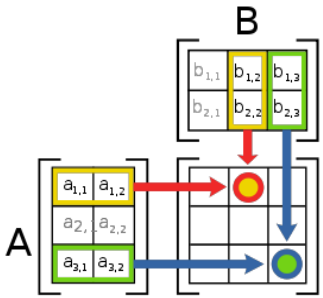
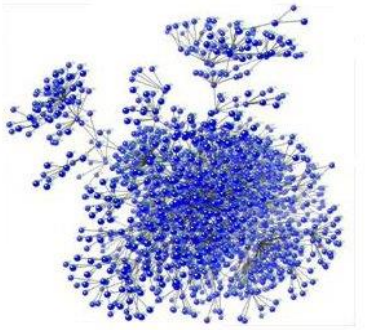
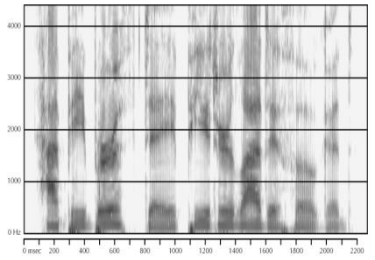






Elicited Imitation: What can it tell us about oral language proficiency?

Alistair Van Moere
Xiaoqiu Xu
Mallory Klungtvedt

Knowledge Technologies, Pearson

Elicited Imitation

- Rationale
- Research questions
- Method
- Data analysis
- Discussion

Assessment Perspective

- Repeating meaningful language is distinct from repeating random words or digits
  - Words that are related in meaning or syntax are chunked
  - Proficient speakers have thousands of hours on-task practice

- Higher proficiency language speakers can repeat longer, complex sentences, as long as the sentence is meaningful to them and the syntax is familiar

(Radloff 1991)

E.g. "Don't forget to feed the cat in the morning"

"forget cat the morning the in don't feed to"

Sociolinguistic Perspective

- Speakers regularly adopt their interlocutor's words and grammar into their own speech.

(Levinson 1983; Brown & Yule 1983)

- Speakers feel a sense of endorsement when their listeners do this. Repeating phrases, or sentences, of other speakers:
  - (a) accomplishes a conversation
  - (b) shows one's response to another
  - (c) shows acceptance of other's utterances & their participation
  - (d) gives evidence of one's own participation

(Tannen 1989)

Pyscholinguistic Perspective

- Incorporating language from an interlocutor is resource efficient: we need not allocate so many resources to formulating syntax and can focus more attention on conceptualizing a response.

(Bygate 2001)

- When listeners listen *passively,* they attend more to *meaning*
- But when they have to contribute in a conversation, they also pay attention to the syntax.

> … they may have to use that syntax as the basis of their own production later.

(Swain 1985)

Pedagogic Perspective

- Rehearsal of 'old input' and enhancement of automaticity should be adopted as learning activities.

  (Hulstijn 2001)

  - Rehearsal
  - Practice
  - Drill
  - Automaticity

  …. These activities often have negative connotations among L2 specialists.

- **However**, several decades of psycholinguistic research show that lexical information must be reactivated regularly for it to be quickly accessible.

  (Hulstijn 2001)

Elicited Imitation

Perspectives:

- Assessment
- Sociolinguistic
- Psycholinguistic
- Pedagogic

Previous Research

- Comparison of 3 tasks: sentence repeats
  sentence completion
  FSI-like interview

	Repeat	Completion	Interview
Repeat	-		
Completion	0.62	-	
Interview	0.75	0.44	-

Sentence repetition tasks:

"despite appearing low in face validity, may surpass interview techniques in overall validity and reliability"

(Henning 1983)

Maximizing the "Reconstructive" Aspect

How can we maximize the reconstructive nature of the task, and minimize dependency on memory?

