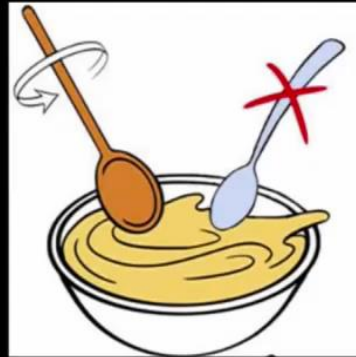


Using professional interactive speech tasks to measure L2 communicative ability: professional task design in task-based language assessment

16 SEPTEMBER 2015
TBLT CONFERENCE
KU LEUVEN
BELGIUM

**Eline van Batenburg
Ron Oostdam
Amos van Gelderen
Martine Braaksma**

KIDS AT WORK



The Wooden Spoon problem

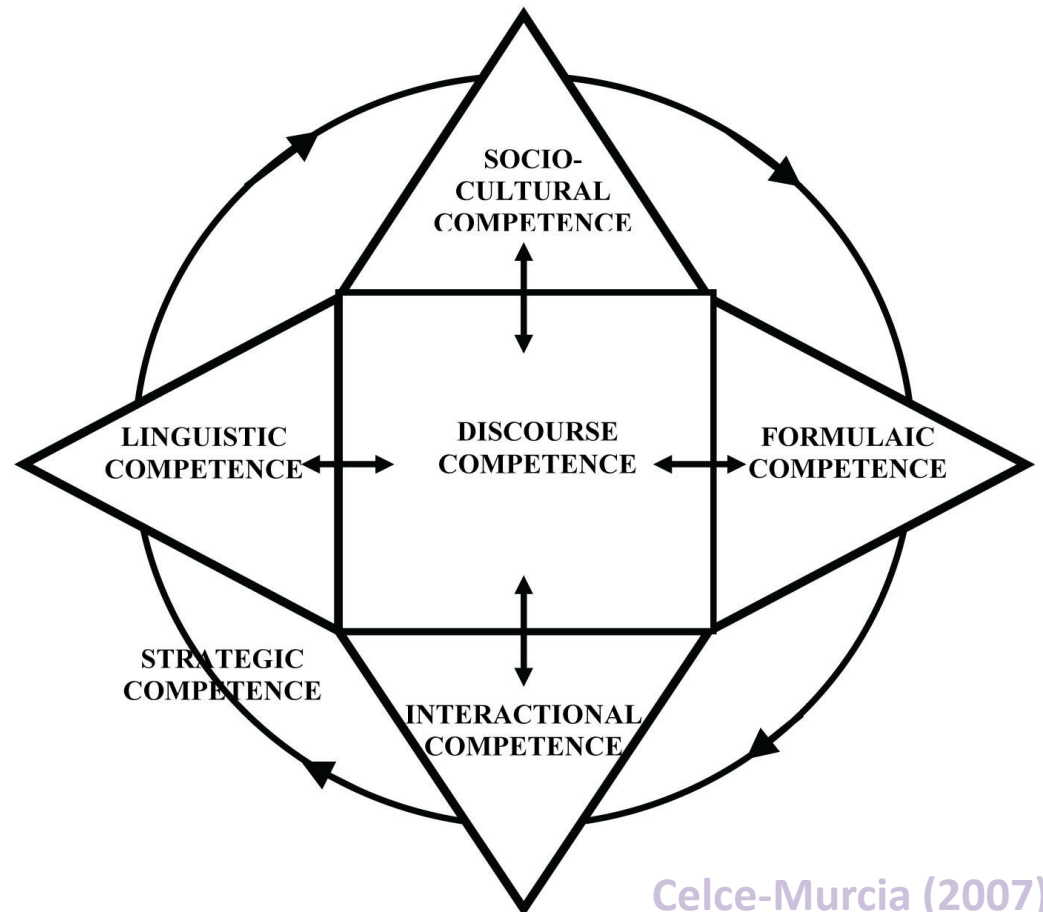
COMMUNICATIVE ABILITY

Self-supporting

- compensation
- meaning negotiation

Other-supporting

- adjust to listener's understanding
- use listener information



DESIGN

- ★ interactive tasks
- ★ authentic professional goals
- ★ functional language use
- ★ interaction strategies
- ★ standardized

