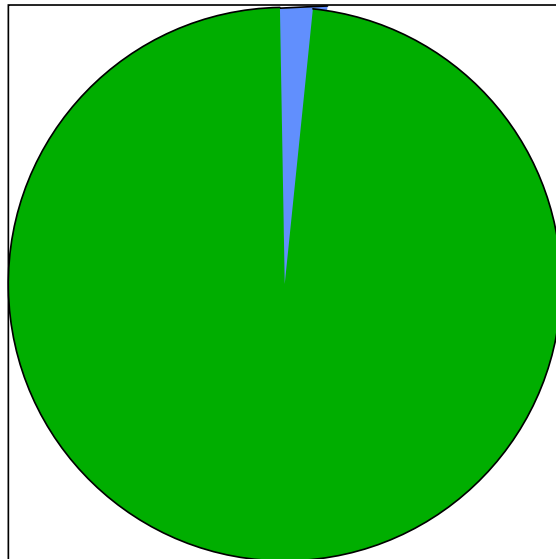
## Characteristics of Automaticity

- Rapid
- Effortless
- Efficient
- Implicit
- Unconscious
- No attention
- Procedural

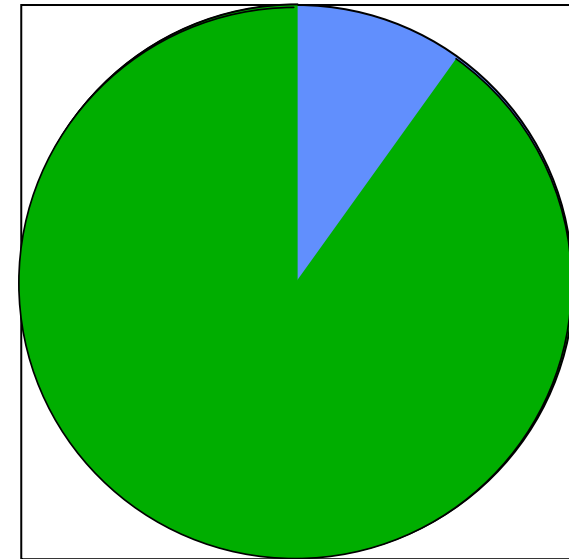
# Automaticity and Spoken Language

- Automatic processing of language allows (L2) speakers to process and produce spoken language in a rapid, unconscious, and efficient manner  
(= timely lexical, syntactic, semantic processing)
- Automatic processing of lower levels of skills frees up attentional resources for higher cognitive orders (e.g. speakers can focus on what they want to say rather than how they can say it)

# Cognitive Processing Hypothesis



*Fluent Native Speaker*



*Fluent L2 Speaker*

1. *Automaticity is independent of accuracy or control of linguistic form.*
2. *Attention to linguistic form saps attention to content and to other complex cognitive tasks.*

# Importance of Assessing “core” language processing

Fulcher (2003, p24)

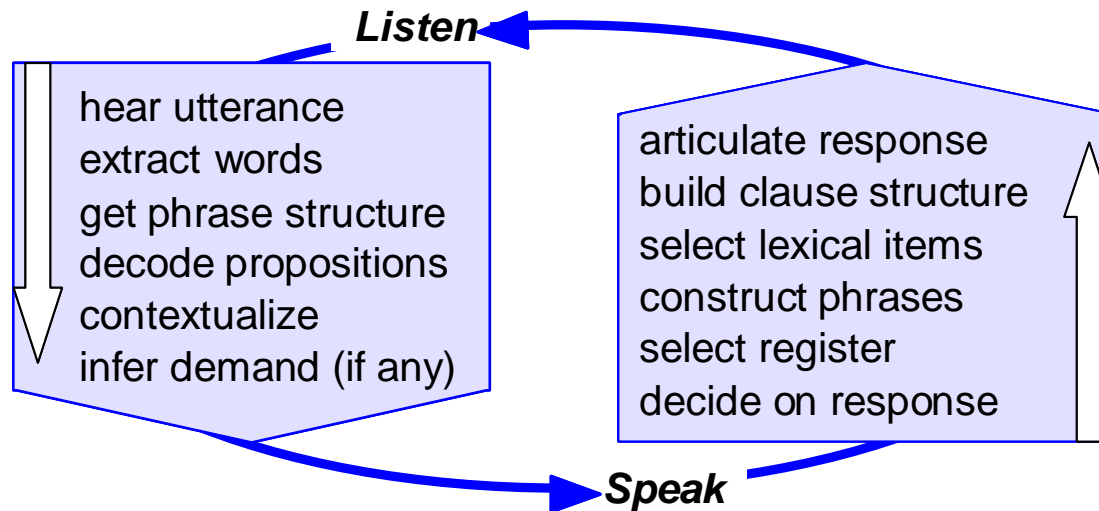
“There is a psychological aspect to speech that needs to be considered. When writing, the writer has time to plan, produce and correct. The speaker cannot tackle the task of speaking in the same leisure manner.

Speech is a ‘real-time’ phenomenon (Bygate, 1987). It has to be **planned**, **formulated** and **articulated with considerable speed**. The speed with which learners can produce speech that is appropriate to the context will depend upon a number of factors, including their control over the structure of the language, their lexical range, the ability to chunk formulaic expression, and summed up in terms of the degree to which the whole process has become ‘**automatic**’, and no longer requires conscious attention. This concept of ‘automaticity’ in speech may be associated with a number of factors, including the complexity of the message the learners wishes to communicate, how familiar they are with the topic area, the speed at which processing is expected to take place, the degree of accuracy required in the context, and perceived ‘penalties’ of getting something wrong.” (boldface, underline added)