…. Insert a distraction between hearing the sentence and repeating it

- Hear a sentence
- Respond to a question
- Repeat the sentence

E.g. *What's the opposite of "soft"?*

Research Questions

1. Does performance on Repeat tasks generalize to performance on constructed speech tasks?

2. Which task – *Repeats* or *Delayed Repeats* – is most reliable for separating test-takers according to proficiency?

Participants

- n = 116
- 16 L1s
  - Mandarin 48%
  - Hindi 16%
  - Korean 13%
  - English 8%
  - Other 19%
- Mean age = 27 (18-46)
- Male=44, female=72
- 98% university-educated

Five Tasks in this Study

Memory

- Digit Span

Repeat Sentence

- Repeat
- Delayed Repeat

Constructed Speech

- Story Retell
- Pragmatics

Tasks

Memory

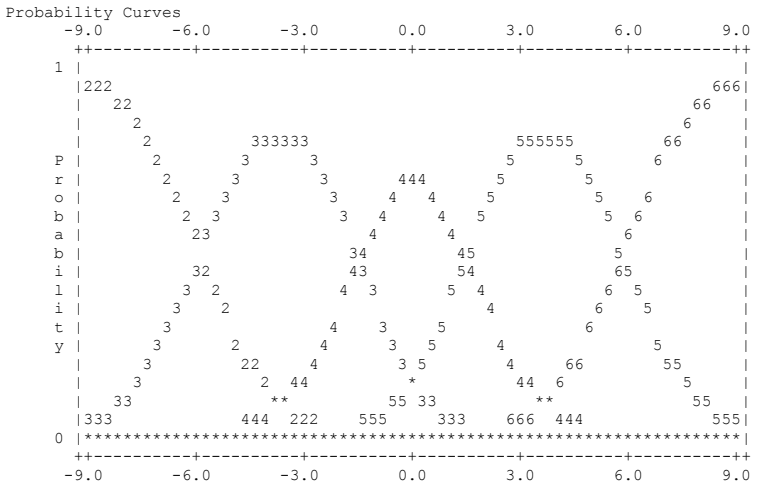
- Digit Span

Repeat Sentence

- Repeat
- Delayed Repeat

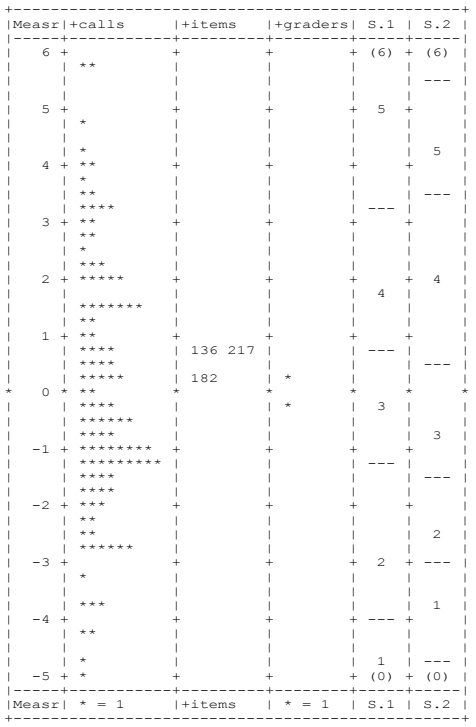
Constructed Speech

- Story Retell
- Pragmatics



"Tom played games on his computer every day. He was the best player at his school. Then Tom's little sister started playing as well. In only two weeks, she was just as good as he was. She even beat him a few times. So Tom quit playing computer games and started playing baseball instead."