- 💣 Interaction is “co-constructed”
- ✓ control interlocutor effects
- ✓ isolate individual contributions

Result

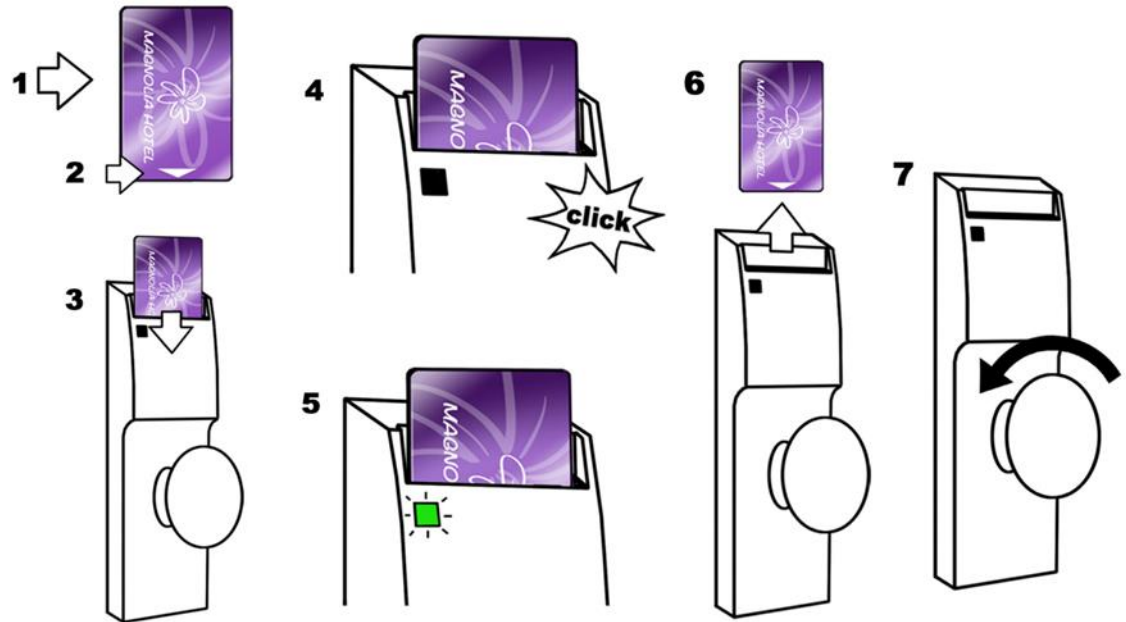
Scripted speech tasks that standardizes linguistic and interactional challenges posed to candidates.

Instruction <i>Explain to me</i>	Advice <i>Advise me on</i>	Sales <i>Convince me to</i>
how to open my door using a hotel key card	the hotel room I should choose	buy a gift in the hotel gift shop
how to bake apple cake	the film I should see	buy your second hand headphones
Connectives; prepositions of time; modals of necessity	Comparatives; zero conditional; modals of recommendation	Superlatives; intensifiers; modals of obligation

Key Card

The hotel where you work as a receptionist no longer works with keys, but with so-called key cards. One of the guests has just tried to open their door, but this did not work.

Can you explain to the guest how they can open their door, using the key card?



Self-supporting strategies

- **COMPENSATION**
familiar concept requiring production of low-frequency vocabulary
arrow, slot, door knob; bowl, wooden spoon, tin; radiator; discount token; drawing, foldable
- **MEANING NEGOTIATION: [difficult language]**
processing low-frequency vocabulary or vernacular
“Would you have a cot for me?”
“Are you telling me that these films are like chalk and cheese?”

Other-supporting strategies

- **CLARIFICATION**
“Sorry, what do I turn left?”
- **CORRECTING MISINTERPRETATION**
“So if I want the Red Room, I need to pay extra for wifi?”