# Ordinate's Test Construct

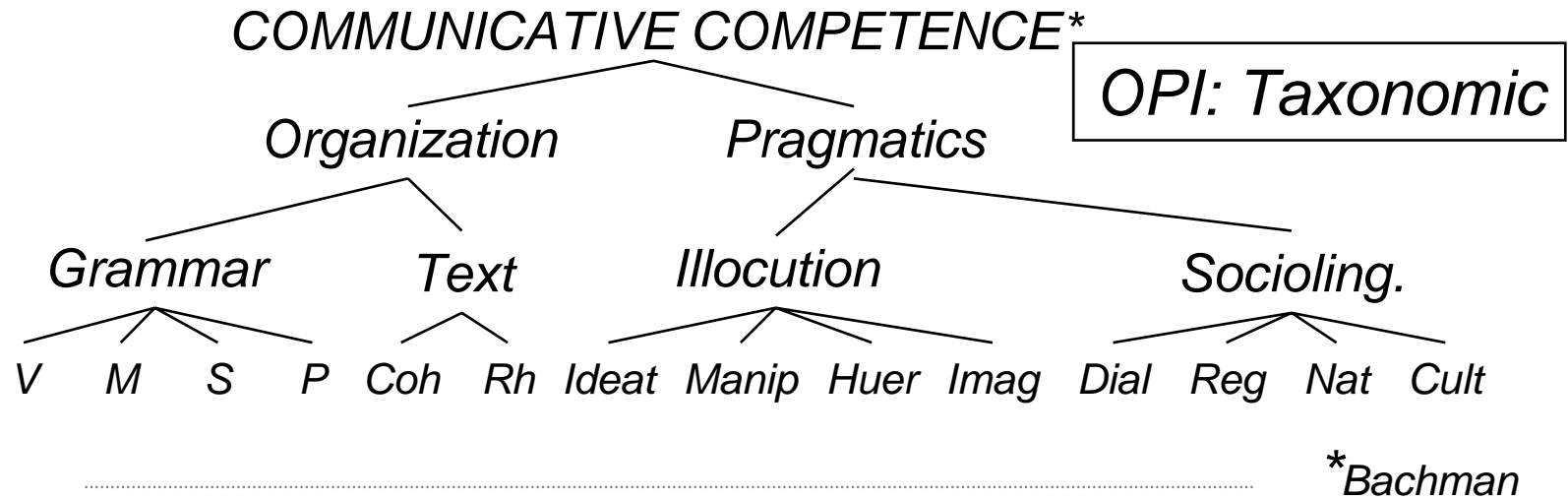
Measures *facility* in spoken language

The ability to understand spoken language on everyday topics and to respond appropriately and intelligibly in the target language at a native-like conversational pace.



*Adapted from Levelt, 1989*

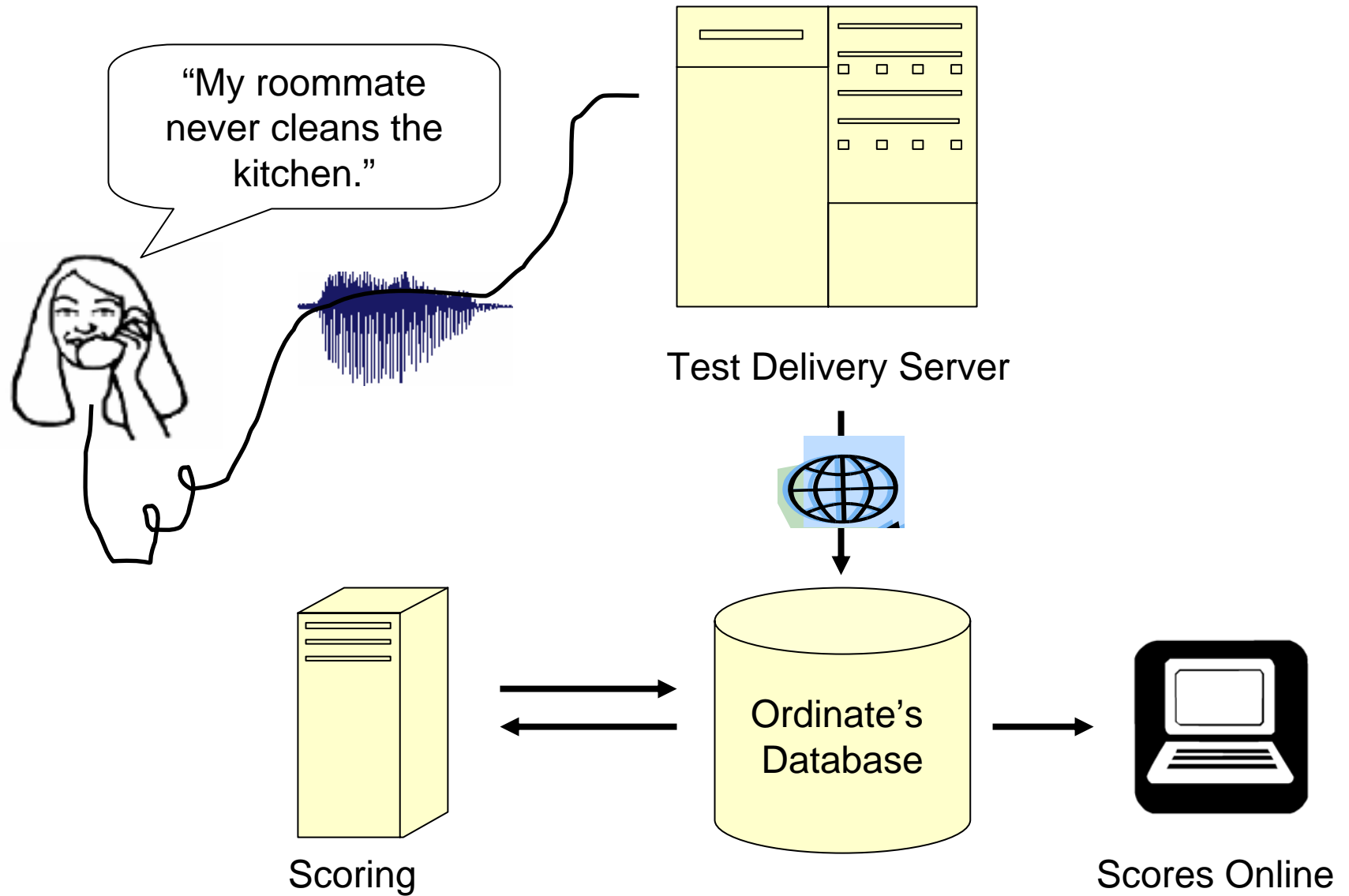
# Construct Comparison



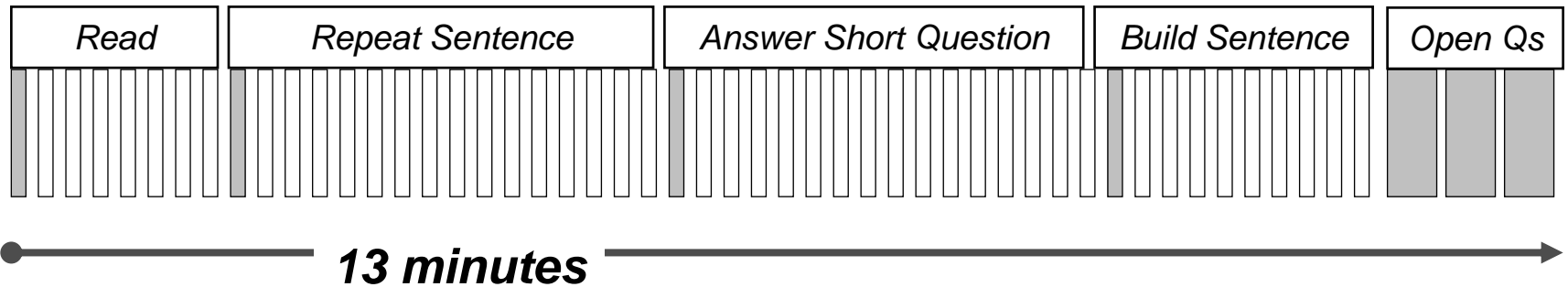
# Ordinate's Automated Spoken Language Tests