Tasks



Memory

- Digit Span

Repeat Sentence

- Repeat
- Delayed Repeat

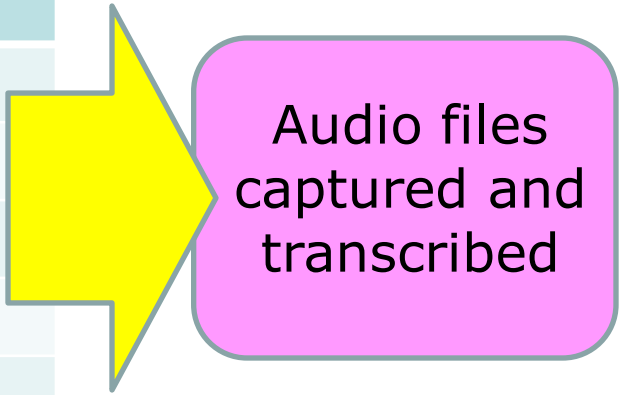
Constructed Speech

- Story Retell
- Pragmatics

You have an exam this afternoon and are studying at the library. A group of students at a nearby table are talking loudly, unaware that they are bothering you. What should you say to them?

Modified Versant English Test

Num items	Task
16	Repeat
16	Delayed Repeat
3	Story Retells
4	Pragmatics
20	Digit span



Measures

Num items	Task	Task Completion	Pronun-ciation	Fluency
16	Repeat	# Word Errors	Ratings	Ratings
16	Delayed Repeat	# Word Errors	Ratings	Ratings
3	Story Retells	Ratings	Ratings	Ratings
4	Pragmatics	Ratings	Ratings	Ratings
20	Digit span	# Correct		

- Double ratings from transcripts
- Pool of 6 judges
- Rasch-scaled

- Double ratings from audio
- Pool of 6 judges
- Rasch-scaled

Rater Agreement

Number of items	Task	Task completion	Pronun-ciation	Fluency
16	Repeat	n.a.	0.91	0.91
16	Delayed Repeat	n.a.	0.95	0.95
3	Story Retell	0.91	0.89	0.90
4	Pragmatics	0.88	0.91	0.92

Raw scores, 2 ratings, n=112

Generalizability Study

			G-coefficients		
Number of items	Time on task	Task	Task completion	Pronun-ciation	Fluency
16	3 mins	Repeat	0.94	0.94	0.94
16	4 mins	Delayed Repeat	0.94	0.96	0.96
3	3 mins	Story Retell	0.79	0.89	0.83
4	4 mins	Pragmatics	0.74	0.91	0.91

Raw scores, 2 ratings, n=112

Point Biserial Correlation

Task Completion

Number of items	Time on task	Task	Person ptbis	Item ptbis
16	3 mins	Repeat	0.65	0.58
16	4 mins	Delayed Repeat	0.67	0.69
3	3 mins	Story Retell	0.15	0.47
4	4 mins	Pragmatics	0.39	0.41

Correlations: Pronunciation

	Repeat	Delayed Repeat	Story Retell	Pragmatics
Repeat				
Delayed Repeat	0.96			
Story Retell	0.93	0.95		
Pragmatics	0.93	0.93	0.93	
Digit Span	-0.27	-0.23	-0.18	-0.24

Rasch-scaled data, n=112

Correlations: Fluency

	Repeat	Delayed Repeat	Story Retell	Pragmatics
Repeat				
Delayed Repeat	0.79			
Story Retell	0.83	0.70		
Pragmatics	0.85	0.76	0.86	
Digit Span	-0.28	-0.26	-0.14	-0.23

Rasch-scaled data, n=112

Correlations: Task Completion

	Repeat	Delayed Repeat	Story Retell	Pragmatics
Repeat				
Delayed Repeat	0.70			
Story Retell	0.72	0.57		
Pragmatics	0.54	0.51	0.60	
Digit Span	-0.30	-0.27	-0.11	-0.32

Rasch-scaled data, n=112

Multitrait Multimethod

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON									
	FLU	0.90								
	ERROR	0.80	0.87							
STORY RETELL	PRON	0.93	0.86	0.77						
	FLU	0.80	0.83	0.74	0.84					
	CONT	0.77	0.81	0.72	0.80	0.82				
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74			
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86		
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	