▪ **TASK KEY CARD**

▪ **Candidate ID:**

▪ **Rater ID:**

▪ **Test leader ID:**

Room for notes

	Episode	Function	Rating
TL	Opens conversation		
C		Greets and offers assistance	1 - 2 - 3
TL	Asks for explanation of key card. "Please explain to me how to open my door using this thing."		
C		Explains the process 1. Purple side (with picture of magnolia) to the front 2. Arrow needs to point downwards 3. Insert the card in the slot 4. Continue to press down until you hear 'click' 5. Green light will flash 6. Remove the card 7. Turn doorknob to the left	999 or: 1 2 3 4 5
TL	Asks for clarification "Sorry, what do I turn left?"		
C		Clarifies that the doorknob needs to be turned	999 or: 1 2 3 4 5
TL	Indicates problem "But I did that. And it didn't work. How is that possible?"		
C		Respond by asking details of how customer used the card; providing explanation; being apologetic about not knowing	999 or: 1 2 3 4 5
TL	Asks "Why do you have such a convoluted procedure in place?"		
C		Responds by <input type="checkbox"/> indicating incomprehension and / or asking for clarification <input type="checkbox"/> providing explanation <input type="checkbox"/> being apologetic about not knowing	999 or: 1 2 3 4 5
TL	Checks understanding of the process		
C		Responds by back-channelling, adding or clarifying information	999 or: 1 2 3 4 5
TL	Misinterprets "The green light flashes and I turn the door knob to the left"		
C		Corrects misinterpretation The card needs to be removed first	999 or: 1 2 3 4 5
TL	Thanks candidate		
C		Closes conversation	999 or: 1 - 2 - 3

Overall impression
Language Accuracy

1 2 3 4 5

Overall impression
Interactional Ability

1 2 3 4 5

Overall impression language accuracy

*On a scale of 1-5, indicate the extent to which the candidate expresses himself in **lexically and grammatically correct** English, and can be understood:*

- 1 = barely speaks English / speaks more Dutch than English.
- 3 = frequently makes mistakes in sentence construction and / or lexical choice, but this does not seriously impede comprehensibility.
- 5 = correct sentence structure and lexical choice. Can be understood easily.

Overall impression interactional ability

*On a scale of 1-5, indicate the extent to which the candidate manages to **convey his message**, and to **overcome potential communication problems**:*

- 1 = is barely able to convey the message. Cannot solve (mutual) communication problems independently in English.
- 3 = the message is not always clear or complete, but attempts to solve (mutual) communication problems in English.
- 5 = the message is conveyed clearly. Communication problems do not / hardly occur.