= Completely automated test of spoken languages  
(e.g. English, Spanish, Dutch)

- Automatic Administration
  - Taken over the telephone
  - Approximately 13-15 minutes to complete
  - Consistent administration on demand
- Automatic Scoring
  - Use speech processing technologies (e.g. ASR)
  - Use computerized automated scoring system
  - Objective scoring
  - Results posted on a secure web site in minutes

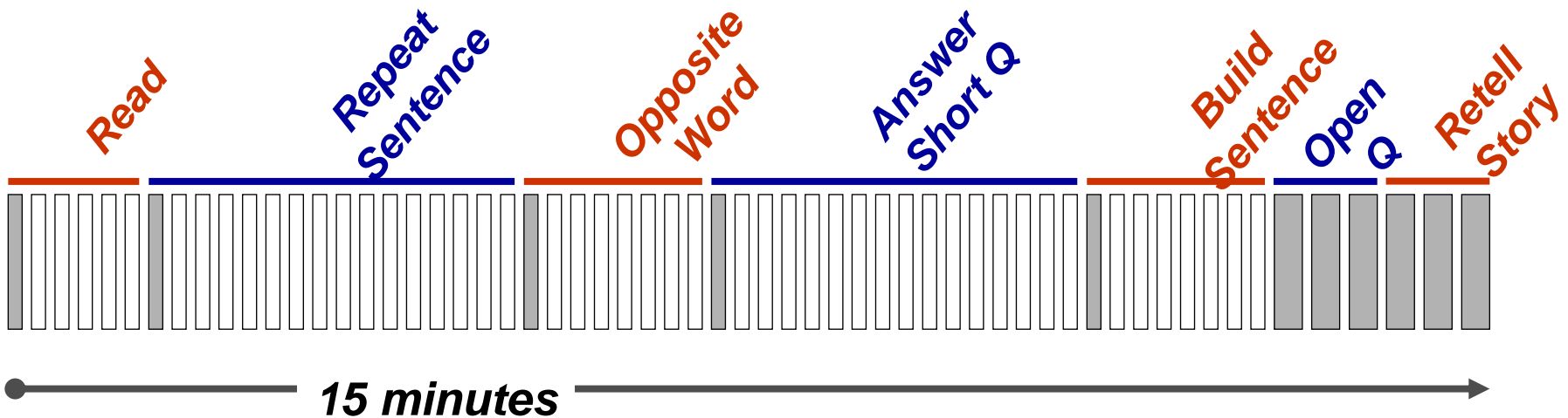


## Versant for English (formally known as PhonePass SET-10)



Gray items not scored

## Versant for Spanish (formally known as SST)

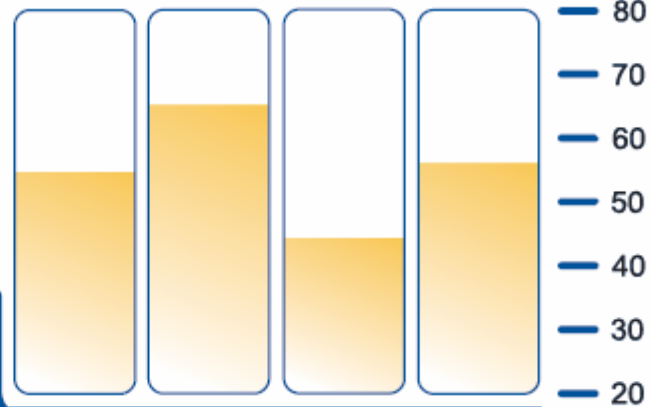


## Characteristics of Test Items

- Relatively frequent vocab based on spoken corpus
- Most items are vocabulary or sentence level
- Integrated “*listen* → *speak*” items
- Recorded by various native speakers
- Decontextualized carefully constructed constrained Items
  - To maximize response density in a given time
  - Needs language-specific abilities
  - Only a few expected responses for accurate scoring

# SET-10<sup>®</sup> Sample Score Report

## SET-10 Test Sample Score Report



55

66

45

57

TEST IDENTIFICATION NUMBER: 12345678

DATE: March 14, 2004

TIME: 2:34 PM (PST)

