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

COMMUNICATIVE COMPETENCE

- Organization
  - Grammar
    - V
    - M
    - S
    - P
    - Coh
    - Rh
    - Ideat
    - Manip
    - Huer
    - Imag
  - Text
  - Illocution
- Pragmatics
  - Socioling.
    - Dial
    - Reg
    - Nat
    - Cult

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.
Levelt’s Model (1989)

Conceptualizer
- Message generation

Preverbal message

Formulator
- Grammatical encoding
- Phonological encoding

Phonetic Plan (Internal Speech)

Articulator

Lexicon

Speech Comprehension System

Parsed Speech

Phonetic String

Audition

Overt speech

Parsed Speech
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)
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.**
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

COMMUNICATIVE COMPETENCE*

- Organization
- Pragmatics

Grammar
- V
- M
- S
- P

Text
- Coh
- Rh
- Ideat

Illocution
- Manip
- Huer
- Imag

Socioling.
- Dial
- Reg
- Nat
- Cult

OPI: Taxonomic

LANGUAGE FACILITY*

- Grammar
- Skill

Grammar
- V
- M
- S
- P

Skill
- Rate
- Fluency

Ordinate: Metric in time

* Bachman

*Ordinate’s Tests
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
“My roommate never cleans the kitchen.”
Versant for English (formally known as PhonePass SET-10)

Read | Repeat Sentence | Answer Short Question | Build Sentence | Open Qs

13 minutes

Gray items not scored

Versant for Spanish (formally known as SST)

Read | Repeat Sentence | Opposite Word | Answer Short Q | Build Sentence | Open Q | Retell Story

15 minutes
Characteristics of Test Items

- Relatively frequent vocab based on spoken corpus
- Most items are vocabulary or sentence level
- Integrated “listen $\rightarrow$ 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® Sample Score Report**

**SET-10 Test**
Sample Score Report

**OVERALL SCORE:** 54

**TEST IDENTIFICATION NUMBER:** 12345678
**DATE:** March 14, 2004
**TIME:** 2:34 PM (PST)

**Scores:**
- Pronunciation: 55
- Fluency: 66
- Vocabulary: 45
- Sentence Mastery: 57
Scoring Logic (Versant for English)

MANNER = how

Pronunciation

Fluency

CONTENT = what

Sentence Mastery

Vocabulary

Read

Repeat Sentence

Answer Short Question

Build Sentence

Open Qs

13 minutes
Speech Recognition

HYPOTHESIS
“it’s supposed to rain tomorrow isn’t it”

Alignment

Content Scoring
- Sentence Mastery
- Vocabulary

Manner Scoring
- Pronunciation
- Fluency
Computation of Scores

50%

Content Subscores = what the speaker said

Manner Subscores = how the speaker said it

20% Vocabulary

Sentence Mastery

30%

Pronunciation

20%

Fluency

30%

Σ

Overall Score

Reported between 20 and 80

Ordinate Corporation

21
# Reliability

<table>
<thead>
<tr>
<th>Score Types</th>
<th>Versant for English</th>
<th>Versant for Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Sentence Mastery</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.88</td>
<td>0.77</td>
</tr>
<tr>
<td>Fluency</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>0.97</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Validity

Native vs. Non-native adults (English)

Native English adults (N = 775)

Non-Native English adults (N = 603)
Validity

Native vs. Non-native adults (Spanish)
Native Scores by Countries

- Argentina: N=135, mean=85.0
- Colombia: N=36, mean=84.1
- Mexico: N=217, mean=84.5
- Puerto Rico: N=21, mean=81.4
- Spain: N=153, mean=85.3

SST Overall Score vs. Percent Cumulative Density [%]
Machine-Human Comparison

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

correlation = 0.94
N = 288
Concurrent Validity

**ACTFL ~ SST**
*Machine ~ Two Raters*  
*Different Material*  
\[ r = 0.86 \]

**SPT OPI ~ SST**
*Two Raters ~ Machine*  
*Different Material*  
\[ r = 0.92 \]
Two Raters ~ Machine
Different Material
$r = 0.90$
Performance Puzzle

COMMUNICATIVE COMPETENCE*

Organization

Pragmatics

Illocution

Socioling.

LANGUAGE FACILITY

Grammar

Text

Grammar

Skill

Ordinate tests contain sufficient material for equivalent rating

Automatic scoring matches test-retest performance of criterion instruments

Ordinate Corporation
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