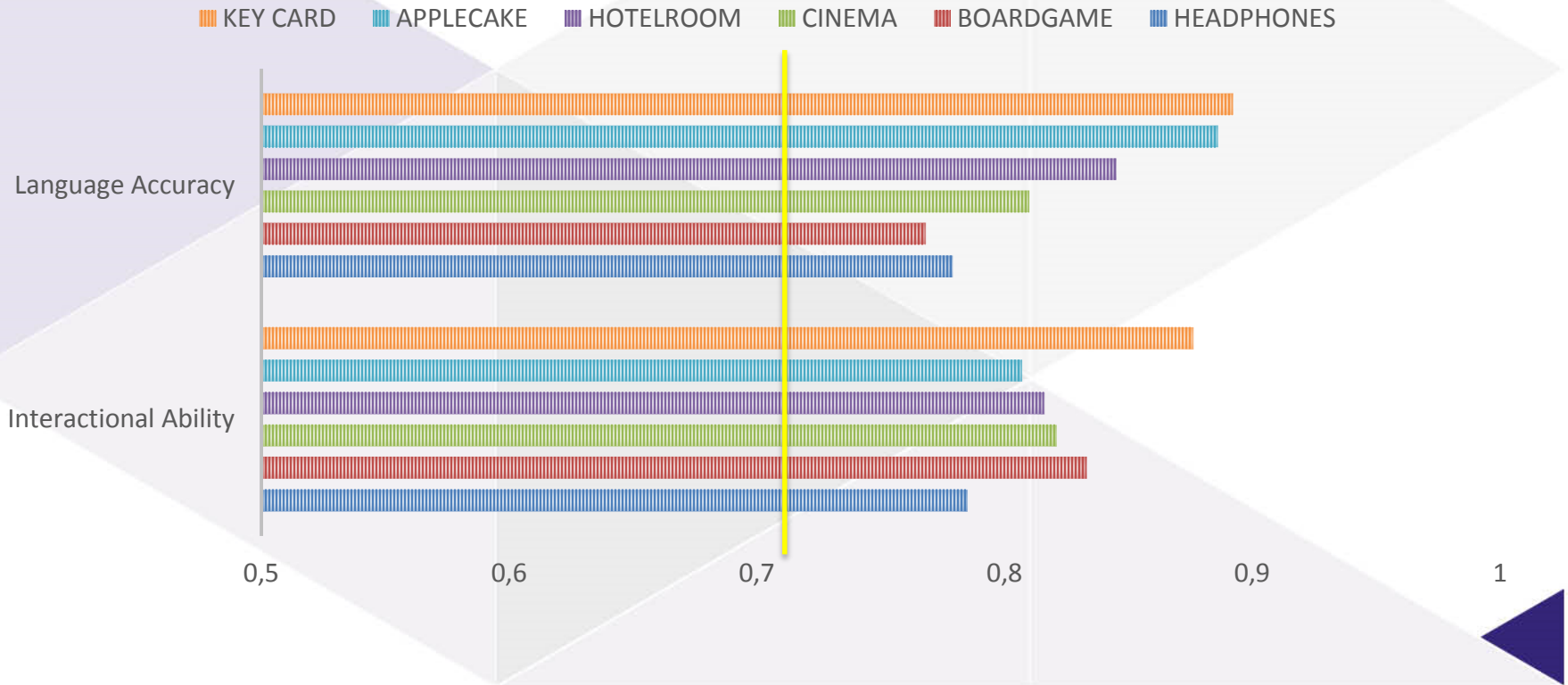
*On a scale of 1-5, indicate the extent to which the candidate provides an **adequate response** during each contribution:*

- 1 = no or predominantly Dutch contribution; very weak / inappropriate
- 2 = weak / inappropriate
- 3 = reasonable
- 4 = strong
- 5 = very strong

RESEARCH QUESTIONS

1. Can pre-vocational learners' interactional performance be measured **reliably** using standardized speech tasks?
2. Can interactional performance be measured **validly** across six tasks?
3. Can interactional performance be **operationalised in specific strategies** that can be measured reliably?