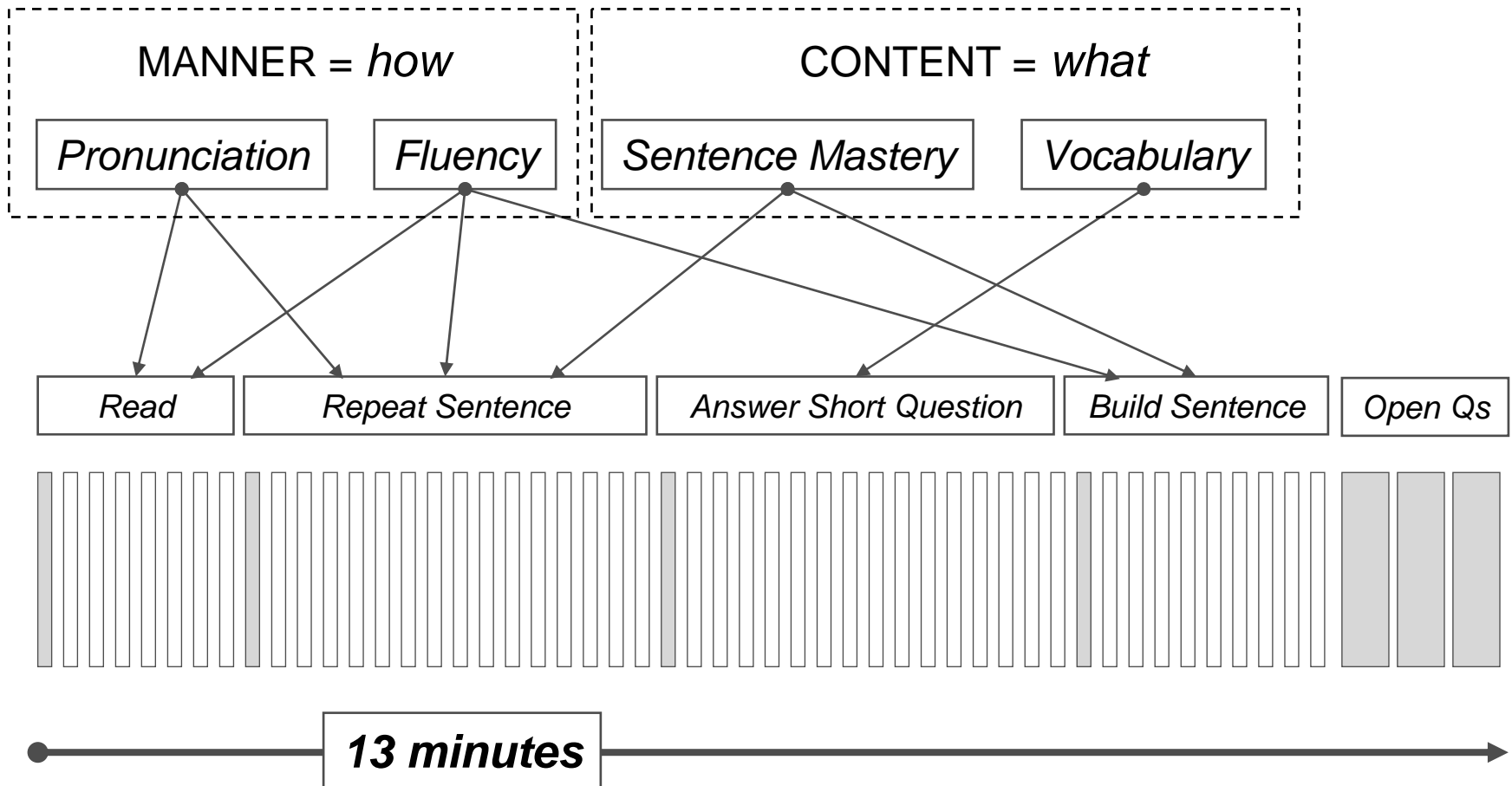
PRONUNCIATION

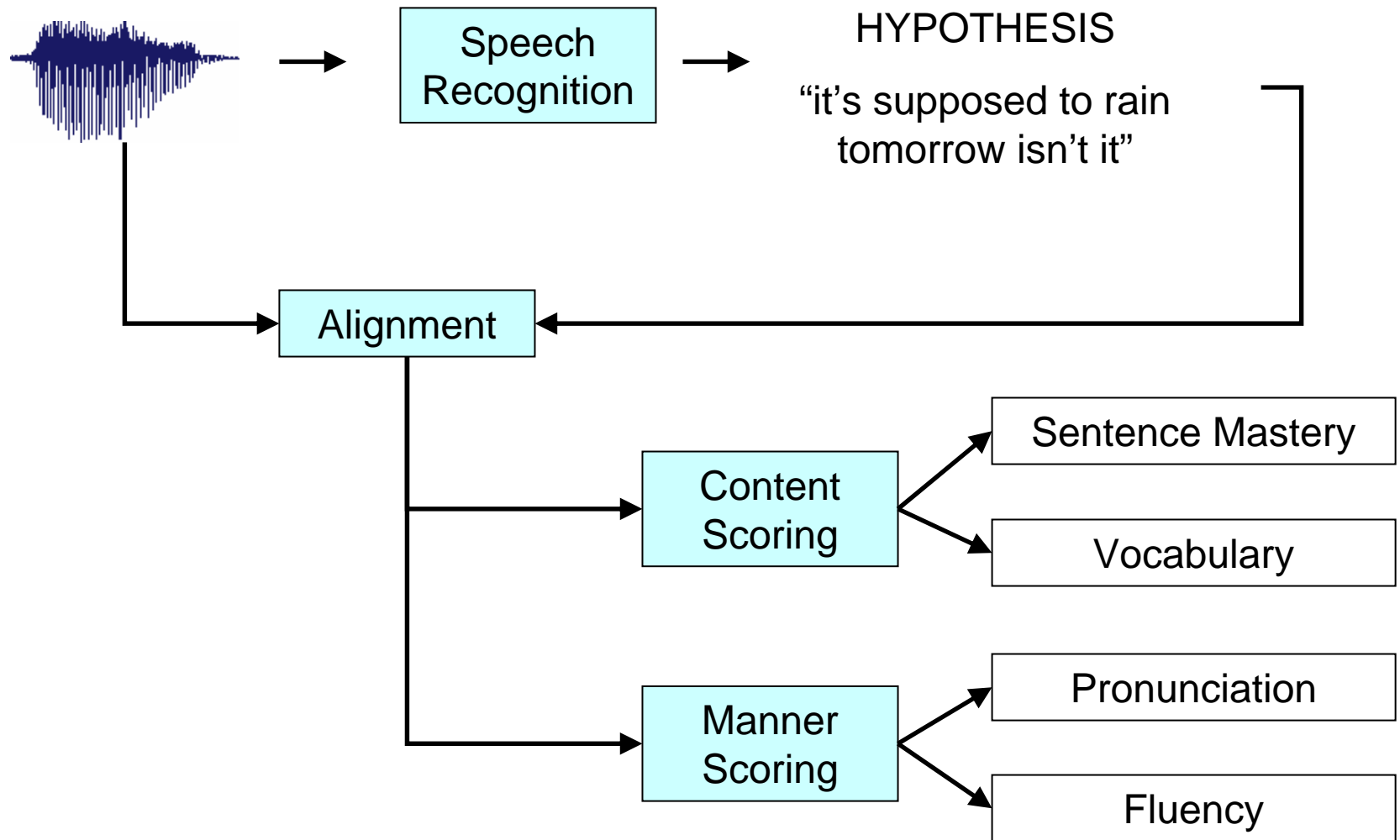
FLUENCY

VOCABULARY

SENTENCE MASTERY

# Scoring Logic (Versant for English)





# Computation of Scores

50%

Content Subscores  
= what the speaker  
said

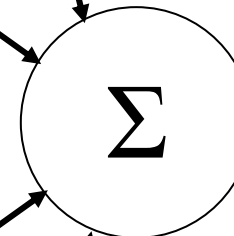
20%  
Vocabulary

30%  
Sentence Mastery

20%

20%  
Pronunciation

30%  
Fluency