MTMM – with Reliability Coefficients

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON	0.94								
	FLU	0.90	0.94							
	ERROR	0.80	0.87	0.94						
STORY RETELL	PRON	0.93	0.86	0.77	0.89					
	FLU	0.80	0.83	0.74	0.84	0.83				
	CONT	0.77	0.81	0.72	0.80	0.82	0.79			
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74	0.91		
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86	0.91	
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	0.74

Hetero-trait Mono-method Triangles

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON	0.94								
	FLU	0.90	0.94							
	ERROR	0.80	0.87	0.94						
STORY RETELL	PRON	0.93	0.86	0.77	0.89					
	FLU	0.80	0.83	0.74	0.84	0.83				
	CONT	0.77	0.81	0.72	0.80	0.82	0.79			
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74	0.91		
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86	0.91	
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	0.74

Validity Diagonals - Pronunciation

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON	0.94								
	FLU	0.90	0.94							
	ERROR	0.80	0.87	0.94						
STORY RETELL	PRON	0.93	0.86	0.77	0.89					
	FLU	0.80	0.83	0.74	0.84	0.83				
	CONT	0.77	0.81	0.72	0.80	0.82	0.79			
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74	0.91		
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86	0.91	
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	0.74

Validity Diagonals - Fluency

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON	0.94								
	FLU	0.90	0.94							
	ERROR	0.80	0.87	0.94						
STORY RETELL	PRON	0.93	0.86	0.77	0.89					
	FLU	0.80	0.83	0.74	0.84	0.83				
	CONT	0.77	0.81	0.72	0.80	0.82	0.79			
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74	0.91		
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86	0.91	
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	0.74

Validity Diagonals - Task Completion

		REPEATS			STORY RETELL			PRAGMATICS		
		PRON	FLU	ERROR	PRON	FLU	CONT	PRON	FLU	CONT
REPEATS	PRON	0.94								
	FLU	0.90	0.94							
	ERROR	0.80	0.87	0.94						
STORY RETELL	PRON	0.93	0.86	0.77	0.89					
	FLU	0.80	0.83	0.74	0.84	0.83				
	CONT	0.77	0.81	0.72	0.80	0.82	0.79			
PRAGMATICS	PRON	0.93	0.85	0.75	0.93	0.78	0.74	0.91		
	FLU	0.84	0.85	0.78	0.84	0.86	0.81	0.86	0.91	
	CONT	0.57	0.58	0.54	0.59	0.54	0.60	0.64	0.68	0.74

Findings

1. Does performance on Repeat tasks generalize to performance on constructed speech tasks?

- Pronunciation – yes, it does
- Fluency – yes, it does
- Task Completion – bearing in mind measurement error, performance on sentence repeats is a good predictor of constructed response tasks.

…. But, of course, the constructs vary.

Findings

2. Which task – *Repeats* or *Delayed Repeats* – is most reliable for separating test-takers according to proficiency?

- Both are highly efficient
- Repeats generalize better to constructed response tasks
  - and have better face validity
  - and are more humane

Discussion

- Tasks which elicit constructed responses (*Story Retells, Pragmatics*) have constructs which are closer to talk-in-interaction, and are therefore preferred by test designers.

  …. However, they are less reliable, have less measurement precision, and there is less control over the grammar/vocabulary selected

- *Elicited imitation* tasks have a more psycholinguistic construct which is not currently in fashion in language testing, despite strong theoretical rationale and superior reliability.

… the two approaches should not be mutually exclusive.

"Hulstijn (2011) argues that core language skills such as linguistic knowledge (vocabulary, grammar, pronunciation) and automaticity (speed of processing) can be reliably and efficiently measured using discrete-point methods.

Once test takers have demonstrated a required level of competence in core proficiency, they can additionally be tested for communicative or organizational competencies in performance-based tasks, which provide alternative but less reliable estimates of ability.

It follows that domain-specific or language-in-use testing should not be the sole basis upon which to make decisions about test taker capabilities, but should be used to complement psycholinguistic assessments. "

Van Moere, A. (online first).
A psycholinguistic approach to oral language assessment, *Language Testing.*

Thank you!