INTER-RATER RELIABILITY (ICC) GLOBAL CATEGORIES



INTER-RATER RELIABILITY (ICC) SPECIFIC CATEGORIES

KEY CARD APPLE CAKE HOTEL ROOM CINEMA BOARDGAME HEADPHONES

Compensation

Clarification

Difficult language

Misinterpretation

0,1

0,2

0,3

0,4

0,5

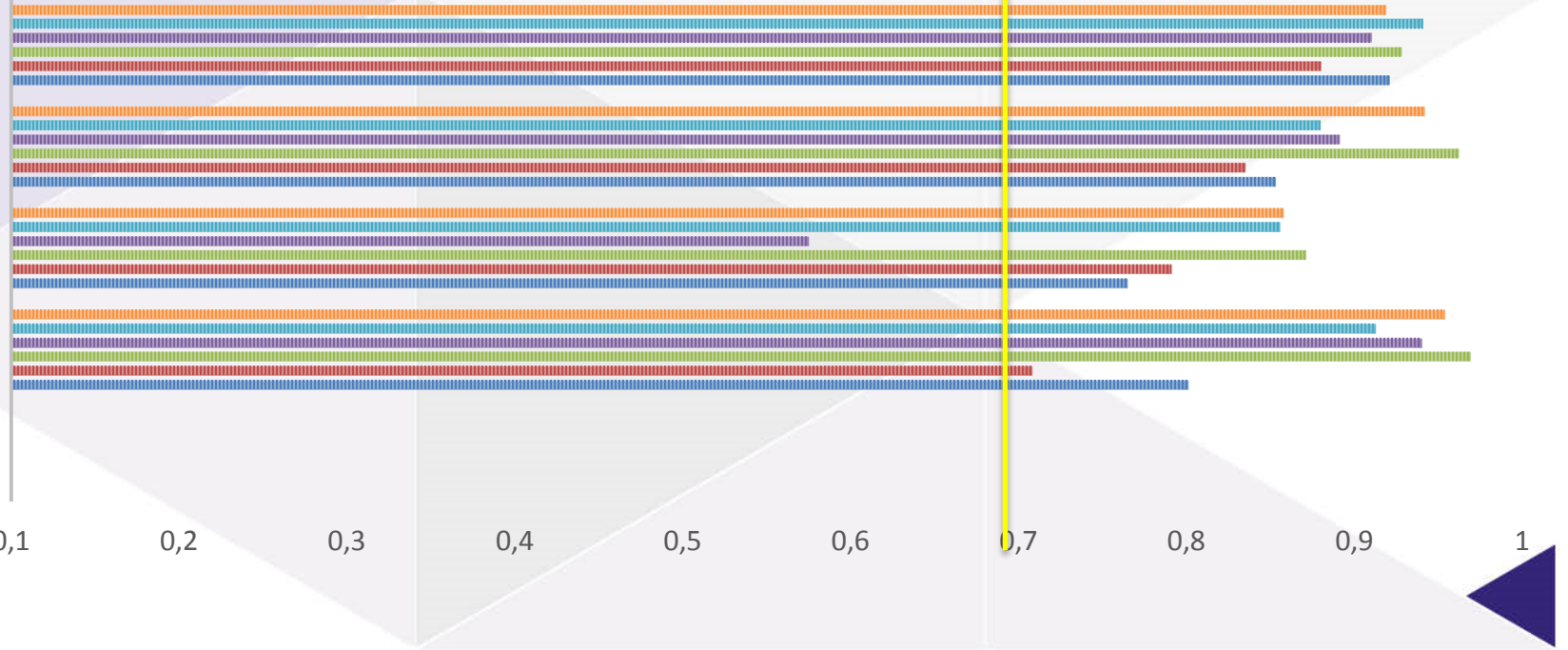
0,6

0,7

0,8

0,9

1



Correlation between scores for **Language Accuracy** across six tasks (Pearson's r)

	Apple Cake	Hotel Room	Cinema	Board game	Headphones
Key card	.904	.801	.824	.788	.807
Apple Cake	-	.816	.850	.788	.833
Hotel Room		-	.791	.868	.867
Cinema			-	.780	.736
Board game				-	.803

$p < 0.01$ (one-tailed)

$\alpha = .963$

Correlation between scores for **Interactional Ability** across six tasks (Pearson's r)

	Apple Cake	Hotel Room	Cinema	Board game	Headphones
Key card	.768	.726	.671	.672	.776
Apple Cake	-	.720	.842	.668	.853
Hotel Room		-	.676	.831	.799
Cinema			-	.738	.782
Board game				-	.799

$p < 0.01$ (one-tailed)

$\alpha = .945$

Correlation between global categories and specific categories (Pearson's r)

	Key Card		Apple Cake		Hotel Room		Cinema		Board games		Headphones	
	<i>Acc</i>	<i>Int</i>	<i>Acc</i>	<i>Int</i>	<i>Acc</i>	<i>Int</i>	<i>Acc</i>	<i>Int</i>	<i>Acc</i>	<i>Int</i>	<i>Acc</i>	<i>Int</i>
Compensation	.946	.835	.948	.902	.516	.388	.828	.879	.479	.300*	.874	.816
Clarification	.810	.822	.897	.892	.430	.289*	.847	.834	.504	.351*	.870	.721
Difficult language	.582	.648	.827	.808	.446	.481	.428	.440	.430	.550	.289*	.322*
Misinterpretation	.586	.640	.797	.745	.613	.481	.693	.771	.699	.614	.457	.423

* not significant
 $p < .001$ (one-tailed)

Reliability of subscales (Cronbach's alpha)

Linguistic Accuracy	.963
Interactional Ability	.945
Compensation	.901
Clarification	*
Meaning negotiation	.701
Misinterpretation	.832