Overall Score

Reported between  
20 and 80

50%

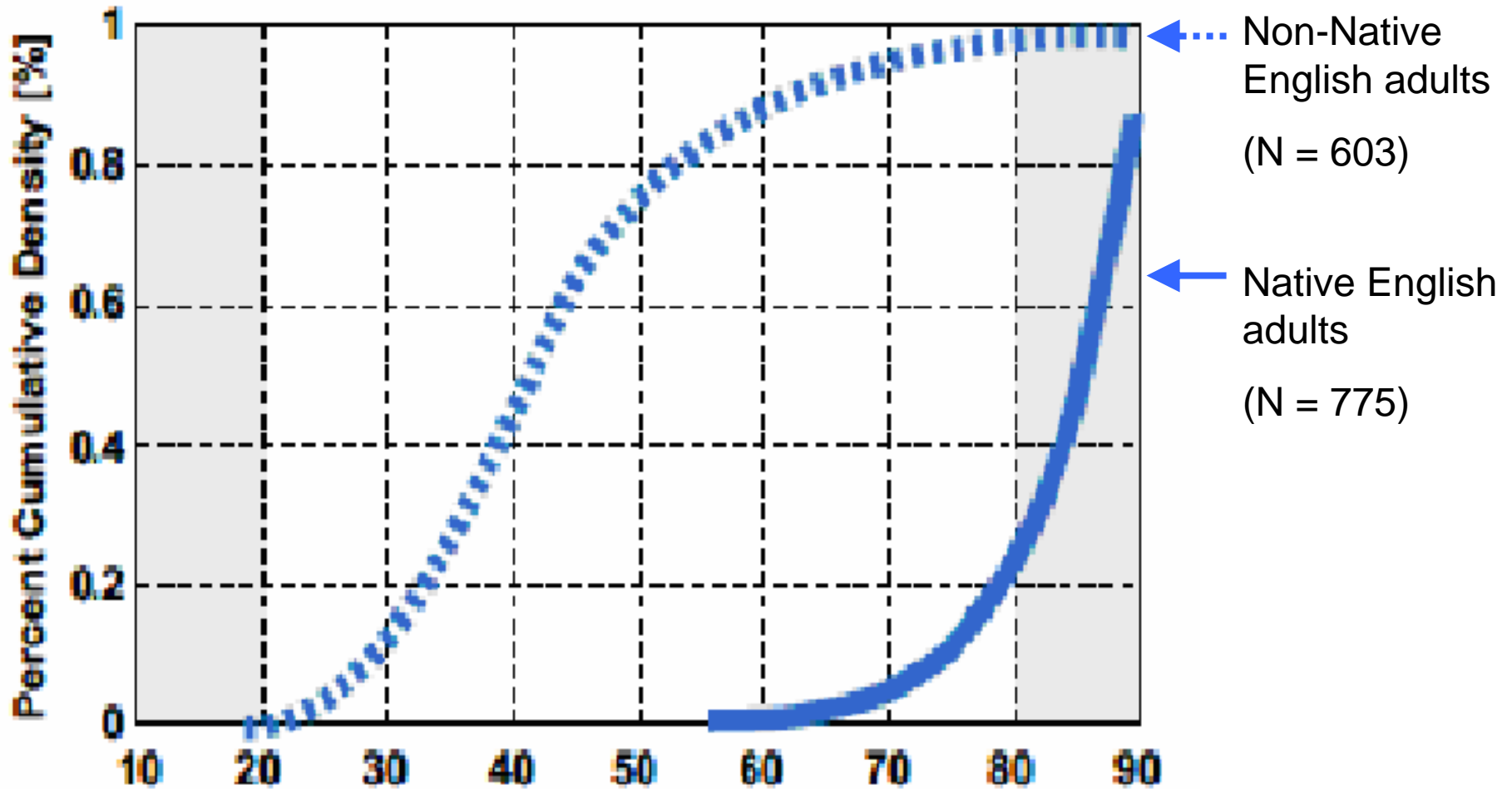
Manner Subscores  
= how the speaker  
said it

## Reliability

<b>Score Types</b>	<b>Versant for English</b>	<b>Versant for Spanish</b>
<b>Overall</b>	0.97	0.96
<b>Sentence Mastery</b>	0.93	0.96
<b>Vocabulary</b>	0.88	0.77
<b>Fluency</b>	0.95	0.94
<b>Pronunciation</b>	0.97	0.95