CONCLUSIONS

1. Can pre-vocational learners' interactional performance be measured **reliably** using standardized speech tasks?
 - ✓ high inter-rater reliability
at global level
at specific levels (except 'difficult language' in task 3)
across three different task types

CONCLUSIONS

2. Can interactional performance be measured **validly** across tasks?

- ✓ Performance on *Linguistic Accuracy* and *Interactional Ability* correlates highly between tasks;
- ✓ *Linguistic Accuracy* and *Interactional Ability* can be measured reliably as two separate measures of L2 communication;
- ✓ Performance on specific strategies correlates positively with global performance....
- ✓ but the strength of the correlation differs per task.

RESEARCH QUESTIONS

3. Can interactional performance be **operationalised in substrategies** that can be measured reliably?

- ✓ Subscales for *Compensation*, *Responding to difficult language* and *Correcting misinterpretation* show high internal consistence.
- ✓ No meaningful result for subscale *Clarification* due to low *n*.

IMPLICATIONS FOR RESEARCH AND EDUCATION

- ☑ For assessment of general interactional ability, tasks suitable for application in both research- and educational settings;
- ☑ Measuring performance of specific strategies provides useful diagnostic information about individual candidates' interactional ability;

but

- ☒ Further research is needed to optimise the manner in which *Clarification* is elicited;
- ☒ Robust testing conditions are required.

LITERATURE (SELECTION)

- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (part 2). *Language Teaching*, 35(2), 79-113.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* Oxford University Press.
- Bygate, M. (2009). Teaching and testing speaking. In C. Doughty, & M. H. Long (Eds.), *Handbook of second and foreign language teaching* (pp. 412-440) New York: Blackwell.
- Bygate, M. (1987). *Speaking*. Oxford University Press.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In Soler, E.A. & Jorda, P.S. (eds) *Intercultural language use and language learning* (pp. 41-57) Springer.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Dörnyei, Z., & Scott, M. L. (1995). Communication strategies: An empirical analysis with retrospection. Paper presented at the *Deseret Language and Linguistic Society Symposium*, , 21. (1) pp. 137-150.
- Fulcher, G. (2003). *Testing second language speaking* Pearson Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book* Routledge.
- Luoma, S. (2004). *Assessing speaking* Cambridge University Press.
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8(2), 127-127-145.
- Taylor, L. (2003). The paired speaking test format: Recent studies. *Perspective*, 55(1), 70-76.
- Weir, C. J. (1993). *Understanding and developing language tests* Prentice-Hall.
- Weir, C. J. (2005). *Language testing and validation. an evidence-based approach* Palgrave MacMillan.

CORRESPONDENCE

Eline van Batenburg

e.van.batenburg@hva.nl

RESEARCH LANGUAGE TEACHING METHODOLOGY

Lesson series *Hotel Magnolia*

pre-vocational learners practice in a vocational context
(business and administration)

9 weeks

3 blocks (instruction, advice and sales)

Design

experimental

$N = 160$

randomly assigned within groups to:

- a **language-oriented** approach (=control group)
- a **meaning-oriented** approach
- a **combination** of both

Table 3

*Correlation between vocabulary size and final score
(Pearson's r)*

	<i>n</i>	<i>r</i>
Task 1 <i>Key Card</i>	33	.708
Task 2 <i>Apple Cake</i>	33	.821
Task 3 <i>Hotel Rooms</i>	27	.575
Task 4 <i>Cinema</i>	25	.610
Task 5 <i>Board Games</i>	29	.620
Task 6 <i>Headphones</i>	29	.655

Table 3

*Correlation between vocabulary size and explanation
(Pearson's r)*

	<i>n</i>	<i>r</i>
Task 1 <i>Key Card</i>	33	.823
Task 2 <i>Apple Cake</i>	33	.849
Task 3 <i>Hotel Rooms</i>	27	.452
Task 4 <i>Cinema</i>	25	.529
Task 5 <i>Board Games</i>	29	.445
Task 6 <i>Headphones</i>	29	.636

p < .001 (one-tailed)