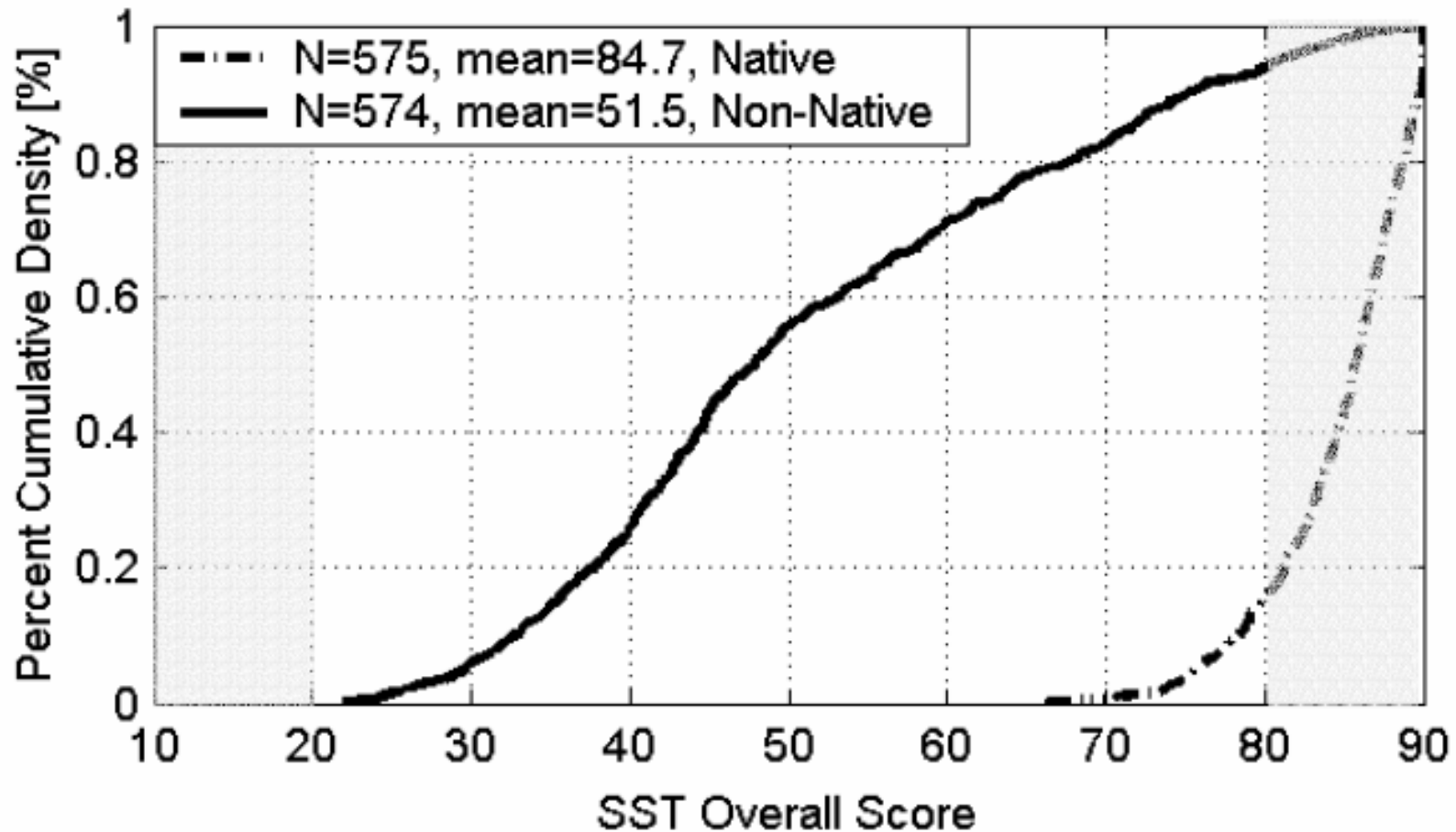
# Validity

## *Native vs. Non-native adults (English)*

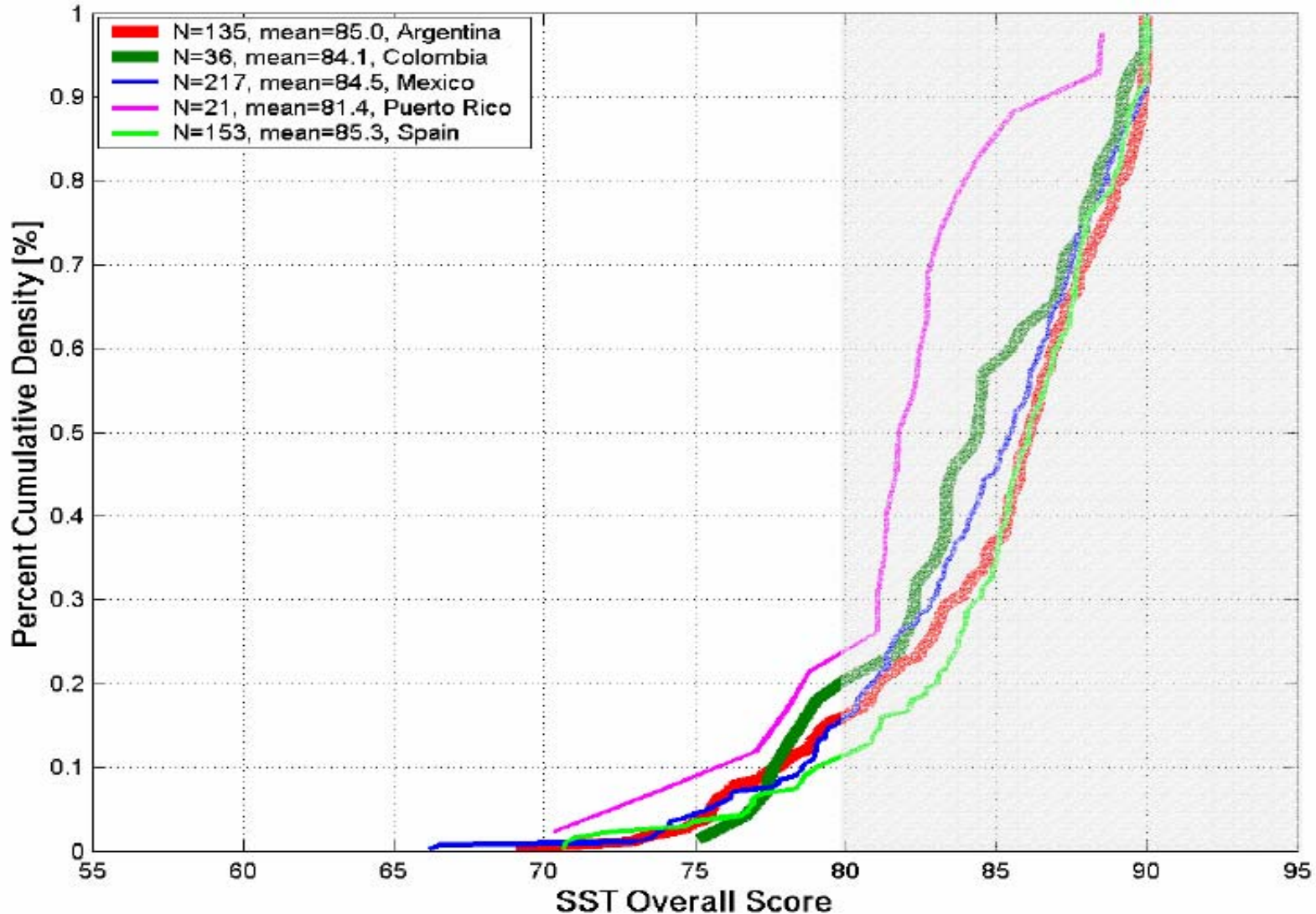


# Validity

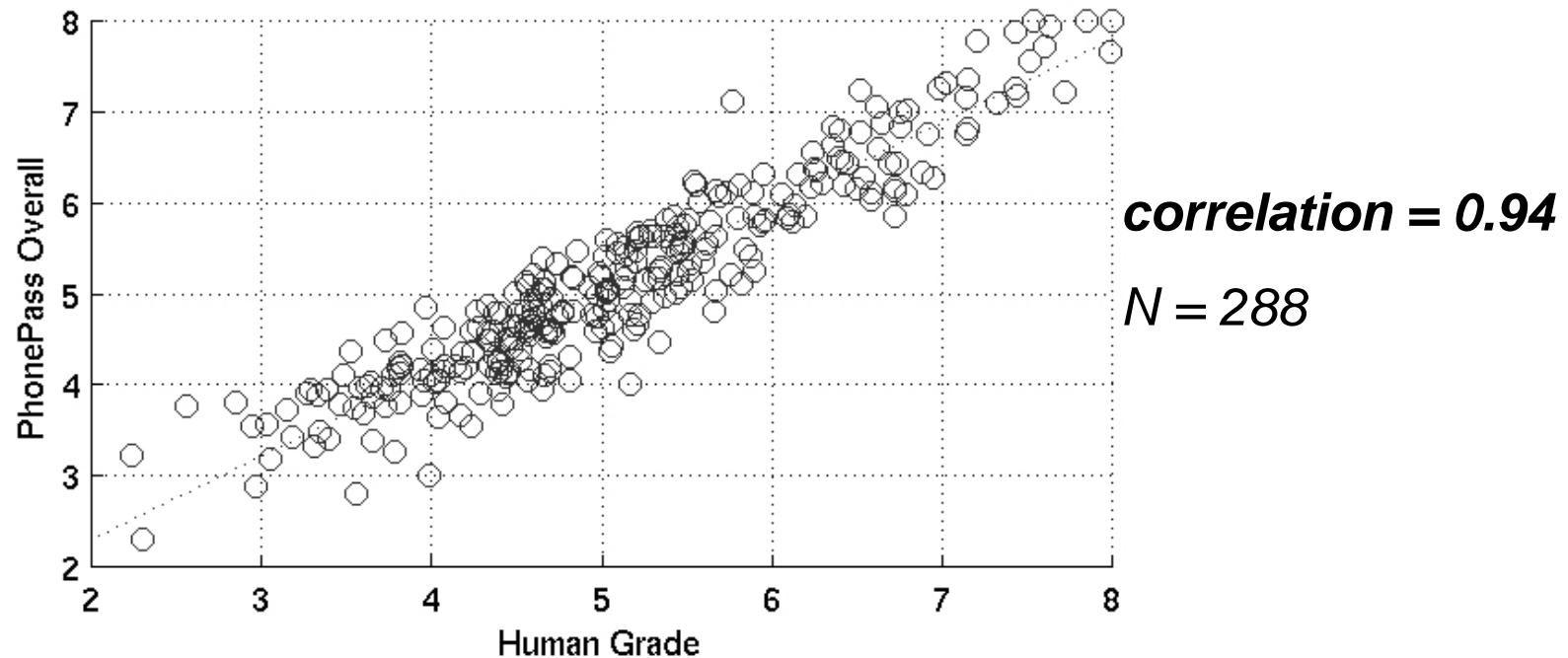
## *Native vs. Non-native adults (Spanish)*



# Native Scores by Countries

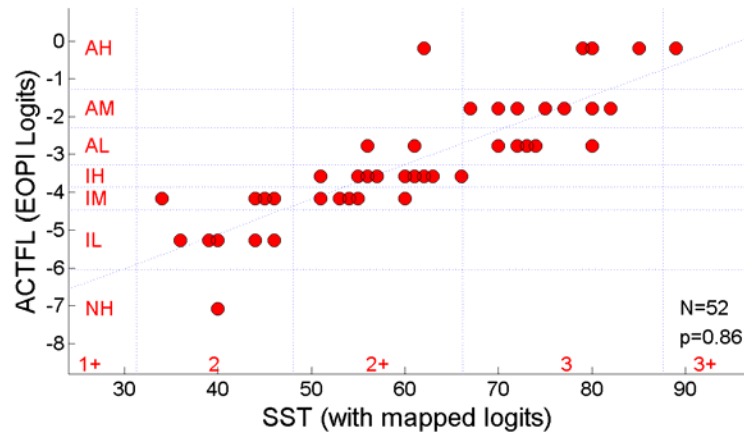


# Machine-Human Comparison



Data from testing at US universities and companies in Europe, Asia, and Latin America.

# Concurrent Validity

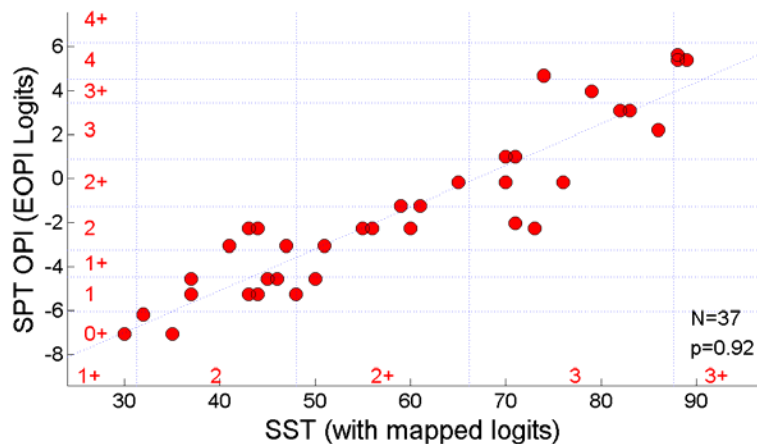


**ACTFL ~ SST**

*Machine ~ Two Raters*

*Different Material*

$$r = 0.86$$



**SPT OPI ~ SST**

*Two Raters ~ Machine*

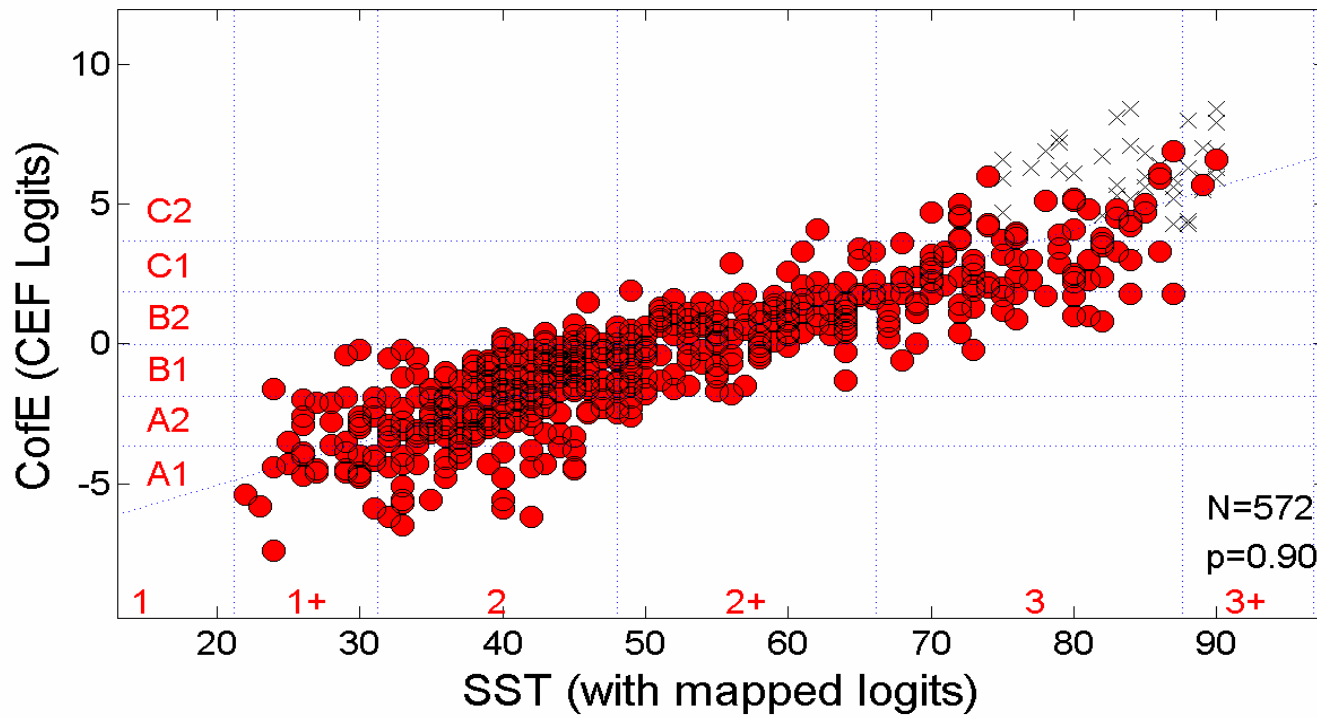
*Different Material*

$$r = 0.92$$

**CEF ~ SST**

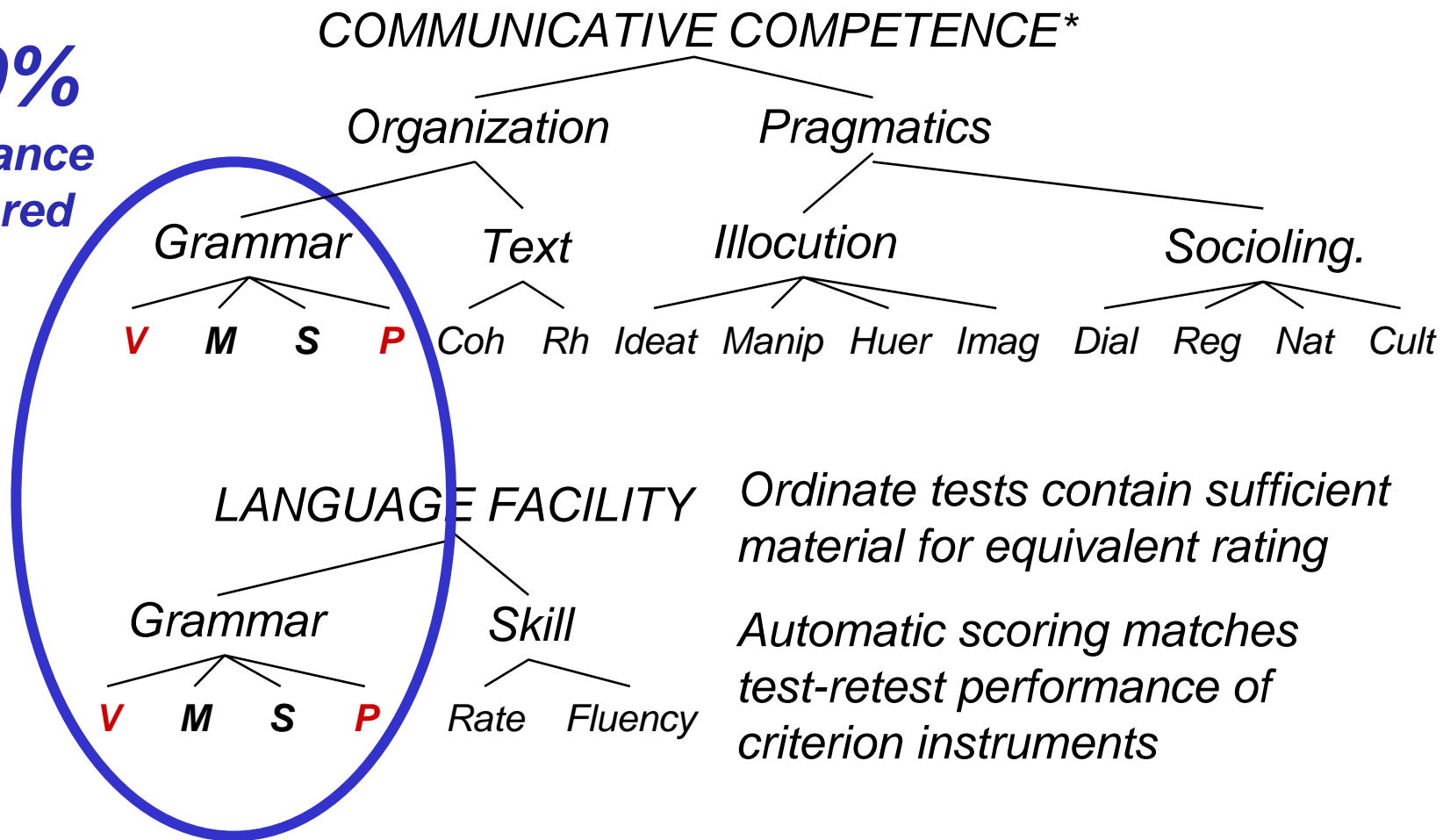
*Two Raters ~ Machine  
Different Material*

$r = 0.90$



# Performance Puzzle

**>80%**  
of variance  
is shared



*Ordinate tests contain sufficient material for equivalent rating*

*Automatic scoring matches test-retest performance of criterion instruments*

## Summary

- Ordinate's tests are psycholinguistic-theory based
- Carefully developed automated tests are reliable and valid
- Performance on integrated Psycholinguistic-based tasks can be a reliable predictor for more complex communicative tasks
- Measuring automaticity of core language processing skills is important in spoken language testing